2026 European Conference on Computing in Construction (EC^3)



Conference Proceedings

Corfu, Greece
12-15 July, 2026

Proceedings of 2026 European Conference on Computing in Construction (EC^3)

12–15 July 2026
Corfu, Greece

Edited by
Dr. Ranjith K. Soman, TU Delft (NL)

Dr. Ekaterina Petrova, TU Eindhoven (NL)

Dr. Jennifer Li, TU Delft (NL)

Dr. Daniel Hall, TU Delft (NL)

European Council for Computing in Construction
***Computing in Construction* series**

Liability: Any statements expressed in these materials are those of the individual authors and do not necessarily represent the views of the EC^3 organisers, who take no responsibility for any statement made herein. The EC^3 organisers make no representation or warranty of any kind, whether express or implied, concerning the accuracy, completeness, suitability, or utility of any information, apparatus, product, or process discussed in this publication, and assume no resulting liability. Anyone using this information assumes all liability arising from such use, including but not limited to infringement of any patent or patents.

Reproduction: No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Furthermore, the authors of each paper retain the right to republish that paper, in whole or part, in any publication of which they are an author or editor, and to make other personal use of the work. Any republication or personal use of the work must, however, explicitly identify prior publication in the EC^3 Proceedings, including page numbers.

Published by the European Council for Computing in Construction.

To cite this publication:
Soman, R. K., Petrova, E., Li, J., & Hall, D. (Eds.). (2026). *Proceedings of the 2026 European Conference on Computing in Construction, July 12–15, 2026, Corfu, Greece*. European Council for Computing in Construction. https://doi.org/10.35490/EC3.2026

Publisher: European Council for Computing in Construction
Series: Computing in Construction
ISSN: 2684-1150
ISBN: 978-90-834513-2-9

All Rights Reserved

© 2026 European Council for Computing in Construction

PREFACE

The 2026 European Conference on Computing in Construction (EC³) was held from July 12–15, 2026, at the Corfu Holiday Palace in Corfu, Greece. The conference brought together researchers, practitioners, and industry professionals from around the world to present their latest work and exchange ideas in the broad field of Computing in Construction. This book contains the papers submitted to the conference and accepted after a rigorous double-blind peer-review process.

The EC³ conference is a core activity of the European Council for Computing in Construction (EC³). The conference organisation is managed independently by a group of volunteers. The present conference marks the 8th edition of bringing together the international Computing in Construction community to discuss emerging research, innovation, and practice.

The EC³ 2026 proceedings present an overview of the conference programme, together with the names of the organisations and individuals who contributed to the technical programme. The peer-review process was conducted in two phases. First, 78 optional abstracts were submitted and reviewed by the respective Area Chairs. This was followed by a rigorous full-paper review process, in which each submission was evaluated by at least two reviewers from the Scientific Committee of international experts. Final acceptance decisions were made jointly by the corresponding Track Chairs and Programme Chairs. Following two rounds of peer review, 174 high-quality full papers, selected from a total of 257 full-paper submissions, were accepted for publication in these proceedings and presented at the conference.

The manuscripts were presented across 23 technical paper sessions, covering topics including:

- Artificial Intelligence in AEC
- Industrialisation, Automation and Robotics
- Decentralised Data, Cybersecurity and Trustworthy Automation
- Sustainability and Climate Action
- Human–Data Interaction
- Digital Twins and Asset Intelligence
- Education, Policy and Standardisation

Additionally, the conference included a poster session, where 24 posters were presented following a peer-reviewed selection process. The proceedings include these poster presentations as extended abstracts in a dedicated section.

Please note: all EC³ proceedings and conference materials are made openly available through the EC³ website (https://ec-3.org/publications/conference/?id=EC32026). Each conference paper is assigned a unique DOI and comprehensively meta-tagged to ensure easy discovery through commonly used academic search and indexing services.

One further track allowed participants to communicate embryonic and thesis-related research, with 21 young researchers participating in the Thesis-in-3 competition. Parallel presentation sessions were followed by a final competition session, providing students with an opportunity to present their research in a concise and engaging format.

Furthermore, the conference included four sessions dedicated to the technical committee work of the EC^3 and these included:

- Data Sensing and Analysis (DSA) Committee
- Modelling and Standards (M&S) Committee
- Education (EDU) Committee
- Human-Data Interaction (HDI) Committee

In addition to the above, a research paper writing workshop for early-career researchers was delivered by Prof. Mirosław J. Skibniewski. The workshop, Academic Writing for Early-Career Researchers, consisted of two complementary parts: Part I: How to Become an Effective Author of Highly Ranked, Peer-Reviewed Scholarly Journal Papers, and Part II: The Research Paper Clinic, where participants received practical guidance on developing high-quality scholarly manuscripts and strengthening their publication strategies.

The conference also featured six technical workshops, covering a diverse range of emerging research topics:

- A Guided Workflow for Integrating BIM and Geospatial Data in Digital Twins
- Aligning "Material" and "Product" Concepts Across Lifecycle Phases and Use Cases
- Temporal Sensor Fusion and State Reconstruction for Actionable Digital Twins
- How to Efficiently Generate RDF Graphs for Automation in AEC with IFC2LBD-Neo
- CoLAB 5.0: Human–Data Interaction for Ethical Electronic Performance Monitoring on Construction Sites
- Building Data Hubs for Circular Material Flows in Construction Logistics

In addition to the technical programme detailed above, the conference provided numerous opportunities for professional networking, collaboration, and fellowship through a variety of informal social events.

We would like to thank the EC³ scientific community, including both academic and industry members, for their contributions and continued support. We hope this conference provided opportunities to strengthen existing collaborations, establish new professional connections, inspire innovative research ideas, and enjoy the unique culture and natural beauty of Corfu. We look forward to welcoming you again at future EC³ conferences.

Dr. Ranjith K. Soman, TU Delft, EC^3 Programme Chair;
Dr. Ekaterina Petrova, TU Eindhoven, EC^3 Conference Chair, and
Dr. Jennifer Li, TU Delft, EC^3 Dissemination Chair.

We would like to thank our industry and academic partners for their generous contributions to the conference.

ORGANISERS:

European Council for Computing in Construction (EC³)

INDUSTRY PARTNERS:

Bentley Education

AUTODESK Research

ACADEMIC PARTNERS:

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

TU Delft Delft University of Technology

EUROPEAN COUNCIL FOR COMPUTING IN CONSTRUCTION

The **European Council for Computing in Construction (EC^3)** is a society of construction professionals, academics, researchers and national Professional Bodies, aspiring to become the leading European forum in the area of information technology in construction engineering and management.

Role of EC^3

The **European Council for Computing in Construction (EC^3)** advances professional knowledge and improves engineering practice in the built environment by fostering research, education and policy in current and emerging computing and information technologies.

EC^3 is founded on the following four pillars and corresponding Technical Committees:

- DSA: Data Sensing & Analysis
- M&S: Modelling & Standards
- HDI: Human Data Interaction
- EDU: Education

EC^3 interacts strongly with other Architecture, Engineering, Construction and Facility Management (AEC/FM) societies in related areas, strengthen the collaborations between academia and industry in topics related to **EC^3**'s mission, spearhead research on such topics, identify and promote effective ways to advance the state of knowledge and the level of education and practice in these topics, assist in the making of policy, and support existing and new related specialty conferences and publications.

Technical Committees

The role of all technical committees is to:

- Gather, maintain and disseminate information on the application of the committee's area to AEC/FM
- Organise and support joint research activities in the committee's area
- Organise and support conference sessions, workshops, and meetings in the committee's area
- Disseminate innovation in the committee's area through position papers, white papers, grand challenges reports and policy work throughout the European spectrum
- Keep the SPAs aware of developments in the committee's area

CONFERENCE COMMITTEE

Dr. Ekaterina Petrova, TU Eindhoven (NL) *(Conference Chair)*
Dr. Ranjith Soman, TU Delft (NL) *(Programme Chair)*
Dr. Jennifer Li, TU Delft (NL) *(Dissemination Chair)*
Dr. Daniel Hall, TU Delft (NL) *(Publications Chair)*

ACKNOWLEDGEMENTS

The organisers would like to thank the Track Chairs and the members of the Scientific Committee of the European Council for Computing in Construction (EC³) for their invaluable contributions to the paper review and selection process. We also extend our sincere appreciation to the members of the EC³ Technical Committees and workshop organisers for their dedication in delivering a rich technical programme that extended well beyond the paper presentation sessions.

We are particularly grateful to the student assistants Krista van Zandwijk, Stella Yongxue Tian, Irfan Čustović, Emilio Murillo Sierra, and Wahib Saif, whose professionalism, enthusiasm, and commitment greatly contributed to the smooth preparation and successful delivery of the conference. We also sincerely thank Sobia Bano for her invaluable support in maintaining the conference website and providing IT support throughout the preparation and delivery of the conference.

Track Chairs

Artificial Intelligence in AEC:
Prof. Jenn McArthur, Toronto Metropolitan University (CA)
Dr. Edlira Vakaj, Birmingham City University (UK)

Industrialisation, Automation and Robotics:
Prof. Daniel Hall, TU Delft (NL)
Dr. Jianpeng Cao, The University of Hong Kong (HK)

Decentralised Data, Cybersecurity and Trustworthy Automation:
Dr. Theo Dounas, Heriot-Watt University (UK)
Dr. Mantha Bharadwaj, University of Sharjah (UAE)

Sustainability and Climate Action:
Prof. Dimitrios Rovas, University College London (UK)
Dr. Kasimir Forth, ETH Zürich (CH)

Human–Data Interaction:
Dr. Diego Calvetti, University of Porto (PT)
Dr. Andrew Sonta, EPFL (CH)

Digital Twins and Asset Intelligence:
Dr. Sebastian Esser, Technical University of Munich (DE)
Dr. Karim Farghaly, University College London (UK)

Education, Policy and Standardisation:
Dr. Dragana Nikolić, University of Reading (UK)
Dr. Marzia Bolpagni, Mace / Northumbria University (UK)

Poster Series:
Dr. Judith Fauth, Eindhoven University of Technology (NL)
Dr. Pedro Meda, University of Porto (PT)

Thesis-in-3 Competition:
Dr. Nicola Moretti, University College London (UK)

Conference Scientific Committee

Dr. Akbarieh Arghavan	Eindhoven University of Technology (NL)
Ms. Al Jadaa Dana	Architectural Science (CA)
Dr. Ali Usman	University College Dublin (IE)
Dr. Allam Amr	Concordia University (CA)
Prof. Amor Robert	University of Auckland (NZ)
Dr. Asut Serdar	TU Delft (NL)
Dr. Awwad Rita	Lebanese American University (LB)
Dr. Babaei Ravandi Amirhossein	University of Toronto (CA)
Prof. Bartels Niels	Technische Hochschule Köln (DE)
Prof. Beach Thomas	Cardiff University (UK)
Prof. Beetz Jakob	RWTH Aachen University (DE)
Ms. Bendiek-Laranjo Ana	ETH Zurich (CH)
Dr. Blay Karen	Loughborough University (UK)
Dr. Bloch Tanya	Technion Israel Institute of Technology (IL)
Ms. Blum Clara Pauline	ETH Zurich (CH)
Dr. Bolpagni Marzia	Northumbria University (UK)
Prof. Borrmann André	Technical University of Munich (DE)
Dr. Byers Brandon S.	Michigan State University (US)
Prof. Calis Gülben	Ege University (TR)
Prof. Calvetti Diego	Pontificia Universidad Católica de Chile (CL)
Dr. Cao Jianpeng	University of Hong Kong (HK)
Prof. Carbonari Alessandro	Università Politecnica delle Marche (IT)
Mr. Carrara Andrea	Technical University of Munich (DE)
Dr. Castronovo Fadi	Okana Global (UK)
Prof. Cerovsek Tomo	University of Ljubljana (SI)
Prof. Chantzaras Christos	Frankfurt University of Applied Sciences (DE)
Dr. Chen Junjie	The University of Hong Kong (CN)
Mr. Chen Li	ETH Zurich (CH)
Dr. Chen Weiwei	University College London (UK)
Dr. Chung Sehwan	Samsung Global Research (KR)

Ms. Coffee Sharon	University College Dublin (IE)
Mr. Custovic Irfan	TU Delft (NL)
Prof. Dallasega Patrick	Free University of Bozen-Bolzano (IT)
Prof. de Wilde Pieter	Lunds University (SE)
Prof. Demian Peter	Loughborough University (UK)
Mr. DiMarco Giancarlo	Xian Jiaotong Liverpool University (CN)
Dr. Donaubauer Andreas	Technical University of Munich (DE)
Mr. Donkers Alex	Eindhoven University of Technology (NL)
Dr. Dounas Theodoros	Heriot-Watt University (UK)
Prof. Dr. Thomas Linner	Fakultät B/Building Lab (DE)
Ms. Durmus Dilan	Università Politecnica delle Marche (IT)
Mr. Elshaboury Hamada Ahmed	Politecnico di Milano (IT)
Ms. Elshani Diellza	University of Stuttgart (DE)
Mr. Elvarsson Arnor	ETH Zurich (CH)
Mr. Emig Julius	Fraunhofer Italia Research (IT)
Prof. Ergen Esin	Istanbul Technical University (TR)
Dr. Esser Sebastian	Technical University of Munich (DE)
Mr. Farcasiu Andrei	Carleton University (CA)
Dr. Farghaly Karim	University Collegel London (UK)
Dr. Fauth Judith	Technical University of Munich (DE)
Ms. Fitkau Isabelle	Technische Universität Berlin (DE)
Dr. Forth Kasimir	ETH Zurich (CH)
Ms. Freifrau von Rössing Lisa	Ruhr-Universität Bochum (DE)
Prof. Frisch Jérôme	RWTH Aachen University (DE)
Dr. Fuchs Stefan	Technical University of Munich (DE)
Prof. Garcia de Soto Borja	New York University Abu Dhabi (AE)
Dr. Gatto Chiara	Politecnico di Milano (IT)
Dr. Gil Daeyoung	Yonsei University (KR)
Prof. Graser Konrad	ZHAW (CH)
Mr. Grauer Dennis	Universität Rostock (DE)
Dr. Hagedorn Philipp	Ruhr-Universität Bochum (DE)
Dr. Hall Daniel	Delft University of Technology (NL)
Mr. Harder Benedict	Technical University of Munich (DE)
Ms. Heise Ina	TU Munich (DE)
Mr. Hellin Sylvain	Technical University of Munich (DE)
Prof. Hjelseth Eilif	Norwegian University of Science and Technology (NO)
Mr. Hoeng Simon K.	Ostbayerische Technische Hochschule Regensburg (DE)
Dr. Hofmeyer Hèrm	Eindhoven University of Technology (NL)
Dr. Hollberg Alexander	Chalmers University of Technology (SE)
Hu Yuqing	Penn State
Mr. Huang Cong	The Hong Kong University of Science and Technology (HK)
Dr. Iranmanesh Sima	Birmingham City University (UK)
Prof. Isaac Shabtai	Ben-Gurion University of the Negev (IL)
Dr. Jang Suhyung	Technical University of Munich (DE)
Dr. Jeong Jaemin	University of Toronto (CA)

Prof. Jeong Jaewook	Seoul National University of Science and Technology (KR)
Dr. K Soman Ranjith	Delft university of Technology (NL)
Ms. Kaltenegger Julia	Eindhoven University of Technology (NL)
Dr. Kandil Mohamed	Toronto Metropolitan University (CA)
Prof. Kassem Mohamad	Newcastle University (UK)
Dr. Kifokeris Dimosthenis	Chalmers University of Technology (SE)
Dr. Kim Joyce	University of Waterloo (CA)
Mr. Kim Yonghan	Technical University of Munich (DE)
Mr. Kirner Lukas	Eindhoven University of Technology (NL)
Prof. Kovacic Iva	TU Wien (AT)
Ms. Kozhevnikova Angelina	TU Delft (NL)
Dr. Krijnen Thomas	AECgeeks (NL)
Ms. Krischler Judith	Bauhaus-Universität Weimar (DE)
Dr. Lauble Svenja	Karlsruher Institute of Technology (DE)
Prof. Lee Ghang	Yonsei University (KR)
Dr. Li Jennifer	TU Delft (NL)
Prof. Li Xiao	The University of Hong Kong (HK)
Dr. Liu Hao	The University of Hong Kong (HK)
Mr. Lombardi Alessio	Buro Happold Engineering (UK)
Dr. Lombardi Davide	Xian jiaoton Liverpool University (CX)
Mr. Lozach Loïs	ETH Zurich (CH)
Dr. Maftei Laura	University of Reading (UK)
Mr. Maibaum Jonas	Ruhr Universität Bochum (DE)
Prof. Maile Tobias	Technische Hochschule Augsburg (DE)
Dr. Mantha Bharadwaj R. K.	University of Sharjah (AE)
Dr. Martani Claudio	ETH Zurich (CH)
Mr. Martin Jakob	Hochschule München University of Applied Sciences (DE)
Dr. Martinez Pablo	Northumbria University (UK)
Dr. McArthur Jenn	Toronto Metropolitan University (CA)
Mr. Meda Pedro	ICS - Instituto para a Construção Sustentável (PT)
Dr. Meschini Silvia	University of Turin (IT)
Dr. Mitropoulou Ioanna	ETH Zurich (CH)
Dr. Moradi Maryam	Toronto Metropolitan University (CA)
Dr. Moretti Nicola	University College London (UK)
Ms. Mousavi Rana	Toronto Metropolitan University (CA)
Dr. Msawil Mahir	Northumbria University (UK)
Dr. Munir Mustapha	University of Salford (UK)
Nam Yeonwoo	Seoul National University
Dr. Navaro-Auburtin Pierre	ETH Zurich (CH)
Prof. Ng Charmaine	Kyoto Institute of Technology (JP)
Dr. Nikolic Dragana	University of Reading (UK)
Mr. Nisbet Nicholas	Bartlett School of Sustainable Construction , University College London (UK)
Dr. Noichl Florian	Technical University of Munich (DE)
Dr. Nousias Stavros	TUM (DE)

Prof. O'Donnell James	University of Limerick (IE)
Prof. Obergrießer Mathias	OTH Regensburg (DE)
Dr. Pauwels Pieter	Eindhoven University of Technology (NL)
Dr. Petrova Ekaterina	Eindhoven University of Technology (NL)
Dr. Pfitzner Fabian	Technical University of Munich (DE)
Dr. Pisu Davide	Univeristy of Hertofordshire (UK)
Prof. Ploennigs Joern	Universität Rostock (DE)
Dr. Raghu Deepika	ETH Zurich (CH)
Dr. Rajabi Farivar	Toronto Metropolitan University (CA)
Mr. Rawat Alok	Tallinn University of Technology (EE)
Prof. Re Cecconi Fulvio	Politecnico di Milano (IT)
Dr. Rezazadeh Azar Ehsan	Toronto Metropolitan University (CA)
Dr. Richter Veronika Elisabeth	RWTH Aachen University (DE)
Dr. Rogage Kay	Northumbria University (UK)
Mr. Roh Hyunsung	Yonsei university (KR)
Prof. Rovas Dimitrios	UCL (UK)
Dr. Saadany Hadeel	Birmingham City University (UK)
Ms. Sadikoglu Emel	Gebze Technical University (TR)
Prof. Scherer Raimar	TU Dresden (DE)
Prof. Schneider-Marin Patricia	NTNU (NO)
Ms. Schwarzkopf Vanessa	ETH Zurich (CH)
Dr. Sonta Andrew	EPFL - Swiss Federal Technology Institute of Lausanne (CH)
Dr. Srećković Marijana	TU Wien (AT)
Dr. Sternal Maximilian	Technische Universität Berlin (DE)
Dr. Sujan Sujesh Francis	Norwegian University of Science and Technology (UK)
Dr. Sydora Christoph	University of Alberta (CA)
Prof. Tan Tan	The University of Hong Kong (HK)
Mr. Theissen Sebastian	BU Wuppertal (DE)
Prof. Vakaj Edlira	Birmingham City University (UK)
Mr. van Berlo Léon	Boundless BIM council (NL)
Ms. Vangelova Silvia	ETH Zurich (CH)
Mr. Vega Castellanos Giuseppe	Loughborough University (UK)
Dr. Vele Jiri	Czech Technical University of Prague / Scoolpt (CZ)
Dr. Verma Sneha	NEWCASTLE UNIVERSITY (UK)
Dr. Veselka Jakub	Czech Technical University in Prague (CZ)
Prof. Vilgertshofer Simon	HM Munich University of Applied Sciences (DE)
Dr. Walzer Alexander N.	ZHAW Zurich University of Applied Sciences (CH)
Dr. Weise Matthias	AEC3 Deutschland GmbH (DE)
Mr. Wolf Nepomuk	Technical University of Munich (DE)
Mr. Wu Jiabin	Technical University of Munich (DE)
Prof. Xenidis Yiannis	Aristotle University of Thessaloniki (GR)
Dr. Xie Xiang	Newcastle University (UK)
Dr. Yang Zhongze	The University of Hong Kong (HK)
Dr. Yao Dongchi	New York University Abu Dhabi (AE)

Dr. Ying Huaquan	Technion-Israel Institute of Technology (IL)
Prof. Yitmen Ibrahim	Jönköping University (Sweden) (SE)
Dr. Zheng Zihao	New York University Abu Dhabi (AE)
Dr. Zucker Gerhard	Austrian Institute of Technology (AT)

TABLE OF CONTENTS

Artificial Intelligence in AEC 1

Prescriptive Maintenance Of Sewer Systems Based On Large Language Models - Thamer Al-Zuriqat, Dhananjay Birmole, Kosmas Dragos, and Kay Smarsly 2

Fast-Asset: A Real Time Scan-To-3D Point Cloud BIM-Ready Asset Generation Method For Small MEP Elements - Urueña, Asier; Eskudero, Ibon; Sagasti, Diego; and Noyé, Sarah 10

Automated Dimension-Aware 2D-To-BIM Reconstruction Through Cross-Modal Text-Geometry Alignment - Xing Liang, Tomohiro Fukuda, and Nobuyoshi Yabuki 18

A BIM-Integrated Adaptive Surrogate Framework For Efficient Uncertainty Quantification In Building Energy Simulation - Arash B. Naamrudi, Reihaneh Aghamolaei 26

Ontology-Supported Bridge Component Design: Implementation Of Bridge Cap Design Rules - Tim Noack, Han Qian, Chongjie Kang and Steffen Marx 34

Exploring Adaptable Office Fit-Out Layouts Using Configurable Floor Plan Graphs And Large Language Models - Mikhael Johanes, Simona Kocour, Siyu Tang, and Catherine De Wolf 42

Few-Shot Building Element Detection With Vision Transformers: Improving Energy Retrofit Matching - Tom Slik, Raphae¨l Gueulet, Christos Lathourakis 50

AI-Assisted Bridge Conceptual Design: A Hybrid Knowledge- And Learning-Based Approach To Automated Span Planning - Han Qian, Tim Noack, Mengyan Peng, Marx Steffen and Chongjie Kang 58

Open-Source VLMs For Engineering Drawing Metadata: QWEN-VL Title-Block Detection And Information Extraction - Mengyan Peng, Angran Cui, Steffen Marx, and Chongjie Kang 66

Automated BIM LOD Inferencing: A Hybrid Approach Using Knowledge Graphs And Multimodal Large Language Models - Taeyoon Kim, Ghang Lee 74

Automated P&ID Compliance Checking Via LLM-Based Code Generation - Heejun Youn, Miyoung Uhm, Juho Han, Justin S. Lee, H. David Jeong, and Ghang Lee 82

Actor-Network Theory For Evaluating BIM-AI Implementation In Design Firms - Federica Crosato, Alireza Borhani, and Luciana Burdi 90

Fabrication-Aware Generative Design For Hybrid Prefabrication Using Stable Diffusion XL - Hang Zhang, Jianpeng Cao, Anton Savov, and Benjamin Dillenburger 98

SYN3D-LLM: Synthetic 3D Point Clouds Generated Through LLM Discussion - Nils Soeren Krause, Joern Ploennigs 105

Bibimbap: A Benchmark For Instructional BIM-Based Automated Programming - Clemens Kujat, Bharathi Kannan Nithyanantham, Stefan Telgmann, Tobias Sesterhenn Jörn Plönnigs, Stefan Lüdtke, Christian Bartelt 113

A Semantic Framework For Formalizing Stakeholder Information Requests In Construction - Timur Weilbach-Eyüboglu, Cornelius Preidel, and Simon Vilgertshofer 121

Exploring The Applicability Of Computer Vision Models For Automated Code Compliance Checking - Asaf Talor Shnay, Guy Austern, Tanya Bloch 129

Aligning Heterogeneous Views: Stakeholder Terminology And Multi-Level Mappings In The AEC Industry - Jakob Martin, Cornelius Preidel, Simon Vilgertshofer, André Borrmann 137

Parametric VPINN Framework For Rapid Design Exploration Of Thin Slabs - Ahmad Arafat, Dimosthenis Kifokeris, and Mohamad Omar Alzokani 145

Potential Of Large Language Models For Construction Quotation Workflows - Anja Brelih, Žiga Krvina, and Robert Klinc 153

Multidimensional Weighted Ranking Recommendation System For Circular Material Matchmaking - Noah Gerber, Ibrahim Al-Hazwani, and Pei-Yu Wu 160

Turning Construction Checklists Into Machine-Readable Inspection Schema - Till Weist, Leon Pfenning, and Prof. Sven Rogalski 168

Vision–Language Models In Construction: Assessing Prompt Robustness In Open-Vocabulary Object Detection - Federica Madaschi, Gelare Taherian, Marco Lorenzo Agostino Trani and Ehsan Rezazadeh Azar 176

Benchmarking Deep Learning Approaches For AEC Engineering Drawing Layout Detection And Information Extraction - Tianyang Huang, Alessio Lombardi, Ahmed Elnagar, Ahmed Zalouk, George Paul, Sepehr Najjarpour Arvid Sigurdsson, Khalid Ismail, Mohamed Ragab, Edlira Vakaj 184

IDS-Constrained Text-To-Graph Querying For IFC Models - Ivan Smirnov, Tanya Bloch, Fulvio Re Cecconi, Sebastiano Maltese, Lavinia Chiara Tagliabue, Silvia Meschini and Shabtai Isaac 192

Automated Schedule Enrichment From Daily Progress Reports: A Bi-Directional Neuro-Symbolic AI Approach - Sudheer Kumar Nanduri, Venkata Santosh Kumar Delhi 200

Assessing The Viability Of LLM Agents For Generating Reusable Compliance Checking Functions - Stefan Fuchs, Sylvain Hellin, and André Borrmann 208

A Multi-Agent RAG-Enabled Knowledge System For National-Scale Whole-Life Carbon Assessment Of Residential Retrofits - Sobia Bano, Usman Ali, Conor Shaw, Tobias Maile, Fatemeh Golpayegani and James O'Donnell 216

Do Buildings Really Need Comprehensive Sensor Networks? A Graph Theory-Based Framework For Physical Consistency Modeling - Yangzhe Chen, Qinglu Zeng, Nan Hu, Thomas Ohlson Timoudas, and Qian Wang 224

Semantic Mapping Of Request For Information (Rfi) To BIM Elements With Large Language Models - Limon Paul Joy, Farzaneh Zarei, and Mazdak Nik-Bakht 232

Explainable AI For Requirement Traceability In Construction Systems Engineering - Timon Kuijper, Ranjith K. Soman, Tong Wang, and Sander van Nederveen 240

Enabling Cross-Study Comparison: A Framework For Automated BIM-Qa Evaluation - Sylvain Hellin, Stefan Fuchs, Stavros Nousias, André Borrmann 248

Comparison Of Segmentation Models For Pothole Detection Using UAV Imagery - Varshney Muskan, Singh Aryan Raj, Garg Sahil 256

Seeing In The Dark: Low-Light Image Enhancement For Modular Construction - Yuanyang Qi, Jie Hong, Tingtian Li, and Xiao Li 262

From General Space Planner To Learning-Based Generation: A Systematic Mapping Of Representation, Objectives, And Control - Suhyung Jang and André Borrmann 270

A Neuro-Symbolic AI Pipeline For Generating SHACL-Based Digital Building Regulations - Alex Donkers and Ekaterina Petrova 278

An LLM-Driven Agent-Based Framework For Cross-Regional Construction Policy Evaluation In The Greater Bay Area - Wentao Zhu, Xiao Li, Juan Wang, and Mingtian Xiong 286

VLM-Augmented RAG Framework With Ontology-Constrained LLM Reasoning For Existing Building Material Documentation - Julia Kaltenegger, Stavros Nousias, Ekaterina Petrova, Andre´ Borrmann and Pieter Pauwels 294

Multi-Class Prediction Of Occupational Accident Types Using XGBoost In Pipeline Construction - Ahmet Eren Ünlü, Recep Özkan, and Onur Behzat Tokdemir 302

Physics-Constrained Reinforcement Learning For Individualized And Biomechanically Feasible Task-Motion Optimization - Kangrui Ren, Ali Golabchi, and Gaang Lee 309

Text-To-AR For Home Remodeling: An End-To-End Pipeline From Natural Language To AR-Compatible 3D Assets - Vamsi Sai Kalasapudi, Bharadwaj R. K. Mantha, Hassan Odeh and Alekhya Seelam 316

A Zero-Shot Approach To Decomposing Graphical Rules In Accessibility Regulations - Yonghan Kim, Ghang Lee, and André Borrmann 324

Shared Embedding Spaces For AEC Ontologies: An Empirical Study Of Discovery And Alignment - Tobias Heimig-Elschner, Jyrki Oraskari, and Jakob Beetz 332

A Semi-Automated Scan-To-BIM Approach To Fire Compartmentation In Existing Buildings - Jorge Magalhães, Rafael Cabral, Jéssica Reis, Rui Gavina, José Aidos Rocha, Vaidotas Sarka, Tatjana Vilutiene, Ricardo Santos and Diogo Ribeiro 1iBuilt, Instituto Superior de Engenharia do Porto, Politécnico do Porto, Porto, Portugal 340

3D Segmentation Agent Integrating 3D Gaussian Splatting And Vision-Language Models For As-Built Analysis - Boyu Wang, Jun Ma, and Borja García de Soto 346

Agentic AI – The Transformation Of Decision-Making Processes Through Artifical Intelligence - Charlotte-Maria Esch, Marijana Srećković2, and Josef Windsperger 353

Neuro-Symbolic Reinforcement Learning For Multi-Category Construction Relationship With Reasoning And Sequence Planning - Akarsth Kumar Singh and Shang Hsien Heish 361

From Building Permit Review Comments To Comment-Based Process Indicators: An Estbert-Based Classification Workflow - Pille-Riin Peet, Ergo Pikas, and Aime Ruus 369

Structured Innovation In Construction Value Engineering Through LLM-Assisted TRIZ Reasoning - Fuhao Zu and Linghui Xie 377

Industrialization, Automation & Robotics 384

Forecasting Machine Working Times In Off-Site Construction: A Machine Learning Benchmark For Supporting Production Planning - Andrea Sbaragli, Xiang Xie, John Hedley and Mohamad Kassem 385

Emerging Business-Model Archetypes Of A Smart Mobile Factory For Construction 4.0 - Alexander N. Walzer, Irfan Čustović2, Konrad Graser and Daniel M. Hall 393

Experimenting With Digital Twins: A Lego®-Based Modular Testbed For Systematic Evaluation - Irfan Čustović 1, Emilio Murillo Sierra, Ranjith K. Soman, Daniel M. Hall 401

Enriching Version Increments With Engineering Intent For Enhanced BIM Model Checking - Sebastian Esser, Jiabin Wu, Suhyung Jang, Lena Fahrner, Robert Amor, and André Borrmann 409

A Computational DEA-Based Decision Support Framework For Green Supply Chain Management In Construction - Erawati Ravindra Kadam, and Ann Francis 417

Defining Demonstrator Projects For Digital Fabrication (DFAB) Adoption In Architecture, Engineering And Construction (AEC) - Jennifer Li, Konrad Graser, Alexander N. Walzer, and Daniel M. Hall 425

A Two-Stage MEP Installation Constructability Optimization System Integrating Dsm-Based Module Division And 4D Simulation - Ang Li, Zhengyi Chen, and Jack C.P. Cheng 432

Emerging Opportunities For Integrating Digital Product Passport Into Lean Construction Workflows: A Conceptual Perspective - Yanh Ribeiro, Pedro Mêda, and Jorge Teixeira 440

Automatic Mountain Tunnel Invert Excavation With Digital Twin Management In Non-Gnss Environments - Nobuyoshi Yabuki, Hitoshi Ishida, Sadatoshi Oomori, Yoichi Moriya, and Shinji Fujita 448

Proactive Safety Management In Earthwork Operations Using IoT-Driven Real-Time Risk Analysis - A N Ledin Ullah, Tulika Choudhury Puja, Jhonattan G. Martinez, Lorenzo Loyola, Istiaque Hossain and Freddy Perez 455

Perceived Benefits Of 5D BIM In Construction Cost Management: A Europe-Wide Survey - Enes Go¨nen and Semra C¸ omu 463

Data-Driven Decision Support For Construction Method Selection In Linear Infrastructure Projects: A Systematic Review - Maik Görgner, Simon Hoeng, Marc Schmailzl, Thomas Linner, Thomas Wolff, Mathias Obergrießer 471

Ontology-Based Cost Management In Public Infrastructure Projects: A Case Study On Transferability And Applicability - Anna Nast, Jacopo Cassandro, Claudio Mirarchi, and Jürgen Melzner 479

Graph-Based Generative Scheduling In Construction Project Management - Jianpeng Cao, Hisham Said 487

Integrated Many-Objective Optimization For Decision-Making And Dynamic Project Control - Aman Shukla, and Sahil Garg 495

Principles For The Industrialisation Of The Infrastructure Renovation And Replacement Assignment - Irfan Pottachola, Alexander Bletsis, and Trijntje Verschuuren Netherlands Organisation for Applied Scientific Research (TNO), Delft, The Netherlands 502

Model To Press: Digital Dataflow For The Design, Rationalization, And Manufacture Of Free-Form Glulam Beams - Tom Svilans, Edgars Rudzītis, Jānis Iejavs, Edvīns Grants, and Kristaps Zīverts 510

Teleoperation Interfaces For Compact Demolition Robots: Results Of A Pilot Study With Construction Workers - Patrick Borges Rodrigues, Zihao Wang, Yiyang Fang, Gale Lucas, Shawn Roll, Burcin Becerik-Gerber, and Lucio Soibelman 518

Automized Planning & Data Exchange For Prefabricated Façade Renovation Processes - Julius Emig, Orjola Braholli, Dietmar Siegele and Dominik T. Matt 526

Evaluation Of Mobile Platforms For Automation Of Metro Tunnel Inspections - Timur Kuzu, Katharina Klemt-Albert 534

Optimizing Linear Construction Processes: A Simulation-Based Approach Towards AI-Supported Decision Making - Anna-Maria Mehringer, Simon K. Hoeng, Jonas Wiederer, and Mathias Obergrießer 542

An Integrated Sensor Data And Machine Learning System For Work Progress Monitoring In Industrialized Construction - Eitan Ballivián, Tomás Guzmán, Caroline Araújo, Pablo Martinez, Tito Arévalo, Beda Barkokebas 550

Designing Human–Robot Collaboration As A Joint Cognitive System: A Sociotechnical Framework For Meaningful Work - Angelina Kozhevnikova and Daniel M. Hall 558

A Three-Layer Taxonomy And Predictive Digital Twin For Role-Aware Human–Robot Collaboration In 3D Concrete Printing - Ibrahim Chahbaz, Reina Chmayssany, Jiany Samara, Noha Ramadan, and Hiam Khoury 566

Decentralized Data, Cybersecurity & Trustworthy Automation 574

Supporting Construction Cybersecurity Decision-Making Through A Crowdsourcing Platform - Muammer Semih Sonkor, and Borja García de Soto 575

Trust Layers For Automated Bridge Management Using Blockchain And Smart Contracts - Leonardo Zunino, Valentina Gatteschi, Marco Domaneschi, Joan Ramon Casas, and Valentina Villa 583

Usage Policies For Fragmented Digital Product Passport Data Retrieval In Aeco Data Spaces - Krista van Zandwijk, Ekaterina Petrova, Rik Eshuis and Pieter Pauwels 590

Social Media Public Response To Critical Infrastructure Cyber Disruptions: A Reddit Based Multi-Dimensional Sentiment Analysis - Bharadwaj R. K. Mantha, Balaji Kesavan, and Johan Ninan 598

Defining A Blockchain-Verified Specimen Passport For 3dcp Standardisation - Giancarlo Di Marco, Davide Lombardi, and Theodoros Dounas 606

Governing Construction Material Reuse As A Blockchain Commons: Ostrom's Design Principles And Eu Digital Product Passports - Theodoros Dounas, Michele Victoria, Magdalena Blasiuzak, Zhe Jiang, and Hico McDonald 614

Blockchain-Enabled Real Estate Ownership Transfer In Belgium: A Hybrid Institutional Framework - Aishuak Kerimzhanova, Theodoros Dounas 621

Sustainability & Climate Action 629

Automated Generation Of Disassembly Graphs Based On Detailed 2D Section Drawings - Kasimir Forth, Mikhael Johanes, Elvire de Clermont Tonnerre, and Catherine De Wolf 630

Towards Standardised Environmental Product Data Integration In IFC - Martina Jakubowska, and Kristoffer Negendahl 638

Graph-Based Instance Traceability For Building Product Reuse - Silvia Vangelova, Clara P. Blum, Kasimir Forth, Jens J. Hunhevicz, Pritha Jacob and Catherine De Wolf 646

Automated Quantity Takeoff And Life-Cycle Costing In Construction Projects Using An openBIM–IFC Approach - Farzad Jalaei, Arash Hosseini Gourabpasi, Abdulwahed Fazeli, Ahmad Jrade 654

Cityretrofit: Development Of A District-Scale Digital Twin - J. J. McArthur, Saeed Ranjbar, Dana Al Jadaa, Masood Shamsaiee, Oriol Gavaldà-Torrellas, Hossein Bagherzadeh, Sanam Dabirian, Jackson Adebisi, Ursula Eicker 662

Breakpoints In Data Continuity Across Construction Reuse Projects - Clara P. Blum, Silvia Vangelova, Jens J. Hunhevicz, Kasimir Forth and Catherine De Wolf 670

Scenario-Based Material-Flow And Circularity Assessment Of Deep Renovation Projects - Mohamed T. Elnabwy, Omar Doukari, and Pablo Martinez 678

Towards Hybrid Digital Twin For Performance-Based Control Evaluation In Hydrogen Facilities - José L. Hernández, Ismael Lozano, Víctor I. Serna, Luis V. Jimeno, Jesús Samaniego and Francisco J. Miguel 686

A Multi-Model Framework Integrating Climate Projections And System Degradation For Climate-Aware Retrofit Planning - Avinash Myakal, Ranjith K. Soman, and Igor Nikolic 694

Heating Demand As An Indicator Of Renovation Depth In Soviet-Era Apartment Buildings - Fariha Harun, Ergo Pikas, Jaanus Hallik 701

Integrating Indicator Review And Social Network Analysis For Sustainable Public Infrastructure - Anagha Divakaran, Ann Francis 709

From Inspection To Digital Passport: Tracing The Provenance Of Testing Data Of Concrete Components For Circular Reuse - Lukas Guntermann, Nina Krautgartner, Hendrik Zielonka, Philipp Hagedorn, Peter Mark, and Markus König 717

Interoperability For A Circular Built Environment: Towards A Roadmap - Mustafa Selçuk Çıdık, Irene Josa, Kell Jones, Nicola Moretti and Karim Farghaly 725

Decision Support Framework For Integrating Construction Consolidation Centre Into Sustainable Waste Management - Jintian Xu, Thijs Nooijen, Qi Han and Dujuan Yang 733

Data Driven Decision Support Framework For Product Level Reuse Prioritisation And Barrier Diagnosis - Qi Han, Jintian Xu, Dujuan Yang, T. van Tongeren 741

Utilize District Heating Smart Meter Data And Building Information To Compare The Heat Load With The Design Guidelines - Martin Frandsen, Qinjiang Yang, Michal Zbigniew Pomianowski, Morten Karstoft Rasmussen, and Michele Tunzi 749

Edge Computing For Federated Learning–Enabled Energy Management And Monitoring Systems - Mustapha Habib, Youssef Elomari, and Qian Wang 757

A Data-Driven Koopman-MPC For Multi-Actuator AHU Control - Youssef Elomari, Mustapha Habib, and Qian Wang 765

Causality-Informed Time-Series Foundation Models For Building Performance Prediction - Dennis Grauer and Joern Ploennigs 770

Adaptive Life Cycle Assessment As An Evolving Environmental Knowledge State For Digital Building Logbooks - Kilian Speiser and Jochen Teizer 778

No Carbon Left Behind: Towards Traceable Biogenic Carbon In Construction Value Chains - Edwin Zea Escamilla, Alexander N. Walzer, and Markus Knecht 786

Household-Level Energy Poverty Assessment In Ireland: A Comparative Analysis Of Indicator-Based Approaches - Afshin Saeedian, Lorna Fitzsimons, James McGrath, Arash B. Naamrudi, Reihaneh Aghamolaei 793

Quantifying The Role Of Asset Maintenance For Resilient Infrastructure - Shyamratna Yadav, Chirag Kothari, and Anushka Singh Rathore 800

Automated Building Classification For Scalable Urban Material Stock Assessment - Elisa Iliste, Ingrid Viskus, Ergo Pikas and Targo Kalamees 808

Digital Climate Indicators For Resilient Design: Examining Completeness Across Carbon And Climate Risk Analytics - Silvia Meschini, Paola Gasbarri, Lavinia Chiara Tagliabue, Stefano Rinaldi, Viviana Vaccaro, and Hira Ovais 816

Beyond Traffic Performance: Assessing Transport Network Resilience Through Access And Equity - Giuseppe Vega, Peter Demian, Georgios M. Hadjidemetriou 824

Automatic Roof Type Classification From Cityjson For Rule-Based Roof Structure's Material Quantity Assessment - Dušan Isailović1, Aleksandra Parezanović1, Zorana Petojević1,2, Nenad Brodić1 830

Exploring Flood Vulnerability Patterns And Residential Flood Impacts In Urban Environment - Hyun Jeong Koo and Vassiliki Demetracopoulou 838

An Ontology-Driven Framework For Intelligent Embodied Carbon Decision-Support In Metro Station Design - Linghui Xie, and Xueqing Zhang 844

Human Data Interaction **852**

Augmented Reality And Computer Vision To Support Scheduling And Monitoring In Construction: A MEP Case Study - Patrick Dallasega, Marco Lanzone, Andrea Revolti, Alessandro Sottovia, and Oswald Lanz 853

Proposal Of Human-Sensored Architecture Framework By Introducing Mixed Reality - Kyung-Tae Lee 861

Chat With Pre-Demolition Audits: Knowledge-Graph Based Information Retrieval - Ioanna Mitropoulou, Beril Önalan, and Catherine De Wolf 869

Collaborative Facade Design With LLM-Based Multi-Agent Systems Using Reclaimed Building Components - Pritha Jacob, Ioanna Mitropoulou, Pei-Yu Wu and Catherine De Wolf 877

Serious Construction Game In Virtual Reality: Embedding High-Resolution Telemetry For Lean Performance Assessment - Kilian Speiser and Jochen Teizer 885

Developing A Linked-Data-Based Framework For Regulation Compliance Checking In Architectural Design - Yutian Yi, Davide Lombardi, Zhelun Zhuand Theodoros Dounas 893

Between Architectural Designers And Developers: Empirical Design Research For Computational Design Support - Elien Vissers-Similon and Theodoros Dounas 901

Supporting Construction Site H&s Inspections With Computer Vision: A Bpmn Process Formalisation - Federica Madaschi, Alessandro Bruttini, and Marco Lorenzo Trani 908

A Decoupled Architecture For Digitally Actionable And Traceable Regulatory Workflows In The AEC Sector - Hans Kristjan Aljas, Ergo Pikas, and Martin Thalfeldt 916

Operationalising Health And Safety Regulations Through Design Science Research - Nicholas Nisbet and Zijing Zhang 924

A Semi-Automated BIM-To-VR Workflow For Teaching: Robust Model Review In Cave And Hmd Environments - Hu Hanwen, Emara Merna, and Menzel Karsten 931

Development Of An IFC-Based Converter For BIM Integration Into Augmented Reality Applications For Building Construction Trades - Christian Annen, Hendrik Morgenstern, and Katharina Klemt-Albert 938

BIM–AR Integration In Practice: Opportunities, Challenges, And Trade-Offs From A Live Construction Case Study - Avar Almukhtar, Muhammad Shahzad, Marri Deepak Kumar, Joseph H.M Tah 946

Space-Driven Knowledge: A Topological Framework For Semantic Inference In Buildings - Isabelle Fitkau, Max Hammer, and Maximilian Sternal 954

The Potential Of Digital Technology To Engage Children With Hospital Experience And Design - Laura Maftei 962

July 1215, 2026 Low-Code Toolchain Development To Accelerate Data-Driven Energy Modelling - Mariam Ali, Cathal Hoare, Gabriel Oyeyemi, Divyanshu Sood, Tiziana Margaria and James O'Donnell 968

Digital Twins & Asset Intelligence 976

Asset-Centric Knowledge Sharing For Permitting And Compliance With Digital Technologies - Alhadi Alsaffar, Professor Thomas Beach, and Professor Yacine Rezgui 977

Bimstruct3d: A Fully Automated Hybrid Learning Scan-To-BIM Pipeline With Integrated Topology Refinement - Mahdi Chamseddine, Fabian Kaufmann, Marius Schellen, Christian Glock, Didier Stricker, and Jason Rambach 985

Information Requirements For Digital Twin Integration In Firefighting: A Survey Study - Iryna Osadcha, Caroline McMullan, Andrius Jurelionis, Robert Howell, and Muhammad Intizar Ali 993

Towards A Semantic Ontology For Occupant-Centric Building Control - Saif Alramahi, Mohammad Mayouf, Elham Del Zendeh, and Khalid Ismail 1001

Ontology-Based Semantic Management Of Bridge Drawings - Xiaoli Song, Mengyan Peng, Steffen Marx, and Chongjie Kang 1009

Enabeling Comprehensive Querying Of Environmental Context And Road Infrastructure Data Using A Graph-Based Approach - Ina Heise, André Borrmann 1017

Greenbimxels: A Data Space Unlocking Environmental Value For Public Facilities In The Construction Sector - Gordo-Gregorio, Paula; Coloma Picó, Eloi; Martínez Ibañez, Manel; Macarulla Martí, Marcel; Angulo, Cecilio 1025

Deep Learning–Based Point Cloud Segmentation And Relationship Graph Construction For Hydropower Dam Digital Twins - Alwyn Mathew, Soheila Kookalani, and Ioannis Brilakis 1033

Data Quality Control For openBIM-Based Digital Twins In Retrofit: A Case Study On IFC Validation And Information Requirements - Sarah Mahbod, Atefeh Hesarbani, Erik Poirier, Farzad Jalaei, Arash Hosseini, and Mazdak Nik-Bakht 1041

Interoperability Imperative: A Built Environment Maturity-Barrier Analysis For Digital Twin Convergence With BIM And IoT - Ubongabasi Inyang, Bamidele Adebisi, and Weizhuo Wang 1049

Open Standards For Data Integration And Visualization Across The Project Lifecycle: Challenges And Opportunities Of IFC5 - Jacopo Cassandro, Alessandro Bruttini, Andrea Bongini, Lukas Guntermann, and Markus König 1057

Symbol Recognition In Piping And Instrumentation Diagrams Via Mask-Aware Template Matching With Spatial Similarity - Juho Han, Miyoung Uhm, Heejun Youn, H. David Jeong, and Ghang Lee 1065

A Comparative Analysis Of IoT Data Pipelines For Real-Time Mixed Reality Visualisation In Unity-Based Digital Twins - Muhammad Shahzad, Joseph H. M. Tah, Avar Almukhtar, and Muhammad Younas 1073

Presenting The LOINxml Schema: Discussing Design Decisions And Intended Application - Martina Mellenthin Filardo, Paul Debus, Sven Zentgraf, Štefan Jaud, Christian Clemen, Ghang Lee, Jakob Beetz and André Borrmann 1082

Structuring Knowledge For AI-Driven Schedule Management In Digital Construction Twins - Tomas Lay Herrera, Matthieu Gillet, Mehrzad Shahinmoghadam and Simon Bee 1090

Connecting Dots: Harnessing Diverse Perspectives For The Integration Of Digital Building Permits And Logbooks - Pedro Mêda, Judith Fauth 1098

Project Management With IFC: Toward An Open-BIM Framework - Riccardo Padoan, Lidia Pinti, Jan Karlshøj 1106

An Extensible GraphQL API For Fine-Grained Access To Building Information Models - Nepomuk Wolf, Sebastian Esser, and André Borrmann 1114

An Information Management Framework For Multi-Use Case Building And District Digital Twins - Joao Patacas, Edgar Segovia, Xiang Xie, Phil James, and Mohamad Kassem 1122

An IFC Framework For Information And Document Management In Construction Processes - Beatrice Maria Toldo, Carlo Zanchetta 1130

BIM-Based Design Change Framework For Assessing MEP System Stabilisation - Mawara Khan, Muhammad Salman Javaid, Eelon Lappalainen, Hisham Abou Ibrahim, Olli Seppänen 1138

Digital Twin Value Creation As A Dynamic System: A Causal Loop Diagram Approach - J Omar Abbas, Nicola Moretti, and Tim Broyd 1146

A Monitoring-Oriented Digital Twin Pipeline For Data-Driven Prediction Of Residential Prosumer Energy Profiles - Amin Rahmani, Mahsa Moradian, Sonia Lupica Spagnolo, and J.J. McArthur 1154

A Decentralized Digital Twinning Approach For Assessing Civil Engineering Structures - Kosmas Dragos, Thamer Al-Zuriqat, and Kay Smarsly 1162

Deep Learning And Reality Capture For Automated Railway Bridge Inspection With BIM Enrichment - Rafael Cabral, Ricardo Santos, Diogo Ribeiro 1170

Automated Generation Of Semantic IFC Models From Legacy Sewer Network Data - Peter Gölzhäuser, Raquel Valles Gomez, Sabine Hartmann, Hendrik Morgenstern, and Katharina Klemt-Albert 1178

Topology-Driven Semantic Integration Of Independent BIM Models For Renovation Analysis - Jacopo Cassandro and Claudio Mirarchi 1186

Latency-Aware Construction Activity Recognition Using Kinematic And Physiological Sensors - Umutcan Kılıç and Semra Çomu 1194

Integrating Time, Cost And Resources In BIM: A Process-Oriented Framework - Davide Avogaro and Carlo Zanchetta 1202

A Semantic Data Model For Interoperable Exchange Of Test Trench Verification Data - Faith Tangara, Léon olde Scholtenhuis, and Ramon ter Huurne 1210

Comparative Analysis Of UWB And IMU For Construction Safety Monitoring - Serdar Bayraktar and Semra Çomu 1217

openBIM-Based Green Public Procurement Code Checking: Rule Mapping, Information Delivery Specifications And Gherkin Tests - Carlo Zanchetta, Valeria Galleran, Thomas Krijnen and Carlo Pavan 1223

Not All Links Are Equal: Link Expressiveness And Discourse Scope In BIM-GIS Integration - Judith Krischler, Sebastian Schilling, Maximilian Sternal, Jakob Taraben and Christian Clemen 1231

Towards A Multi-Domain Semantic Architecture For Graph-Based Orchestration Of Decentralised Additive Manufacturing - Lukas Kirner, Alex Donkers, and Ekaterina Petrova 1239

A Multimodal Hybrid Retrieval Framework For Digital Twins In Civil And Structural Engineering - Marcel Aubart Jyrki Oraskari, Jakob Beetz and Jörg Blankenbach 1247

Toward A BIM-Centric Digital Twin For Sustainable Museum Operation: The MACA Living Lab Case Study - Silvia Meschini, Viviana Vaccaro, Lavinia Chiara Tagliabue, Silvana Dalmazzone, Ferruccio Damiani, Gianluca Torta Stefano Rinaldi 1255

Towards Semantic Voxel-Based Reasoning For Robotic Fleet Management In Construction - Chu Han Wu, Pranav Shevkar, and Sigrid Brell-Cockan 1264

Structural Health Monitoring Of Bridges: Integrating openBIM, GIS, And IoT In The Breganze Bridge Case Study - Idriss Tchaheu Tchaheu,Francesca da Porto,Andrea Gennaro, Giorgia Marcellino, Carlo Zanchetta 1272

A Utility-Aware Knowledge Graph To Support Multi-Scale Residential Energy Modelling - Cathal Hoare, Sharon Coffee, Divyanshu Sood, Tiziana Margaria and James O'Donnell 1280

Education, Policy & Standardisation **1288**

Towards An Industry-Academia Collaboration Framework For The Development Of Digital Geospatial Surveying Micro-Credentials - Brian Graham, Bryan Hallissey, and Shane Murtagh 1289

Object-Level Multi-Dimensional Analysis Of Level Of Information Requirements Across Project Lifecycle - Pablo Martinez and Marzia Bolpagni 1297

From Analysis To Design: Teaching Algorithm-Aided Design In Structural Engineering - Marcin Luczkowski, Anders Rønnquist 1305

Macro Analysis Of Computing In Construction Education In Europe - Pablo Martinez, Mark Baldwin, Niels Bartels, Karen Blay, Sebastian Esser, and Tobias Maile 1313

Towards A Streamlined Digital Permitting Process: A Swedish Case Study - Nischal Bhattarai, Annika Moscati, Ulf Seigerroth, Alexandros Tzanetos, Erik Bergström 1321

IFC5 For Collaborative And Traceable Circularity In Construction - Tong Wang and Thomas Krijnen 1329

Proposal For A Modern Foundation For A Data-Driven Built Environment - Léon van Berlo, Thomas Krijnen, Tom van Diggelen, Artur Tomczak, Evandro Alfieri, David de Koning, Greg Schleusner, Helga Tauscher and Dennis Shelden 1337

Teaching BIM As A Computational, Queryable, And Analyzable Model: A Project-Based Framework Integrating APIs, Programming, And Early Performance Reasoning - Tobias Maile, Burcu Güldür Erkal 1345

Narratives Of Digital Twins: How Framings Script Governance - Johan Ninan and Ranjith K. Soman 1353

Building Energy Simulation Module: The Design, Implementation And Evaluation Of A Dual-Assessment Masters Module - Divyanshu Sood, Donal Finn, James OD´ onnell 1360

Extended Abstracts: Poster Presentations 1368

Analysing Green And Digital Skills In The Twin Transition - Farina Koller, Niels Bartels, and Claudia Ziller 1369

Blockchain-Enabled Digital Twins For Enhanced Lifecycle Management In Building Handover - Yutian Yi, Davide Lombardi and Theodoros Dounas 1371

From Capture To Co-Construction: A Scot- Informed Analysis Of Digital Twin Practices In Cemetery Scanning - Ekpenyong Ekpo, Nikolic Dragana, and Green Martin 1373

Parsing Construction Standards Into Machine-Processable Format - Mohamed Helmy Ammar and Merna Emara 1375

A Workflow For FEM Generation From BIM Models Based On Geometric Verification - Villem Vaktskjold, Marcin Luczkowski, and Lars Olav Toppe 1377

A BIM-Based Diagnostic Digital Twin For Preventive Seismic Retrofit Of Occupied School Buildings - Lavinia Chiara Tagliabue, Alessandro Zichi, Silvia Meschini, Stefano Rinaldi, and Hira Ovais 1379

A Large Language Model-Based Multi-Agent System For Enhanced Information Retrieval In Civil Engineering - Sebastian Kunz, Benedikt Justin Kandler, and Uwe Rüppel 1381

From Cultural Data To Policy Narrative: A Semantic Framework For Urban Transformation - Hira Ovais, Lavinia Chiara Tagliabue, Silvia Meschini, and Vincenzo Lombardo 1383

Predictive Framework For Dominant Delay Risks In Metallurgical Projects - Shivam Pandey, Sahil Garg 1385

A Systematic Review Of Distributed Reinforcement Learning For Building-To-District Energy Networks - Katharina Steib, Thomas Ohlson Timoudas, and Qian Wang 1387

Towards Automated Building Maintenance And Renovation: A Multi-Agent LLM-Based Approach - Jiaxin Ling and Weiwei Chen 1389

An Architectural Workflow Database For Generating Evidence-Based Healthcare Facilities' Design Via Transfer Learning - Julia Zieleniewska and Dimosthenis Kifokeris 1391

RGB-D Based Work Area Recognition Model For Fireproofing Spray Robot - Sangmin Lee, Kangmin Bae, Minseung Cha, Sebeen Yoon, and Taehoon Kim 1393

Human-Data Interaction (HDI) Technical Committee - Diego Calvetti, Mahir Msawil, Xiang Xie, Jennifer Li, Dimosthenis Kifokeris, Gulben Calis, Peter Demian, Nick Nisbet, Dragana Nikolic, Kay Rogage, May Shayboun, Sujesh Sujan, Alexandre Mugnai, Ranjith K. Soman, Mustapha Munir. 1395

Generating IFC-Based Thermal Models For Industrial Facilities - Adolfo Sánchez, Eduardo Adán, and Antonio Adán 1397

A Self-Verification Framework Toward Reliable Text-To-BIM Generation - Tobias Sesterhenn, Bharathi Kannan Nithyanantham, Stefan Lüdtke, and Christian Bartelt 1399

Integrating Parametric Modeling, Computer Vision, And Mixed Reality: Real-Time Quality Control In Craftsmanship - Simon K. Hoeng, Christoph J. Holzfurtner, Anne-Sophie Saffert, Florian Weininger, Thomas Linner and Mathias Obergriesser 1401

Categorizing Recurring Tasks For Faster And Consistent Embodied Carbon Assessments - Eleni Alexandraki, Giulia Scagliotti, Calvin Kam, Joy Li, and Martin Fischer 1403

Thickness Simulation With Kinematic Constraints For Robotic Fireproofing Spray - Kangmin Bae, Sangmin Lee, Minseung Cha, Sebeen Yoon and Taehoon Kim 1406

RGB-Only Work-Area Recognition For Fireproofing Spray Robot - Minseung Cha, Kangmin Bae, Sangmin Lee, Sebeen Yoon, and Taehoon Kim 1408

Anomaly Detection In Indoor Environmental Sensor Data Based On A Digital Twin - Arezoo Fathollahizenooz, Cornelius Preidel, and Simon Vilgertshofer 1410

Bimchained: A Blockchain Supported Digital Framework For Real Time Construction Site Monitoring - Eleonora Vendetti, Giuseppe Piras 1412

A Cross-Domain Framework For Explainable Structural Inspection - Vedant Girish Dalvi, Cornelius Preidel, and Simon Vilgertshofer 1414

Evaluating Local Small Language Models For Architectural Meeting Transcription Summarization And Action Extraction - Sean Bayly, Sneha Verma, Mohamad Kassem and Huizhi Liang 1416

Artificial Intelligence in AEC

Sub-Tracks:

Lifecycle Value & Predictive Maintenance

Augmented Design Intelligence

Scan-to-BIM & Reality Capture

NLP & Large Language Models

AI-Driven Project Planning & Risk Analytics

AI-Enhanced BIM & Automated Compliance Checking

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

PRESCRIPTIVE MAINTENANCE OF SEWER SYSTEMS BASED ON LARGE LANGUAGE MODELS

Thamer Al-Zuriqat[1], Dhananjay Birmole[1], Kosmas Dragos[1], and Kay Smarsly[1]
[1]Institute of Digital and Autonomous Construction, Hamburg University of Technology,
Blohmstraße 15, 21079 Hamburg, Germany

Abstract

Sewer systems constitute urban infrastructure critical for public health, environmental protection, and flood prevention. Decision-making in sewer systems maintenance often fails to effectively and efficiently interpret heterogeneous and multimodal inspection data into actionable maintenance strategies. To improve decision-making in sewer systems maintenance, this paper introduces a prescriptive maintenance (RxM) framework, in which a pre-trained large language model (LLM) is adapted to the domain of sewer systems using adapter-based fine-tuning. As will be shown in this paper, the LLM is capable of interpreting heterogeneous, multimodal inspection data and recommending stepwise repair procedures of sewer systems, following the paradigm of RxM.

Introduction

Civil infrastructure plays an important role in shaping modern societies by providing services essential for public well-being, such as transportation, water supply, energy, and communication networks. Maintenance of civil infrastructure is vital to ensure long-lasting, healthy, and functional infrastructure. Sewage systems comprise infrastructure for collecting and transporting sewage ("sewer systems"), and facilities for treating and disposing sewage ("wastewater treatment plants") (Davis & Cornwell, 2018). Sewer systems consist of pipelines, manholes, and pumping stations, integrated into unified networks. The operational reliability of sewer systems may be compromised by degradation due to aging, increasing hydraulic loads, and overstressing, driven by urban development and climate change (Hartmann et al., 2024).

Maintenance of sewer systems is carried out using either reactive or proactive maintenance strategies (Hartmann et al., 2025). Reactive maintenance is performed after the occurrence of significant damage, emergencies, or functional disruptions in a sewer system. By contrast, proactive maintenance comprises measures implemented independently of damage, emergencies, or functional disruptions. Proactive maintenance strategies include (i) preventive maintenance, which relies on scheduled interventions independent of condition assessment, (ii) condition-based maintenance, which is triggered by condition-assessment observations, (iii) predictive maintenance, which incorporates forecasts of degradation progression to support maintenance planning (Errandonea et al., 2020), and (iv) prescriptive maintenance, which utilizes predicted system states together with technical, operational, and contextual constraints to recommend specific maintenance actions (Smarsly et al., 2025). Within proactive maintenance strategies, reliable decision-making is crucial, particularly in prescriptive maintenance, where actions depend on the interpretation of heterogeneous and multimodal inspection data. Such multimodal inspection data includes sensor data, closed-circuit television (CCTV) footage, inspection reports, and expert assessments, which must be translated into actionable stepwise repair procedures.

A variety of approaches toward decision-making in sewer system maintenance have been proposed and can be classified into (i) traditional methods and (ii) artificial-intelligence (AI)-based methods. Traditional methods comprise visual inspections (Yang et al., 2011), standardized "defect" (i.e. damage) coding schemes, and periodic condition assessment protocols (Cao et al., 2023), which have formed the foundation of early sewer system maintenance practices. Nonetheless, traditional methods are constrained in processing heterogeneous and multimodal data, exhibit subjectivity in defect classification, and lack systematic mechanisms for integrating sensor data, visual inspection data, and textual documentation into coherent maintenance actions. AI-based methods have gained increasing attention in recent years due to advances in sensing technologies, computational resources, and data availability. AI-based methods, such as convolutional neural networks (Wang et al., 2023), "you-only-look-once" (YOLO) models (Zhang et al., 2023), single-shot multi-box detectors (Shen et al., 2023), and semantic segmentation models (Wu et al., 2025) have demonstrated effectiveness in detecting defects, classifying damage, and forecasting degradation in sewer systems. However, AI-based methods remain constrained by training-data imbalance, sensitivity to noisy inspection data, limited interpretability, and insufficient integration with established maintenance standards, resulting in outputs hardly amenable to translation into consistent, stepwise repair procedures of sewer systems. Consequently, despite the importance of reliable decision-making in prescriptive maintenance, traditional decision-making approaches often fall short of effectively and efficiently interpreting heterogeneous and

multimodal inspection data into actionable stepwise repair procedures.

To improve decision-making in sewer system maintenance, this paper introduces a prescriptive maintenance (RxM) framework, in which a pre-trained large language model (LLM) is adapted to the domain of sewer systems maintenance. In particular, the LLM is adapted to the domain of sewer systems maintenance using adapter-based fine-tuning techniques, which involves training an adapter module using a dataset consisting of repair and rehabilitation standards, subsequently integrated into the pretrained LLM. Thereupon, the LLM is capable of interpreting heterogeneous, multimodal inspection data and recommend detailed, stepwise repair procedures, adhering to the RxM paradigm. Contrary to approaches limited to prompt engineering, this study performs domain-specific fine-tuning of the Mistral-7B-Instruct-v0.2 LLM using the low-rank adaptation (LoRA) process, thereby, embedding codified normative knowledge, stemming from sewer systems maintenance repair and rehabilitation standards, into the internal representations of the LLM. The remainder of this paper is structured as follows: First, the design of the RxM-LLM framework is illuminated. Then, the implementation of the of the RxM-LLM framework is presented, followed by a validation test using a dataset not used during fine-tuning the LLM. Next, the results of the validation test are presented and discussed. Finally, the summary and conclusions as well as an outlook on potential future research directions are provided.

Design of the RxM-LLM framework

The rationale behind using LLMs for RxM of sewer systems lies in the capabilities of LLMs to self-learn complex relationship within heterogeneous and multimodal inspection data, as well as procedures mentioned in repair and rehabilitation standards. In this section, the design of the RxM-LLM framework is presented, comprising two phases, (i) preparing the fine-tuning dataset, and (ii) fine-tuning the LLM. Each phase comprises several steps, flowcharted in Figure 1.

Phase 1: Preparing the fine-tuning dataset

1. In the first step, repair and rehabilitation standards related to sewer systems maintenance are selected. Standards providing guidelines associated with inspection, repair, rehabilitation, and operational procedures for sewer systems are considered. The purpose of this step is to identify codified normative documents with technical content that can be used to form the foundation of the fine-tuning dataset.
2. In the second "preprocessing" step, standards, typically available in document formats, such as portable document format (PDF) files, that are not directly interpretable by machines are converted into a machine-readable dataset $D_{mr} = \{d_1, d_2, \ldots, d_i\}$, where d_k denotes the kth standard series part (e.g. DIN EN 1986-4) with $k = 1...i$. Each standard series part d_k is decomposed into page-wise text chunks using context-aware extraction. Specifically, each page j of d_k is represented as a text chunk $c_{k,j}$, forming the set $d_k = \{c_{k,1}, c_{k,2}, \ldots, c_{k,n}\}$, where n denotes the total number of pages in d_k. Furthermore, figures and tables appearing on page j are referenced within the corresponding text chunk $c_{k,j}$, thus preserving contextual integrity and ensuring that D_{mr} is coherent and consistently formatted.

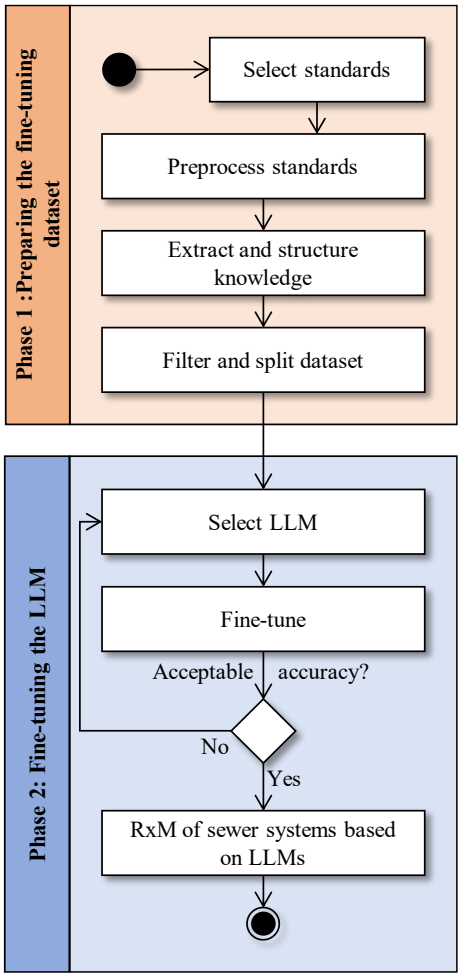


Figure 1: Flowchart of the RxM-LLM framework.

3. In the third "knowledge extraction and structuring" step, a question-answer (Q&A) dataset $D_{Q\&A}$ is generated using a two-stage strategy for fine-tuning the LLM.
   a. In the first stage, a document-level summary S_k is generated for each d_k in D_{mr}. S_k may capture the overarching scope, dominant technical topics, and normative intent of d_k. Furthermore, S_k will serve as a high-level representation of d_k and is not intended to replace the original text in d_k. Instead, S_k is used as contextual guidance to ensure that subsequent processing steps remain aligned with the overall normative context of d_k.
   b. In the second stage, each text chunk $c_{k,j}$, derived from d_k, is processed to generate Q&A pairs according to a domain-specific schema, comprising material type, damage type, damage

cause, damage location, damage severity, and potential maintenance procedures. The domain-specific schema enables a structured generation of $D_{Q\&A}$, ensuring that the Q&A pairs are context-aware and aligned with the requirements for RxM of sewer systems.

4. In the fourth step "filter and split dataset", $D_{Q\&A}$ is filtered to identify and remove similar (duplicate) entries, yielding a "filtered" Q&A dataset ($\tilde{D}_{Q\&A}$) with no redundant Q&A pairs, thereby improving the dataset diversity during fine-tuning. Next, the $\tilde{D}_{Q\&A}$ is split into a training dataset $\tilde{D}_{Q\&A(training)}$, a validation dataset $\tilde{D}_{Q\&A(validation)}$ (to avoid overfitting issues while fine-tuning the LLM), and a testing dataset $\tilde{D}_{Q\&A(testing)}$ (to eventually test the training performance with input data previously unused).

Phase 2: Fine-tuning the LLM

1. In the first step of the phase 2, candidate LLMs are selected for fine-tuning and adapting to the domain of sewer systems RxM. In general, LLMs, under the umbrella of AI systems, are trained on "big" datasets consisting of text from sources such as books, journals, and web pages, assisting the LLMs to comprehensively extract linguistic patterns, semantics, and grammar structures. Thus, LLMs are capable of analyzing, comprehending, and producing human-like language through advanced deep learning techniques, rendering LLMs well-suited for RxM applications. The candidate LLMs are assessed according to criteria, including, (i) parameter scale, relative to available computational resources, (ii) support for instruction tuning and parameter-efficient fine-tuning methods, (iii) model openness and reproducibility, and (iv) compatibility with structured Q&A generation. Then, based on the aforementioned criteria, a single LLM is selected for subsequent fine-tuning and adaptation in the next step.
2. In the second step, the LLM is fine-tuned using the $\tilde{D}_{Q\&A}$ dataset. During training, a large number of LLM parameters (sometimes in the order of billions) for predicting the next word in a sequence are adjusted, gradually learning to produce accurate, context-aware outputs. As a result, training a new LLM for sewer system maintenance would be time-consuming, and fine-tuning, i.e. retraining an already trained LLM, is employed as an alternative, essentially adapting the model to the domain of sewer systems maintenance.

3. In the final step, the accuracy of the fine-tuned LLM is checked using cross-entropy loss and mean "token" accuracy. Each token expresses the smallest unit of text processed by the LLM, obtained through sub-word tokenization, and may represent a complete word, a word fragment, a punctuation mark, or a special symbol. On the one hand, the cross-entropy loss represents the prediction error by measuring the divergence between the distribution of tokens predicted by the LLM and the ground-truth distribution, derived from the training dataset. Fine-tuning entails minimizing the cross-entropy loss between the probabilities of tokens predicted by the LLM and the reference tokens, provided by the ground truth. Low loss values indicate that the distribution of tokens predicted by the LLM closely matches the ground-truth distribution. On the other hand, the mean token accuracy serves as the principal metric for assessing fine-tuning performance, expressing the fraction of tokens predicted correctly out of all tokens in each "mini-batch" (subset of the batch, i.e. training data, processed by the LLM in a single training iteration). The mean token accuracy metric provides a granular indication of the sequence-level ability of the LLM to reproduce linguistic patterns. High mean token accuracy indicates that the LLM is capable of generating outputs closely aligned with realistic sewer systems maintenance procedures.

The combination of the two phases results in actionable information, facilitating real-time decision making in sewer system maintenance. In what follows, the implementation of the RxM-LLM framework is presented.

Implementation of the RxM-LLM framework

The implementation is presented in accordance with the phases of the RxM-LLM framework, previously described, i.e. "preparing the fine-tuning dataset" and "fine-tuning the LLM". In this context, standards on sewer systems repair and rehabilitation, issued by the German Institute for Standardization (Deutsches Institut für Normung, DIN) and harmonized as European standards (Europäische Normen, EN), as well as standards issued by the German Association for Water, Wastewater and Waste (Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall, DWA) are incorporated into the implementation. Furthermore, the programming language Python is employed for implementing both phases.

Phase 1: Preparing the fine-tuning dataset

Figure 2 illustrates the standard series, selected in step 1, used in this study. The parts from the standards series selected, specifically 33 parts, are acquired through the databases "DWA Regelwerk" (Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall (DWA), n.d.) and the standards database "Nautos" (Nautos, n.d.) as PDF files. The standards include:

- The DIN EN 13508 series, which constitute a Europe-wide-harmonized basis for governing inspection, assessment, and documentation of sewer and drainage systems (DIN EN 13508),
- The DIN EN 1986 series, which regulates the planning, design, construction, operation, and maintenance of drainage systems located on private properties, ensuring alignment between private and municipal sewer systems (DIN EN 1986),
- The DWA-M 143 series, which defines technical procedures for inspecting, assessing, and rehabilitating sewer systems (DWA-M 143),

- The DWA-M 144 series, which specifies technical requirements, procedures, and evaluation principles for structural condition assessment of sewer pipelines and manholes, providing a uniform basis for damage classification and maintenance planning (DWA-M 144), and
- The DWA-M 149 series, which provides methods and technical criteria for hydrological, hydraulic, and performance-based assessment of sewer systems, supporting reliable network analysis, monitoring, and operational planning (DWA-M 149).

Data preprocessing (step 2) for converting the PDF files into the machine-readable dataset $D_{mr} = \{d_1, d_2, \ldots, d_{33}\}$ is achieved via a Python preprocessing pipeline. Within the pipeline, textual content contained in the PDF files is extracted into a textual corpus, using optical character recognition (OCR). The textual corpus is then segmented into page-wise text chunks, such that each page j of the standard part d_k corresponds to one text chunk $c_{k,j}$. In parallel, figures contained in the PDF files are exported as separate portable network graphics (PNG) files with automatically assigned captions, and tables are reconstructed as spreadsheets. Figures and tables identified on page j are referenced within the corresponding $c_{k,j}$ to preserve contextual integrity between textual content and visual or tabular information.

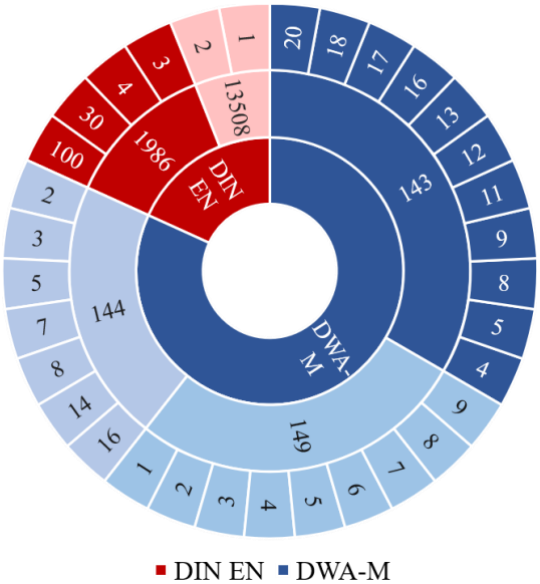


Figure 2: Standards series selected for the implementation of the RxM-LLM framework.

Knowledge extraction and structuring (step 3) for transforming D_{mr} into $D_{Q\&A}$ is accomplished as follows: First, a generation prompt based on Chat GPT-4o-mini LLM is used to produce document-level summaries $S = \{S_1, S_2, \ldots, S_{33}\}$ of D_{mr}, where the summary S_k corresponds to the kth standard part. Next, for standard part d_k, a dedicated generation prompt based on GPT-4o-mini (33 generation prompts in total) is used for generating Q&A pairs using the page-wise text chunks c_{jk}. The GPT-4o-mini generation prompts are constrained by the domain-specific schema, defined in the second stage of step 3, as well as by the document-level summaries $S = \{S_1, S_2, \ldots, S_{33}\}$. As a result, a total of 9,509 Q&A pairs are generated for $D_{Q\&A}$. Figure 3 presents a breakdown of the total Q&A pairs generated from each standard series using the GPT-4o-mini generation prompts.

In the final step of the first phase, filtering and splitting, $D_{Q\&A}$ is filtered to identify and remove duplicate Q&A pairs. Following user review and a cosine-similarity filtering, a total of 476 Q&A pairs are removed. As a result, $\tilde{D}_{Q\&A}$ comprises 9,033 Q&A pairs, which are used for fine-tuning the adapter in the next step. Finally, $\tilde{D}_{Q\&A}$ is split into 80% for training ($\tilde{D}_{Q\&A,training}$) (7,226), 15% for validation ($\tilde{D}_{Q\&A,validation}$) (1,355), and 5% for testing ($\tilde{D}_{Q\&A,testing}$) (452). From $\tilde{D}_{Q\&A,testing}$ a total of 200 Q&A pairs are held-out (not used to evaluate the fine-tuning process) and used later in the validation section. Thus, $\tilde{D}_{Q\&A,testing}$ comprises 252 Q&A pairs.

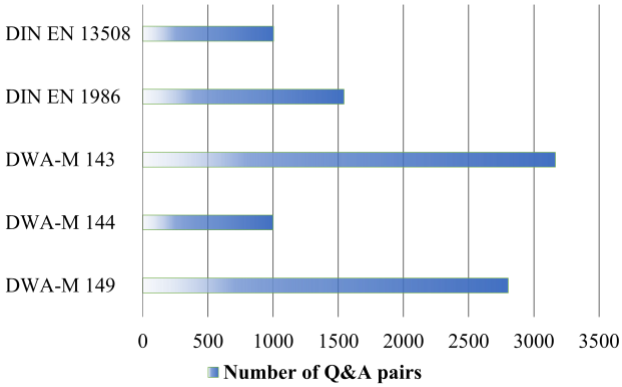


Figure 3: Number of Q&A pairs generated per standard series.

Phase 2: Fine-tuning the LLM

In the LLM selection step, which constitutes the first step of phase 2, two candidate LLMs are initially considered, namely DeepSeek-VL2 and Mistral-7B-Instruct-v0.2 because of multimodal processing and instruction-based text generation, respectively. DeepSeek-VL2 is a multimodal-vision LLM, considered due to the ability of the model to process visual inputs, such as CCTV data, figures from standards, and textual inspection reports. Mistral-7B-Instruct-v0.2 is an instruction-tuned LLM, designed for instruction following and structured text generation. Based on the criteria defined in step 1 of phase 2, the following rationale governs the selection of the LLM: While DeepSeek-VL2 provides a native multimodal processing, the parameter scale and hardware requirements needed for DeepSeek-VL2 exceed the available computational resources. By contrast, Mistral-7B-Instruct-v0.2 supports LoRA through open-source frameworks and enables efficient fine-tuning through libraries such as Hugging Face Transformers, supervised fine-tuning frameworks, and bitsandbytes, under constrained hardware conditions. As a result, the Mistral-7B-Instruct-v0.2 model is selected.

Fine-tuning (step 2 of phase 2) of the Mistral model is achieved via parameter-efficient LoRA. The Fine-tuning is conducted on a single NVIDIA RTX 4090 GPU. The baseline model, i.e. Mistral-7B-Instruct-v0.2, is loaded

using 4-bit NF4 with double quantization, while computations are carried out in bfloat16 to reduce memory consumption. Optimization is performed using the paged AdamW optimizer in 8-bit mode with a learning rate of $2 \cdot 10^{-4}$ and a linear learning-rate schedule. A warm-up phase covering 3% of the total training steps and a weight decay of 0.01 are applied. Training of the adapter is conducted using gradient accumulation to obtain an effective batch size of 32 token sequences, based on a per-device mini-batch size of four token sequences. Input sequences are packed to a maximum length of 2,048 tokens. The training process is configured for a maximum of three epochs, with gradient checkpointing enabled to reduce memory consumption. Moreover, early stop is applied once validation loss (from applying the $\tilde{D}_{Q\&A(validation)}$ dataset) fails to improve two consecutive times. As shown in Figure 4, the fine-tuning process converges to a stable state, with the validation cross-entropy loss stabilizing between 0.32 and 0.34 and the mean token accuracy reaching approximately 88%. The results of the fine-tuning indicate that a consistent and stable adaptation of the Mistral model to the domain-specific task of sewer systems maintenance.

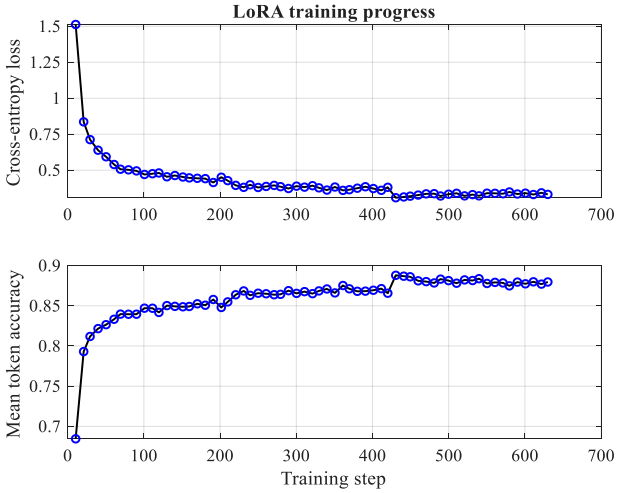


Figure 4. Performance of the LLM fine-tuning, in terms of cross-entropy loss and mean token accuracy.

In the final step, the dataset $\tilde{D}_{Q\&A(testing)}$ is used upon completing training to verify the accuracy of the fine-tuning process of the "adapted" Mistral model, based on the cross-entropy loss and the mean token accuracy, showcasing that 91.82% of the tokens have been correctly predicted. In the following section, the adapted Mistral model, i.e., the Mistral model equipped with the LoRA adapter, is validated and compared against a baseline Mistral model (Mistral-7B-Instruct-v0.2) without adapters.

Validation test

In this section, the RxM-LLM framework is validated by assessing the capabilities of the adapted Mistral model, developed in the implementation section, to interpret heterogeneous, multimodal inspection data, previously unseen, and generate detailed stepwise repair procedures of sewer systems maintenance. In the validation test, a total of 200 Q&A pairs are selected from a held-out dataset not used during fine-tuning. The Q&A pairs cover distinct damage scenarios, damage locations, and repair severity levels. Furthermore, answers generated by the adapted Mistral model are compared with those produced by a baseline Mistral model without adapters, thereby enabling an assessment of how well the adapted Mistral model fits to the domain of sewer systems maintenance.

Both the adapted Mistral model and the baseline Mistral model are evaluated using a domain-specific scoring rubric, which serves as the criterion for assessing the suitability of the models for sewer systems RxM. The scoring rubric is designed to capture requirements of sewer systems maintenance workflows and comprises the following criteria: (i) Structural clarity of responses, assessing whether answers are presented in stepwise and operationally executable formats; (ii) conformity with standards, evaluating explicit references to relevant repair and rehabilitation standards series; (iii) material and damage specificity, assessing the differentiation between pipeline materials (e.g., concrete, vitrified clay, polyvinyl chloride, and glass reinforced plastic) and common damage types (e.g., corrosion, root intrusion, and infiltration); (iv) actionability, evaluating the inclusion of repair and rehabilitation measures, such as CCTV inspection, high-pressure cleaning, injection grouting, short-liner techniques, or cured-in-place pipe lining; (v) consideration of safety and operational constraints, including confined-space entry, bypass pumping, and gas monitoring; (vi) verification and quality assurance measures, such as leak testing, deflection testing, post-rehabilitation CCTV inspection, or as-built documentation; (vii) numerical specificity, assessing the presence of quantitative information included in the repair and rehabilitation standards, such as tolerances or permissible deformation limits; and (viii) a hallucination penalty, applied to answers introducing non-domain-specific or implausible repair and rehabilitation procedures, such as irrelevant inspection technologies or incoherent procedural fragments. The scoring rubric quantifies the performance of each model on a scale from 0 to 10. The results of the validation test are presented and discussed in the following section.

Results and discussion

The quantitative performance of the both models for the 200 Q&A pairs is summarized in Table 1. The evaluation scores are computed using a LLL acting as a judge (LLM-as-a-judge), specifically GPT-5o, which assesses the generated responses based on the predefined rubric criteria introduced in the previous section. The baseline Mistral-7B-Instruct-v0.2 model exhibits an average score of 4.8/10 (±1.2), reflecting frequent fragmentation and limited actionability despite occasional references to standards. Meanwhile, the adapted Mistral model has achieved a higher overall performance with an average score of 7.6/10 (±1.0). The adapted Mistral model demonstrates a strong domain specificity and structural clarity, with correct material and damage references present in 71% of responses.

Table 1: Performance of the adapted Mistral model as well as the baseline Mistral model for the 200 Q&A pairs

Model	Score (1-10)	Standard deviation
Baseline Mistral model	4.8	±1.2
Adapted Mistral model	7.6	±1.0

In addition, more than 65% of the outputs generated by the adapted Mistral model, included at least three distinct repair and rehabilitation measures. A detailed breakdown of the performance of the two models across the individual evaluation criteria, mentioned above, is illustrated in Figure 5.

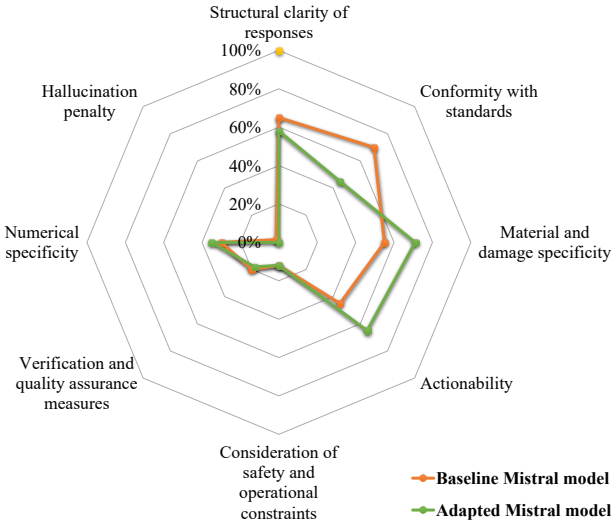


Figure 5. Performance of the two models, based on individual criteria.

Finally, to qualitatively illustrate differences between the baseline Mistral model and the adapted Mistral model, an exemplary prescriptive maintenance query is evaluated across both models. The example addresses the rehabilitation of radial cracks in vitrified clay pipes using short glass-fiber liners, representing a typical maintenance task that requires material awareness, damage-specific repair selection, and operational sequencing. The responses generated by both models are shown in Figure 6.

The baseline Mistral model demonstrates general instructional capability but lacks the technical insight required for RxM of sewer systems. The responses generated by the baseline Mistral model remain abstract and primarily describe general principles rather than actionable procedures. In addition, in outputs generated by the baseline Mistral model, damage-repair coupling is weak, and outputs frequently lack material-specific adaptation and operational details. Furthermore, stepwise repair procedures execution is incomplete, and references to safety measures and quality assurance are often omitted. Finally, the baseline Mistral model provides a superficial overview of repair and rehabilitation concepts without sufficient procedural specificity, rendering the output unsuitable for real-world RxM applications of sewer systems.

The adapted Mistral model shows a clear improvement over the baseline Mistral model, driven by damage-repair mappings introduced through structured fine-tuning. The outputs generated by the adapted Mistral model demonstrate strong alignment with technical procedures and high domain awareness. Furthermore, the adapted Mistral model generates detailed, stepwise repair and rehabilitation procedures, and consistently accounts for sewer pipes material, defect types, and conditions severity. In addition, the adapted Mistral model produces outputs resembling technical repair and rehabilitation reports rather than generic advisory text. In addition, repair and rehabilitation techniques such as cured-in-place pipe lining, short-liner installation, robotic sealing, injection grouting, packer-based repairs, and reinstatement are frequently and appropriately referenced. The adapted Mistral model exhibits accurate matching between damage mechanisms and repair methods, indicating strong alignment with sewer systems repair and rehabilitation practices.

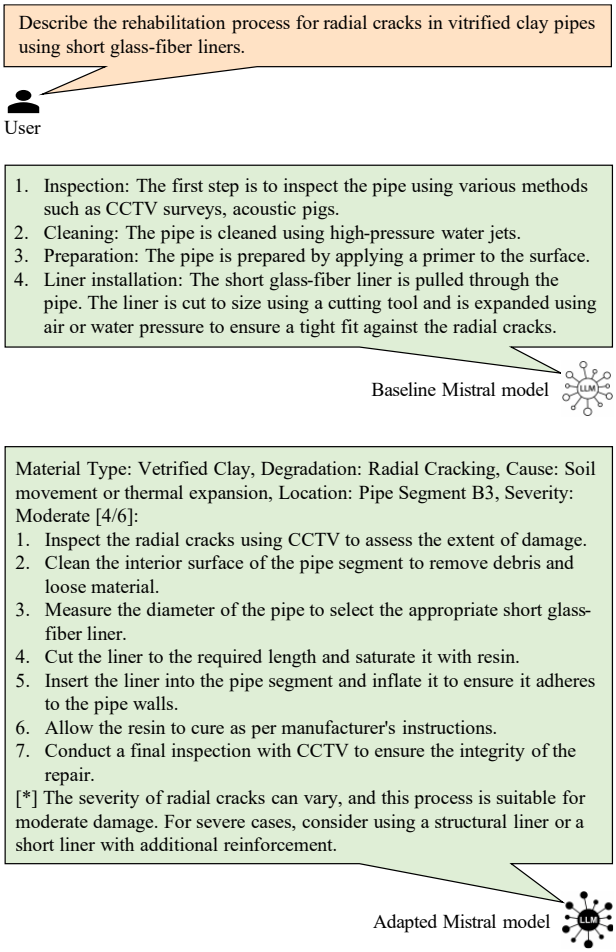


Figure 6: Example conversation of a user with the baseline Mistral model and the adapted Mistral model.

Summary and conclusions

Sewer systems constitute urban infrastructure critical for public health, environmental protection, and flood prevention. Sewer systems maintenance is vital to ensure long-lasting, healthy, and functional infrastructure.

However, decision-making in sewer systems maintenance has failed to effectively interpret heterogeneous and multimodal inspection data into actionable maintenance strategies. This paper has presented a prescriptive maintenance framework, in which, a pre-trained large language model has been adapted to the domain of sewer systems maintenance. The RxM-LLM framework comprises two phases, (i) preparing the fine-tuning dataset, and (ii) fine-tuning the LLM, and it has been implemented based on sewer systems standards defined under the DIN EN and DWA standards. In the implementation of the RxM-LLM framework, a generation prompt based on the ChatGPT-4o-mini model has been employed to generate document-level summaries for each standard. Thereupon, the summaries, together with a domain-specific schema, have been used to constrain downstream generation prompts for producing structured Q&A pairs. The dataset has served as an input for low-rank adaptation fine-tuning of the Mistral-7B-Instruct-v0.2 model. In the validation test, capabilities of the adapted Mistral model in interpreting heterogeneous, multimodal inspection data, and generate detailed stepwise repair procedures of sewer systems maintenance have been evaluated and compared to a baseline Mistral model. The validation test results has shown that the adapted Mistral model outperformed the baseline Mistral model. The adapted Mistral model has demonstrated enhanced structural clarity, consistently generating material-aware, damage-specific, and actionable repair and rehabilitation procedures. The results have demonstrated that adapter-based fine-tuned LLMs are capable of correctly interpreting heterogeneous and multimodal inspection data and recommend stepwise repair procedures, thus supporting decision-making in sewer systems maintenance. Future work may involve further developing the RxM-LLM framework to improve uncertainty handling and additional validation tests using real-world sewer systems inspection data.

Acknowledgments

This work is funded by the German Research Foundation (DFG) under grant SM 281/33-1 and by the German Federal Ministry of Transport (BMV) under grant 01FV2059C. The financial support is gratefully acknowledged. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of BMDV or DFG.

References

Cao, G., Guo, S., Wei, J., Huang, R., & Li, M. (2023). Comparison of four sewer condition assessment protocols based on field data. Water, 15(21), 3790.

Davis, M.L. & Cornwell, D.A. (2013). Introduction to environmental engineering (5th ed.). McGraw-Hill Education. New York, NY, USA.

Deutsches Institut für Normung (DIN). DIN EN 13508: Investigation and assessment of drain and sewer systems outside buildings – Parts 1 and 2. Berlin, Germany: Beuth Verlag.

Deutsches Institut für Normung (DIN). DIN EN 1986: Drainage systems for buildings and land – Parts 3, 4, 30, and 100. Berlin, Germany.

Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall (DWA). DWA-M 143: Sanierung von Entwässerungssystemen außerhalb von Gebäuden – Teile 4, 5, 8, 9, 11, 12, 13, 16, 17, 18, und 20. Hennef, Germany.

Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall (DWA). DWA-M 144: Zusätzliche Technische Vertragsbedingungen (ZTV) für die Sanierung von Entwässerungssystemen außerhalb von Gebäuden – Teile 2, 3, 5, 7, 8, 14, und 16. Hennef, Germany.

Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall (DWA). DWA-M 149: Zustandserfassung und -beurteilung von Entwässerungssystemen außerhalb von Gebäuden – Teile 1, 2, 3, 4, 5, 6, 7, 8, und 9. Hennef, Germany.

Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall (DWA). DWA-Regelwerk: Technische Regeln Wasserwirtschaft, Hennef, Germany. Available at: https://de.dwa.de/de/regelwerk-fachpublikationen.html (Accessed: 30 April, 2025).

Errandonea, I., Beltrán, S., & Arrizabalaga, S. (2020). Digital Twin for maintenance: A literature review. Computers in Industry, 123(2020), 103316.

Hartmann, S., Al-Zuriqat. T., Peralta, P., Valles, R., Gölzhäuser, P., Al-Hakim, Y., Mackenbach, S., Klemt-Albert, K., & Smarsly, K. (2024). Digital twins for sewer systems maintenance: A systematic literature review. In: Proceedings of the 24th International Conference on Construction Applications of Virtual Reality (CONVR). Sydney, Australia, November 4-6, 2024, pp.345-357.

Hartmann, S., Valles, R., Schmitt, A., Al-Zuriqat, T., Dragos, K., Gölzhäuser, P., Jung, J. T., Villinger, G., Varela Rojas, D., Bergmann, M., Pullmann, T., Heimer, D., Stahl, C., Stollewerk, A., Hilgers, M., Jansen, E., Schoenebeck, B., Buchholz, O., Papadakis, I., Merkle, D. R., Jäkel, J.-I., Mackenbach, S., Klemt-Albert, K., Reiterer, A., & Smarsly, K. (2025). Digital-twin-based management of sewer systems: Research strategy for the KaSyTwin project. Water, 17(3), 299.

Nautos (n.d.) Nautos database for standards, technical regulations and guidelines. Berlin, Germany. Available at: https://nautos.de (Accessed: 30 April, 2025).

Shen, D., Liu, X., Shang, Y. and Tang, X. (2023). Deep learning-based automatic defect detection method for sewer pipelines. Sustainability, 15(12), 9164.

Smarsly, K., Chmelnizkij, A., Dragos, K., Peralta, J., Al-Zuriqat, T., Ahmad, M. E., Chillón Geck, C., Peralta, P., & Seitz, L. (2025). Decision making in structural

health monitoring using large language models. In: Proceedings of the 15th International Workshop on Structural Health Monitoring (IWSHM). Stanford, CA, USA, September 09-11, 2025, pp.641-648.

Wang, X., Thiyagarajan, K., Kodagoda, S. and Zhang, M. (2023). PIPE-CovNet: Automatic in-pipe wastewater infrastructure surface abnormality detection using convolutional neural network. IEEE Sensors Letters, 7(4), pp.1-4.

Wu, Z., Li, M., Han, Y. and Feng, X. (2025). Semantic segmentation of 3D point cloud for sewer defect detection using an integrated global and local deep learning network. Measurement, 253(2025), 117434.

Yang, M.D., Su, T.C., Pan, N.F., &Yang, Y.F. (2011). Systematic image quality assessment for sewer inspection. Expert Systems with Applications, 38(3), pp.1766-1776.

Zhang, J., Liu, X., Zhang, X., Xi, Z. and Wang, S. (2023). Automatic detection method of sewer pipe defects using deep learning techniques. Applied Sciences, 13(7), 4589.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

FAST-ASSET: A REAL TIME SCAN-TO-3D POINT CLOUD BIM-READY ASSET GENERATION METHOD FOR SMALL MEP ELEMENTS

Urueña, Asier[1]; Eskudero, Ibon[1]; Sagasti, Diego[1]; and Noyé, Sarah[1]
[1]FUNDACION TECNALIA RESEARCH & INNOVATION, Spain

Abstract

Building Information Modeling (BIM) requires timely as-is updates, but existing scan-to-BIM pipelines based on laser scanning or photogrammetry demand specialised hardware and offline processing, limiting frequent small-asset updates in digital-twin maintenance. This paper introduces Fast-Asset, a real-time on-site framework integrating mixed-reality capture, learning-based detection, and incremental 3D reconstruction to generate BIM-ready, pre-segmented point clouds of small MEP assets in a single walkthrough. Instant visual feedback enables in-situ validation and re-capture. Experiments demonstrate a 3-to-10× reduction in capture-to-model time versus laser scanning and photogrammetry, with comparable surface roughness and low background contamination, collapsing multi-stage offline processes into a single field operation.

Introduction

Building Information Modeling (BIM) has become the digital backbone of the construction value-chain, linking design, construction, and facility-management activities (Yilmaz et al., 2023). Although owners increasingly rely on BIM for digital-twin services such as life-cycle assessment, operation, inspection, and predictive maintenance (Boje et al., 2023), most legacy models only reflect the as-designed state. Reality inevitably diverges from design during construction, refurbishments and operation, so decision-makers need rapid, low-cost updates of the as-built geometry and semantics (Wijeratne et al., 2023). Within the digital-twin lifecycle, keeping the virtual replica synchronised with the physical asset is essential for reliable monitoring and maintenance. Yet the cost of re-scanning makes frequent synchronisation impractical for small components. By enabling decentralised, operator-driven model updates, frameworks such as the one proposed in this paper can reduce dependency on specialist surveying teams and shorten the governance cycle for BIM lifecycle management.

Obtaining such updates typically requires expensive static laser scanners or high-end LiDAR rigs, specialist crews, and a multi-day offline processing chain (cleaning, registration, segmentation and modelling). When only a few small industrial assets (valves, pumps, switch-gear, etc.) need to be captured, the cost and latency of conventional scan-to-BIM workflows become prohibitive.

Research over the last decade has gradually shifted from heavy terrestrial laser scanning toward lighter, hybrid capture solutions and from fully manual pipelines to increasingly automated ones.

Scan-to-BIM automation. Automated pipelines that generate semantically labelled point clouds from large-scale scans have reached centimetre-level accuracy for transportation assets (Ding et al., 2024). For construction, guides exist for the integration of such scans into BIM. However, these approaches still rely on bulky LiDAR devices and require offline processing on workstations or cloud servers.

Point-cloud generation and completion. Early efforts focused on reconstructing dense surfaces from laser data. Recent work targets the completion of sparse, noisy data captured with consumer-grade depth sensors. Deep generative models can hallucinate missing geometry, yet they demand offline GPU-based training and inference that is unsuitable for field deployment (Zhuang et al., 2024).

Small-object capture. The literature distinguishes between detecting small assets and generating high-fidelity point clouds for them. Abreu et al. (2023) review scan-to-BIM pipelines and conclude that procedural modelling is essential for objects below 0.1 m. Building on this insight, Pan et al. (2022) combine terrestrial LiDAR with AI-guided photogrammetry to enrich digital twins with sockets, switches and fire-safety fixtures, achieving sub-15 mm accuracy in under three minutes per room. Chen et al. (2023) further demonstrate depth-completion networks that reconstruct fine steel rebars (≈ 20 mm diameter) from multi-view stereo, highlighting the demand for methods that preserve small-object topology.

Mixed-reality capture. Head-mounted devices equipped with RGB-D sensors (e.g., Microsoft HoloLens 2) provide real-time spatial mapping, on-device inference and intuitive visual feedback (Navares-Vázquez et al., 2023). These platforms enable operators to place semantic tags in situ, reducing segmentation error to 2.4 ± 1.8 cm. Nevertheless, the raw point clouds remain cluttered and three-dimensional object detection is still computationally demanding; most approaches still require a desktop step to export IFC-ready geometry.

Detection and segmentation of tiny objects. Clustering-based methods such as adaptive DBSCAN can improve recall for small targets in low-density point clouds (Duan et al., 2024), but they do not integrate with mixed-reality

capture hardware. Lightweight 2D detectors (YOLO family) have proven effective for real-time proposal generation, yet their translation to 3D segmentation remains an open research question.
Despite these advances, three critical gaps prevent a truly on-site, real-time *as-is* BIM update for small-scale MEP components:

1. **On-device processing.** Existing pipelines either offload point-cloud cleaning and segmentation to a workstation or require GPU clusters, contradicting the goal of a single-walkthrough workflow.
2. **Small-object focus.** Methods that achieve sub-centimetre accuracy for tiny assets generally depend on multi-view acquisition or heavy LiDAR, and they rarely produce asset-level point clouds that are already semantically linked to BIM entities.
3. **Immediate usability.** Current mixed-reality tools improve semantic tagging but still require a post-processing stage to obtain a clean, exportable model, delaying the availability of up-to-date information at the construction site.

These limitations hinder the adoption of scan-to-BIM technology in time- and cost-sensitive contexts where maintaining up-to-date digital twins is critical for operational decisions and maintenance planning.
To address those limitations, this paper introduces Fast-Asset, a modular, real-time framework for generating clean, segmented 3D point clouds of small assets directly on site. Fast-Asset integrates four tightly coupled stages:

1. **Live RGB-D capture** that maintains a globally consistent pose while the operator scans the asset.
2. **Lightweight 2D detection** to produce per-frame object masks and class predictions.
3. **Incremental point-cloud fusion**, where masked depth data are merged into clean, class-labelled point clouds whose centroids are verified in situ.
4. **Instant export** of each segmented cloud together with metadata (type, instance ID, acquisition time) ready to be converted to an IFC-compliant schema.

The contribution of the Fast-Asset framework is that the operator receives instant visual feedback on every capture and can re-scan if necessary, reducing revisits and manual post-processing. Because every asset is captured, validated and geo-referenced during a single walkthrough, facility managers can import the generated model snippets directly into the master BIM and obtain an *as-is* representation without manual cropping, registration or annotation. The paper evaluates Fast-Asset against photogrammetric and laser-scanned baselines with respect to both capture-to-model time and geometric accuracy, using intrinsic surface-quality indicators to benchmark lightweight scan-to-BIM systems. By collapsing a multi-day laser-scan and modelling loop into a single on-site operation, Fast-Asset extends recent small-object enrichment approaches such as (Pan et al., 2022) and learning-based completion pipelines such as Chen et al. (2023).

Methodology

General approach

Fast-Asset is a modular, real-time scan-to-asset framework whose pipeline (Figure 1) divides the process into independent yet connected phases: capture, alignment, detection, reconstruction and cleaning. Each block can be updated or swapped without redesigning the rest, ensuring continued relevance as sensors, algorithms or standards evolve.
Unlike laser scanning or deep-learning reconstructions that produce dense scene-wide point clouds requiring costly post-processing, Fast-Asset fuses RGB and depth images to combine scanning and segmentation in a single, lightweight step executable in real time, avoiding the computational burden of techniques such as NeRF or Gaussian splatting and the cost of high-precision hardware.

Framework phases

Built environment capture

The capture phase acquires aligned RGB and depth images to build an RGBD representation of the target asset. While laser scanners provide high accuracy, they are slow and expensive; standalone RGB cameras are affordable but lack native depth. Fast-Asset proposes a device-based capture framework with a built-in depth sensor that delivers pre-aligned RGBD frames, although it could be replaced by any source of RGBD imagery, including monocular depth estimation.

2D-3D alignment

Fast-Asset relies on sensor fusion from a head-mounted device with factory-calibrated RGB and depth sensors, yielding pre-aligned RGB-D frames and eliminating costly software-based 2D-3D alignment required in prior work. Given the depth camera intrinsics $\mathbf{K}_d$ and the extrinsic transform $[\mathbf{R} \mid \mathbf{t}]_{d\to\mathrm{rgb}}$, each depth pixel (u_d, v_d) with depth d is first back-projected into 3D in the depth sensor frame and then mapped to the RGB image plane via the RGB intrinsics $\mathbf{K}_{\mathrm{rgb}}$:

$$\mathbf{p}_d = d\,\mathbf{K}_d^{-1}\begin{bmatrix} u_d \\ v_d \\ 1 \end{bmatrix}, \quad \mathbf{p}_{\mathrm{rgb}} = \mathbf{R}\,\mathbf{p}_d + \mathbf{t} \tag{1}$$

IMU measurements compensate for head motion and maintain temporal synchronisation, so each RGB-D frame remains geometrically consistent for subsequent reconstruction (Dibene and Dunn, 2022).

Asset identification

The proposed method relies on 2D deep-learning detectors to recognise assets on RGB images rather than on 3D point clouds, providing three advantages: (i) lighter computation enabling real-time segmentation, (ii) faster and

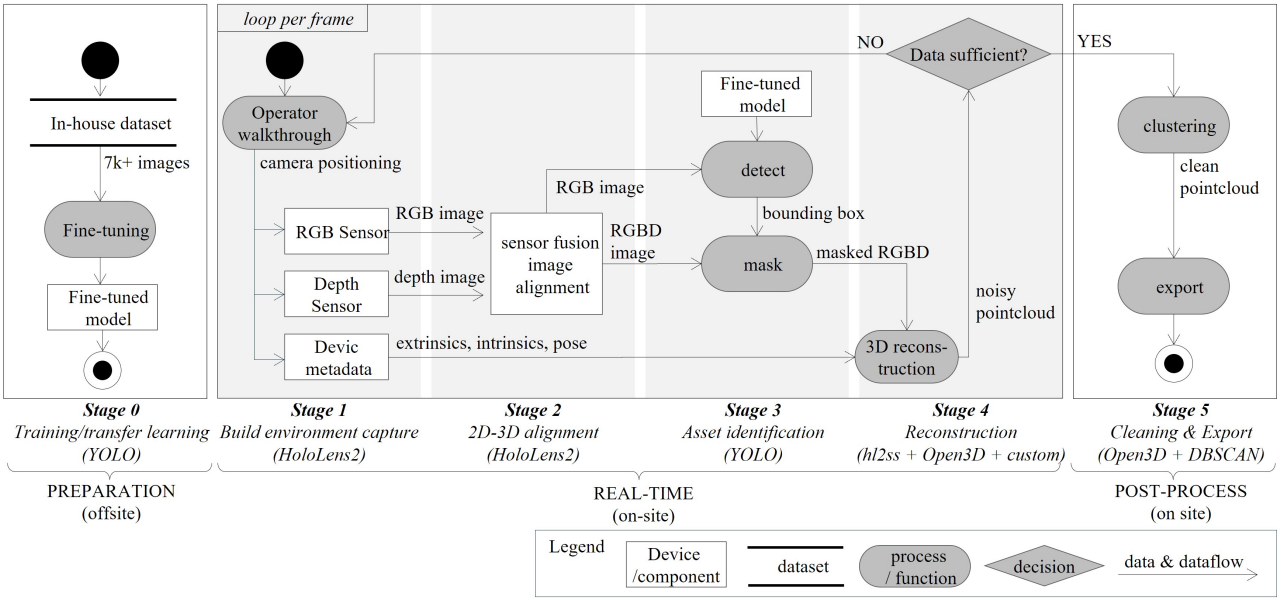


Figure 1: Framework overview showing needed devices/elements and needed tasks for each stage

cheaper dataset creation thanks to 2D image annotation labeling, and (iii) reuse of existing image-based models via transfer learning. By leveraging pre-trained weights from foundation models, the system can be adapted to new asset classes with minimal retraining effort.

Reconstruction

The system operates on a temporal sequence of frames $\{(\mathbf{I}_t, \mathbf{D}_t, \mathbf{T}_t, \mathbf{M}_t)\}_{t=1}^{T}$, where $\mathbf{I}_t$ is the RGB image, $\mathbf{D}_t$ the depth map, $\mathbf{T}_t \in SE(3)$ the camera pose and $\mathbf{M}_t$ the binary detection mask. For each masked pixel (u, v) with $\mathbf{M}_t(u, v) = 1$ and depth $d = \mathbf{D}_t(u, v) > 0$, a 3D point in the global frame is obtained as:

$$\mathbf{p}^w = \mathbf{T}_t \, d \, \mathbf{K}^{-1} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{2}$$

The fused point cloud is the union $\mathscr{P} = \bigcup_t \{\mathbf{p}_{t,i}^w\}$, aggregating partial views from multiple viewpoints and progressively densifying the asset representation. The method merges per-frame contributions without explicit overlap validation, deferring class-based filtering to the cleaning phase.

Cleaning

A dedicated cleaning module eliminates noise, artefacts and background geometry. In the current implementation, DBSCAN (Ester et al., 1996) clusters the fused cloud by defining a point $\mathbf{p}$ as a *core point* when its ε-neighbourhood $N_\varepsilon(\mathbf{p}) = \{\mathbf{q} \in \mathscr{P} : \|\mathbf{p} - \mathbf{q}\| \leq \varepsilon\}$ contains at least *MinPts* neighbours. Mutually density-reachable core points are grouped into clusters and only the largest cluster is retained as the final asset representation, discarding isolated noise points and spurious blobs. The parameters ε and *MinPts* are set proportionally to the median nearest-neighbour distance of the cloud, ensuring robustness across varying scan densities. The module can be adapted to specific asset types or detection models with varying false-positive rates.

Taken together, these phases provide a complete specification of the Fast-Asset framework, from data capture to BIM-ready point-cloud output. The different phases could be updated with improved technologies without altering the framework.

Implementation

Overview

The implementation performed couples live RGB-D capture from a Microsoft HoloLens 2, a YOLOv7-based detector fine-tuned for switches and outlets, and a Python back-end that performs incremental point-cloud fusion and clustering with Open3D. As the operator walks through the room, detections are streamed, masked depth is converted to 3D points and fused into a single, segmented point cloud for each target object. At the end of a short walkthrough, the operator receives a cleaned point cloud per asset, exported as `.pcd` for further BIM integration. The four main stages of this workflow (training, capture, streaming+detection, and reconstruction+cleaning) are summarised in Figure 2.

Hardware and capture setup

All experiments were run using a mobile mixed-reality device and a standard workstation. The processing machine was a Dell Precision-class laptop with an Intel Core i7 CPU, 32 GB RAM, SSD storage and a mid-range NVIDIA GPU. These specifications represent a realistic configuration for on-site or near-site processing in construction projects.

Microsoft HoloLens 2 was used as the capture device. It integrates an RGB camera and a long-range depth sensor with factory calibration, providing aligned RGB-D frames

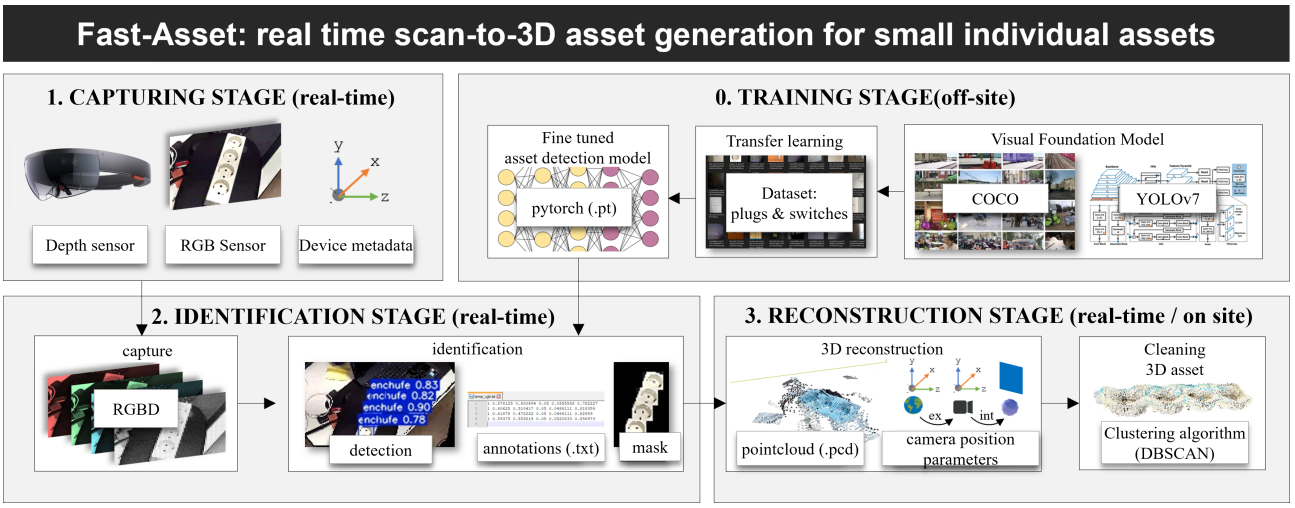


Figure 2: Implemented architecture for streaming detection and point-cloud segmentation

in real time as the user moves. Only the front RGB camera and the long-throw depth sensor were used, while other sensors (IMU, eye tracking, etc.) were leveraged implicitly by the HoloLens runtime for pose estimation but did not require explicit handling in our code. The device was connected to the laptop via WiFi in a shared local network, streaming RGB, depth and pose information through the `hl2ss` library (Dibene and Dunn, 2022). This configuration mirrors the intended deployment context of Fast-Asset, in which an operator walks through a facility with a head-mounted device and offloads the heavier processing to a nearby computer.

Software stack

The implementation was developed in Python 3.10. The HoloLens sensor streams were accessed through the open-source `hl2ss` project, which exposes RGB and depth frames together with camera intrinsics and extrinsics in a client–server setup. YOLOv7 was selected as the object-detection backbone because it combines real-time performance with good accuracy on embedded and edge devices. The publicly available COCO-pretrained weights (Lin et al., 2015) were used as a starting point and then fine-tuned via transfer learning for the target classes (outlets and switches). Open3D (Zhou et al., 2018) was used as the 3D processing back-end, providing utilities for RGB-D handling, point-cloud generation, incremental fusion and visualization. For point-cloud cleaning, we relied on the DBSCAN clustering algorithm (Ester et al., 1996) as implemented in Open3D, retaining only the densest cluster to discard background noise and stray points.
The three-component stack (YOLOv7, `hl2ss`, Open3D) keeps external dependencies minimal, making the implementation straightforward to deploy on different machines or to adapt to alternative detectors and capture devices.

Dataset and transfer learning

The detector focuses on two MEP-related classes: power outlets and light switches. A custom dataset was assembled to reflect realistic indoor conditions. First, 568 images of outlets and switches were captured by the research team in different environments (offices, corridors and homes), covering a variety of background textures, lighting conditions and mounting configurations.
To increase robustness and reduce annotation effort, an additional outlet-focused dataset was imported from the Roboflow platform. This contributed 2 263 base images plus augmented variants (rotations), resulting in a total of 6 789 external samples. Combined with the in-house images, the final dataset comprised 7 329 labelled images, of which 20 % were held out for validation. An example of the annotation format is illustrated in Figure 3, where each bounding box is stored as a class index and normalised coordinates in a YOLO-style text file.
Transfer learning on the combined dataset improved outlet detection recall from 64 % to 92 %, while switch detection remained above 80 % (Table 1). The detector also achieved mAP@0.5 = 0.902 (AP@0.5: 0.841 for switches, 0.962 for outlets). There were no inter-class false detections; only the background generated false positives or negatives. These values were considered sufficient to support real-time mask generation in the target use case.

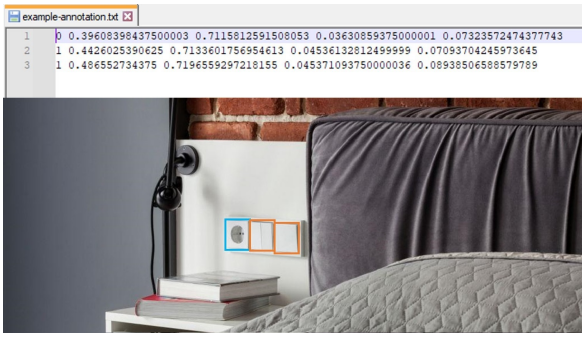


Figure 3: Example of annotations of one of the photographs, where three objects are detected and each line records the class index followed by normalised bounding-box coordinates in $[0, 1]$

Table 1: Normalized confusion matrix (Predicted vs. True).

		True		
		switch	power outlet	background FP
Predicted	switch	0.83	–	0.50
	power outlet	–	0.92	0.50
	background FN	0.17	0.08	–

Normalised per ground-truth class confusion–matrix values. FP false positives, FN false negatives. Computed at a fixed operating point (confidence ≥ 0.4, match IoU ≥ 0.5 after NMS) to reflect the real-time deployment setting and to balance false positives and missed detections while retaining the standard IoU$= 0.5$ matching criterion for a fair, interpretable per-class recall comparison.

Online processing pipeline

The end-to-end implementation follows the methodological stages described earlier but focuses on a practical, streaming-oriented design. For each RGB-D frame delivered by `hl2ss`, the following steps are executed:

1. **Capture and pre-processing.** The client receives an RGB image, a depth map and the current camera pose from the HoloLens 2. Basic consistency checks (resolution, timestamp) are performed, but no heavy pre-processing is required because the device already delivers aligned RGB and depth.
2. **2D detection.** The RGB frame is passed through the YOLOv7 model, which returns bounding boxes, class labels and confidence scores for each detected outlet or switch. Detections below a confidence threshold are discarded to reduce false positives.
3. **Masking and RGB-D construction.** For each retained bounding box, a binary mask is created and applied to the RGB and depth images, zeroing out background pixels. The resulting masked RGB-D frame isolates the candidate asset region while preserving depth information for the visible surfaces.
4. **Point-cloud generation and fusion.** For reconstruction, the masked RGB-D data, together with the camera intrinsics and pose, are back-projected into 3D following Equation 2 and merged into a single global point cloud using Open3D. Multiple viewpoints gradually densify the representation.
5. **Real-time visualization.** An optional Open3D visualizer displays the aggregated point cloud and updates it as new frames arrive. This provides visual feedback so the operator can decide whether additional passes are required to fill occlusions or refine details.

This streaming loop runs continuously while the user scans the environment. Since detection and fusion are relatively lightweight, they can keep up with the HoloLens frame rate on the chosen hardware, enabling an interactive experience: the operator sees both 2D detections and a growing 3D reconstruction within seconds. The timing and resolution metrics obtained from this implementation are later summarised and compared with laser scanning and photogrammetry.

Cleaning and export

Raw fused point clouds inevitably contain residual noise from imperfect detections, depth artefacts and background geometry that slipped through the masks. To obtain a compact, BIM-ready representation, a final cleaning step is applied to each asset-level cloud. In our implementation, DBSCAN clustering is used to segment the cloud into dense regions and outliers. Only the largest, most coherent cluster is retained as the final asset representation, removing isolated points and small spurious blobs around the target object.

The cleaned point cloud is then exported as a `.pcd` file together with basic metadata such as the detected class, a unique instance identifier and the acquisition timestamp. This format can be consumed directly by downstream tools for meshing, parametric fitting or IFC element creation, for instance, by fitting a bounding geometry to the cloud, assigning an `IfcBuildingElementProxy` type and writing the result via IfcOpenShell. In parallel, the implementation logs timing information for the main stages (detection, fusion, cleaning) and the total number of points, enabling quantitative comparison with alternative workflows.

This implementation demonstrates the Fast-Asset concept in practice: small MEP assets can be detected, reconstructed and exported as clean point clouds within a single, short walkthrough, without requiring heavy post-processing or specialised scanning equipment.

Results and discussion

The proposed Fast-Asset workflow was compared against two widely used scan-to-point-cloud baselines, terrestrial laser scanning and photogrammetry, on two representative objects: a power strip, chosen as a typical MEP component that is normally absent from image-based models and a computer keyboard, included as a control object for our transfer-learning pipeline since it is a standard category in COCO-trained YOLO detectors (class ID66). Both objects are medium-sized, abundant in indoor environments and share key traits with typical MEP components: small footprint, surface-mounted installation and prismatic geometry with recesses and holes. Their matte materials avoid specular reflections that hinder photogrammetry and their working distance ($< 2\,\mathrm{m}$) lies within the effective

range of HoloLens 2, making them suitable proxies for everyday BIM-relevant assets.

Time and workflow analysis

Table 2 summarises the scan time, generation time, total time and point-cloud resolution for all combinations of method and object. Laser scanning (Leica P40) required 17 min 30 s per object, yielding dense but unsegmented clouds (≈690k and 659k points). Photogrammetry was run from the same HoloLens 2 captures. While it produced highly detailed models (1.1M–3.0M points), processing times ranged from 4 min to over 50 min and the resulting point clouds include extensive background geometry without semantic segmentation. By contrast, Fast-Asset fixed the scan duration to one minute (to simulate realistic short walkthrough) and performed detection, fusion and cleaning online, yielding total times of 4 min 45 s to 6 min 24 s with pre-segmented, asset-level point clouds. Figure 4 qualitatively illustrates these effects. Photogrammetry and laser scanning retain large amounts of background geometry, requiring post-segmentation, whereas Fast-Asset delivers a pre-segmented cloud concentrated on the target asset.

Geometric quality analysis

Geometric quality is evaluated through two complementary lenses: (i) *cross-method distance metrics* between Fast-Asset and photogrammetry, which share identical acquisitions, using photogrammetry as a dense-reconstruction baseline; and (ii) *intrinsic quality indicators* that characterise each method's reconstruction fidelity independently of a reference model. All point clouds were normalised to a unit bounding-box diagonal before analysis.

Table 3 reports four intrinsic quality indicators computed on the normalised point clouds. **Surface roughness** measures the local planarity residual via PCA on k-neighbourhoods; values of 0.001–0.005 across all methods confirm that all three produce geometrically smooth surfaces, with Fast-Asset achieving the lowest roughness for the power strip (0.0011) and comparable values for the keyboard (0.0033 vs. 0.0026–0.0032). **Normal consistency** reflects how uniformly surface normals are oriented among neighbours. Laser scanning yields the most consistent normals (12–17°), as expected from a high-precision sensor, while Fast-Asset shows higher angular variance (21–30°). This gap reflects the lower spatial resolution of the HoloLens 2 depth sensor relative to laser optics—a hardware constraint, not a pipeline limitation; adopting a higher-resolution depth camera would directly reduce this deviation. **Background ratio** quantifies the proportion of points classified as outliers by statistical outlier removal, serving as a proxy for segmentation effectiveness. Fast-Asset achieves the lowest background for the power strip (1.4 % vs. 3.8–4.1 %), confirming that YOLO detection combined with DBSCAN cleaning effectively suppresses irrelevant geometry; for the keyboard, the moderately higher ratio (5.0 %) results from bounding-box over-inclusion in the denser scan and is readily removable via tighter post-filtering. **Density uniformity** (CV of k-NN distances) shows Fast-Asset achieving the best uniformity for the keyboard (CV = 0.14 vs. 0.16–0.23 for the baselines); the elevated CV for the power strip (1.14) reflects fewer contributing viewpoints during the one-minute scan, addressable through longer or guided acquisition. Overall, Fast-Asset is competitive with the baselines on surface smoothness and excels on at least one object per segmentation-related indicator, while the lower normal consistency is a predictable trade-off of real-time portable capture that does not compromise suitability for BIM-level asset documentation.

Cross-method comparison between Fast-Asset and photogrammetry (both acquired from the same HoloLens 2 session) yielded a normalised RMSE of 0.007 (0.7 % of the object diagonal) and a mean point-to-point distance of 0.006 for the power strip after ICP registration. The keyboard comparison produced a higher RMSE (0.167), partly attributable to cluster-extraction difficulty when isolating the object from the dense photogrammetric cloud. The power-strip result, where cluster extraction was more reliable, therefore provides the more representative accuracy estimate and confirms that the real-time pipeline preserves surface geometry comparable to offline reconstruction, with deviations well within typical BIM tolerance thresholds for small MEP elements.

Error propagation and failure modes

The pipeline performance depends on the accuracy of the 2D detection stage: false positives introduce background fragments into the fused cloud, while missed detections cause incomplete coverage. DBSCAN cleaning mitigates part of these errors by discarding small, isolated clusters; however, when a false detection overlaps spatially with the true asset, the spurious geometry is merged into the main cluster and cannot be removed without additional validation. Conversely, systematic under-detection (e.g. due to occlusion or unusual viewpoints) reduces the surface completeness of the reconstruction. The confusion matrix (Table 1) shows that inter-class misclassification is absent in the current model, so error propagation is limited to background false positives and missed detections against background, both of which have bounded impact on the final asset geometry. Future work could mitigate these issues through multi-view detection consistency checks, 3D instance segmentation to produce tighter masks than bounding boxes, and frame-level ICP validation to reject geometrically inconsistent contributions before fusion.

Scalability

In its current form, Fast-Asset processes one asset class per walkthrough, but the architecture supports multi-class parallel detection because YOLOv7 already outputs bounding boxes for all trained classes simultaneously. Extending to multi-asset environments requires maintaining one fused cloud per detected instance and associating new

Table 2: Scanning time, generation time, total time and point-cloud resolution for the three tested methods.

Methods	Scan time	Generation time	Total time	Total points	
Laser scanner	Keyboard	00:03:30	00:14:00	00:17:30	690 309
Photogrammetry[1]	Keyboard	00:01:00	00:03:02	00:04:02	1 147 792
Fast-Asset[2]	Keyboard	00:01:00	00:05:24	00:06:24	674 708
Laser scanner	Power strip	00:03:30	00:14:00	00:17:30	659 489
Photogrammetry[1]	Power strip	00:01:00	00:49:21	00:50:21	2 955 940
Fast-Asset[2]	Power strip	00:01:00	00:03:45	00:04:45	38 852

[1] Photographs were provided from the HoloLens experiment.
[2] For Fast-Asset, the 3D asset is generated in scan-time, which could be reduced. Generation time can be parallelized with sequential scanning tasks.



Figure 4: Resulting model point clouds using photogrammetry, laser scanner and the proposed method

detections with existing instances via spatial proximity in the global frame. Since detection and back-projection are per-frame operations whose cost is independent of the number of accumulated instances, the system scales linearly with the number of frames rather than with the number of assets, making it suitable for room-level deployments with tens of small objects.

Limitations

The current experiments focus on two objects within the optimal range of HoloLens 2. While the keyboard is not itself an MEP component, pairing it with a power strip allowed us to evaluate both transfer learning on a well-represented class and adaptation to a category absent from standard image models, demonstrating the feasibility of extending Fast-Asset to new asset types. Nonetheless, the choice of proxy objects does not fully capture the diversity of installed MEP components. Generalising to larger assets or full rooms will require multi-scale strategies, additional object categories and tests in operational environments. A rigorous comparison of acquisition cost, training effort and on-site ergonomics is also outside the scope of this paper.

Despite these caveats, the results confirm that Fast-Asset substantially reduces the overall scan-to-BIM time, achieving 4-6 min per asset versus 17–50 min for conventional methods, while delivering pre-segmented, geometrically adequate point clouds with integrated real-time feedback, modular architecture and low background contamination.

Conclusion

The proposed Fast-Asset method produces clean, pre-segmented point clouds of small objects with surface roughness comparable to laser scanning and photogrammetry while reducing total capture-to-model time by a factor 3 to 10 and enabling real-time on-site model generation. For low- to medium-resolution, real-time scanning of individual assets, Fast-Asset better matches practical requirements than conventional approaches. Laser scanners are optimised for large environments and lose efficiency at close range, while photogrammetry, although capable of high-resolution output, lacks real-time feedback and requires offsite reconstruction. In contrast, Fast-Asset generates and visualises a segmented model during acquisition and automates subsequent cleaning once sufficient detail is reached. The geometric quality analysis confirms that the method achieves low background contamination and surface roughness on par with established techniques, while the lower normal consistency, attributable to the depth-sensor resolution rather than the pipeline itself, represents an acceptable trade-off for asset-level BIM updates, supporting shorter governance cycles in which

Table 3: Intrinsic geometric quality indicators for all three methods (unit-normalised point clouds).

Method	Object	Roughness[a]	Normal consist.[b]	Background[c]	Density CV[d]
Laser scanner	Keyboard	0.0026	16.8°	3.0 %	0.16
Photogrammetry	Keyboard	0.0032	14.0°	3.4 %	0.23
Fast-Asset	Keyboard	0.0033	20.9°	5.0 %	0.14
Laser scanner	Power strip	0.0031	12.0°	3.8 %	0.18
Photogrammetry	Power strip	0.0047	19.4°	4.1 %	0.25
Fast-Asset	Power strip	0.0011	30.0°	1.4 %	1.14

[a] Mean PCA planarity residual (σ): lower values indicate smoother local surfaces.
[b] Mean angular deviation between neighbouring normals: lower values indicate more consistent surface orientation.
[c] Percentage of points classified as outliers by statistical outlier removal (SOR, k=20, α=2).
[d] Coefficient of variation of local k-NN distances: lower values indicate more uniform point density.

non-specialist operators can maintain the digital twin autonomously.

Future work will focus on improving detection robustness and extending applicability. Expanding and refining the dataset, including negative examples, is expected to improve recall and precision. Extending the pipeline to fully multiclass, parallel processing and integrating advanced segmentation models are also planned. Additionally, a full point-cloud-to-IFC conversion workflow, fitting parametric primitives to the cleaned clouds and exporting them via IfcOpenShell, will be developed to close the gap between BIM-ready point clouds and native BIM elements. In-situ visual guidance via headset display and evaluation on wall-mounted MEP fixtures in operational environments are further planned extensions.

Acknowledgments

This research was funded by the Departamento de Desarrollo Económico, Sostenibilidad y Medio Ambiente of the Basque Government under the ELKARTEK programme (grant KK-2022/00065, LANVERSO project).

References

Abreu, N., Pinto, A., Matos, A., and Pires, M. (2023). Procedural point cloud modelling in scan-to-bim and scan-vs-bim applications: A review. ISPRS International Journal of Geo-Information, 12(7):260.

Boje, C., Hahn Menacho, J., Marvuglia, A., Benetto, E., Kubicki, S., Schaubroeck, T., and Navarrete Gutiérrez, T. (2023). A framework using BIM and digital twins in facilitating LCSA for buildings. Journal of Building Engineering, 76:107232.

Chen, S., Gao, F., and Li, J. (2023). Improving completeness and accuracy of 3d point clouds by using deep learning for applications of digital twins to civil structures. Advanced Engineering Informatics, 58:102196.

Dibene, J. C. and Dunn, E. (2022). Hololens 2 sensor streaming.

Ding, Y., Yin, M., Wei, R., Brilakis, I., Liu, M., and Luo, X. (2024). Scan-to-BIM for as-built roads: Automatic road digital twinning from semantically labeled point cloud data. arXiv preprint 2406.12404.

Duan, Z., Shao, J., Zhang, M., Zhang, J., and Zhai, Z. (2024). A small-object-detection algorithm based on LiDAR point-cloud clustering for autonomous vehicles. Sensors, 24(16):5423.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In Knowledge Discovery and Data Mining.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P. (2015). Microsoft coco: Common objects in context.

Navares-Vázquez, J. C., Balado, J., Arias, P., and Díaz-Vilariño, L. (2023). Virtual seeds for indoor point cloud segmentation with HoloLens 2. SSRN preprint.

Pan, Y., Braun, A., Brilakis, I., and Borrmann, A. (2022). Enriching geometric digital twins of buildings with small objects by fusing laser scanning and AI-based image recognition. Automation in Construction, 140:104375.

Wijeratne, P. U., Gunarathna, C., Yang, R. J., Wu, P., Hampson, K., and Shemery, A. (2023). Bim enabler for facilities management: A review of 33 cases. International Journal of Construction Management.

Yilmaz, G., Akcamete, A., and Demirors, O. (2023). BIM-CAREM: Assessing the BIM capabilities of design, construction and facilities management processes in the construction industry. Computers in Industry, 147:103861.

Zhou, Q.-Y., Park, J., and Koltun, V. (2018). Open3D: A modern library for 3D data processing. arXiv:1801.09847.

Zhuang, Z., Zhi, Z., Han, T., Chen, Y., Chen, J., Wang, C., et al. (2024). A survey of point cloud completion. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, PP:1–22.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATED DIMENSION-AWARE 2D-TO-BIM RECONSTRUCTION THROUGH CROSS-MODAL TEXT-GEOMETRY ALIGNMENT

Xing Liang[1], Tomohiro Fukuda[1], and Nobuyoshi Yabuki[2,3]
[1]Division of Sustainable Energy and Environmental Engineering, Graduate School of Engineering, The University of Osaka, Osaka, 565-0871, Japan
[2]Advanced Research Laboratories, Tokyo City University, Setagaya-ku, Tokyo, Japan
[3]The University of Osaka, Suita, Osaka, Japan (Professor Emeritus)

Abstract

Digitising existing buildings supports energy-efficient retrofitting and lifecycle analysis, yet many are documented only by legacy drawings with dimensions embedded as textual annotations. Existing methods often rely on heuristic global scaling over metric fidelity. This paper presents a vision-based framework integrating optical character recognition (OCR), OCR-guided denoising, semantic segmentation, and grid-based alignment. The primary contribution is a cross-modal text–geometry alignment strategy that reconstructs a metric reference grid from annotations and links it to building geometry. A second contribution is an Industry Foundation Classes-based evaluation framework. Results demonstrate reliable recovery of absolute dimensions for dimension-aware BIM reconstruction.

Introduction

The building sector accounts for nearly 40% of global energy consumption (El Hafdaoui et al., 2023), highlighting the need to rapidly digitise existing building stock for energy-efficient retrofits, structural analysis, and lifecycle applications. Many advanced architecture, engineering, and construction (AEC) workflows rely on semantically rich and geometrically precise building information models (BIMs) (Schönfelder et al., 2023). However, most existing buildings predate BIM adoption and are documented primarily by legacy two-dimensional (2D) floorplan drawings (Lu and Lee, 2017). These drawings contain critical geometric, semantic, and dimensional information. Consequently, automated 2D-to-BIM reconstruction has become an important research topic (Lim et al., 2018).

Typical 2D-to-BIM pipelines involve floorplan analysis followed by BIM instantiation. While BIM generation can often be handled deterministically once reliable inputs are available, the floorplan analysis stage remains challenging. Many existing methods prioritise topological correctness over engineering-level geometric precision, often relying on heuristic global scaling or normalised coordinates that are insufficient for millimetre-sensitive BIM reconstruction (Wang and Pajarola, 2025).

Achieving engineering-grade precision requires not only detecting walls and openings, but also correctly interpreting dimensional annotations and associating them with the corresponding building geometry. This task is difficult because engineering drawings contain diverse drafting conventions, rotated or occluded texts, and annotation-induced noise (Pizarro et al., 2022).

Existing studies often assume that relative geometry is sufficient or use annotations only for coarse scale estimation, without establishing explicit component-level correspondences between dimension texts and building elements (Lee et al., 2025). Moreover, reconstruction quality is usually assessed with computer-vision-oriented metrics such as intersection over union (IoU), which are insensitive to millimetre-level geometric deviations. As a result, a model may appear visually correct while remaining inadequate for engineering use.

To address these limitations, this paper proposes an automated framework for dimension-aware 2D-to-BIM reconstruction from annotated raster floorplans. Its primary methodological contribution is a cross-modal text–geometry alignment strategy that reconstructs a metric reference grid from dimensional annotations and links it to segmentation-derived architectural geometry at the component level. Its second main contribution is an IFC-based evaluation framework for engineering-level assessment of reconstruction quality at the object and parameter level.

Related work

Research on 2D-to-BIM reconstruction for existing buildings generally follows a two-stage paradigm: (1) automated floorplan analysis, which extracts geometric and semantic information from legacy drawings, and (2) BIM instantiation, which converts these extracted entities into structured, parameterised building components. BIM instantiation is relatively mature and can often be handled deterministically once reliable inputs are available, using commercial platforms, programmatic APIs, or open-standard libraries (Lu et al., 2020). As a result, the main technical divergence among existing 2D-to-BIM approaches lies in floorplan analysis, where drawing format, quality, and interpretation strategy strongly affect reconstruction accuracy and robustness.

The performance of automated floorplan analysis is influenced by drawing format. Vector-based CAD drawings preserve editable information such as layers, blocks, and dimension objects, enabling rule-based or API-driven methods to achieve high-fidelity geometric reconstruction and component classification (Zhao et al., 2021). However, in practical engineering contexts, access to native CAD files is often restricted by software

dependencies, version incompatibilities, or confidentiality constraints (Dixit et al., 2019). As a result, floorplans are more commonly exchanged as raster images, such as scanned PDFs, which irreversibly lose embedded semantics and introduce noise, resolution variability, and drafting inconsistencies. These limitations have driven increasing interest in learning-based methods designed specifically for raster floorplans.

Existing research on raster-based floorplan analysis largely falls into three parallel streams: (1) geometric analysis, which reconstructs architectural structures such as walls and room boundaries; (2) symbolic object detection and recognition, which identifies standardised elements such as doors, windows, and fixtures; and (3) text detection and recognition, which extracts annotations including room names and dimensional values.

In geometric analysis, wall detection is typically treated as the core task and formulated as pixel-level semantic or instance segmentation to generate wall masks (Sun et al., 2025). More recent approaches attempt to directly predict vectorised primitives, such as corners and edges, using end-to-end models that produce machine-readable spatial layouts (Hu et al., 2024). Nevertheless, most geometric methods operate in normalised image coordinates or relative spatial representations, deriving absolute dimensions only through assumed global scale factors or omitting metric information altogether (Su et al., 2023). Although such representations are sufficient for layout understanding or navigation-oriented applications, they fall short of the millimetre-level geometric accuracy required for engineering-grade BIM reconstruction.

In parallel, symbolic object detection leverages general-purpose frameworks, such as YOLO variants, to localise and classify architectural symbols including doors, windows, and sanitary fixtures (Xu et al., 2024; Jamieson et al., 2025). These models typically output class-labelled bounding boxes or segmentation masks, enabling structured interpretation of floorplan content. However, they are often evaluated on simplified or schematic drawings with sparse or intentionally removed textual annotations. As a result, they are not designed to handle the dense and overlapping annotations commonly found in real-world engineering drawings, which limits their applicability to practical 2D-to-BIM scenarios.

In contrast, text-centric methods follow the standard scene text understanding pipeline, detecting text regions followed by character recognition, and have demonstrated success in handling arbitrarily oriented and stylised annotations (Schönfelder et al., 2024). Yet, their integration into geometry-aware reasoning remains limited. Textual annotations are frequently overlaid on or adjacent to walls, and auxiliary lines, such as dimension leaders, can substantially interfere with structural segmentation and geometric inference.

Consequently, most existing pipelines either ignore textual information entirely or use it only for coarse global scaling, without establishing explicit, component-level associations between dimensional annotations and building elements (Pokuciński and Mrozek, 2024). This reveals a critical gap: although geometric, symbolic, and textual analyses have each advanced substantially, they remain largely decoupled in practice. Existing workflows still rely heavily on heuristic assumptions, without explicitly linking dimensional annotations to building geometry or evaluating reconstruction quality at the engineering level. This limitation reduces metric fidelity and constrains the practical applicability of 2D-to-BIM reconstruction for engineering use.

Methodology

Framework overview

The proposed methodology introduces a multi-stage, vision-based framework for the automated analysis of annotated floorplans to achieve geometry-accurate 2D-to-BIM reconstruction. As illustrated in Figure 1, the framework integrates OCR, OCR-guided denoising, semantic segmentation, and geometric alignment to recover component-level dimensional information from legacy raster floorplans.

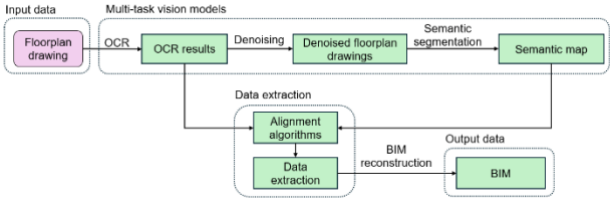


Figure 1: Overall workflow of proposed method

The methodology emphasises the cross-modal integration of heterogeneous visual cues, including textual annotations, auxiliary reference lines, and architectural geometry. This design addresses a critical bottleneck in AEC workflows, namely the lack of robust mechanisms for translating disparate recognition outputs into geometric representations.

The workflow takes raster floorplans as input and does not assume pre-cleaned or simplified documents, as shown in Figure 2(a). Textual and geometric information are processed in parallel and subsequently aligned through a dedicated data extraction and association stage, enabling staged BIM automation with dimensionally grounded component reconstruction.

OCR training and configuration

OCR is applied directly to the input floorplan to detect and recognise textual elements, including dimensional annotations. A two-stage OCR pipeline is employed, consisting of a CRAFT text detector (Baek et al., 2019) and a TPS-STN + CRNN recogniser (Shi et al., 2016) trained with CTC loss.

The OCR model is fine-tuned using 1,000 annotated floorplan images sampled from CubiCasa5k (Kalervo et al., 2019) and CVC-FP (de las Heras et al., 2015), with an 8:1:1 train/validation/test split; ResBIM (Liang et al., 2026) is used exclusively for external validation. Overlapping tiling and print-scan-style augmentations are adopted to handle large drawing canvases and small text

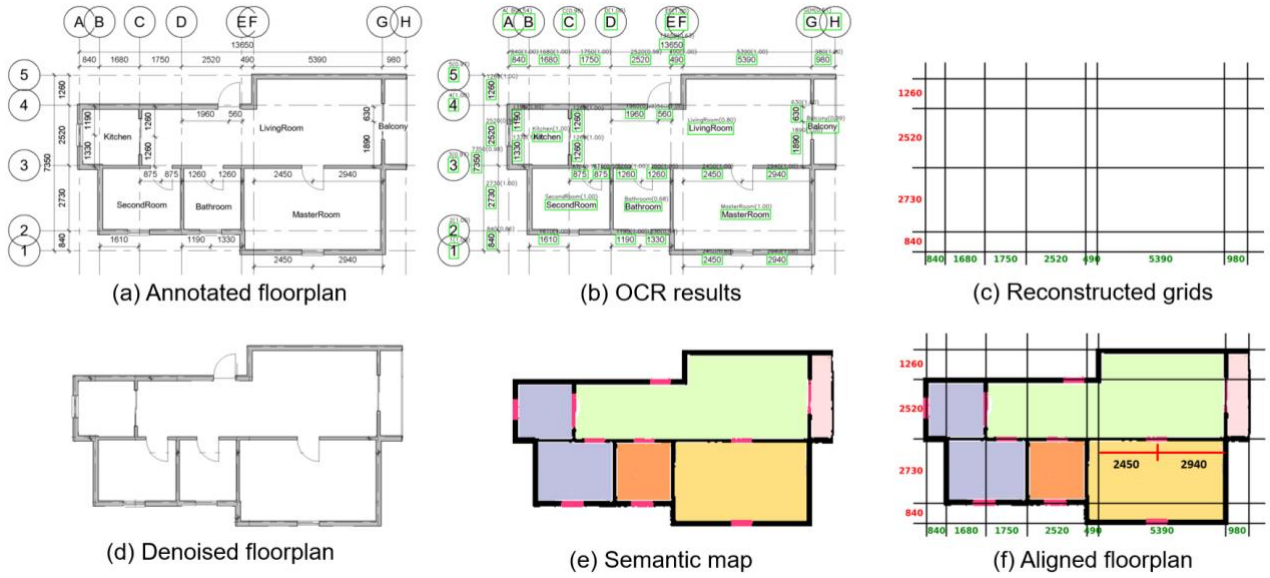


Figure 2: Overall process of multi-stage vision-based framework. (a) Input annotated floorplan; (b) OCR detection and recognition results; (c) reconstructed metric reference grid from dimensional annotations; (d) OCR-guided denoised floorplan; (e) semantic segmentation map of the denoised floorplan; and (f) final aligned floorplan obtained by aligning (c) and (e).

instances. Recognised texts are normalised and converted into structured representations, including numeric values, orientations, bounding-box coordinates, and confidence scores. These structured OCR outputs are treated as intermediate dimensional cues and are subsequently used in later cross-modal alignment and geometric reasoning stages.

OCR-guided denoising

Prior to semantic segmentation, an OCR-guided denoising step is applied to suppress annotation-induced artifacts that may interfere with geometric extraction. OCR outputs provide spatial and directional cues that localise regions associated with dimensional annotations. Within these regions, subtraction-based operations are applied to remove annotation lines and leaders that are geometrically aligned with detected text.

To further reduce non-structural artifacts, a thin-line removal procedure is applied to suppress auxiliary reference lines, including grids and leaders, which are typically rendered with thinner strokes than architectural components. Morphological and geometric filtering is used conservatively to preserve wall and opening continuity. The resulting denoised image retains the structural geometry required for segmentation while significantly reducing false positives caused by annotation-related noise, as shown in Figure 2(d).

Semantic segmentation

Semantic segmentation of architectural components is formulated as a multi-task learning problem that jointly models architectural boundaries and interior spatial regions. The adopted U-net encoder-decoder architecture (Ronneberger et al., 2015) employs a shared convolutional encoder and two task-specific decoder branches, enabling efficient feature sharing across related segmentation tasks.

The first decoder branch focuses on boundary segmentation, producing pixel-level masks for structural elements such as walls and openings. The second branch performs region-level semantic segmentation of interior spaces. To enhance structural consistency, boundary features are incorporated into the region prediction process through a boundary-guided attention mechanism at multiple feature scales. This design helps preserve thin geometric elements and enforces spatial alignment between room regions and their enclosing structures. Direction-aware feature aggregation is applied to better capture dominant architectural orientations.

The network is fine-tuned using the R2V (Liu et al., 2017) and R3D (Liu et al., 2015) datasets, which provide pixel-level annotations for both boundaries and regions. Training employs a multi-task, class-balanced loss to mitigate severe data imbalance between large interior regions and thin structural components. Although both tasks are retained during training to leverage shared structural priors, only the boundary-related outputs, specifically wall and opening masks, are propagated to downstream geometric alignment and BIM reconstruction.

In the proposed pipeline, semantic segmentation serves as a geometric sensing module rather than an end objective. Its role is to provide topology-consistent and geometrically reliable structural primitives that can be aligned with OCR-derived dimensional annotations, thereby supporting engineering-grade geometric reconstruction.

Alignment algorithms

The alignment stage aims to establish explicit and metrically consistent correspondences between dimensional annotations extracted by OCR and the architectural geometry obtained from semantic segmentation. Instead of assuming a global scale factor,

the proposed method reconstructs a metric reference grid directly from dimensional annotations embedded in the floorplan drawings. This grid serves as an intermediate geometric representation that bridges raster-level geometry and real-world distances, enabling component-level metric reconstruction.

As illustrated in Algorithm 1, the process begins by analysing OCR outputs, as shown in Figure 2(b), where each detected text instance is represented by its recognised string, bounding box, orientation, and confidence score. High-confidence numeric annotations with consistent orientation are first selected as candidates potentially associated with grid spacing. Drawing-convention-aware spatial heuristics are then applied to identify annotations that likely describe distances between parallel grid lines. Specifically, vertical grid spacings are inferred from horizontally oriented dimension texts whose bounding boxes lie at similar vertical positions, reflecting the common drafting practice in which distances between vertical elements are annotated using horizontal dimension strings placed between the corresponding grid lines.

Candidate annotations are subsequently grouped based on the proximity of their bounding-box coordinates and sorted according to spatial order. Adjacent annotations within each group are interpreted as describing consecutive grid intervals, yielding an ordered set of grid spacings associated with both pixel-domain distances and OCR-derived metric values.

Because OCR outputs may contain recognition errors, a local verification and correction mechanism is introduced prior to global consistency checks. The algorithm estimates a pixel-to-metric conversion ratio by computing the median ratio between OCR-reported distances and their corresponding pixel gaps measured from bounding-box positions or dimension leader extents. Each grid spacing is then re-evaluated by comparing its OCR-derived value with the geometry-inferred distance. When deviations exceed a predefined tolerance, OCR values are locally corrected using geometry-consistent estimates, thereby enforcing metric consistency at the grid-interval level.

After individual grid spacings are verified and corrected, the algorithm detects total grid-length annotations by comparing detected dimension values with cumulative grid lengths. Grid lines are then placed cumulatively according to the corrected distances, forming a globally consistent metric grid in real-world units, as illustrated in Figure 2(c). The semantic map is subsequently scaled along the horizontal and vertical axes to match the grid extent and aligned to the grid coordinate system, as demonstrated in Figures 2(e) and 2(f). Semantic labels for doors, windows, and room regions are transferred to the aligned floorplan, enabling component-level metric interpretation.

The current alignment strategy incorporates several drafting-convention-aware assumptions. It assumes that dimension texts associated with a given grid interval exhibit consistent orientation, preserve relatively regular spatial ordering, and remain sufficiently close to the referenced structural location for local grouping and verification to be meaningful. These assumptions are generally consistent with common engineering drafting practice. However, robustness may decrease for non-standard layouts, such as offset dimension texts, long auxiliary leader lines, overlapping dimension chains, or highly irregular annotation structures.

Algorithm 1: Grid alignment with OCR-guided verification

Input: OCR detections $\mathcal{O} = \{o_i\}_{i=1}^{N}$, $o_i = (s_i, b_i, \theta_i, c_i)$ where s_i is recognized text, $b_i = (x_i, y_i, w_i, h_i)$ is the bounding box, θ_i is text orientation, c_i is confidence; semantic segmentation map $\mathcal{S}$

Output: Metric vertical grid $\mathcal{G}^v$ and grid-aligned geometry

Step 1: Select candidate vertical grid spacings;
Initialize candidate set $\mathcal{C}^v \leftarrow \emptyset$;
foreach $o_i \in \mathcal{O}$ **do**
  if $c_i \geq \tau_c$ *and* $\theta_i = 0$ **then**
    $\mathcal{C}^v \leftarrow \mathcal{C}^v \cup \{o_i\}$; // horizontal text, high confidence

Step 2: Group candidates by vertical alignment;
Partition $\mathcal{C}^v$ into groups $\{\mathcal{C}_k^v\}$ such that $|y_i - y_j| < \tau_y$ for $o_i, o_j \in \mathcal{C}_k^v$; // similar height

Step 3: Extract vertical grid spacings;
Initialize spacing set $\Delta^v \leftarrow \emptyset$;
foreach *group* $\mathcal{C}_k^v$ **do**
  Sort $\mathcal{C}_k^v$ by x coordinate;
  for $j = 1$ **to** $|\mathcal{C}_k^v| - 1$ **do**
    $\delta x_j = |x_{j+1} - x_j|$; // pixel distance between adjacent grid leaders
    $d_j^{ocr} = \mathrm{val}(s_j)$; // OCR-derived spacing value (metric)
    $\Delta^v \leftarrow \Delta^v \cup \{(\delta x_j, d_j^{ocr})\}$; // store pixel gap and OCR distance

Step 4: Verify and correct OCR-derived distances;
Estimate pixel-to-metric ratio $\rho = \mathrm{median}\left(\frac{d_j^{ocr}}{\delta x_j}\right)_{(\delta x_j, d_j^{ocr}) \in \Delta^v}$; // distance per pixel estimated from grid spacings
foreach $(\delta x_j, d_j^{ocr}) \in \Delta^v$ **do**
  $d_j^{px} = \rho \cdot \delta x_j$; // distance inferred from pixel gap
  if $|d_j^{px} - d_j^{ocr}| > \epsilon$ **then**
    $d_j^{ocr} \leftarrow d_j^{px}$; // correct OCR error using geometric consistency

Step 5: Detect total vertical grid length (if available);
Compute cumulative length $L^v = \sum_{(\delta x_j, d_j^{ocr}) \in \Delta^v} d_j^{ocr}$; // sum of verified vertical grid spacings
foreach $o_k \in \mathcal{O}$ **do**
  if $|val(s_k) - L^v| < \epsilon$ **then**
    Mark o_k as total vertical grid length annotation;

Step 6: Construct metric vertical grid and align geometry;
Construct vertical grid $\mathcal{G}^v = \mathrm{cumsum}(\{d_j^{ocr}\}_{(\delta x_j, d_j^{ocr}) \in \Delta^v})$; // metric grid positions
Align semantic segmentation map $\mathcal{S}$ to $\mathcal{G}^v$; // assign coordinates to segmented map

Finally, the aligned geometry and associated semantic information are used to instantiate building elements as IFC entities. The resulting BIM preserves the topological relationships derived from semantic segmentation while incorporating explicit, grid-consistent dimensional information. By using the reconstructed grid as an intermediary between textual annotations and raster geometry, the proposed alignment strategy reliably recovers absolute measurements without relying on heuristic global scaling, thereby providing a robust metric foundation for accurate 2D-to-BIM reconstruction and downstream BIM-based analysis.

Results and discussion

Quantitative results

The quantitative evaluation addresses three complementary aspects: OCR recognition performance, semantic segmentation accuracy, and engineering-level BIM reconstruction quality.

For OCR evaluation, commonly used metrics in computer vision and document analysis include the character error rate, CER, and string similarity measures such as the FuzzyWuzzy ratio. These metrics assess character-level recognition accuracy and are sensitive to minor transcription errors, making them suitable for algorithm

tuning and benchmarking. However, they are not well suited to engineering-oriented reconstruction tasks. In the context of dimensional annotations, even a single-digit error may cause substantial geometric distortion in the reconstructed BIM, making partial string similarity inadequate for downstream evaluation.

Therefore, a strict OCR accuracy metric is adopted to reflect whether recognised dimension values are directly usable for geometric reconstruction. OCR accuracy is defined as:

$$OCR\ accuracy = \frac{N_{correct}}{N_{GT}} \quad (1)$$

where $N_{correct}$ is the number of correctly recognised dimension annotations and N_{GT} is the total number of ground-truth dimension annotations. A prediction is considered correct only if the numeric value is fully matched. Under this strict criterion, the OCR module achieves an accuracy of 90.2% on the test set and 85.7% on the ResBIM dataset, indicating stable cross-dataset generalisation.

Semantic segmentation performance is evaluated using the IoU metric:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

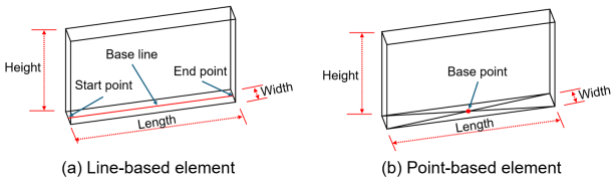
where A and B represent the predicted and ground-truth masks, respectively. The segmentation model achieves an IoU of 92.1% for wall and opening classes, demonstrating reliable extraction of structural geometry from denoised raster floorplans.

IFC-based evaluation

Although OCR accuracy and IoU are standard evaluation metrics in computer vision, they do not directly reflect reconstruction effectiveness at the engineering level, where the geometric correctness of BIM elements is critical. To address this gap, IFC-based evaluation metrics are introduced to assess reconstruction quality at the object and parameter levels.

The ResBIM dataset provides paired annotated floorplans and corresponding reference BIM models, enabling one-to-one comparison between reconstructed and ground-truth elements. Both reconstructed and reference BIMs are converted into IFC format, which provides a platform-independent representation in which each architectural element is encoded as an object with placement and geometric parameters. This object-centric representation enables direct and interpretable element-level comparison. Although this dataset supports rigorous IFC-level comparison, it does not cover the full range of drafting conventions encountered in practice. Therefore, the present evaluation should be interpreted as an initial engineering-level validation under the currently available benchmark setting.

Because the input data consist of 2D floorplans, the evaluation focuses on planar reconstruction accuracy. For walls, the origin location corresponds to the start point of the wall baseline, as shown in Figure 3(a), and the length is defined by the baseline length encoded in the IFC geometry. For doors and windows, the origin is defined by the opening base point, as shown in Figure 3(b), and dimensional parameters are retrieved from the corresponding IFC geometric definitions.



(a) Line-based element (b) Point-based element

Figure 3: Two methods for representing element locations in IFC-based BIM models in the ResBIM dataset

To comprehensively evaluate geometry-accurate 2D-to-BIM reconstruction at the engineering level, three complementary IFC-based metrics are proposed to assess numerical completeness, dimensional correctness, and strict geometric accuracy.

The count error Err_{cnt} measures numerical consistency between reconstructed and ground-truth models.

$$Err_{cnt} = \frac{|N_{pred} - N_{gt}|}{N_{gt}} \quad (3)$$

where N_{pred} and N_{gt} denote the numbers of reconstructed and ground-truth IFC elements, respectively.

The dimension-level precision metric $IFC_{\text{dim_precision}}$ evaluates whether element dimensions are correctly reconstructed, independent of absolute location:

$$IFC_{dim_precision} = \frac{TP_{dim}}{TP_{dim} + FP_{dim}} \quad (4)$$

where TP_{dim} refers to elements whose dimensional parameters are reconstructed within a predefined tolerance, and FP_{dim} corresponds to elements whose dimensions exceed this threshold. This metric evaluates dimensional correctness independently of absolute spatial alignment and is particularly sensitive to OCR and dimension-geometry association errors.

Finally, the strict geometric precision metric $IFC_{precision}$ assesses whether both dimensional parameters and origin locations are simultaneously correct:

$$IFC_{precision} = \frac{TP}{TP + FP} \quad (5)$$

where TP denotes elements whose dimensions and origin locations are simultaneously reconstructed within tolerance, and FP denotes elements failing either criterion. This metric reflects whether reconstructed elements are directly usable for downstream BIM applications without manual correction.

A tolerance of 1 mm is applied to wall dimensions and locations only to absorb minor numerical inconsistencies introduced by IFC export, unit normalisation, and floating-point conversion, rather than to relax geometric evaluation. For openings, a larger tolerance of 20 mm is adopted as a benchmark-specific evaluation threshold. In ResBIM, openings are annotated by centre-point distances rather than explicit physical widths, so opening dimensions and positions must be inferred indirectly from raster geometry and host-wall context before being

converted into IFC opening parameters. This introduces higher annotation and representation uncertainty than for walls, whose baselines and lengths are more directly defined. Because openings are also smaller elements, they are more sensitive to minor segmentation or alignment shifts. The 20 mm threshold therefore accounts for annotation granularity and benchmark-specific uncertainty, rather than implying a looser engineering tolerance.

Experimental results on a randomly selected subset of 200 floorplans from ResBIM are presented in Table 1. The count error Err_{cnt} is 2.45%, indicating high numerical consistency between reconstructed and ground-truth models. At the IFC entity level, the dimension-level precision $IFC_{\text{dim_precision}}$ is 82.4%, and the strict geometric precision $IFC_{precision}$, which requires both correct dimensions and locations, is 71.1%. Notably, in 62 of the 200 floorplan reconstruction tasks, all building components are correctly predicted in both location and dimensional attributes. The average processing time per floorplan is approximately 6.5 minutes.

Table 1: Quantitative evaluation results of the proposed dimension-aware 2D-to-BIM reconstruction framework

Metric	Value
Err_{cnt}	2.5%
$IFC_{\text{dim_precision}}$	82.4%
$IFC_{precision}$	71.1%
Exact dimension match (count)	62/200
Total processing time	21.5 hours

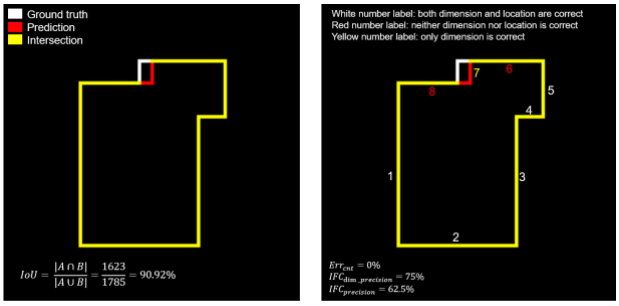
Error analysis and discussion

This study evaluates the proposed vision-based 2D-to-BIM reconstruction framework using both conventional computer vision metrics and the proposed IFC-based evaluation metrics. The OCR module achieves 90.2% accuracy on the in-domain test split and 85.7% on the ResBIM dataset for external validation, while the semantic segmentation model achieves an IoU of 92.1%. These results indicate strong performance when each visual task is evaluated independently using standard computer vision criteria.

However, when the same outputs are assessed using IFC-based reconstruction metrics, a noticeable performance drop is observed. The dimension-level accuracy and strict geometric precision reach only 82.4% and 71.1%, respectively, even after applying the pixel-level correction strategy described in Algorithm 1. This discrepancy reflects a fundamental difference between computer-vision-oriented evaluation and engineering-oriented BIM reconstruction requirements.

Conventional metrics such as OCR accuracy and IoU evaluate predictions in isolation, assessing each text string or pixel region independently. In contrast, BIM reconstruction is inherently relational and connectivity-driven. Walls act as hosts for openings, adjacent walls are expected to connect seamlessly, and local geometric inconsistencies can propagate through topological relationships. As a result, dimensional values extracted by OCR are not independent in the reconstruction context but mutually constraining.

Figure 4 provides a representative qualitative example of this discrepancy. In this case, the predicted wall mask achieves an IoU of 90.92% against the ground truth, and the reconstructed model contains the correct number of wall entities, 8 versus 8. However, only 6 of the 8 wall entities are dimensionally correct, yielding a dimension-level accuracy of 75%, and only 5 of the 8 are correct in both dimension and location, yielding a strict geometric precision of 62.5%. This example illustrates why strong image-level overlap does not necessarily translate into equally high IFC-level correctness.



(a) Pixel-based exterior-wall prediction result and IoU (b) Vectorised exterior-wall and IFC-based metric calculation

Figure 4: Representative exterior-wall example illustrating the difference between conventional computer vision metrics and engineering-oriented IFC metrics.

IFC-based evaluation is stricter because it requires each reconstructed entity to satisfy dimensional correctness, spatial placement, and geometric consistency simultaneously. Therefore, small OCR assignment errors or minor boundary shifts that have little influence on OCR accuracy or IoU may still cause reconstructed IFC entities to be judged incorrect once their baselines, origins, or dimensions exceed the evaluation tolerance. These effects are further amplified during BIM instantiation because local errors do not remain isolated at the image level but propagate through dependent geometric relationships. This explains both the drop from the higher OCR and IoU scores to lower IFC-based metrics and the additional gap between dimension-level precision, 82.4%, and strict geometric precision, 71.1%. In other words, conventional computer vision metrics tend to overestimate reconstruction quality in dimension-sensitive BIM workflows because they do not capture entity-level placement errors and topology-dependent error propagation.

Limitations

Despite its advantages, the proposed pipeline has several limitations. First, the alignment module relies on drafting-convention-aware spatial assumptions, including locally interpretable annotation placement, consistent text orientation, and regular interval ordering. Its robustness may therefore decrease for drawings with highly irregular annotation layouts, overlapping dimension chains, or long auxiliary leader lines that displace dimension texts far

from the referenced geometry. In addition, the alignment algorithm degrades noticeably when OCR accuracy falls below approximately 50%, as may occur in highly cluttered or severely occluded drawings. To maintain computational efficiency, the semantic segmentation model also operates on downsampled inputs, which may amplify geometric distortions in the semantic map and reduce alignment quality.

Second, IFC-based evaluation is constrained by the limited availability of datasets that provide paired floorplans and reference BIM models (Pizarro et al., 2022; Liang et al., 2026). ResBIM enables direct object-level comparison, but relying primarily on a single benchmark limits the diversity of drafting conventions and annotation styles represented in the evaluation. Accordingly, the reported results should be interpreted as benchmark-based evidence rather than a complete demonstration of robustness across all real-world floorplan types.

Third, the current study is limited to planar reconstruction from floorplans and does not recover elevation-related dimensions or richer semantic attributes from multi-view drawings. In addition, although the pipeline automates several reconstruction stages, the present work does not directly quantify labour or time savings through a user study or a comparison against manual or semi-automatic workflows.

Conclusions

This paper presents a multi-stage, vision-based framework for the automated analysis of annotated architectural floorplans to enable dimension-aware 2D-to-BIM reconstruction. The main methodological contribution is the proposed cross-modal text–geometry alignment strategy, which reconstructs a metric reference grid from dimensional annotations and supports component-level metric interpretation. A second key contribution is the IFC-based evaluation framework, which enables direct object-level assessment of reconstruction quality from an engineering perspective.

The results show that conventional computer vision metrics, such as OCR accuracy and segmentation IoU, are insufficient for characterising engineering-level BIM reconstruction quality. Although the proposed workflow supports automated, dimension-aware reconstruction from annotated raster floorplans, the present study remains limited to planar, benchmark-based validation and does not directly quantify labour or time savings through user studies or comparisons with manual or semi-automatic workflows.

Future directions

Several limitations remain and motivate future research. First, the current framework focuses on planar reconstruction from floorplans and does not recover elevation-related dimensions or richer semantic attributes from multi-view drawings. Future work should therefore incorporate additional drawing types, such as elevations and sectional views, to support more complete BIM reconstruction.

Second, the present alignment strategy remains partly rule-based and relies on drafting-convention-aware assumptions. Learning-based alignment methods trained on annotated geometry–text correspondences may improve robustness to OCR errors, irregular annotation layouts, and heterogeneous drawing styles.

Third, the current IFC-level evaluation is constrained by the limited availability of paired floorplan–BIM datasets. Extending the evaluation to more diverse real-world drawings and additional benchmarks will be important for testing generalisability across different drafting conventions and building standards.

Together, these directions support more comprehensive and reliable automation of existing-building digitisation in real-world AEC workflows.

Acknowledgements

This research was partially supported by the Telecommunications Advancement Foundation's overseas travel expense assistance.

References

Baek, Y., Lee, B., Han, D., Yun, S. & Lee, H. (2019) Character Region Awareness for Text Detection. arXiv [Preprint].

Dixit, M.K., Venkatraj, V., Ostadalimakhmalbaf, M., Pariafsai, F. & Lavy, S. (2019) Integration of facility management and building information modeling (BIM). Facilities, 37(7/8), pp. 455–483.

El Hafdaoui, H., Khallaayoun, A. & Ouazzani, K. (2023) Activity and efficiency of the building sector in Morocco: A review of status and measures in Ifrane. AIMS Energy, 11(3), pp. 454–485.

Hu, S., Wu, W., Su, R., Hou, W., Zheng, L. & Xu, B. (2024) Raster-to-Graph: Floorplan Recognition via Autoregressive Graph Prediction with an Attention Transformer. Computer Graphics Forum, 43(2).

Jamieson, L., Moreno-Garcia, C.F. & Elyan, E. (2025) Towards fully automated processing and analysis of construction diagrams: AI-powered symbol detection. International Journal on Document Analysis and Recognition, 28(1), pp. 71–84.

Kalervo, A., Ylioinas, J., Häikiö, M., Karhu, A. & Kannala, J. (2019) Cubicasa5k: A dataset and an improved multi-task model for floorplan image analysis. In Image Analysis: 21st Scandinavian Conference, SCIA 2019, Norrköping, Sweden, June 11–13, 2019, Proceedings 21. Springer, Cham, pp. 28–40.

de las Heras, L.-P., Terrades, O.R., Robles, S. & Sánchez, G. (2015) CVC-FP and SGT: a new database for structural floor plan analysis and its groundtruthing tool. International Journal on Document Analysis and Recognition (IJDAR), 18, pp. 15–30.

Lee, H., Jang, S., Lee, J., David Jeong, H. & Lee, G. (2025) Automated BIM generation from inconsistent

multi-view raster architectural drawings with missing dimensions. SSRN Preprint [Preprint].

Liang, X., Yabuki, N. & Fukuda, T. (2026) Fully automated synthetic BIM dataset generation using a deep learning-based framework. Automation in Construction, 181, p. 106584.

Lim, J., Janssen, P. & Stouffs, R. (2018) Automated generation of BIM models from 2D CAD drawings. In CAADRIA 2018 - 23rd International Conference on Computer-Aided Architectural Design Research in Asia: Learning, Prototyping and Adapting. The Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), pp. 61–70.

Liu, C., Schwing, A.G., Kaustav Kundu, ⇤–, Urtasun, R. & Fidler, S. (2015) Rent3D: Floor-Plan Priors for Monocular Layout Estimation. In In Computer Vision and Pattern Recognition (CVPR), Boston, 2015.

Liu, C., Wu, J., Kohli, P. & Furukawa, Y. (2017) Raster-to-vector: Revisiting floorplan transformation. In Proceedings of the IEEE International Conference on Computer Vision, pp. 2195–2203.

Lu, Q., Chen, L., Li, S. & Pitt, M. (2020) Semi-automatic geometric digital twinning for existing buildings based on images and CAD drawings. Automation in Construction, 115, p. 103183.

Lu, Q. & Lee, S. (2017) Image-Based Technologies for Constructing As-Is Building Information Models for Existing Buildings. Journal of Computing in Civil Engineering, 31, p. 4017005.

Pizarro, P.N., Hitschfeld, N., Sipiran, I. & Saavedra, J.M. (2022) Automatic floor plan analysis and recognition. Automation in Construction, 140, p. 104348.

Pokuciński, S. & Mrozek, D. (2024) Methods and Applications of Space Understanding in Indoor Environment—A Decade Survey. Applied Sciences (Switzerland). Multidisciplinary Digital Publishing Institute (MDPI).

Ronneberger, O., Fischer, P. & Brox, T. (2015) U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241.

Schönfelder, P., Aziz, A., Faltin, B. & König, M. (2023) Automating the retrospective generation of As-is BIM models using machine learning. Automation in Construction, 152, p. 104937.

Schönfelder, P., Stebel, F., Andreou, N. & König, M. (2024) Deep learning-based text detection and recognition on architectural floor plans. Automation in Construction, 157, p. 105156.

Shi, B., Wang, X., Lyu, P., Yao, C. & Bai, X. (2016) Robust Scene Text Recognition with Automatic Rectification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4168–4176.

Su, J.-W., Tung, K.-Y., Peng, C.-H., Wonka, P. & Chu, H.-K. (2023) SLIBO-Net: Floorplan Reconstruction via Slicing Box Representation with Local Geometry Regularization. In Advances in Neural Information Processing Systems, 2023.

Sun, Q., Fang, C., Liu, S., Sun, Y., Shang, Y. & He, Y. (2025) PolyGraph: A Graph-Based Method for Floorplan Reconstruction From 3D Scans. IEEE Transactions on Visualization and Computer Graphics, 31(10), pp. 7350–7362.

Wang, S. & Pajarola, R. (2025) Eliminating Rasterization: Direct Vector Floor Plan Generation with DiffPlanner. IEEE Transactions on Visualization and Computer Graphics, 31.

Xu, Z., Jha, N., Mehadi, S. & Mandal, M. (2024) Multiscale object detection on complex architectural floor plans. Automation in Construction, 165.

Zhao, Y., Deng, X. & Lai, H. (2021) Reconstructing BIM from 2D structural drawings for existing buildings. Automation in Construction, 128, p. 103750.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A BIM-INTEGRATED ADAPTIVE SURROGATE FRAMEWORK FOR EFFICIENT UNCERTAINTY QUANTIFICATION IN BUILDING ENERGY SIMULATION

Arash B. Naamrudi[1], Reihaneh Aghamolaei[1]
[1] School of Mechanical & Manufacturing Engineering, Faculty of Engineering and Computing, Dublin City University, Whitehall, Dublin 9, Ireland

Abstract

Uncertainty analysis is critical for reliable building energy predictions, yet classical Monte Carlo simulation is computationally prohibitive for detailed physics-based models. This study proposes a BIM-integrated adaptive surrogate uncertainty analysis framework combining adaptive sampling with sparse Polynomial Chaos Expansion to efficiently propagate uncertainty and perform global sensitivity analysis. The approach is demonstrated on four Irish residential archetypes with uncertain envelope thermal parameters. The adaptive surrogates achieve stable accuracy (CV-RMSE 0.74–2.05 kWh/m² for annual EUI) while reducing computational cost by around 99% compared to Monte Carlo simulation, enabling scalable and interpretable decision support for building design and retrofit.

Introduction

Buildings are a significant contributor to global energy use and carbon emissions, accounting for about 34–37% of total energy consumption worldwide (United Nations Environment Programme, 2024). In the European Union, residential buildings consume nearly 40% of total energy and are responsible for roughly 33% of energy-related greenhouse gas emissions, largely due to an ageing and inefficient building stock with low renovation rates (Energy Performance of Buildings Directive, 2025). Improving the energy performance of existing buildings is therefore critical to reducing energy demand and achieving a fully decarbonised building stock by 2050 (Energy Performance of Buildings Directive, 2025). In response, energy standards such as ASHRAE Standard 90.1 (ASHRAE Standard, 2022) and the International Energy Conservation Code (IECC) (International Code Council (ICC), 2021) have progressively tightened efficiency requirements, increasing the need for accurate yet computationally efficient Building Energy Models (BEMs) to support compliance, retrofit evaluation, and design optimisation (ElSayed, Shultz and Kurtz, 2025). BEM is a computational simulation approach used to predict the energy performance of buildings under defined operating conditions. The typical inputs to a BEM can be broadly categorised into five groups, including geometric and spatial parameters, envelope thermal properties, occupancy and internal load schedules, HVAC system parameters and boundary conditions. The primary outputs of interest in building energy analysis include the annual and monthly Energy Use Intensity (EUI, in kWh/m²), total heating and cooling demand (kWh), peak load (kW), and thermal comfort indicators.

Building Information Modelling (BIM) has increasingly been integrated with BEM to improve data consistency and automation in energy analysis (Heydari and Heravi, 2023). BIM provides a structured, data-rich representation of buildings, incorporating geometry alongside material properties and performance attributes (Shadram *et al.*, 2016; Wang *et al.*, 2022), which can be translated into energy simulations (Arbulu, Oregi and Etxepare, 2024; Forth and Borrmann, 2024). Within BIM-based energy modelling workflows, building performance is most commonly evaluated using deterministic simulation approaches, in which all model inputs are assigned fixed, single-point values (Alwan *et al.*, 2021; Heydari and Heravi, 2023). These values are typically derived from design specifications, standards, or nominal assumptions, implicitly assuming that building components behave exactly as prescribed and that variability arising from construction quality, ageing, occupant behaviour, and environmental conditions is negligible (Liu *et al.*, 2022; Prataviera *et al.*, 2022). While deterministic BIM-based simulations provide a streamlined pathway from design data to energy performance prediction, they are inherently limited by their reliance on single-point input assumptions, which obscure the effects of uncertainty and variability in building parameters.

To address this limitation and consider variability and uncertainty in building parameters, uncertainty-aware modelling frameworks have been developed that explicitly represent uncertain inputs using probabilistic or stochastic descriptions and propagate their effects through building energy models. Within this context, Forward Uncertainty Analysis (FUA) defines the analytical task of quantifying how uncertainty in model inputs translates into uncertainty in energy performance outputs.

FUA can be implemented using a range of numerical and surrogate-based techniques, including Monte Carlo simulation, Latin Hypercube sampling, and surrogate models such as Polynomial Chaos Expansion or Gaussian process regression. Yet, existing approaches face significant computational constraints since they require a high number of simulations over fixed and predefined samples. For example, Monte Carlo Simulation (MCS), the conventional benchmark for FUA, propagates

uncertainties through repeated random sampling (Liu *et al.*, 2020; Chadly *et al.*, 2023; Prilliman *et al.*, 2023). Although theoretically robust, MCS converges slowly, with accuracy improving only with the inverse square root of sample size, often requiring thousands of simulations (Zhao, Lee and Augenbroe, 2016). Given the high computational cost of individual BEM evaluations, this renders MCS impractical for many real-world applications.

To improve the efficiency MCS, static sampling strategies such as Latin Hypercube Sampling (LHS) have been proposed. LHS can reduce the required number of simulations by approximately 50% compared to MCS for similar accuracy (Tian *et al.*, 2020; Koniorczyk *et al.*, 2022). However, LHS allocates samples uniformly across the input domain, irrespective of model sensitivity or non-linearity, leading to inefficient use of computational resources in complex regions of the input space. Adaptive sampling strategies address this limitation by sequentially selecting new samples based on output variance, achieving reported simulation reductions of 25–68% relative to static designs (Yu *et al.*, 2025). Despite these gains, adaptive sampling becomes challenging when applied to high-fidelity BEMs with multiple outputs and temporal resolutions, where the dimensionality of objectives significantly increases computational burden.

Surrogate modelling offers a pathway to overcoming these computational barriers by approximating expensive BEM simulations with inexpensive metamodels trained on limited high-fidelity data (Zhao, Lee and Augenbroe, 2016; Sood *et al.*, 2025). Among surrogate approaches, Polynomial Chaos Expansion (PCE) is particularly attractive due to its explicit polynomial representation of input–output relationships, enabling analytical computation of statistical moments and Sobol' sensitivity indices directly from model coefficients (Tian, 2024). This analytical structure provides both uncertainty and sensitivity information within a single framework and enhances interpretability through direct examination of polynomial terms. However, classical PCE suffers from the curse of dimensionality, as the number of basis terms grows combinatorially with input dimension and polynomial order, making it impractical for building models with moderate to high numbers of uncertain parameters (Thapa, Mulani and Walters, 2020).

To address this challenge, adaptive PCE methods have been developed that iteratively refine the polynomial basis by retaining only influential terms and pruning negligible contributions, reducing required simulations by 40–60% compared to full-order PCE (Cheng and Lu, 2018; Thapa *et al.*, 2024). When combined with sequential sampling strategies, adaptive PCE further improves efficiency by allowing both the surrogate structure and training data to evolve based on observed model behaviour (Thapa, Mulani and Walters, 2020). Despite these advances, adaptive PCE has not been systematically integrated into BIM-based BEM workflows, particularly for simultaneous multi-output and multi-timescale analysis.

The literature, therefore, reveals three critical gaps. First, no existing framework unifies adaptive sampling and adaptive PCE within automated BIM-to-BEM workflows, with current approaches relying on either computationally expensive Monte Carlo methods or static surrogate training designs. Second, most uncertainty and sensitivity studies focus on single outputs and single temporal scales, failing to capture how parameter importance varies across different performance metrics and timescales, a limitation that can misinform retrofit decisions. Third, the scalability of adaptive PCE for multi-output, multi-timescale building energy analysis remains largely unvalidated. Addressing these gaps, this study proposes a unified BIM–BEM framework that integrates adaptive sampling with adaptive PCE to efficiently quantify uncertainty and sensitivity across multiple outputs and temporal resolutions and benchmarks its performance against Monte Carlo simulation.

Methodology

BIM-based BEM Transformation

The proposed framework (Figure 1) integrates BIM-based building modelling with adaptive surrogate-assisted uncertainty and sensitivity analysis. A BEM is first derived from a BIM representation, focusing on envelope-related parameters, including wall, roof, and window U-values and window solar heat gain coefficient (SHGC). These parameters are treated as uncertain inputs and assigned probability distributions. An adaptive PCE surrogate, coupled with sequential sampling, is then constructed to propagate uncertainties and evaluate their influence on multiple performance metrics. The framework employs a two-level adaptivity strategy, combining variance-based basis refinement with sequential experimental design. At each iteration, surrogate accuracy is improved through targeted basis enrichment and adaptive sampling, while convergence is assessed using distribution-based criteria.

The BIM-based building model was developed by identifying geometric and non-geometric parameters that define the physical and thermal characteristics of the building, including story configuration, ceiling height, openings, construction assemblies, and thermal properties. Parameter values were obtained from literature and established databases to ensure representativeness. Autodesk Revit was selected as the BIM authoring platform due to its parametric data structure, interoperability, and industry adoption. Thermal attributes such as U-values, layer thicknesses, and SHGC were extracted directly using a single-step BIM-to-BEM approach to minimise data loss and improve reliability.

To bridge BIM and energy simulation domains, a transitional energy model was generated in HBJSON format using the Pollination plugin. This intermediate representation preserves geometry, zoning, and aperture definitions while enabling quality control of spatial and thermal data. Aggregating spaces with similar thermal behaviours defined thermal zones. Fenestration elements were mapped directly from Revit window families, and

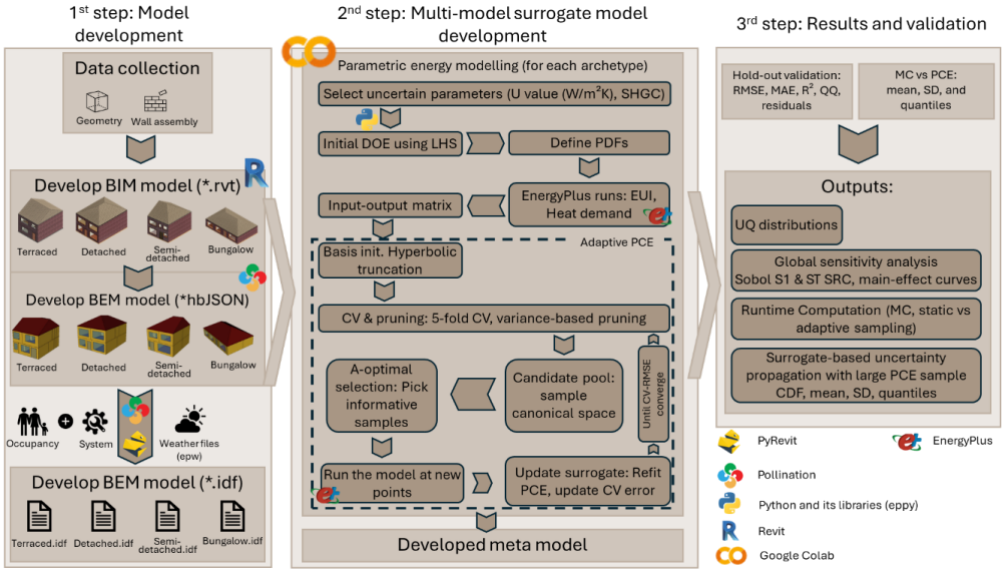


Figure 1. Overarching methodology diagram of the proposed framework.

boundary locations were assigned to wall centrelines to ensure surface alignment and shared interfaces. The transitional model was subsequently validated to correct envelope discontinuities, geometric gaps, and orthogonality issues. Once validated, the model was exported in EnergyPlus IDF format and enriched with material properties, weather data (EPW), occupancy schedules, HVAC parameters, and internal loads, forming the final simulation-ready BEM (Figure 2).

Adaptive Uncertainty Quantification Framework

Uncertainty in envelope parameters was propagated through the BEM using an adaptive PCE framework. Uncertain inputs were modelled as independent normally distributed random variables and transformed into standard normal space to satisfy orthogonality requirements. The building energy response was approximated using a truncated generalised PCE, where multivariate orthogonal basis functions are constructed as tensor products of univariate polynomials. Statistical moments of the output, including mean and variance, were obtained analytically from the PCE coefficients.

When using PCE, the uncertainty is represented through a set of independent random variables $\mathbf{X} = (X_1, \dots, X_d)$, each characterised by a normal probability distribution. In order to allow the construction of a generalised Polynomial Chaos Expansion (gPCE), each variable is transformed to a standardised form $\xi_i = (X_i - \mu_i)/\sigma_i$, such that $\xi_i \sim \mathcal{N}(0,1)$ and μ_i, and σ_i are the mean and standard deviation of the i-th parameter, respectively. The mapping between physical parameters $\mathbf{x}$ and standard normal variables $\boldsymbol{\xi}$ ensures the orthogonality properties required for PCE convergence (Thapa, Mulani and Walters, 2020; Thapa *et al.*, 2024).

In PCE, $u(x(\xi))$, which is a second-order stochastic response with finite variance and dependent on a set of independent random inputs $x = \{x_i\}_{i=1}^{N}$, is represented as a spectral expansion using multivariate orthogonal polynomials $\psi_k(\xi)$ and deterministic PCE coefficients a_k as in Eq 1.

$$u(x(\xi)) \approx \sum_{k=1}^{\infty} a_k \, \Psi_k(\boldsymbol{\xi}) \qquad 1$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_d)$, is a multi-index with non-negative integer components, $a_{\boldsymbol{\alpha}}$ are the PCE coefficients to be determined, $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ are multivariate orthonormal polynomial basis functions. More detailed descriptions and formulas of adaptive PCE are presented in the related literature (Sudret, 2008, 2008; Blatman and Sudret, 2011).

To address the curse of dimensionality, an adaptive PCE formulation (VARPCE) was employed, in which polynomial basis functions are introduced incrementally in small subsets. Variance-based pruning was applied to remove basis terms with negligible contributions while ensuring that the relative loss of total variance remained below a predefined tolerance. Polynomial order selection and overfitting prevention were managed using k-fold cross-validation, with model complexity increased only when cross-validation error showed meaningful improvement.

To validate the adaptive PCE surrogate, a hold-out set of N = 50 direct EnergyPlus simulations, drawn from an independent random seed withheld from the training process, was used as the ground truth. Moreover, the computational efficiency of the adaptive PCE surrogate was benchmarked against direct Monte Carlo simulation using EnergyPlus as the ground truth solver.

Adaptive sampling was integrated with adaptive PCE to efficiently enrich the experimental design. Initial samples were generated using Latin Hypercube Sampling to ensure global coverage of the input space. As the active basis evolved, additional samples were selected using an A-optimal design criterion, which minimises the variance of PCE coefficient estimates and targets regions of high surrogate uncertainty. This strategy balances global exploration with local refinement, reducing redundant simulations and improving surrogate stability (Figure 3). The adaptive sampling procedure is:

1- Generate an initial Latin Hypercube Sample (LHS) of size N_0
2- Evaluate computational model at initial design points
3- For polynomial order $p = 1,2, \dots, p_{\max}$:
   a. Fit PCE using current data
   b. Apply variance-based pruning
   c. Compute cross-validation error
   d. If the CV convergence criterion is met, stop
   e. Otherwise, for $\ell = 1, \dots, L_{\text{adapt}}$ (here $L_{\text{adapt}} = 5$):
      - Generate a candidate set of size N_{cand} via LHS
      - Select n_{new} points (here $n_{\text{new}} = 3$) using the A-optimal criterion
      - Evaluate model at new points
      - Update PCE and check CV convergence

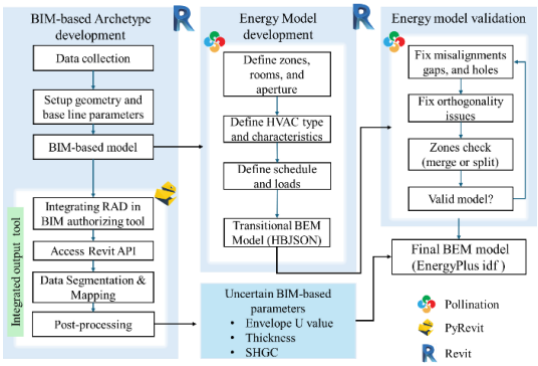


Figure 2. Developing a BIM-based energy-simulation-ready file format.

Sensitivity Analysis

Global sensitivity analysis was performed analytically using the final PCE coefficients. First-order and total-effect Sobol' indices were computed to quantify both main effects and interaction contributions of each uncertain input parameter. Convergence of the adaptive process was primarily assessed using the Kullback–Leibler divergence, which measures stability of output probability distributions across iterations and aligns directly with uncertainty quantification objectives. Secondary validation metrics, including leave-one-out error and the coefficient of determination (Q^2), were used to confirm predictive accuracy. Once the convergence criteria were satisfied, uncertainty and sensitivity results were finalised and benchmarked against Monte Carlo simulation to evaluate computational efficiency and accuracy.

Case study and data preparation

The proposed methodology is applied to four representative Irish residential archetypes: terraced, bungalow, detached, and semi-detached houses. These archetypes are chosen as they have been tested and validated in previous studies (Ali *et al.*, 2024; Sood *et al.*, 2025) and are among the most common archetypes in Ireland. BIM models for each archetype are developed in Revit (Figure 4) to capture both geometric and non-geometric attributes, including floor areas, ceiling heights, fenestration, and thermal envelope properties.

A PyRevit-based workflow is used to directly extract data via the Revit API, enabling a unified, automated, and user-friendly BIM data integration. Envelope-related thermal parameters are treated as uncertain inputs with normal distributions, while deterministic parameters are fixed at literature-based mean values.

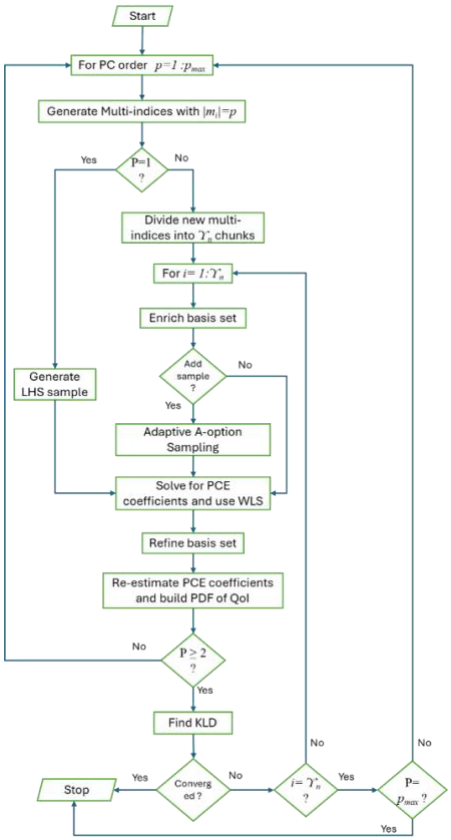


Figure 3. Flowchart of integrating adaptive sampling and adaptive PCE.

The uncertain inputs comprised envelope-related parameters, including U-values of the external walls, roof, and floor, the window U-value, and the solar heat gain coefficient (SHGC), each assigned normal probability distributions as described in the previous Section. Deterministic parameters, such as setpoint temperatures and lighting density, were fixed at mean values derived from the literature (Table 1).

Initial uncertainty propagation was performed using 30 Latin Hypercube samples to generate the training dataset. The adaptive PCE framework was then applied, with polynomial order incrementally increased up to the fourth order. A variance-based pruning strategy was employed to discard basis terms with negligible contribution, while new input samples were iteratively added using an A-optimal sequential design until convergence was achieved. Convergence was determined when the Kullback–Leibler divergence (KLD) between successive output distributions fell below 0.01 and the coefficient of determination (Q^2) exceeded 0.9. The outputs of interest were annual and monthly energy use intensity and annual and monthly heating demand (kWh).

After convergence, global sensitivity analysis was performed using Sobol' indices derived directly from the

final PCE coefficients. Both first-order and total-order Sobol' indices were computed for each uncertain parameter, providing interpretable measures of their individual contributions and interaction effects on the outputs. In the final stage, the computational cost of this method will be compared with that of the classic Monte Carlo.

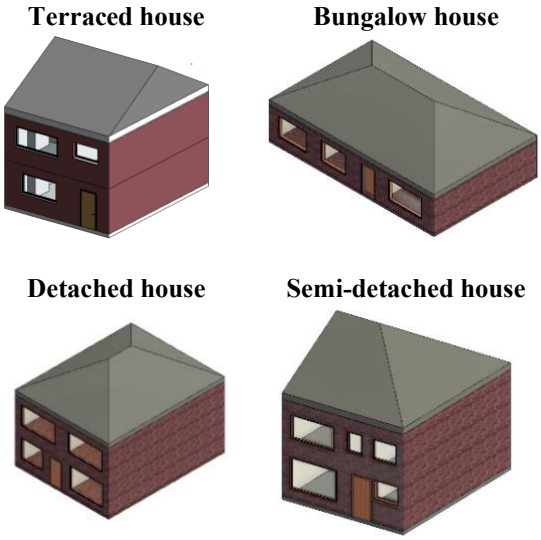


Figure 4. Common archetypes for Irish residential buildings in Revit.

Results and discussion

As shown in Table 2, the convergence history demonstrates that the adaptive sparse PCE model reaches stability very quickly while requiring only a small number of EnergyPlus simulations. At first order with 30 samples, for Bungalow as an example, the meta model contains six terms and already achieves a low Cross-Validation Root Mean Square Error (CV-RMSE) of 2.05 kWh/m², showing that annual EUI is dominated by linear effects, which is typical for U-value perturbations in building envelope analysis. When the adaptive routine adds new samples, the temporary rise in CV-RMSE reflects normal cross-validation variability rather than deteriorating model quality, and the consistent retention of only six terms confirms that higher-order interactions remain negligible.

Table 1. Deterministic parameter values and range of the variables with uncertainty for enriching final energy models (Egan et al.*, 2018; Ali* et al.*, 2024; Sood* et al.*, 2025).*

HVAC and system	Utility energy consumption (kWh/year.m2)			3
	Heating setback (°C)			12
	HVAC efficiency			2.2
	Heating setpoint (°C)			20.5
	DHW (l/m2/day)			2
	Lighting density (W/m2)			5
Internal loads	Occupants (person)			3.5
	Appliance density (W/m2)			11
Envelope	External wall U value (W/m2K)	0.1	1.1	
	Floor U value (W/m2 K)	0.1	1.1	
	Roof U value (W/m2 K)	0.1	0.9	
	Windows U value (W/m2 K)	0.6	4.3	
	SHGC of glass	0.1	0.9	
	Infiltration (ACH at 50 Pa)			1.85

After increasing the polynomial order to two, the meta model for semi-detached retains just 11 terms following pruning, indicating that most quadratic and interaction components do not meaningfully contribute to output variance. This modest growth in active terms aligns with the weak nonlinearities expected from EnergyPlus under small parameter perturbations. Throughout all iterations, the stability of total variance, which remains within the range of approximately 255–284, shows that the model captures the essential variance structure early and preserves it across refinements.

Table 2. Analytical variation of PCE convergence history.

Archetype	Order	No. terms	No. samples	CV-RMSE	Action	Itr.
Semi-detached	1	6	30	1.82	pp	
	1	6	33	2.44		1
	1	6	36	2.92		2
	2	11	36		pp	
	2	11	39	0.88		1
	2	11	42	0.82		2
	2	11	45	0.80		3
	2	11	48	0.78		4
	2	11	51	0.72		5
Detached	1	6	30	1.84	pp	
	1	6	33	2.26		1
	1	6	36	2.25		2
	2	11	36		pp	
	2	11	39	0.81		1
	2	11	42	0.77		2
	2	11	45	0.87		3
	2	11	48	0.70		4
Terraced	1	6	30	1.865	pp	
	1	6	33	2.48		1
	1	6	36	2.23		2
	2	11	36		pp	
	2	11	39	0.83		1
	2	11	42	0.85		2
	2	11	45	0.82		3
Bungalow	1	6	30	2.05	pp	
	1	6	33	2.27		1
	1	6	36	2.20		2
	2	11	36		Pp	
	2	11	39	1.06		1
	2	11	42	0.85		2
	2	11	45	0.84		3
	2	11	48	0.82		4
	2	11	51	0.74		5

*pp: post-prune; Itr: Iteration

The final CV-RMSE of the surrogate model for semi-detached, detached, terraced and bungalow are 0.72, 0.70, 0.82, 0.74 kWh/m² at second order, respectively confirming that the surrogate has reached its optimal balance between accuracy and complexity. Additional samples or higher-order terms do not improve performance, signalling classical convergence in adaptive PCE. Overall, the results show that the relationship between envelope parameters and annual EUI is smooth and predominantly linear, allowing an accurate surrogate to be built using approximately 50 EnergyPlus

simulations for the semi-detached archetype with excellent computational efficiency.

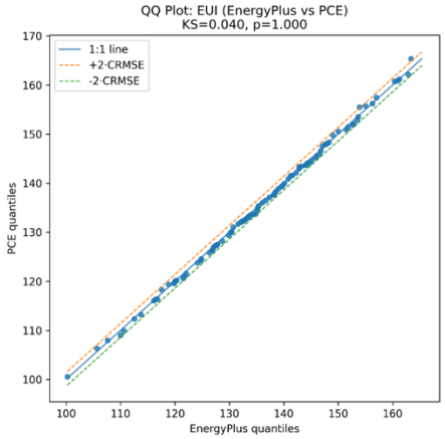


Figure 5. QQ comparison between the semi-detached surrogate predictions and the Energy Plus results.

Figure 5 shows excellent distributional agreement between PCE surrogate predictions and EnergyPlus outputs, with QQ points closely aligned to the 1:1 line across the full EUI range. The KS test (KS = 0.040, p = 1.000) confirms that both distributions are statistically indistinguishable. Nearly all points lie within ±2×CRMSE bounds, indicating small, well-controlled errors that are predominantly random rather than systematic, demonstrating high-fidelity probabilistic emulation by the surrogate.

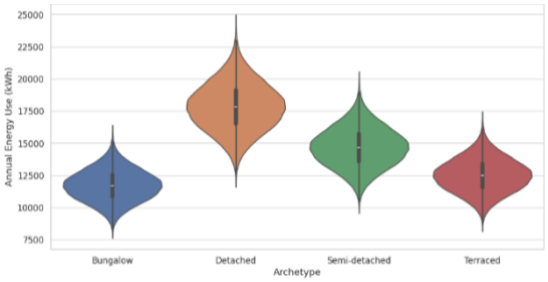


Figure 6. Violin plot of annual energy use for different archetypes.

The violin plot in Figure 6 illustrates the distribution of annual energy use (kWh) across four distinct building archetypes. This illustrates how input uncertainties in envelope U-values and SHGC propagate into annual energy variability. Notably, Detached properties exhibit the highest median annual energy consumption, as well as the broadest distribution, indicating a wider range of energy use within this archetype. Semi-detached properties show a slightly lower median energy use compared to detached, with a similarly wide but slightly less dispersed distribution. Terraced houses consistently demonstrate the lowest median annual energy use, with a comparatively tighter distribution, suggesting more uniform energy consumption patterns. Bungalows, while generally having lower energy use than Detached and Semi-detached, present a distribution that is somewhat wider than Terraced homes but narrower than detached properties. These findings suggest a strong correlation between building archetype (and implicitly, average floor area) and annual energy demand, with larger, standalone properties typically requiring more energy.

The similarity between distributions is because all four archetypes were generated under the same uncertainty space and parameter ranges from identical probabilistic assumptions. As a result, differences in built form alone introduce only second-order effects on annual energy use, leading to overlapping distributions with highly comparable central tendency and spread.

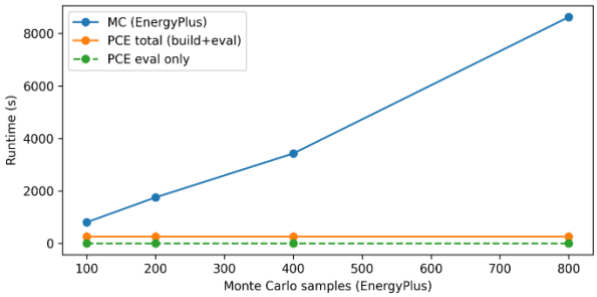


Figure 7. Runtime comparison between direct MCS using EnergyPlus and adaptive PCE surrogate for annual EUI.

Figure 7 compares the computational runtime of direct MCS with that of a PCE surrogate for annual EUI. The MC runtime increases approximately linearly with the number of samples, reflecting the fixed cost of repeated high-fidelity EnergyPlus evaluations and rapidly becoming computationally prohibitive as the number of runs grows. In contrast, the total PCE runtime, comprising surrogate construction and evaluation, remains nearly constant across different numbers of simulations, indicating that the computational cost is dominated by the one-time surrogate construction. Once constructed, the PCE surrogate can be evaluated at negligible cost, enabling large-scale uncertainty propagation and statistical analysis with minimal additional runtime. These results highlight the fundamentally different scaling behaviour of surrogate-based uncertainty quantification and demonstrate the suitability of PCE for applications requiring extensive sampling, repeated analyses, or computationally efficient sensitivity studies. However, the final decision on which method to choose depends on the trade-off between computational budget and required accuracy.

The monthly Sobol' sensitivity analysis shows a strong seasonal pattern in how envelope uncertainties influence energy demand (Table 3). During the heating-dominated months (January–March and September–December), Uwall is the main driver of variability, explaining 47-66% of the monthly energy variance. In contrast, during April–July, Uwin becomes the dominant factor, reflecting the greater role of glazing heat transfer and solar gains in warmer periods. Across all months, first-order and total-order indices are nearly identical (difference < 0.003), indicating that model behaviour is almost entirely additive with minimal interaction effects. This also confirms the strong convergence and stability of the adaptive sparse

PCE surrogate. Overall, the results demonstrate that the dominant envelope parameter changes seasonally, and that the PCE model reliably captures these monthly dynamics for interpreting energy behaviour and guiding retrofit priorities.

Table 3. Top first-order and total Sobol monthly indices.

Month	Top S_1 variable	Top S_1 value	Top S_T variable	Top S_T value
Jan	U_{wall}	0.66	U_{wall}	0.66
Feb	U_{wall}	0.53	U_{wall}	0.53
Mar	U_{wall}	0.47	U_{wall}	0.47
Apr	U_{win}	0.43	U_{win}	0.43
May	U_{win}	0.40	U_{win}	0.40
Jun	U_{win}	0.38	U_{win}	0.38
Jul	U_{win}	0.41	U_{win}	0.41
Aug	U_{wall}	0.40	U_{wall}	0.40
Sep	U_{wall}	0.53	U_{wall}	0.53
Oct	U_{wall}	0.53	U_{wall}	0.53
Nov	U_{wall}	0.57	U_{wall}	0.57
Dec	U_{wall}	0.58	U_{wall}	0.58

Conclusions

This study introduced a unified BIM–BEM uncertainty framework that integrates adaptive PCE with sequential A-optimal sampling. Unlike conventional BIM-based energy analyses that rely on deterministic simulations or Monte Carlo simulations, the proposed framework embeds uncertainty quantification and global sensitivity analysis directly into the modelling workflow. The results demonstrate that adaptive basis refinement combined with informed sampling is sufficient to capture the dominant variance structure of residential building energy models with high fidelity.

The adaptive sparse PCE surrogate converged using approximately 50 EnergyPlus simulations per archetype, achieving RMSE values below 0.8 kWh/m² and an R^2 exceeding 0.99 for annual EUI. In contrast, a Monte Carlo simulation would require several orders of magnitude more model evaluations to reach comparable distributional accuracy. This confirms that the energy response surface induced by envelope uncertainty is smooth and low-dimensional enough to be efficiently approximated by a variance-pruned polynomial basis, validating the core hypothesis of this study. Additionally, analytical Sobol' indices derived from the final PCE coefficients revealed a clear hierarchy of envelope parameter influence. External wall U-value and window U-value jointly explain the majority of annual EUI variance, while monthly analysis exposes a strong seasonal shift in dominance from opaque envelope components in winter to glazing properties in shoulder and summer months.

Future research should extend the uncertainty space beyond envelope properties to include occupant behaviour, infiltration variability, HVAC details, and system efficiencies, where stronger nonlinearities and interaction effects are expected. It is also valuable to check the effectiveness of LHS vs adaptive sampling and normal PCE vs adaptive PCE. Finally, comparing the results with the measured data could increase the credibility and robustness of the model's validation.

Acknowledgments

This publication has emanated from research supported partly by a grant from the Sustainable Energy Authority of Ireland (SEAI) under the Funding Program Grant No. 22/RDD/822. The opinions, findings, and conclusions expressed in this material are those of the author(s) and do not necessarily reflect the views of the SEAI.

References

Ali, U. et al. (2024) 'Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach', Energy and Buildings, 303, p. 113768. Available at: https://doi.org/10.1016/J.ENBUILD.2023.113768.

Alwan, Z. et al. (2021) 'Framework for parametric assessment of operational and embodied energy impacts utilising BIM', Journal of Building Engineering, 42, p. 102768. Available at: https://doi.org/10.1016/J.JOBE.2021.102768.

Arbulu, M., Oregi, X. and Etxepare, L. (2024) 'Parametric simulation tool for the enviro-economic evaluation of energy renovation strategies in residential buildings with life cycle thinking: PARARENOVATE-LCT', Energy and Buildings, 312. Available at: https://doi.org/10.1016/j.enbuild.2024.114182.

ASHRAE Standard (2022) 'Standard 90.1–2022: Energy standard for sites and buildings except low-rise residential buildings', American Society of Heating, Refrigerating and Air-conditioning Engineers, Atlanta, GA [Preprint].

Blatman, G. and Sudret, B. (2011) 'Adaptive sparse polynomial chaos expansion based on least angle regression', Journal of Computational Physics, 230(6), pp. 2345–2367. Available at: https://doi.org/10.1016/j.jcp.2010.12.021.

Chadly, A. et al. (2023) 'Techno-economic assessment of energy storage systems in green buildings while considering demand uncertainty', Energy and Buildings, 291, p. 113130. Available at: https://doi.org/10.1016/j.enbuild.2023.113130.

Cheng, K. and Lu, Z. (2018) 'Adaptive sparse polynomial chaos expansions for global sensitivity analysis based on support vector regression', Computers & Structures, 194, pp. 86–96. Available at: https://doi.org/10.1016/j.compstruc.2017.09.002.

Egan, J. et al. (2018) 'Definition of a useful minimal-set of accurately-specified input data for Building Energy Performance Simulation', Energy and Buildings, 165, pp. 172–183. Available at: https://doi.org/10.1016/j.enbuild.2018.01.012.

ElSayed, M., Shultz, J. and Kurtz, J. (2025) 'User-friendly AI-driven automation for rapid building energy model

generation', Energy and Buildings, 345, p. 116092. Available at: https://doi.org/10.1016/j.enbuild.2025.116092.

Energy Performance of Buildings Directive (2025) An official website of the European Union. Available at: https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en (Accessed: 7 May 2025).

Forth, K. and Borrmann, A. (2024) 'Semantic enrichment for BIM-based building energy performance simulations using semantic textual similarity and fine-tuning multilingual LLM', Journal of Building Engineering, 95, p. 110312. Available at: https://doi.org/10.1016/j.jobe.2024.110312.

Heydari, M.H. and Heravi, G. (2023) 'A BIM-based framework for optimization and assessment of buildings' cost and carbon emissions', Journal of Building Engineering, 79, p. 107762. Available at: https://doi.org/10.1016/J.JOBE.2023.107762.

International Code Council (ICC) (2021) 2021 International Energy Conservation Code (IECC), International Code Council, Inc. Available at: https://codes.iccsafe.org/content/IECC2021P2 (Accessed: 7 October 2024).

Koniorczyk, M. et al. (2022) 'Stochastic energy-demand analyses with random input parameters for the single-family house', Building Simulation, 15(3), pp. 357–371. Available at: https://doi.org/10.1007/s12273-021-0812-9.

Liu, Z. et al. (2020) 'Surrogate modeling based on resampled polynomial chaos expansions', Reliability Engineering & System Safety, 202, p. 107008. Available at: https://doi.org/10.1016/j.ress.2020.107008.

Liu, Z. et al. (2022) 'Impacts of uncertainty in building envelope thermal transmittance on heating/cooling demand in the urban context', Energy and Buildings, 273, p. 112363. Available at: https://doi.org/10.1016/J.ENBUILD.2022.112363.

Prataviera, E. et al. (2022) 'Evaluation of the impact of input uncertainty on urban building energy simulations using uncertainty and sensitivity analysis', Applied Energy, 311, p. 118691. Available at: https://doi.org/10.1016/j.apenergy.2022.118691.

Prilliman, M. et al. (2023) Quantifying Uncertainty in PV Energy Estimates Final Report. NREL/TP-7A40-84993, 1961370, MainId:85766, p. NREL/TP-7A40-84993, 1961370, MainId:85766. Available at: https://doi.org/10.2172/1961370.

Shadram, F. et al. (2016) 'An integrated BIM-based framework for minimizing embodied energy during building design', Energy and Buildings, 128, pp. 592–604. Available at: https://doi.org/10.1016/j.enbuild.2016.07.007.

Sood, D. et al. (2025) 'Machine learning based multi-objective optimisation of energy consumption, thermal comfort and CO2 concentration in energy-efficient naturally ventilated residential dwellings', Building and Environment, 267, p. 112255. Available at: https://doi.org/10.1016/j.buildenv.2024.112255.

Sudret, B. (2008) 'Global sensitivity analysis using polynomial chaos expansions', Reliability Engineering & System Safety, 93(7), pp. 964–979. Available at: https://doi.org/10.1016/j.ress.2007.04.002.

Thapa, M. et al. (2024) 'Classifier-based adaptive polynomial chaos expansion for high-dimensional uncertainty quantification', Computer Methods in Applied Mechanics and Engineering, 422, p. 116829. Available at: https://doi.org/10.1016/j.cma.2024.116829.

Thapa, M., Mulani, S.B. and Walters, R.W. (2020) 'Adaptive weighted least-squares polynomial chaos expansion with basis adaptivity and sequential adaptive sampling', Computer Methods in Applied Mechanics and Engineering, 360, p. 112759. Available at: https://doi.org/10.1016/j.cma.2019.112759.

Tian, W. et al. (2020) 'Sensitivity analysis of building energy performance based on polynomial chaos expansion', Journal of Green Building, 15(4), pp. 173–183. Available at: https://doi.org/10.3992/jgb.15.4.173.

Tian, W. (2024) 'Towards advanced uncertainty and sensitivity analysis of building energy performance using machine learning techniques', Journal of Building Performance Simulation, 17(6), pp. 655–662. Available at: https://doi.org/10.1080/19401493.2024.2387071.

United Nations Environment Programme, G.A. for B. and C. (2024) 'Global Status Report for Buildings and Construction - Beyond foundations: Mainstreaming sustainable solutions to cut emissions from the buildings sector'. Available at: https://wedocs.unep.org/20.500.11822/45095.

Wang, H. et al. (2022) 'BIM-based automated design for HVAC system of office buildings—An experimental study', Building Simulation, 15(7), pp. 1177–1192. Available at: https://doi.org/10.1007/S12273-021-0883-7/METRICS.

Yu, G. et al. (2025) 'A novel global prediction framework for multi-response models in reliability engineering using adaptive sampling and active subspace methods', Computer Methods in Applied Mechanics and Engineering, 433, p. 117506. Available at: https://doi.org/10.1016/j.cma.2024.117506.

Zhao, F., Lee, S.H. and Augenbroe, G. (2016) 'Reconstructing building stock to replicate energy consumption data', Energy and Buildings, 117, pp. 301–312. Available at: https://doi.org/10.1016/j.enbuild.2015.10.001.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

ONTOLOGY-SUPPORTED BRIDGE COMPONENT DESIGN: IMPLEMENTATION OF BRIDGE CAP DESIGN RULES

Tim Noack[1], Han Qian[1], Chongjie Kang[1] and Steffen Marx[1]
[1]TUD Dresden University of Technology, Dresden, Germany

Abstract

The bridge design process is a complex and time-consuming task for bridge engineers. To reduce repetitive tasks, automating the application of guidelines and standards is desirable. Currently, automation attempts are mostly based on table-based calculations. This work proposes a new approach to automated, code-compliant structural design: an exemplary automation process for dimensioning the bridge cap. In this process, an ontology describing bridge components and boundary conditions is created. German road bridge design rules are implemented using the Semantic Web Rule Language (SWRL). The aim is to enable a fully automatic geometric dimensioning of bridge caps based on several boundary conditions.

Introduction

When designing a bridge, the bridge engineer encounters a challenging task. There is a vast number of boundary conditions, requirements and structural mechanics to consider. This leads to long planning and testing phases. Guidelines and standards have been created to support the engineers' decisions and to ensure the quality of bridge design. Their application is often repetitive and laborious because of the many required steps and decisions. Therefore, automating these tasks is desirable. In recent decades, this has primarily been achieved through table-based calculations and, in rare cases, by implementing parts of the standards in programming code (Jeske et al., 2022). Both approaches lack flexibility and are unsuitable for integration into an artificial intelligence-supported bridge design process, which will be essential for a future-proof and efficient bridge engineering sector.

Therefore, a new approach to automated, code-compliant structural design is proposed: an ontology-based bridge cap dimensioning process. Ontologies, as part of the Semantic Web, represent a way to structure semantic knowledge and make it machine-interpretable (Staab and Studer, 2009) and are therefore well suited for representing guidelines and standards. The dimensioning process of a bridge cap was chosen because it is highly standardized in German road guidelines. A bridge engineer needs only a limited set of boundary conditions to determine, without further decision-making, the appropriate bridge-cap type and its dimensions.

In the first step of this work, existing ontologies in the field of bridge engineering and construction guidelines are reviewed. Next, German bridge-engineering guidelines and the process of extracting design rules from them are described. Bridge-cap dimensioning rules are implemented in an ontology using the Semantic Web Rule Language (SWRL) together with existing and newly created ontology structures. The proposed bridge cap dimensioning process is carried out for an exemplary bridge. Finally, the integration of the approach into a Building Information Modeling (BIM)-based process is discussed.

State of the Art

Extensive research on ontologies has been conducted in the broader Architecture, Engineering and Construction (AEC) industry as well as in the more specialized field of bridge engineering. Farghaly et al. (2023) found that between 2000 and 2021, more than 140 journal articles on ontologies, the Semantic Web or Linked Data were published in the AEC sector. They were categorized into ten application fields, such as process, compliance, smart cities and sustainability. Most of the ontologies were developed for managing and interpreting infrastructure or building data. Others aim to create a semantic interoperability between different sets of information. The most established ontologies in this field are the IfcOWL ontology (Beetz et al., 2009) and the Building Topology Ontology (BOT) (Rasmussen et al., 2019, 2020). The purpose of both ontologies is to enable the mapping from Industry Foundation Classes (IFC) to an ontological representation. Thus, with these ontologies, building data from IFC files can be transformed into an ontological database. Since the IFC version 4.3, the standard also includes objects for infrastructure representation, specifically bridge elements and the new basic class IfcBridge (International Organization for Standardization (ISO), 2024). This is why an alignment of new bridge ontologies to the IFC standard or at least to IfcOWL or BOT is recommended.

The objective of this work is to create a new bridge component based on the bridge's boundary conditions and relevant guidelines and standards. Therefore, the ontology application field "compliance" is most important here and requires further investigation. Ontologies in this field have primarily been used for code-compliance evaluation, which is also addressed in this paper. Pauwels et al. (2011, 2017) describe the advantages of mapping building data from IFC-files (explicit building information) into a Se-

mantic Web representation. By using rule languages such as SWRL or N3Logic together with an inference engine, implicit building information can be inferred. For example, the compliance of the building to a specific standard paragraph can be checked automatically. This was used in other research (Salama and El-Gohary, 2013; Jędrzejewska et al., 2026) but only for code-compliance checking, not for deriving a structural design directly from boundary conditions.

In the field of bridge engineering, a similar situation emerges. The application fields of developed ontologies are risk management (Hassan and Le, 2020), Structural Health Monitoring (Li et al., 2021), inspection (Erdene et al., 2020; Liu and El-Gohary, 2022; Zhang et al., 2022), diagnostics (Jäkel et al., 2025), bridge rehabilitation (Wu et al., 2021), and maintenance (Ren et al., 2019; Göbels and Beetz, 2025). These domains involve collecting, managing, and analyzing existing bridge data using various bridge ontologies. Hamdan and Scherer (2020) describe a foundational bridge ontology, the Bridge Topology Ontology (BROT), with several extensions. These ontologies can represent a bridge, its materials, components, and associated damages. BROT is based on BOT and therefore suitable for an integration into a BIM-based process.

Only few research studies have addressed ontologies in the field of bridge design. Aqlan et al. (2025) proposed an ontology for structuring data such as dimensions, structural details, boundary conditions, and images from bridge databases, making it accessible for machine-learning-based bridge design. Zhang et al. (2021) used an ontology for the holistic design of energy piles in a bridge deck deicing system. Although the ontology supports the design of these structures, the described principle is not an automatic design method based on the bridge's boundary conditions. Instead, it performs a database query to select which of 12 predefined design solutions best fits a given condition. Therefore, according to the author, no existing research has addressed automatic bridge-component design or dimensioning based on guideline-derived rules, despite the suitability of Semantic Web technologies for such tasks.

Methodology

Rules from Guidelines and Standards

Standards and guidelines contain information in four main formats: text, formulas, tables, and images (photos, drawings, diagrams). To digitalize and automate the document's content, the information must be interpreted and translated into *if-then* rules, formulas, or algorithms. The interpreted information can then be encoded and integrated into a digital process. The goal is to define boundary conditions as inputs and generate outputs —- such as parameter values or text —- after the process is executed.

This process will be demonstrated in this work using a highly standardized bridge component: the bridge cap. Figure 1 illustrates the regular concrete road slab bridge cross section, including two bridge caps. The caps are a component that generally consists of reinforced concrete and carries the railings, traffic barriers and the foot- or cycleway.

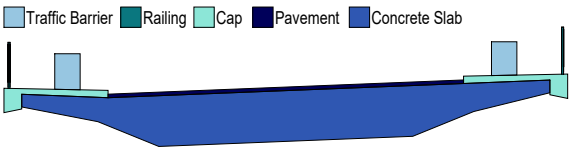


Figure 1: Superstructure cross-section: Main components of a road slab bridge.

In German standards and guidelines, there are several documents defining the design of concrete bridge caps. The RE-ING of the BASt (Bundesanstalt für Straßen- und Verkehrswesen, 2024) is a guideline describing the composition of road bridges. The RiZ-ING, also published by the BASt (Bundesanstalt für Straßen- und Verkehrswesen, 2023) is a collection of road bridge component drawings including twelve types of bridge caps. The Eurocode 2 (CEN/TC 250 Structural Eurocodes et al., 2023) and the DIN 1045 (DIN-Normenausschuss Bauwesen (NABau) et al., 2023) are specifying the concrete mixture and design details such as the concrete cover.

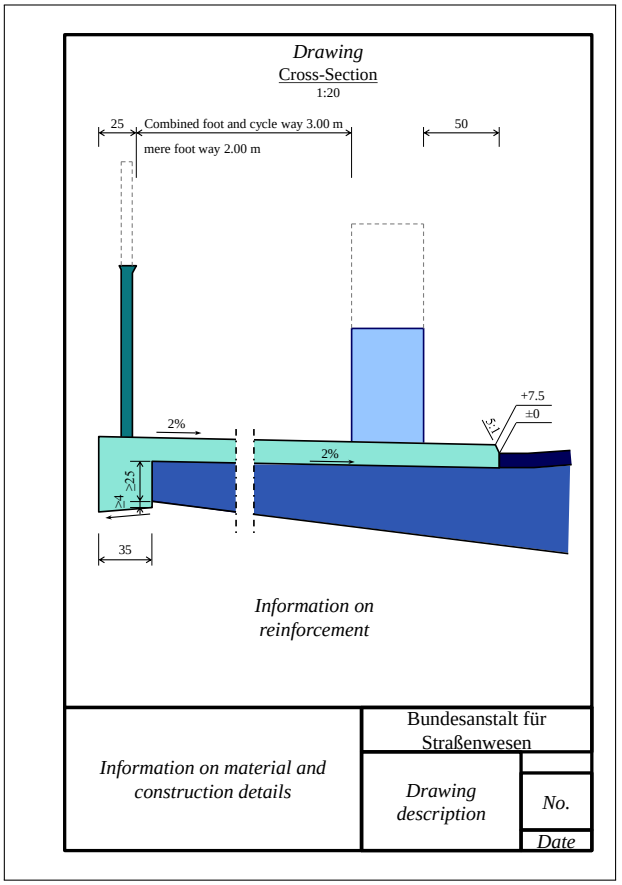


Figure 2: Composition of a RiZ-ING drawing with bridge cap measurements of drawing number "Kap 1 Blatt 3" (Bundesanstalt für Straßen- und Verkehrswesen, 2023)

This work primarily uses the RiZ-ING, since it focuses

on early-stage design where material details and reinforcement configuration are not yet critical. The RiZ-ING drawings are generally A4-pages containing a drawing of the component and its adjacent components. In Figure 2, the composition of such a page is shown. Besides the drawing it contains indicated component dimensions and location as well as additional information in text-form in the drawing and the title block. Each drawing has a title describing the component and a number e.g., "Kap 1 Blatt 3" for the third page of the first cap drawing (Bundesanstalt für Straßen- und Verkehrswesen, 2023). From the information contained in these drawings, rules can be derived that describe the step-by-step dimensioning of bridge caps. These rules can then be used for automated bridge cap dimensioning. Because SWRL is used for implementation, all rules are expressed as *if-then* constructs. Some of these rules are described in the Tables 2, 3 and 4.

The information contained in the RiZ-ING cap drawings can be grouped into three categories:

A) Geometric information on the cap
B) Relational information between the cap and adjacent components
C) Additional information on the cap's purpose, its material and its detailed design

To implement this information into an automated design process, relational and additional information is particularly critical. A minimum requirement would be implementing the rules in object-oriented programming (OOP) to correctly include all the information. OOP would be a good solution for the implementation of rigid *if-then* rules and the process in this work in general. However, OOP is not suitable when rules have to be implemented that are softer or contradicting and need to be weighted against each other. Here, an ontological structure would be appropriate with these rules described in the SPARQL Protocol And RDF Query Language (SPARQL) or the Shapes Constraint Language (SHACL) and applied using a Large Language Model (LLM). Therefore, this work uses ontologies to implement the automated bridge-cap dimensioning process and establish the foundation for future extensions.

Ontology

The ontological structure used in this work is based on the Bridge Topology Ontology (BROT) and the Bridge Component Ontology (BRCOMP) described by Hamdan and Scherer (2020), both of which build upon BOT introduced by Rasmussen et al. (2020). In addition, several *ObjectProperties* of the Relative Location Ontology (RELOC) by Göbels and Beetz (2024) have been incorporated.

Table 1 lists the ontologies used in this work. The Bridge Core Ontology (BCore) and the Bridge Element Design Rule Ontology (BEDRO) were created for this work in the Web Ontology Language (OWL 2) using Protégé (Musen, 2015). BCore extends BROT, BMAT and BRCOMP to represent the bridges, their components and their materials. BCore and BEDRO were created in a bottom-up approach. Thus, the *Classes* and properties

Table 1: Ontologies used in this work

Ontology	Prefix	IRI
Bridge Core Ontology	BCore	`https://w3id.org/bcore#`
Bridge Element Design Rule Ontology	BEDRO	`https://w3id.org/bedro#` *
Building Topology Ontology	BOT	`https://w3id.org/bot#`
Bridge Topology Ontology	BROT	`https://w3id.org/brot#`
Bridge Component Ontology	BRCOMP	`https://w3id.org/brcomp#`
Building Material Ontology	BMAT	`https://w3id.org/bmat#`
Relative Location Ontology	RELOC	`https://w3id.org/reloc#`

*The publication process for this website is not yet complete at the time of submission of this work.

necessary for the implementation of the RiZ-ING design rules were added to the basic framework of BROT, BMAT and BRCOMP. BCore includes structures for representing boundary conditions relevant to bridge design – the environment where the bridge is to be built. Additionally, it contains *Classes* and *ObjectProperties* describing agents (Persons or Companies) and media files (images, technical documents) that are connected to the specific bridge. These structures are not used in this paper, unlike the *Classes* or *ObjectProperties* displayed in Figure 3.

Here, the *Classes* needed for the dimensioning of the bridge caps are shown. A `Bridge` contains `Components`, such as the `EquipmentComponents` `Cap`, `TrafficBarrier` and `Pavement`. These components are also shown in Figure 1. Their relative locations are represented using the RELOC *ObjectProperties* `meetTop`, `meetBottom`, `meetLeft` and `meetRight`. This allows representing category B information. Newly created *Classes* are `InfrastructureRoute` and its subclasses `FootCycleRoute` and `RoadRoute`. The foot or cycle way can run to the left or to the right of a road. This information belongs to category C.

Several *DataProperties* were added to describe category A information, mainly geometric dimensions: the height of the pavement `PAVh`, the width of the traffic barrier `TRBb` and the cap's geometric parameters shown in Figure 4. All these parameters have the assigned *Datatype* `xsd:decimal`. The `FootCycleWay` has a `xsd:boolean`

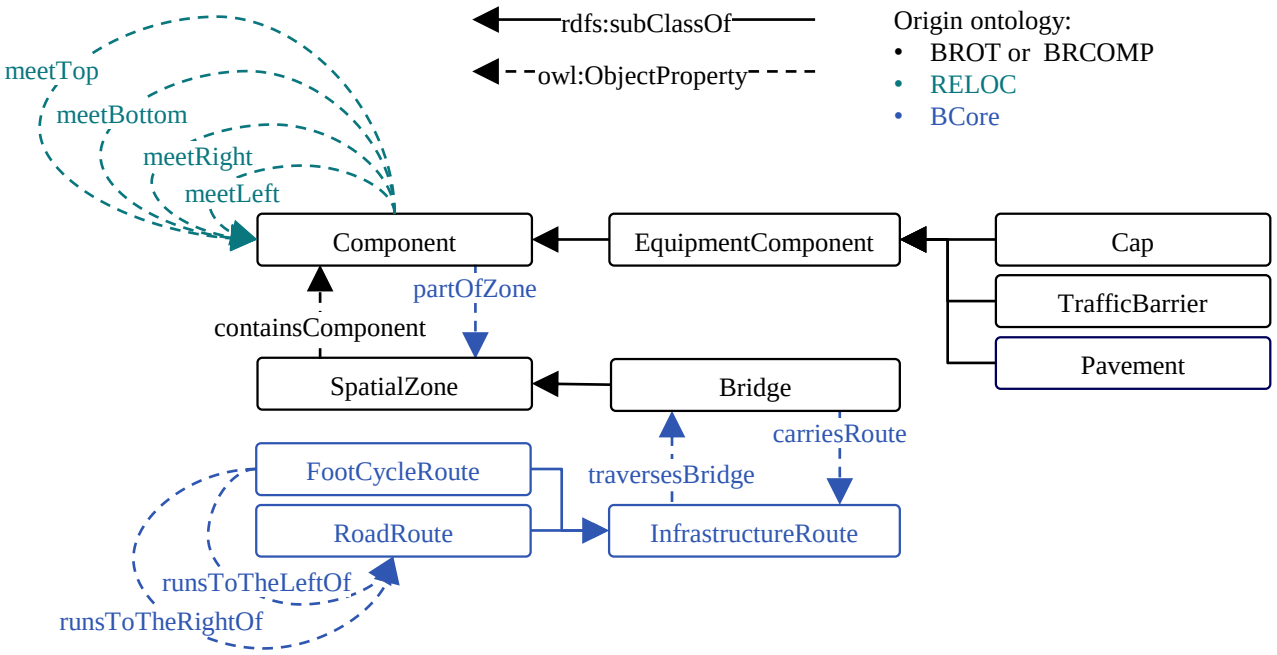


Figure 3: Diagram of Classes *from Hamdan and Scherer (2020) (BROT and BRCOMP, black), with added axioms of BCore (blue) and* ObjectProperties *from Göbels and Beetz (2024) (RELOC, green)*

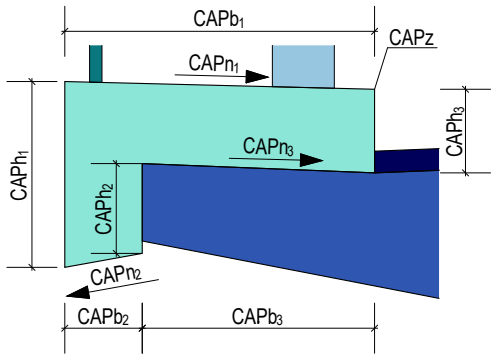


Figure 4: Geometric parameters in a bridge cap cross-section.

property `FootAndCycleWay`, which is `true` if it is a *combined* foot and cycle way. If it is only a foot way, it the value is `false`. The *DataProperty* `Position` of `Component` specifies, where the component is positioned in the bridge superstructure. The value 1 means left (relative to the direction of the route), the value 99 stands for right. If there are more than two components of the same type, they are numbered from left to right starting with 1, where right means always `Position` = 99. There are two more important *DataProperties* of `Cap`: the `CAPType`, indicating which cap design is assigned to the specific cap, and the `CAPCurb`, a `xsd:boolean`-Property describing if the cap has a curb or not. For example, a cap requires a curb if no traffic barrier is located above it, and vice versa.

This rule and other rules have been implemented in an ontology named BEDRO, an extension of BCore containing *if-then* rules extracted from the RiZ-ING guideline (Bundesanstalt für Straßen- und Verkehrswesen, 2023). The rules were first written in natural language before being translated into SWRL (O'Connor et al., 2005).

Three of these rules are contained in the tables below. They depict the cap dimensioning process. First, the boundary conditions of the bridge and the cap have to be checked to decide which type of cap to chose (Table 2). The type of cap refers to the different RiZ-ING drawings. For example, the drawing in Figure 2 dictates the measurements of an outer cap carrying a foot or cycle way and a traffic barrier. Inner caps are just needed if the bridge consists of multiple partial structures.

Table 2: Rule cap type*: description and SWRL-implementation*

If a road traverses a bridge and a combined foot and cycle way runs to the right of it and on the right side of the bridge it contains a cap with a traffic barrier on top, the cap's type is an outer cap with foot and cycle way and traffic barrier ("OuterCap_FCW_TRB")

`RoadRoute(?r) ^FootCycleRoute(?fcw)`
`^runsToTheRightOf(?fcw,?r)`
`^FootAndCycleWay(?fcw,true)`
`^Bridge(?b) ^carriesRoute(?b,?r)`
`^carriesRoute(?b,?fcw)`
`^containsComponent(?b,?c) ^Cap(?c)`
`^Position(?c,99) ^meetBottom(?c,?trb)`
`^TrafficBarrier(?trb)`
`-> CAPType(?c,"OuterCap_FCW_TRB")`

In the second step, the values contained in the RiZ-ING drawing are assigned to the *DataProperties* of the specific cap (Tables 3 and 4). They differ from cap type to cap type and are sometimes dependent on the dimensions of the adjacent components (`Pavement` and `TrafficBarrier`).

Table 3: Rule "Kap 1 Blatt 3" with combined foot- and cycle way*: description and SWRL-implementation*

If a cap has the type "OuterCap_FCW_TRB" (Outer cap with combined foot- and cycle way and traffic barrier) and it lies directly below a traffic barrier, the cap does not need a curb, its dimensions are declared as shown in Figure 2 and the top width is calculated as follows $CAPb_1 = 0.25 + 3.00 + TRBb + 0.5$ (unit: m)
`Cap(?c) ^CAPType(?c,"OuterCap_FCW_TRB")` `^meetBottom(?c,?trb) ^TrafficBarrier(?trb)` `^TRBb(?trb,?trbb)` `^swrlb:add(?b01,0.25,3.00,?trbb,0.50)` `-> CAPb1(?c,?b01) ^CAPb2(?c,0.35)` `^CAPn1(?c,0.02) ^CAPn3(?c,0.02)` `^CAPh2(?c,0.29) ^CAPCurb(?c,false)`

Table 4: Rule cap without curb*: description and SWRL-implementation*

If a bridge contains a pavement and a cap without curb, the cap's height 3 ($CAPh_3$) is 7.5 cm bigger than the pavement's height ($PAVh$), as shown in Figure 2
`Bridge(?b) ^containsComponent(?b,?p)` `^Pavement(?p) ^PAVh(?p,?ph)` `^containsComponent(?b,?c) ^Cap(?c)` `^CAPCurb(?c,false)` `^swrlb:add(?ch3,?ph,0.075)` `-> CAPh3(?c,?ch3)`

The third step is to calculate the rest of the cap's dimensions (see Figure 4) using the following equations (for better readability, the prefix "CAP" has been removed):

$$n_2 \approx 7.5\ \% \tag{1}$$

$$b_3 = b_1 - b_2 \tag{2}$$

$$h_1 = h_2 + h_3 + b_1 n_1 + b_2 n_2 - b_3 n_3 \tag{3}$$

These equations can also be implemented using SWRL and its math built-ins. As soon as all the rules are implemented and are not contradicting, the ontology can be used for an automated bridge cap dimensioning process.

Automated Design Process

Bridge design is typically an iterative process involving repeated refinement of the bridge structure and the components. An exemplary process is shown in Figure 5. It begins with the assessment of the bridge's boundary conditions, followed by a rough design of the superstructure, substructure, and equipment. Then, it is checked if the boundary conditions are met and wether the design meets the requirements related to cost, environmental impact, and aesthetics.

This work proposes automating one part of this process – the equipment planning, specifically the dimensioning of bridge caps. The created ontology with the implemented cap design rules, BEDRO, can be used as follows:

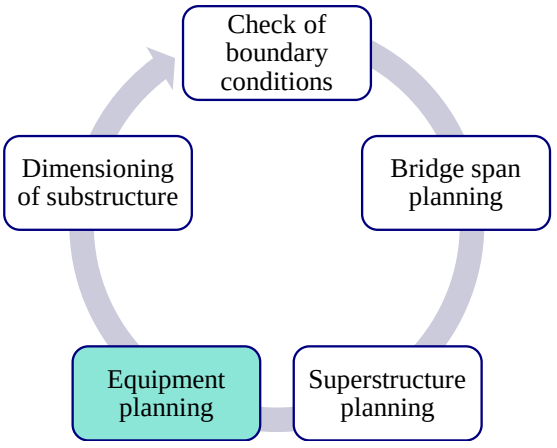


Figure 5: Iterative bridge design process.

1. Create instances of `Bridge`, `RoadRoute` and possibly one or two instances of `FootCycleRoute`.
2. Connect `RoadRoute` and `FootCycleRoute` instances to the `Bridge` via `traversesBridge` and assign either `runsToTheRightOf` or `runsToTheLeftOf` as between the routes.
3. Create an instance of `Pavement`, two instances of `Cap` and possibly instances of `TrafficBarrier`.
4. Assign *DataProperties* values for the components. Each cap and traffic barrier needs a value for its `Position`. The left component is assigned a "1", the right one a "99". As no design rules for `Pavement` or `TrafficBarrier` are implemented yet, their dimensions (e.g., `PAVh`, `TRBb`) must be provided manually.
5. Instantiate `partOfZone` as property of the components and the `Bridge`.
6. Run a SWRL-capable reasoner on the ontology and the dimensions of the bridge cap get automatically assigned values. In Protégé, the reasoner *Pellet* is recommended, as it handles SWRL effectively (O'Connor et al., 2005).

In Table 5, exemplary instances of a bridge and its corresponding routes and equipment components are described as well as their *ObjectProperties* and *DataProperties*. After implementing them in BEDRO and running the *Pellet*-reasoner, new relations and attributes have been created. Figure 6 shows the instance "B1Cap1", which is the left cap of the bridge "B1". The reasoner has inferred two *ObjectProperties*, describing that the cap is directly below the traffic barrier "B1TrB1" and directly left of the bridges pavement "B1Pav1". Also, it has inferred that the type of the cap is "OuterCap_FCW_TRB", following the rule in Table 2. The other inferred values are assigned by executing the other rules described above. "B1Cap99" does not have the same type as "B1Cap1" and hence has other values for its properties.

Since not many bridge engineers are experienced in working with ontologies or Protégé, the process must be made more user-friendly for practical adoption. A graphical user-interface (GUI) may be implemented,

Table 5: Exemplary instances and assigned properties for an automated dimensioning of the bridge caps "B1Cap1" and "B2Cap99"

Class	Instance	ObjectProperty/ DataProperty	Value
`Bridge`	B1		
`Road`	R1	`traversesBridge`	B1
`FootCycleWay`	FC1	`FootAndCycleWay`	true
		`runsToTheLeftOf`	R1
		`traversesBridge`	B1
`Cap`	B1Cap1	`Position`	1
		`partOfZone`	B1
`Cap`	B1Cap99	`Position`	99
		`partOfZone`	B1
`Pavement`	B1Pav	`PAVh`	0.07
		`partOfZone`	B1
`TrafficBarrier`	B1TrB1	`Position`	1
		`TRBb`	0.5
		`partOfZone`	B1

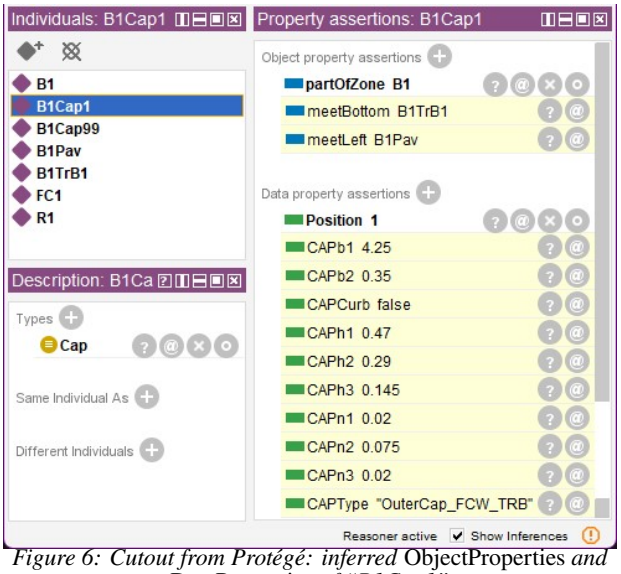


Figure 6: Cutout from Protégé: inferred ObjectProperties *and* DataProperties *of "B1Cap1"*

where the engineer can insert the boundary conditions of the respective bridge. Via an application programming interface (API), the *Classes* and *Properties* are then created, the reasoner is run and the generated cap dimensions are displayed as in Figure 7. Also, an export of these dimensions to a CSV-file would be feasible.

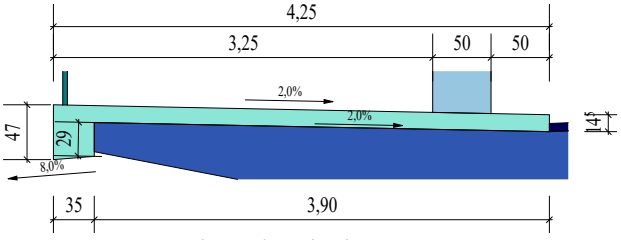


Figure 7: Rule-conform bridge cap cross-section.

A step further would be to implement the whole process into a combined GIS-BIM-approach. Here, it would be possible to extract the relevant boundary conditions from geographical information system (GIS)-data. The boundary conditions are then integrated into the ontology via an API. After the application of the rules using a reasoner, the generated component attributes can be extracted. With this data, a parametric building information model as described by Jeske et al. (2022) is populated, to obtain rule-conform dimensions and material information. This model can be used for further analyses such as quantity or cost estimation, as well as for visualization of the created bridge components.

Discussion

This work proposes a method for implementing design rules from a German guideline, the RiZ-ING, into an ontology. This ontology can be used to automatically generate rule-conform bridge cap dimensions for specific bridge boundary conditions. For that, SWRL rules were derived and can be applied using a reasoner. This process is limited in its user friendliness, applicability and scalability.

As stated before, bridge engineers do not usually work with ontologies and even when they do, many manual steps are still required before the cap dimensions can be generated. A solution could be a GUI, or integration into a GIS-BIM-application. Implementing a GUI or API would be simpler if the entire process were programmed in OOP. No API to the ontology would be needed and *if-then* rules can also be implemented in OOP. SWRL also has limitations: although it can in principle create new instances or properties, most reasoners do not support this functionality. Also, the Open World Assumption of OWL 2 makes it hard to check for non-existent instances (O'Connor et al., 2005).

The integrated rules apply only to new German road bridge caps. Thus, when designing a railway bridge or a French bridge, the ontology cannot be used. Also, the rules from the RiZ-ING the RiZ-ING rules are recommenda-

tions rather than legal requirements i.e., it is possible to deviate from these rules. Fixed *if-then* rules do not allow such deviations which also complicates transferring the approach to other components or another structure. In this case, a highly standardized component is needed, where every dimension can be derived using rules. In the field of bridge design, this condition mainly applies to equipment components e.g., the pavement, railing or traffic barrier. But, possibly, a manual intermediate step is needed here for a complete component design. For an automated design of the whole bridge, generative artificial intelligence is needed, but can be supported by rules implemented in an ontology.

Nevertheless, this exemplary automated dimensioning shows the potential of ontology applications in the construction sector. As stated above, in the AEC sector, ontologies and semantic rules have not yet been used to create new structural designs from the structure's boundary conditions. Especially in the field of bridge engineering, using ontologies for bridge design and guideline implementation is a new topic.

In principle, it is possible to transfer the process to other bridge components if they are as highly standardized as the bridge cap. The implementation of different kinds of rules still needs to be tested. Also, a human in the loop-implementation is imaginable if not every design decision can be automatized. When embedded in a well-adapted application, it could save bridge engineers significant time by avoiding repeatedly dimensioning the bridge cap or other standardized components every time for a new round of design circle (see Figure 5) but just to enter the boundary conditions and to generate the dimensions automatically. With that, they can concentrate on the more important decisions in the substructure and superstructure planning.

Conclusions

Bridge design consists of a vast amount of boundary conditions to consider and of decisions to make. This work proposes a way of automating the dimensioning of highly standardized bridge components, to reduce the engineer's workload. German road bridge guidelines are translated into *if-then*-rules defining the dimensions of the bridge caps for specific boundary conditions. The rules are implemented in an ontology using SWRL. The ontology in combination with a reasoner can be used to automatically generate the cap's dimensions. The process was applied to an exemplary imaginary bridge. The findings of this work are:

1. Information in guidelines and standards is contained in different forms and formats. For an automated design process, the information needs to be translated into *if-then*-rules, which can be transformed into programming code and applied to the respective design task. In this work, the *if-then*-rules were transformed into SWRL-rules.
2. It is possible to generate the bridge cap's dimensions using an ontology, SWRL and a reasoner. For that, the boundary conditions of the bridge and the bridge components need to be created as instances in the ontology. The proposed process can be transferred to other components, as long as they are highly standardized and the content of these standards can be translated into *if-then*-rules.
3. There are several limitations of this process: user friendliness, applicability and expandability. These limitations can be met using different state of the art-technologies such as a GIS-BIM-application, generative AI and newly developed APIs or by creating a well-adapted GUI and manual intermediate steps and intermediate decisions into the bridge design process.

Acknowledgments

The research project "KI-basierter Brückengenerator mit Wissens- und Erfahrungsdaten und früher Bürgerbeteiligung - HyBridGen" and the research project "Nutzung digitaler Regelwerke zur effizienten Planung und Prüfung von BIM-basierten Brückenentwürfen für die Eisenbahn - SmartBridgeRules" are funded by the German Federal Ministry for Transport within the framework of the mFUND funding program (Grant No. 01FV2067A and 01F1228B).

References

Aqlan, S., Hamdan, A.-H., Marx, S., and Kang, C. (2025). Ontology-driven Database for Integrating Bridge Data from Multiple Open Sources.

Beetz, J., Van Leeuwen, J., and De Vries, B. (2009). IfcOWL: A case of transforming EXPRESS schemas into ontologies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 23(1):89–101.

Bundesanstalt für Straßen- und Verkehrswesen (2023). Richtzeichnungen für Ingenieurbauten - RiZ-ING.

Bundesanstalt für Straßen- und Verkehrswesen (2024). Richtlinien für den Entwurf, die konstruktive Ausbildung und Ausstattung von Ingenieurbauten - RE-ING.

CEN/TC 250 Structural Eurocodes, CEN/TC 250/SC 2 Eurocode 2: Design of concrete structures, and NA 005-07-01 AA Design of concrete structures (national mirror committee for CEN/TC 250/SC 2) (2023). EN 1992 - Eurocode 2 - Design of concrete structures.

DIN-Normenausschuss Bauwesen (NABau), DIN Standards Committee Building and Civil Energeering, and NA 005-07-01 AA Design of concrete structures (national mirror committee for CEN/TC 250/SC 2) (2023). DIN 1045 - Concrete Structures.

Erdene, K., Kwon, T. H., and Lee, S.-H. (2020). Integration of Extended IFC-BIM and Ontology for Information Management of Bridge Inspection. Journal of the

Computational Structural Engineering Institute of Korea, 33(6):411–417.

Farghaly, K., Soman, R. K., and Zhou, S. A. (2023). The evolution of ontology in AEC: A two-decade synthesis, application domains, and future directions. Journal of Industrial Information Integration, 36:100519.

Göbels, A. and Beetz, J. (2024). Relative Location Ontology: An ontological model for representing directional topological relationships between spatial entities in oriented space. In Proceedings of the 12th Linked Data in Architecture and Construction Workshop, pages 91–104, Bochum, Germany. CEUR-WS.

Göbels, A. and Beetz, J. (2025). Graph-based method for extracting spatial information from semi-formal text to derive 3D bridge and damage models from legacy maintenance data. Automation in Construction, 178:106418.

Hamdan, A.-H. and Scherer, R. J. (2020). Integration of BIM-related bridge information in an ontological knowledgebase. In 8th Linked Data in Architecture and Construction Workshop - Proceedings, Online.

Hassan, F. U. and Le, T. (2020). Ontology-Based Decoding of Risks Encoded in the Prescriptive Requirements in Bridge Design Codes. Kitakyushu, Japan.

International Organization for Standardization (ISO) (2024). ISO 16739 - Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries.

Jeske, A., Scheil, M., and Castrischer, U. (2022). Parametrisierte Brückenmodellierung mit PythonParts in Allplan. Bautechnik, 99(7):565–574.

Jäkel, J.-I., Heinlein, E., Morgenstern, H., Kim, H., and Klemt-Albert, K. (2025). System- and Data-integrated linking of digital 3D models of existing bridge structures with Knowledge Graphs of non-destructive diagnostic methods. Journal of Information Technology in Construction, 30:603–630.

Jędrzejewska, A., Leonardi, M. L., Ramonell Cazador, C., Granja, J., Chacón, R., Oliveira, D. V., and Azenha, M. (2026). Eurocode Core Ontology: Building the semantic framework for structural design assistants. Advanced Engineering Informatics, 70:104218.

Li, R., Mo, T., Yang, J., Jiang, S., Li, T., and Liu, Y. (2021). Ontologies-Based Domain Knowledge Modeling and Heterogeneous Sensor Data Integration for Bridge Health Monitoring Systems. IEEE Transactions on Industrial Informatics, 17(1):321–332.

Liu, K. and El-Gohary, N. (2022). Bridge Deterioration Knowledge Ontology for Supporting Bridge Document Analytics. Journal of Construction Engineering and Management, 148(6):04022030.

Musen, M. A. (2015). The protégé project: a look back and a look forward. AI Matters, 1(4):4–12.

O'Connor, M., Knublauch, H., Tu, S., Grosof, B., Dean, M., Grosso, W., and Musen, M. (2005). Supporting Rule System Interoperability on the Semantic Web with SWRL. In The Semantic Web – ISWC 2005, volume 3729, pages 974–986. Springer Berlin Heidelberg, Berlin, Heidelberg. Series Title: Lecture Notes in Computer Science.

Pauwels, P., De Farias, T. M., Zhang, C., Roxin, A., Beetz, J., De Roo, J., and Nicolle, C. (2017). A performance benchmark over semantic rule checking approaches in construction industry. Advanced Engineering Informatics, 33:68–88.

Pauwels, P., Van Deursen, D., Verstraeten, R., De Roo, J., De Meyer, R., Van De Walle, R., and Van Campenhout, J. (2011). A semantic rule checking environment for building performance checking. Automation in Construction, 20(5):506–518.

Rasmussen, M. H., Lefrançois, M., Pauwels, P., Hviid, C. A., and Karlshøj, J. (2019). Managing interrelated project information in AEC Knowledge Graphs. Automation in Construction, 108:102956.

Rasmussen, M. H., Schneider, G. F., and Pauwels, P. (2020). BOT: the Building Topology Ontology of the W3C Linked Building Data Group. Semantic Web Journal.

Ren, G., Ding, R., and Li, H. (2019). Building an ontological knowledgebase for bridge maintenance. Advances in Engineering Software, 130:24–40.

Salama, D. A. and El-Gohary, N. M. (2013). Automated Compliance Checking of Construction Operation Plans Using a Deontology for the Construction Domain. Journal of Computing in Civil Engineering, 27(6):681–698.

Staab, S. and Studer, R., editors (2009). Handbook on ontologies. International handbooks on information systems. Springer, Berlin, 2 edition.

Wu, C., Wu, P., Wang, J., Jiang, R., Chen, M., and Wang, X. (2021). Ontological knowledge base for concrete bridge rehabilitation project management. Automation in Construction, 121:103428.

Zhang, P., Cui, C., Li, C., Zhang, C., and Liu, H. (2021). Holistic Design of Energy Pile Bridge Deicing System With Ontology-Based Multiobjective Decision Making. Frontiers in Materials, 8:710404.

Zhang, Y., Liu, Y., Lei, G., Liu, S., and Liang, P. (2022). An Enhanced Information Retrieval Method Based on Ontology for Bridge Inspection. Applied Sciences, 12(20):10599.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

EXPLORING ADAPTABLE OFFICE FIT-OUT LAYOUTS USING CONFIGURABLE FLOOR PLAN GRAPHS AND LARGE LANGUAGE MODELS

Mikhael Johanes[1], Simona Kocour[1,2], Siyu Tang[1], and Catherine De Wolf[1]
[1]ETH Zurich, Zurich, Switzerland
[2]CTU in Prague, Czech Republic

Abstract

Conventional office fit-outs are designed for fixed requirements, which limits their long-term adaptability. This paper introduces a workflow that integrates configurable floor-plan graphs with large language models (LLMs) to enable adaptable office layout exploration. Interior space is represented as a graph that encodes geometry, adjacency, and accessibility, and is parameterised by binary partition decisions. For each candidate configuration, an LLM infers program assignments across multiple scenarios, and solutions are evaluated in a multi-objective combinatorial optimisation. Results show configurations with feasible performance across scenarios, suggesting that LLM-based program assignment can serve as a flexible semantic component that complements quantitative solver-based methods.

Introduction

Offices are constantly evolving. Changes in organisational structures, workforce patterns, and work styles require interiors to adapt. However, this adaptation often comes with a cost: frequent renovations lead to significant environmental impacts, with emissions and waste from repeated fit-out cycles exceeding those of the building itself during its lifetime (Casas Arredondo, 2021; Forsythe and Wilkinson, 2015).

Circular fit-out design has emerged as a response to the environmental consequences of short fit-out lifecycles. In circular construction, adaptability is a core design requirement that increases the ability to accommodate changes in use, environment, and specification throughout time without extensive refurbishment, enabling longer building lifespans and lower embodied greenhouse gas emissions (Watt et al., 2023). Rather than optimising interiors for a single functional state, adaptable fit-outs explicitly account for change over time, by extending the design scope to encompass successive use scenarios within the same spatial system.

Computational generative-design methods offer a promising means to support such adaptability by systematically generating and evaluating alternative spatial configurations under varying requirements. However, most existing generative-design paradigms remain orientated toward single-scenario or fixed specifications. Implicitly assuming stable functional requirements, they optimise layouts for a single, fixed programmatic state. As a result, adaptability is typically treated as an evaluative afterthought rather than a generative principle embedded in the design process itself.

This research addresses the limitation of single-scenario design by decoupling spatial configuration from programmatic assignment. We investigate how floor plan graphs, in which spaces are modelled as nodes and connectivity is encoded as edges, can support large language models (LLMs) in assigning program labels for a given spatial configuration under different scenarios. Multi-objective optimisation is employed to find configurations that maintain acceptable performance across multiple programmatic scenarios.

The hypothesis is threefold. First, separating spatial configuration from programmatic assignment enables the modelling of multifunctionality, in which a single space can accommodate multiple uses. Second, floor plan graphs act as an intermediate representation bridging geometric and semantic interpretation. Third, LLM-based programmatic assignment may provide a flexible semantic labelling step that complements quantitative constraint-based solvers and helps relate underspecified preferences to optimisation objectives.

The proposed workflow supports finding spatial configurations that can accommodate multiple programmatic scenarios (Figure 1). By operationalising adaptability in the workflow, this work provides a method for exploring spatial configurations that remain viable across multiple future use scenarios, supporting longer fit-out lifecycles.

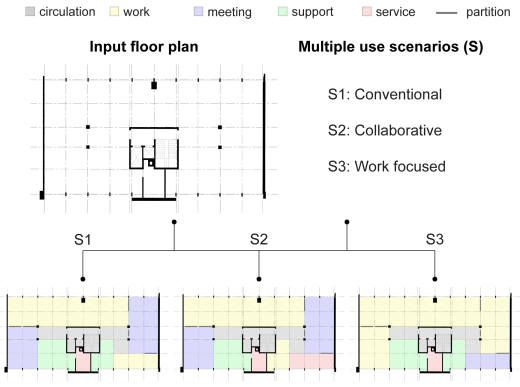


Figure 1: Adaptive layout objective. Given an input floor plan and multiple scenario requirements, we seek layouts that accommodate multiple programs without reconfiguration.

Background

Modelling spatial adaptability in circular fit-out

Adaptability operates on multiple temporal scales, ranging from short-term adjustments such as moving furniture, through medium-term reconfiguration of spatial layouts, to long-term transformations including changes in building use (Schmidt-III and Austin, 2016; Watt et al., 2023). For office interiors, adaptability is primarily situated on a short- to medium-term scale, where frequent organisational and operational changes require interior interventions. To minimise environmental impact, we posit that adaptability at these scales depends on the latent capacity of a spatial layout to support multiple configurations with minimal structural intervention.

The spatial capacity represented by the dimensions of space and the position of durable elements plays a significant role in the definition of the future adaptability (Durmišević, 2018). Graph-based representations offer an opportunity to model such spatial capacity. For example, Pacquée et al. (2025) developed a graph-based method to encode the topology and geometry of a building, allowing comparison between reference projects to assess their potential for future reuse. At the scale of architectural layouts, Herthogs et al. (2019) modelled floor plans as spatial graphs and quantified their capacity for change through measures of topological depth and path redundancy as a proxy of the layout's ability to support multiple programmatic scenarios over time.

Building on this body of work, we adopt spatial graphs not primarily as an evaluative tool to measure adaptability but as a configurable representation to systematically explore alternative layout configurations under varying programmatic scenarios.

Generative design of an adaptable office

Generative spatial layout design has been widely studied in computational design and space-planning research, which commonly distinguishes between knowledge-driven and data-driven approaches (Yan et al., 2025). Knowledge-driven methods generate layouts by explicitly structuring the design through rules, constraints, and objectives. They encompass rule-based systems, heuristic and mathematical optimisation, and reinforcement learning. Architectural intent is explicitly encoded at different stages of the generative process: rule-based methods formalise the design intent through predefined spatial rules and discretization schemes (Sheijani et al., 2024), parametric workflows couple parametric models with multi-objective evolutionary optimisation to explore trade-offs among evaluation criteria (Nagy et al., 2017), and hybrid approaches combine constraint programming with evolutionary search to generate feasible layouts within fixed envelopes (Laignel et al., 2021). More recently, reinforcement learning has been used to incorporate spatial constraints and objectives directly into state, action, and reward formulations (Zhang et al., 2024b). Although considerably effective, these approaches rely on extensive explicit knowledge encoding, and thus face scalability limits for real-world problems that are often open-ended and incomplete.

Data-driven approaches generate spatial layouts by learning patterns from given datasets; for example, they map partial design input such as text prompts, bubble diagrams, or footprint constraints to complete spatial layouts. Data-driven approach includes statistical generative models, deep neural networks, and example-based or case-based generation methods. Representative examples include models based on the generative adversarial network (GAN) such as HouseGAN (Nauata et al., 2020), diffusion-based approaches such as HouseDiffusion (Shabani et al., 2023), and transformer-based models such as MaskPLAN (Zhang et al., 2024a). However, the limited availability of large, high-quality architectural datasets has mostly confined these methods to residential typologies.

Beyond these limitations, most generative-design methods optimise layouts for a single design state, limiting future adaptability. We address this limitation by decoupling spatial configuration from programmatic assignment. This separation allows multiple program interpretations to be inferred from the same spatial configuration in response to different use scenarios.

LLM graph-based spatial reasoning in generative design

Recent advances in LLMs have shown growing capabilities in symbolic and relational tasks, motivating their integration into generative-design workflows for spatial layout problems. Rather than directly generating precise geometry, most approaches use LLMs as high-level semantic or planning components operating over abstract spatial representations, such as graphs, symbolic relations, or procedural descriptions. At the interior scale, LayoutGPT frames layout generation as a compositional planning problem using in-context examples, in which LLMs translate natural language prompts into structured descriptions of objects and their spatial properties in a CSS-like format (Feng et al., 2023). FlairGPT uses LLMs to generate zones, objects, and inter-object constraints, which are organised into a layout constraint graph and solved using standard optimisation algorithms (Littlefair et al., 2025). At the building scale, Text2BIM extends this direction through a multi-agent framework that generates structured procedural instructions for BIM authoring tools (Du et al., 2026). Collectively, these works suggest that LLMs are most effective when applied to high-level semantic and relational tasks, while detailed geometric realisation is delegated to rule-based, procedural, or optimisation-based systems.

Method for graph-based fit-out design explorations

Overview

We propose a generative workflow for an adaptable layout design that integrates spatial-configuration optimisation with scenario-driven programmatic assignment (Figure 2). The workflow starts with two inputs: an input floor

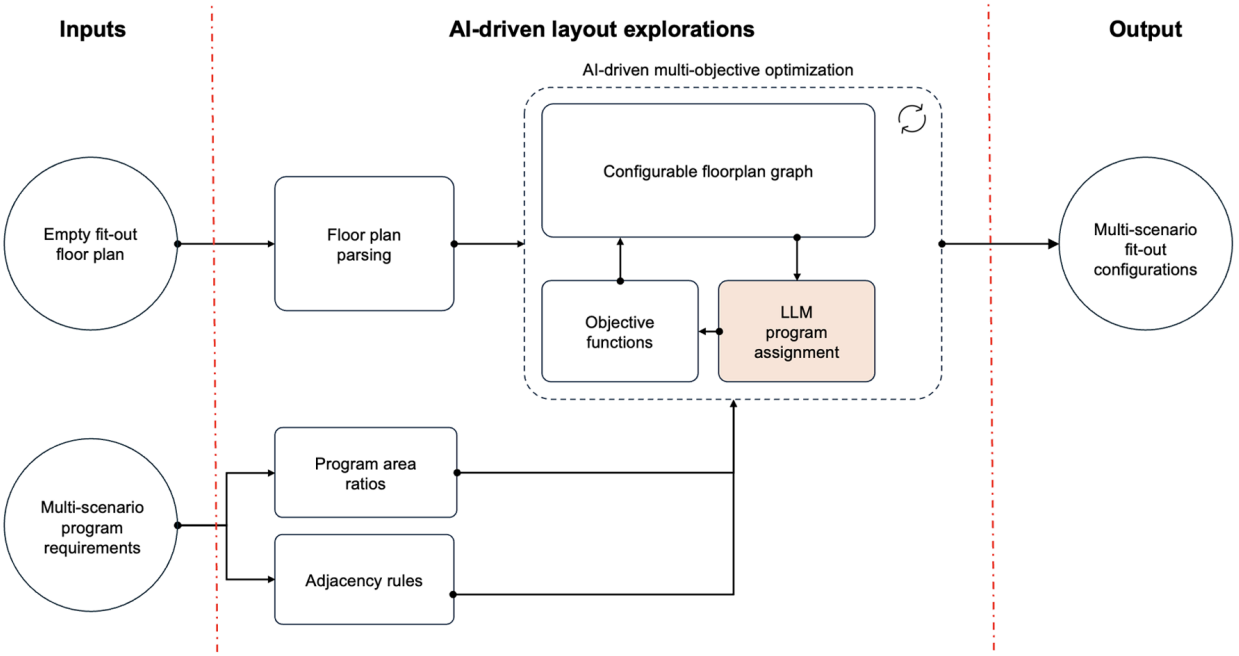


Figure 2: Overview of graph-based fit-out layout explorations with LLM

plan and a set of use scenarios. The floor plan is parsed into a configurable floor plan graph that defines a fixed topological layout independent of the program. The programmatic requirements of each scenario are delegated to an LLM, which assigns zone labels to the floor plan graph according to the area-ratio targets and adjacency rules. A multi-objective evolutionary algorithm optimises the spatial configuration using these criteria, yielding layouts that can accommodate multiple scenarios.

Parsing the input floor plan into a configurable graph

The input floor plan contains geometric information on the existing structure, walls, doors, and windows as well as predefined zones, such as the entrance and the service areas, and an underlying planning grid used to discretise the layout. The parsing module converts this input into a parametric representation by discretising the floor area into cells and introducing binary decision variables, x, that govern the placement of partitions between adjacent cells (Figure 3). From this representation, an intermediate visibility graph is constructed, and strongly connected cells are clustered to produce a configurable floor plan graph, $G_{fp}(x)$. Changing x, the optimisation algorithm explores alternative spatial configurations, each of which is reflected in a corresponding instance of $G_{fp}(x)$. The resulting graph representation then serves as input for the LLM modules to infer programmatic assignments under different use scenarios.

Programmatic requirements for multi-scenario inputs

Programmatic requirements are specified as a set of discrete use scenarios, each of which is translated into two complementary specifications: (1) target program area ratios and (2) adjacency rules between functional zones.

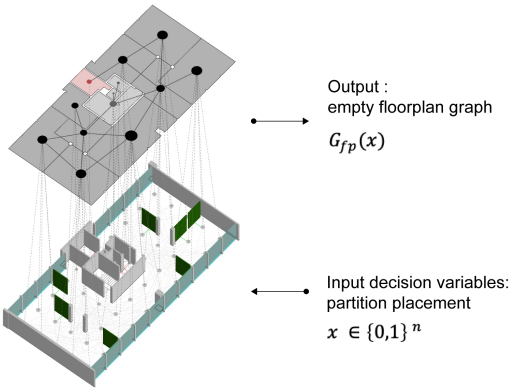


Figure 3: The empty floor plan is parsed into a configurable graph, where binary decision variables defined on the planning grid control the placement of partitions.

This formulation allows multiple future programs to be evaluated against a shared spatial configuration. In this preliminary study, each scenario defines target area ratios for five abstract zone types: Work (W), Meeting (M), Support (Su), Service (Se), and Circulation (C) (Table 1). In parallel, qualitative spatial relationships are encoded as hop-based constraints on the floor plan graph (Table 2). Required adjacencies (R_1) and forbidden adjacencies (F_1) are enforced as hard constraints, defined in terms of graph distance between zones, while preferred proximities (P_2) are modelled as soft objectives that bias solutions toward functional coherence without over-constraining the layout. Circulation is constrained to maintain direct adjacency to all other zone types, ensuring global accessibility across configurations.

For each scenario, s, the programmatic requirements are

represented by a pair (A_s, R_s), where A_s specifies the ratios of the target program area, and R_s specifies the adjacency rules. These scenario specifications serve two functions in the workflow. First, they are translated into a structured prompt context for the LLM, which assigns zone labels to the nodes of the input floor plan graph for each scenario. Second, they parameterise the objective functions used during evolutionary optimisation. Specifically, deviations from A_s and violations of R_s are quantified through scenario-wise evaluations, which are subsequently aggregated across scenarios to assess the adaptability of a given spatial configuration.

Table 1: Program scenarios as target area ratios (sum = 1.0)

Scenario	W	M	Su	Se	C
S1	0.50	0.20	0.10	0.10	0.10
S2	0.40	0.30	0.10	0.10	0.10
S3	0.65	0.10	0.10	0.05	0.10

Table 2: Adjacency and proximity constraints encoded as hop-based relations

From\To	W	M	Su	Se	C
W	–	P_2	P_2	F_1	R_1
M	P_2	–	–	F_1	R_1
Su	–	–	–	P_2	R_1
Se	F_1	F_1	P_2	–	R_1
C	R_1	R_1	R_1	R_1	–

LLM programmatic inference

Given a candidate floor plan graph, G_{fp}, and a scenario-specific program configuration (A_s, R_s), we construct a textual prompt, P_s, consisting of a program block and a rules block. The floor plan graph, G_{fp}, is provided to the LLM as a structured JSON representation encoding global, node, and edge properties, including spatial attributes such as area, perimeter components, bounding boxes, aspect ratios, and interface characteristics between adjacent spaces. The program block derives the total usable area from the node attributes in G_{fp} and specifies the target area for each zone type in A_s, reported in absolute and relative terms. The rules block serialises the adjacency constraints in R_s as hard and soft relations with associated predicates and targets. A structured system instruction and output template guide the LLM to assign zone labels within a predefined schema without modifying the underlying spatial topology. The instruction embeds the scenario context, P_s, together with the list of valid zone types. Outputs that do not conform to the expected schema or label set are discarded, and the original unlabelled floor plan graph is retained for evaluation. Analysis of the reported experiments suggests that such fallback was negligible, occurring in at most 1 of 5000 evaluated rows for the LLM-based setups. The final prompt is formed by concatenating the two blocks (1), and the LLM receives G_{fp} and P_s to assign zone labels, producing an updated graph, G_{fp}^{llm} (2).

$$P_s(G_{fp}, A_s, R_s) = P_{\text{prog}}(G_{fp}, A_s) \parallel P_{\text{rules}}(R_s) \quad (1)$$

$$G_{fp}^{\text{llm}} \leftarrow \text{LLM}(G_{fp}, P_s) \quad (2)$$

Multi-objective optimisation

Optimisation variables are the binary partition decisions, x, defined in the planning grid. For each candidate, x, a floor plan graph, $G_{fp}(x)$, is generated and tested for connectivity. Only connected configurations invoke LLM inference; disconnected graphs are excluded using a hard feasibility constraint.

For each scenario, $s \in S$, the LLM assigns zone labels to $G_{fp}(x)$ using the corresponding program area ratios and adjacency rules, producing a set of inferred graphs, $\{G_{fp}^{\text{llm}}(s)\}$. Then, two objective functions are evaluated and aggregated across scenarios: program-ratio fitness and rule compliance. The fitness of the program ratio measures the deviation between the achieved and target area ratios using the mean absolute error (MAE). Rule compliance evaluates the adherence to required, forbidden, and preferred adjacencies using hop-based scoring, normalised to a score in the range $[0, 100]$. The resulting bi-objective optimisation problem is defined as:

$$f_1(x) = \frac{1}{|S|} \sum_{s \in S} \text{MAE}_{\text{area}}\left(G_{fp}^{\text{llm}}(s), A_s\right), \quad (3)$$

$$f_2(x) = -\frac{1}{|S|} \sum_{s \in S} \text{Score}_{\text{rules}}\left(G_{fp}^{\text{llm}}(s)\right), \quad (4)$$

$$\text{subject to } G_{fp}(x) \in \mathcal{G}_{\text{conn}}. \quad (5)$$

We solve this bi-objective optimisation problem using NSGA-II (Deb et al., 2002). The initial population is sampled as sparse binary vectors to promote diversity in partition configurations. Optimisation terminates after a fixed number of generations or when changes in decision variables and objective values fall below a predefined tolerance.

Case study and discussion

Based on an input office floor plan with a total area of approximately 400 m^2 (Figure 1) and the multi-scenario programmatic specifications in Tables 1 and 2, we conducted a series of experiments to evaluate the proposed workflow using different LLMs and a baseline approach, assessing their performance across optimisation quality, spatial regularity and topological complexity of the resulting layouts.

Experiment setup

LLM-based programmatic inference is executed using a locally hosted Ollama server accessed through the Python `ollama` client, keeping all inference local to the experimental environment. We evaluate three large language

models with increasing capacity: `phi4:14b`, `deepseek-r1:32b`, and `gpt-oss:120b`, where model size (in billions of parameters) provides a rough indication of overall model capability.

To ensure reproducibility, deterministic decoding is used where possible. Greedy decoding (temperature = 0.0) is applied for `phi4:14b` and `deepseek-r1:32b`, producing consistent outputs across runs. For `gpt-oss:120b`, temperature = 0.0 proved unstable in our Ollama setup; therefore, we retain the model's default sampling configuration (temperature = 1.0) and set `thinking=low`, leaving other decoding parameters unchanged.

As a baseline, we implement a constraint-programming model using OR-Tools CP-SAT (Perron et al., 2023). Following established discrete space allocation formulations (Laignel et al., 2021), the solver assigns a single zone label to each node of the floor plan graph using a one-hot encoding while keeping the topology fixed and respecting predefined labels. The objective minimises deviations from target area ratios and penalties for soft spatial preferences, such as proximity and path coherence. Hop-based adjacency rules are converted into weighted penalties. Circulation accessibility is approximated by enforcing connectivity among circulation-labelled nodes using a predecessor-tree formulation.

Optimisation of spatial configuration is performed using NSGA-II implemented in an existing Python package (Blank and Deb, 2020), with a population size of 50 evolved over 100 generations. The initial population is generated through Bernoulli sampling with an open-edge probability of 0.1, biasing the search toward sparse partition configurations in early generations. Results are reported from a single optimisation run per setup with a fixed random seed. For each optimisation run, LLM inference consumed an estimated 49.9 million tokens in total, including 46.0 million input tokens and 3.9 million output tokens.

We evaluate the results in three complementary dimensions. First, optimisation performance is assessed by comparing Pareto fronts across setups using hypervolume (HV) (Fonseca et al., 2006) and inverted generational distance plus (IGD+) (Ishibuchi et al., 2015). Both indicators are computed on normalised objective values, with objectives scaled to $[0, 1]$ across all experiments. For HV, a fixed reference point of $(1.1, 1.1)$ is used to ensure dominance over all feasible solutions. IGD+ is computed with respect to the union of non-dominated solutions obtained across all setups, serving as a common reference set for convergence comparison. These metrics jointly quantify trade-off quality and convergence with respect to program area-ratio fitness and rule compliance. Second, the shape complexity of the generated layouts is evaluated following Brinkhoff et al. (1995). The measure captures the irregularity of the boundaries and the convexity of the shapes of individual rooms, weighted by room area, providing an indicator of spatial regularity and constructibility. Third, the topological complexity of Pareto-optimal floor plan graphs is evaluated using the cyclomatic number (McCabe, 1976), which captures loop redundancy in adjacency and circulation networks and indicates the potential for alternative routing and future adaptability.

Pareto fronts

The comparison of Pareto fronts across different setups is shown in Figure 4 and Table 3. A clear distinction emerges between constraint-based and LLM-based programmatic inference. The CP-SAT baseline consistently achieves higher HV and lower IGD+, reflecting its ability to exploit a fully specified optimisation model under explicitly encoded objectives. In contrast, all LLM-based setups exhibit lower Pareto-front quality, with broader but less extreme fronts, indicating weaker convergence within the defined objective space.

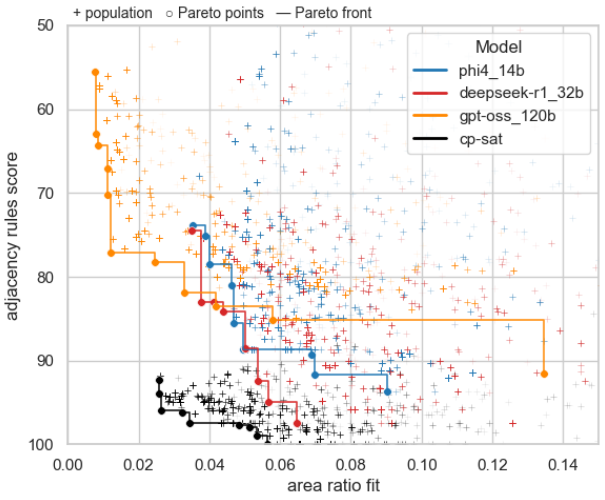


Figure 4: Pareto fronts from the optimisation results

Table 3: Pareto-front quality metrics across setups

Setup	HV ↑	IGD+ ↓
Phi-4	1.051	0.079
DeepSeek-R1	1.092	0.060
GPT-OSS	1.086	0.070
CP-SAT	1.153	0.018

Among the LLM-based configurations, optimisation performance does not consistently improve with larger model size. DeepSeek-R1:32B achieves the highest HV and lowest IGD+ among the LLM variants, suggesting a more balanced trade-off between program area-ratio fitness and adjacency rule compliance. GPT-OSS:120B tends to favour program area-ratio fitness, achieving lower area deviation at the expense of adjacency rule compliance, while Phi-4:14B shows the weakest overall performance across both objectives. Overall, LLM-based programmatic inference remains substantially below the CP-SAT baseline and is better understood as a flexible program-assignment mechanism rather than a replacement for exact optimisation.

These observations point to a hybrid direction for future work, in which hard constraints are handled by classical

solvers while LLMs support more qualitative or underspecified aspects of program assignment.

Floor plan complexity

To further characterise the generated layouts, we use shape complexity and cyclomatic number as proxies for geometric and topological complexity, respectively. Both metrics are evaluated for the Pareto-optimal solutions of each setup, with their distributions shown in Figure 5.

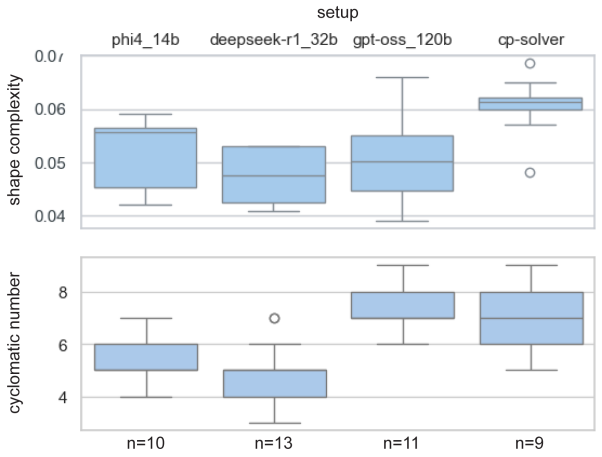


Figure 5: Comparison of room shape and topological complexity of Pareto-optimal floor plan graphs generated by different setups

The results reveal clear differences between the CP-SAT baseline and the LLM-based setups. While the CP-SAT solver identifies Pareto-optimal solutions that more closely satisfy the explicitly defined objectives, it also produces layouts with more irregular room geometries. This indicates a geometric trade-off associated with closely fitting the optimisation objectives.

A separate pattern appears in the topological metrics. The higher distribution of cyclomatic complexity indicates that CP-SAT also explores configurations with more complex adjacency and circulation structures. This is broadly consistent with (Herthogs et al., 2019), who related spatial-graph measures such as topological depth and path redundancy to a layout's capacity for change across programmatic scenarios. From this perspective, higher topological complexity may indicate greater capacity for change across programmatic scenarios, as increased loop redundancy can support alternative spatial relationships.

In contrast, LLM-based setups tend to favour structurally simpler layouts. Phi-4:14B and DeepSeek-R1:32B both produce layouts with simpler room geometries and lower cyclomatic complexity, with DeepSeek-R1:32B showing the strongest reduction in complexity. However, from the perspective of graph-based adaptability, this simplification may also imply lower path redundancy and reduced capacity to accommodate alternative programmatic scenarios. GPT-OSS:120B departs from this pattern, exhibiting higher topological complexity and greater variability in room shape complexity, suggesting that stronger optimisation of program area ratios may again introduce more complex spatial structures.

Multi-scenario layout configurations

Figure 6 presents representative multi-scenario layout configurations obtained from each setup. Each column shows a single spatial configuration evaluated in three scenarios, together with summary metrics on geometric complexity, topological complexity, and objective performance. The qualitative observations are consistent with the quantitative analysis reported earlier.

The CP-SAT baseline is a strong benchmark when program requirements are clearly defined. By directly optimising area ratios and adjacency constraints, it consistently achieves the best objective performance, underscoring the value of explicit formulations: when objectives and constraints can be clearly encoded, exact optimisation is most effective. In contrast, LLM-based setups use floor plan graph features and in-context prompting to assign program labels across scenarios. They do not reliably match CP-SAT's objective values and show variable geometric and topological complexity, but they offer an alternative assignment method when scenario-specific formulations are underspecified or difficult to encode explicitly.

Representative samples from `phi4:14b` and `deepseek-r1:32b` balance objective performance with relatively low topological complexity, with `phi4:14b` showing higher geometric complexity than `deepseek-r1:32b`. `gpt-oss:120b` instead produces layouts with lower geometric complexity but more rooms and connections, increasing topological complexity.

Separating spatial configuration search from programmatic assignment allows adaptable layout exploration, since one configuration can be evaluated across scenarios while optimisation targets layouts that support high-quality assignments. In this setup, explicit constraint solving is effective when program requirements are clearly specified, as in the CP-SAT baseline. LLM-based inference is generally less optimal but provides an alternative when scenario requirements are less fully specified or more difficult to encode explicitly.

Limitations

The proposed workflow is intended as an exploratory prototype and several limitations should be addressed in future work. Although the graph-based representation enables LLM-based programmatic inference, a clear performance gap remains compared to the constraint-based baseline, and embedding LLM inference within the optimisation loop introduces substantial computational overhead. This finding motivates more selective hybrid workflows in which well-defined objectives and hard constraints are handled by solvers, while LLM-based programmatic inference is applied only to a smaller set of promising candidate configurations. In addition, the reported experiments are limited to a single optimisation run on one small office floor plate and a set of simplified programmatic sce-

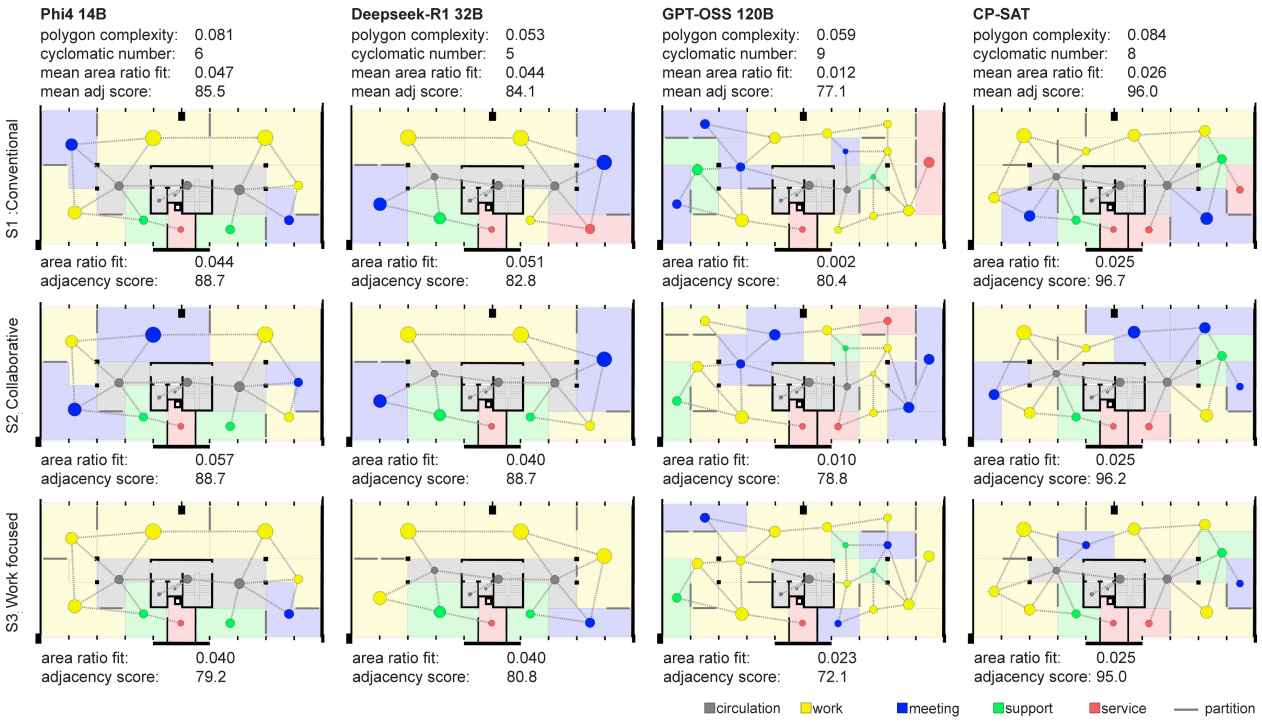


Figure 6: Representative sample of generated multi-scenario floor plans from across setups

narios with only modest structural differences, so the findings should be read as exploratory rather than generalisable. Finally, the evaluation focuses on proxy metrics to assess the feasibility of the results; more comprehensive and practice-orientated measures, including direct indicators of adaptability, circularity, and environmental impact, are required to evaluate the effectiveness of the method in reducing the embodied greenhouse gas emissions associated with office refurbishments.

Conclusion and outlook

This paper presented a graph-based computational workflow for exploring adaptable office fit-out layouts across multiple future use scenarios. By decoupling spatial configuration from programmatic assignment and delegating scenario-specific program assignment to LLMs, the approach supports the systematic exploration of spatial layouts that remain viable across changing requirements. The results show that, while LLM-based programmatic inference does not match the optimisation performance of fully specified constraint-based models, it can generate diverse configurations without requiring exhaustive re-encoding of programmatic constraints for each scenario. This suggests that LLMs are better understood as flexible semantic components within a hybrid optimisation workflow rather than as replacements for exact optimisation.

Looking ahead, the findings point toward more selective hybrid generative-design workflows in which classical optimisation methods enforce hard constraints and quantitative objectives, while LLM-based programmatic inference is applied to a reduced set of promising candidate layouts. Future work should extend the method to larger and more heterogeneous floor plans, richer and more realistic program definitions, and a broader set of evaluation criteria relevant to adaptability, circularity, and environmental impact. Incorporating downstream fit-out feasibility and reuse-related assessments would further support practical application in circular fit-out design.

Acknowledgements

This work was supported through a Design++ fellowship program at ETH Zurich together with Integral Design-Build AG and Halter Group AG. A collaborating author was supported by the Czech Science Foundation (GACR) EXPRO (grant no. 23-07973X). The authors thank Dr. Roxanne Goldberg for editing support.

References

Blank, J. and Deb, K. (2020). pymoo: Multi-objective optimization in python. IEEE Access, 8:89497–89509.

Brinkhoff, T., Kriegel, H.-P., Schneider, R., and Braun, A. (1995). Measuring the Complexity of Polygonal Objects. In ACM-GIS, volume 109.

Casas Arredondo, J. M. (2021). Circular economy and office fit-out: an analysis for office fit-out processes based on material flows. Doctoral, UCL (University College London).

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algo-

rithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2):182–197.

Du, C., Esser, S., Nousias, S., and Borrmann, A. (2026). Text2BIM: Generating Building Models Using a Large Language Model-Based Multiagent Framework. Journal of Computing in Civil Engineering, 40(2):04025142. Publisher: American Society of Civil Engineers.

Durmišević, E. (2018). Reversible building design. In Designing for the Circular Economy. Routledge. Num Pages: 16.

Feng, W., Zhu, W., Fu, T.-J., Jampani, V., Akula, A., He, X., Basu, S., Wang, X. E., and Wang, W. Y. (2023). LayoutGPT: Compositional Visual Planning and Generation with Large Language Models. In Advances in Neural Information Processing Systems, volume 36, pages 18225–18250. Curran Associates, Inc.

Fonseca, C., Paquete, L., and Lopez-Ibanez, M. (2006). An Improved Dimension-Sweep Algorithm for the Hypervolume Indicator. In 2006 IEEE International Conference on Evolutionary Computation, pages 1157–1163.

Forsythe, P. and Wilkinson, S. (2015). Measuring office fit-out changes to determine recurring embodied energy in building life cycle assessment. Facilities, 33(3-4):262–274.

Herthogs, P., Debacker, W., Tunçer, B., Weerdt, Y. D., Temmerman, N. D., Herthogs, P., Debacker, W., Tunçer, B., Weerdt, Y. D., and Temmerman, N. D. (2019). Quantifying the Generality and Adaptability of Building Layouts Using Weighted Graphs: The SAGA Method. Buildings, 9(4).

Ishibuchi, H., Masuda, H., Tanigaki, Y., and Nojima, Y. (2015). Modified Distance Calculation in Generational Distance and Inverted Generational Distance.

Laignel, G., Pozin, N., Geffrier, X., Delevaux, L., Brun, F., and Dolla, B. (2021). Floor plan generation through a mixed constraint programming-genetic optimization approach. Automation in Construction, 123:103491.

Littlefair, G., Dutt, N. S., and Mitra, N. J. (2025). FlairGPT: Repurposing LLMs for Interior Designs. Computer Graphics Forum, 44(2):e70036. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.70036.

McCabe, T. (1976). A Complexity Measure. IEEE Transactions on Software Engineering, SE-2(4):308–320.

Nagy, D., Lau, D., Locke, J., Stoddart, J., Villaggi, L., Wang, R., Zhao, D., and Benjamin, D. (2017). Project discover: an application of generative design for architectural space planning. In Proceedings of the Symposium on Simulation for Architecture and Urban Design, SIMAUD '17, pages 1–8, San Diego, CA, USA. Society for Computer Simulation International.

Nauata, N., Chang, K.-H., Cheng, C.-Y., Mori, G., and Furukawa, Y. (2020). House-GAN: Relational Generative Adversarial Networks for Graph-Constrained House Layout Generation. In Vedaldi, A., Bischof, H., Brox, T., and Frahm, J.-M., editors, Computer Vision – ECCV 2020, Lecture Notes in Computer Science, pages 162–177, Cham. Springer International Publishing.

Pacquée, R., Mueller, C., and Rinke, M. (2025). Envisioning alternative buildings: Graph tool shows spatial reuse capacity and informs open-ended design interventions. In Structures and Architecture. CRC Press. Num Pages: 8.

Perron, L., Didier, F., and Gay, S. (2023). The cp-sat-lp solver. In Yap, R. H. C., editor, 29th International Conference on Principles and Practice of Constraint Programming (CP 2023), volume 280 of Leibniz International Proceedings in Informatics (LIPIcs), pages 3:1–3:2, Dagstuhl, Germany. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

Schmidt-III, R. and Austin, S. (2016). Adaptable Architecture. Routledge, 0 edition.

Shabani, M. A., Hosseini, S., and Furukawa, Y. (2023). HouseDiffusion: Vector Floorplan Generation via a Diffusion Model with Discrete and Continuous Denoising. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 5466–5475, Vancouver, BC, Canada. IEEE.

Sheijani, S. S., Momenaei, A., and Hassanzade, H. (2024). Automating modular open-plan office layouts with performance-based generative design. Automation in Construction, 167:105692.

Watt, H., Davison, B., Hodgson, P., Kitching, C., and Densley Tingley, D. (2023). What should an adaptable building look like? Resources, Conservation & Recycling Advances, 18:200158.

Yan, S., Wu, C., and Zhang, Y. (2025). Generative design for architectural spatial layouts: a review of technical approaches. Journal of Asian Architecture and Building Engineering, 0(0):1–21. Publisher: Routledge _eprint: https://doi.org/10.1080/13467581.2025.2512235.

Zhang, H., Savov, A., and Dillenburger, B. (2024a). MaskPLAN: Masked Generative Layout Planning from Partial Input. pages 8964–8973.

Zhang, Z., Guo, Z., Zheng, H., Li, Z., and Yuan, P. F. (2024b). Automated architectural spatial composition via multi-agent deep reinforcement learning for building renovation. Automation in Construction, 167:105702.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

FEW-SHOT BUILDING ELEMENT DETECTION WITH VISION TRANSFORMERS: IMPROVING ENERGY RETROFIT MATCHING

Tom Slik[1], Raphaël Gueulet[1], Christos Lathourakis[1]
[1]Netherlands Organisation for Applied Scientific Research

Abstract

Traditional computer vision struggles with domain-specific few-shot building element detection in street-view imagery, creating a bottleneck for extracting the building information needed for data-driven retrofit matching. This paper proposes a few-shot framework that combines a pre-trained Vision Transformer with a Random Forest patch classifier and adaptive cropping to address domain shift and limited annotations. With only 21 labeled images per element, the method achieves mean detection accuracy greater than 0.95 across seven elements and enriches building stock data at address level, enabling more accurate matching of retrofit measures.

Introduction

Accelerating energy retrofits is critical for climate neutrality, yet scaling standardized solutions across millions of dwellings remains challenging due to limited address-level data on retrofit feasibility.

Recent innovation projects (e.g., Energiesprong and INFINITE) and research initiatives have demonstrated that accelerating energy retrofit requires a shift from custom, project-based approaches toward standardized or product-based methodologies that enable scalability through industrialization (van Groesen, 2022; Bjelland et al., 2024).

Large-scale renovation studies often rely on building archetypes, where dwellings are typically grouped by construction period and type. An archetype can be associated with renovation measures that correspond to characteristic insulation and energy performance levels—supporting macro-scale renovation planning, scenario analysis, and cost and energy estimations (Uihlein and Eder, 2010; Coffee et al., 2025; Mata et al., 2014). A comprehensive review of archetype-based modeling underlines both their central role in building stock analysis and their methodological limits, encouraging more adaptive, data-driven approaches (Shen and Wang, 2024).

Despite their practicality, archetypes mask substantial intra-class heterogeneity (Rivallain et al., 2020). The feasibility of a retrofit often depends on factors that archetypes overlook. For instance, clustering a Swiss village dataset yielded more precise representative buildings for realistic large-scale retrofit planning (Geyer et al., 2017), while a machine learning classifier predicted renovation measure applicability across New York City's building stock using detailed building attributes (Marasco and Kontokosta, 2016). Such methods exploit fine-grained characteristics—construction details and heating system attributes—beyond archetype averages and open-data limitations.

This creates a knowledge gap: how to predict retrofit applicability at address level without coarse archetype grouping or manual audits, using scalable feature extraction?

In the Dutch housing stock (8M dwellings), we apply a renovation-centered approach to identify dwellings suitable for standardized renovation solutions at scale, minimizing the need for per-building customization.

To predict retrofit applicability for individual addresses, we created datasets combining open data sources (e.g., cadastral records) for detailed building characteristics with supplier-provided training labels—historical records of hundreds to thousands of addresses labeled with binary applicability values for their standardized solutions. Using these supplier-labeled datasets as ground truth and open data characteristics as input features, we trained machine learning classifiers designed to extrapolate retrofit applicability predictions across the entire national housing stock, provided they achieve sufficient performance.

The accuracy of address-level retrofit predictions depends on how well building features capture aspects relevant to renovation applicability. Relying solely on readily available open-data features, such as cadastral information, provided limited precision for many retrofit solutions. Facade and roof components may strongly influence the feasibility of certain renovation measures. Therefore, we employed computer vision methods to extract building features from street-level imagery. These vision-derived features extend the feature set beyond current open-data sources, thereby enabling more precise applicability predictions.

In this research, we developed a scalable computer vision methodology to extract building element features from street-view imagery, addressing critical knowledge gaps in the Dutch housing stock for data-driven retrofit matching. Our objective is to investigate whether the automated detection of building elements, including solar panels, dormers, chimneys, roof windows, ventilation units, removable panels, and balconies, can improve predictions of renovation solution applicability at the address level. This methodology leverages self-supervised Vision Transformers (ViTs) that learn semantic representations without man-

ual annotation, using a patch-based classification approach on dense patch embeddings. We combined this with an adaptive cropping strategy that maximizes relevant patch samples from bounding box annotations, enabling effective few-shot learning with limited labeled data that reflect real-world scenarios where domain-specific building element datasets are scarce.

This study makes two principal contributions to the existing literature. First, we demonstrated a practical computer vision pipeline requiring minimal labeling across the entire workflow: we extracted building element features from street-view imagery and evaluated detection performance per feature. Second, we evaluated whether augmenting building stock databases with these vision-derived features improved the predictive accuracy of retrofit solution applicability models, providing evidence of computer vision's value in data-driven retrofit matching.

Background and Related Work

Automatic detection of building elements from street-view imagery remains a challenging task in computer vision. This section establishes the motivation, technical foundations, and related work for patch-level detection using self-supervised ViTs. We begin by examining existing approaches to collecting building information from street-view imagery and identify gaps where scalable, data-efficient methods are needed. Next, we explore how self-supervised ViTs have emerged as a promising approach and how their dense patch-level representations can capture semantic information almost directly off the shelf. Finally, we survey how recent research leveraged ViTs patch embeddings for detection and classification.

Collecting Building Information from Street-View Imagery and CNN-based Facade Parsing

Several facade datasets were created for building element detection. The ECP dataset contains 104 images, whereas the eTRIMS 8-Class database comprises 60 images (Liu et al., 2017; Korc and Förstner, 2009). The CMP Facade Database is larger, providing 606 images from diverse architectural styles (Tyleček and Šára, 2013). These datasets cover basic facade components such as windows, walls, balconies, doors, shops, sky, chimneys, roofs, and background elements, with CMP introducing finer architectural details across 12 classes including molding, cornice, and pillar. The large-scale ADE20K dataset extends coverage to diverse scene types and includes certain building elements, such as dormers and chimneys (Zhou et al., 2018). However, annotation consistency analysis revealed missing object labels throughout ADE20K, with objects labeled inconsistently across images. Moreover, certain elements crucial for assessing retrofit applicability remain absent from the existing dataset classes. Consequently, relying only on existing datasets restricts the usefulness of facade imagery for extracting detailed domain-specific information relevant to renovation solution matching.

The existing literature approaches facade parsing using classical deep learning architectures. DeepFacade employs FCN-8s with symmetric regularization for pixel-wise segmentation, and more recent work apply encoder-decoder architectures such as Xception-based networks with pyramidal multi-scale fusion (Liu et al., 2020; Chen et al., 2023). These methods rely on CNNs and are usually pretrained on ImageNet, followed by extensive fine-tuning. However, research has demonstrated that transfer learning with CNNs struggles to generalize across datasets (Zhang et al., 2022; Lin et al., 2021; Li et al., 2020). Distribution shifts, whether from imaging conditions, architectural styles, or geographic contexts, often cause performance degradation. This limitation motivated the exploration of more robust, large-scale pretraining approaches that do not depend on supervised label distributions.

Self-Supervised Vision Transformers for Dense Prediction

Self-supervised ViTs have emerged as powerful alternatives to supervised learning, providing versatile and transferable image representations without requiring labeled data. It was demonstrated that self-supervised ViT features contain explicit semantic information about the image structure; its features are sufficiently dense that attention maps directly reveal semantic correspondences and segmentation patterns (Caron et al., 2021). Building on this foundation, DINOv2 was introduced to scale self-supervised pretraining using a dataset of 142 million diverse images, achieving state-of-the-art performance on dense prediction tasks (Oquab et al., 2023). Next, a key innovation addressed artifacts in ViT attention maps, introducing learnable register tokens that eliminate high-norm outliers and produce cleaner, more interpretable feature maps without performance degradation (Darcet et al., 2023). Most recent advances, including DINOv3, have introduced regularization to stabilize dense feature learning at scale, essentially ensuring that patch representations are well-distributed in the embedding space (Siméoni et al., 2025).

In addition, recent studies on pretrained ViTs have investigated the use and refinement of patch-level representations. STEGO performs unsupervised semantic segmentation by clustering DINO features (Hamilton et al., 2022). In contrast, NeCo employs a retraining approach that refines patch embeddings for nearest-neighbor retrieval of semantically similar patches across and within images (Pariza et al., 2024). As illustrated in Figure 2 and 4 in STEGO and NeCo, respectively, individual patch embeddings can serve as queries to identify semantically similar regions without requiring labels. This capability provides a foundation for the data-efficient building element detection method from street-view imagery proposed in this study.

Classification with Patch Embeddings

Self-supervised ViTs produce rich patch-level representations applicable to downstream tasks, although existing methods typically require additional processing or archi-

tectural modifications to extract class-relevant information. Class-Aware Patch Embedding Adaptation was introduced to adapt patch embeddings via learnable class-aware embeddings (Hao et al., 2023), while Class-relevant Patch Embedding Selection was proposed to filter patches by similarity to a single class embedding (Jiang et al., 2024). Both methods aggregate class-relevant patches through global image representations and succeed on image-level few-shot benchmarks (miniImageNet, tiered-ImageNet, CIFAR-FS, and FC-100), where the entire image belongs to one class, and patches collectively support the image classification. However, building element detection requires patch-level classification of multiple elements present simultaneously in different image regions. Global patch selection biases the model toward overall building features rather than fine-grained architectural elements, filtering out local information critical for precise element classification. Moreover, existing few-shot object detection methods require extensive base class training on thousands of annotations before adapting to novel classes contradicting truly annotation-scarce scenarios (Zhang et al., 2023; Bulat et al., 2023), while vision-language alternatives like GroundingDINO require large-scale pretraining on millions of image-text pairs (Liu et al., 2024).

We adopted a simpler approach: direct classification of patch embeddings using standard machine learning classifiers (k-Nearest Neighbors (kNN), Random Forest (RF), Multilayer Perceptron (MLP), and Support Vector Machine (SVM)) applied to DINOv2 embeddings. While the use of kNN with dense ViT representations have been explored, techniques that use more complex non-linear classifiers on patch embeddings for independent patch-level classification remain, to our knowledge, underexplored. By treating each patch as an independent instance for classification, we enable patch-level predictions without requiring global image representation or modifications to the pretrained backbone. This approach provides fine-grained control with a minimal training overhead and achieves reliable detection with only 21 labeled images total per building element.

Methodology

The full pipeline, from street-view image annotation to building element inference, is described in the five steps below. Figure 1 visualizes steps 1 (annotation), 3 (patch embedding extraction), and 4 (classifier training); steps 2 and 5, adaptive cropping and inference on unseen images, are omitted from the figure but described in their respective subsections below.

Step 1: Annotation

Rather than drawing bounding boxes around entire elements, we annotate smaller bounding boxes *within* the elements of interest, providing precise control over the area to be extracted by the ViT, well-suited to its square patch structure. Polygon annotations were also explored in Slik (2025) but yielded lower detection performance,

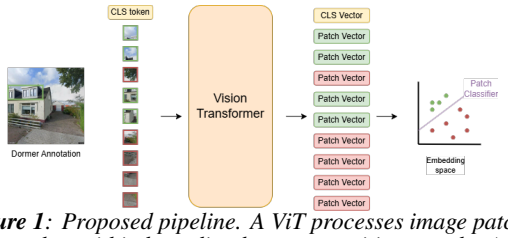


***Figure 1**: Proposed pipeline. A ViT processes image patches, with patches within bounding boxes as positive samples (green) and others as negatives (red), training a classifier per class.*

making axis-aligned bounding boxes the preferred choice. Areas without an element of interest are marked with "negative" bounding boxes. Two examples of this annotation style are shown in Figure 2. For this study we focused on seven building elements: solar panels, dormers, chimneys, roof windows, roof ventilation, removable panels, and balconies. Per element, 21 images with bounding box annotations are used for training, and 100 images carrying only image-level presence labels—50 positive and 50 negative—are used for testing. All 121 images per element depict distinct buildings, ensuring no building appears more than once across the training and test sets of a given element.



(a) Balcony Annotation (b) Solar Panel Annotation

***Figure 2**: Annotations to collect patch embeddings.*

Step 2: Adaptive Cropping

A ViT processes an image as a grid of patches, where each patch is a visual token. Bounding box annotations are often misaligned with this patch grid, reducing the number of usable positive patches. To address this, we apply adaptive cropping: the image is strategically cropped so the annotation aligns with the patch grid, enabling selection of significantly more relevant patches. Patches that partially overlap the bounding box along its right and bottom edges are also included, ensuring coverage when the annotation is smaller than the patch size. Figure 3 illustrates the worst-case scenario without alignment (left) versus the optimized configuration (right), increasing extracted patches from 4 to 12.

As shown in Table 1, the patch yield at least doubles for every element class: chimneys go from 0 to 57 patches, making classification possible where it would otherwise fail entirely.



(a) Without Alignment (b) With Adaptive Cropping

Figure 3*: Adaptive cropping impact on patch selection. Left: misaligned grid extracts 4 patches. Right: adaptive cropping aligns bounding box with patch grid, extracting 12 patches.*

Table 1*: Number of patch embedding collected from example images* ***without*** *and* ***with*** *adaptive cropping.*

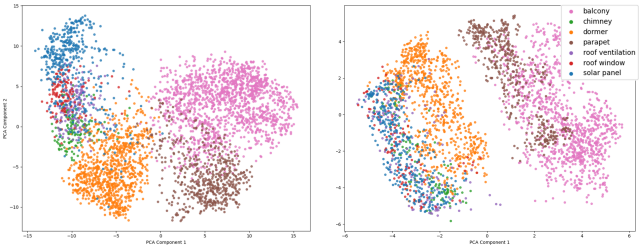
Class	**without**	**with**
Solar Panel	103	244
Dormer	234	539
Chimney	0	57
Roof Window	2	69
Roof Ventilation	8	75
Removable Panels	107	530
Balcony	368	867

Step 3: Patch Embedding Extraction

Images used have a resolution of 640×640 pixels before being passed to the ViT, which operates with a fixed patch size of 14×14 pixels, yielding a spatial grid of 45×45 patches per image. The ViT processes each cropped image and produces one embedding vector per patch. The choice of pretrained ViT significantly affects task behavior. We visualized the per-element patch embedding spaces of DINOv2 and DINOv3 models after PCA reduction to two dimensions (Figure 4). DINOv2 embeddings (Figure 4a) exhibit greater class separation, with distinct regions for different building components. In contrast, DINOv3 embeddings (Figure 4b) are more compressed and overlapping reflecting the model's patch-level spatial regularization, which prioritizes neighborhood consistency over discriminative separation. This trade-off, DINOv2 for element distinctiveness versus DINOv3 for spatial coherence, directly influences patch-level classification performance as evaluated in Step 4.

Step 4: Patch Classifier Training

For each element of interest, a binary one-vs-rest classifier is trained on the collected patch embeddings: positive examples are patches from that class's bounding boxes; negative examples are patches from negative bounding boxes and all other element classes. We compare four lightweight classifiers well suited to the high-dimensional ViT feature space: kNN as a distance-based baseline commonly used with ViT embeddings; and MLP, SVM with an RBF kernel, and RF as non-linear alternatives.



(a) DINOv2 (b) DINOv3

Figure 4*: First two principal components after PCA of patch embeddings obtained using different ViT models.*

Step 5: Inference

During inference, the ViT extracts patch embeddings from a new image, which the class-specific binary classifiers evaluate independently. For each classifier, if at least one patch is classified as positive, the element of interest is considered present in that image. This patch-level output can be visualized and provides inherent explainability, as this can be visually inspected to verify whether patches correctly cover regions of a target class.

Results and Discussion

This section evaluates our patch-level classification approach across seven building elements using four lightweight classifiers (kNN, MLP, SVM, RF) applied to the DINOv2 and DINOv3 embeddings. We aim to compare a kNN classifier, commonly used in ViT patch application tasks, with MLP, SVM, and RF classifiers that can capture more complex decision boundaries. We report Quantitative Results on class distribution balance, few-shot training set size, and per-class detection performance, then provide Qualitative Results on spatial prediction behavior.

Quantitative Results

Figure 5 and 6 reveal distinct performance patterns across the classifiers, backbones, and training configurations.

Class Distribution Balance (Figure 5): The redistribution ratio adjusts class balance in the patch classifier training set by undersampling the majority class to reach the target positive-to-negative ratio. DINOv2+RF achieved the highest performance at a redistribution ratio of 0.10, reaching 95.8% balanced accuracy. RF and SVM consistently performed better with DINOv2 than with DINOv3 across all distribution ratios. In contrast, kNN and MLP show a crossover pattern: DINOv2 outperforms DINOv3 for redistribution ratios below 0.05, but DINOv3 becomes superior above this threshold. kNN shows poor overall performance with substantial fluctuation, demonstrating high sensitivity to class distribution. MLP achieved 92.6% with DINOv2 (ratio 0.02) compared to 89.7% with DINOv3 (ratio 0.02), indicating that MLP and kNN classifiers benefit from different patch embedding structures.

Few-Shot Training Set Size (Figure 6): Performance trends diverge across classifier types as training examples grow from one to 21 images. RF and SVM

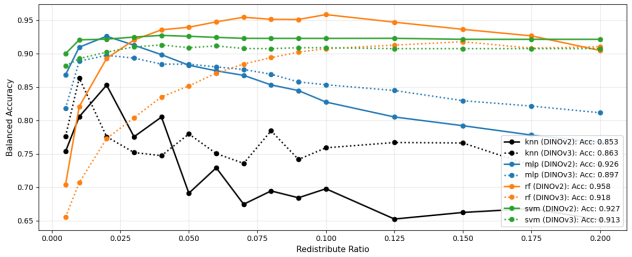


***Figure 5**: Performance scores **per redistribution ratio**, for different patch classifiers and dinov2 and dinov3 backbones.*

showed gradual improvement with additional training data: DINOv2+RF achieved the highest performance at 95.8% with 21 examples, substantially outperforming DINOv3+RF (91.2%)—a 4.7 percentage point advantage. Similarly, DINOv2+SVM with 21 examples outperformed DINOv3+SVM (+1.8 percentage points). In contrast, kNN and MLP exhibited more unexpected behavior – peaking at 3-5 examples before declining – seeming to be more sensitive to the chosen training samples than RF and SVM.

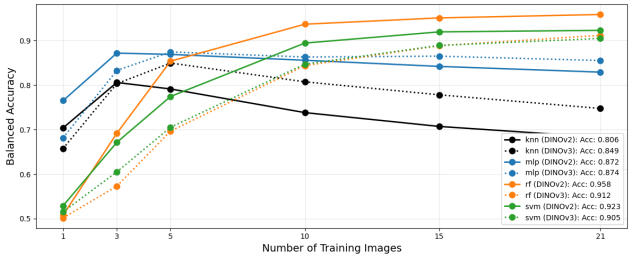


***Figure 6**: Performance scores **per number of example images**, for different patch classifiers and dinov2 and dinov3 backbones.*

These results reveal a trade-off between DINOv2 and DINOv3 embeddings. DINOv3 demonstrates gains exclusively with classifiers kNN and MLP at higher training sizes ($\geq$ 5 examples) and balanced distributions ($\geq$ 0.05 ratio), suggesting that its patch-level regularization primarily benefits kNN and MLP methods. However, with the sophisticated non-linear classifiers RF and SVM, DINOv2 consistently achieves superior performance across all conditions, with 4.6 and 1.8 percentage point advantages respectively at optimal settings. This indicates that while DINOv3 regularizes the embedding space for simple paradigms, DINOv2 retains greater representational capacity for complex decision boundaries, achieving the highest performance for patch-level building element detection. The results show that DINOv2 embeddings with RF offer the highest performance for patch-level building element detection.

Per-class Performance: **Table 2** presents metrics per-class using DINOv2+RF, 21 training images, and a redistribution ratio of 0.10. Roof-mounted elements (solar panels, dormers, chimneys, and roof windows) achieve exceptional accuracy (98–99%) with perfect precision, demonstrating that large, distinctive elements are reliably detected. Roof ventilation presents the most challenging case at 84.0% accuracy, because of its small size and high visual variability. Removable panels and balconies achieve strong performance (96% accuracy) with balanced precision-recall trade-offs. The method achieved 95.8% mean accuracy across all classes, with precision (96.8%) slightly exceeding recall (94.7%), indicating slightly conservative decision boundaries. Although the approach achieves strong performance for most element classes, it reveals limitations when elements exhibit high visual variability or appear at small scales, where patch embeddings struggle to capture sufficient discriminative information.

***Table 2**: Detection performance per facade element, mean over all elements, 100 samples test set.*

Class	Accuracy	Precision	Recall
Solar Panel	0.990	1.000	0.981
Dormer	0.991	1.000	0.981
Chimney	0.990	1.000	0.981
Roof Window	0.981	1.000	0.963
Roof Ventilation	0.840	0.889	0.784
Removable Panels	0.955	0.943	0.962
Balcony	0.961	0.943	0.980
Mean	**0.958**	**0.968**	**0.947**

Qualitative Results

Beyond quantitative metrics, the visual inspection of patch classification predictions provides interpretability. Figure 7 presents three examples illustrating both the strengths and failure modes of the approach.

Figure 7a shows a True Positive detection for roof windows: red patches adequately cover the regions containing the roof windows, demonstrating that the classifier correctly localizes the element across its spatial extent. Figure 7b shows a False Positive for removable panels, where a patch around the window frame is incorrectly highlighted. This is a plausible confusion: the removable panels in our dataset frequently occur in between window frames, and this positively highlighted patch is attached to a window frame too. Figure 7c illustrates a False Negative for roof ventilation: a small ventilation unit is visible on the roof but goes undetected. Given the fixed patch size of the ViT, small elements near patch boundaries are unlikely to be represented in a patch embedding, making detection for this small element in this configuration of pixel density and patch size less reliable. Overall, the approach correctly localizes elements of interest, though False Positives can arise from visual similarity to the target element class, and False Negatives from elements too small to be captured at the current patch resolution.

Case Study: Improving Data-Driven Retrofit Matching

The building element detection approach developed in this study enables practical applications. This section demonstrates how detected building elements from street-view imagery can improve a data-driven retrofit matching system, evaluates performance improvements across five retrofit datasets and discusses the broader implications for stakeholders in the building renovation sector.



(a) Roof Window TP (b) Removable Panel FP (c) Roof Ventilation FN

***Figure 7**: Qualitative results: Red patches indicate positive classifications.*

Added Value of Image Based Features

To evaluate the energy retrofit matching approach, we used multiple datasets from renovation advisors' assessments. Each dataset represents a renovation solution, with samples corresponding to individual dwellings and binary labels indicating applicability (per advisor judgment). Alongside these labels, we incorporated 66 features derived from open data — including construction year, building type, geometry, energy label, and occupant-related variables. Table 3 summarizes the datasets used. Removable panels are window-integrated facade elements which can be renovated by replacing them as a single unit with active ventilation.

***Table 3**: Energy retrofit datasets with label distributions (SFB: single-family building, MFB: multiple-family building).*

Dataset	Samples	Negative class	Positive class
Floor insulation SFB	7344	5217	2127
Cavity wall insulation SFB	6149	4940	1209
Roof insulation SFB	948	759	189
Removable panels SFB	255	189	66
Removable panels MFB	69	24	45

For each dataset, an RF classifier was trained and evaluated using grouped 5-fold cross-validation, where groups—defined by building complexes at the neighborhood scale—are kept within the same fold to prevent data leakage given the high uniformity within Dutch housing complexes. Cross-validation was repeated with different folds to yield 95% confidence intervals on balanced accuracy. The default 0.5 decision threshold was retained; threshold tuning confirmed this as a good trade-off between minimizing missed opportunities (low false negatives) and maximizing conversion rates (high precision), thereby avoiding costly false positives deferred to inspector review.

We then used our seven building element detectors to infer image-based features for all addresses in the datasets using Google Street View. Imagery was accessed via the Google Street View Static API under terms permitting non-commercial research use; coverage is extensive for the Dutch housing stock, and alternatives such as Cyclomedia exist should access or licensing constraints arise in other contexts. ViT inference takes < 50ms per image and the Street View Static API has a practical rate limit of 5 requests per second; the 14,765 addresses used in this study can be processed in under one hour. Querying the full Dutch housing stock (±5.5 million buildings) takes less than 15 days.

We evaluated three feature configurations: (i) 66 open-data features as a baseline ("Image features: no"), (ii) open-data features combined with seven Boolean element presence features derived from our detectors ("Image features: yes"), and (iii) open-data features combined with DINOv2's CLS token as a global image embedding, reduced to 20 dimensions via PCA ("Image features: CLS"). Results for all three configurations are reported in Table 4.

***Table 4**: Retrofit matching performance across three feature configurations: open data / no image features (baseline), open data + global image representation (CLS token), and open data + explicit building elements (seven Boolean features). Balanced accuracy ±95% CI.*

Dataset	Image features	Bal. Acc.	Gain
Floor insulation SFB	no	**62.3** ± 0.1	
	CLS	61.4 ± 0.1	
	yes	62.2 ± 0.1	
Cavity wall insulation SFB	no	63.1 ± 0.1	
	CLS	**63.5** ± 0.2	+0.4
	yes	63.1 ± 0.1	
Roof insulation SFB	no	58.6 ± 0.5	
	CLS	62.0 ± 0.5	
	yes	**63.4** ± 0.6	+4.8
Removable panels SFB	no	72.9 ± 0.6	
	CLS	77.6 ± 0.6	
	yes	**78.8** ± 0.5	+5.9
Removable panels MFB	no	55.8 ± 1.9	
	CLS	**67.4** ± 1.7	+11.6
	yes	56.6 ± 1.5	

The most substantial improvements occur for renovation solutions strongly correlated with visually detectable building elements. For instance, roof insulation shows a large improvement (+4.8 percentage points) when image-based Boolean features are added, reaching a balanced accuracy of 63.4%. This aligns with the contribution of roof-related element detection (e.g. dormers and roof windows). This improvement translates to gains in precision (55% to 60%) and recall (42% to 54%), reducing unnecessary advisor inspections of inapplicable housing by 5% while decreasing missed opportunities by 12%.

Removable panel detection yields particularly interesting results. For the "removable panels, single-family building" classifier, the image-based Boolean features achieve the best performance (78.8%, +5.9 percentage points), confirming that direct removable panel detection enhances the corresponding model. In contrast, the "removable panels, multi-family building" classifier exhibits a different pattern: CLS vector features lead to the largest improvement

(+11.6 percentage points), while the image-based Boolean features provide no significant gain. This result is based on a limited number of samples and should be confirmed using a larger dataset. Furthermore, it may be attributed to differences between single-family buildings and multi-family buildings, where removable panels are often more difficult to detect. Our results do illustrate that the CLS vector provides an indication of whether relevant visual information is present in a corresponding image. Targeted element detection then captures this global signal more explicitly, further boosting classifier performance.

Cavity wall and floor insulation show little to no improvement from image-based features. This is likely due to the small size of visual cues associated with these renovation measures. Presence of a cavity wall can be inferred from weep holes in the masonry. Presence of crawl space, which is necessary for floor insulation, can be inferred from small ventilation grilles at the bottom of facades. These details are not detectable at the resolution provided by the standard Google Street View API. However, sophisticated query strategies could provide the necessary detail for detection, enabling future performance gains for these classifiers. Overall, these results validate the practical value of integrating building-element detection into data-driven energy retrofit matching.

Impact of Data-driven Energy Retrofit Matching

When classifiers perform sufficiently well, this approach benefits multiple stakeholders. For renovation contractors offering standardized solutions, macro-level analyses of classifier predictions at regional or national scale can support market sizing, production planning, investment decisions, and industrialization, while micro-level, address-based information can help them reach suitable dwellings without extensive customization. Municipalities can use these tools to more effectively target homeowners of specific dwelling groups, increasing conversion rates for energy renovation. Housing associations can use these models to manage their stock and plan retrofit programs more efficiently.

Conclusion and Outlook

In this paper, we presented a scalable few-shot methodology for building element detection from street-view imagery through Vision Transformer patch classification. This work was based on the Master's thesis work of Slik (2025). The source code and labeled image datasets used in this study are available from the authors upon request.

From the ViT backbones evaluated, DINOv2 combined with RF or SVM yielded substantial performance improvements, with RF achieving 95.8% balanced accuracy compared to 91.8% with DINOv3. The adaptive cropping strategy increases positive patch samples by up to 10×. Overall, the methodology demonstrates potential in using ViT embeddings for domain-specific detection, with practical applications in energy retrofit matching, where it improves prediction accuracy by 4.8 percentage points for roof insulation applicability, effectively addressing data gaps in building stock databases.

Future work could improve performance through higher-resolution imagery or multi-scale detection strategies with shifted patch grids. Concretely, zoomed Street View API queries targeting facade details, super-resolution preprocessing, or tiled inference at multiple scales offer direct paths to detecting fine building details such as cavity wall weep holes, crawl space ventilation grilles, and roof exhaust units that remain below the current patch resolution. Moreover, rather than using a single frozen ViT configuration for all tasks, different backbones and classifiers could be selected by evaluating performance in specific scenarios, enabling task-specific optimization for address-level element detection within a comprehensive pipeline. Furthermore, researching building outlining techniques using cadastral data and segmentation models could strengthen confidence that detected elements belong to the target address. This methodology, including its extensions, readily scales to other national housing stocks—particularly those with limited open data—and to non-energy domains requiring building element inference.

References

Bjelland, D., Brozovsky, J., and Hrynyszyn, B. D. (2024). Systematic review: Upscaling energy retrofitting to the multi-building level. Renewable and Sustainable Energy Reviews, 198:114402.

Bulat, A., Guerrero, R., Martinez, B., and Tzimiropoulos, G. (2023). Fs-detr: Few-shot detection transformer with prompting and without re-training. In Proceedings of the IEEE/CVF international conference on computer vision, pages 11793–11802.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., and Joulin, A. (2021). Emerging properties in self-supervised vision transformers. In Proceedings of the IEEE/CVF international conference on computer vision, pages 9650–9660.

Chen, D., Li, X., Hu, F., Mathiopoulos, P. T., Di, S., Sui, M., and Peethambaran, J. (2023). Edpnet: An encoding–decoding network with pyramidal representation for semantic image segmentation. Sensors, 23(6):3205.

Coffee, S., O'Donnell, J., and Hoare, C. (2025). Comparative machine learning and deep learning study of energy predictions in urban and rural buildings. In Proceedings of the 2025 European Conference on Computing in Construction, volume 6 of Computing in Construction, Porto, Portugal. European Council on Computing in Construction.

Darcet, T., Oquab, M., Mairal, J., and Bojanowski, P. (2023). Vision transformers need registers. arXiv preprint arXiv:2309.16588.

Geyer, P., Schlüter, A., and Cisar, S. (2017). Application of clustering for the development of retrofit strategies for large building stocks. Advanced Engineering Informatics, 31:32–47. Towards a new generation of the smart built environment.

Hamilton, M., Zhang, Z., Hariharan, B., Snavely, N., and Freeman, W. T. (2022). Unsupervised semantic segmentation by distilling feature correspondences. arXiv preprint arXiv:2203.08414.

Hao, F., He, F., Liu, L., Wu, F., Tao, D., and Cheng, J. (2023). Class-aware patch embedding adaptation for few-shot image classification. In Proceedings of the IEEE/CVF international conference on computer vision, pages 18905–18915.

Jiang, W., Cui, H., and He, K. (2024). Class-relevant patch embedding selection for few-shot image classification. arXiv preprint arXiv:2405.03722.

Korc, F. and Förstner, W. (2009). etrims image database for interpreting images of man-made scenes. Dept. of Photogrammetry, University of Bonn, Tech. Rep. TR-IGG-P-2009-01.

Li, Y., Wang, T., Kang, B., Tang, S., Wang, C., Li, J., and Feng, J. (2020). Overcoming classifier imbalance for long-tail object detection with balanced group softmax. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 10991–11000.

Lin, C., Yuan, Z., Zhao, S., Sun, P., Wang, C., and Cai, J. (2021). Domain-invariant disentangled network for generalizable object detection. In Proceedings of the IEEE/CVF international conference on computer vision, pages 8771–8780.

Liu, H., Xu, Y., Zhang, J., Zhu, J., Li, Y., and Hoi, S. C. (2020). Deepfacade: A deep learning approach to facade parsing with symmetric loss. IEEE Transactions on Multimedia, 22(12):3153–3165.

Liu, H., Zhang, J., Zhu, J., and Hoi, S. C. H. (2017). Deepfacade: A deep learning approach to facade parsing. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17), pages 2301–2307.

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., et al. (2024). Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In European conference on computer vision, pages 38–55. Springer.

Marasco, D. E. and Kontokosta, C. E. (2016). Applications of machine learning methods to identifying and predicting building retrofit opportunities. Energy and Buildings, 128:431–441.

Mata, , Sasic Kalagasidis, A., and Johnsson, F. (2014). Building-stock aggregation through archetype buildings: France, germany, spain and the uk. Building and Environment, 81:270–282.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. (2023). Dinov2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193.

Pariza, V., Salehi, M., Burghouts, G., Locatello, F., and Asano, Y. M. (2024). Near, far: Patch-ordering enhances vision foundation models' scene understanding. arXiv preprint arXiv:2408.11054.

Rivallain, M., Agapoff, S., Boisson, P., Foucquier, A., Lee, Y., and la Vallée, M. (2020). Clustering as a simplification tool for the decision-making process on building stock renovation.

Shen, P. and Wang, H. (2024). Archetype building energy modeling approaches and applications: A review. Renewable and Sustainable Energy Reviews, 199:114478.

Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al. (2025). Dinov3. arXiv preprint arXiv:2508.10104.

Slik, T. (2025). Few-shot building element detection via self-supervised vision transformer patch classification. Master's thesis, University of Amsterdam.

Tyleček, R. and Šára, R. (2013). Spatial pattern templates for recognition of objects with regular structure. In German conference on pattern recognition, pages 364–374. Springer.

Uihlein, A. and Eder, P. (2010). Policy options towards an energy efficient residential building stock in the eu-27. Energy and Buildings, 42(6):791–798.

van Groesen, W. (2022). Industrialized Renovations: Towards Mass-Customization in the Construction Industry. Engd thesis, Technische Universiteit Eindhoven. EngD thesis.

Zhang, X., Liu, Y., Wang, Y., and Boularias, A. (2023). Detect everything with few examples. arXiv preprint arXiv:2309.12969.

Zhang, X., Xu, Z., Xu, R., Liu, J., Cui, P., Wan, W., Sun, C., and Li, C. (2022). Towards domain generalization in object detection. arXiv preprint arXiv:2203.14387.

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., and Torralba, A. (2018). Semantic understanding of scenes through the ade20k dataset.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AI-ASSISTED BRIDGE CONCEPTUAL DESIGN: A HYBRID KNOWLEDGE- AND LEARNING-BASED APPROACH TO AUTOMATED SPAN PLANNING

Han Qian[1], Tim Noack[1], Mengyan Peng[1], Marx Steffen[1] and Chongjie Kang[1]
[1]Technische Universität Dresden, Dresden, Germany

Abstract

Bridge conceptual design is challenged by heterogeneous boundary conditions, strict engineering constraints, and large solution spaces. This paper presents an AI-assisted pipeline for automated bridge conceptual design, focusing on early-stage span layout planning. The proposed approach integrates ontology-based semantic knowledge representation with reinforcement learning to enable automated decision-making while preserving engineering consistency. By embedding engineering constraints into the learning environment, the method supports efficient exploration of feasible span configurations. The results indicate that the proposed pipeline provides a robust and extensible foundation for AI-supported bridge planning in realistic engineering scenarios.

Introduction

Bridge conceptual design is a pivotal early-stage task in which key decisions such as span layout and structural system selection are made under heterogeneous boundary conditions, such as site conditions of terrain and crossed obstacles, clearance requirements, environmental restrictions, and project-specific constraints (Troitsky, 2019). While decisions at this stage strongly influence feasibility, cost, and downstream design flexibility, the workflow is still dominated by manual iterations and experience-driven heuristics. In today's digitalized and automation-oriented construction industry, this creates two persistent challenges: (i) early-stage information is often incomplete, symbolic, and distributed across different sources, and (ii) the design space is combinatorial, yet must remain consistent with engineering constraints.

Artificial Intelligence (AI) is increasingly adopted across civil engineering and construction to improve efficiency, consistency, and scalability of decision-making, which can be a viable path to address the fragmented nature of early-stage design information. Machine Learning (ML) has been used to support structural engineering beyond inspection, including design-related prediction and rapid evaluation. Supervised Learning (SL) and Deep Learning (DL) models are often adopted as surrogate predictors to approximate the mapping between design boundary conditions and design variables (Qian et al., 2024), or between design variables and structural responses (Zheng et al., 2020; Balmer et al., 2024). These ML-based methods enable efficient screening of conceptual alternatives and reduce repeated high-cost simulations. Recent work also explores transfer learning to improve robustness of learned predictors under limited data, which is common in civil engineering applications (Zhang et al., 2025). Beyond prediction, Reinforcement Learning (RL) is being investigated for automating design decisions that can be formulated as sequential optimization, e.g., learning to automatically design reinforced concrete structures (Jeong et al., 2021), and handling construction design coordination tasks such as automated reinforcement clash resolution in Building Information Modeling (BIM) (Liu et al., 2020). Nevertheless, purely ML-based approaches (SL/DL or RL) often lack explicit mechanisms to guarantee strict feasibility under hard engineering constraints and may remain difficult to interpret and justify in safety-critical conceptual design.

To address these limitations, knowledge-based and rule-based systems have been developed to preserve engineering consistency and enable transparent reasoning. Ontology-driven approaches can formalize domain entities, constraints, and relationships, thereby supporting deterministic inference and validation. Ontology- and rule-based systems have been used to encode engineering knowledge in an explicit, machine-interpretable form, enabling early constraint checking and reducing iterative design cycles (Cao et al., 2022). In bridge engineering, most ontology efforts have focused on rehabilitation and maintenance, e.g., knowledge bases for maintenance decision support, while comparatively few studies address ontology integration in bridge design workflows (Ren et al., 2019; Zhang et al., 2022; Aqlan et al., 2025). Recent work also emphasizes that systematic rule structuring and ontology-based knowledge organization remain important for transparency and traceable rule management in digital construction (Wen et al., 2024). However, knowledge-based systems alone are often conservative in practice. While they are strong at enforcing hard constraints and providing explainable logic, they typically lack flexible optimization capabilities when multiple soft objectives (e.g., economy, trade-offs between competing objectives) must be balanced under diverse boundary conditions.

These complementary strengths and weaknesses motivate hybrid AI strategies that combine knowledge-based constraint enforcement with learning-based exploration.

Accordingly, this paper proposes a hybrid AI-assisted framework for bridge conceptual design, with a focus on automated span planning. In contrast to purely constraint-guided RL and existing bridge layout automation methods, it uses ontology-based reasoning to translate boundary conditions into a structured feasibility space and employs RL within this constrained environment to explore and optimize span configurations under soft engineering preferences. By integrating explicit knowledge with adaptive decision-making, the framework aims to improve automation while maintaining engineering consistency, transparency, and extensibility for realistic bridge planning scenarios.

Methods

AI-assisted pipeline for bridge conceptual design

In bridge conceptual design, the involved tasks differ substantially in their data availability, constraint structure, and decision complexity, making the use of a single AI technique across all stages neither efficient nor reliable. This work therefore adopts a task-oriented AI-assisted pipeline in which different design subtasks are supported by AI methods that match their intrinsic characteristics.

As shown in Figure 1, the proposed pipeline decomposes bridge conceptual design into three main stages: (i) interpretation and structuring of boundary conditions, (ii) planning of bridge spans, and (iii) planning of superstructure cross-sections. Boundary conditions originate from diverse sources, including site conditions, regulatory requirements, and project-specific constraints. These inputs are predominantly semantic and rule-driven rather than data-intensive, and are therefore represented using ontology-based knowledge models. Ontologies enable consistent semantic interpretation of boundary conditions, support rule-based reasoning, and provide structured inputs for subsequent design steps. Bridge span planning constitutes the central layout decision within the pipeline. It involves selecting the number of spans, their lengths, and pier locations under complex geometric and functional constraints. This task exhibits a large combinatorial search space and strong interdependencies between design variables, making it unsuitable for purely rule-based enumeration. At the same time, strict engineering constraints must be satisfied to ensure feasibility. For this reason, span planning is addressed using a hybrid approach that combines symbolic design rules with RL, allowing automated exploration of feasible layouts while preserving engineering validity. Once a span layout is determined, the pipeline proceeds to the planning of superstructure cross-sections. In contrast to span planning, this task benefits from the availability of existing bridge data and relatively stable input-output relationships between boundary conditions and structural dimensions. As a result, supervised learning techniques are employed to predict key geometric parameters of superstructure components, while rule-based methods are applied where normative prescriptions are explicit and deterministic.

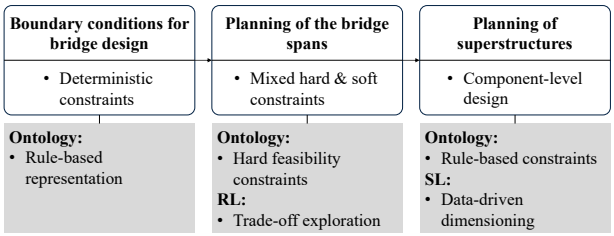


Figure 1: AI-assisted pipeline for bridge conceptual design

Hybrid methodology for bridge span planning

Within the AI-assisted conceptual design pipeline, bridge span planning represents the most critical decision-making task. The selection of span numbers, span lengths, and pier locations determines not only the overall structural layout but also construction feasibility, cost efficiency, and the quality of downstream design tasks.

Purely rule-based approaches have traditionally been used in early-stage bridge layout planning to ensure compliance with standards and engineering best practices (Troitsky, 2019). While such approaches provide strong guarantees of feasibility, they tend to be conservative and lack flexibility when confronted with complex or non-standard boundary conditions. Conversely, learning-based optimization techniques, such as RL, offer the ability to explore large design spaces and identify efficient solutions, but cannot inherently guarantee adherence to engineering rules and safety constraints. In bridge span planning, violations of such constraints are unacceptable, even at the conceptual design stage.

To address these complementary limitations, the proposed hybrid methodology combines rule-based reasoning with RL in a tightly coupled hybrid framework, as illustrated in Figure 2. The first step is to simplify the complex side conditions from three-dimensional (3D) to 2D topography. Explicit design rules and boundary conditions are then formalized using an ontology-based representation and used to encode domain knowledge and constraint logic, defining permissible, restricted, and forbidden regions for pier placement. RL is employed to perform sequential decision-making within this feasibility space defined by the ontology and further encoded design rules, because local support adjustments influence subsequent feasible actions and the overall quality of the final span layout. It enables automated exploration and optimization while maintaining engineering validity. Therefore, rule-based reasoning and RL fulfill distinct but interdependent roles. Semantic rules in Category A (see Table 1) act as constraints that shape the environment in which exploration takes place, reducing the effective search space and preventing infeasible solutions. RL, in turn, provides adaptability by optimizing layout decisions under these constraints and resolving trade-offs between competing objectives based on the rules in Category B (see Table 1), such as minimizing pier numbers while respecting economic span ranges. Rather than encoding all constraints only as reward penalties, the proposed framework externalizes hard, boundary-condition-dependent feasibility constraints through ontology-based

reasoning and uses RL to optimize softer layout preferences within the admissible design space.

The hybrid methodology is structured to support modular extension and transparent interpretation. The following sections detail the categorization of design rules, their ontology-based representation, and the RL model used to realize this hybrid approach.

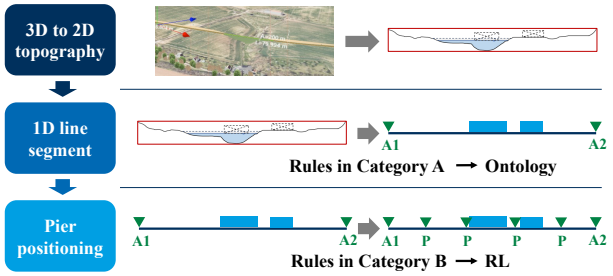


Figure 2: Hybrid methodology for bridge span planning

Design rules for bridge span planning

Design rules form the backbone of engineering knowledge in bridge span planning and play a central role in constraining and guiding the hybrid methodology. In the context of conceptual design, they are particularly important as detailed structural verification is not yet available, making explicit knowledge-based constraints essential for ensuring feasibility. For this work, span planning rules were systematically collected from multiple sources, such as German design standards, project-specific requirements, and engineering practice. The collected rules were then analyzed and structured with respect to their functions, rather than their original textual form. This functional analysis enables a clear distinction between rules that directly restrict geometric feasibility and those that provide qualitative or economic guidance for span layout decisions.

Table 1: Economical span ranges for different bridge types

Category A	Topography-dependent rules
Rule sources	• German design standards (e.g., DIN, RiZ-ING) • site-specific planning documents • engineering practice
Rule characteristics	• Hard constraints • Deterministic • Directly linked to terrain and obstacles
Typical rules	• Required clearance heights and widths for waterways and roads • Restrictions on pier placement in water bodies • Restrictions on pier placement in built-up areas and private property
Category B	Topography-independent rules
Rule sources	• Engineering handbooks • Guidelines • Expert knowledge
Rule characteristics	• Soft constraints • Heuristic and preference-based • Potentially conflicting
Typical rules	• Symmetry requirements of pier layout • Main span ≥ side spans • Trade-off between span length and number of piers • Economical span ranges for different bridge types

Based on this analysis, the design rules are categorized into two main groups according to their dependency on boundary conditions, which is demonstrated in Table 1. Category A rules are strongly dependent on topography and crossed obstacles, and define spatial constraints on span planning, such as clearance requirements over waterways and roads, and geometric limitations imposed by terrain conditions. These rules primarily determine whether a pier location is permissible, restricted, or forbidden along the bridge axis. Conversely, Category B rules are largely independent of local topography and describe general engineering principles for bridge span layout, e.g., symmetry requirements, and economical span length ranges for different bridge types. Unlike Category A rules, these rules do not impose absolute feasibility constraints but influence the quality and efficiency of candidate span layouts.

The separation of design rules into Category A and Category B supports their differentiated use within the hybrid span planning framework. Category A rules are primarily used to define hard constraints that shape the feasible design space, while Category B rules provide soft constraints and optimization guidance during layout generation. This structured rule system forms the basis for both ontology-based reasoning and RL, enabling explicit integration of engineering knowledge into automated bridge span planning.

Ontology-based rule representation and reasoning

To enable systematic and machine-interpretable application of topography-dependent design rules, an ontology-based representation is adopted for bridge span planning. Ontologies provide a formal and extensible framework for structuring heterogeneous engineering knowledge and are particularly suitable for representing explicit boundary conditions and hard feasibility constraints. In the proposed methodology, ontology-based reasoning is deliberately applied to the Category A rules.

The ontology used in this work is called Bridge Core Ontology (BCore), an extension of the Bridge Topology Ontology (BROT) proposed by Hamdan and Scherer (2020) which is again based on the Building Topology Ontology (BOT) described by Rasmussen et al. (2020). BCore was created with the ontology editor Protégé in the Web Ontology Language (OWL 2). BCore has multiple purposes in the field of bridge engineering, amongst them the semantic description of bridge boundary conditions. The Category A rules are implemented using the Semantic Web Rule Language (SWRL). Based on the bridge boundary conditions, they determine whether it is possible, inadvisable, or not possible to place a support (pier or abutment) at a specific point of the bridge axis.

The outputs of ontology-based reasoning consist of structured feasibility representations that define the admissible design space for bridge span planning. These representations serve as validated inputs to the RL environment, where Category B rules are subsequently integrated as optimization guidance. In this way, the ontology acts as a semantic filter that enforces hard boundary-condition-driven constraints, while leaving efficiency-oriented layout principles in Category B to be handled by learning-based decision making.

Learning-based decision making using RL

Based on the ontology-derived feasibility space, RL is used to generate bridge span layouts under complex boundary conditions. While ontology-based reasoning enforces hard constraints via Category A rules, Category B rules are embedded in the learning process to guide optimization toward efficient and balanced configurations. Operating within this rule-constrained environment, RL resolves trade-offs that are difficult to encode deterministically, enabling flexible yet engineering-consistent span planning with interpretable decision logic at the conceptual stage.

Problem formulation

Bridge span planning is formulated as a sequential decision-making problem in a custom Gymnasium (Towers et al., 2024) environment and modeled as a Markov Decision Process (MDP) (Garcia et al., 2013), defined by the tuple

$$M = \langle S, A, P, R, \gamma \rangle \quad (1)$$

where S denotes the state space, A the action space, P the state transition dynamics, R the reward function, and $\gamma \in (0, 1)$ the discount factor. The problem formulation in RL is illustrated in Figure 3.

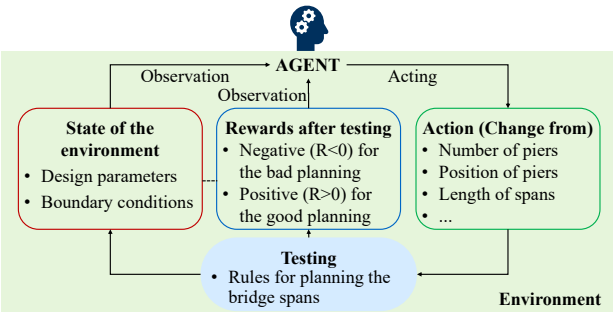


Figure 3: RL framework for bridge span planning

At each step, the learning agent interacts with an environment that represents the current bridge layout under given boundary conditions. The environment state $s \in S$ includes structured information derived from ontology-based reasoning, such as total bridge length L, admissible regions and forbidden zones that encode hard feasibility constraints (Category A rules). Additional state variables describe the current number of piers n, span lengths l_i, and pier positions x_i along the bridge axis. All quantities are normalized with respect to the total bridge length to ensure numerical stability during learning. The action space A comprises discrete layout edits, primarily moving piers forward or backward by a fixed step, and adding or removing piers within admissible regions. At each step, the environment executes the selected action, updates the state, computes a composite reward, and checks termination. State transitions P are deterministic given the selected action and current layout configuration. As shown in Figure 3, by observing a current state representing the bridge layout and boundary conditions, the agent is guided by current learned policy and selects an action that modifies pier positions or span layout and aims to maximize the cumulative reward. Then a testing procedure is applied in the environment, which evaluates the action against predefined rules and feasibility constraints. Valid actions lead to deterministic updates of the bridge layout and derived state variables, resulting in the next state. Actions violating hard feasibility constraints are rejected or penalized, so that infeasible states are not generated during learning.

Integration of design rules into the learning process

Category B rules encode soft principles and are incorporated into the RL framework as optimization guidance. Due to their qualitative nature and potential conflicts, these rules are not deterministic but are reflected in the reward formulation.

The reward function is defined as a weighted aggregation of multiple rule-related components:

$$R(s, a) = \sum_{i=1}^{N} \omega_i R_i(s, a) \quad (2)$$

where $R_i(s, a)$ denotes an individual reward component under specific state s after action a associated with a specific layout principle, and ω_i is the corresponding weighting factor. This formulation allows different engineering preferences to jointly influence the optimization process while preserving flexibility in resolving trade-offs between competing objectives.

In the current implementation, these reward components capture span validity, symmetry, uniformity, and span proportion preferences derived from Category B rules. The reward shaping provides continuous feedback that guides the learning agent toward high-quality span layouts. The reward function is designed to penalize layouts while rewarding solutions that achieve feasible span distributions. In the following, one representative reward, the component symmetry reward R_{sym}, is presented to illustrate how the Category B rule is translated into quantitative learning signals.

Symmetry of pier placement is evaluated based on the spatial distribution of pier positions along the bridge axis. For layouts with an odd number of piers, an additional requirement is imposed on the central pier. Let x_c denote the position of the middle pier, then its deviation from the bridge center is measured as:

$$e_{\text{center}} = |x_c - L/2| \quad (3)$$

For all layouts, symmetry of pier pairs is evaluated by comparing the sum of symmetric pier positions with the bridge length. For each mirrored pier pair (x_i, x_{n+1-i}), the pairwise symmetry error is defined as:

$$e_i = |(x_i + x_{n+1-i}) - L| \quad (4)$$

The overall symmetry error is computed as:

$$e_{\text{sym}} = \begin{cases} \frac{1}{\lfloor n/2 \rfloor}\left(\sum_{i=1}^{\lfloor n/2 \rfloor} e_i + e_{\text{center}}\right), & \text{if n is odd} \\ \frac{1}{\lfloor n/2 \rfloor}\sum_{i=1}^{\lfloor n/2 \rfloor} e_i, & \text{if n is even} \end{cases} \quad (5)$$

where $\lfloor n/2 \rfloor$ denotes the number of mirrored pier pairs. This formulation ensures that both mirrored pier alignment and, if applicable, the central pier position contribute to the symmetry test, while normalization prevents bias toward layouts with more piers.

Based on the computed symmetry error e_{sym}, the symmetry reward is defined as:

$$R_{\text{sym}}(e_{\text{sym}}) = \left(1 - \frac{e_{\text{sym}}}{T_{\text{sym}}}\right) R_{\text{sym,max}} \quad (6)$$

where T_{sym} represents symmetry threshold defining "acceptable" asymmetry and $R_{\text{sym,max}}$ denotes maximum achievable symmetry bonus magnitude. This reward provides a graded positive bonus when the layout is sufficiently symmetric (error below threshold) and a proportional penalty once the symmetry error exceeds the threshold, thereby encoding symmetry as a soft optimization objective rather than a hard constraint.

Beyond reward shaping, Category B rules are also used to define episode termination. An episode ends when a feasible span layout is achieved (success) or when the maximum step limit is reached. In the implementation, a solution is marked as DONE when all spans fall within the prescribed range and either (i) the layout meets a uniformity threshold, or (ii) it satisfies a symmetry threshold with the center span being the largest. This criterion reflects, that acceptable conceptual solutions may arise through different quality pathways rather than a single rigid condition.

Deep Q-Network (DQN) architecture and training

A DQN framework in RL (Mnik et al., 2015) is adopted to approximate the optimal action-value function $Q^*(s,a)$ for bridge span planning. Instead of explicitly modeling transition probabilities, learning is performed based on sampled transitions generated through interaction with the environment. Accordingly, the Bellman optimality principle is expressed in its sample-based form:

$$Q^*(s,a) = \mathbb{E}[r_t + \gamma \max Q^*(s_{t+1}, a') \,|s_t, a_t] \quad (7)$$

where s_t denotes the current state representing the bridge layout and boundary conditions at decision step *t*, while s_{t+1} is the subsequent state, a_t is the selected action modifying the current layout, r_t is the immediate reward obtained after applying action a_t, a' denotes an admissible action in the next state, and γ is the discount factor weighting future rewards.

In practice, the expectation is approximated using an experience replay buffer. During training, a separate target network with parameters θ is maintained to compute the temporal-difference target Q-values:

$$y_t = r_t + \gamma \max Q(s_{t+1}, a'; \theta) \quad (8)$$

where $Q(\cdot;\theta)$ denotes the target network used for stable target estimation. The main network parameters are updated by minimizing the squared temporal-difference error between the predicted Q-value and the target y_t. Then the target network parameters are periodically synchronized with the main network. This delayed update mechanism improves convergence robustness.

The network maps the environment state to Q-values for all admissible actions and is implemented as a multilayer perceptron (MLP) with two fully connected hidden layers and ReLU activations. Action selection follows an epsilon-greedy policy with probability ε, a random admissible action is chosen for exploration, otherwise, the action with the highest estimated Q-value is selected. The exploration rate ε is gradually decayed over training episodes. Training is conducted across diverse boundary conditions to promote generalization and to reliably produce feasible and efficient span layouts.

Results and case study

Results of ontology-based rule representation and reasoning

The ontology-based knowledge representation provides a structured and interpretable foundation for bridge span planning under heterogeneous boundary conditions. Design rules related to topography, crossed obstacles, clearance requirements, and pier feasibility are formalized as semantic entities and relations, enabling explicit reasoning over permissible (support-possible), restricted (support-not-recommended), and forbidden (support-not-possible) regions along the bridge axis.

The important classes of BCore are displayed in Figure 4. Every *Bridge carriesRoute InfrastructureRoute* which is composed of *RoutePoint*s. These points, being a *rdfs:subClassOf LocationPoint*, can be *locatedInZone. Zone*s can be, amongst others, a *Forest, WaterBody, Estate, Building* or another *Bridge*. Important *DataProperties* (attributes) in BCore are the coordinates of *RoutePoint* (*LocalX*, *LocalY*, *LocalZ*, *Station*; all of *Datatype xsd:decimal*) and its *xsd:boolean*-type properties *SupportNotPossible* and *SupportNotRecommended*. An *InfrastructureRoute* has the *DataProperties InfrastructureName* (*xsd:string*) and *InfrastructureWidth* (*xsd:decimal*).

With these semantics, the bridge's boundary conditions can be represented, as well as the Category A rules. Two examples of these rules are shown and explained in Table 2 and Table 3. They serve to ascertain that a pier or other bridge support cannot be placed on another *InfrastructureRoute*. All rules are implemented in SWRL and are thus hard *if-then*-rules. Other Category A rules include different types of *Zone*s, where a *RoutePoint* is located in. Depending on the type of *Zone*, either a support is not possible or not recommended at the respective *RoutePoint*. For example, in a *WaterBody*, a support is not recommended, while in a *Building*, it is not possible.

After creating instances of the *Bridge*, the carried route(s), surrounding *Zone*s and crossing routes, the bridge axis (= carried route) is divided into a high number of sections

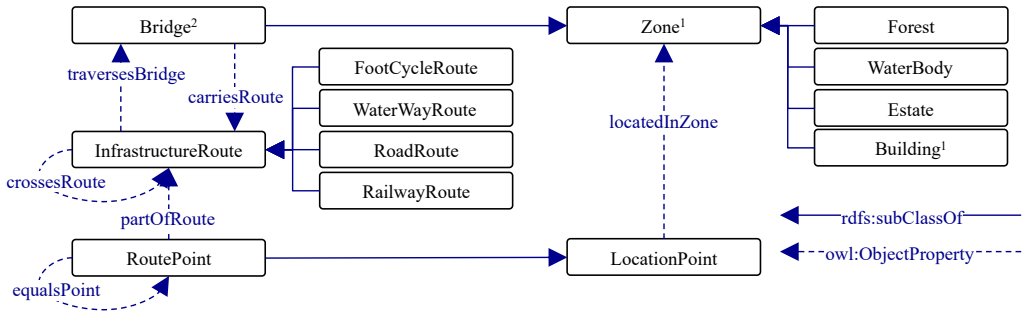


Figure 4: Diagram of BCore parts adapted from Rasmussen et al. (2020) for Zone [1] and Hamdan and Scherer (2020) for Bridge [2]

of the same length. A section length between 10 cm and 1 m is recommended. At the end of each section, a *RoutePoint* is created with its respective x-, y- and z-coordinates and its *Station* (kilometrage). In the proximity of the bridge, the other routes need their name, width, and *RoutePoint*s with their coordinates. Each *RoutePoint* that is located in a *Zone* needs to be connected to that *Zone* via the *locatedInZone*-property. After creating all these instances, a SWRL-capable reasoner (e.g., *Pellet*) needs to be run. For every point inside a zone or on a crossing route, the values of properties *SupportNotPossible* and *SupportNotRecommended* are inferred.

The points and their respective support-conditions can be exported or queried from the ontology.

Table 2: Rule route crossing point: description and SWRL-implementation

If two points from different routes have the same x- and y-coordinates, they are "equal" and their routes cross.
`InfrastructureRoute(?r1) ^InfrastructureRoute(?r2) ^InfrastructureName(?r1,?n1) ^swrlb:notEqual(?n1,?n2) ^InfrastructureName(?r2,?n2) ^RoutePoint(?rp1) ^partOfRoute(?rp1,?r1) ^RoutePoint(?rp2) ^partOfRoute(?rp2,?r2) ^LocalX(?rp1,?x) ^LocalY(?rp1,?y) ^LocalX(?rp2,?x) ^LocalY(?rp2,?y)`
`-> equalsRoutePoint(?rp1,?rp2) ^crossesRoute(?r1, ?r2)`

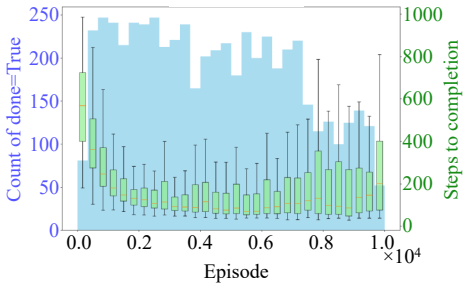
Table 3: Rule route crossing support: description and SWRL-implementation

If a route point is located in the area of another route, a support is not possible at this point.
`InfrastructureRoute(?r1) ^InfrastructureRoute(?r2) ^crossesRoute(?r1,?r2) ^RoutePoint(?rp1) ^partOfRoute(?rp1,?r1) ^RoutePoint(?rp2) ^partOfRoute(?rp2, ?r2) ^equalsRoutePoint(?rp1,?rp2) ^RoutePoint(?rp3) ^partOfRoute(?rp3,?r1) ^Station(?rp1,?s1) ^Station(?rp3,?s3) ^swrlb:subtract(?diff,?s1,?s3) ^swrlb:abs(?abs,?diff) ^InfrastructureWidth(?r2,?b) ^swrlb:divide(?bhalf,?b,2) ^swrlb:greaterThanOrEqual(?bhalf,?abs)`
`-> SupportNotPossible(?rp3,true)`

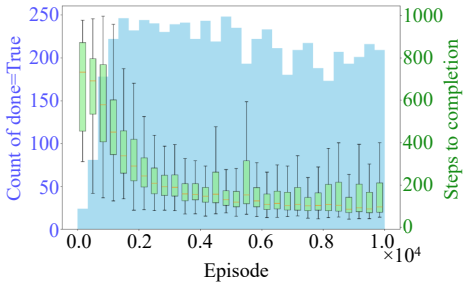
RL training and hyperparameter analysis

To achieve stable and efficient learning behavior in bridge span planning, a systematic hyperparameter tuning process was conducted. A broad range of parameters was explored, including training-related parameters (e.g., number of episodes and maximum steps per episode), DQN architecture parameters (e.g., network depth, neuron numbers, and replay memory size), and environment parameters (e.g., forbidden zone configurations and action step lengths), and reward-related parameters reflecting engineering objectives.

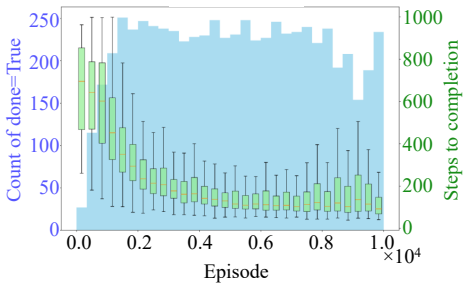
Special attention in this paper was given to parameters governing the exploration-exploitation balance, namely the epsilon target percent (etp) and epsilon decay percent (edp). These parameters define the decay schedule of the ε-greedy exploration strategy, where ε decays to etp within the first edp portion of training episodes. To evaluate training performance, two complementary metrics were adopted: (i) count of done, representing the number of episodes that successfully reached a valid span configuration within fixed episode intervals, and (ii) steps to completion, indicating the number of action steps required to reach termination in each episode. Together, these metrics provide insight into both learning effectiveness and planning efficiency.



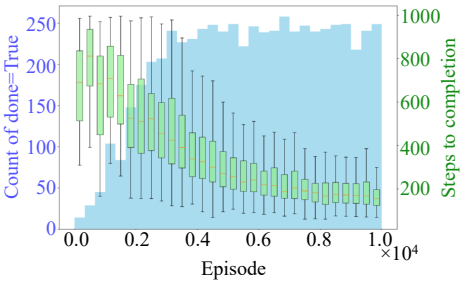
(a) etp = 40 %, edp = 30 %



(b) etp = 40 %, edp = 80 %



(c) etp = 70 %, edp = 30 %



(d) etp = 70 %, edp = 80 %

Figure 5: Training evolution of DQN model with the examples under four representative parameter pairs of etp and edp

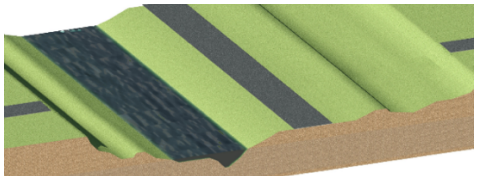
Figure 5 presents the training evolution with the examples under four representative parameter pairs of etp and edp, highlighting the interaction between exploration decay and learning stability. Apparently, for small values of (etp = 40%, edp = 30%) in Figure 5(a), the agent exhibits a rapid increase in successful episodes during early training stages. However, this improvement is not sustained, as the success rate decreases in later stages, and the variance of steps to completion increases. This pattern indicates early-stage overfitting, where the agent prematurely commits to suboptimal policies and loses robustness as training proceeds.

By comparing Figure 5(b) to (a) or Figure 5(d) to (c), increasing the decay period to edp = 80% significantly improves long-term stability. When comparing configurations with identical edp (see Figure 5(a) and (c), Figure 5(b) and (d)), a larger target exploration rate (etp = 70%) consistently yields more stable success rates and lower variance in step counts. These observations demonstrate that slower exploration decay and higher residual exploration are essential for avoiding local optima in bridge span planning tasks with complex constraints. Based on this comparative analysis, the parameter combination (etp = 70%, edp = 80%) was selected for subsequent experiments, as it provides the best balance between convergence speed, solution quality, and robustness.

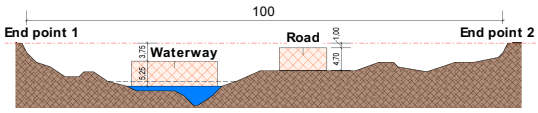
Case study

To demonstrate the practical applicability of the proposed pipeline, a simple case study was conducted, as shown in Figure 6. Starting from a simulated 3D terrain model in Autodesk Revit (Figure 6(a)), the terrain data were first transformed into a 2D longitudinal section (Figure 6(b)), from which ontology-based reasoning identified forbidden zones and feasible support regions considering terrain geometry and crossed obstacles (Figure 6(c)).

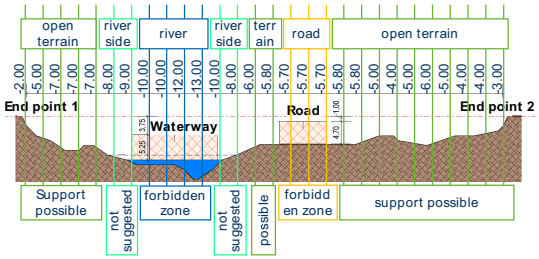
Based on the resulting 1D design space derived from ontology, the trained RL agent was then applied to this scenario to generate an optimized bridge span layout (green marks in Figure 6(d)). The resulting configuration satisfies all feasibility constraints while achieving a balanced span distribution consistent with engineering design principles in Category B. The generated span layout was subsequently mapped back into the 3D terrain model (Figure 6(e)), producing a conceptual bridge span design aligned with the terrain and boundary conditions.



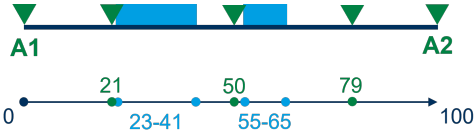
(a) 3D terrain model



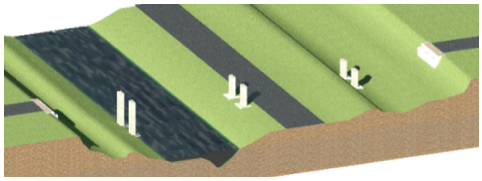
(b) Longitudinal 2D terrain



(c) Schema of ontology-based reasoning results



(d) Generated span layout through trained RL agent



(e) Bridge span layout in the 3D terrain model

Figure 6: A case study of bridge span planning using developed hybrid method

Conclusions and outlook

This paper presents an AI-assisted pipeline for bridge conceptual design, with a focus on automated span planning. Design rules are systematically collected and separated into topography-dependent feasibility rules and general layout principles related to efficiency and economy. The former are formalized through ontology-based reasoning to define the feasible design space, while RL is used to optimize span layouts under softer engineering preferences. The results demonstrate the feasibility of the proposed hybrid reasoning-and-learning framework for AI-assisted bridge span planning.

Despite these promising results, several limitations remain. Overall, the present study is intended as a proof-of-concept validation of the hybrid reasoning-and-learning framework rather than a comprehensive benchmark against alternative search strategies. From the ontology perspective, practical adoption is constrained by the effort and expertise required to develop and maintain

ontologies with specialized languages and tools, as well as the need for engineers to formalize rules, configure the knowledge base, and interpret outputs. This highlights an important direction for future research, i.e., improving automation and user-friendliness at the knowledge-engineering level. In this context, Large Language Model (LLM)-based agents may help lower this barrier by supporting semi-automatic rule formalization, ontology population, and natural-language interaction with the system. From the RL perspective, the model performance is also sensitive to the formulation of the reward function. Although rule-guided reward shaping enables the incorporation of engineering knowledge into the learning objective, the selection, weighting, and interaction of reward components strongly affect training dynamics and can encode subjective trade-offs, limiting generalization across boundary conditions and bridge typologies. Future work will therefore explore adaptive or hierarchical reward strategies, automated tuning of reward weights, and broader evaluation across unseen boundary-condition scenarios, including comparisons with constrained-search and heuristic baselines, to improve robustness and reduce manual calibration.

In summary, this work demonstrates that combining ontology-based reasoning with reinforcement learning enables a robust and extensible framework for AI-assisted bridge conceptual design. The proposed framework will be extended to the downstream design task of superstructure cross-section design in our research project by combining ontology-based rule reasoning and SL. The pipeline is intended as a decision-support tool that improves efficiency, consistency, and transparency in early-stage bridge planning, while laying the groundwork for more automated and user-friendly design workflows.

Acknowledgments

This research was supported by the HyBridGen project (project number 01FV2067A) funded by the German Federal Ministry of Transport (BMV) within the framework of the mFUND funding program.

References

Aqlan, S., Hamdan, A. H., Marx, S., & Kang, C. (2025, July). Ontology-driven Database for Integrating Bridge Data from Multiple Open Sources. In EC3 Conference 2025 (Vol. 6, pp. 0-0).

Balmer, V., Kuhn, S. V., Bischof, R., Salamanca, L., Kaufmann, W., Perez-Cruz, F., & Kraus, M. A. (2024). Design space exploration and explanation via conditional variational autoencoders in meta-model-based conceptual design of pedestrian bridges. Automation in Construction, 163, 105411.

Cao, J., Vakaj, E., Soman, R. K., & Hall, D. M. (2022). Ontology-based manufacturability analysis automation for industrialized construction. Automation in Construction, 139, 104277.

Garcia, F., & Rachelson, E. (2013). Markov decision processes. Markov Decision Processes in Artificial Intelligence, 1-38.

Hamdan, A.-H. & Scherer, R. J. (2020). Integration of BIM-related bridge information in an ontological knowledgebase. In 8th Linked Data in Architecture and Construction Workshop - Proceedings.

Jeong, J. H., & Jo, H. (2021). Deep reinforcement learning for automated design of reinforced concrete structures. Computer-Aided Civil and Infrastructure Engineering, 36(12), 1508-1529.

Liu, J., Liu, P., Feng, L., Wu, W., Li, D., & Chen, Y. F. (2020). Automated clash resolution for reinforcement steel design in concrete frames via Q-learning and Building Information Modeling. Automation in Construction, 112, 103062.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... & Hassabis, D. (2015). Human-level control through deep reinforcement learning. nature, 518(7540), 529-533.

Qian, H., Panagiotou, E., Peng, M., Ntoutsi, E., Kang, C., & Marx, S. (2024). A novel dataset and feature selection for data-driven conceptual design of offshore jacket substructures. Ocean Engineering, 303, 117679.

Rasmussen, M. H., Lefrançois, M., Schneider, G. F., & Pauwels, P. (2020). BOT: The building topology ontology of the W3C linked building data group. Semantic Web, 12(1), 143-161.

Ren, G., Ding, R., & Li, H. (2019). Building an ontological knowledgebase for bridge maintenance. Advances in Engineering Software, 130, 24-40.

Towers, M., Kwiatkowski, A., Terry, J., Balis, J. U., De Cola, G., Deleu, T., ... & Younis, O. G. (2024). Gymnasium: A standard interface for reinforcement learning environments. arXiv preprint.

Troitsky, M. S. (2019). Conceptual bridge design. In Bridge engineering handbook (pp. 1-1). CRC Press.

Wen, Y., Wang, M., Ariyachandra, M., Wei, R., & Xiao, I. B. (2024). Knowledge Driven Rule-Based Building Geometric Digital Twin Construction–State of Art Review. EC3 2024 proceeding.

Zhang, P. F., Zhang, D., Zhao, X. L., ... & Zhao, Q. (2025). Natural language processing-based deep transfer learning model across diverse tabular datasets for bond strength prediction of composite bars in concrete. Computer-Aided Civil and Infrastructure Engineering, 40(7), 917-939.

Zhang, Y., Liu, Y., Lei, G., Liu, S., & Liang, P. (2022): An Enhanced Information Retrieval Method Based on Ontology for Bridge Inspection. In: Applied Sciences 12 (20), S. 10599.

Zheng, H., Moosavi, V., & Akbarzadeh, M. (2020). Machine learning assisted evaluations in structural design and construction. Automation in Construction, 119, 103346.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

OPEN-SOURCE VLMS FOR ENGINEERING DRAWING METADATA: QWEN-VL TITLE-BLOCK DETECTION AND INFORMATION EXTRACTION

Mengyan Peng, Angran Cui, Steffen Marx, and Chongjie Kang
Dresden University of Technology, Institute of Concrete Structures
Corresponding author: mengyan.peng@tu-dresden.de

Abstract

Title-blocks in engineering drawings contain essential metadata, but automated interpretation remains difficult because of historical scans and heterogeneous layouts. This paper presents a vision-language model (VLM)-based workflow for title-block detection and structured information extraction using the open-source Qwen2.5-VL backbone. The workflow combines promptable visual grounding on full drawings with schema-constrained extraction from cropped title-block images. The study focuses on three representative fields: scale, drawing name, and route. Results show promising performance for both tasks and indicate that open-source VLMs can provide a simplified alternative to conventional multi-stage pipelines for architecture, engineering, and construction (AEC) document analysis.

Introduction

Engineering drawings remain a central information carrier throughout architecture, engineering, and construction (AEC) workflows. Despite the increasing adoption of digital design tools, large volumes of legacy drawings, often available only as scanned images, are still routinely accessed for maintenance, renovation, compliance checking, and digital asset management. Within these drawings, the title-block plays a critical role, as it encodes essential metadata such as the drawing title, scale, drawing number, creation date, author, and responsible organizations, making reliable extraction important for indexing, version control, and building information modeling (BIM) integration.

However, automated extraction of title-block information remains challenging in practice. Historical scans often exhibit low image quality, heterogeneous layouts, and inconsistent typography. Engineering drawings are visually complex documents in which textual content is embedded in spatially structured layouts and mixed with graphical elements, making reliable parsing difficult (Khan et al., 2025; Wang et al., 2024). Moreover, many older drawings contain handwritten or semi-structured annotations, further complicating recognition. While modern optical character recognition (OCR) systems perform well on clean, machine-printed text, their performance can decline for handwritten engineering-drawing content and under challenging imaging conditions (Peng et al., 2023; Peng et al., 2026).

To address these challenges, existing approaches to engineering drawing information extraction have often relied on rule-based heuristics or multi-stage learning pipelines. Rule-based methods often require manual customization and may become less robust when layouts deviate from expected templates (Gölzhäuser et al., 2023; Faltin et al., 2024). More recent data-driven approaches often adopt detector-based pipelines, sometimes augmented with large language models (LLMs) or vision-language models (VLMs) for post-processing and reasoning (Lombardi et al. 2025). While such hybrid solutions can improve extraction accuracy, they often involve complex multi-model workflows, may depend on proprietary components, and can require additional adaptation when applied to new document styles, particularly in privacy-sensitive AEC environments.

Recent advances in VLMs suggest that such models can align visual content with textual queries and generate structured outputs from complex images (Li et al., 2025; Shinde et al., 2025). However, their use for engineering drawings remains limited, particularly for workflows that employ the same open-source VLM backbone family for both title-block localization and structured information extraction. This gap motivates the VLM-based workflow proposed in this paper.

In this paper, we investigate a VLM-based workflow for title-block detection and structured information extraction using the open-source Qwen2.5-VL backbone. The two tasks are fine-tuned and evaluated separately: detection is formulated as promptable visual grounding on full engineering drawings, while extraction is performed on cropped title-block images. By using the same backbone family in both stages, the workflow avoids a heterogeneous multimodal toolchain and reduces pipeline complexity. Experimental results on bridge engineering drawings show promising performance and motivate further study of open-source VLMs for AEC document understanding.

Related Work

Engineering Drawing Information Extraction

Engineering drawing information extraction has attracted attention for decades, particularly due to the need for digitization and archival management (Scheibel et al., 2021; Kashevnik et al., 2023). Early work focused on

title-block localization and metadata extraction, a key step for indexing and retrieval from technical drawing archives (e.g., Najman et al., 2001). Traditional OCR-based pipelines often combine optical character recognition with layout-dependent heuristics, geometric rules, and manually designed templates to locate and crop title-blocks before field-level text recognition and cleaning (e.g., Ondrejcek et al., 2009). While often effective on standardized layouts, such methods may become less reliable on diverse templates and degraded scans. More recent work has explored hybrid pipelines that integrate detection models with LLMs to improve both localization and structured extraction on complex or noisy drawings (Lombardi et al., 2025).

Multi-model-based Pipelines

To overcome the limitations of rule-based systems, recent work has adopted learning-based pipelines that combine object detection with OCR. In such approaches, detectors such as You Only Look Once (YOLO) (Khanam and Hussain, 2024) or Faster R-CNN (Liu et al., 2017) are used to localize regions of interest, followed by text recognition and post-processing (Faltin et al., 2024). More recently, LLMs and VLMs have been integrated with visual pipelines to improve semantic interpretation and schema mapping (Li et al., 2019; Huang et al., 2022; Lombardi et al., 2025). Although such pipelines can improve accuracy, they often involve multiple interconnected components, which may increase deployment and maintenance effort.

VLMs in Document Understanding

Recent advances in VLMs support visual-text alignment and structured output generation from complex images. Models such as LayoutLM have shown strong performance in document understanding and information extraction by jointly modeling text and visual features (Xu et al., 2020; Xu et al., 2021; Huang et al., 2022). Benchmarks such as DocVQA further validate the usefulness of joint visual-textual representations in complex document tasks (Mathew et al., 2021). More recently, general-purpose VLMs such as GPT-4 and Qwen-VL have extended these capabilities beyond document-specific architectures and support prompt-based visual grounding, multimodal reasoning, and flexible output generation (Achiam et al., 2023; Bai et al., 2025). However, their use for engineering drawings remains limited, particularly for workflows that employ the same open-source VLM backbone family for both title-block localization and structured information extraction. This gap motivates the VLM-based workflow proposed in this paper.

Methodology

Overall Workflow

We propose a VLM-based workflow for title-block detection and structured information extraction using the same Qwen2.5-VL backbone family for both tasks, as shown in Figure 1.

At inference time, the workflow operates in two prompting steps. First, the complete engineering drawing is prompted to localize the title-block via visual grounding. Second, the cropped title-block region is prompted to extract structured information. This design avoids the use of standalone detectors or OCR engines within the proposed workflow.

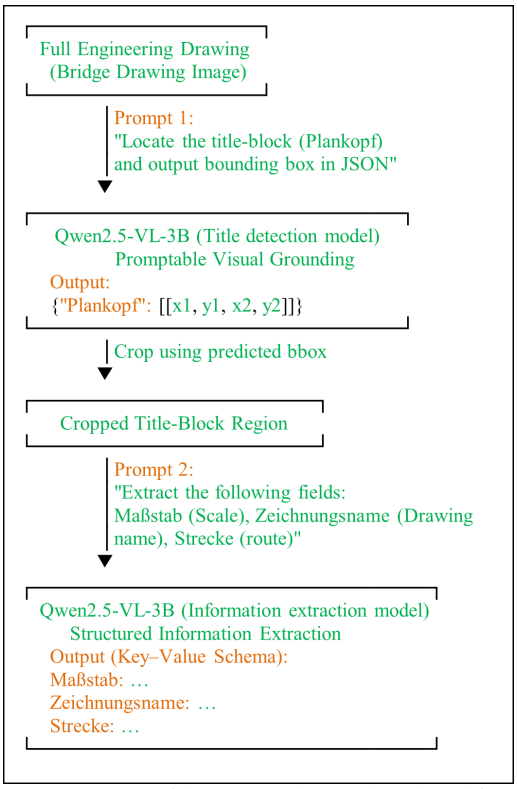


Figure 1: Overview of the proposed VLM-based workflow with separate task-specific stages for detection and structured extraction.

Title-Block Detection via Promptable Visual Grounding

Instead of training a standalone object detector, we formulate title-block localization as a promptable visual grounding task within a VLM. Given a full engineering drawing, the model receives the image together with a natural-language instruction (e.g., “locate the title-block” or “locate the Plankopf”) and directly outputs bounding-box coordinates in a structured JSON format.

This formulation avoids detector-specific training pipelines and allows localization targets to be specified flexibly through language. By performing localization within the same VLM backbone family used for the extraction stage, the predicted bounding boxes can be directly used to crop the title-block region for subsequent processing.

To support multilingual engineering documentation, we employ both English (“title-block”) and German (“Plankopf”) prompts during training and inference and observe consistent grounding performance under both formulations.

Structured Information Extraction

Following title-block detection, the cropped region is re-prompted to perform structured information extraction. We formulate this step as a prompt-based generation task, where the model is instructed to extract metadata fields and return them in a fixed key–value schema. In this study, we focus on three representative attributes commonly present in title-blocks of bridge engineering drawings, namely scale (Maßstab), drawing name (Zeichnungsname), and route (Strecke), as an initial step toward broader metadata extraction.

By explicitly constraining both the target attributes and the output format in the prompt, the model is used to perform OCR-like text recognition and semantic interpretation within a single prompted inference step, without relying on rule-based parsing or post-processing. This design supports structured extraction across layout variations and mixed printed–handwritten content commonly found in historical engineering drawings.

Model Fine-Tuning Strategy

To adapt the pre-trained VLM to the domain of engineering drawings, we apply parameter-efficient fine-tuning based on Low-Rank Adaptation (LoRA). While Qwen2.5-VL-3B-Instruct provides strong general-purpose visual and textual understanding, engineering drawings differ substantially from natural images in terms of visual structure, symbol usage, and domain-specific terminology. Fine-tuning is therefore used to better address this domain gap and improve performance in title-block detection and metadata extraction.

During fine-tuning, the vision tower is frozen to preserve stable low-level visual representations learned during large-scale pre-training. Trainable LoRA adapters, together with the multimodal projection layer, are introduced into the language model component to specialize the model for domain-specific tasks. This strategy significantly reduces the number of trainable parameters and the computational cost, while still supporting practical adaptation under limited annotated data. Although both tasks use the same Qwen2.5-VL-3B-Instruct backbone family, they are fine-tuned separately because they differ in input granularity, output format, and supervision type: detection uses full drawings with box-level supervision, whereas extraction uses cropped title-block images with field-level supervision.

Dataset and Annotation

Dataset Overview

For title-block detection, we use 500 full bridge engineering drawings with box-level title-block annotations for training and validation, comprising 450 training drawings and 50 validation drawings. In addition, an independent test set of 64 full engineering drawings, separate from the training and validation data, is used for evaluation.

For structured information extraction, we use a crop-level development dataset of 467 title-block images for training and validation, comprising 420 training samples and 47 validation samples. In addition, evaluation is conducted on two independent crop-level test sets consisting of 50 machine-printed and 50 handwritten title-block crops, respectively. These test sets are separate from the development data.

The dataset exhibits substantial variation in layout, typography, and image quality. Machine-printed title-blocks often contain denser information, tabular structures, and standardized fonts, whereas handwritten title-blocks typically have simpler layouts but introduce challenges such as writing-style variability, fading ink, and scanning artifacts. Together, these categories support evaluation across both contemporary and legacy engineering documents.

Annotation Schema

Each sample is annotated at two complementary levels for detection and information extraction.

```
{
 "image": "<image>",
 "conversation": [
  {
   "role": "user",
   "content": "Locate the Plankopf (title-block) in this
image and output its bounding-box coordinates in JSON
format."
  },
  {
   "role": "assistant",
   "content": "{\"Plankopf\": [[965, 1090, 1268, 1262]]}"
  }
 ]
}
```

Figure 2: Example annotation of title-block detection.

```
{
 "image": "<image>",
 "conversation": [
  {
   "role": "user",
   "content": "Please extract the following information
from the image and output it in the specified
format:\nMaßstab:\nZeichnungsname:\nStrecke:"
  },
  {
   "role": "assistant",
   "content": "Maßstab: 1:50\nZeichnungsname:
Widerlager Achse 20\nStrecke: Hamburg – Berlin, km
0,628"
  }
 ]
}
```

Figure 3: Example annotation of structured information extraction.

First, box-level annotations define the spatial extent of the title-block within the full engineering drawing, as illustrated in Figure 2. For this task, the model is prompted to locate the title-block, and the corresponding ground-truth response provides rectangular bounding-box coordinates in JSON format. These annotations are used

to supervise and evaluate title-block localization performance.

Second, field-level annotations specify the ground-truth values for three target metadata attributes, namely scale, drawing name, and route within the title-block, as illustrated in Figure 3. Given a cropped title-block image, the model is instructed to extract these attributes and return them in a fixed key–value schema.

By representing both detection and extraction annotations as conversational image-text pairs with structured outputs, the dataset enables direct supervision of heterogeneous tasks within the same VLM backbone family. This design reduces the need for task-specific post-processing rules or layout assumptions and facilitates the combined use of the two models at inference time.

Experiments and Results

Experimental Setup

We evaluate the proposed VLM-based workflow on two related tasks: title-block detection and structured information extraction. Both tasks use the same Qwen2.5-VL-3B-Instruct backbone family but are fine-tuned separately because they differ in input granularity and output targets. During preliminary development, we explored a small set of LoRA configurations and selected the final settings based on stable convergence and validation performance. For title-block detection, the final LoRA configuration used rank 32, scaling factor 16, and dropout 0.05. For structured information extraction, the final configuration used rank 32, scaling factor 64, and dropout 0.05. The model was optimized using AdamW with an initial learning rate of 1×10^{-4}, a cosine learning-rate scheduler, and 70 warmup steps. Mixed-precision training with bfloat16 (bf16) was used to reduce memory consumption. For title-block detection, the model was fine-tuned for 8 epochs, and the best-performing checkpoint was selected based on validation loss. For structured information extraction, the final checkpoint was selected after 4 epochs, consistent with the observed validation-loss minimum around the fourth epoch.

Baselines

For title-block detection, we compare the fine-tuned Qwen2.5-VL-3B-Instruct model against two references: (1) YOLOv11-medium as a conventional detector baseline and (2) zero-shot Qwen2.5-VL-3B-Instruct without task-specific fine-tuning. All models are evaluated on the same 64-drawing test set using precision, recall, F1, and mean IoU at an IoU threshold of 0.5.

For structured information extraction, we compare the fine-tuned Qwen2.5-VL-3B-Instruct model against the corresponding zero-shot configuration without task-specific fine-tuning. Both models are evaluated on the same crop-level test sets of machine-printed and handwritten title-block images using Word Accuracy (WA) and Character Accuracy (CA). Per-field results are additionally reported for Maßstab, Zeichnungsname, and Strecke.

Title-Block Detection

We evaluate title-block detection using precision, recall, F1, and mean Intersection over Union (IoU) at an IoU threshold of 0.5. Predictions above the threshold are counted as true positives (TP), lower-overlap detections are treated as false positives (FP), and missed title blocks are treated as false negatives (FN).

Table 1: Comparative Performance for Title-Block Localization

Metric	Fine-tuned Qwen2.5-VL-3B	YOLOv11 -medium	Zero-shot Qwen2.5-VL-3B
Precision	93.8%	91.0%	0.0%
Recall	93.8%	95.3%	0.0%
F1	93.8%	93.1%	0.0%
Mean IoU	0.871	0.951	---
TP	60	61	0
FP	4	6	0
FN	4	3	64
Output Validity Rate	100%	98.4%	100%

Table 1 compares the fine-tuned Qwen2.5-VL-3B-Instruct model with the YOLOv11-medium baseline and the zero-shot Qwen2.5-VL-3B-Instruct baseline on the 64-drawing test set. The fine-tuned Qwen model achieves 93.8% precision and 93.8% recall, with an F1 score of 93.8% and a mean IoU of 0.871. This performance is comparable to YOLOv11-medium (F1=93.1%), although YOLO yields a higher mean IoU (0.951). YOLO produced a small number of duplicate or mislocalized predictions, which explains why the number of detections exceeded the number of test images. In contrast, the zero-shot Qwen model does not localize title blocks successfully in this setting, despite producing syntactically valid structured outputs. This highlights the importance of domain-specific fine-tuning for engineering drawings.

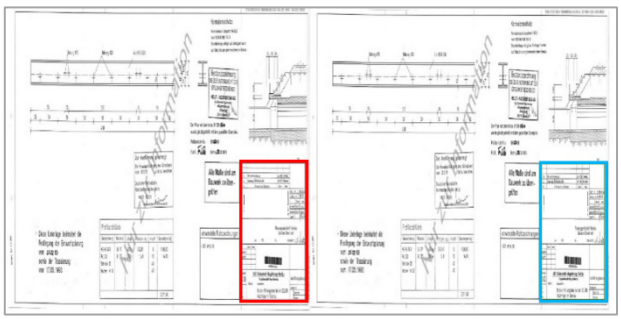


Figure 4: Representative result of title-block detection. Comparison between ground-truth annotation (red) and model prediction (blue).

For comparison, YOLO detections were converted into the same structured output format. The fine-tuned Qwen achieved a 100% output validity rate, meaning that valid structured outputs were generated for all test samples. As shown in Figure 4, the predicted bounding box closely

aligns with the ground-truth annotation in a representative drawing. Most detection failures occur in cases with unconventional title-block placement, severe visual degradation, or ambiguous boundaries between the title block and adjacent tabular regions. In a smaller number of cases, visually similar tabular areas are mistaken for title blocks, or the predicted box only partially covers the true title-block extent, leading to IoU values below the acceptance threshold.

Structured Information Extraction

To evaluate structured information extraction, we report both Word Accuracy and Character Accuracy, which capture complementary aspects of extraction quality. The metrics are computed directly on the extracted key-value outputs by comparing each predicted field value with the corresponding ground-truth value, without additional post-processing, normalization, or special handling of empty fields. Word Accuracy reflects semantic-level correctness and sensitivity to word-boundary errors, while Character Accuracy provides a more fine-grained assessment of character-level robustness.

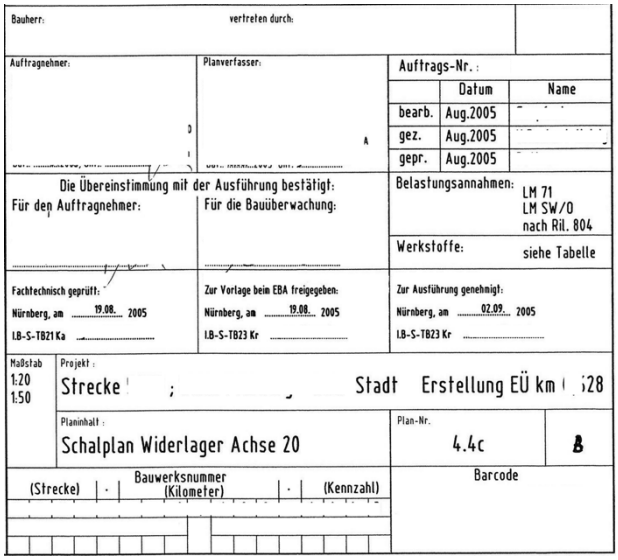


Figure 5: Example of machine-printed title-block. (Sensitive project-specific information has been anonymized for confidentiality.)

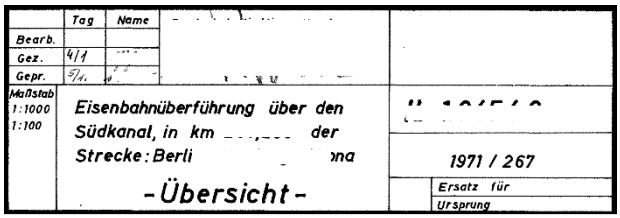


Figure 6: Example of predominantly handwritten title-block. (Sensitive project-specific information has been anonymized for confidentiality.)

We evaluate extraction performance on two independent crop-level test sets constructed from ground-truth title-block crops, as described in the Dataset Overview. One set contains 50 machine-printed title-blocks with dense layouts (Figure 5), and the other contains 50 predominantly handwritten title-blocks with relatively simple layouts but diverse handwriting styles (Figure 6). These results reflect crop-level extraction performance and are reported independently of the upstream detection stage. All results reported in this section correspond to the final configuration selected based on validation performance, as described in the Experimental Setup.

On the machine-printed title-block test set, the fine-tuned model achieves a WA of 0.78 and a CA of 0.76. On handwritten title-blocks, higher performance is observed, with WA reaching 0.83 and CA reaching 0.86. These results indicate promising extraction accuracy for both modern and historical title-block images.

Table 2: Comparative performance for structured information extraction on handwritten title-block crops (Field 1: Maßstab; Field 2: Zeichnungsname; Field 3: Strecke).

Metric	Fine-tuned Qwen2.5-VL-3B	Zero-shot Qwen2.5-VL-3B
Field 1 WA	0.9474	0.6023
Field 1 CA	0.9495	0.5865
Field 2 WA	0.7216	0.1868
Field 2 CA	0.8479	0.4320
Field 3 WA	0.8545	0.2836
Field 3 CA	0.8671	0.3536

To assess the effect of domain-specific adaptation in more detail, we further compare the fine-tuned Qwen model with the corresponding zero-shot configuration at the field level. Tables 2 and 3 report per-field WA and CA for the three target attributes (Maßstab, Zeichnungsname, and Strecke) on handwritten and machine-printed title-block crops, respectively.

Table 3: Comparative performance for structured information extraction on machine-printed title-block crops (Field 1: Maßstab; Field 2: Zeichnungsname; Field 3: Strecke).

Metric	Fine-tuned Qwen2.5-VL-3B	Zero-shot Qwen2.5-VL-3B
Field 1 WA	0.9900	0.5721
Field 1 CA	0.9892	0.5510
Field 2 WA	0.7132	0.1716
Field 2 CA	0.7510	0.3217
Field 3 WA	0.7398	0.2520
Field 3 CA	0.7117	0.3061

Per-field analysis reveals a clear difference in extraction difficulty across the three target attributes. Maßstab is consistently the easiest field, achieving the highest WA and CA in both handwritten and machine-printed title-blocks. This is likely due to its short and highly regular format. Zeichnungsname is the most challenging field, especially at the word level, which can be attributed to its greater length, frequent line breaks, and higher sensitivity to surrounding layout complexity. Strecke shows intermediate difficulty, although its performance decreases substantially in machine-printed title-blocks,

suggesting that dense tabular layouts and closely positioned neighboring text increase ambiguity in field boundaries. Across all three fields, the fine-tuned model substantially outperforms the corresponding zero-shot baseline, highlighting the importance of domain-specific adaptation for engineering drawings.

The higher overall extraction accuracy on handwritten title-blocks is consistent with the per-field comparison. Although Maßstab achieves slightly better results in machine-printed title-blocks, both Zeichnungsname and Strecke are extracted more accurately from handwritten samples. This indicates that, for these semantically richer fields, layout simplicity and reduced visual clutter may be more important than text regularity alone. By contrast, machine-printed title-blocks often contain denser tabular structures and more closely positioned text elements, which increase ambiguity in field boundaries and make extraction more difficult, especially for drawing names and route-related content.

Maßstab (scale): 1:20, 1:50
Zeichnungsname (drawing name): Erstellung EÜ Schalplan Widerlager Achse 20
Strecke (route): …Stadt km …28

Figure 7: Representative result of structured information extraction for a machine-printed title-block, corresponding to the title-block shown in Figure 5.

Maßstab (scale): 1:1000, 1:100
Zeichnungsname (drawing name): Eisenbahnüberführung über den Südkanal Übersicht
Strecke (route): Berli…ona, km …

Figure 8: Representative result of structured information extraction for a predominantly handwritten title-block, corresponding to the title-block shown in Figure 6.

Figures 7 and 8 present representative qualitative examples of structured information extraction on machine-printed and handwritten title-blocks, respectively. In each example, the model correctly identifies and extracts multiple metadata fields, including scale, drawing name, and route, despite variations in layout, typography, and content density.

Overall, the results indicate that a fine-tuned VLM can support OCR-like text recognition and structured information extraction for title-block content without relying on external OCR engines or rule-based post-processing. This highlights the practical potential of the proposed VLM-based workflow for engineering document analysis.

Small-Scale Pipeline Evaluation

We additionally evaluated the workflow on 20 real full bridge engineering drawings on an NVIDIA RTX A5000 GPU with 24 GB of memory. The set reflects challenging and practically relevant legacy-archive conditions, including dark, speckle-noisy scans and mixed machine-printed layout fields with handwritten technical annotations. Title-block detection achieved Precision/Recall/F1 = 85%/85%/85%, with a mean IoU of 0.633. On the corresponding 20 title-block crops, extraction from ground-truth crops achieved a WA of 0.6458 and a CA of 0.7475, whereas extraction from Qwen-predicted crops achieved a WA of 0.6771 and a CA of 0.7712. Per-field trends remained consistent with the main benchmark: Maßstab was the easiest field, Zeichnungsname the most difficult, and Strecke intermediate. Notably, extraction from Qwen-predicted crops did not lead to a performance drop in this small-scale evaluation and even showed slightly higher WA and CA, suggesting that the predicted boxes were sufficiently accurate for downstream extraction and may, in some cases, reduce surrounding visual clutter. The average runtime was 2.97 s per drawing for title-block detection and 8.91 s per title-block crop for structured information extraction, with peak allocated GPU memory below 10 GB in all settings. These results indicate that the workflow is practically feasible for offline archive-style processing on a 24 GB-class GPU.

Discussion and Limitations

The experimental results indicate that the same open-source VLM backbone family can be used for both title-block detection and structured information extraction on engineering drawings. Unlike conventional detector-based pipelines, the proposed VLM-only workflow uses the same VLM backbone for both localization and interpretation, reducing system complexity while showing promising performance across heterogeneous drawing styles. In the present study, the two tasks are fine-tuned separately and combined at inference time.

One factor that may contribute to this behavior lies in the nature of title-blocks themselves. Although title-blocks exhibit substantial variation in layout, typography, and language across organizations and time periods, they are typically semantically well-defined regions that encode standardized metadata. Promptable visual grounding may help the model leverage semantic cues rather than relying solely on geometric or appearance-based patterns, which may contribute to localization performance even under unconventional layouts or degraded scan quality. This differs from traditional object detectors, which may require additional adaptation under distribution shifts or limited training samples.

Another advantage of the proposed approach is the conceptual consistency between detection and extraction through the same VLM backbone and prompt-based interaction paradigm, which may reduce intermediate conversions and error propagation in multi-stage workflows.

The information extraction results further indicate that the model shows promising performance across different title-block styles. Higher accuracy is observed on handwritten title-blocks, which can be attributed to their typically simpler layouts and lower information density compared with modern machine-printed title-blocks. In contrast, printed title-blocks often contain dense tables, multiple scale entries, and additional graphical elements, increasing extraction complexity. These observations

suggest that layout complexity may be an important challenge, in addition to handwriting, for structured extraction in engineering drawings.

From a practical perspective, the proposed VLM-only workflow offers several potential advantages for deployment in AEC environments. The use of a single open-source model reduces engineering overhead, improves reproducibility, and facilitates on-premise deployment, which may be beneficial in contexts involving privacy, regulatory compliance, and long-term maintainability. These properties may make the approach relevant for real-world engineering archives, where heterogeneous document collections and limited annotation budgets are common.

Despite its promising performance, the proposed workflow has several limitations. First, the current study focuses on three representative metadata fields, namely scale, drawing name, and route, which, while practically relevant, do not cover the full range of information present in title-blocks. Extending the schema to additional fields, such as drawing numbers, revision indices, dates, and responsible organizations would further improve applicability.

Second, although promptable visual grounding performs well on most drawings, detection errors still occur in edge cases involving unconventional layouts, severe visual degradation, or ambiguous title-block boundaries. Addressing such cases may require additional contextual prompts, multi-step reasoning, or lightweight layout priors.

Third, the workflow does not explicitly handle non-textual elements commonly found in engineering drawings, such as stamps, signatures, symbols, or dimension annotations. Integrating symbol understanding or multimodal reasoning over graphical primitives remains an open challenge.

Finally, the current dataset is limited to a bridge-engineering context, and broader generalization to other engineering domains, regions, and historical conventions remains to be validated. While the small-scale pipeline evaluation (n = 20) provides encouraging results, systematic validation on a larger full-drawing test set is still needed. Future work will therefore examine the complete workflow on full drawings using predicted title-block localization, extend the approach to other drawing domains such as buildings, Mechanical, Electrical, and Plumbing (MEP) systems, and other infrastructure types, and investigate broader baseline comparisons, including detector-plus-OCR pipelines or document-specialized models. Extending the workflow to new domains will likely require additional domain-specific data at both stages, including full-drawing box annotations for title-block localization and crop-level key-value annotations for structured extraction, as well as adaptation of the target metadata schema, prompt design, and task-specific fine-tuning. Larger-batch throughput analysis beyond the current single-sample sequential runtime setting also remains future work.

Conclusions

This paper shows that a single open-source VLM backbone family can provide a promising basis for unified title-block detection and structured information extraction in engineering drawings. By formulating title-block detection as a promptable visual grounding task and performing metadata extraction through prompt-based generation, the proposed approach combines detection and interpretation within a VLM-based workflow built on the same backbone family without relying on standalone detectors or external OCR engines.

The results across modern and historical drawings indicate that VLM-based approaches show promising performance across heterogeneous layouts and visual conditions, suggesting their potential for broader use in engineering document analysis workflows.

Overall, this work highlights the potential of open-source VLMs as a simplified and reproducible alternative to conventional multi-model pipelines for engineering drawing understanding. It also offers a useful starting point for future research on richer metadata schemas, deeper semantic understanding, and scalable digitization of engineering archives using open-source VLMs.

Data and Code Availability

The study uses the publicly available Qwen2.5-VL model (https://huggingface.co/Qwen/Qwen2.5-VL-3B-Instruct). The base architecture was not modified. The engineering drawing dataset is confidential but may be shared upon reasonable request for non-commercial research purposes, subject to institutional approval.

Acknowledgments

The authors acknowledge the support of the German Federal Ministry of Transport (BMV) within the project HyBridGen (Project No. 01FV2067A).

Declaration of generative AI and AI-assisted technologies in the Writing Process

During manuscript preparation, the authors used ChatGPT and DeepL under human oversight solely to improve language quality. The authors reviewed and revised the content as needed and take full responsibility for the final version.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). GPT-4 technical report. arXiv preprint arXiv:2303.08774. https://doi.org/10.48550/arXiv.2303.08774.

Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., ... & Lin, J. (2025). Qwen2.5-VL technical report. arXiv preprint arXiv:2502.13923. https://doi.org/10.48550/arXiv.2502.13923.

Faltin, B., Schönfelder, P., Gann, D., & König, M. (2024). Reconstructing as-built beam bridge geometry from construction drawings using deep learning-based symbol pose estimation. Advanced Engineering

Informatics, 62, 102808. https://doi.org/10.1016/j.aei.2024.102808.

Gölzhäuser, P., Peng, M., Jäkel, J. I., Klemt-Albert, K., & Marx, S. (2023). Approach to generate a simple semantic data model from 2D bridge plans using AI-based Text Recognition. In Proceedings of the 33rd European Safety and Reliability Conference (Vol. 3, pp. 2023-07). https://doi.org/10.3850/978-981-18-8071-1_P424-cd.

Huang, Y., Lv, T., Cui, L., Lu, Y., & Wei, F. (2022, October). LayoutLMv3: Pre-training for document ai with unified text and image masking. In Proceedings of the 30th ACM international conference on multimedia (pp. 4083-4091). https://doi.org/10.1145/3503161.3548112.

Kashevnik, A., Shilov, N., Teslya, N., Hasan, F., Kitenko, A., Dukareva, V., ... & Blokhin, D. (2023). An approach to engineering drawing organization: Title block detection and processing. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3244603.

Khan, M. T., Yong, Z., Chen, L., Tan, J. M., Feng, W., & Moon, S. K. (2025). Automated Parsing of Engineering Drawings for Structured Information Extraction Using a Fine-tuned Document Understanding Transformer. arXiv preprint arXiv:2505.01530. https://doi.org/10.48550/arXiv.2505.01530.

Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725. https://doi.org/10.48550/arXiv.2410.17725.

Li, L. H., Yatskar, M., Yin, D., Hsieh, C. J., & Chang, K. W. (2019). VisualBERT: A simple and performant baseline for vision and language. arXiv preprint arXiv:1908.03557. https://doi.org/10.48550/arXiv.1908.03557.

Li, Z., Wu, X., Du, H., Liu, F., Nghiem, H., & Shi, G. (2025). A survey of state of the art large vision language models: Benchmark evaluations and challenges. In Proceedings of the Computer Vision and Pattern Recognition Conference (pp. 1587-1606). https://doi.org/10.48550/arXiv.2501.02189.

Liu, B., Zhao, W., & Sun, Q. (2017, October). Study of object detection based on Faster R-CNN. In 2017 Chinese automation congress (CAC) (pp. 6233-6236). IEEE. https://doi.org/10.1109/CAC.2017.8243900.

Lombardi, A., Duan, L., Elnagar, A., Zaalouk, A., Ismail, K., & Vakaj, E. (2025). Title block detection and information extraction for enhanced building drawings search. arXiv preprint arXiv:2504.08645. https://doi.org/10.48550/arXiv.2504.08645.

Mathew, M., Karatzas, D., & Jawahar, C. V. (2021). DocVQA: A dataset for VQA on document images. In Proceedings of the IEEE/CVF winter conference on Applications of computer vision (pp. 2200-2209). https://doi.org/10.48550/arXiv.2007.00398.

Najman, L., Gibot, O., & Barbey, M. (2001, September). Automatic title block location in technical drawings. In Proceedings of 4th IAPR International Workshop on Graphics Recognition, Kingston, Ontario (Canada) (pp. 19-26).

Ondrejcek, M., Kastner, J., Kooper, R., & Bajcsy, P. (2009). Information extraction from scanned engineering drawings. Image Spatial Analysis Group, National Center for Supercomputing Applications, NCSA-ISDA09-001.

Peng, M., Kang, C., & Marx, S. (2023). Text Recognition for 2D Bridge Plans Using OCR-Algorithms. ce/papers, 6(5), 661-666. https://doi.org/10.1002/cepa.2077.

Peng, M., Qian, H., Marx, S., & Kang, C. (2026). Optimizing 2D Bridge Engineering Drawing Digitization: A Comparative Study of Text Recognition Tools and Development of Lightweight Post-Recognition Structured Information Extraction Methods. Results in Engineering, 110186. https://doi.org/10.1016/j.rineng.2026.110186.

Scheibel, B., Mangler, J., & Rinderle-Ma, S. (2021). Extraction of dimension requirements from engineering drawings for supporting quality control in production processes. Computers in Industry, 129, 103442. https://doi.org/10.1016/j.compind.2021.103442.

Shinde, G., Ravi, A., Dey, E., Sakib, S., Rampure, M., & Roy, N. (2025). A Survey on Efficient Vision-Language Models. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 15(3), e70036. https://doi.org/10.1002/widm.70036.

Wang, D., Raman, N., Sibue, M., Ma, Z., Babkin, P., Kaur, S., ... & Liu, X. (2024, August). DocLLM: A layout-aware generative language model for multimodal document understanding. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 8529-8548). https://doi.org/10.18653/v1/2024.acl-long.463.

Xu, Y., Xu, Y., Lv, T., Cui, L., Wei, F., Wang, G., ... & Zhou, L. (2021, August). LayoutLMv2: Multi-modal pre-training for visually-rich document understanding. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers) (pp. 2579-2591). https://doi.org/10.18653/v1/2021.acl-long.201.

Xu, Y., Li, M., Cui, L., Huang, S., Wei, F., & Zhou, M. (2020, August). LayoutLM: Pre-training of text and layout for document image understanding. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 1192-1200). https://doi.org/10.1145/3394486.3403172.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATED BIM LOD INFERENCING: A HYBRID APPROACH USING KNOWLEDGE GRAPHS AND MULTIMODAL LARGE LANGUAGE MODELS

Taeyoon Kim[1], Ghang Lee[1]
[1]Yonsei University, Seoul, Rupblic of Korea

Abstract

Automated Level of Development (LOD) assessment in Building Information Modeling (BIM) is hindered by semantic ambiguity in specifications and geometric complexity of objects. Existing rule-based and vision-based methods lack contextual reasoning or struggle with fragmented, occluded geometry. This study proposes a hybrid framework combining a three-layered knowledge graph—built top-down via LLM requirement extraction and bottom-up via visual population—with a geometric pipeline reassembling meshes and exposing occluded components. A multimodal LLM cross-references graph-derived criteria with multi-view geometry, producing explainable assessments. A four-condition ablation study achieves 83.47% accuracy while reducing false negatives, confirming feasibility for explainable BIM quality assurance.

Introduction

Research Background

As digital transformation accelerates within the construction industry, Building Information Modeling (BIM) has established itself as a core database for managing information throughout the entire project lifecycle, moving beyond simple 3D geometric modeling. To maximize the utility of BIM data, it is essential to verify the accuracy of the information and ensure compliance with the Level of Development (LOD) standards required for each project phase. However, existing LOD verification methods heavily rely on manual inspection by experts, which incurs significant time and costs. Furthermore, conventional rule-based automation tools lack the flexibility to handle various geometric variations and complex exceptions. While recent studies have attempted to leverage Artificial Intelligence (AI) to address these issues, two fundamental barriers remain: the "semantic gap" and "geometric complexity."

Problem Statement

The first challenge is the semantic gap between human language and machine understanding. LOD specifications are written in natural language, possessing linguistic characteristics that are difficult for machines to interpret intuitively. For instance, umbrella terms such as "All Reinforcement" do not map one-to-one with specific objects (e.g., main bars, stirrups) within the model. Additionally, context-dependent terms like "Connection" exhibit polysemy, where the target object for inspection changes depending on whether the system is Steel or Precast Concrete. Simple text-matching algorithms may utilize static dictionaries, but they fail to dynamically infer these hierarchical and contextual relationships, preventing the scalable application of automated verification.

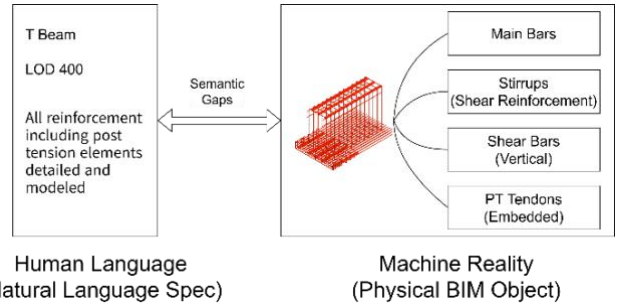


Figure 1 The semantic gap in LOD verification.

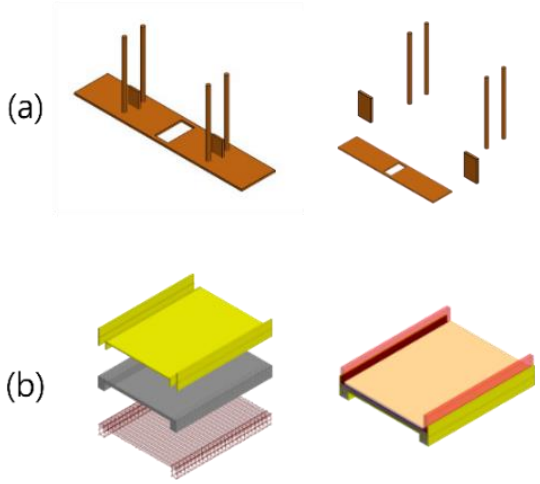


Figure 2 Visual illustration of geometric complexity in BIM objects. (a) Fragmentation: A single embedded component is geometrically constructed from multiple disconnected meshes. (b) Occlusion: A baseplate connection hides internal reinforcement behind opaque surfaces, impeding visual identification.

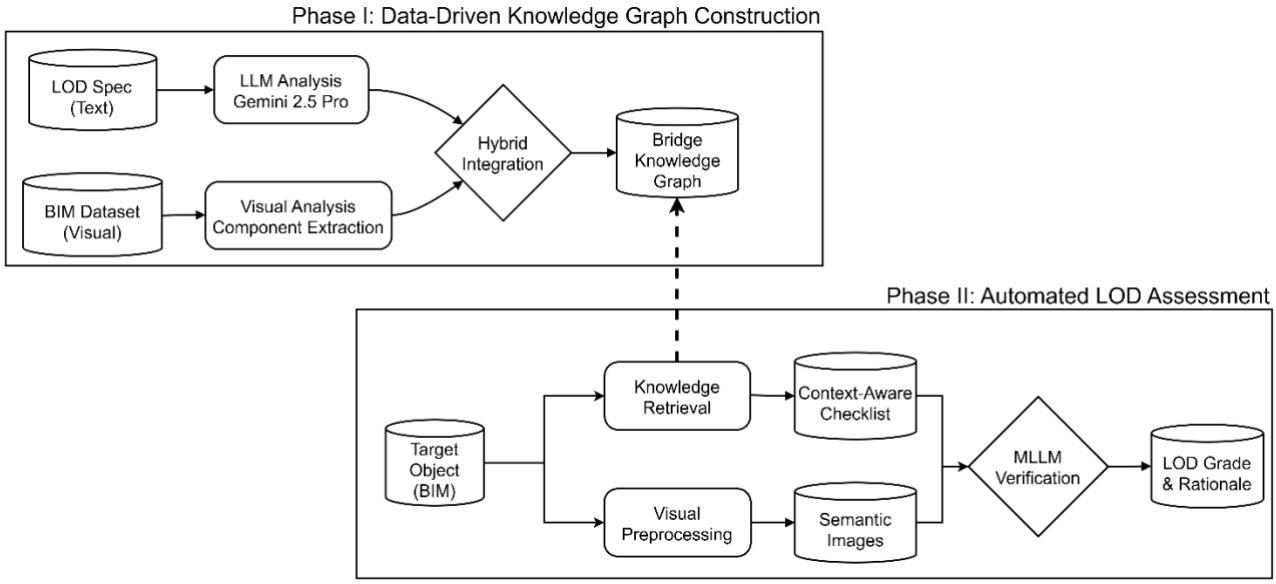


Figure 3 The architecture of the proposed dual-phase LOD assessment framework

The second challenge is geometric complexity. Although a BIM object may logically represent a single element (e.g., a Precast Concrete T-beam), its sub-components often exhibit geometric fragmentation. For instance, a single embedded steel part is frequently composed of multiple disconnected meshes—where the base plate and studs are modeled as separate geometric entities—lacking a unified topological structure (fragmentation). Conversely, objects often contain critical internal components, such as reinforcement within a baseplate connection, that are visually obscured by external surfaces (occlusion). If such objects are simply captured as raw images, the AI model is likely to misinterpret the fragmented meshes as separate unrelated objects, while failing to recognize the hidden internal components.

Research Objective and Organization

The objective of this research is to establish an intelligent quality assessment system that integrates domain knowledge with visual reasoning. Unlike black-box AI approaches, this study proposes a knowledge graph-based framework to ensure explainable assessments. Specifically, the system aims to (1) resolve linguistic ambiguities in LOD standards by constructing a logical reasoning engine, and (2) overcome visual occlusion and fragmentation through semantic geometric preprocessing. By bridging these semantic and geometric gaps, the framework enables MLLMs to perform rigorous and traceable LOD verification.

Related Works

This section reviews existing literature on BIM object classification and quality assessment, categorizing them into (1) geometry and metadata-based approaches, (2) deep Learning-based visual classification, and (3) Recent MLLM applications. It highlights the limitations of each approach and defines the differentiation of the proposed research.

Geometry and Metadata-based Approaches

Early studies focused on utilizing geometric features or textual metadata of BIM objects for classification. (Wu and Zhang, 2019) proposed a hybrid algorithm combining geometric features with keyword matching to enhance BIM interoperability. However, this method relies heavily on pre-defined rules, lacking flexibility for irregular objects that deviate from standard specifications. Similarly, (Abualdenien and Borrmann, 2022) utilized ensemble learning to classify the Level of Geometry (LOG) based on geometric complexity (e.g., area, volume). While effective for assessing geometric precision, this approach fails to semantically identify the presence of specific components (e.g., bolts, nuts), which is the core requirement for actual LOD assessment.

Vision-based Deep Learning and Data Dependency

To overcome the limitations of metadata reliance, research on converting 3D BIM objects into 2D images for classification using Convolutional Neural Networks (CNN) has gained traction. (Yang et al., 2026) improved classification performance for architectural and MEP objects using multi-view images. However, these Supervised Learning approaches entail a significant critical limitation: they require the costly pre-construction of massive, labeled datasets containing thousands of images.

To address performance issues,(Yu et al., 2023) and (Utkucu et al., 2024) proposed Ensemble Deep Learning techniques and Random Forest integration to enhance accuracy and compatibility with classification systems like Uniclass. Despite these improvements, they still fail to overcome the fundamental "Data Hunger" problem; whenever a new LOD standard or a new object type is introduced, new data must be collected, and the model must be re-trained. Furthermore, relying solely on exterior visual features prevents these models from identifying

occluded components—such as rebar inside concrete—which are essential for detailed LOD verification.

Multimodal LLMs and Ontology Approaches in Generative Interpretation

Recently, studies utilizing MLLMs have emerged to achieve Zero-shot classification and reduce data construction costs. (Wang, X. et al., 2025) demonstrated the use of models like GPT-4V to detect surface cracks on steel bridges, and (Zeng et al., 2024) estimated building age solely from facade images. While these studies show the potential of MLLMs, they are limited to surface-level image analysis. They have not attempted to integrate domain knowledge (knowledge graph) to evaluate objects with complex hierarchies and contexts, such as BIM elements. Without structured knowledge, general-purpose MLLMs are prone to hallucinations and fail to grasp the specific engineering nuances required for LOD assessment.

Recent advancements in Natural Language Processing (NLP) and ontology have significantly improved the structuring of unstructured construction data, primarily enhancing information extraction and retrieval tasks. For instance, (Nahri et al., 2025) utilized transformer-based models to extract explicit requirements from technical specifications, while (Yin et al., 2023) developed an ontology-aided approach to translate natural language queries for retrieving information from existing BIM models. Similarly, (F.h et al., 2025) formalized the ISO 19650 standard into the BIM-OIM ontology to support process compliance. While these studies effectively parse explicit text or validate established processes, the domain of generative modeling presents an opportunity for further exploration regarding the "Semantic gap." Current methods typically focus on direct mapping, and there is potential to extend these approaches to include the inferential reasoning required to translate abstract "Human Standards" (e.g., vague LOD descriptions) into concrete "Deep Standards" (e.g., specific component hierarchies). This study aims to build upon these foundations by introducing a semantic bridge to infer implied domain concepts—such as deducing required Connection Assemblies from a text that mentions anchor rods—thereby facilitating the automated decision-making process in BIM object creation.

Methodology

Framework Overview

This study proposes a comprehensive framework designed to bridge the semantic gap between human requirements and machine interpretation for automated LOD verification. The proposed system is structurally divided into two primary phases: (1) Phase I: Data-Driven knowledge graph Construction and (2) Phase II: Automated LOD Assessment. In the first phase, a Bridge knowledge graph (KG) is constructed as a foundational reasoning engine. Instead of relying on static rule-based definitions, this knowledge base is developed through a hybrid data-driven process where natural language requirements are parsed by LLM, specifically Gemini 2.5 Pro, while physical component vocabularies are derived from the visual analysis of high-fidelity BIM data. In the second phase, the system executes the assessment by cross-referencing this pre-validated knowledge with the geometric data of the target BIM object. This dual-phase approach ensures both the scalability of knowledge acquisition and the consistency of real-time verification.

Phase I: Data-Driven knowledge graph Construction

Three-Layer Graph Architecture

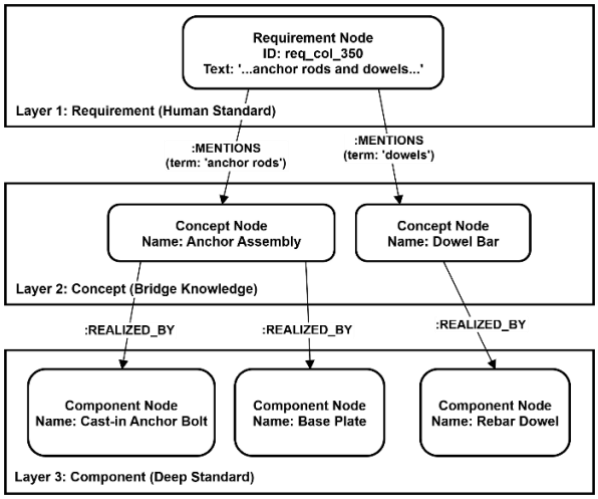


Figure 4 The three-layered architecture of the Bridge knowledge graph.

A critical limitation of existing LOD verification methods is the inability to rigorously map abstract natural language requirements to specific 3D objects. In this study, we distinguish between two levels of specification: "Human Standards," defined as the natural-language requirements stated in LOD specifications (e.g., "Include anchor rods and dowels"), which are intended for human interpretation and often underspecified for automated inspection. In contrast, "Deep Standards," denote a machine-interpretable representation of these requirements, expressed as structured inspection criteria that explicitly define each required sub-component in terms of its identity and expected geometric form (e.g., "Cast-in Anchor Bolt: cylindrical rod embedded in concrete"). The term “Deep” emphasizes that these standards capture latent, operational constraints implicit in human-readable specifications but not explicitly stated. The core objective of the proposed knowledge graph is to bridge this gap by systematically translating Human Standards into Deep Standards, enabling reliable reasoning and verification by downstream visual and geometric analysis modules.

To achieve this, we designed a three-layer graph architecture, as illustrated in Figure 4. The top layer, Requirement Nodes, encapsulates the raw textual requirements extracted from standard specifications. For example, as shown in the diagram, a requirement for LOD 350 Columns might state: "Include anchor rods and dowels." The bottom layer, Component Nodes, consists of concrete physical elements existing in the modeling

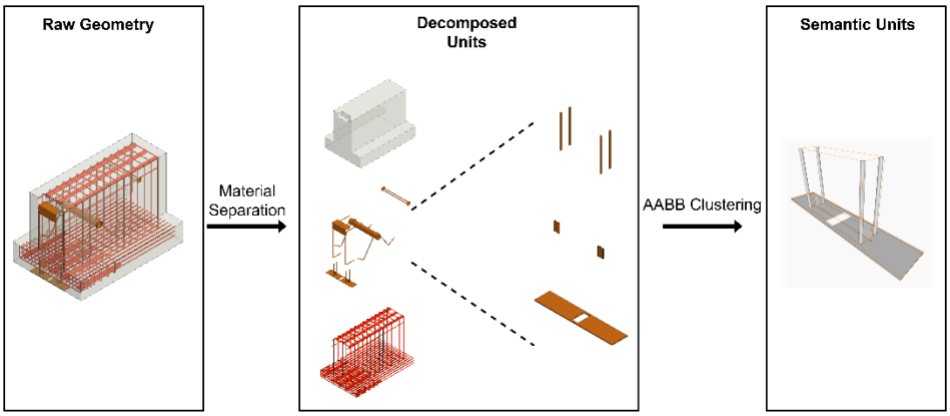


Figure 5 Transformation from raw geometry to semantic units. Visualization of the preprocessing steps: Raw input, Decomposed units via material separation, and Re-assembled semantic units via AABB clustering.

environment, such as "cast-in anchor bolt," "base plate," and "rebar dowel." Connecting these two is the Concept Node Layer (Bridge Layer). This intermediate layer acts as a semantic hub that abstracts ambiguous terms into system-specific engineering concepts. In Figure 4, the term "anchor rods" from the requirement is mapped to the concept "anchor assembly" (via the "mentions" relationship), which is then physically realized by components like "Cast-in Anchor Bolt" and "Base Plate" (via the "realized_by" relationship). This structure effectively resolves linguistic ambiguity by integrating diverse expressions into unified technical definitions.

It should be noted that the semantic mapping between layers in the current knowledge graph was manually curated. While LLMs assisted in extracting requirements and suggesting candidate concepts, the final alignment was performed through manual validation. Consequently, when a new LOD standard is introduced, the Requirement layer must be reconstructed, whereas the Concept and Component layers—representing reusable engineering vocabulary—can be largely retained and extended incrementally.

Hybrid Learning Pipeline

To populate this three-layer structure, we employed a Hybrid Learning Pipeline that synthesizes both textual and visual data. First, adopting a top-down approach, we utilized GPT-5.1 to analyze unstructured PDF standard documents. The model extracts core requirements and infers implied engineering concepts (implied concepts) to construct the Requirement and Concept layers. Simultaneously, a Bottom-Up approach was applied using Visual Analysis techniques on a dataset of high-fidelity (LOD 400) BIM models. By scanning these models, the system extracted a "Visual Vocabulary" of actual existing components to populate the Component Layer. Finally, through a Semantic Alignment process, these two streams were mapped to complete a validated knowledge graph that eliminates the discrepancy between "what is required" (Human Standards) and "what is modeled" (Deep Standards).

Phase II: Context-Aware & Explainable Assessment

Context-Aware Knowledge Retrieval

During the verification phase, the system utilizes the pre-constructed knowledge graph as a deterministic reasoning engine. When a user selects a BIM object, the system queries the graph using the object's classification code (e.g., OmniClass). The query engine performs Context-Aware Retrieval, dynamically interpreting ambiguous terms based on the object's system type—for instance, mapping the term "Connection" to "Anchor Bolts" for a steel column, while mapping it to "Shear Tabs" for a beam. This process instantly generates a precise, object-specific checklist without redundant inference computations.

Semantic Geometric Preprocessing

In typically complex BIM environments, internal components are often obscured by finishing materials (Occlusion) or exist as fragmented meshes (Fragmentation). To address this, the visual preprocessing module transforms raw geometry into machine-readable "Semantic Units," as illustrated in Figure 5.

The process begins with Raw Geometry, where internal reinforcement and embed plates are hidden within the concrete mass (Left). The module first executes Material Separation, decomposing the object into distinct layers—separating opaque concrete from internal steel elements—to generate Decomposed Units (Middle). This step effectively creates "X-ray Views" that reveal previously occluded structures.

Subsequently, the system employs Axis-Aligned Bounding Box (AABB) Clustering to reassemble these scattered parts. By analyzing spatial proximity, fragmented meshes (e.g., separate bolts and plates) are merged into unified Semantic Units (Right). This final representation provides the MLLM with a clear, obstacle-free view of the target components, ensuring accurate identification.

Evidence-Based MLLM Verification

Finally, Google's Gemini 2.5 Pro is employed as the central Multimodal LLM (MLLM) verification engine,

evaluating LOD compliance by cross-referencing the logical checklist derived from the knowledge graph with the preprocessed visual evidence. To prevent logical circularity—where the same model validates its own reasoning—this framework utilizes a cross-model verification strategy: OpenAI's GPT-5.1 was employed during the initial phase to generate the logical rationales and Knowledge Graph, while Google's Gemini 2.5 Pro is used independently for the final inspection phase. Unlike traditional black-box AI models that output opaque probability scores, the proposed system generates traceable rationales for its decisions. For instance, instead of merely labeling an object as "LOD 350," the system explicitly validates whether specific criteria—such as the visual presence of "Lifting Anchor Assemblies"—are met.

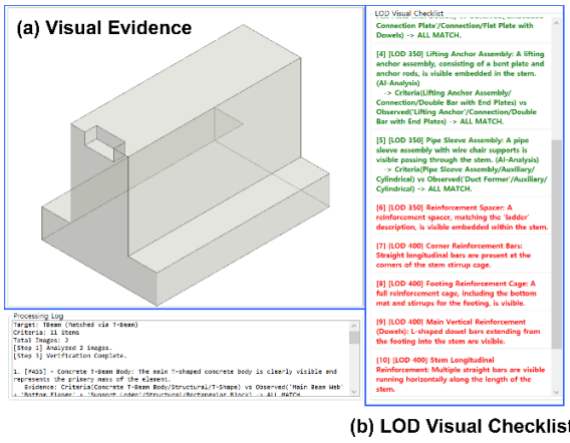


Figure 6 The integrated verification dashboard within Autodesk Revit.

To ensure seamless integration into the engineer's workflow, this verification logic is embedded directly within the BIM authoring environment via the Revit API, as illustrated in Figure 6. The system provides a real-time LOD Visual Checklist panel that displays the itemized inspection results derived from the knowledge graph. Compliant items are intuitively highlighted in green with an "ALL MATCH" tag, allowing users to instantly verify the satisfaction of specific requirements. This interactive feedback mechanism transforms the framework from a theoretical model into a practical decision-support tool, offering granular transparency beyond simple binary classifications.

Experimental Design and Validation Strategy

This section details the experimental framework established to assess the feasibility and performance of the proposed MLLM based LOD evaluation system. The composition of the dataset and the comparative scenarios designed to verify the impact of knowledge graph integration are described below.

Experimental Setup and Dataset

Dataset Composition

To evaluate the system's versatility across diverse construction disciplines, this study utilized the Level of Geometry (LOG) dataset (Abualdenien & Borrmann., 2022), a widely recognized benchmark in LOD research. The dataset encompasses a broad range of BIM families from structural, architectural, and MEP domains.

For the experimental scope, 31 representative families were selected, covering four LOD levels (LOD 200, 300, 350, and 400). This selection resulted in a total of 124 verification cases. These objects were specifically chosen to test the system's robustness against geometric complexities, including intricate internal components such as reinforcement cages, anchor bolts, and embed plates, which are rigorously modeled to comply with high-LOD standards.

Comparative Scenarios

To quantitatively measure the contribution of the proposed methodology, we conducted a comparative analysis between two distinct scenarios:

Baseline (Vanilla MLLM): This scenario represents the conventional verification method where raw text requirements from standard LOD Specifications are directly input into the MLLM without semantic processing. Under this baseline condition, an average of 4 criteria were applied per level, totaling 496 verification items. These criteria relied primarily on text-based descriptions, limiting the verification to abstract specifications rather than quantitative geometric analysis.

Proposed Method (Proposed KG+MLLM Approach): This scenario utilizes the Bridge knowledge graph constructed in this study. It applies the "deep standards checklist"—derived from the hybrid ontology learning pipeline—along with semantic visual preprocessing. This approach automatically generated 253 Concept Nodes and 626 Detailed Features. Consequently, the proposed method provides approximately 5.1 times more data points than the baseline, significantly enhancing the granularity of verification by grounding the process in actual model geometry.

Implementation Details

Model Implementation Details All experiments were conducted using the Gemini 2.5 Pro model via the API. This model was selected for its superior multimodal reasoning capabilities and extended context window, which are essential for processing the complex interplay between visual inputs (semantic units) and textual logic (knowledge graph queries).

Evaluation Metrics and Ground Truth

The quantitative evaluation employed confusion matrix metrics (TP, TN, FP, FN), determined based on the Cumulative Nature of LOD. This principle dictates that geometric features present in a lower LOD must persist in higher LOD levels.

To ensure reliability, the Ground Truth (GT) for comparison was established through a hybrid validation protocol. This approach utilized GPT-5.1 for initial automated verification, followed by rigorous manual inspection by researchers to finalize the dataset. The specific metrics are defined as follows:

True Positive (TP): System correctly classifies a Valid Model (e.g., Model 400) as Pass.

True Negative (TN): System correctly classifies an Invalid Model (e.g., Model 300 claiming to be 400) as Fail.

False Positive (FP): System incorrectly classifies an Invalid Model as Pass.

False Negative (FN): System incorrectly classifies a Valid Model as Fail.

Results and Discussion

Calibration: Threshold Sensitivity Analysis

Before performing the comparative analysis, a calibration experiment was conducted to determine the optimal "Decision Boundary" for the MLLM-based verification. Unlike rule-based systems, vision-based AI verification involves semantic ambiguity due to occlusion or resolution limits. Therefore, selecting an appropriate pass rate threshold is critical to balance sensitivity (Recall) and specificity (Precision).

Table 1: Sensitivity Analysis of Feature Emergence LOD Thresholds

Threshold	Self-Check (TP Rate)	Rejection (TN Rate)	Balance Score (TP+TN)
20%	99.2%	64.1%	1.63
40%	94.3%	85.3%	1.80 (Max)
60%	85.4%	93.5%	1.79
80%	65.9%	98.4%	1.64
100%	50.4%	98.9%	1.49

Table 1 presents the system's performance across pass rate thresholds ranging from 20% to 100%. The "Self-Check" score indicates the probability of a valid model passing its own standard, while the "Rejection" score indicates the probability of a lower-LOD model failing a higher standard.

The analysis reveals that the 40% threshold serves as the optimal operating point, maximizing the Balance Score (1.80). At this threshold, the system correctly identifies 94.3% of valid models while successfully rejecting 85.3% of invalid claims. Conversely, stricter thresholds (80-100%) caused a sharp decline in Recall (down to 50.4%), confirming that requiring a 100% feature match is impractical due to visual occlusion in 2D renders.

Consequently, this study adopts 40% as the standard threshold for the final comparative analysis.

Comparative Performance Analysis

Table 2 presents the ablation study results under the calibrated 40% threshold across four conditions: Vanilla MLLM (baseline), KG Only, Cluster IMG Only, and KG+Cluster IMG (proposed).

The Vanilla baseline exhibited a Recall of only 21.61% with 243 False Negatives, confirming that unstructured LOD descriptions cause the MLLM to default to negative predictions. Its Precision of 100% (zero FPs) reveals an extreme conservative bias rather than genuine accuracy.

Introducing semantic decomposition alone (KG Only) raised Recall to 51.61% by breaking abstract criteria into ~5.1 concrete inspection items, but plateaued at 67.34% accuracy due to insufficient visual resolution from a single overview image. Conversely, providing cluster images without structured criteria (Cluster IMG Only) yielded the worst performance (Recall: 17.74%), even below the baseline—demonstrating that detailed visual inputs without a semantic framework for interpretation lead to near-systematic rejection.

The proposed KG+Cluster IMG framework achieved the highest performance across all metrics (Accuracy: 83.47%, F1: 86.15%, Recall: 82.26%), confirming a complementary synergy: the Knowledge Graph provides what to inspect, while cluster images provide where to

Table 2: Performance Comparison between Baseline and Proposed Framework

Metric	Vanilla MLLM	KG Only	Cluster IMG Only	KG+Cluster IMG
Accuracy	51.01%	67.34%	48.19%	83.47%
F1 Score	35.54%	66.39%	29.97%	86.15%
Precision	100.00%	93.02%	96.49%	90.43%
Recall	21.61%	51.61%	17.74%	82.26%
True Positive (TP)	67	160	55	255
True Negative (TN)	186	174	184	159
False Positive (FP)	0	12	2	27
False Negative (FN)	243	150	255	55

inspect. Neither alone is sufficient, but their combination boosts Recall by 30.65 percentage points over KG Only.

The ablation results further reveal that FPs originate primarily from semantic decomposition rather than visual input: the KG-Only case produced 12 FPs, compared with only 2 for the Cluster-IMG-Only case. The combined framework yields 27 FPs, representing a favorable trade-off, as it recovers 188 TPs at a 1:7 TP-TP ratio. This balance is operationally desirable in engineering quality assurance, where undetected deficiencies (FN) pose substantially greater risk than over-acceptance (FP).

Overall, the F1 Score surged from 35.54% to 86.15%, signifying a qualitative leap from unusable baseline performance to engineering-deployable reliability. The ablation confirms that automated LOD verification requires both structured semantic criteria and spatially decomposed visual evidence.

Conclusion

This study addressed the critical challenges in automated BIM Quality Assurance—specifically the "Semantic Gap" arising from linguistic ambiguities in LOD specifications and the "Geometric Complexity" caused by fragmentation and occlusion in 3D models. To bridge these gaps, we proposed a novel hybrid assessment framework that integrates a Bridge knowledge graph with a Semantic Geometric Preprocessing pipeline, validated through a MLLM (Gemini 2.5 Pro).

The experimental results quantitatively demonstrate the superiority of the proposed "Deep Standards" approach over conventional text-based methods. By translating abstract human requirements into system-specific visual inspection items, our framework achieved an overall accuracy of 83.47%, marking a significant improvement of 32.46% compared to the baseline. Notably, the recall rate increased dramatically from 21.61% to 82.26%, proving that the integration of the knowledge graph effectively mitigates the issue of missing feature detection inherent in traditional approaches. A four-condition ablation study further revealed that neither semantic decomposition (KG Only: 67.34%) nor geometric preprocessing (Cluster IMG Only: 48.19%) alone achieves comparable performance—confirming that the two components exhibit a complementary synergy, where the knowledge graph provides "what to inspect" and cluster images provide "where to inspect." The sensitivity analysis further confirmed that a feature Emergence threshold of 40% offers the optimal balance between precision and recall, ensuring robust performance even with complex geometric variations.

The qualitative contribution of this research lies in its transition from "Black-box" AI to an "Explainable" verification system. Unlike purely vision-based deep learning models that require massive labeled datasets, our system generates traceable rationales by cross-referencing logical criteria with semantic visual evidence. The implementation of "Semantic Units"—reassembled from fragmented meshes and visualized through material-based layer separation—enables the MLLM to accurately identify internal components previously obscured by occlusion.

In conclusion, this framework successfully automates the translation of "Human Standards" into machine-interpretable "Deep Standards," providing a scalable and transparent solution for LOD verification. Future work will focus on expanding the dataset through systematic LOD augmentation—selectively adding or removing sub-components from existing high-fidelity families to generate boundary-condition models at each LOD transition—thereby enabling more rigorous evaluation of generalizability. Additionally, the "Visual Vocabulary" of the knowledge graph will be extended to cover a broader range of construction disciplines, and real-time inference speed will be optimized for large-scale infrastructure projects. This research paves the way for the next generation of intelligent BIM auditing tools, where domain knowledge and visual reasoning operate in synergy.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-23324132) and the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (RS-2024-00407028).

References

Abualdenien, J. and Borrmann, A. (2022) ‘Ensemble-learning approach for the classification of levels of geometry (LOG) of building elements’, Advanced Engineering Informatics, 51, p. 101497.

F.h, A., A, A., V.v, T. and J.h.m, T. (2025) ‘BIM ontology for information management (BIM-OIM)’, Journal of Building Engineering, 107, p. 112762.

Nahri, I., Pinquié, R., Véron, P., Bus, N. and Thorel, M. (2025) ‘Extracting structured requirements from unstructured building technical specifications for building information modeling’, arXiv.

Utkucu, D., Ying, H., Wang, Z. and Sacks, R. (2024) ‘Classification of architectural and MEP BIM objects for building performance evaluation’, Advanced Engineering Informatics, 61, p. 102503.

Wang, X., Yue, Q. and Liu, X. (2025) ‘Crack image classification and information extraction in steel bridges using multimodal large language models’, Automation in Construction, 171, p. 105995.

Wu, J. and Zhang, J. (2019) ‘New automated BIM object classification method to support BIM interoperability’, Journal of Computing in Civil Engineering, 33(5), p. 04019033.

Yang, M., Hei, X., Meng, H., Chen, K., Tong, X., Li, Y. and Zhao, Q. (2026) ‘Lightweight framework for IFC element classification using multi-view and heterogeneous graph with decision-level fusion’, Automation in Construction, 181, p. 106660.

Yin, M., Tang, L., Webster, C., Xu, S., Li, X. and Ying, H. (2023) ‘An ontology-aided, natural language-based approach for multi-constraint BIM model querying’, Journal of Building Engineering, 76, p. 107066.

Yu, Y.S., Kim, S.H., Lee, W.B. and Koo, B.S. (2023) ‘Ensemble-based deep learning approach for performance improvement of BIM element classification’, KSCE Journal of Civil Engineering, 27(5), pp. 1898–1915.

Zeng, Z., Goo, J.M., Wang, X., Chi, B., Wang, M. and Boehm, J. (2024) ‘Zero-shot building age classification from facade image using GPT-4’, arXiv.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATED P&ID COMPLIANCE CHECKING VIA LLM-BASED CODE GENERATION

Heejun Youn[1], Miyoung Uhm[1], Juho Han[1], Justin S. Lee[1], H. David Jeong[2], and Ghang Lee[1,3]
[1]Yonsei University, Seoul, Korea, Republic of South Korea
[2]Texas A&M University, College Station, TX, USA
[3]Technical University of Munich, Munich, Germany

Abstract

Manual P&ID review is time-consuming, while existing image-based or hard-coded automation methods lack efficiency. We propose a hybrid framework that extracts P&ID data from DXF files and converts natural language rules into FOL-based Python verification modules. By restricting the LLM to code generation, the framework ensures deterministic rule verification via formal logic. Evaluated on four industrial drawings (140 objects, 587 checks), our approach achieves a 97.5% F1-score, compared to 55.9% for the image-based LLM baseline. This enables non-experts to automate design, conditional, and engineering rules, advancing QA in digital twin environments.

Introduction

Piping and Instrumentation Diagrams (P&IDs) are critical engineering documents in plant facility design, containing information on equipment, piping systems, instrumentation, and process flows. Quality assurance of P&IDs is essential for plant safety, operational efficiency, and compliance with industry standards. However, the current review process still relies heavily on experienced engineers who must manually cross-reference thousands of drawings against multiple design specifications. This process is labor-intensive and vulnerable to human errors, particularly when dealing with complex and project-specific design rules.

The recent emergence of Large Language Models (LLMs) has opened new possibilities for bridging the gap between natural language specifications and executable code. Yet the direct application of LLMs for P&ID verification raises serious concerns: hallucinations during inference and the non-deterministic nature of outputs pose significant risks in safety-critical engineering contexts where reproducibility and accuracy are required. These limitations highlight the need for a hybrid approach that leverages the flexibility of LLMs for code generation while delegating the actual verification to deterministic reasoning engines.

Existing approaches to automated P&ID quality assurance fall into three broad categories, each with notable limitations. Image-based methods using computer vision achieve high symbol recognition accuracy but cannot verify compliance with semantic design rules and are sensitive to drawing quality (Alimin et al., 2025; Kim et al., 2025). Rule-based systems provide deterministic validation but require substantial expert effort for rule encoding and lack flexibility when adapting to project-specific requirements (Balhorn et al., 2025). Machine learning approaches such as Transformer-based semantic analysis show promise on large datasets but raise concerns about interpretability and portability across heterogeneous engineering environments (Gómez-Vega et al., 2025). A detailed review of these studies is provided in Section 2.

To address these gaps, this study proposes an automated P&ID verification framework that combines DXF-based data extraction with First-Order Logic (FOL)-based rule verification, restricting the role of LLMs to code and parser generation rather than direct validation. The core contributions are as follows:

1) We present a hybrid verification framework that extracts engineering objects directly from DXF-format CAD files through layer analysis and pattern recognition and transforms natural language design rules into executable Python verification functions via FOL formalization and LLM-assisted code generation. By confining LLM involvement to the code generation phase, the framework ensures that safety-critical verification outcomes remain deterministic and reproducible.
2) We formalize nine categories of Annotation Rules (AR) and Basic Design Rules (BR) into FOL expressions that serve as unambiguous, machine-verifiable specifications. This neuro-symbolic approach enables users with limited coding expertise to modularize complex design rules without manual hard-coding.
3) We validate the framework on four industrial P&ID drawings comprising 140 engineering objects and 587 rule checks. The proposed approach achieves an overall extraction F1 of 97.5% with a 100% rule compliance pass rate, compared to 55.9% F1 for the image-based LLM baseline. Equipment and Line extraction maintain 100% F1 across all four drawings without parser modification, demonstrating generalization beyond a single test case.

Related Work

Rule-Based Verification Systems

Schulze Balhorn et al. (2025) proposed a systematic approach that converts P&IDs into DEXPI-based structured data or graph representations to detect design rule violations through explicit pattern matching. Their method demonstrated clear advantages in identifying specific violation patterns and maintaining compatibility with industry standards but requires substantial effort for rule definition and lacks flexibility in processing natural language specifications. Additionally, scalability is limited when adapting to diverse project requirements.

Machine Learning-Based Approaches

Gómez-Vega et al. (2025) introduced Smart P&ID, which extracts text and attribute data from databases and CSV files, and transforms piping information into token sequences processed by Transformer-based models such as DistilBERT. Their approach showed high accuracy in detecting semantic inconsistencies in large datasets and demonstrated promising adaptability to rule changes. The proposed method faces challenges in interoperability and portability to other engineering environments, as it is heavily dependent on specific Smart PID data structures and does not fully account for legacy CAD systems or image-oriented PDF formats commonly used in the field.

Image-Based Detection Methods

Alimin et al. (2025) employed computer vision techniques including YOLO, Faster R-CNN, and SIFT/ORB algorithms for symbol detection and geometric verification in P&IDs. While achieving over 90% recognition accuracy in symbol identification, these methods are limited to detecting visual inconsistencies and cannot verify compliance with semantic design rules. Furthermore, the performance of these approaches is highly sensitive to image quality and resolution, limiting their applicability in real-world scenarios where drawing formats and quality vary. Beyond these three categories, Additional related work in DEXPI-based interoperability, graph-based P&ID analysis, and digital twin-integrated validation will be addressed in the journal extension of this study.

Methodology

This study develops an automated rule verification system for P&ID drawings through a two-phase approach. First, we conduct a comparative test to evaluate data extraction performance between image-based LLM approaches and DXF-based direct parsing. Second, based on the extraction results, we implement a FOL-based rule verification framework that transforms natural language design specifications into executable verification modules. This section describes experimental design, data extraction methodologies, and rule verification architecture.

Rule Set Composition

P&ID design review rules can be categorized as follows:

- Annotation Rules (AR): Rules related to notation and compliance with international standards.
- Basic Design Rules (BR): Rules related to the presence of essential design information.
- Conditional Rules (CR): Rules related to conditional design requirements.
- Design Rules (DR): Rules related to compound design principles.
- Engineering Rule (ER): Rules related to engineering logic and calculations.

These rules verify the presence or absence of required information elements and check basic formatting compliance, without involving complex conditional logic, engineering calculations, or cross-drawing relationships. Since this study focuses on establishing baseline extraction and verification capabilities for AR and BR rules, selecting a drawing that comprehensively tests these rule categories provides the most robust validation of the proposed framework's fundamental capabilities.

For the test, nine design rules were implemented from a total of eleven rules defined across the AR and BR categories, excluding rules that involve inter-rule dependencies. The AR rules (AR-1 to AR-5) verify drawing number uniqueness, entity identification schemes, pipe nominal diameter, pipe number, and material class. The BR rules (BR-1 to BR-4) verify name,

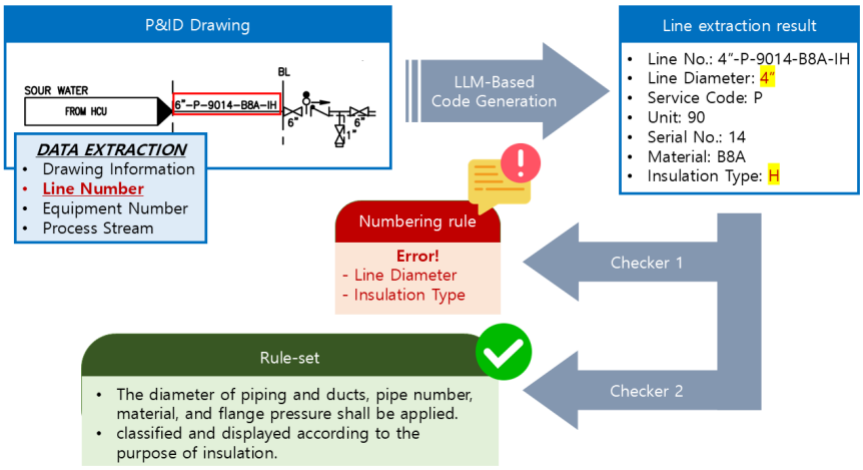


Figure 1. P&ID Compliance Checking via LLM-based Code Generation Framework

specification, type, and references to connected drawings (Table 1).

Table 1. Selected AR and BR rules

Class	Code	No.	Keyword
Annotation	AR	01	Drawing Number
	AR	02	Object ID
	AR	03	Line Diameter
	AR	04	Line Number
	AR	05	Line Material
	AR	06	Valve Failure mode
Basic Design	BR	01	Equipment ID
	BR	02	Equipment Spec.
	BR	03	Equipment Type
	BR	04	Process Flow
	BR	05	Manhole/Tray

These nine rules are applied to all extracted entities, including equipment, piping, instruments, and connected drawings.

Ground truth annotations were manually created. All equipment tags, line numbers, service codes, material classes, and instrumentation tags were extracted and cross-verified against the original design specification documents. Each extracted item is classified as True Positive (TP), False Positive (FP), or False Negative (FN) against the ground truth. Precision, Recall, and F1 Score are computed per category and macro-averaged across categories. These metrics are employed not as machine learning performance indicators but as standard information retrieval measures to quantify extraction completeness and correctness under a unified evaluation framework for both approaches.

P&ID Drawing

Seven P&ID drawings from an industrial Sour Water Stripper unit (A39002-500-A 101 through A107) were available for this study. Drawings A101–A103 are Symbol & Legend sheets defining project-wide naming conventions and serve as the reference for numbering rule verification. The remaining four drawings (A104–A107) are P&IDs, each containing equipment, piping lines, and process stream information subject to AR and BR rules (Table 2).

Table 2. Reviewable rules in each P&ID drawing

P&ID No.	Class	Adaptable Rules
A39002-500-A 101	Standard	1
A39002-500-A 102	Standard	1
A39002-500-A 103	Standard	1
A39002-500-A 104	Drawing	11
A39002-500-A 105	Drawing	9
A39002-500-A 106	Drawing	10
A39002-500-A 107	Drawing	11

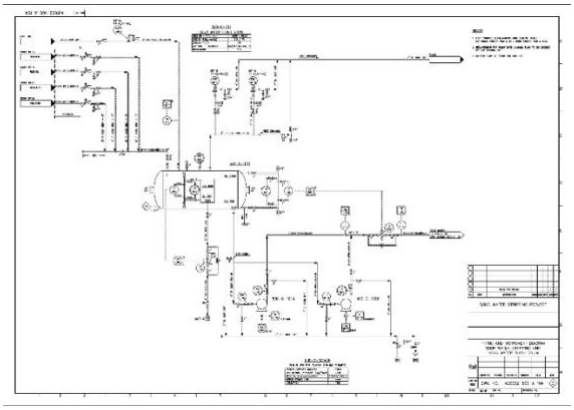


Figure 2. A39002-500-A 104 P&ID drawing

Among these, Drawing A104 (Sour Water Flash Drum) was obtained in both PDF and DXF formats to enable a fair comparison between the image-based LLM extraction (Approach A) and DXF-based parsing (Approach B). Drawings A105–A107 were obtained in DXF format only and were used together with A104 to evaluate the generalization of the proposed framework across drawings of varying complexity. The four P&IDs collectively contain 180 items

Table 3. Test data summary across four drawings

P&ID No.	Drawing Info	Equipment	Lines	Process Streams	Total
104	10	2	26	7	45
105	10	3	19	5	37
106	10	4	22	6	42
107	10	5	34	7	56
Total	40	14	101	25	180

Experiments

Approach A: Image-based LLM Extraction

The first approach utilizes a LLM equipped with vision-language capabilities to directly extract design information from P&ID drawings rendered as PDF images. This approach represents the prevalent image-based LLM methodology in recent automated P&ID interpretation research and serves as the baseline for comparing the performance of the proposed DXF-based automated review framework.

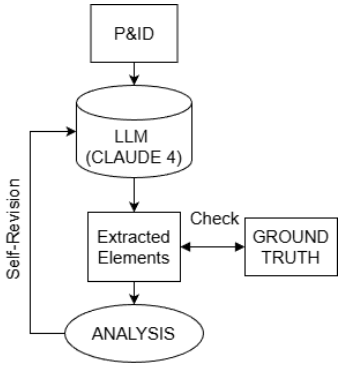


Figure 3. The flow of Image-based LLM Extraction

In this methodology (Figure 3), P&ID drawings are converted into high-resolution images and used as the primary input. Without the aid of dedicated CAD parsing or rule-based engines, the model simultaneously performs text and symbol recognition to extract engineering data, such as equipment, piping, and instrumentation. The interpretation relies solely on the model's visual reasoning and internal domain knowledge. To ensure structured outputs, a specific prompt restricts the response format to JSON, explicitly defining key data fields—including drawing numbers, equipment tags, and piping specifications—required for Annotation Review (AR) and Basic Review (BR).

Claude Sonnet 4.5 (Anthropic, 2025) was employed for this extraction task as a widely accessible vision-enabled LLM. The Opus model was excluded due to prohibitive token costs and significantly slower inference speeds for large image inputs. Since Approach A aims to establish a representative baseline rather than to maximize LLM performance, Sonnet 4.5 was deemed sufficient for this purpose. The model processes the drawing image and a structured prompt concurrently to return engineering data in a predefined JSON schema. The extracted data are then evaluated against the ground truth for a quantitative comparison with the proposed method.

However, this approach possesses inherent structural limitations. Since drawing legends cannot be explicitly fixed as external knowledge, consistency in symbol interpretation is not guaranteed. Furthermore, it lacks the capability to verify complex logical structures, such as Conditional Rules (CR) or Detailed Design Rules (DR). Most notably, the occurrence of hallucinations during the inference process and the non-deterministic nature of the outputs pose significant challenges for safety-critical design reviews that demand high reliability. By establishing Approach A as a baseline, this study aims to highlight these deficiencies and demonstrate the superior reliability of the proposed DXF-based framework utilizing FOL.

Approach B: DXF-Based Direct Parsing

The second approach leverages the original CAD file format (DXF) to extract engineering information directly from vector data without image conversion (Figure 4).

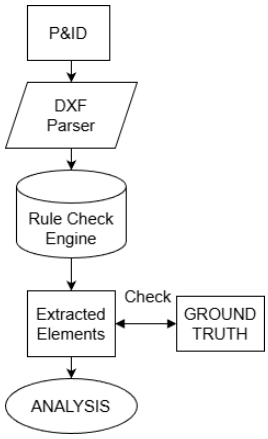


Figure 4. DXF-based data extraction pipeline

The DXF-based extraction approach offers several critical advantages over conventional image-based methods in terms of data integrity and analytical precision. By eliminating degradation from image conversion and vision interpretation errors, this method ensures lossless information preservation and facilitates spatial relationship analysis. Furthermore, it allows for direct access to the CAD organizational structure, including layer-based filtering, block definitions, and embedded metadata, thereby enhancing attribute richness. Beyond its technical accuracy, the DXF approach is highly scalable as processing time remains independent of image resolution. Most importantly, it yields deterministic results, ensuring reproducible extraction without the stochastic variation typically associated with image-based processing.

The role of LLMs in the framework is strictly limited to code and parser generation, not direct rule verification. This architectural choice mitigates hallucination risks while leveraging LLM flexibility.

The DXF-based data extraction process is structured into three distinct stages to ensure high fidelity and structural integrity.

- Stage 1: Entity Parsing. The ezdxf library for Python is employed to parse the DXF file, extracting all text-bearing entities while preserving their original Cartesian coordinates and layer associations. This stage captures raw textual content directly from the CAD environment without information loss.
- Stage 2: Pattern-Based Classification. Extracted text elements are categorized using domain-specific regular expressions derived from project-specific naming conventions formulated in accordance with identification of equipment tags, line numbers, and other standardized engineering identifiers.
- Stage 3: Coordinate-Based Attribute Association. Spatial analysis links equipment tags with their corresponding specifications through a deterministic search algorithm operating within a defined radius (default: 150 DXF units) around each identified tag. To ensure association accuracy, invalid patterns—including instrument tags, specification table headers, and control keywords—are filtered through rigorous exclusion rules. This mitigates data misinterpretation risks arising from spatial proximity of unrelated elements.

The verification engine consists of two checking modules.

Checker 1(Numbering Rules) verifies the formal integrity of the extracted information based on the numbering scheme defined in the Symbol & Legend sheet (A39002-500-A 101-103) of the P&ID drawing (Figure 5).

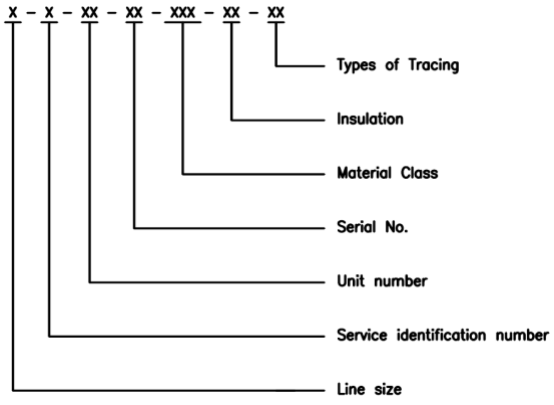


Figure 5. Line Identification

Checker 2 (AR & BR): This module examines whether mandatory information required by each rule is present and evaluates attribute completeness. Inspired by (Pan et al. 2023) neuro-symbolic framework, we adopt a similar three-stage paradigm—formulation, reasoning, interpretation—but adapt it specifically for P&ID verification. To establish unambiguous verification criteria, design rules are formalized using FOL.

FOL is essential for P&ID rule verification for three reasons. First, it eliminates ambiguity in natural language rules through precise semantic expression. For instance, the vague statement "Appropriate insulation shall be used for high-temperature piping." becomes:

$\forall x\,(\text{Pipe}(x) \wedge \text{HighTemperature}(x) \geq 65 \rightarrow \text{Insulated}(x)$.

Second, FOL systematically represents complex conditional rules and nested logical structures, enabling integration of multi-condition design rules (e.g., rules involving operating pressure, temperature, material, and fluid properties) into unified logical expressions. Third, formally expressed rules enable mechanical verification through automated theorem provers or model checkers, allowing pre-validation of logical consistency and conflict detection among rules.

Experimental Result

LLM Extraction Method Performance

To maximize extraction performance, the methodology underwent three iterations with progressive improvements:

Table 4. F1 Score Comparison by Data Categories

Category`	Test 1	Test 2	Test 3
Drawing info	100%	100%	100%
Equipment	100%	100%	100%
Lines	48%	48%	48%
Process streams	71.4%	71.4%	71.4%
Overall	55.9%	55.9%	55.9%

Table 5. Iterative Refinement Process

Class	Revision Detail	Result
Iteration 1 (Baseline)	Direct extraction using the base prompt without post-processing.	F1 55.88
Iteration 2 (Post processing)	Applied automated post-processing includes text normalization and format verification against P&ID naming conventions.	F1 55.88
Iteration 3 (Error Pattern Correction)	Analyzed common error patterns from previous iterations and incorporated specific correction rules including size extraction, service code verification, and missing entity detection.	F1 55.88

The image-based approach demonstrated several systematic failures:

1. Text recognition errors: Misrecognition of characters in low-contrast regions

2. Spatial relationship inference: Difficulty in correctly associating spatially disconnected text elements with their corresponding equipment or piping symbols, leading to attribute misassignment.

3. Overlapping element confusion: Inability to disambiguate text when overlapping with graphical symbols, leader lines, or dense annotations, resulting in incomplete or erroneous extraction.

4. Implicit attribute loss: Failure to extract attributes not explicitly labeled in text form, such as connection relationships inferred from line continuity or material specifications derived from line style conventions.

These limitations align with fundamental challenges of image-based approaches documented in prior studies (Alimin et al., 2025), demonstrating the inherent trade-offs between vision model flexibility and extraction reliability.

DXF-based Method Performance

The proposed DXF-based extraction and rule verification pipeline was applied to all four process drawings (A104–A107). Table.6 presents the extraction accuracy compared with the ground truth across 180 items. Drawing Info, Equipment, and Lines achieved 100% F1 across all four drawings. The three unmatched items occurred exclusively in A107 Process Streams, where F1 dropped to 53.3% (Precision 50.0%, Recall 57.1%), while A104–A106 maintained 100% in all categories.

Table 6. Extraction accuracy compared with Ground Truth

Category	Ground Truth	Extracted	Matched	F1
Drawing Info	40	40	40	100%
Equipment	14	14	14	100%
Lines	101	101	101	100%
Process streams	25	26	22	86.3%
Total	180	181	177	97.5%

Error analysis of A107 Streams identified 4 true positives, 4 false positives, and 3 false negatives. All errors originated from streams lacking explicit FROM/TO boundary markers in the DXF source: (1) anchor text absence preventing stream detection (1 case), (2) label-boundary mismatching causing line number misreading or false extraction (2 cases), and (3) incorrect direction assignment due to ambiguous spatial layout (1 case). These are inherent limitations of text-based spatial matching, not parser logic defects.

Table 7 presents the rule verification results. Checker 1 performed 119 numbering-format checks and Checker 2 conducted 468 AR/BR rule checks; all 587 items passed with zero errors. Thirteen warnings were issued for material class codes not registered in the rule-set list and missing equipment attributes in the DXF source. These warnings indicate incomplete source data rather than rule violations and therefore do not affect the pass/fail determination.

Table 7. AR/BR rule verification results across four drawings

Rule	Keyword	Checked Items	Warning
AR-1	Drawing Number	4 Drawings	0
AR-2	Object ID	140 IDs	0
AR-3	Line Diameter	101 Lines	0
AR-4	Line Number	101 Lines	0
AR-5	Line Material	101 Lines	10
BR-1	Equipment ID	14 Equipment	0
BR-2	Equipment Spec.	14 Equipment	2
BR-3	Equipment Type	14 Equipment	1
BR-4	Process Flow	25 connections	0

The AR rules (AR-1 to AR-5) verify formal integrity of drawing numbers, entity identification schemes, and piping attributes including nominal diameter, line number format, and material class. The BR rules (BR-1 to BR-4) validate the completeness of equipment information and process flow references to connected drawings. All nine rules passed across the four drawings, confirming compliance with the project naming conventions defined in the Symbol & Legend sheets (A39002-500-A 101–103).

Analysis of LLM Extraction vs. DXF parsing

In the context of P&ID information extraction, LLM–based approach demonstrated effectiveness in identifying textual elements and equipment tags within drawings. However, several structural limitations were observed in the extraction of piping and process-related data. First, due to the nature of image-based inputs, text misdetection and information loss occurred depending on drawing resolution and visual noise. In addition, probabilistic inference during numerical data extraction introduced the possibility of incorrect outputs. Furthermore, frequent false detections were observed when spatially adjacent elements visually interfered with each other, resulting in mixed or conflated data. These issues can be attributed to the inherent sensitivity of vision-based algorithms and probabilistic models to complex physical layouts and visual variability within engineering drawings.

- Data Integrity: By eliminating image conversion, information loss and visual interpretation errors are fundamentally avoided.
- Geometric Precision: Direct utilization of numerical coordinate data enables sub-millimeter–level accuracy in spatial relationship analysis.
- Structural Accessibility: CAD-native metadata, including layer hierarchies, block definitions, and line types, allows for explicit identification of semantic relationships among design elements.
- Computational Efficiency and Reliability: The approach maintains consistent processing performance independent of image resolution and guarantees reproducibility through deterministic, non-probabilistic outcomes.

Nevertheless, the DXF-based approach has its own limitations. When explicit boundary markers such as FROM/TO anchor texts are absent from the source drawing, the coordinate-based stream detection relies on spatial heuristics that can misidentify stream boundaries or flow directions. This was observed in A107 Process Streams, where the absence of boundary markers led to line number misreading, false extraction, and direction assignment errors, resulting in an F1 of 53.3%. These errors are inherent to text-based spatial matching rather than defects in the parser logic, and they define the current boundary condition of the framework.

Consequently, for design rule checking in plant engineering where high reliability and strict reproducibility are critical, the DXF-based structural approach is considered to provide a more stable and trustworthy data processing foundation than an LLM-only model. Table 8 presents this comparison: on the same drawing (A104), Approach B achieved 100% F1 across all categories, whereas Approach A reached 55.9%. When Approach B was extended to four drawings without parser modification, Overall F1 remained at 97.5%, with the reduction attributable solely to process stream detection in A107.

Table 8. Performance and iteration comparison table to match ground truth

Class	Approach A (LLM)	Approach B (Code)	
Drawing	A104	A104	A104-107
Input Data	PDF	DXF	
Method	Vision LLM	DXF Parsing	
Iteration	3	1	
Drawing info F1	100%	100%	100%
Equipment F1	100%	100%	100%
Line F1	48%	100%	100%
Process F1	71.4%	100%	86.3%
Overall F1	55.9%	100%	97.5%

Discussion

The experimental results demonstrate that the proposed framework achieves high accuracy in extracting Equipment and Line information from DXF-format P&ID drawings, successfully passing all 587 design rule checks across four test drawings containing 140 engineering objects. Notably, the parser and rule engine developed for A104 were applied to A105–A107 without modification, and Equipment and Lines maintained 100% F1 across all drawings. This provides evidence that the DXF parsing algorithm and FOL-based rule verification engine can generalize effectively across real-world plant engineering drawings of varying complexity. By structuring high-level and complex rules from the P&ID rule definition document into formal logic (FOL), the proposed approach mitigates the hallucination issues frequently observed in standalone LLM-based methods and enables deterministic and reproducible verification results.

When LLMs were applied independently, they exhibited strong reasoning capabilities but failed to provide reliable verification outcomes due to persistent hallucination problems. In the plant engineering domain, where design quality is directly linked to safety incidents, probabilistic models alone are insufficient to guarantee the level of accuracy and consistency required for design verification. Notably, AR and BR rules that could not be reliably validated by standalone LLMs were verified with high accuracy through the proposed framework, which combines LLM-based code generation with symbolic logic verification. This result highlights the practical effectiveness of the proposed approach.

Compared to traditional manual drawing review processes, the proposed automated data extraction and rule verification system significantly reduces review time and effectively minimizes the risk of human error. By integrating the flexibility of LLM-based code generation with the deterministic reliability of formal logic-based verification, the proposed approach overcomes key limitations of existing automated P&ID quality verification studies.

The proposed framework addresses LLM hallucination issues through a carefully designed separation of concerns. LLMs are restricted to two specific roles: (1) transforming natural language rule specifications into structured FOL representations, and (2) generating Python verification code from these formal logic structures. The actual verification process is performed entirely by deterministic symbolic reasoning engines, eliminating the probabilistic uncertainty inherent in direct LLM-based verification. This architectural design ensures that while the system benefits from LLM flexibility in code generation, the safety-critical verification outcomes remain reproducible and deterministic.

However, the framework's current boundary condition was revealed in A107 Process Streams, where F1 dropped to 53.3%. All errors occurred in streams lacking explicit FROM/TO boundary markers, indicating that the coordinate-based spatial matching is sensitive to the completeness of anchor text in the source DXF. This limitation does not affect Equipment or Line extraction, which rely on self-contained tag patterns, but it suggests that process stream detection requires either enhanced spatial heuristics or supplementary metadata to achieve robustness comparable to other categories.

Conclusions

This study proposes an automated data extraction and design rule compliance verification system for P&ID drawings in the plant engineering domain. The proposed system parses DXF-format P&ID drawings to extract engineering objects such as Equipment, Lines, and Instruments as structured data and automatically verify compliance with design standards using FOL-based rule engine. Experimental validation on four P&ID drawings containing 140 engineering objects demonstrates that the system achieves 100% F1 for Equipment and Line extraction and 97.5% Overall F1 and successfully performs exhaustive verification of 587 design rule checks across nine categories of AR and BR rules.

The main contributions of this research are as follows. First, a DXF-based P&ID parsing algorithm was developed to automatically extract engineering objects through layer structure analysis and text pattern recognition. Second, an FOL-based rule verification engine was implemented to formally represent plant design standards and perform systematic compliance verification. Third, an end-to-end automation pipeline was established that integrates drawing input, data extraction, rule verification, and result report generation, thereby substantially reducing the manual review burden on design engineers.

This study acknowledges several limitations that warrant further investigation. First, although validation was extended to four P&ID drawings comprising 140 engineering objects, all drawings originate from a single industrial project, which limits conclusions about generalization performance across diverse projects, drawing conventions, and industrial contexts. Second, the current implementation addresses nine categories of AR and BR rules, whereas real-world projects require more

comprehensive coverage including equipment-specific requirements (e.g., manhole and tray installation specifications) and failure mode documentation standards. Future research will address these limitations through a phased approach:

- **Immediate Extension**: The scope of validation will be expanded to various industrial plant types, including petrochemical and power generation, and water treatment facilities. By securing real-world project drawings, the research will broaden the application of AR and BR rules to verify the system's generalization performance and derive industry-specific adaptation requirements.
- **Medium-term Goals**: Technical capabilities will be developed to support CRs, DRs and ERs. This involves advancing context-aware verification logic based on FOL expressions, integrating domain-specific calculation engines for pressure drop and heat balance verification, and establishing a cross-drawing consistency-checking mechanism.
- **Long-term Vision**: Achieving Digital Twin integration with real-time verification capabilities. This encompasses developing bidirectional synchronization between P&ID models and Digital Twin platforms, implementing continuous quality assurance during design iterations, and establishing automated compliance reporting aligned with international standards (ISO, ASME, API).

This study demonstrates the practical feasibility of automated P&ID quality verification in plant engineering contexts. The successful validation of AR and BR rules on industrial drawings, combined with the framework's user-centric design that enables non-programmers to define rules in natural language, suggests significant potential for industry adoption.

However, successful deployment in engineering organizations will require addressing several practical considerations: (1) integration with existing CAD workflows and document management systems, (2) establishment of verification protocols for LLM-generated code before production use, and (3) development of user training programs for effective rule specification and result interpretation.

Despite these implementation challenges, the adoption of automated verification systems is expected to deliver substantial benefits including reduced design review time (potentially 50-70% reduction based on pilot testing), improved consistency in rule application, and prevention of human errors in repetitive checking tasks. These improvements directly contribute to enhanced safety outcomes and economic efficiency in plant engineering projects, while freeing experienced engineers to focus on high-level design decisions and complex problem-solving activities.

In the future, this study lays the foundation for extending beyond single-project verification to automated quality assurance encompassing CRs, DRs and ERs. This extension will contribute to comprehensive quality assurance automation within digital twin environments, supporting broader digital transformation initiatives in plant engineering and operations.

Acknowledgments

This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (No. RS-2024-00407028) and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. RS-2025-25436231).

References

Balhorn, L. S., Seijsener, N., Dao, K., Kim, M., Goldstein, D. P., Driessen, G. H. M., & Schweidtmann, A. M. (2025). Rule-based autocorrection of Piping and Instrumentation Diagrams (P&IDs) on graphs (arXiv:2502.18493). arXiv. https://doi.org/10.48550/arXiv.2502.18493

Alimin, A. A., Goldstein, D. P., Balhorn, L. S., & Schweidtmann, A. M. (2025). Talking like Piping and Instrumentation Diagrams (P&IDs) (arXiv:2502.18928). arXiv. https://doi.org/10.48550/arXiv.2502.18928

Gómez-Vega, F. S., Acuña, O., Camargo, A. C., Jimenez, J. D., Galeano, S. M., Franco, I. E., Lozano, L. L., Vásquez, J., & Puertas, E. (2025). Enhancing consistency in piping and instrumentation diagrams using DistilBERT and smart PID systems. Systems and Soft Computing, 7, 200373. https://doi.org/10.1016/j.sasc.2025.200373

Pan, L., Albalak, A., Wang, X., & Wang, W. Y. (2023). Logic-LM: Empowering Large Language Models with Symbolic Solvers for Faithful Logical Reasoning (arXiv:2305.12295). arXiv. https://doi.org/10.48550/arXiv.2305.12295

Jianwu, L. W., Yew, L. S., On, L. K., Keong, T. C., Yuan Sheng, R. T., Sani, S. B., & Juan Agnes, T. H. (2024). Artificial intelligence-enabled evaluating for computer-aided drawings (AMCAD). International Journal of Mechanical Engineering Education, 52(1), 3–31. https://doi.org/10.1177/03064190231175231

Dzhusupova, R., Banotra, R., Bosch, J., & Olsson, H. H. (2023). Using artificial intelligence to find design errors in the engineering drawings. Journal of Software: Evolution and Process, 35(12), e2543. https://doi.org/10.1002/smr.2543

Kim, J.-B., Moon, Y., Han, S.-T., & Mun, D. (2025). Automated inspection of P&ID object recognition using deep learning. Scientific Reports, 15(1), 39031. https://doi.org/10.1038/s41598-025-25506-2

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

ACTOR-NETWORK THEORY FOR EVALUATING BIM-AI IMPLEMENTATION IN DESIGN FIRMS

Federica Crosato1, Alireza Borhani2, and Luciana Burdi2
[1]Iuav University of Venice, Venice, Italy, IT
fcrosato1@iuav.it
[2]Massachusetts Port Authority, Boston, Massachusetts, MA
aborhani@massport.com, lburdi@massport.com

Abstract

Artificial Intelligence (AI) is assuming a central role in the AECO industry, supporting automation, workflow optimization, and decision-making. Despite rapid technological advances, the use of AI in design practice remains underexplored, particularly regarding the sociotechnical dynamics that favor or hinder its adoption within BIM processes. This study introduces Actor-Network Theory (ANT) to the field of BIM-AI research, focusing on design firms. The goal is to develop an ANT-based method to guide researchers in analyzing how actor networks can be mobilized to support the implementation of AI in BIM design workflows.

Introduction

In recent years, Building Information Modeling (BIM) has become a central component of information management in the construction sector, supported by increasingly sophisticated software technologies (Eastman et al., 2008). Understanding how AI may influence BIM processes requires examining the technological maturity of design firms and their ability to select, integrate, and coordinate heterogeneous digital tools. In particular, mapping internal information flows is essential for assessing how new technologies modify or optimize the circulation of data across departments, operational staff, clients, and external consultants (Cavka et al., 2015). Although the literature recognizes the strategic importance of BIM, its effectiveness depends on the clear definition of project objectives and BIM uses, which guide responsibilities and workflow organization (Raj et al., 2025). At the same time, organizational change is shaped by external environmental, economic, and policy conditions (Pettigrew, 1987), and the availability of standards, protocols, and best practices remains a key enabler of BIM implementation (Hochscheid & Halin, 2018), supporting more advanced environmental and performance analyses (Alnaser et al., 2024). AI tools encompass a broad domain of techniques dedicated to automation, with machine learning (ML) and deep learning (DL) representing the most widely adopted approaches (Alves et al., 2025). Despite their potential, the effective integration of AI in the AEC sector depends on professionals' perceptions, competencies, and attitudes (Albukhari et al., 2025). As highlighted in a recent AIA report (2024), AI is now used across a wide spectrum of activities - from chatbots supporting operational tasks to the generation of visual content in early design phases (Xu et al., 2024), as well as linguistic analysis, text revision, and automated transcription of meetings and workplace communications. AI therefore influences both the technological dimension and the organizational processes of design firms (Wang et al., 2025), making it essential to identify the factors that enable or hinder its adoption. Despite growing interest in AI applied to BIM, the existing literature remains fragmented. Current studies tend to focus on specific applications (Al-Muntaser et al., 2024), project-based case studies (Attia, 2025), perceived risks (Bolpagni et al., 2021), or benefits (Chen et al., 2024; Alves et al., 2025; Albukhari, 2025). However, a qualitative sociotechnical perspective capable of systematically examining the drivers of BIM-AI integration - and how these are translated into professional practice - has not yet been developed. In particular, a holistic framework that guides researchers in studying the BIM-AI domain is still lacking. Actor-Network Theory (ANT) offers a promising lens for addressing this gap. ANT conceptualizes both "non-human" actors (artefacts, processes, standards, regulations) and "human" actors (roles, professionals, institutions) as entities capable of shaping action within a network (Latour, 2005). This perspective enables researchers to examine how technology implementation is translated and adapted within specific institutional, organizational, or project contexts. Previous studies have applied ANT to analyze BIM maturity development (Edirisinghe, 2021), planning practices (Rydin, 2013), BIM implementation in large organizations (Lindblad & Vass, 2015), municipal land-allocation processes involving BIM, GIS, and Digital Twins (Hamarashid, 2021), sustainability certification networks (Frank et al., 2022), and collaboration in industrialized housing (London & Pablo, 2017). Yet, despite these contributions, the perspective of design firms - particularly regarding the integration of AI into everyday BIM-based practice - remains underexplored. This study addresses this gap by investigating and systematizing the key components influencing the integration of AI into BIM processes in design firms. By adopting a structured methodology, the research aims to clarify how technological, organizational, and environmental conditions shape AI

adoption. More specifically, the literature review conducted in this study seeks to: (1) identify existing frameworks for BIM implementation; (2) examine current applications of AI in design; (3) extract the key factors relevant to design firms; and (4) develop an ANT-based diagram to guide future data collection and analysis.

Methodology

This paper presents the initial phase of a broader research project that investigates BIM-AI implementation in design firms from a comparative international perspective. The study is structured into two sequential phases. The first phase - reported in this article - consists of a systematic literature review aimed at identifying existing frameworks, methodological approaches, and thematic areas relevant to BIM-AI integration. Guided by this objective, the present contribution addresses two research questions:

How is BIM implementation conceptualized within design firms?

What are the key components for evaluating BIM-AI implementation in design?

To answer these questions, the study adopts a qualitative methodology grounded in a structured literature review articulated into four phases (Figure 1). The first phase involved planning and defining the review protocol, including research questions, keywords, and inclusion criteria:

- Types: articles and conference papers;
- Period: 2010-2025;
- Language: English;
- Open access;
- Keywords: Table 1.

Table 1: Keywords used in the literature reviews

"BIM implementation" OR "BIM adoption"	AND	"design firm" OR "design firms" OR "design organization" OR "design organizations" OR "SME"
"AI implementation" OR "AI adoption"	AND	"design firm" OR "design firms" OR "design organization" OR "design organizations" OR "SME"

Applying these criteria, 132 documents were initially identified. After screening for relevance, 18 publications focusing on BIM implementation in design firms and 23 addressing AI in design were selected for in-depth analysis. The subsequent sections examine the research methods and thematic areas emerging from these studies. Based on this analysis, the factors related to AI within BIM-based design processes are extracted, categorized, and used to support the construction of the actor-network that informs the ANT-based framework proposed in this research.



Figure 1: Synthesis of methodology

BIM implementation in design firms

BIM implementation is commonly defined as the set of organizational activities undertaken to prepare for, initiate, or enhance digital deliverables and related workflows (Succar & Kassem, 2015). As a consolidated methodology and a driver of innovation in the construction sector (Eastman et al., 2008), BIM has become a central component of contemporary design practice. However, its interpretation and application remain highly heterogeneous across countries and even among firms operating within the same market (El Masry & Chronéer, 2025). The reviewed studies confirm this variability, showing that BIM implementation in design firms has been examined through a wide range of methodological approaches (Table 2). A substantial portion of the literature adopts survey-based methods to assess readiness, perceived benefits, barriers, and organizational needs. Examples include studies on BIM usage in Hong Kong (Chan, 2014), BIM adoption in Indian architectural firms using the TOE framework (Ahuja et al., 2016), BIM-Lean integration and standardization practices (Bolpagni et al., 2017), knowledge and skills gaps (Abd Hamid et al., 2020), maturity factors in Italian firms (Troiani et al., 2020), and barriers and benefits in Algerian practices (Tehami & Seddiki, 2023). Large-scale surveys, such as Wang et al. (2022), extend this perspective by examining national trends and regulatory pressures through TOE, DOI, and time-series analyses involving over one thousand Chinese firms. Sadeh et al. (2021) further show how MSMEs perceive BIM benefits across operational and informational dimensions while facing persistent constraints related to expertise and interoperability. Other contributions adopt mixed-methods or qualitative approaches to explore organizational dynamics in greater depth. Liu et al. (2020) combine questionnaires and interviews to investigate micro-level BIM implementation in UK design firms, while Tong et al. (2021) develop an organizational readiness framework integrating the MIT90s model and BIM maturity literature. Action Research (AR) is used to examine BIM as an organizational transformation process, as in Arayici et al. (2011) for an architectural SME and in Hochscheid & Halin (2020) for French architecture firms. Additional studies explore collaboration between BIM and Lean experts (Davoudabadi et al., 2022), obstacles to BIM implementation in Middle Eastern contexts using Grounded Theory (Altassan et al., 2023), and multi-level workflow analysis through process mapping and Social Network Analysis (Bonanomi, 2019). Taken together, these contributions show that BIM implementation in design firms is not a uniform or linear process but a complex organizational transformation, typically examined in the early and intermediate stages of digital transition. The literature increasingly recognizes the value of socio-technical perspectives, often supported by mixed methods (Figure 2), to understand the interplay between practices, technologies, and organizational structures. Within this context, the introduction of Actor-Network Theory (ANT) represents an innovative and still

underexplored contribution, offering a framework capable of capturing the heterogeneous networks of human and non-human actors that shape BIM implementation in design firms.

Table 2: Research methods of design-firm studies

Research methods	Authors
Literature review + questionnaire	Chan (2014), Ahuja et al. (2016), Bolpagni et al. (2017), Abd Hamid et al. (2020), Troiani et al. (2020), Tehami & Seddiki (2023), AlBalkhy et al. (2023), Sadeh et al. (2021)
TOE framework + DOI	Wang et al. (2022)
Literature review + questionnaire + interviews	Liu et al. (2020)
Organizational readiness framework + MIT90s framework + BIM maturity literature + interviews	Tong et al. (2021)
Literature review + interviews	Travaglini et al. (2014)
AR + case study	Arayici et al. (2011)
AR + case study + interviews	Hochscheid & Halin (2018, 2020)
Secondary research + case study + interviews	Davoudabadi et al. (2022)
GT + interviews	Altassan et al. (2023)
Literature review	Bonanomi (2019)

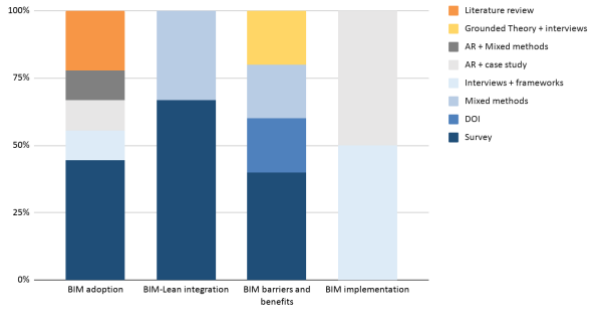


Figure 2: Thematic areas identified across the selected studies

AI application in design

The rapid growth of interest in AI-based solutions in recent years is largely driven by the widespread availability of tools such as ChatGPT, Claude, and Gemini, which have popularized AI among a broad audience despite being primarily individual, user-oriented systems. In contrast, organizations require AI solutions capable of operating within structured workflows, supporting collaborative environments, and integrating with existing information systems. Ma et al. (2025) highlight three organizational conditions that strongly influence AI adoption: collaboration with advanced technology providers, internal R&D capabilities, and continuous staff training. Similarly, Aiudi et al. (2025) show that AI reshapes SMEs' strategic decision-making by increasing decision speed and formalization, although financial, technical, and cultural constraints still limit full integration. These findings indicate that effective AI implementation requires formalized workflows, clear access and usage permissions, alignment with organizational objectives, appropriate technical enablers, human-centered considerations, and external support structures (Proietti, 2025). Within the design domain, AI is progressively shifting practice from intuition-based decision-making toward data-driven and performance-oriented processes. Applications include spatial planning, parametric modeling, generative design, and performance-based analysis, enabling designers to explore broader solution spaces while simultaneously addressing functional, structural, and environmental constraints (Albukhari, 2025). Recent studies highlight the growing influence of generative AI in early design phases, where it supports concept exploration, layout generation, and the rapid production of multiple alternatives, accelerating the translation of project requirements, site conditions, and user intentions into visual and formal proposals (Jang et al., 2025). BIM-based datasets provide a structured foundation for machine learning and deep learning models, supporting applications ranging from facility management and sustainability assessment to energy optimization, cost prediction, Heritage BIM, object-based modeling, and generative design (Alves et al., 2025). Empirical studies confirm this trend: AI-enabled BIM environments can assist designers in evaluating radiation exposure, sun hours, shadow performance, LEED-related indicators, and energy consumption at early stages (Al-Muntaser & Ashur, 2024). Despite these advancements, AI applications remain unevenly distributed across design phases, with design development and construction documentation still comparatively underexplored (Jang et al., 2025). This diversification of AI applications is illustrated in Figure 3, which synthesizes the breadth of AI research across the AEC and design domains in selected studies. The largest segments correspond to consolidated areas such as AI for decision-making and management, AI-BIM for performance simulation, and generative design workflows, confirming the centrality of AI in supporting analytical and early-stage design tasks. Other segments represent emerging or specialized domains - including bio-inspired computational design, knowledge-driven generative modeling - indicating a progressive expansion of AI beyond purely technical tasks toward cognitive, pedagogical, and biomimetic applications. The presence of categories such as AI for construction processes, computer vision for the built environment, and AI for BIM lifecycle management demonstrates that AI is increasingly being applied across the entire project lifecycle.

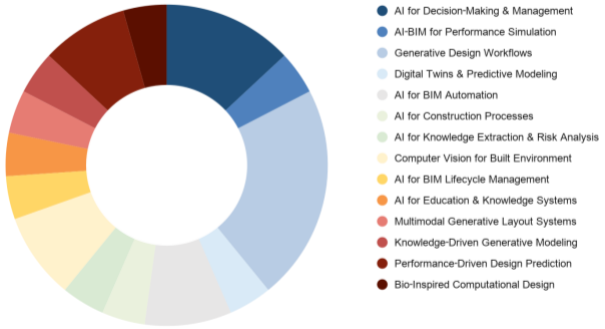


Figure 3: AI Application in design in selected studies

BIM-AI key factors

To compare how the selected studies address the factors influencing BIM-AI implementation, each publication was coded using a four-level scale reflecting the degree of thematic relevance: primary focus (3), substantially discussed (2), mentioned (1), and not addressed (0). This coding system enables a comparative reading of the literature by showing not only whether a theme appears, but also the extent to which it is developed within each study. Figure 4 synthesizes these results, illustrating the relative prominence of the key factors across the corpus and highlighting the thematic patterns that characterize current research on BIM-AI integration in design firms.Technology stands out as the most emphasized factor, with the highest proportion of "Primary" coding (60%), confirming that technological readiness - particularly in terms of tools, systems, and digital infrastructures - remains the dominant concern in the literature. Closely related factors such as interoperability-compatibility- complexity and information management also receive high levels of attention, reflecting persistent challenges associated with fragmented digital ecosystems, inconsistent modelling practices, and limited standardization. In the context of AI, these technological constraints become even more critical, as AI-enabled processes depend on structured, accessible, and high-quality data. Organizational factors display a similarly strong presence. Roles and expertise, training, and top management commitment register substantial levels of primary and secondary coding, indicating that human capital, leadership, and internal capability-building are consistently recognized as essential enablers of digital transformation. These findings align with prior research showing that BIM adoption - and, by extension, BIM-AI integration- depends heavily on organizational vision, staff competencies, and structured change management processes. In contrast, market-related factors such as client demand, standards and policies, and market pressure appear with more variability. Their uneven distribution suggests that external influences on BIM-AI adoption differ significantly depending on national regulatory environments, procurement cultures, and the maturity of the surrounding ecosystem. Nonetheless, when these factors are addressed, they emerge as decisive drivers - particularly in contexts where mandates, standards, or procurement requirements act as catalysts for digital transformation. Project-related factors, including project delivery systems, lean principles, and process and project management, appear less frequently in the literature. Their lower coding values highlight a gap in current research, despite the fact that contractual structures, coordination routines, and project governance models can significantly influence how digital tools are mobilized within project networks. Similarly, pilot projects and external partners show the lowest levels of primary coding, indicating that empirical experimentation and inter-organizational collaboration remain underexplored dimensions of BIM-AI research.

Table 3: Coding scale for thematic relevance

Code	Label	Color
3	Primary focus	
2	Substantially discussed	
1	Mentioned	
0	Not addressed	

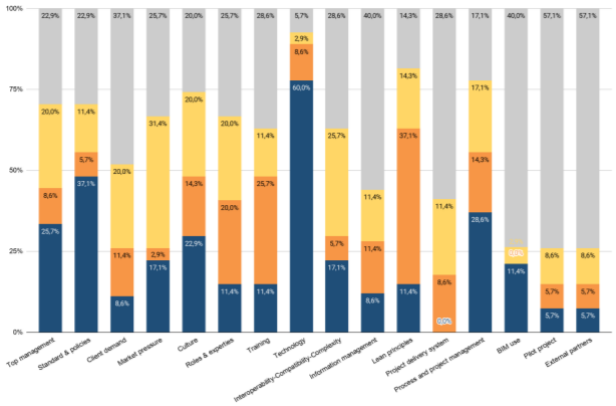


Figure 4: BIM-AI Key factors of the selected studies

The ANT diagram

This study proposes a socio-technical methodology based on ANT to study the BIM-AI implementation in design firms. The proposed ANT diagram is structured into four macro-areas that synthesize the main key factors identified in the literature. These key factors are organised into stakeholders characteristics ("human"), market, project, and technology characteristics ("non-human") (Figure 5), and are further articulated into two additional subgroups:

1. Stakeholders characteristics: internal and external;
1. Market characteristics: standard & regulations and environmental & socio-cultural factors;
2. Project characteristics: project delivery systems and project management;
3. Technology characteristics: BIM and AI.

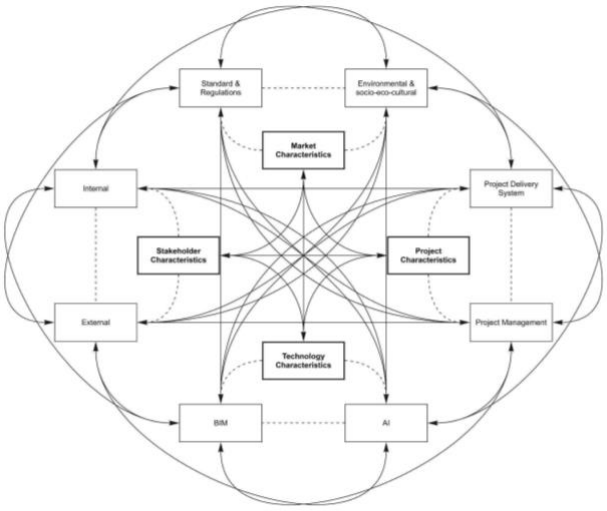


Figure 5: ANT diagram

This structure supports future analysis and enables the examination of social composition and decomposition of key actors. Also, the ANT diagram helps the research team to use the connections between the identified actors as a basis to formulate the questions for semi-structured interviews and case studies.

Stakeholder characteristics

As part of the ANT diagram development, the research team classified stakeholders involved in design, construction, and operation processes. Table 4 provides a summary of the identified human actors categorized into an "Internal stakeholders" and "External stakeholders".

Table 4: Stakeholder characteristics components

Internal stakeholders	External stakeholders
Public/Private owners	Local/National authorities
Trade unions	Regulatory/Permitting agencies
Architect	Insurance companies
Engineer	Financial institutions
Project manager	Environmental agencies
BIM professionals	Public community
General contractor	Technology vendors
Consultants	Real estate stakeholders
Surveyors/GIS experts	Utility providers
Fabricators/Suppliers	Academic/Professional training providers
Facility operators	
ICT/IT department	

Market characteristics

Table 5 provides a summary of the identified Market characteristics ("non-human" actors) categorized into "Standard & regulations" and "Environmental & socio-cultural factors".

Table 5: Market characteristics components

Standard & regulations	Environmental & socio-cultural factors
National legislation	Sustainability & environmental policies
Mandatory regulations	Education & training culture
Technical standards/codes	Professional/Digital readiness
BIM standards	Cultural attitudes
Design and construction guidelines	Local demand
Facility management protocols	Economic factors
Government Digitalization Programs	
Data Management Standards	

Project characteristics

Although the literature has extensively examined organizational, technological, and human-related factors, the role of the Project Delivery System remains relatively underexplored. For this reason, its inclusion among the key components of the network represents a relevant and innovative contribution. Table 6 reflects this advancement by categorizing project characteristics into "Project management" and "Project delivery system".

Table 6: Project characteristics components

Project management	Project delivery system
Project objectives	Design Build (DB)
Design management	Design Bid Build (DBB)
Quality control	Construction Manager at Risk (CM@R)
Model-based coordination practices	Integrated Project Delivery (IPD)
Lean principles	
Integrated information flows	
Decision-making	

Technology characteristics

Table 7 provides a summary of the identified Technology characteristics ("non-human" actors) categorized into "BIM" and "AI".

Table 7: Technology characteristics components

BIM Use Case	AI General use
Concept	Generation, review and synthesis of texts and documents
Rendering	Generation of images and graphic content
Design and Modelling	Transcription and analysis of meeting notes

Analysis and Simulation	Workflow automation
Collaboration & Coordination	Data analytics & BI
Time Scheduling (4D)	**AI Domain-specific use**
Cost Estimation (5D)	Generative design and design optioning
	Automated documentation
	Visual programming & parametric modeling
	Security, analytics and computer vision
	Rendering
	Model coordination
	Scan-to-BIM

Discussion

The literature on BIM implementation within design firms has predominantly examined barriers, adoption levels, benefits, and internal organizational dynamics. While these studies offer valuable insights, they often adopt an inward-looking perspective that frames BIM primarily as a technical or managerial issue. By contrast, the ANT lens adopted in this study shifts attention toward the heterogeneous network of relations that shape the implementation of new processes, technologies, and standards. Through ANT, BIM-AI integration emerges as a socio-technical reconfiguration involving negotiations, translations, and alignments among multiple actors. The ANT diagram developed in this research operationalizes this perspective by making visible the connections across stakeholder groups, technological components, market conditions, and project characteristics. These connections provide the analytical basis for investigating how specific relationships influence BIM-AI adoption. For example, linking internal stakeholders (e.g., BIM professionals) with AI general uses (e.g., workflow automation) highlights how modelling practices may be reshaped by automated processes and what organizational conditions support such transitions. Similarly, connecting standards and regulations with AI domain-specific uses (e.g., security and analytics) underscores the importance of protocols, data governance, and regulatory frameworks in enabling responsible AI integration. The interplay between project management practices (e.g., lean principles) and AI-enabled analytics illustrates how information flows, decision-making routines, and continuous improvement processes evolve when AI tools are introduced. Likewise, examining project delivery systems alongside BIM-AI objectives clarifies how contractual structures and collaboration models influence technological adoption. Finally, connecting internal stakeholders with AI domain-specific uses supports an investigation of how organizations plan to embed AI within BIM workflows, redistribute responsibilities, and develop new competencies. Overall, these insights address RQ1 by showing that BIM implementation in design firms is best understood as a dynamic socio-technical process and by demonstrating how ANT can structure their analysis. They address RQ2 by identifying the key factors that influence BIM-AI integration. These findings provide a structured foundation for the empirical phase of the research, where the ANT-based framework will be tested and refined across different organizational contexts.

Conclusions

This study proposes an ANT-based diagram to support researchers in analyzing how AI becomes integrated within BIM processes in design firms. The main contribution lies in the development of a structured methodology that applies an ANT lens to interpret the sociotechnical relationships that enable or hinder BIM-AI implementation. By positioning human and non-human actors at the same analytical level, ANT provides a robust and transferable framework capable of capturing how technologies, standards, roles, and organizational routines interact during implementation. The proposed diagram serves as the conceptual foundation for the next phase of the research, which will involve empirical investigations through semi-structured interviews and case studies across different organizational contexts. This subsequent phase will allow the research team to test and refine the framework, assessing its ability to identify enabling conditions, tensions, and translation processes within BIM-AI networks. The expected outcomes will advance current knowledge on BIM implementation by offering an interpretative tool that supports both academic inquiry and professional practice in the AECO sector. Overall, the study contributes a novel sociotechnical perspective to the emerging field of BIM-AI integration and provides a methodological pathway for future empirical work.

Acknowledgments

This doctoral research is funded under the Italian National Recovery and Resilience Plan (PNRR) by the European Union through the NextGenerationEU, and co-funded by a design firm. The author wishes to express her gratitude to the organization that hosted the research stay abroad for the support provided to the activities.

References

Abd Hamid, A. B., Embi, M. R., Mohd Taib, M. Z., & Abdul Razak, A. H. N. (2020). Enhancing the knowledge and proficiency for interior designers in Malaysia through the implementation of Building Information Modelling. IOP Conference Series: Materials Science and Engineering, 713, 012047.

Ahuja, R., Jain, M., Sawhney, A., & Arif, M. (2016). Adoption of BIM by architectural firms in India: Technology-organization-environment perspective. Architectural Engineering and Design Management, 12(4), 311–330.

Aiudi, A., Francioni, B., Kyrdoda, Y., & Amerio, S. (2025). The role of artificial intelligence in international strategic decision-making for SMEs. Journal of Strategy and Management. Advance online publication.

Al-Muntaser, N. S., & Ashur, S. A. (2024). Assessment of the integration of artificial intelligence (AI) into Building Information Modeling (BIM) for smart construction management and decision-making. In Proceedings of the 2024 ASEE North Central Section Conference. American Society for Engineering Education.

AlBalkhy, W., Chaize, E., Morael, V., Lafhaj, Z., & Iordanova, I. (2023). Lean construction and Building Information Modeling (BIM) in design management: Survey results from French companies. In Proceedings of the 31st Annual Conference of the International Group for Lean Construction (IGLC31) (pp. 220–229).

Albukhari, I. N. (2025). The role of artificial intelligence (AI) in architectural design: A systematic review of emerging technologies and applications. Journal of Umm Al-Qura University for Engineering and Architecture. Advance online publication.

Alnaser, A. A., Hassan Ali, A., Elmousalami, H. H., Elyamany, A., & Gouda Mohamed, A. (2024). Assessment framework for BIM-Digital Twin readiness in the construction industry. Buildings, 14(1), 268.

Altassan, A., Othman, M., Elbeltagi, E., Abdelshakor, M., & Ehab, A. (2023). A qualitative investigation of the obstacles inherent in the implementation of Building Information Modeling (BIM). Buildings, 13(3), 700.

Alves, T., Goia, F., Mahdjoubi, L., Delgado, J. M. D., Vieira, A. S., & Bragança, L. (2025). Towards an integrative framework for BIM and artificial intelligence capabilities in smart architectural and urban design. Sustainable Cities and Society, 113, 105010.

Arayici, Y., Coates, P., Koskela, L., Kagioglou, M., Usher, C., & O'Reilly, K. (2011). BIM adoption and implementation for architectural practices. Structural Survey, 29(1), 7–25.

Attia, A. R. (2025). The impact of integrating artificial intelligence and Building Information Modeling (BIM) systems on the development of construction methodologies. Journal of Umm Al-Qura University for Engineering and Architecture. Advance online publication.

Bolpagni, M., Burdi, L., & Ciribini, A. L. C. (2017). The implementation of Building Information Modelling and Lean Construction in design firms in Massachusetts and its correlation with client's requirements. In K. Walsh, R. Sacks, & I. Brilakis (Eds.), Proceedings of the 25th Annual Conference of the International Group for Lean Construction (IGLC) (pp. 235–242).

Bolpagni, M., Bartoletti, I., & Ciribini, A. L. C. (2021). Artificial intelligence in the construction industry: Adoption, benefits and risks. In Proceedings of the CIB W78 Conference 2021. Luxembourg.

Bonanomi, M. M. (2019). A systematic methodology for change management: Proposal for a digital transformation support. In Digital transformation of multidisciplinary design firms (pp. 37–47). Springer.

Cavka, H. B., Staub-French, S., & Pottinger, R. (2015). Evaluating the alignment of organizational and project contexts for BIM adoption: A case study of a large owner organization. Buildings, 5(4), 1265–1300.

Chan, C. T. W. (2014). Barriers of implementing BIM in construction industry from the designers' perspective: A Hong Kong experience. Journal of System and Management Sciences, 4(2), 24–40.

Chen, K., Zhou, X., Bao, Z., Skibniewski, M. J., & Fang, W. (2024). Artificial intelligence in infrastructure construction: A critical review. Frontiers of Engineering Management. Advance online publication.

Davoudabadi, S., Pedo, B., Tezel, A., & Koskela, L. (2022). A cognitive review for improving the collaboration between BIM and Lean experts. In Proceedings of the 30th Annual Conference of the International Group for Lean Construction (IGLC30) (pp. 669–680).

Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2008). BIM handbook: A guide to building information modeling for owners, managers, designers, engineers, and contractors (1st ed.). John Wiley & Sons.

Edirisinghe, R., Pablo, Z., Anumba, C., & Tereno, S. (2021). An actor–network approach to developing a life cycle BIM maturity model (LCBMM). Sustainability, 13(23), 13273.

El Masry, M., & Chronéer, D. (2025). Obstacles to BIM adoption in construction production: A study of Swedish construction contractors' experiences. International Journal of Construction Management. Advance online publication.

Frank, E. W., Dahy, H., & Sánchez Vibæk, K. (2022). Challenges in creating a sustainable building certificate for single-family housing in Denmark through an Actor-Network Theory (ANT) lens. Current Research in Environmental Sustainability, 4, 100144.

Hamarashid, R. (2021). Challenges of implementing digital tools in a municipality: The case of BIM in the land allocation process (Master's thesis). Jönköping University, School of Engineering.

Hochscheid, E., & Halin, G. (2018). BIM implementation in architecture firms: Interviews, case studies and action research used to build a method that facilitates implementation of BIM processes and tools. In Proceedings of the 36th eCAADe Conference (pp. 231–240).

Hochscheid, E., & Halin, G. (2020). Generic and SME-specific factors that influence the BIM adoption process: An overview that highlights gaps in the literature. Frontiers of Engineering Management, 7(1), 119–130.

Jang, J., Kim, H., Lee, S., & Kim, J. (2025). Generative AI in architectural design: Application, data, and evaluation methods.

Latour, B. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford University Press.

Lindblad, H., & Vass, S. (2015). BIM implementation and organisational change: A case study of a large Swedish public client. Procedia Economics and Finance, 21, 178–184.

Liu, Z., Wang, Y., Osmani, M., & Demian, P. (2020). Leveraging micro-level Building Information Modeling for managing sustainable design: United Kingdom experience. Advances in Civil Engineering, 2020, Article 3641950.

London, K., & Pablo, Z. (2017). An actor-network theory approach to developing an expanded conceptualization of collaboration in industrialized building housing construction. Construction Management and Economics, 35(8–9), 553–577.

Ma, L., Zhao, X., Jiang, R., Wu, C., Liao, L., Yang, Z., & Tan, J. (2025). Adopting large language models in the construction industry: Drivers, barriers, and strategic implications from China. Buildings, 15(23), 4296.

Pettigrew, A. M. (1987). Context and action in the transformation of the firm. Journal of Management Studies, 24(6), 649–670.

Proietti, S., & Magnani, R. (2025). Assessing AI adoption and digitalization in SMEs: A framework for implementation. arXiv.

Raj, K., Mastrolembo Ventura, S., Comai, S., & Ciribini, A. L. C. (2025). Toward a sustainable and efficient design process: A BIM-based organisational framework for public agencies—An Italian case study. Sustainability, 17(15), 6716.

Rydin, Y. (2013). Using Actor-Network Theory to understand planning practice: Exploring relationships between actants in regulating low-carbon commercial development. Planning Theory, 12(1), 23–45.

Sadeh, H., Mirarchi, C., & Pavan, A. (2021). BIM implementation for micro, small and medium-sized enterprises. In Proceedings of the 2021 European Conference on Computing in Construction (EC3) (pp. 1–8).

Succar, B., & Kassem, M. (2015). Macro-BIM adoption: Conceptual structures. Automation in Construction, 57, 64–79.

Tehami, M., & Seddiki, M. (2023). Investigation toward the adoption of Building Information Modelling in Algeria from architects' perspective. Journal of Construction in Developing Countries, 28(2), 329-352.

The American Institute of Architects. (2024). The architect's journey to specification: Artificial intelligence adoption in architecture firms-Opportunities & risks. AIA.

Tong, N., & Phung, Q. (2021). Developing an organizational readiness framework for BIM implementation in large design companies. International Journal of Sustainable Construction Engineering and Technology, 12(3), 57–67.

Troiani, E., Mahamadu, A.-M., Manu, P., Kissi, E., Aigbavboa, C., & Oti, A. (2020). Macro-maturity factors and their influence on micro-level BIM implementation within design firms in Italy. Architectural Engineering and Design Management, 16(3), 209–226.

Wang, Z., Liu, Z., & Liu, J. (2022). Innovation strategy or policy pressure? The motivations of BIM adoption in China's AEC enterprises. Journal of Asian Architecture and Building Engineering, 21(4), 1578–1589.

Wang, X., Zhao, Y., Zhang, W., Li, Y., Shi, X., Xia, R., Su, Y., Li, X., & Xu, X. (2025). Artificial Intelligence-Based Architectural Design (AIAD): An influence mechanism analysis for the new technology using the hybrid multi-criteria decision-making framework. Buildings, 15(21), 3898.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

FABRICATION-AWARE GENERATIVE DESIGN FOR HYBRID PREFABRICATION USING STABLE DIFFUSION XL

Hang Zhang[1], Jianpeng Cao[2], Anton Savov[1], and Benjamin Dillenburger[1]
[1]ETH Zurich, Zurich, Switzerland
[2]The University of Hong Kong, Hong Kong S.A.R., China

Abstract

Existing generative design approaches for floorplan generation rarely incorporate fabrication constraints, limiting their applicability in industrialized construction. This study presents a fabrication-aware generative design framework for hybrid prefabrication based on Stable Diffusion XL. The framework integrates spatial, functional, and manufacturing constraints to balance design flexibility and production efficiency. A two-stage training strategy generates both conceptual layouts and module-integrated solutions. Experiments demonstrate the generation of boundary-compliant layouts and standardized module allocation under diverse boundary conditions and text-based requirements. An interactive workflow supports convergence from conceptual designs to fabrication-aligned solutions, bridging early-stage architectural intent with downstream fabrication requirements.

Introduction

While data-driven layout generation has accelerated early-stage concept exploration, architectural workflows frequently treat downstream fabrication as an afterthought. Recent surveys highlight significant advancements across rule-based, optimization-driven, and data-driven floorplan synthesis, yet they consistently underscore bottlenecks regarding constraint satisfaction and practical constructability (Weber et al., 2022). In parallel, modular building automation has progressed by leveraging BIM-centric data representations and specialized configurators to explicitly systematize modules alongside their spatial relationships (Liu et al., 2024). Merging these high-capacity generative priors with standardized prefabrication rules presents a clear opportunity for industrialized construction.

Yet existing approaches remain split. Mainstream generative floorplan frameworks, such as HouseDiffusion, (Shabani et al., 2023; Zhang & Zhang, 2025), operate primarily on idealized, rectangular room abstractions and do not account for prefabricated assemblies. Conversely, modular-layout methods, such as CoGAN (Ghannad & Lee, 2022) and graph-based BIM pipelines (Gan, 2022b), enforce modular structure but often offer limited design variety and are not integrated with generative frameworks.

To bridge this gap, the goal of this research is to develop a Fabrication-aware generative design framework for prefabrication that accepts predefined modules as input and incorporates them into the customized floor plan generation. The framework aims to increase the adoption of standard modules within a boundary-compliant layout, enabling efficient exploration of feasible configurations while preserving production efficiency in industrialized construction.

Literature review

This section reviews computational methods for automatic floor plan generation. The references are limited to industrialized construction. In addition to top-down and bottom-up methods, relevant BIM-based parametric methods are also included. To maintain the high quality, the analyzed studies are only sourced from journal publications.

Top-down methods

The top-down approach follows a decomposition workflow, where a floor plan is divided into discrete modular units, or referred to as modularization. Lin et al. formulate the modularization process by first segmenting the original space into smaller regions and then merging them to obtain an optimal layout under shape and size constraints. The generation is evaluated via two indicators, number of modules and the diversity of modules with different dimensions (Lin et al., 2025). The generation achieved optimal modular key plan convergence in 50 iterations, with automated runtimes averaging only 7.4 seconds. Cao et al. translate the decomposition process as the decomposition process as matching predefined modules to the given floor plan. The difference compared to Lin's approach is that the floor plan is fixed, while the prefabrication strategy, including panelized units and modular units, is determined by evaluating the construction cost and durations (Cao et al., 2024). The study demonstrated that shifting to a fully volumetric strategy over a fully panelized one can accelerate construction schedules by up to 23%, although it concurrently inflated material and shipping costs by roughly 22%.

Bottom-up methods

The bottom-up approach starts with the definition of standard modules and establishes the generative rules for module placement. Guided by objective functions, the optimal layout is created. Zheng et al. develops a generative design method for the floor layout of modular buildings. The method uses room module, corridor module and stair module as building blocks. The meta-heuristic rules are provided by geometric and functional relationship between module units. The objective functions include project cost and layout penalties (Zheng et al., 2024). The generation could escape local minima to discover globally optimized volumetric configurations within 42 generations (under 870 seconds). Sheijani et al. adopt a bottom-up approach to automate layout generation for office buildings. The generation is framed as identifying the most inclusive path for placing modules (Sheijani et al., 2024). The results show that the approach can condense the standard front-end schematic design phase from a typical timeframe of multiple weeks down to just 5 to 7 days.

BIM-based parametric methods

Beyond these two methods, there are BIM-based generative design approaches that accommodate a broader range of building topologies. Gan developed a customized BIM object library where elements with different embodied carbon and cost index could be matched with the model geometry and material information to create the alternative designs (Gan, 2022a). The generative framework by Gan (2022a) can create between 1,000 and 1,500 new design options in approximately 30 minutes. Li et al. proposed a parametric generative design method for precast buildings using BIM and graph convolutional neural networks. The layout is explicitly parameterized through a set of spatial and geometric features, and the resulting designs are evaluated and optimized with respect to wind pressure performance (Li et al., 2024). The method achieves a highly accurate layout prediction rate of 98.79% (R^2) on testing sets for evaluating surface-averaged wind pressure coefficients. However, the BIM-based approach is less efficient during the layout generation, when design exploration is primarily driven by functional program, rather than specific building elements.

In short, the top-down strategy tends to defer module design to the final stage, which may result in numerous irregular modules, thereby reducing manufacturing efficiency and increasing overall project costs. Although prior studies have considered fabrication constraints and modularization efficiency (Cao et al., 2024; Lin et al., 2025) in the post-design phase, they overlook the opportunity for early manufacturer engagement. Delayed engagement might lead to prolonged lead times for procuring building materials before fabrication. In contrast, the bottom-up approach integrates standard modules directly within the generative process, potentially overcoming these limitations (Sheijani et al., 2024; Zheng et al., 2024). However, the generation process is usually driven by a set of rules which limits their scalability and usually consumes longer iterations. Last but not least, the bottom-up approach is not flexible with respect to predefined building boundaries and room adjacency requirements.

Methodology

Dataset preparation

Training data are derived from the SwissDwellings (Van Engelenburg et al., 2025) database, which provides apartment layouts as vector room outlines. Figure 1 summarizes our two-stage preprocessing pipeline. Stage 1 converts each vector layout into a semantic raster target and a set of conditional inputs, including the site context, a partial layout graph, and a structured prompt. Stage 2 augments this dataset with fabrication labels by sampling candidates from the Stage 1 model, filtering invalid instances using geometric checks, and matching them to standardized wet-core modules.

Problem Formulation

We define fabrication-aware floorplan generation as learning the parameters θ of a conditional generative model, $p\theta(F^{img} \mid G^{img}, T_{desc}, T_{fab}, V_{Module}, Mode)$. As illustrated in Figure. 1, given multi-modal inputs that encode spatial, functional, and fabrication constraints, the model synthesizes a floorplan image $F^{img} \in R^{\{512\times512\times3\}}$. The objective is to maximize the likelihood of generating coherent, feasible, and architecturally valid designs that are consistent with both functional program and manufacturing realities. The input modalities are defined as follows:

Layout Graph Image (G^{img}) A 2D raster image that encodes the spatial topology as adjacency graph $G^{vec} = (V, E)$, where V denotes functional room values of types and locations, and E denotes the adjacencies to be satisfied. The graph input is rasterized on canvas $\mathbf{G^{img}}$, and it may be partial ($G_{partial}$), specifying only key relationships, or complete ($G_{complete}$), defining the full topology.

Module (V_Module): We predefined five module prototypes with various dimensions, as shown in Module Library in Figure 1.

Floorplan Description (T_desc): A structured prompt of language specification, describing the functional program, including the room types and locations. For example: "Floorplan with a Northeast Bedroom, a West Corridor." The description may be partial or complete, depending on the level of details the user wishes to impose.

Fabrication Information (T_fab): A structured prompt of language specification, describing manufacturing constraints that ground the design in the selected construction system. For example: "use bath-kitchen module, less panels, medium doors, medium windows." This input may also be partial or complete, allowing different degrees of fabrication control.

Mode: A mode switcher that identify whether the model generate layout with only design constraints or generate layout that be aware of fabrication constraints.

Meanwhile, the model output is defined as follows:

Floorplan Image (F): A high-resolution raster image of a complete floorplan. The generated plan must jointly satisfy the constraints imposed by all input modalities. Specifically, it should:

(1) Respect the spatial topology encoded by the layout graph image G.

(2) Fulfill the functional program (room types and locations) described in T_desc.

(3) Comply with the fabrication constraints, including clear placement of predefined modules (e.g., bathroom, kitchen) and compatibility of remaining spaces with the stated construction system (e.g., timber panels).

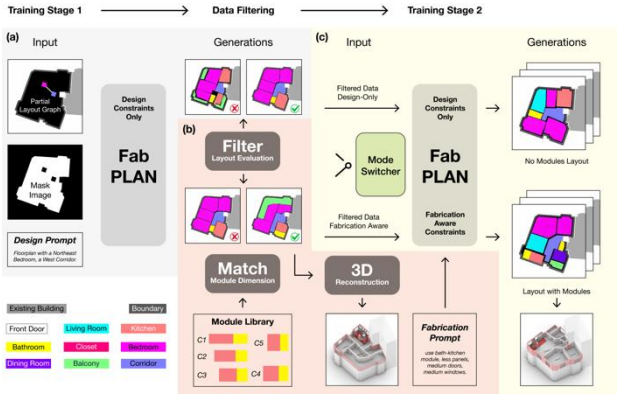


Figure 1: Framework of Fab-aware GD Framework

Fab-aware GD Framework using SDXL

As also described in Figure 1, we propose a Fab-aware GD framework built on Stable Diffusion XL (SDXL) (Podell et al., 2023), where the UNet, the core denoising network responsible for transforming random noise into coherent images, is fine-tuned with hybrid conditioning to support both design exploration and fabrication aware generation. To maintain general design capability while enforcing fabrication constraints, we introduce a mode-switching strategy. When under design mode, the model is conditioned on **G** and **T_desc**, encouraging broad layout diversity within the boundary and adjacency constraints. When under fabrication-aware mode, the model is additionally conditioned on the module specification **V_Module** and **T_fab**, biasing generation toward layouts that can accommodate standardized modules and compatible panel logic. This dual-path conditioning prevents overfitting to modular templates while enabling reliable module-aware outputs.

During fine-tuning, both modes are sampled within each training epoch using filtered design-only data and synthetically enriched fabrication-aware data. This yields a single UNet that can be toggled at inference time to produce either unconstrained conceptual layouts or fabrication-ready layouts that explicitly integrate standardized modules. In effect, the framework unifies early-stage architectural exploration and downstream manufacturing feasibility within a single SDXL-based model.

Data Preparation and Two-Stage Training Pipeline

To generate layouts that are both spatially valid and compatible with fabrication constraints, as demonstrated in Figure 2, we adopt a two-stage training pipeline with dataset filtering and synthetic fabrication-aware enrichment.

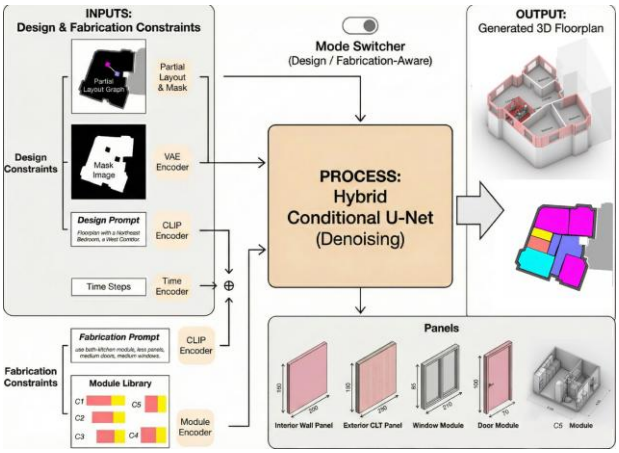


Figure 2: Data Preparation and Two-Stage Training Framework

Training Stage 1: Design-Centric Initialization

We first train the base model with design constraints only to establish robust spatial priors. The model receives partial layout graphs (room types, locations, and adjacency), mask images (the region model should denoise), and structured prompt design prompts. It then generates candidate floorplans without fabrication inputs.

Data Filtering and Module Matching

Stage 1 outputs are screened to remove invalid or low-quality layouts. The retained plans are passed to a module dimension matching procedure that fits standardized components from a module library (C1–C5, e.g., bathroom and kitchen units). The selection of five module prototypes is motivated by the geometric variation of standard residential wet-core units, ranging from compact square layouts to elongated narrow configurations. These modules represent bathrooms and kitchens, which are among the most service-intensive and costly elements in prefabricated housing. While the number of possible module types is not inherently limited, expanding the library would increase the complexity of module embedding and training, whereas practical industrial systems typically rely on a relatively small family of standardized products. The five-module library therefore provides a balanced test set that captures representative variation while remaining computationally and practically tractable."

For successful matches, the system reconstructs the layout in 3D and automatically generates fabrication prompts describing the applied module and panel constraints. This yields a synthetic fabrication-aware dataset pairing geometry with explicit manufacturing information.

To ensure high-quality training data for the diffusion model, raw generated layouts are passed through an automated rule-based validation pipeline. A generated

layout is classified as invalid or low-quality if it fails any of the following four objective criteria:

- Room Shape Convexity: Deformed or highly irregular room polygons are rejected using a compactness index (ratio of room area to its axis-aligned bounding box area). Layouts falling below a mean threshold of 0.85 are discarded.
- Aspect Ratios: Rooms with extreme, unbuildable proportions (e.g., long, thin slithers) are flagged by checking side-length ratios.
- Connectivity: A room-contact graph is built by analyzing polygon intersections. A layout is invalid if a room has no feasible spatial connection to the rest of the dwelling (e.g., a bathroom that is isolated and cannot be accessed from a corridor or living room).
- Essential Room Types: To be valid, a residential layout must contain a core set of functional zones: a kitchen, a bathroom, and a living or sleeping area.
- Module Feasibility: Bathroom, kitchen, and combined total area ratios are computed against the module's specifications. A module is considered feasible if all three ratios fall within a 15% tolerance threshold.

Training Stage 2: Fabrication-Aware Refinement

After assembling the filtered design-only dataset and the synthetically enriched fabrication-aware dataset, we fine-tune the SDXL UNet with a dual-path strategy controlled by a mode switcher. (1) Mode A (design-only): The model is conditioned on the layout graph image and design prompt (and boundary masks, if applicable). This path preserves spatial generality and layout diversity, and prevents overfitting to modular templates. (2) Mode B (fabrication-aware): The training target is the same floorplan image, but the model is additionally conditioned on module specifications and fabrication prompts. This path enables dimensional feasibility and module accommodation. Mini-batches are routed between the two modes using a fixed sampling policy, yielding a single unified model. At inference, the same UNet can be toggled to generate either concept-level layouts or module-integrated layouts that better satisfy fabrication constraints.

Results

We evaluate the proposed fabrication-aware generative framework in three experiments: (1) controlled comparisons between design-only and fabrication-aware outputs under identical boundary and partial inputs, (2) robustness to different boundary conditions and partial-input completeness, and (3) iterative, user-in-the-loop customization including module freezing. All results are generated from partial inputs to reflect early-stage design practice.

Fabrication-Aware Generation

Figure 3 compares design-only and fabrication-aware generation using the same boundary condition. Design-only outputs satisfy the partial topology and program intent, but do not explicitly allocate space for standardized modules, often implying partition or opening patterns that are difficult to align with fabrication systems. With fabrication-aware conditioning enabled, the model consistently introduces module-compatible allocations and clearer service clustering while maintaining boundary compliance and the intended spatial logic.

Using identical inputs, we further test prompt-level controllability within fabrication mode across three variants: baseline module integration, reduced panel/door/window complexity, and increased panel/door/window richness. The results show that the model can shift between assembly-simplified and architecture-enriched solutions without breaking module fit or topology.

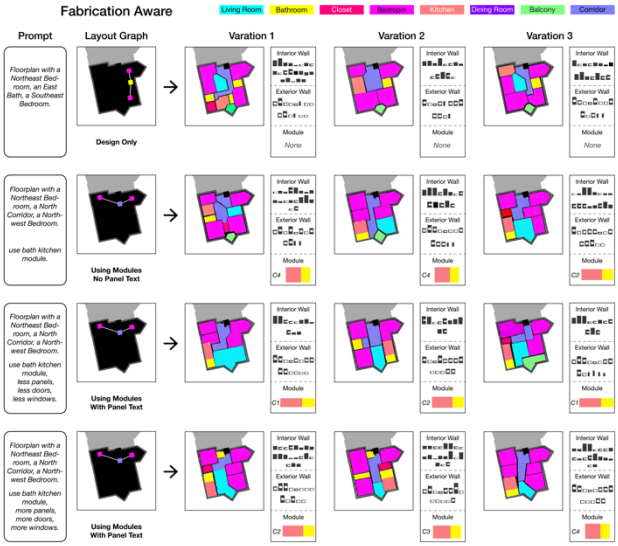


Figure 3: Design-only vs fabrication-aware outputs with identical boundary and partial inputs

Various Partial Inputs

As illustrated in figure 4, we evaluate fabrication-aware generation across diverse boundary conditions and varying levels of partial input. For both regular and irregular envelopes, the model preserves boundary compliance and maintains plausible module placement aligned with the provided adjacency cues. When the partial input is sparse, the module specification acts as a stabilizing prior, producing a broad set of feasible configurations. As the partial input becomes more detailed, the output distribution narrows and more consistently reflects user intent. These results indicate that Fab-aware generation remains robust under both geometric variability and incomplete early-stage requirements.

Iterative Customization

Figure 5 demonstrates an interactive workflow in which users progressively refine partial inputs rather than specifying all constraints upfront. Starting from an initial partial graph and boundary, the model produces multiple candidate solutions. The user then selects a preferred outcome and adjusts the partial input to steer the next iteration.

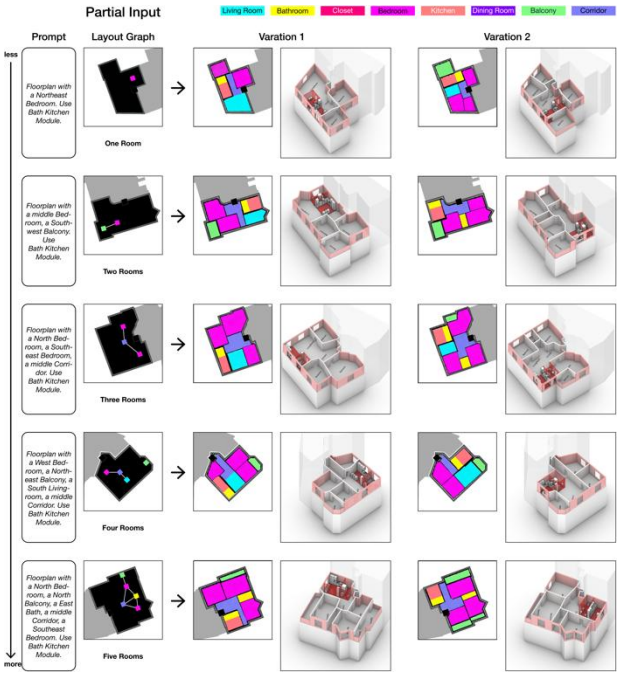


Figure 4: Fabrication-aware results across different boundaries and increasing partial-input completeness.

Modifying constraints directly shapes the design space. For instance, adding constraints yields more precise, less diverse outcomes; editing constraints corrects local functional issues while preserving global structure; and removing constraints increases diversity under the same boundary and module rules. Finally, users can freeze accepted module positions and dimensions, prompting the model to regenerate surrounding spaces while respecting the locked design decisions. This supports gradual convergence from conceptual exploration to fabrication-aligned layouts.

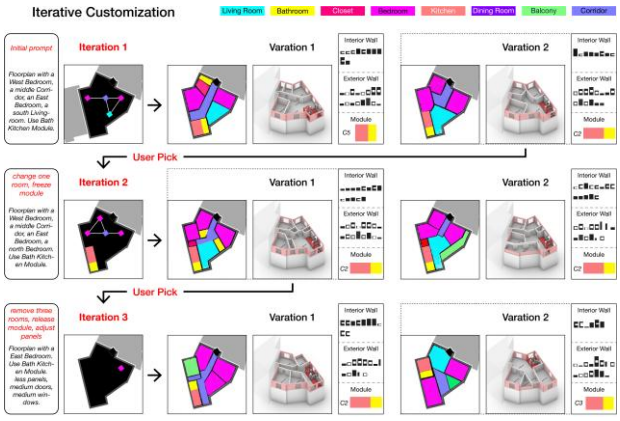


Figure 5: Iterative customization: initial partial input, user selection, add/edit/remove constraints, and module freezing with regenerated layouts.

Evaluation

Our framework is designed for floor plan generation where designers progressively specify site boundaries, partial layouts, and wet-core modules. There is no comparable prior SOTA methods that support the same combination of partial layout input and module-aware generation within a single generator. The framework was quantitatively evaluated on a held-out test split of 1,000 site conditions and prompts. We tracked perceptual visual similarity via Fréchet Inception Distance (FID). A lower FID is better, as it means the generated images are indistinguishable from real drawings. Besides, Module Feasibility is recorded as the percentage of samples whose decoded wet-core matches the specified module.

Table 1 presents a comparative benchmarking of the full framework against three key ablations, quantifying how each component affects architectural validity and module integration. In the default configuration, the full framework achieves the strongest overall performance, attaining a high spatial realism score (FID=5.68) and highest geometry pass rate and module feasibility (85.6%). Deactivating the mode switcher leads to a noticeable deterioration in performance, with module feasibility dropping to 64.2% and FID numbers increasing from 5.68 to 6.54. Omitting the data pipeline's geometric screening yields the poorest performance across the board. Under this scenario, module feasibility drops to its lowest point at 57.8%, while the FID worsens to 8.21. In particular, fine-grained geometric checks like shape convexity drop from 91.8% to 82.6%.

Table 1 Fabrication-aware ablation test

	FID	Mod Fea %	Geometry Pass Rate %			
			Conv	Ratio	Conn	Type
Ours (Full)	5.68	85.6	91.8	85.4	81.4	94.2
w/o Switcher	6.54	64.2	89.4	83.8	69.6	92.7
w/o Filter	8.21	57.8	82.6	77.1	72.4	90.8

Discussion

This study presents a fabrication-aware floorplan generation framework that unifies early-stage spatial reasoning with downstream manufacturability. By fine-tuning the UNet under hybrid conditioning and a mode-switching scheme, the model can generate both concept-level layouts and module-integrated solutions from partial inputs while respecting a fixed boundary. The results indicate that diffusion-based models can serve as a practical bridge between architectural intent and fabrication constraints without relying on brittle, hand-crafted placement rules.

A key contribution is the two-mode training strategy. The design-only path preserves broad generative capability and avoids narrowing the model to a small set of modular templates. The fabrication-aware path introduces explicit module and panel constraints, improving dimensional feasibility and constructability. This separation appears to be critical: without the design-only channel, the model risks reduced diversity and over-regularized outputs; without the fabrication-aware channel, module fit remains inconsistent. The combined approach yields a single

deployable model that can be toggled by conditioning, which is advantageous for real design workflows where the level of constraint evolves over time.

The results also highlight the value of the synthetic enrichment pipeline. Module matching and fabrication-prompt generation convert raw design outputs into a structured dataset that injects manufacturing knowledge into training. This offers a scalable alternative to manually annotated fabrication-aware datasets, which are expensive and scarce. The observed robustness across boundaries and partial-input amounts suggests that the enrichment strategy effectively teaches module accommodation as a generalizable behavior rather than a memorized pattern.

However, several limitations remain. First, representing layouts solely as raster outputs may obscure geometric precision needed for detailed coordination, especially for tolerance-sensitive assemblies. Second, the module matching success rate and its dependency on library richness likely influence overall performance; limited module typologies may bias the generator toward a narrower solution space. Third, the current framework infers constructability primarily through textual and library-based constraints rather than explicit structural, MEP routing, code compliance, or egress simulation. These aspects may require additional conditioning signals or post-generation validation.

Conclusion

Overall, our proposed Fab-aware GD framework demonstrates that diffusion models can support boundary-compliant, module-aware floorplan generation from partial inputs, offering a scalable approach to balancing design flexibility with fabrication efficiency. The mode-switching paradigm and synthetic fabrication enrichment provide a practical foundation for hybrid prefabrication workflows where manufacturability must be introduced progressively rather than imposed upfront.

Future work could focus on (1) integrating vector or parametric representations for higher geometric fidelity, (2) expanding module libraries and supporting multi-objective optimization across cost, carbon, and schedule, (3) incorporating explicit code and performance-aware validators in the loop, and (4) evaluating the system with professional designers to quantify usability, time savings, and trust. A promising direction is a closed-loop pipeline where diffusion-based generation, rule-based checking, and module library updates continuously inform one another.

References

Cao, J., Said, H., Savov, A., & Hall, D. (2024). Graph-Based Evolutionary Search for Optimal Hybrid Modularization of Building Construction Projects. Journal of Construction Engineering and Management, 150(8), 04024098. https://doi.org/10.1061/JCEMD4.COENG-14687

Gan, V. J. L. (2022a). BIM-Based Building Geometric Modeling and Automatic Generative Design for Sustainable Offsite Construction. Journal of Construction Engineering and Management, 148(10), 04022111. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002369

Gan, V. J. L. (2022b). BIM-based graph data model for automatic generative design of modular buildings. Automation in Construction, 134, 104062. https://doi.org/10.1016/j.autcon.2021.104062

Ghannad, P., & Lee, Y.-C. (2022). Automated modular housing design using a module configuration algorithm and a coupled generative adversarial network (CoGAN). Automation in Construction, 139, 104234. https://doi.org/10.1016/j.autcon.2022.104234

Li, K., Gan, V. J. L., Li, M., Gao, M. Y., Tiong, R. L. K., & Yang, Y. (2024). Automated generative design and prefabrication of precast buildings using integrated BIM and graph convolutional neural network. Developments in the Built Environment, 18, 100418. https://doi.org/10.1016/j.dibe.2024.100418

Lin, X., Su, P., Lu, W., & Guo, H. (2025). ModulePacking: A Top-Down Generative Design Approach for Modular Key Plans. Journal of Computing in Civil Engineering, 39(1), 04024047. https://doi.org/10.1061/JCCEE5.CPENG-6078

Liu, J., Qiu, Z., Wang, L., Liu, P., Cheng, G., & Chen, Y. (2024). Intelligent floor plan design of modular high-rise residential building based on graph-constrained generative adversarial networks. Automation in Construction, 159, 105264. https://doi.org/10.1016/j.autcon.2023.105264

Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., Penna, J., & Rombach, R. (2023). SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2307.01952

Shabani, M. A., Hosseini, S., & Furukawa, Y. (2023). HouseDiffusion: Vector Floorplan Generation via a Diffusion Model with Discrete and Continuous Denoising. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 5466–5475. https://doi.org/10.1109/CVPR52729.2023.00529

Sheijani, S. S., Momenaei, A., & Hassanzade, H. (2024). Automating modular open-plan office layouts with performance-based generative design. Automation in Construction, 167, 105692. https://doi.org/10.1016/j.autcon.2024.105692

Van Engelenburg, C., Mostafavi, F., Kuhn, E., Jeon, Y., Franzen, M., Standfest, M., Van Gemert, J., & Khademi, S. (2025). MSD: A Benchmark Dataset for Floor Plan Generation of Building Complexes. In A. Leonardis, E. Ricci, S. Roth, O. Russakovsky, T. Sattler, & G. Varol (Eds.), Computer Vision – ECCV 2024 (Vol. 15116, pp. 60–75). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-73636-0_4

Weber, R. E., Mueller, C., & Reinhart, C. (2022). Automated floorplan generation in architectural design: A review of methods and applications. Automation in Construction, 140, 104385. https://doi.org/10.1016/j.autcon.2022.104385

Zhang, H., & Zhang, R. (2025). Generating accessible multi-occupancy floor plans with fine-grained control using a diffusion model. Automation in Construction, 177, 106332. https://doi.org/10.1016/j.autcon.2025.106332

Zheng, Z., Li, Y., & Torres, J. (2024). An integrated method of automated layout design and optimization for modular construction. Engineering, Construction and Architectural Management, 31(3), 1016–1036. https://doi.org/10.1108/ECAM-04-2022-0329

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

SYN3D-LLM: SYNTHETIC 3D POINT CLOUDS GENERATED THROUGH LLM DISCUSSION

Nils Soeren Krause[1], Joern Ploennigs[1]
[1]University of Rostock, Germany
AI for Sustainable Construction

Abstract

Creating synthetic datasets that capture the diversity of real-world sites and measured point clouds remains challenging. The paper introduces SYN3D-LLM, a dialog-driven large language model framework for generating semantically annotated synthetic 3D point clouds by producing executable code through an iterative designer-critic loop. Experimental evaluations of the framework demonstrate that it improves the pass rate by discussing and testing generated code and by assessing the geometrical consistency between the initial prompt and the generated data. The paper demonstrates that SYN3D-LLM is versatile in generating point clouds for various cases.

Introduction

Most Scan-to-BIM, SLAM, or semantic segmentation models in the Architecture, Engineering, Construction, and Operations (AECO) sector are trained on real measured point clouds. This has the benefit of representing real-world conditions and complexities such as noise, occlusions, and varying point densities.

On the other hand, it requires extensive data collection efforts, including scanning diverse environments and manual annotation of point clouds, which can be labor-intensive and time-consuming. Furthermore, survey papers indicate that many Scan-to-BIM approaches are primarily evaluated on proprietary or project-specific datasets, hindering objective comparison within the State-of-the-Art (Abreu et al., 2023). This lack of standardized benchmarks limits reproducibility and comparability of workflows (Xu and Stilla, 2021; Luo et al., 2026) and the development of generalized models.

To address this challenge, researchers are exploring the use of synthetic point clouds (Noichl et al., 2021) as a complement to real measures to introduce a larger variety of data. Synthetic point clouds can provide a reliable foundation for benchmarking, as demonstrated by synthetic datasets such as ModelNet in Wu et al. (2015), which is widely used for object recognition. Studies such as those by Yue et al. (2025) show that synthetic point clouds are effectively augmenting real-world data, leading to improved performance of the trained model. The authors used a diffusion model to generate synthetic data, which, as before, requires training on large amounts of real data. Limiting this is the potential diversity of the generated data. Similar challenges arise in approaches that generate synthetic point clouds from IFC models Zou et al. (2026), which must be modeled and are available in only a limited variety.

To overcome these limitations, scalable and diverse synthetic point cloud generation methods are needed. We present SYN3D-LLM as an approach for generating synthetic point cloud data using large language models (LLMs) that do not require pre-training on real-world point clouds. While SYN3D-LLM does not introduce a new methodology, its contribution lies in integrating LLM-based self-refinement for AECO-specific point cloud generation. Beyond demonstrating that it is possible to generate synthetic point clouds with LLMs, we want to investigate the reliability and accuracy of the generated data in engineering contexts. We therefore focus on two key research questions:

- **RQ1:** How reliably does the proposed SYN3D-LLM framework generate executable point cloud code under iterative designer-critic refinements?
- **RQ2:** How accurately do the generated point clouds comply with the geometric dimensions and semantic constraints specified in the input prompts?

To address those research questions, SYN3D-LLM does not generate point clouds as a one-shot solution, but it uses an iterative framework to critically review and improve the point cloud drafts. To remove the requirement of data for training that generative models like Generative Adversary Networks (GAN) or Diffusion Models (DM) have, the approach relies on the LLMs capability to generate code from natural language prompts. For this SYN3D-LLM introduces a review-loop where the LLM assumes two roles: (i) as a code-generating designer and (ii) a code-reviewing critic. This enables SYN3D-LLM to generate point clouds of a wide complexity across a variety of construction types. The main contributions of this paper are:

Proposing SYN3D-LLM framework: Introduction of a dialog-based self-discussion framework in which an LLM generates synthetic point clouds as a programming designer.

Scalable, labeled data generation: An approach for generating different building classes, and directly annotated point cloud scenes for the AECO sector.

Evaluation of SYN3D-LLM: Investigation of the Pass rate and increasing token cost of iterations. As well as

constraining checks on the input and output information for geometrical consistency, and the semantic labels of the required information from the input prompts.

First, we introduce SYN3D-LLM in the SOTA context and distinguish it from existing work of synthetic point clouds for Scan-to-BIM workflows and semantic segmentation. Next, we describe the design and the functioning of the proposed framework in the methodology section. This is followed by data evaluation, a discussion of the results, an outlook, and a conclusion.

State-of-the-Art

The motivation for synthetic point clouds in the Scan-to-BIM context stems from the limitations of classical workflows and the growing need for benchmarks, scalable point cloud training data for machine learning and AI applications (Abreu et al., 2023).

Traditional Scan-to-BIM workflows consist of several processing steps, starting with point cloud measurement, followed by geometry segmentation and segment classification, and culminating in the reconstruction of 3D models or IFC data. The problem is that most approaches are project-specific and work primarily on the data for which they were developed. As data acquisition and pre-processing, including labeling, are time-consuming (Abreu et al., 2023), most approaches are limited in their generalizability.

To address these problems, simulation-based approaches generate synthetic point clouds from existing 3D models. Virtual laser scanning of BIM or CAD models can produce labeled point clouds automatically (Noichl et al., 2021). These approaches create synthetic point clouds by sampling points from the surfaces of explicit geometric representations, with semantic labels inherited from the underlying CAD or BIM structure (Pierdicca et al., 2019). In particular, IFC-based environments enable the simulation of laser-scanning processes to generate semantically annotated point clouds for Scan-to-BIM applications (Ma et al., 2021). Experimental results demonstrated in that case that synthetic point clouds enhanced the semantic segmentation tasks and reduced the amount of manually annotated real-world data required for training and semantic segmentation (Xiao et al., 2022; Stefanelli et al., 2022). But in these approaches, the variability of the generated data is constrained by the availability and diversity of predefined geometric models used for the simulated generation of synthetic point clouds.

Machine learning approaches usually target generative latent-space models. Achlioptas et al. (2018) introduced a deep generative model for point clouds by training autoencoders (AE) and modeling the latent space using Generative Adversarial Networks (GANs) and Gaussian Mixture Models (GMMs). This approach demonstrated the ability to reconstruct previously unseen point cloud samples from the ModelNet dataset.

More recent work uses 2D diffusion models to generate architectural images Paananen et al. (2024), floor plans He et al. (2023), or building designs Zhong et al. (2024). These applications remain 2D, limiting their applicability to 3D point cloud generation. The use of diffusion models for 3D point cloud generation is presented in Yue et al. (2025), where the authors propose a method to generate synthetic point clouds to augment real-world training data. This method still relies on large amounts of real-world data to train the diffusion model and, therefore, does not address the core limitations.

In addition, generative text models such as Large Language Models (LLMs) are increasingly being used in the AECO domain. Zheng and Fischer (2023) uses an LLM for natural language processing to develop an assistant for information retrieval from BIM models, and Wei and Li (2025); Du et al. (2026) use LLMs to generate floor plans. In context of point clouds, Hong et al. (2023) and Ye and Gao (2024) use LLMs for 3D interpretation of point cloud data to solve tasks such as captioning, question answering, navigation, and guidance via prompting. Xu et al. (2024) presents PointLLM, which combines point clouds and text for classification and description. Similarly, Yang et al. (2025) proposes LiDAR-LLM, which uses this combination to process LiDAR data. Further applications include point cloud registration with MiniGPT-3D Tang et al. (2024) and segmentation with SegPoint He et al. (2024). While these works demonstrate the potential of LLMs for point cloud tasks, the generation of point clouds with LLMs remains underexplored (Liu et al., 2024).

The challenge with LLMs like GPT-4 Achiam et al. (2023) is that they can analyze and generate text, but they cannot directly generate point clouds, as these data are typically unstructured data, stored in binary or other non-textual formats. LLMs have demonstrated strong capabilities in generating code from natural language Poldrack et al. (2023), and various libraries for processing and creating point cloud data exist, such as PyVista, Open3D, PyntCloud, or Laspy. Yet, using these libraries requires domain knowledge and additional manual programming effort. This opens an alternative path for point cloud generation, in which LLMs generate executable code that produces point cloud data directly, without requiring pre-existing 3D models or real-world training data.

While LLMs are good at code generation, they may lack domain knowledge and knowledge about the point cloud libraries mentioned above (Gu et al., 2025). In addition, LLMs are prone to hallucinations, have limited spatial understanding, and exhibit limited reasoning about causal constraints (Ploennigs et al., 2025). Therefore, special precautions are needed to ensure that code-based point cloud generation yields feasible results. The key advantage of the LLM-based approach is its ability to translate natural language into executable code, enabling prompt-driven scene generation without predefined rule sets.

These limitations motivate the exploration and design of a discussion-based LLM framework called SYN3D-LLM, for generating synthetic, semantically point clouds.

Methodology of the Framework

The workflow of the framework is presented in Figure 1. It starts with a *user prompt* that specifies what kind of point cloud should be generated. This user prompt is extended by the Designer with a *system prompt* that provides additional instructions for generating the point cloud. This encompasses strict code generation constraints, including the output schema, programming language (Python), and a restriction that external libraries are not allowed. It also provides information on the helper functions that the LLM can use to generate the point clouds. For example, helper functions exist for creating wall or ceiling structures filled with surface points. After the user prompt and the Designer system prompt are combined, the Designer generates a first code version. This code is not directly executed, but is sent to another LLM, called the Critic.

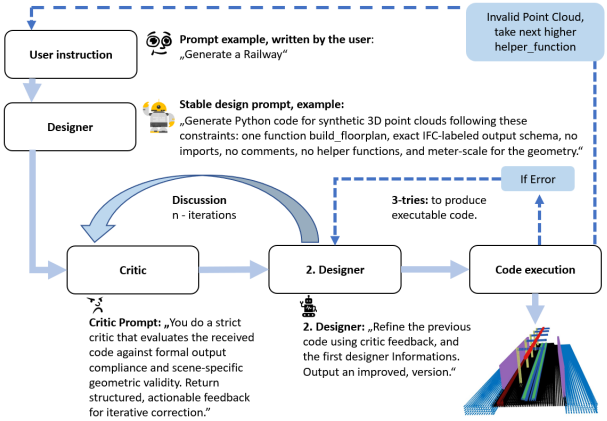


Figure 1: The proposed framework of SYN3D-LLM with its self-discussion functionality to obtain executable code through an iterative process for generating synthetic point cloud data.

The Critic receives the Designer's drafted code, another system prompt defining how it should review the code, and the information from the user prompt. The Critic's response is then sent, together with the previously created code, to the next iteration of the exchange between the Designer and Critic. The Designer then creates an improved code version based on the Critic's feedback and the user prompt. The framework performs a fixed number of Designer and Critic iterations, set to two by default, and the first design draft. With each draft and iteration, the framework has multiple tries to generate executable code (default: 3). The iteration setup represents a controlled experimental design rather than an optimisation stopping strategy. Increasing the number of iterations can improve refinements, but it also increases computational cost. The code's execution is checked by running it automatically in a Python environment. If execution succeeds and a constraint checker verifies that the point cloud contains points, the framework proceeds with the export. Over the iterations, the generated point cloud code becomes progressively more detailed. These improvements are not driven by changes in the *user prompt*, but emerge from the iterative refinement process between the Designer and Critic.

While writing the code, the LLM is instructed in the system prompt to label the individual objects and components based on the scene description.

The discussions between the Designer and Critic are automatically logged to *.CSV* and *.JSON* files for evaluation.

After all iterations are complete, the final code is executed to generate the point cloud. It is saved in three different formats: *.e57*, *.xyz*, and *.ply*.

The framework is complemented by a Graphical User Interface (GUI) in Open3D. It provides access to the settings (e.g., number of iterations), system prompts, and all scene descriptions and variations to generate the data. In the GUI, the discussions between the Designer and Critic are logged in real-time. The GUI also allows the selection of different building types and infrastructure scenarios. For each scenario, a different prompt templates exist that contains the system prompt for the Designer, along with helper functions and the system prompt for the Critic. Currently, templates exist for high-rise buildings, industrial halls, streets (with traffic signs), railway tracks, bridges, and apartments in different shapes. The code is available on GitHub [1].

Validation

The evaluation of SYN3D-LLM focuses on the framework itself and on the plausibility of the point clouds generated from the initial geometric dimensions specified by the combination of the **user prompt** and the **system prompt**. Standard baseline metrics were not applied, since commonly used measures such as Intersection over Union (IoU) or Chamfer Distance (CD) require ground-truth data, which are not available for the proposed SYN3D-LLM framework. The evaluation is exploratory, focusing on plausibility rather than strict geometric accuracy. Instead,

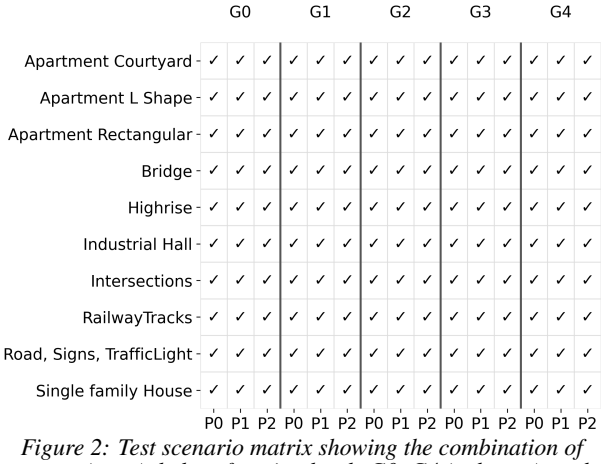


Figure 2: Test scenario matrix showing the combination of scenes (rows), helper function levels G0–G4 (columns), and prompt variations P0–P2 used for evaluation.

we constructed a structured matrix of test scenarios to systematically evaluate the generation process across different

[1] https://github.com/NILSK94/SYN3D-LLM-SYNTHETIC-3D-POINT-CLOUDS-GENERATED-THROUGH-LLM-DISCUSSION

building and infrastructure types in the AECO domain, as illustrated in Figure 2.

Scenes: On the left side of Figure 2 are the scenes we generated, which are described in the code as part of the user input. For example: *A high-rise building with a central core and repeating floors, with an approximate footprint of 18 m × 18 m. It has at least six storeys, with floor levels at z = 0, 3, 6, 9, 12, and 15 m, and a top (roof) level at z = 18 m. The core consists of walls with additional inner walls and columns, windows in the outer walls, and a few corridor doors. Required IFC elements: IfcWall, IfcSlab, IfcColumn, IfcWindow.*

Helper functions G0–G4: At the top of Figure 2, we show the helper functions that are activated during generation. These helper functions act as modular constraints and processing helpers, guiding the LLM by providing code snippets and predefined functions for point cloud generation. The helper functions consist of the objects listed in Table 1.

Table 1: Helper functions for shapes and objects

Helper function	Includes
G0_NoHelper	None
G1_Minimal	Plane Surface
G2_BoxSurface	Rectangle Box
G3_BoxPlusRound	Rectangle Box, Round
G4_Specialized	Rectangle Box, Round, Specialized Objects

Prompt variations P0–P2: In addition, the bottom row of the matrix in Figure 2 shows the prompt variations *P0*, *P1*, and *P2*, which are used to steer the generation and evaluate how prompts influence LLM behavior. While the geometry profiles control which helper functions are available, the prompt variants modify the textual instructions given to the model, ranging from an unconstrained baseline to more specific and robustness-oriented testing. This test is derived from findings on the influence of role allocation in prompt engineering, which can improve LLM performance. The prompt variants are summarized in Table 2.

Table 2: Prompt variants for Apartments as an example

Prompt Variant	Instruction Add-on
P0_Base	None (baseline prompt)
P1_More_Specific	Be explicit: use clear axis-aligned geometry, avoid random noise, and keep openings placed on walls.
P2_Robustness	Robustness: include mild occlusion gaps and slight density variation across elements.

Evaluation metrics

In addition to evaluating the framework, we conducted quantitative analyses, including lightweight checks directly derived from the framework's generated point clouds and log files, to underline the qualitative analysis.

Pass Rate: measures the number of Designer-Critic iterations required until executable Python code is generated. An iteration is considered a pass if the generated code runs without errors and produces valid output. The metric is used across all scenes and configurations.

CountS (Count Successful): verifies each iteration produces a non-empty point cloud, both overall and for required IFC classes where applicable. Reporting the iterations that yield actual point sets rather than empty results.

B–IoU (Bounding-box IoU): measures the overlap between the bounding box of the generated point cloud and the target dimensions specified in the scenario prompt. Values range from 0 to 1, serving as a geometric plausibility indicator for how closely the generated cloud matches the intended dimensions.

Evaluation and Results

For the evaluation, we ran every variation shown in Figure 2. This comprises all scenes and configurations: fifteen runs per scene, each consisting of an initial design step followed by two Designer–Critic iterations. In total, we evaluated more than 300 runs of the framework using an OpenAI GPT-5.2 model, with fixed temperature settings of 0.6 for the designer and 0.4 for the critic. The resulting point clouds were subsequently analyzed. The evaluation starts with a qualitative analysis of the functionality, then moves to the quantitative metrics. To explain the results, we first focus on variations in the helper functions (*G0–G4*) across configurations of the rectangular apartment scene to illustrate how each variation refines the scene design.

Helper function G0–G4 evaluation: Figure 3 shows the result for *G0* (no helper function) and the refinement of the scene across iterations. Over the iterations, the model was able to place doors and windows in plausible positions, so that every room is accessible and has at least one window; The placement of supports remained unreliable. Overall, this highlights the importance of the iterative loop, which progressively refines the design.

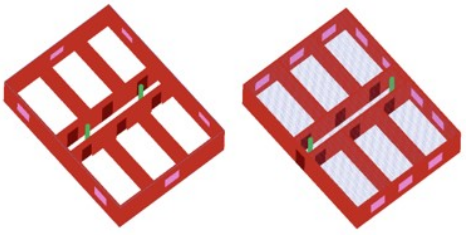


Figure 3: Apartment rectangular layout, prompt variation (P0), helper (G0), iteration 1 (left), iteration 2 (right)

In addition, we analyzed the helper functions provided within the SYN3D-LLM framework by examining their influence when invoked in configurations *G1* through *G4*. For the rectangular apartment layout, this influence was found to be minimal. Nevertheless, the results indicate that the LLM was capable of generating code autonomously without relying on helper functions, and that outputs generated without them tend to exhibit creative variance.

Result pass rate: When considering the pass rate of executable code generation, the results differ across the evalu-

ated helper function variants. Our analysis across all configurations shows that the helper function is crucial for producing executable code in the first iteration. The results are shown in Figure 4. The best pass rate was achieved with the helper function *G2*, which was designed to generate rectangular boxes filled with points on their surfaces. In this case, the function appears to guide the LLM toward producing more structured code output that can be executed directly. In contrast to runs without a helper function (*G0*), the LLM required more iterations to produce executable code.

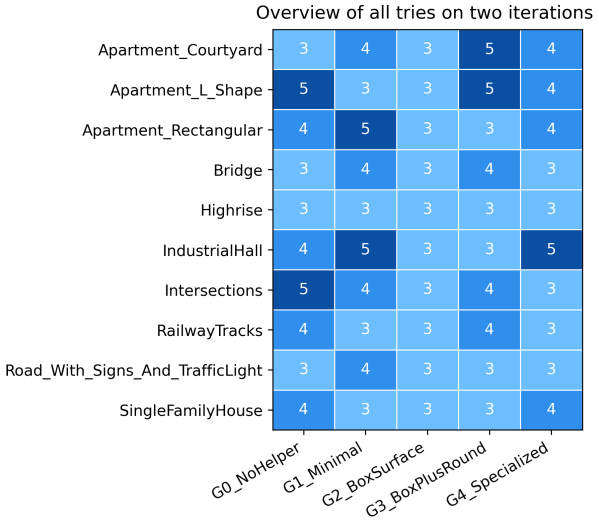


Figure 4: Heatmap of pass rates (two iterations and first design draft required) across scenes and configurations for executable code generation.

CountS: Across all executed iterations in Figure 4, CountS was counted in the Framework, indicating that no empty point cloud outputs occurred in generations because the maximum of 6, by a set default to 3, was not reached.

Prompt variation P0–P2 evaluation: To present the results for prompt variations *P0–P2* (shown for the apartment example in Table 2), we selected the bridge generation scenario to illustrate the influence of prompt variation. The first variant, *P0*, generated the scene shown in Figure 5.

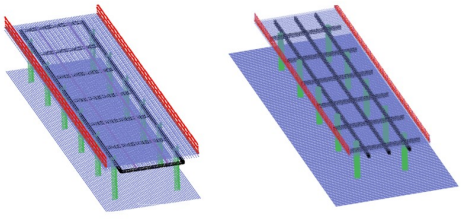


Figure 5: Bridge scene, prompt variation (P0), helper (G0), iteration 1 (left), iteration 2 (right)

Figure 5 shows the baseline prompt *P0* without any additional prompt instructions. While the LLM generates a bridge-like scene, the result is not among the strongest outcomes of the approach, as the geometry remains partially inconsistent and structurally incomplete. In particular, the placement of the supports is not well integrated into a coherent load-bearing bridge structure.

The run with the more specific prompt variant *P1* for the bridge addressed some of the issues observed in *P0*, as shown in Figure 6. The supports are better aligned with the bridge's load-bearing structure, and the LLM generates more detail, particularly in the cut and in the railings or other load-bearing elements. In addition, the model demonstrates the ability to introduce openings in the railings and to incorporate random noise into the generated point clouds. A specific noise setting is also available in the SYN3D-LLM GUI.

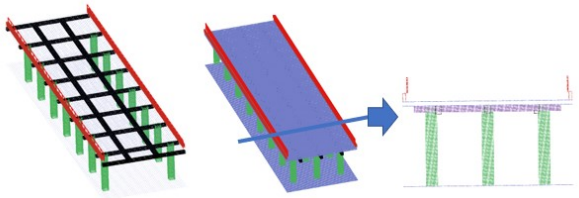


Figure 6: Bridge scene, prompt variation (P1), helper (G0), iteration 1 (left), iteration 2 (middle), cut of iteration 2 (right)

The results obtained with the robustness-oriented prompt variant *P2* are shown in Figure 7. The outcome clearly indicates that the LLM followed the instructions by introducing mild occlusion gaps and slight density variations across elements in the generated point cloud. Beyond improving the baseline prompt's arrangement of supports and load-bearing structure in the generated bridge scenes, *P2* also leads to greater object variation. For example, the LLM attempts to introduce foundations at the ends of the bridge, which are not yet properly connected to the main structure. In addition, the bridge girders beneath the deck are modeled as I-beams, representing substantial structural detail. This can be seen in the cut on the right side of Figure 7. Overall, these results demonstrate that prompt variation can produce meaningful improvements and more plausible outputs within the SYN3D-LLM framework.

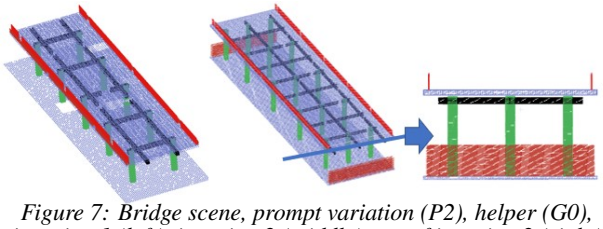


Figure 7: Bridge scene, prompt variation (P2), helper (G0), iteration 1 (left), iteration 2 (middle), cut of iteration 2 (right)

Prompt variant P2 realism effects: Figure 8 presents a qualitative comparison between a synthetic bridge point cloud generated using the prompt variant *P2* (left) and a real-world point cloud acquired with a BLK2GO mobile laser scanner (right). In both cases, mild occlusion gaps and line-like sampling patterns are observable. As illustrated in Figure 7, spatially varying point densities occur across structural elements, particularly along the bridge deck, as well as occlusion-induced shadow regions on the

ground behind supporting columns. In the synthetic example, these effects are generated through the coded point cloud. In real scans, similar characteristics arise from viewpoint dependencies, scanner motion, and the partial occlusions of mobile laser scanning. This visual analysis indicates that SYN3D-LLM generates geometric imperfections like those observed by mobile laser scanners, without explicitly simulating the sensor behavior.

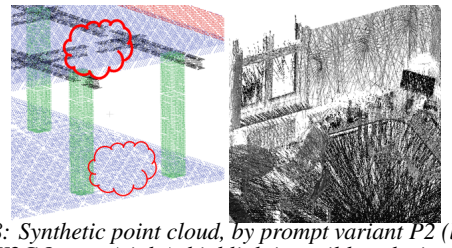


Figure 8: Synthetic point cloud, by prompt variant P2 (left) and a BLK2GO scan (right), highlighting mild occlusion gaps, line-like sampling patterns.

Geometrical check B–IoU: The radar plot in Figure 9 shows how closely the generated synthetic point clouds match the scene dimensions specified in the prompt. Overall, the alignment of the bounding boxes with the target dimensions is minimal.

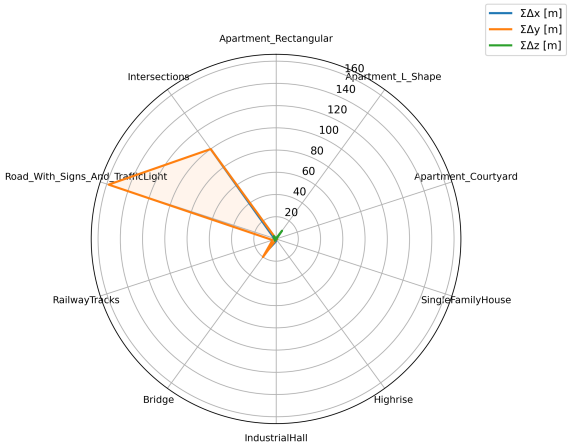


Figure 9: Radar plot showing B-IoU along each axis, comparing generated point clouds to target scene dimensions.

The LLM struggled to generate street scenes with signs and traffic lights, as well as intersection scenes. This can also be observed in the generated data. Figure 10 shows that roads and intersections can be generated, but that there are still major issues in representing these scenes correctly.

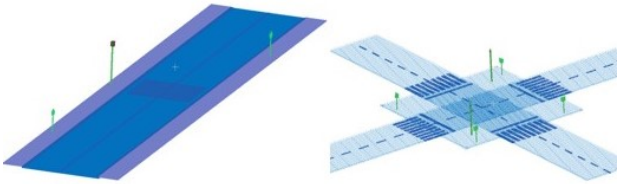


Figure 10: Generated road scene with signs and zebra crossing (left), intersection scene (right).

Semantic labels: The SYN3D-LLM framework stores labels for individual IFC objects as RGB color values in the point cloud. The input scene includes information on the required and permitted IFC classes. The evaluation compares the stored RGB color values in the point cloud with the required and permitted IFC classes specified in the input prompts. The results show a 100 % match for the required and permitted IFC classes explicitly specified in the input prompts. This verifies class presence only and does not assess instance-level correctness or spatial relationships.

Computational cost (Tokens): While the qualitative results demonstrate that iterative refinement substantially improves the structural plausibility and semantic consistency of the generated point clouds, this improvement comes at an increased computational cost.

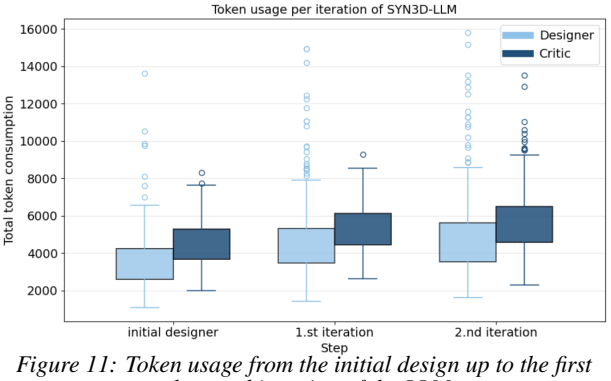


Figure 11: Token usage from the initial design up to the first and second iteration of the LLMs.

Figure 11 shows the token usage per iteration. The results show an increase in the number of tokens required from the initial generation to the first and second iterations. Overall, the Critic needed more tokens to evaluate the design. The round dots in the plot indicate outliers, which are typical for complex tasks.

Discussion

The aim of the paper was to investigate whether an iteration-based, self-discussing framework is capable of generating semantically labeled synthetic point clouds and to examine how reliably the framework follows its input prompts. The results of the proposed SYN3D-LLM framework demonstrate that it successfully generates semantically labeled point clouds. The analysis indicates that the generated point clouds capture the scene's structural elements, but precise geometric alignment is limited in evaluation because the generation is performed without relying on ground-truth data.

The helper functions tested in the framework primarily improve the framework's pass rate and reduce the APIs costs. We can therefore conclude that integrating helper functions is beneficial. But there are limitations, the helper functions may also constrain the LLMs independent creative freedom in generating scenes. Same applies to the predefined input prompts in the framework. Hallucinations are difficult to assess. We should note that the bounding boxes examined using the geometric B-IoU show deviations for certain classes, such as bridges and roads. The

labels indicate that the LLM did not incorporate any new objects into the generated data without being prompted to do so. While semantic hallucinations were not observed, minor geometric deviations in dimensioning cannot be fully ruled out, as they would not be captured by this method. When interpreting the experiments carried out and using the proposed framework, several limitations must be taken into account. The framework was not developed with the aim of generating point cloud data that faithfully reproduces real scan conditions. The absence of an explicit sensor model may introduce a synthetic-to-real gap and limits generalisation to real-world data. The evaluation shows scan patterns that reflect some essential characteristics of real data acquisition, but we need to implement it in future developments. The framework focuses on the variance of point cloud generation. The paper shows that the challenge is not data generation alone, but the controlled, systematic variation of data. Even with model-based approaches, the variance remains in the initial models. In contrast, the variance in SYN3D-LLM is constrained by user prompts, creating a wider variation space.

In particular, for evaluating post-processing steps in point cloud pipelines. Algorithmic methods can be applied to the generated data from SYN3D-LLM and evaluated on constrained data, for example, in the context of ceiling reconstruction.

Outlook

Now we demonstrate our generated data in a short:

Exploratory qualitative case study: that uses the Scan-to-BIM reconstruction approach from Zbirovskỳ and Nežerka (2025) to reconstruct ceilings. We have not yet achieved any results for walls or other components. The SYN3D-LLM framework still needs improvement in this regard. For the Cloud2BIM approach to work, a density of 0.002 meters should be used during generation. The results can be seen in Figure 12. The reconstructed lower ceiling matched the point cloud accurately, but the upper ceiling did not.

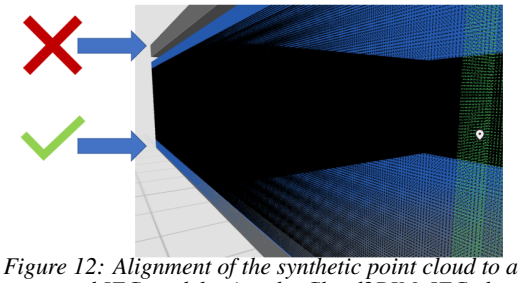


Figure 12: Alignment of the synthetic point cloud to a reconstructed IFC model using the Cloud2BIM. IFC elements shown in grey, point cloud shown in blue, black, and green.

Future work should investigate stopping criteria for controlled refinements in the iteration process and for optimising the computational cost. Furthermore, agent-based approaches are being considered for future development, similar to the Model Context Protocol approach for IFC-based modeling by Nithyanantham et al. (2025). Our next approach could leverage tools that parameterize different point cloud types, enabling more detailed data generation based on user inputs.

Conclusions

This paper presents the SYN3D-LLM framework and demonstrates the potential of synthetic point cloud generation via self-discussing LLMs.

Regarding **RQ1**, the evaluation shows that the framework reliably generates executable code across all tested configurations. The results indicate that helper functions, particularly the G2 configuration, improve first-iteration pass rates. Configurations without helper functions (G0) require additional iterations but still converge to valid outputs within the allowed number of tries.

Regarding **RQ2**, the generated point clouds show plausible structure, but limited geometric precision in complex scenes. Semantic accuracy reached 100% for the required IFC classes specified in the input prompts. But the infrastructure scenes, such as roads and intersections, remain challenging, with larger geometric deviations observed.

Overall, SYN3D-LLM represents a promising approach for generating synthetic point clouds at scale, variations, and for testing reconstruction workflows in the AECO domain. Further optimization is needed, particularly for complex infrastructure scenarios.

References

Abreu, N., Pinto, A., Matos, A., and Pires, M. (2023). Procedural point cloud modelling in scan-to-BIM and scan-vs-BIM applications: A review. ISPRS Int. J. of Geo-Information, 12(7):260.

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al. (2023). GPT-4 technical report. arXiv:2303.08774.

Achlioptas, P., Diamanti, O., Mitliagkas, I., and Guibas, L. (2018). Learning representations and generative models for 3D point clouds. In ICML, pages 40–49. PMLR.

Du, C., Esser, S., Nousias, S., and Borrmann, A. (2026). Text2BIM: Generating building models using a large language model-based multiagent framework. J. of Computing in Civil Engineering, 40(2):04025142.

Gu, X., Chen, M., Lin, Y., Hu, Y., Zhang, H., Wan, C., Wei, Z., Xu, Y., and Wang, J. (2025). On the effectiveness of large language models in domain-specific code generation. ACM Trans. on Software Engineering and Methodology, 34(3):1–22.

He, S., Ding, H., Jiang, X., and Wen, B. (2024). SegPoint: Segment any point cloud via large language model. In ECCV. Springer.

He, Z., Wang, Y.-H., and Zhang, J. (2023). Generative AIBIM: An automatic and intelligent structural design pipeline integrating BIM and generative AI.

Hong, Y., Zhen, H., Chen, P., Zheng, S., Du, Y., Chen, Z., and Gan, C. (2023). 3D-LLM: Injecting the 3D world into large language models. NeurIPS, 36.

Liu, D., Huang, X., Hou, Y., Wang, Z., Yin, Z., Gong, Y., Gao, P., and Ouyang, W. (2024). Uni3D-LLM: Unifying point cloud perception, generation and editing with large language models.

Luo, J., Yang, Z., Shi, P., and Ye, Q. (2026). Indoor scan-to-BIM automation: From mobile perception to 3D building modelling. Automation in Construction.

Ma, J. W., Han, B., and Leite, F. (2021). An automated framework for generating synthetic point clouds from as-built bim with semantic annotation for scan-to-bim. In 2021 winter simulation conference (WSC), pages 1–10. IEEE.

Nithyanantham, B. K., Sesterhenn, T., Nedungadi, A., Garijo, S. P., Zenkner, J., Bartelt, C., and Lüdtke, S. (2025). MCP4IFC: IFC-based building design using large language models. arXiv preprint arXiv:2511.05533.

Noichl, F., Braun, A., and Borrmann, A. (2021). BIM-to-scan for scan-to-BIM: Generating realistic synthetic ground truth point clouds based on industrial 3D models. In EC3, volume 2, pages 164–172. European Council on Computing in Construction.

Paananen, V., Oppenlaender, J., and Visuri, A. (2024). Using text-to-image generation for architectural design ideation. Journal of Architectural Computing, 22.

Pierdicca, R., Mameli, M., Malinverni, E. S., Paolanti, M., and Frontoni, E. (2019). Automatic generation of point cloud synthetic dataset for historical building representation. In International Conference on Augmented Reality, Virtual Reality and Computer Graphics, pages 203–219. Springer.

Ploennigs, J., Berger, M., Wortmann, T., Kirchner, J., Beetz, J., Roitberg, A., Menzel, K., and Ommer, B. (2025). Building foundation models - potentials, challenges and research directions for using LLM and LVM in AEC. In EC3.

Poldrack, R. A., Lu, T., and Beguš, G. (2023). AI-assisted coding: Experiments with GPT-4. arXiv:2304.13187.

Stefanelli, M. G., Jain, A. R., Jalui, S. K., and Agapaki, E. (2022). Synthetic point cloud generation for class segmentation applications. arXiv preprint arXiv:2205.03738.

Tang, Y., Han, X., Li, X., Yu, Q., Hao, Y., Hu, L., and Chen, M. (2024). MiniGPT-3D: Efficiently aligning 3D point clouds with large language models using 2D priors. In ACM Multimedia.

Wei, Y. and Li, X. (2025). Text-to-code generation for modular building layouts in building information modeling. arXiv preprint arXiv:2509.23713.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. (2015). 3D shapenets: A deep representation for volumetric shapes. In CVPR, pages 1912–1920.

Xiao, A., Huang, J., Guan, D., Zhan, F., and Lu, S. (2022). Transfer learning from synthetic to real lidar point cloud for semantic segmentation. In Proceedings of the AAAI conference on artificial intelligence, volume 36, pages 2795–2803.

Xu, R., Wang, X., Wang, T., Chen, Y., Pang, J., and Lin, D. (2024). PointLLM: Empowering large language models to understand point clouds. In ECCV. Springer.

Xu, Y. and Stilla, U. (2021). Toward building and civil infrastructure reconstruction from point clouds: A review on data and key techniques. IEEE J. of Selected Topics in Applied Earth Observations and Remote Sensing, 14:2857–2885.

Yang, S., Liu, J., Zhang, R., Pan, M., Guo, Z., Li, X., Chen, Z., Gao, P., Li, H., Guo, Y., et al. (2025). LiDAR-LLM: Exploring the potential of large language models for 3D LiDAR understanding. In AAAI, volume 39.

Ye, Y. and Gao, W. (2024). LLM-PCGC: Large language model-based point cloud geometry compression. arXiv preprint arXiv:2408.08682.

Yue, H., Wang, Q., Huang, L., and Zhang, M. (2025). Enhancing point cloud semantic segmentation of building interiors through diffusion-based scene-level synthesis. Automation in Construction.

Zbirovskỳ, S. and Nežerka, V. (2025). Cloud2BIM: An open-source automatic pipeline for efficient conversion of large-scale point clouds into IFC format. arXiv preprint arXiv:2503.11498.

Zheng, J. and Fischer, M. (2023). Dynamic prompt-based virtual assistant framework for BIM information search. Automation in Construction, 155:105067.

Zhong, C., Yi'an Shi, L., and Wang, L. (2024). AI-enhanced performative building design optimisation and exploration. In Computer-Aided Architectural Design Research, volume 1.

Zou, Y., Liang, T., Chen, W., Ren, Z., and Wen, Y. (2026). A BIM-derived synthetic point cloud (SPC) dataset for construction scene component segmentation. Data, 11(1):16.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

BIBIMBAP: A Benchmark for Instructional BIM-Based Automated Programming

Clemens Kujat[1], Bharathi Kannan Nithyanantham[2], Stefan Telgmann[1], Tobias Sesterhenn[3]
Jörn Plönnigs[1], Stefan Lüdtke[2], Christian Bartelt[3]
[1]AI for Sustainable Construction, University of Rostock, Germany
[2]Institute for Visual and Analytic Computing, University of Rostock, Germany
[3]Institute for Software and Systems Engineering, TU Clausthal, Germany

Abstract

Large Language Models (LLMs) are emerging as natural-language interfaces for Building Information Modeling (BIM), successfully achieving information retrieval, but no benchmark exists to evaluate their ability to edit Industry Foundation Classes (IFC). We introduce BIBIMBAP, a Text-to-BIM benchmark with 100 curated tasks covering Create-Read-Update-Delete (CRUD) operations across spatial, geometric, topological, numeric, and conceptual categories. Each task includes a natural-language prompt, an IFC model, expected structured outputs, and test scripts for automated evaluation. Baseline results show limited performance (best: 50.2%) and frequent failures in constraint preservation and spatial reasoning, with some correct outputs achieved despite flawed reasoning.

Introduction

Architectural planning can be seen as an iterative process in which design ideas develop through repeated cycles of interpretation and action rather than through linear refinement (Schon and Wiggins, 1992). This process ranges from fundamental conceptual revisions to detailed modifications of façades, individual spaces, or technical components. In BIM-based planning (Building Information Modeling) with computer-aided design (CAD) tools, this is reflected in a dynamic workflow in which individual elements or component groups, such as walls, slabs, and openings, are iteratively created, moved, deleted, or duplicated to explore alternatives of design trade-offs. The resulting solutions aim to satisfy functional and architectural requirements and comply with regulatory constraints.

Several approaches have been proposed to automate the design workflow. More recently, LLM-based approaches have been investigated that enable Text-to-BIM generation, e. g., Du et al. (2024); Deng et al. (2025). These approaches follow an *automated programming* paradigm in which LLMs generate code from a user prompt to create IFC files. While such one-shot approaches may produce simple models, they do not support the iterative refinement process typical of architectural planning. This has changed with the emergence of tool-augmented LLM frameworks, e.g., using the Model Context Protocol (MCP), which define communication interfaces that allow LLMs to access Application Programming Interfaces (APIs) that tools often provide. LLMs thereby understand which functions are available and how to use them. This unlocks LLMs' access to single CAD tools themselves and, more importantly, enables iterative Text-to-BIM workflows.

An example of such a future dialog-oriented assistive workflow is shown in figure 1. Within this agent-based Text-to-BIM workflow, planners can iteratively develop design proposals, refine layouts, or elaborate specific design details through natural language interaction. The LLM can not only provide suggestions but also execute incremental modifications to the building model directly. Over successive design revisions, planners can gradually converge toward a coherent design, reflecting the iterative nature of established architectural planning practices.

However, an iterative design process requires new levels of spatial and contextual understanding by the LLM, because it needs to understand the current design context from the targeted object, design intents, and causal implications. It is debated in the literature whether LLMs exhibit these robust geometric, spatial, topological, and numerical reasoning capabilities required for engineering (Dong et al., 2025; Ploennigs et al., 2025) and architecture concepts (Ching, 2023). Therefore, it is necessary not only to demonstrate what may be possible but also to identify the current limitations of Text-to-BIM systems and to determine whether these limitations stem from the underlying LLMs or from system-level constraints.

Benchmarks are an important instrument for uncovering these limitations and for establishing baselines against which future research can measure improvements. As discussed, end-to-end test do not serve well the iterative refinement approach as they do not ensure to cover all actions, particular update actions. Therefore, we propose to *benchmark the underlying atomic actions*, as their reliable execution is a prerequisite for any iterative Text-to-BIM workflow. Evaluating whether LLMs can execute individual CRUD actions correctly provides: (i) relevant insights into where and why LLMs fail with simple actions; (ii) allow to recombine those actions into iterative workflows; and (iii) avoid instable actions in such workflows.

The state-of-the-art analysis in the next section will show that, to the best of the authors' knowledge, no LLM benchmark exists that systematically evaluates all CRUD actions (Create, Read, Update, and Delete) on IFC models. To address this gap, we present a **B**enchmark for **I**nstructional **BIM**-**B**ased **A**utomated **P**rogramming (**BIBIMBAP**), which is systematically designed to eval-

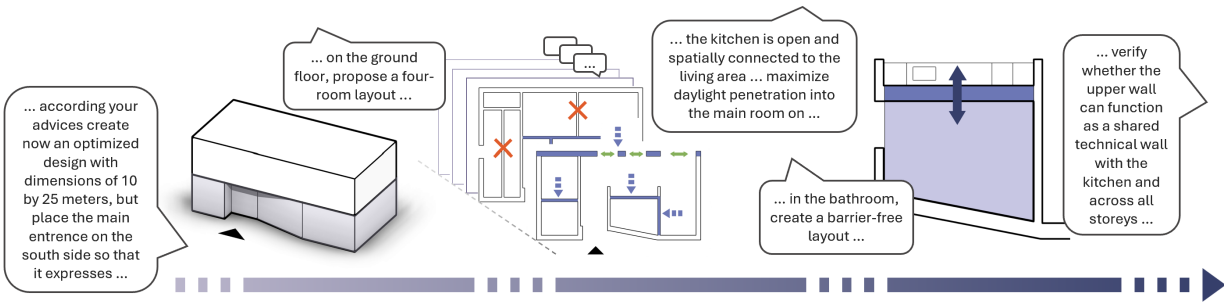


Figure 1: Example of an agentic Text-to-BIM process. Through stepwise interaction with the planner, the LLM provides suggestions, plans changes, and executes the actions in the model, without the planner using the CAD tool itself.

uate Text-to-BIM systems across a curated suite of BIM tasks spanning the full CRUD spectrum. BIBIMBAP emphasizes (i) manipulation tasks that require spatial/geometric and topological reasoning within IFC models, (ii) structured outputs for retrieval tasks that enable direct comparison to ground truth, and (iii) automated tests and evaluation metrics that diagnose where systems fail (e. g., existence, location, dimension, and integrity constraints) without relying on human interpretation. By providing an automated benchmark with tasks, reference IFC models, and executable tests, BIBIMBAP aims to make Text-to-BIM research more reproducible and to support objective comparisons of both language models and the BIM tool interfaces they operate through. Due to its atomic nature, the supported BIM tool interfaces, can be low level frameworks, tool APIs as well as MCPs to test all LLM-based automated programming approaches. Therefore, BIBIMBAP's contributions are:

- A set of Text-to-BIM tasks spanning the full CRUD operations and multiple reasoning categories.
- A prompt grammar (operator/operand/context specifiers) that standardizes task formulation and supports systematic analysis and extensibility.
- An automated pipeline with procedural tests and fine-grained evaluation metrics (existence, location, dimensions, correctness, and IFC integrity constraints).
- Baseline experiments: A validation of the benchmark across different LLMs.

Related Work

This section reviews existing benchmarks and systems for natural-language interaction with BIM, with a focus on how they evaluate retrieval and manipulation tasks. While a growing body of work addresses isolated aspects of Text-to-BIM, like information retrieval, specific generation scenarios, or object recognition, there is currently no widely adopted benchmark that systematically evaluates *all* BIM CRUD operations within a unified framework. Existing benchmarks can be distinguished into: information retrieval benchmarks, datasets related to creation and modification, and end-to-end BIM CRUD approaches. BIBIMBAP is positioned with respect to the remaining gaps across these categories.

Information Retrieval Benchmarks

IFC-Bench-v1 provides curated IFC reference models paired with natural-language queries to assess BIM comprehension and reasoning in an agentic NL-to-IFC retrieval setting, primarily covering the *Read* operation (Hellin et al., 2025). Tasks are grouped according to whether the requested information is directly, indirectly, or insufficiently available in the model. The released dataset lacks explicit class annotations, which requires manual re-categorization, and many tasks rely on open-ended answers evaluated via an LLM-as-a-judge approach, which can be costly and unreliable for free-form responses (Wei et al., 2024). In contrast, our benchmark emphasizes structured outputs, explicit evaluation criteria, and a task taxonomy spanning all CRUD operations and multiple reasoning perspectives (e. g., geometric and topological), enabling finer-grained and reproducible evaluation.
BIMCoder (Liu and Chen, 2025) introduces 1,680 retrieval tasks grouped into object-, location-, and relationship-based queries, which largely correspond to basic retrieval scenarios in our taxonomy. Its evaluation is based on exact matching of generated IFCQL queries and execution accuracy. Other retrieval-focused datasets include *BIN-LQ* (Wang et al., 2022), which contains 2,065 direct questions about element properties or quantities, and *BIM-NLQ* (Yin et al., 2023), which supports ontology-aided querying but lacks explicit ground-truth answers, limiting quantitative evaluation. *BuildingQA* (Mulayim et al., 2025) addresses natural-language question answering over building knowledge graphs represented in RDF/TTL format, but it operates on a different data representation and does not target IFC-based manipulation.

Further AI-Driven BIM Benchmarks and Datasets

Beyond retrieval, several benchmarks address aspects of BIM creation and modification without focusing on Text-to-BIM interaction. *IFCNet* provides single-entity IFC files spanning many IFC classes, enabling controlled evaluation of geometry- and semantics-based methods, but it targets object-level recognition rather than multi-step authoring or manipulation workflows (Emunds et al., 2021). In as-built reconstruction, *BIMNet* offers a scan-to-BIM dataset with metrics that explicitly capture geometric and topological reconstruction quality (Liu et al., 2025). Al-

though it is not driven by natural language, its emphasis on explicit constraint checking aligns with the requirements of reliable IFC editing. *ResBIM* contributes a synthetic paired dataset of 2D floor plans and corresponding BIM models for reference-based generation (Liang et al., 2026).

BIM CRUD Approaches

Text-to-BIM systems translate natural-language intent into executable modeling actions, often via multi-agent decomposition and tool or API calling, enabling the generation of editable, semantically rich building models (Du et al., 2024; Dong et al., 2025). Recent work frames BIM authoring as a computer-use problem: BIMgent employs multimodal LLM agents to plan and execute GUI-level actions in BIM software (Deng et al., 2025). More recently, MCP-based approaches aim to standardize tool access for LLMs to query and edit IFC models directly, as exemplified by *MCP4IFC* (Nithyanantham et al., 2025).

Despite increasing capabilities across Create, Update, and Delete operations, evaluation remains fragmented. Existing studies typically report task-specific success rates or qualitative outcomes, but provide limited standardized, fine-grained diagnostics of geometric correctness, topological validity, and constraint integrity across the full CRUD spectrum, motivating the need for a dedicated benchmark.

Benchmark

We introduce **BIBIMBAP**, a benchmark of 100 tasks for evaluating Text-to-BIM systems on both information retrieval and BIM model manipulation. Tasks are organized along two dimensions: CRUD operation (Create, Read, Update, Delete) to reflect practical BIM workflows, and reasoning requirement, including direct, spatial-geometrical, topological, numeric-calculative, and conceptual tasks. Beyond evaluating LLM capability, BIBIMBAP also evaluates the Text-to-BIM system itself by checking whether the required BIM operations are available and exposed in a way that models can use reliably. To support broad adoption, the benchmark framework provides a unified interface that allows easy integration of different system backends, including direct code-generation pipelines, MCP-based tool interfaces, and custom implementations. The evaluation operates on native IFC, for both input and output, as it is an open standard that most proprietary BIM systems can import and export, making BIBIMBAP applicable to any Text-to-BIM system regardless of implementation details.

Benchmark Architecture

Each task consists of (i) a natural-language prompt, (ii) a task-specific IFC file containing only the elements needed for the task, (iii) a Python test script that computes evaluation metrics, and (iv) for retrieval tasks, a structured output schema. The IFC files were created specifically for individual tasks and include only the information needed to solve them. At the same time, they are designed so that the task is not trivial; for example, retrieval tasks may include multiple competing candidate objects. Figure 2 illustrates the execution flow. Tasks are executed independently by design, as they represent atomic steps within an iterative design process, such as creating a single wall, inserting a door, moving one element, or retrieving one quantity. This is a deliberate methodological choice: the benchmark does not aim to evaluate the solving of complete multi-step design objectives, such as generating an entire house from a single prompt, but rather the model's ability to reliably perform individual, well-defined atomic actions that are prerequisites to reliably fulfill multi-step objectives.

For automated evaluation, the system prompt discourages the LLM from asking clarification questions. The benchmark framework allows users to configure both the evaluated LLM and the number of samples per task. All model outputs are logged, and modified IFC files are stored for Create, Update, and Delete tasks. Evaluation is deterministic and depends on the task type: retrieval tasks expect structured outputs from the LLM with explicit schemas for direct comparison with ground truth, avoiding LLM-as-a-judge evaluation Hellin et al. (2025). Create, Update, and Delete tasks use task-specific test scripts written in Python, using IfcOpenShell, which analyze the original and modified IFC files to compute the metrics. Each test script was validated by a second person to assure correctness.

Task Categorization

As Text-to-BIM systems do not fail uniformly across different tasks, we use distinct categories to systematically separate different error modes, enabling not only to diagnose that tasks fail, but rather understand failure origins. The benchmark is therefore structured into five main categories: direct tasks, spatial-geometrical reasoning, topological reasoning, numeric-calculative reasoning, and conceptual understanding. Each category contains a total of 20 tasks. The first three categories are subdivided according to the four CRUD operations, with five tasks assigned to each one, while the numeric-calculative and conceptual understanding categories consist of 20 retrieval tasks each. By structuring the benchmark along these categories, we make explicit which aspects of BIM understanding current LLM-based systems struggle with, rather than reporting a single aggregated performance score, thereby enabling to assess a models specific strengths and limitations.

Task Complexity

The general difficulty of individual task categories results from the information provided in the prompt. Direct tasks contain explicit information about the targeted elements, so that the LLM can identify them concretely by their ID. Geometric-spatial, topological, numeric-calculative and conceptual tasks contain a prompt component with implicit information about the targeted elements. Based on a contextualized description (e.g. "element that is enclosed by"), the LLM has to derive the targeted elements through model properties and spatial relation-

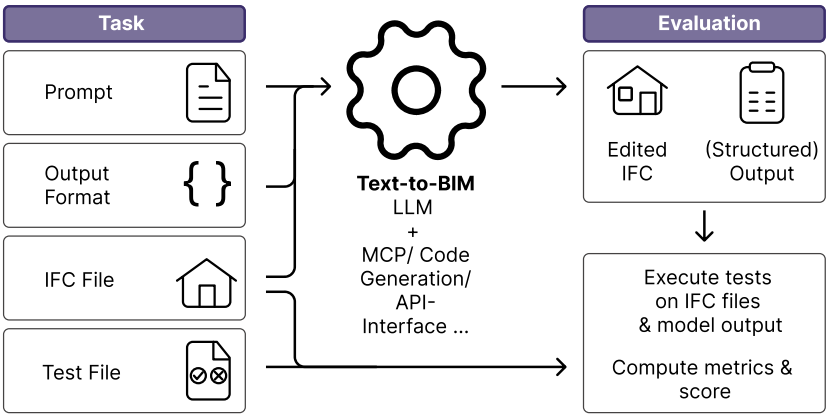


Figure 2: Benchmark architecture: the Text-to-BIM (LLM) takes a prompt and an IFC file (with an output structure for retrieval tasks) and produces an edited IFC file and optional structured output, which are evaluated using the test file and the original IFC model.

ships. Across all CRUD operations, domain-specific contexts, properties and expressions are applied to address realistic and practice-oriented tasks. This is particularly evident in numeric-calculative and conceptual tasks: Numeric-calculative tasks require mathematical reasoning and domain-specific knowledge for their solution, while conceptual tasks require the capacity to assess abstract architectural relationships.

Direct Tasks

These tasks test LLM capabilities for basic BIM operations and examine whether fundamental planner intentions can be understood and correctly executed without requiring complex reasoning. Each prompt includes all required details (e. g., element IDs, coordinates, and dimensions), so the task can be solved directly, as all information for understanding the scene is explicitly provided in the prompt. Examples include moving objects within a local coordinate system, deleting objects specified by explicit element IDs, and adding objects with numerically defined dimensions, placement, and orientation (e. g., ‘‘Add a wall of length 15 m starting at (0, −10)‘‘).

Spatial-Geometrical Tasks

Spaces are configurations of surfaces, volumes, and openings whose effects emerge through placement, orientation, scale, and angular relations, and in building design these configurations determine how spaces, functionality, and spatial perception are produced. Spatial-geometrical tasks therefore test whether an LLM can reason about such spatial and geometric relationships. The model must identify the correct elements based on shape cues (e. g., ‘‘the room that is not a perfect square‘‘) by computing properties such as lengths, perimeters, or bounding-box dimensions. It should also follow spatial instructions (e. g., ‘‘the door between doors X and Y‘‘) by understanding relations such as between, next to, or opposite. These tasks assess whether an LLM actually reasons about the spatial alignment and require the model to construct an internal spatial representation and derive properties such as relative position, betweenness, and shape characteristics.

Topological Tasks

Topological reasoning tasks test whether an LLM can understand functional and technical relationships such as adjacency, enclosure, openings, connectivity, and contact. In building design, these topological relations generate functionality and spatial organization, for example, whether a wall fully separates a kitchen from a living room or defines a shared space. While some relations are explicitly stored in IFC (e. g., a column assigned to the space it belongs to), many important relations are not represented directly and must be inferred through higher-level topological reasoning. These tasks, therefore, require the model to reason over collections of elements, such as identifying how multiple walls jointly form an enclosure or determining which walls separate interior from exterior. Examples include creating a connection between an interior space and the outside by adding a window.

Numeric-Calculative Tasks

In building design and planning, a robust understanding and mathematically correct handling of quantities, areas, volumes, and perimeters is essential, as building regulations frequently specify numerical constraints, and such quantities form the basis for cost estimation, and rentable area calculations used in economic evaluations. These tasks require the LLM to retrieve numeric properties from IFC elements and perform multi-step calculations to produce a final result. They often involve selecting the correct elements (by ID or filter), choosing the appropriate measurement (net vs. gross), aggregating across multiple elements, and applying deductions such as subtracting openings from wall surface areas. As such, they test multi-step reasoning rather than direct value lookup. For example, ‘‘Provide the total net wall area of walls A, B, and C‘‘ requires extracting each wall’s surface area after accounting for openings and then summing the values.

Conceptual Tasks

Architectural design builds on an understanding of how architectural concepts emerge from the interplay of spatial, geometric, and topological properties. For example,

functionality is often produced through the organization and hierarchization of spaces, while spatial perception can arise from geometric form, placement, and light. These tasks, therefore, require an understanding of formal and functional architectural concepts that emerge from spatial configuration rather than from explicitly stored properties. Guided by principles of architectural perception and design (Ching, 2023), we distinguish four categories: spatial–geometric relationships, organizational structures, openings and light, and ordering principles. Solving these tasks often requires combining multiple types of reasoning described earlier. For instance, identifying a corridor requires understanding orientation, proportion, and connectivity patterns, even if the space is not explicitly labeled.

Prompt Construction

To ensure a high CRUD task coverage and systematic testing of LLM reasoning capabilities, all prompts follow a structured grammar (figure 3) that allows to create variants of prompts. Each prompt includes an *operator* that denotes the intended operation (e. g., create, rotate, provide); an *operand* that denotes the target building elements or the desired information subject to the task; *operation specifiers* that constrain the request (e. g., required properties, relations, or filters), and *context specifiers* that specify the relevant context elements either implicitly or explicitly. This allows similar tasks to address different reasoning categories by replacing or omitting individual parts of the prompt. For example, a task could be direct with the explicit context specifier "with <id>", while by using the context specifier "which is parallel to wall with <id>", the task would be a spatial-geometrical one.

Evaluation Metrics

To quantify performance, we define evaluation metrics. Because tasks differ, it is impractical to define a fixed set of metrics for all of them. Therefore, we define a set of metrics for each CRUD category. The set of metrics and their assignment to task categories is described in table 1. For each task t, we calculate $\mathrm{score}(t)$, and the overall benchmark score is the average of $\mathrm{score}(t)$ over all tasks.

$$\mathrm{score}(t) = \frac{\text{no. of metrics satisfied for } t}{\text{no. of metrics evaluated for } t}.$$

These metrics provide more insights into where an LLM fails, rather than simply determining whether a task was completed correctly. For example, the task "Create a door with the size of 2 by 1 meters in the wall" is evaluated using the metrics *Object exists*, *Correct location*, *Correct dimension*, and *Integrity assured* (a corresponding opening must exist that voids the wall and is filled by the door). The integrity metric is task-specific. For example, inserting a wall to enclose a room requires topological relationships to the other walls. Deleting an element requires that IFC relationships (e. g., *IfcRelContainedInSpatialStructure*) are also removed. Furthermore, the integrity metric verifies that no clashes are introduced, the spatial hierarchy is preserved, and non-target elements remain unchanged. Retrieval tasks use *Correct answer*, comparing the structured output against ground truth.

Table 1: Evaluation metrics and applicable task categories.

Metric	Description	Categories
Object exists	Required object exists in the edited IFC file	Create
Object does not exist	Referenced object does not exist in the edited IFC file	Delete
Correct location	Object of interest is in the correct location	Create, Update
Correct dimension	Object of interest has the correct size	Create, Update
Integrity assured	No constraints are violated	Create, Update, Delete
Correct answer	Model provided the correct answer	Retrieval

Experiments

We evaluate Claude Opus 4.6, GPT-5.4, Gemini 3 Flash, GPT-5.2, GPT-5 Mini, and GPT-4.1 in a minimal tool-based setup implemented with the LangChain framework. Each model is given access to a single tool, `execute_ifc_code`, which runs Python code with IfcOpenShell in a sandboxed environment. The models may call this tool multiple times to solve a task. To prevent infinite loops, we limit the number of tool calls at 25 per task. No task-specific examples or hints are provided, which ensures a fair baseline setting. All models use the same system prompt, agent framework, and tool access. To demonstrate that the top-performing models produce consistent results across runs, we run the three newest models (Claude Opus 4.6, GPT-5.4, and Gemini 3 Flash) three times each and report mean scores with standard deviations. The remaining three models (GPT-5.2, GPT-5 Mini, and GPT-4.1) are evaluated with a single run.

Code-Execution Pipeline

The benchmark evaluates LLMs through a code-execution pipeline rather than direct IFC text editing. When the LLM receives a task prompt, it generates Python code that uses the IfcOpenShell library. This code is then passed to the `execute_ifc_code` tool, which runs it in a sandboxed subprocess with the IFC file already loaded as the variable `ifc`. For editing tasks, the LLM must call `commit()` to save the changes. Geometry is modified through the IfcOpenShell API, for example by creating representations with `ifcopenshell.api.run()`. The LLM may make several tool calls in sequence, first to inspect the model and then to apply the required changes. For example, consider the prompt: *''Delete all doors on the first floor.''* The LLM may first generate code to inspect the IFC model and find which storey corresponds to the first floor and which doors belong to it. After receiving the result, it then gener-

Operator	Operand	Operation Specifiers	Context Specifiers	
Create	Door	with the size 2 by 1 meters	in the wall	with id <id>
Rotate	Wall	by +45° around its center	which is parallel	to wall with id <id>
Provide	net floor area	of the room	with id <id>	

Figure 3: Visualization of the prompt structure using the grammar with three examples.

ates a second code snippet to delete those doors and calls `commit()` to save the changes. All the generated code is executed through the `execute_ifc_code` tool.

Our goal is to establish strong baseline results that can serve as a reference point for future improvements. Failures on particular task types highlight open challenges (e. g., geometric reasoning, IFC relationship handling, and BIM-aware representations), while strong performance indicates tasks that current LLMs already support reliably.

Results

Table 2 summarizes the overall model performance with a breakdown by category. Overall, the newer models achieve higher scores; however, the performance still remains low, which shows that Text-to-BIM manipulation is still challenging for an LLM. Claude Opus 4.6 achieves the best overall average score (50.2%), followed by GPT-5.4 (31.3%), Gemini 3 Flash (30.8%), GPT-5.2 (30.5%), GPT-5 Mini (25.0%), and GPT-4.1 (23.9%). The top three models show consistent results across repeated runs. Even the best model solved only about half the tasks on average, which highlights the difficulty of reliably using LLMs for CRUD and fulfilling all evaluation criteria.

Performance differences across categories show clear trends. Direct tasks are relatively easy because they often provide explicit identifiers, coordinates, or dimensions. Spatial-geometrical and topological tasks are harder because they require selecting the correct target elements from context and reasoning about relationships or placements. Numeric-calculative tasks are the most difficult for all models, with consistently low scores. These tasks typically require aggregating information across multiple elements and computing quantities from geometry (e. g., areas, volumes, and totals). Models often produce partially correct code but miss elements, apply an incorrect quantity definition, use wrong unit scale or return an incorrect value. The consistently low scores on numeric-calculative tasks suggest that BIM reasoning remains a major limitation for the used LLMs.

Tool usage patterns differ across models. All models rely heavily on the tool interface to solve tasks. Claude Opus 4.6 makes the most tool calls, averaging around 4.9 per task, followed by Gemini 3 Flash with about 3.5 calls, GPT-4.1 with 2.7 calls, GPT-5 Mini with 2.5 calls, GPT-5.2 with 2.2 calls, and GPT-5.4 with only 1.67 calls. Notably, Claude Opus 4.6 achieves the highest overall score while also using the most tool calls, suggesting that its iterative strategy contributes to its stronger performance.

Table 3 shows the performance of the LLMs by CRUD operation. Create is the easiest operation in the current benchmark, since most creation tasks focus on simple building elements with explicit parameters. Claude Opus 4.6 reaches 81.1% on Create tasks. Retrieve is the most difficult because it requires both reliably identifying the correct target elements and computing or formatting the answer. Update tasks require precise geometric modifications, while Delete tasks can appear easy, but many failures arise from integrity requirements: deleting an element often requires removing related relationships or associated elements (e. g., when a door is deleted, its associated opening also needs to be deleted). As a result, a model may remove the main element but still fail the task if the IFC remains inconsistent.

Figure 4 plots the accuracy across the evaluation metrics and highlights a consistent limitation: *integrity_constraint* is the most challenging criterion for all models. In contrast, the *object_exists* and *object_not_exists* constraints achieve higher scores, indicating that models can often create or delete an entity but still struggle to maintain correct IFC relationships and constraints. Similarly, *correct_location* and *correct_dimensions* remain low for most models, which is critical for practical BIM authoring because a created element is only useful if it is placed correctly and has the correct dimensions. For retrieval tasks, *correct_answer* remains low, suggesting that even when the model can access the relevant elements, it often fails to compute or properly format the final answer.

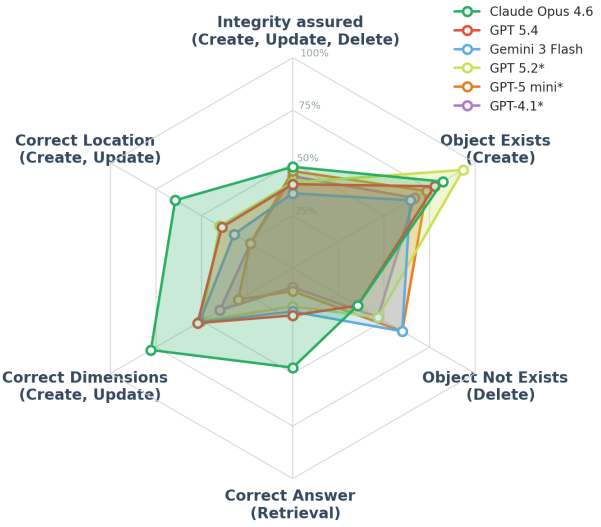


Figure 4: Each model's performance across the six evaluation criteria.

Overall, these results highlight both the promise and the limitations of current LLMs. Current models can of-

*Table 2: Model performance by category and overall average score (%). The top three models are averaged over three runs (± std. dev.). Models marked with * are single-run results.*

Model	Direct (20)	Spatial (20)	Topological (20)	Numeric (20)	Conceptual (20)	Overall Score
Claude Opus 4.6	65.83 ± 0.16	63.89 ± 0.22	53.06 ± 0.19	38.33 ± 0.24	30.0 ± 0.21	**50.22 ± 0.22**
GPT-5.4	62.15 ± 0.40	42.08 ± 0.24	20.56 ± 0.10	26.67 ± 0.22	5.00 ± 0.00	31.29 ± 0.04
Gemini 3 Flash	65.42 ± 10.4	40.97 ± 1.73	21.11 ± 3.37	20.00 ± 5.00	6.67 ± 2.33	30.83 ± 1.91
GPT-5.2*	52.50	34.58	35.42	5.0	25.0	30.50
GPT-5 Mini*	50.94	25.83	25.83	0.0	25.0	25.00
GPT-4.1*	38.33	30.00	31.25	0.0	20.0	23.92

Table 3: CRUD success rate by model (%).

Model	Create	Read	Update	Delete
Claude Opus 4.6	81.1±0.2	47.3±0.3	61.5±0.2	18.9±0.1
GPT-5.4	56.0±0.3	22.4±0.0	50.4±0.3	20.0±0.2
Gemini 3 Flash	45.6±8.4	20.6±1.1	46.7±2.2	37.8±6.9
GPT 5.2*	58.9	18.2	51.1	26.7
GPT-5 Mini*	31.2	10.9	55.5	43.3
GPT 4.1*	29.4	9.0	60.0	36.6

ten perform basic Create and Delete actions, and larger, stronger models benefit from iterative tool use. However, robust BIM manipulation requires more than producing valid code: it also requires correct element selection, correct computations, and consistent IFC integrity constraints. The results show, that no model manages to address all those aspects.

Examples

To illustrate task complexity, we give one example each for success and failure cases. For each, we show the IFC models alongside the corresponding prompts, as well as the resulting IFC models.

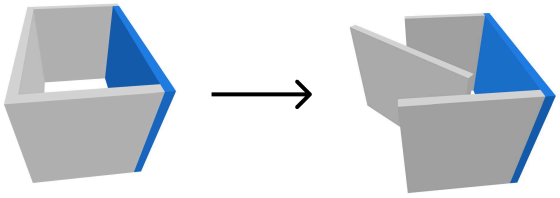


Figure 5: Successfully solved spatial-geometrical task of Claude Opus 4.6

Figure 5 shows a spatial-geometrical task that was successfully solved by Claude Sonnet 4.5. The main challenges are identifying the correct wall to rotate and rotating it about its center while ensuring the correct placement after rotation. The LLM first extracted the direction of the reference wall. It then extracted the directions of the other walls to find the one with the same direction, computed its rotation center, and finally derived the new direction and location. This process required six tool calls in total.

Figure 6 shows an example of a conceptual task. The task is to identify the door that is perceived as the entrance only by reasoning about the architectural concept. GPT-5.4 tried to find keywords in door objects that indicate their

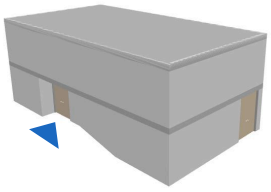


Figure 6: Example of a conceptual task. GPT-5.4 solved it by choosing the right door randomly.

use as the entrance. Because no such keyword was present, the model selected the first door as the answer, which happened to be correct by chance, even though the model was not able to solve the task in a principled way. In other tasks in this category, the LLMs exhibited similar issues and produced errors after making too many tool calls.

Another common failure was incorrect unit handling: models sometimes used wrong units or returned incorrect results despite explicit unit specifications.

Limitations

The current version of the benchmark consists of a limited number of 100 prompts. Future versions aim to expand the task pool to improve statistical robustness and broader coverage. Although the benchmark covers all four CRUD operations, retrieval tasks represent the largest share. This is partly by design: the numeric-calculative and conceptual understanding categories consist exclusively of retrieval tasks, and providing quantifiable, directly evaluable answers for these is itself a contribution of this work. Future versions of the benchmark will increase the proportion of Create, Update, and Delete tasks to achieve a more balanced distribution across operation types.

The IFC files used in the evaluation are generally applicable to other tasks, but they primarily serve as task-specific test environments that include only the elements required for the respective task. We therefore selected small models to keep the benchmark focused on individual tasks and to ensure a clear evaluation. However, this also means the benchmark does not fully reflect realistic BIM authoring settings, which are typically larger and noisier and exhibit higher structural complexity (e. g., more storeys and a wider variety of building elements).

Our tasks are instructions for atomic operations. In real use, users often interact with an AI system over multiple turns to refine requests and refer to earlier actions

(e. g., "Undo the last change" or "Now make that window larger"). Those interaction sequences are not part of this work, as models already fail on simple steps.
The evaluation metrics are detailed, but they do not cover all aspects of the BIM authoring standards. The benchmark focuses on hard constraints (geometry, existence, and relationships), but does not evaluate, e.g. whether element naming conventions are followed. We also do not test whether non-geometric properties are preserved in Update tasks; for example, deleting an element and recreating it can count as a correct update even if properties are lost.

Conclusion

This paper introduces BIBIMBAP[1], a benchmark for evaluating Text-to-BIM systems on IFC-based tasks spanning the full CRUD spectrum (Create, Read, Update, Delete) across multiple reasoning categories (direct, spatial-geometrical, topological, numeric-calculative, and conceptual). We report baseline results for several state-of-the-art LLMs in a tool-based setup. Our results indicate that current models can often complete simple actions when parameters are explicit, but they frequently fail on tasks that require geometric or topological reasoning, multi-step computation, or preserving IFC relationships and constraints. These findings highlight that reliable BIM manipulation remains challenging and that progress should be measured with reproducible tests. Overall, BIBIMBAP provides an automated evaluation framework for Text-to-BIM approaches, identifying failure modes and tracking progress over time. We hope it supports both researchers and practitioners in developing more reliable BIM assistants and motivates future work on BIM-aware LLMs and domain-specific training.

Acknowledgments

This research was partially funded by the German Ministry of Research, Technology and Space (BMFTR).

References

Ching, F. D. (2023). Architecture: Form, space, and order. John Wiley & Sons.

Deng, Z., Du, C., Nousias, S., and Borrmann, A. (2025). BIMgent: Towards autonomous building modeling via computer-use agents. arXiv preprint arXiv:2506.07217.

Dong, Y., Zhan, Z., Hu, Y., Doe, D. M., and Han, Z. (2025). AI BIM coordinator for non-expert interaction in building design using LLM-driven multi-agent systems. Automation in Construction, 180:106563.

Du, C., Esser, S., Nousias, S., and Borrmann, A. (2024). Text2BIM: Generating building models using a large language model-based multi-agent framework. arXiv preprint arXiv:2408.08054.

[1]The benchmark is available at `https://nbharathik.github.io/bibimbap/`

Emunds, C., Pauen, N., Richter, V., Frisch, J., and Van Treeck, C. (2021). IFCNet: A benchmark dataset for IFC entity classification. In EG-ICE, volume 166.

Hellin, S., Nousias, S., and Borrmann, A. (2025). Natural language information retrieval from BIM models: An LLM-based agentic workflow approach. In Proceedings of the European Conference on Computing in Construction (EC3), pages 1–8.

Liang, X., Yabuki, N., and Fukuda, T. (2026). Fully automated synthetic bim dataset generation using a deep learning-based framework. Automation in Construction, 181:106584.

Liu, B. and Chen, H. (2025). BIMCoder: A comprehensive large language model fusion framework for natural language-based BIM information retrieval. Applied Sciences, 15(14):7647.

Liu, Y., Huang, H., Gao, G., Ke, Z., Li, S., and Gu, M. (2025). Dataset and benchmark for as-built BIM reconstruction from real-world point cloud. Automation in Construction.

Mulayim, O. B., Anwar, A., Saka, U. M., Paul, L., Prakash, A. K., Fierro, G., Pritoni, M., and Bergés, M. (2025). Buildingqa: A benchmark for natural language question answering over building knowledge graphs. In buildSys, pages 65–75.

Nithyanantham, B. K., Sesterhenn, T., Nedungadi, A., Peral Garijo, S., Zenkner, J., Bartelt, C., and Lüdtke, S. (2025). MCP4IFC: IFC-based building design using large language models. arXiv preprint arXiv:2511.05533.

Ploennigs, J., Berger, M., Wortmann, T., Kirchner, J., Beetz, J., Roitberg, A., Menzel, K., and Ommer, B. (2025). Building foundation models – potentials, challenges and research directions for using LLM and LVM in AEC. In EC3.

Schon, D. A. and Wiggins, G. (1992). Kinds of seeing and their functions in designing. Design studies, 13(2):135–156.

Wang, N., Issa, R. R., and Anumba, C. J. (2022). Transfer learning-based query classification for intelligent building information spoken dialogue. Automation in Construction, 141:104403.

Wei, H., He, S., Xia, T., Liu, F., Wong, A., Lin, J., and Han, M. (2024). Systematic evaluation of llm-as-a-judge in llm alignment tasks: Explainable metrics and diverse prompt templates. arXiv preprint arXiv:2408.13006.

Yin, M., Tang, L., Webster, C., Xu, S., Li, X., and Ying, H. (2023). An ontology-aided, natural language-based approach for multi-constraint BIM model querying. J. Building Engineering, 76:107066.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

A Semantic Framework for Formalizing Stakeholder Information Requests in Construction

Timur Weilbach-Eyüboglu, Cornelius Preidel, and Simon Vilgertshofer
Hochschule München University of Applied Sciences, Munich, Germany

Abstract

The construction industry relies on answering recurring questions, but required information is fragmented across heterogeneous data sources, including digital models, schedules, and documentation. This paper introduces a semantic request framework implementing stakeholder questions as information objects that capture templates, bindings, constraints, and provenance for re-execution and diagnostics. Reusable question templates derived from an exemplar construction process are aligned with established domain ontologies such as *DiCon*, *ifcOWL*, and *BOT*, then operationalized through parameterized graph query patterns over federated knowledge graphs. Instantiated requests bind templates to project instances, enabling evidence-based evaluation of readiness, approvals, constraints, and resource availability.

Introduction

Construction projects require stakeholders to obtain situation-specific information on demand to make informed decisions. Such information needs arise dynamically, depending on the current project state, the actor's role, and the decision context. Typical questions include *"Is this area ready for the next activity?"*, *"Are the required approvals available?"*, or *"Can this task be released under the current conditions?"*. While their wording varies, the underlying intent of these questions is highly recurrent: across projects and roles, stakeholders repeatedly ask structurally similar questions but expect context-specific, evidence-based answers.

These questions are central to production coordination and decision-making. In Lean Construction, they directly relate to constraint analysis, work packaging, and reliable promise-making (Ballard, 2020; Koskela, 2000). From a situation awareness perspective, they express the information required to perceive, comprehend, and anticipate the state of the work system (Endsley, 1995). As such, they determine whether work can be executed safely, correctly, and without disruption. Despite their importance, answering such questions remains difficult in practice. Applied research and industry-informed studies consistently report that the evidence required to assess process readiness or approval conditions is distributed across heterogeneous systems, maintained by different stakeholders, and represented in various formats (Sacks et al., 2020; Čustović et al., 2025). This observation is further reinforced by practice-driven analyses that we conducted in preparation for this paper in close collaboration with industrial partners, based on structured workshops and expert interviews, in which even routine operational inquiries were found to require manual coordination across BIM models, schedules, quality assurance records, safety documentation, and domain-specific planning tools. As no single system provides a unified representation of process dependencies and current states, situational awareness is fragmented, and decision-making becomes time-consuming and error-prone. The core challenge is therefore not a lack of data, but the absence of a mechanism that consistently translates recurring operational questions into verifiable, cross-system information requests. Although construction projects repeatedly pose the same types of questions, these questions are reformulated anew for each system and context, preventing systematic reuse and comparability of answers.

This paper addresses this gap by proposing a semantic request layer for recurring operational information needs in construction. The approach separates stable question semantics from project-specific bindings, allowing the same underlying request structure to be specified consistently across heterogeneous systems and reused across projects. Three contributions are made: first, a formal representation of recurring stakeholder questions as reusable request templates; second, a binding and constraint mechanism that links these templates to project-specific entities and evidence requirements; and third, a prototype design for operationalizing such requests over federated graph-based project data. Execution-time readiness and work-release decisions serve as the exemplar application. Automated natural-language interpretation is discussed as a prospective interface to this layer, but is treated as future work rather than part of the present evaluation.

Related Work

From a lean production-control perspective, questions such as whether work can be released or whether an area is ready for the next task belong to make-ready planning, constraint analysis, and reliable promise-making (Ballard, 2020; Ballard and Howell, 1998; Koskela, 2000). These decision situations depend on multiple prerequisites spanning workforce, information, materials, equipment, prerequisite work, safety, and external conditions.

At the same time, BIM and construction digital twin research consistently reports that the required evidence is distributed across domain tools, making the problem one of information access rather than data availability (Sacks et al., 2018, 2020; Čustović et al., 2025). Ontology-based approaches address this fragmentation by providing shared semantic foundations. Foundational ontologies such as *ifcOWL* (buildingSMART International, 2025) and the *Building Topology Ontology (BOT)* (Rasmussen et al., 2020) support graph-based representations of building elements and spatial structures. Workflow-oriented ontologies extend this scope toward construction processes, resources, states, and information objects, most notably the *DiCon* ontology suite (Zheng et al., 2021). Related work further specializes these ideas for production control, safety, and digital twin integration (Farghaly et al., 2024; Speiser et al., 2025; Schlenger et al., 2025). While these ontologies provide rich, machine-interpretable models of construction entities, processes, and states, their focus remains on the supply side of information modeling. They describe what data exists and how it can be linked. However, they are not designed to explicitly represent the structure of stakeholder questions or the decision logic for combining evidence from different domains. Competency questions used in ontology development primarily serve validation and illustrative purposes, but they are not formalized as reusable models of operational information needs (Zheng et al., 2021; Sacks et al., 2020). As a result, recurring process-oriented inquiries, such as readiness or approval checks, remain implicit and system-specific. In parallel, significant research has focused on improving information retrieval from BIM models. Early approaches introduced domain-specific query languages and visual query tools that operate directly on IFC or *ifcOWL* schemas, such as *BIMQL* (Mazairac and Beetz, 2013) and *QL4BIM* (Daum and Borrmann, 2014). These languages enable expressive, schema-aware queries for building models. However, they require users to be familiar with underlying data schemas and primarily address model-internal structures, geometric properties, and element relationships. More recent work increasingly uses natural language processing to translate user queries into formal graph queries. Ontology-aided semantic parsers and question-answering systems enable multi-constraint queries over BIM models for non-expert users (Yin et al., 2023; Nabavi et al., 2023). Recent work also explores LLM-based natural-language retrieval directly over IFC-encoded BIM models through agentic workflows (Hellin et al., 2025). These approaches improve access to BIM-centered information, but they generally focus on individual questions rather than on model data or closely related semantic extensions. By contrast, the present work formalizes recurring operational question types as reusable semantic request templates that can be instantiated across project contexts and resolved against federated cross-domain evidence. Thus, the contribution is not a direct NL-to-query interface, but an intermediate semantic request layer that such interfaces could use. To enable cross-system access, research on federated knowledge graphs proposes technical architectures for linking and querying heterogeneous, domain-specific graphs (Teclaw et al., 2024; Pauwels, 2014). Such federated approaches provide a robust infrastructure for cross-domain querying and align well with the digital twin vision. Nevertheless, they primarily address schema alignment and data availability. They do not specify how operational questions should be represented, nor how complex dependencies across federated sources can be captured in a reusable and decision-oriented manner. Users are still required to formulate explicit queries or rely on system-specific interfaces. Across these research streams, a common limitation becomes apparent. Existing ontologies, query systems, and federated graph architectures excel at modeling and accessing construction data, but they do not provide a structured semantic layer for representing recurring stakeholder questions. In particular, operational inquiries that combine spatial, temporal, procedural, and administrative dependencies, such as readiness to release work in lean production systems (Ballard and Howell, 1998), are not supported at this stage. This gap motivates the present work, which introduces a semantic request layer that formalizes recurring operational questions as reusable request templates, links them explicitly to distributed information dependencies, and connects lean work-release logic with existing ontologies and query infrastructures.

Methodology

We objectify stakeholder information needs by representing them as explicit *InformationRequest* entities in a semantic request layer, rather than ad-hoc *queries* across heterogeneous systems. Each request is a persistent, addressable object with an identifier and links to its *RequestTemplate*, bindings, constraints, and provenance, supporting reproducibility and auditability. We refer to (i) the natural-language question, (ii) its formal representation as an *InformationRequest*, and (iii) the corresponding graph-based retrieval specification (e.g., federated SPARQL). Because such questions recur in specific process situations (readiness, approval, compliance, coordination), we elicit reusable request types and associated constraints using an exemplar foundation process chain while keeping the framework process-agnostic; the foundation and readiness examples serve to instantiate the method, not to define its full scope. The methodology therefore proceeds from observed execution-time questions and process checkpoints to formal representations in a request layer that can be aligned with existing ontologies, such as *DiCon*, *ifcOWL*, *BOT*, and digital twin information models. Figure 1 summarizes how an *execution-time question* is mapped to a template, instantiated via placeholder bindings, and specified for graph-based retrieval over federated project data. Unlike NL-to-query approaches, the request layer specifies the decision condition and admissible evidence paths, not a single system-specific retrieval proce-

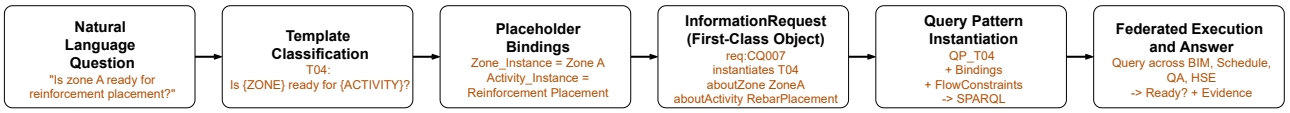


Figure 1: From a natural-language question to an operationalizable graph-based request via the proposed request framework

dure. A future natural-language front-end is envisioned as a candidate-generation layer rather than the locus of semantic interpretation itself. Its role is to rank likely request templates and propose bindings for typed placeholders (e.g., {Zone}, {Activity}, {Entity}) based on the user's utterance and the project context. The semantic request layer defined here then validates these candidates against ontology-based type constraints and admissible relation patterns before execution. Where multiple interpretations remain plausible, the system is intended to request clarification rather than execute an underspecified request.

At the core of the approach is the notion of a *RequestTemplate* (a controlled-natural-language question template). Practitioners may ask many different *execution-time questions* in natural language, but many of these can be understood as instances of a compact set of underlying question types. For example, *"Has the excavation been completed?"*, *"Has the blinding layer been completed?"* and *"Has the reinforcement placement been completed?"* are phrased differently but share the same underlying structure: asking whether a given activity has reached a state of completion. Similarly, *"Has the blinding layer been inspected and approved?"*, *"Has the reinforcement for Footing "F1" been inspected and approved for concreting?"* and *"Has the formwork been inspected and approved?"* conform to a pattern that inquires whether an entity has passed through an approval process in a specific execution context. The methodology, therefore, begins by abstracting from individual wordings to generic question types.

In four workshops with two industry partners (design–build and general contracting), we collected execution-time questions along an exemplar foundation process, clustered them into request types, and iteratively refined template formulations and argument roles with practitioner feedback. The outcome is an initial template library covering completion, approval, readiness, and selected cross-domain request types, designed to be extended through additional templates, mappings, and query patterns. These workshops supported template elicitation and refinement, but they do not constitute a performance validation of the framework.

Each recurring execution-time question type is expressed as a *RequestTemplate*; each template references one or more query patterns (*QP_**) that are intended to become executable after placeholder binding. A template is written in a controlled natural language form with explicit placeholders for variable elements. For instance, a completion template may have the form *"Has **{Activity}** been completed?"*, a tolerance template may be written as *"Is the **{Property}** of **{Component}** within **{Requirement}**?"*, and a readiness template takes the form *"Is **{Zone}** ready for **{Activity}**?"*. The placeholders denote the template's arguments. They are not mere textual gaps but correspond to semantic roles that must be bound to specific project instances, such as activities, components, zones, requirements, or documents. Templates may introduce additional roles where required (e.g., resources, suppliers, time windows, hazards), each of which is typed and described as part of the template metadata. Each template is enriched with an explicit description of its arguments, specifying for each placeholder the type of entity it refers to.

Table 1 shows the resulting *RequestTemplate* representation for a readiness question type (T34), including typed arguments, ontology mappings, and selected template metadata such as flow type and target state.

Table 1: Request template T34 for readiness checking.

Request Template	
Context	Template_ID: T34
Question_Template	*Is {Entity} ready for {Activity}?*
Arguments	Entity: Entity Activity: Activity
Mapping	Entity: `ifcowl:IfcProduct` → `dice:Object` Activity: `dicp:Activity`
Flow_Type	Safety / Workspace; Prerequisite Work; Material; Equipment; External Condition
Target_State	Ready; Approved

Decision logic is captured in a configurable *FlowConstraint* profile linked to each template/request. At runtime, an *InformationRequest* instantiates the template by binding argument roles to project instances, e.g., CQ013 binds Entity=FootingF1, Activity=Concreting, Zone=Zone_FootingF1. Argument types are anchored in shared semantic representations through ontological mapping: activities to *DiCon* Processes (dicp:); components to *DiCon* Entities (dice:) or ifcOWL products (ifc:); and zones to BOT spaces (bot:) or *DiCon* locations (dice:). Similarly, documents map to *DiCon* Information (dici:), resources to equipment/material concepts (dice/dicp:), and actors to *DiCon* Agents (dica:). This structure enables binding placeholders to project-specific instances, such as schedule tasks, model identifiers, zone definitions, or supplier records.

In addition to the argument structure, each template is associated with a characterization of the dependency it addresses and the state being checked. The methodology employs a concise vocabulary of flow types and target states, informed by lean construction practice and represented us-

ing *DiCon* constraint and state constructs (dicv:, dice:). In particular, the flow vocabulary is derived from the eight main flows used in Lean Construction constraint analysis: people, information, shared understanding, equipment, materials, prerequisite work, safe workplace, and external conditions (buildingSMART Deutschland, Fachgruppe BIM und Lean Construction, 2025). In the present paper, this flow vocabulary serves as an exemplar structuring device for one important class of operational requests, rather than as a restriction on the generality of the request layer itself. Target states specify whether the required condition is completion, approval, conformity, cleanliness, safety, availability, confirmed booking, or acceptable external conditions. For approval *requests*, the template semantics reflect that approvals are often granted for a subsequent activity, for example, "approved for concreting", rather than representing a context-free approval status. This improves alignment with work-release decisions, in which the relevant interpretation of an approval is tied to the next planned operation.

These elements are formalized in a request ontology. In this ontology, *RequestTemplates* are represented as instances of a *RequestTemplate* class, with properties that capture argument roles, associated flow types, and target states, as well as links to the corresponding classes in *DiCon*, *ifcOWL*, and *BOT*. The request ontology sits on top of the existing domain ontologies and defines the structure of information requests independently from the particular data instances of a project. It does not introduce new construction entities but explicitly represents the demand side. This allows the framework to refer to question types in the same way that *DiCon* relates to activity or entity types, and it enables reasoning about which data and states must exist to answer a given question type.

A second element of the methodology is to define graph-based query patterns for each template. While the request ontology defines the semantic structure of the *InformationRequest*, the query pattern describes how an instance of that *template type* is intended to be resolved over a set of federated knowledge graphs. For each template, the methodology identifies the relevant paths through the *DiCon* modules and related ontologies that correspond to the flows and states referenced by the template. For a completion *template*, the pattern follows the link from a process to its state and checks whether it is *Completed*. For a tolerance *template*, it follows from a component to a variable representing a measured property. It compares it to a constraint that encodes an allowed range and provides traceability to the originating requirement or specification reference. For booking and availability queries, patterns integrate process time windows with resource allocation or booking states and verify that commitments cover the relevant period.

For readiness *requests*, the pattern becomes more complex because readiness is defined as a conjunction of multiple flow conditions. In the framework, readiness is operationalized as an explicit set of constraints attached to a request, each specifying a flow type, a required target state, and the topic it concerns. The readiness query pattern evaluates these constraints by searching for evidence paths in the federated graph. This supports evidence-based assessment of work release and enables identification of which constraint lacks supporting evidence when a request cannot be satisfied.

These query patterns are defined as reusable graph-traversal templates and, for selected question types, are specified in SPARQL. They specify which node and relationship types must be traversed in the graph, but they do not hard-code project-specific identifiers. Instead, they are designed so that, once a concrete information request binds the template placeholders to specific project entities, the pattern can be instantiated into a graph query specification in SPARQL or another query language.

Scope and limitations. The methodology targets *execution-time* decision support in construction, where stakeholders assess readiness, approvals, constraints, resource commitments, and safety/compliance conditions under short-horizon operational planning. While the request layer is process-agnostic by design, the elicitation and exemplar instantiations in this study focus on foundation construction; extending coverage to additional project phases (e.g., design), trades, or process types primarily requires adding further templates, mappings, and query patterns.

The methodology stops short of implementing the full machinery for automatically translating arbitrary natural-language questions into template instances and for extracting evidence from all operational systems. Instead, it focuses on defining the semantic structure and logical content of templates and patterns, demonstrating how they can be aligned with existing ontologies and support cross-system reasoning about readiness, approvals, constraints, and resource availability.

By proceeding from observed questions in practice through abstraction into templates, formalization in a request ontology, and definition of graph-based query patterns, the methodology provides a coherent framework for representing stakeholder information needs in a manner compatible with ontology-based data integration efforts. It establishes a layer between human questions and machine-interpretable models, serving as a foundation for subsequent work on natural language interfaces, interactive disambiguation of underspecified requests, automated query generation, and integration with digital twin platforms.

Prototype Design and Illustrative Instantiation

This section outlines the current prototype design for operationalizing *InformationRequest*s over federated construction knowledge graphs and illustrates how selected request types are intended to be instantiated. The companion repository (see Data Availability) provides the request ontology, template library, example request instances, and

an evolving query-pattern collection that together specify the proposed execution logic. At the current stage, these artifacts demonstrate how recurring operational questions can be formalized, bound to project-specific entities, and linked to graph-based evidence paths, but they do not yet constitute a fully implemented end-to-end demonstrator. In particular, automated execution across multiple distributed sources, persistence of intermediate check results, and comparative evaluation against manual information retrieval remain future work. Accordingly, the contribution of this section is to make the proposed operationalization concrete and inspectable rather than to report a completed system evaluation. The examples that follow are therefore intended as illustrative instantiations of the request logic, not as an empirical evaluation of end-to-end system performance.

A central design choice is that the framework does not assume universal, one-size-fits-all definitions of readiness, approvals, constraints, and resource availability. Instead, each template is paired with a *constraint profile* derived from decision checkpoints in an exemplar process chain, as described in the Methodology. This keeps request semantics stable across projects while allowing project-specific specialization of what counts as sufficient evidence (e.g., adding/removing constraints or tightening evidence requirements).

Consider the natural language question *''Is zone A ready for reinforcement placement?''*. The framework does not treat *ready* as a literal attribute to retrieve. Instead, the utterance is mapped to template T04, *''Is {Zone} ready for {Activity}?''*. This readiness mechanism is defined for multiple scopes, including zone-level readiness (T04) and entity-level readiness (T34, Figure 2). In the current repository state, zone-level readiness (T04) is specified more concretely than the entity-level readiness pattern (T34), whereas the latter currently remains illustrative. The request ontology constrains this mapping by typing the placeholders: *{Zone}* must resolve to a spatial entity, for example a *BOT* space or an aligned spatial entity concept, and *{Activity}* must resolve to a process concept in the workflow model, for example a *DiCon* process instance or type. These types act as semantic guardrails. They ensure that query construction relies on relations and states that exist in the aligned graph model, rather than producing queries that are formally valid but semantically ungrounded.

The template becomes operationalizable through bindings. For the corresponding request instance, *Zone_Instance = Zone A* and *Activity_Instance = Reinforcement Placement* are bound to concrete graph nodes such as *ex:ZoneA* and *ex:RebarPlacement_ZoneA*. The binding step links the domain language to project identifiers, including schedule IDs, identifiers from digital building models (for example, IFC GUIDs or curated element sets), and document references. In practice, this does not require authoring each request by hand: zones can be resolved from a location breakdown structure or spatial definitions, while activities can be resolved from schedule activities, work packages, or Last Planner tasks. What matters is that bindings persist as explicit links, preserving traceability from the question text to the graph entities used in execution. Operationally, readiness is evaluated as a *set of prerequisites* rather than a single status value: an *InformationRequest* is satisfied only if each required prerequisite can be justified by an evidence path in the federated graph. Accordingly, each prerequisite presupposes that corresponding evidence objects (e.g., inspection records, delivery confirmations, reservations, or safety documents) are maintained in the relevant source systems with sufficient identifier quality and timeliness to support decision-making. In Lean Construction constraint analysis, such prerequisites are commonly organized into recurring flow categories (often referred to as the ''eight flows''): *People* (qualified workforce), *Information* (execution documents), *Shared Understanding* (aligned interpretation and commitments), *Equipment*, *Materials*, *Prerequisite Work* (prerequisite work and trade interfaces), *Safe Workplace* (space availability and safety measures), and *External Conditions*. Table 2 summarizes the mapping from these flow categories to *FlowConstraint* types and typical evidence objects in a federated graph. Within

Table 2: Mapping of Lean ''8 main flows'' to controlled FlowConstraint types and typical evidence objects.

8 main flows	**Flow Constraint Type**	**Evidence Object Types (examples)**
People	Staffing	Crew allocation; Shift record
Information	Information	Latest drawings; Permits
Shared Understanding	Commitment	Weekly plan promise; Handover criteria
Equipment	Equipment	Reservation; Availability confirmation
Materials	Material	Delivery confirmation; On-site stock record
Prerequisite Work	Prerequisite Work	Completion state; Predecessor handover
Safe Workplace	Safety / Workspace	Risk assessment; Traffic setup
External Conditions	External Condition	Weather check; Site restrictions

Note: Flow types and target states are controlled vocabulary terms; labels shown here are human-readable.

the request ontology, these flow categories are encoded via *FlowConstraint* objects attached to an *InformationRequest*. Each *FlowConstraint* specifies a flow type, a target state, and the topic it concerns, with approvals modeled

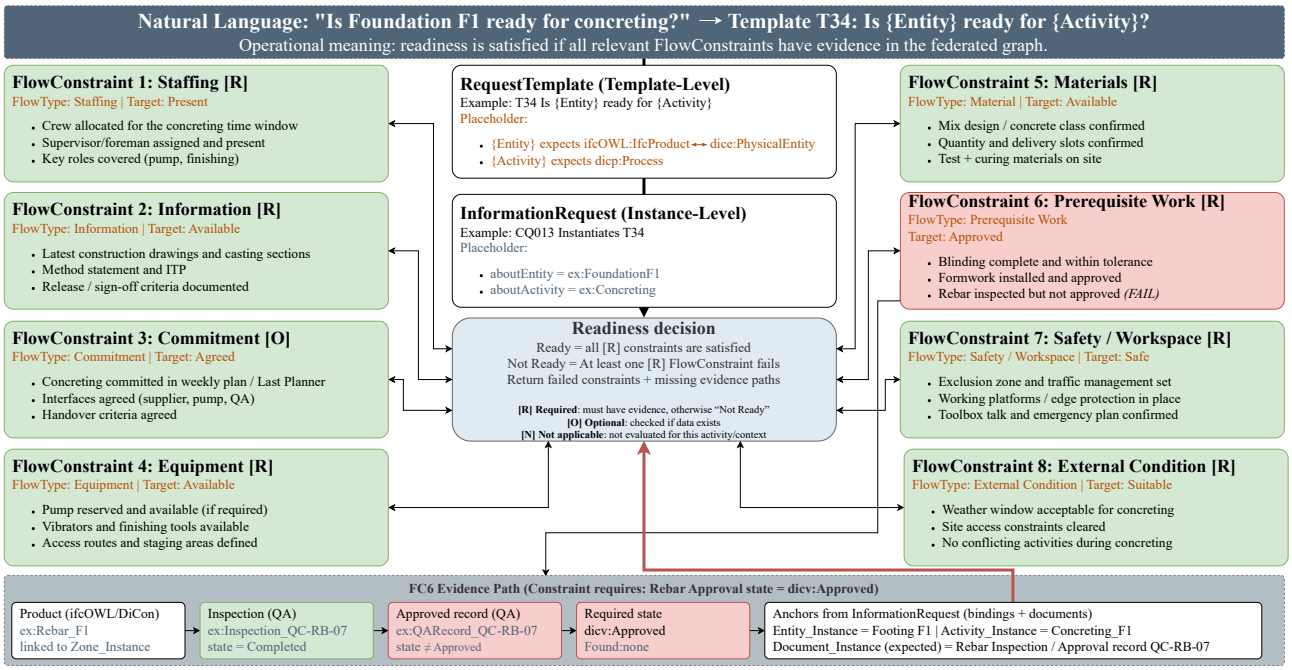


Figure 2: Illustrative template-driven readiness specification with evidence-path diagnostics at object level.

as required target states of the relevant evidence objects (e.g., inspection records) rather than as a separate flow type. For reinforcement placement in a foundation zone, this can include (i) the completion state of prerequisite work (e.g., blinding), (ii) an approval state for the relevant inspection record, and (iii) an implemented state for required safety measures. In this design, a *constraint profile* serves as a reusable default for a given template, specifying which flow constraints are required at a particular checkpoint. An *InformationRequest* then references the applicable profile and materializes the corresponding *FlowConstraint* instances for request-specific evaluation logic; project-specific specialization is expressed by adding, removing, or tightening these constraints at the request level (while keeping the template semantics stable). The key point is that the constraint profile is not constructed at query time; it is derived from the exemplar process checkpoints and refined based on partner feedback, so that the template encodes reusable readiness logic while remaining adaptable to project-specific requirements.

To make evaluation and diagnostics explicit, Figure 2 illustrates the same mechanism as shown in Figure 1, but on an object-level readiness request for concreting (T34) instead of zone-level readiness (T04), in which a single failing constraint can be traced to missing evidence that should justify work release. This highlights a practical requirement of operational question answering: a negative outcome should be explainable in terms of identifiable blockers rather than an undifferentiated ‘‘not ready’’ label.

In the intended execution logic, the bound request is evaluated against the attached constraint set. A readiness request is satisfied only if each required constraint can be justified by a matching evidence path in the federated graph; otherwise, the corresponding unsatisfied constraint and expected evidence type are identified. In this way, negative outcomes remain actionable by exposing specific blockers rather than an undifferentiated ‘‘not ready’’ label.

In the pattern library, analogous blocks are specified to resolve state checks (e.g., approved inspections), safe workplace, information, and logistics constraints through the corresponding document, resource, and process links in the aligned ontologies. Because constraints are explicit objects, the same request can, in principle, return diagnostics by selecting the constraints that fail and exposing missing evidence paths. Conceptually, a response can therefore be treated as a structured answer object that includes (i) the boolean result (ready / not ready), (ii) the constraint set evaluated, (iii) minimal evidence paths for satisfied constraints, and (iv) for unsatisfied constraints an expected path signature indicating what evidence type is missing (e.g., an inspection record in state *Approved* for a specific object and checkpoint). This turns negative outcomes into actionable blockers and preserves traceability back to the originating question, bindings, and sources.

Bindings determine which concrete project instances a request refers to (e.g., which zone, activity, object, or document). In practice, bindings can be obtained through three complementary mechanisms: (i) user-assigned bindings, for example by selecting a zone, activity, and checkpoint in an interface; (ii) context- or rule-derived bindings from existing project structures, such as linking scheduled activities to locations via a location breakdown structure, mapping work packages to curated element sets, or resolving document instances from metadata (e.g., inspection type and revision); and (iii) language-assisted proposals that extract candidate entities from the utterance and surrounding

context. In ambiguous cases, proposed bindings are expected to be confirmed or clarified by the user.

To preserve traceability and support evidence-based diagnostics, each binding is explicitly stored in the *InformationRequest* and annotated with provenance metadata that capture its origin and creation context (source system/-dataset, timestamp, and confirmer). At minimum, this includes the source system or dataset from which the binding was obtained, the identifier/URI used in the graph, a timestamp, and (if applicable) the role or actor who confirmed the binding. Thereby, the provenance metadata supports reproducibility and auditing of decision-relevant answers, and helps distinguish missing evidence in the data from underspecified or incorrectly bound requests. Beyond readiness, the same request semantics covers other execution-time decision questions across the eight flows. The template library includes approval and release checks (linking components or activities to inspection records and release documents), tolerance and requirement conformance queries (linking measured variables to specification constraints with traceability to the originating document), and logistics and resource commitment requests (linking activities and time windows to bookings for equipment, deliveries, or crews). These requests follow the same operational pattern: placeholders are bound to project instances, and the corresponding query pattern searches for satisfiable evidence paths, potentially across federated sources such as scheduling systems, Quality Assurance (QA) tools, health-and-safety management systems, procurement, or supplier dispatch data. When evidence is missing or inconsistent, the request returns unsatisfied constraints and their missing paths, enabling targeted follow-up actions rather than manual, system-by-system search.

Discussion and Future Work

The proposed framework highlights that many execution-time questions are fundamentally representation problems: decision-level inquiries require combining evidence across domain systems. The framework addresses these information needs through reusable requests: templates capture stable semantics, while bindings link them to project-specific instances, and the intended response structure exposes evidence-grounded blockers.

The readiness example illustrates why this separation is useful. A question such as *"Is zone A ready for reinforcement placement?"* cannot be reliably answered by searching a single system or treating readiness as a single status value. It requires explicit aggregation of prerequisite conditions aligned with lean constraint analysis flows, and it benefits from diagnostics that expose which prerequisites lack supporting evidence. Representing prerequisites as typed constraints and evaluating them via evidence paths enables the system to return actionable blocker traces, rather than undifferentiated negative answers. At the same time, the approach is compatible with federated settings, where evidence remains in its native systems, while requests provide a uniform mechanism to verify whether the required information is present and consistent.

The current scope also implies limitations. Templates and query patterns were derived from an exemplar process chain and refined with partner feedback, but broader generalization requires validation across additional trades, contracting arrangements, and delivery settings. Practical deployment depends not only on semantic formalization but also on data governance: binding resolution requires sufficiently stable identifiers, classifications, and contextual cues across tools, and missing records must not be conflated with negative evidence. In federated settings, evaluation must also account for access constraints, latency, and conflicting versions across sources. We have not yet evaluated the approach through end-to-end execution on noisy real-world project data; current validation is limited to curated example graphs and synthetic instances in the companion repository. Robustness under missing, conflicting, or partially linked evidence across sources remains future work. The intended behavior under such conditions is to return explicit blockers (missing evidence paths) rather than produce ungrounded answers silently.

In this sense, the framework can also be understood as an explicit specification of project information obligations: it makes visible which evidence must be available, from which source, and at what level of reliability to support a given operational decision.

Future work will focus on closing the loop from natural language to operationalizable, evidence-grounded requests. This includes template selection and placeholder binding from user utterances and project context, clarification of underspecified requests, and connectors for retrieving and linking evidence from real project systems. Evaluation across multiple projects will assess the correctness of answers, time-to-answer, and the practical usefulness of diagnostics for resolving blockers.

Conclusion

This paper addressed the persistent difficulty of formalizing execution-time process questions in construction when relevant evidence is distributed across heterogeneous systems. It introduced a semantic request layer that represents recurring operational information needs as reusable request templates, separates stable request semantics from project-specific bindings, and links these requests to graph-based evidence structures across federated sources. Lean work-release and readiness questions served as an exemplar application domain for eliciting and structuring the required dependencies.

The proposed request layer complements supply-side ontologies such as *DiCon*, *ifcOWL*, and *BOT* by explicitly representing the demand side of project information: which question is being asked, which entities and constraints it refers to, and which evidence paths would be required to support an answer. The illustrative examples show how natural-language questions can be grounded in explicit bindings and operationalized as evidence-oriented request specifications, providing a more transparent and

traceable basis for decision support. Automated natural-language interpretation and end-to-end execution over live project systems remain future work.

Data availability statement

The data supporting this study's findings are available at `https://gitlab.lrz.de/tweilbach/semantic_request_framework`,
ensuring that they are Findable, Accessible, Interoperable, and Reusable (FAIR).

Acknowledgments

The authors gratefully acknowledge the support and resources provided by the Bavarian Construction Industry Association (Bayerischer Bauindustrieverband), enabling our research and its practical application.
In addition, we would like to thank Vollack Gruppe GmbH & Co. KG and Riedel Bau AG.

References

Ballard, G. (2020). The last planner system. In Lean construction, pages 45–53. Routledge.

Ballard, G. and Howell, G. (1998). Shielding Production: Essential Step in Production Control. Journal of Construction Engineering and Management, 124(1):11–17.

buildingSMART Deutschland, Fachgruppe BIM und Lean Construction (2025). BIM und Lean Construction. Number Schriftenreihe 1.0. buildingSMART Verlag.

buildingSMART International (2025). ifcOWL. `https://github.com/buildingsmart-community/ifcOWL`. Accessed: 2025-11-27.

Čustović, I., Cao, J., Hall, D. M., and Wamelink, H. (2025). Enhancing Situation Awareness of Construction Managers using Human-centered Digital Twins. Advanced Engineering Informatics, 67:103557.

Daum, S. and Borrmann, A. (2014). Processing of Topological BIM Queries using Boundary Representation Based Methods. Advanced Engineering Informatics, 28(4):272–286.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors, 37(1):32–64.

Farghaly, K., Soman, R., and Whyte, J. (2024). cSite ontology for production control of construction sites. Automation in Construction, 158:105224.

Hellin, S., Nousias, S., and Borrmann, A. (2025). Natural Language Information Retrieval from BIM Models: An LLM-Based Multi-Agent System Approach. In Proceedings of the 2025 European Conference on Computing in Construction, volume 6 of Computing in Construction, Porto, Portugal. European Council on Computing in Construction.

Koskela, L. (2000). An exploration towards a production theory and its application to construction: Dissertation. PhD thesis, Aalto University, Finland.

Mazairac, W. and Beetz, J. (2013). BIMQL – An open query language for building information models. Advanced Engineering Informatics, 27(4):444–456.

Nabavi, A., Ramaji, I., Sadeghi, N., and Anderson, A. (2023). Leveraging Natural Language Processing for Automated Information Inquiry from Building Information Models. Journal of Information Technology in Construction, 28.

Pauwels, P. (2014). Supporting Decision-Making in the Building Life-Cycle Using Linked Building Data. Buildings, 4(3):549–579.

Rasmussen, M. H., Lefrançois, M., Schneider, G. F., and Pauwels, P. (2020). BOT: The building topology ontology of the W3C linked building data group. Semantic Web, 12(1):143–161.

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S., and Girolami, M. (2020). Construction with digital twin information systems. Data-Centric Engineering, 1:e14.

Sacks, R., Eastman, C., Lee, G., and Teicholz, P. (2018). BIM handbook: A guide to building information modeling for owners, designers, engineers, contractors, and facility managers. John Wiley & Sons.

Schlenger, J., Pluta, K., Mathew, A., Yeung, T., Sacks, R., and Borrmann, A. (2025). Reference architecture and ontology framework for digital twin construction. Automation in Construction, 174:106111.

Speiser, K., Seiß, S., Boukamp, F., Melzner, J., and Teizer, J. (2025). From fragmented data to unified construction safety knowledge: A process-based ontology framework for safer work. Automation in Construction, 176:106293.

Teclaw, W., O'Donnel, J., Kukkonen, V., Pauwels, P., Labonnote, N., and Hjelseth, E. (2024). Federating cross-domain BIM-based knowledge graph. Advanced Engineering Informatics, 62:102770.

Yin, M., Tang, L., Webster, C., Xu, S., Li, X., and Ying, H. (2023). An ontology-aided, natural language-based approach for multi-constraint BIM model querying. Journal of Building Engineering, 76:107066.

Zheng, Y., Törmä, S., and Seppänen, O. (2021). A shared ontology suite for digital construction workflow. Automation in Construction, 132:103930.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

EXPLORING THE APPLICABILITY OF COMPUTER VISION MODELS FOR AUTOMATED CODE COMPLIANCE CHECKING

Asaf Talor Shnay[1], Guy Austern[2], Tanya Bloch[1]
[1]Faculty of Civil and Environmental Engineering, Technion-Israel Institute of Technology, Haifa, Israel
[2]Faculty of Architecture and Town Planning, Technion-Israel Institute of Technology, Haifa, Israel

Abstract

Automated Compliance Checking (ACC) traditionally relies on rule-based systems operating on fully parametric Building Information Models (BIM). In practice, 2D drawings remain dominant for communicating building information. Recent advances in Computer Vision (CV) enable reliable detection and segmentation of 2D drawings, but remain largely unexplored for ACC. We propose a CV-based ACC framework for 2D drawings. Through a case study focusing on the instructions for sanitary installations, we describe the full computational pipeline, from drawing preprocessing and detection to compliance checking. Results indicate that CV technology is a viable computational bridge between traditional drawing-based engineering practices and ACC.

Introduction

Design review is a time-consuming task that requires high attention to detail and extensive domain knowledge. Automating design review has been a subject of research for many years. Significant efforts have focused on formalizing regulations into a computer readable format (Preidel and Borrmann, 2016; Wang et al, 2023; Fuchs et al, 2024), as well as on processing BIM files so that they contain the needed information for regulatory checks (Bueno and Bosché, 2024). Despite their potential, existing ACC systems such as Solibri (Solibri, 2024) face several limitations. These tools are BIM based, require adherence to standardized modeling practices, often involve a steep learning curve, and are cumbersome to use. These poses substantial barriers to the adoption of advanced ACC workflows. While overall BIM adoption is increasing, it is unevenly adopted across regions and organizations (Chowdhury et al., 2024). BIM adoption is considerably low in the global south, with the Middle east and Africa typically showing very low scores at the end of the previous decade (Zahedi et al., 2022). Furthermore, BIM use is notably limited within municipalities and local planning authorities (Ullah et al., 2022), where design review still mostly relies on traditional 2D documentation and manual work. Recent advances in computer vision (CV) provide an opportunity to revisit the role of 2D drawings in ACC. Recent image segmentation and recognition methods have reached a level of accuracy that enables the extraction of meaningful geometric information from floor plans. Since 2D floor plans are heavily embedded in the construction culture, leveraging them for an ACC system can offer several advantages. In this paper, we demonstrate how these advances can be used to analyze 2D floorplans with sufficient accuracy to support compliance checking.

We propose a CV approach that treats visual recognition as an enabling step toward regulatory verification, which is performed using traditional, heuristic algorithms. This approach can serve as an intermediate solution for organizations transitioning toward more advanced digital workflows, or a potential alternative to BIM-based ACC systems in contexts where BIM adoption is limited. This will also allow designers to continue working within the familiar environments and workflows they are accustomed to. Alternatively, it may function as a complementary subsystem within predominantly BIM-based environments, supporting compliance checks that do not require fully structured data.

Given the extensive regulatory landscape governing floor plan design and code compliance, we narrow the scope to architectural spatial regulations related to bathroom design. By addressing a well-defined subset of rules that are both spatially complex and critical for compliance checking, we demonstrate the applicability and potential of CV algorithms to support ACC.

Following a review of related work in ACC and floor plan analysis, we describe our computational pipeline, including data processing, object detection, and rule formulation. The proposed approach is then validated through a case study focused on sanitary regulations, analyzing the system's ability to bridge the gap between 2D documentation and automated verification. The study concludes with a discussion of the results and the potential for scaling this technology to broader regulatory contexts.

Background

ACC requires two fundamental components: the design and building regulations. To enable an automated evaluation, both need to be translated into a machine-readable format. A significant body of research has been focused on the interpretation of regulations. Particularly rule formalization, Natural Language Processing (NLP) and Machine Learning (ML) techniques for extracting computable rules from regulatory documents (Ullah et al. 2022; Alnuzha and Bloch, 2025; Zheng et al. 2024). As for the design, one of the most significant enablers of

ACC tools is BIM, which provides a computer readable representation of the design information through explicit geometric and semantic descriptions of building elements. However, BIM has not yet been fully adopted by the AEC industry, especially in the public sector (Yusuf et al., 2024). As a result, the design review process continues to rely on two-dimensional drawings as the primary source of information. Despite the volume of research in the area of ACC, using rule based approaches (Aydin, 2022), graph based approaches (Bloch et al, 2023), ML based approaches (Alnuzha and Bloch, 2025), and more recently, LLM based workflows (Ying and Sacks, 2024), to the best of the author's knowledge, implementation of CV has not been explored for supporting the ACC workflow.

The use of computer vision to process 2D floorplans

Machine Learning is a subset of Artificial Intelligence (AI) that learns to make decisions from a given detailed data set without being explicitly programmed for every decision (Russell and Norvig, 2020; Datta et al, 2024). ML is extensively researched, and applied for real world applications, including in the AEC sector. CV is a branch of ML that involves training algorithms to interpret and classify visual data, such as images and videos. In recent years, Convolutional Neural Networks (CNNs) became a pivotal CV technique, used for image classification, object detection, and segmentation (Chen et al., 2021). CNNs are a deep learning architecture designed to process images by utilizing mathematical convolution operators on their grid data structure. They are designed with specific layers that aid in shape and texture detection and edges. YOLO (Jocher et al., 2023) is a good example for this type of model and is utilized here as an object detection model. Previous research has demonstrated the potential of CV and image processing techniques for extracting meaningful architectural information from 2D floor plan images.

Early work by Ahmed et al. (Ahmed et al., 2012) applied traditional image processing to rasterized floor plans, achieving 82% accuracy in generating structured, computer-readable data. Other studies focused on drawings derived from CAD files (Huang et al., 2008) bridging the gap between architectural drawings and automated analysis. Lv et al. (2021) used DeepLabv3+ combined with specialized loss functions and OCR-based text detection to generate 3D reconstructions from 2D floor plans with over 90% accuracy. Zeng et al. (2019) introduced a dual-decoder network to simultaneously predict room boundaries and types, effectively handling complex wall topologies with an 89% accuracy rate. More recent advancements leverage multi-task learning and deep structural understanding. MuraNet (Huang et al, 2023), for instance, uses relation attention mechanisms to enhance both object detection and segmentation performance, treating rooms and fixtures as interrelated components.

Despite these advancements, the majority of efforts have focused on using the extracted information to generate 3D models or BIM-style reconstructions (Lv et al, 2021; Zeng et al, 2019). However, the data extracted from 2D floorplans has not been leveraged for regulatory code compliance. While floor plan parsing has matured, the downstream use of this information, particularly in evaluating spatial relationships or conformance to design rules, remains underexplored.

Aims and methods

The primary aim of this work is to assess whether information extracted from 2D floor plans using CV techniques can support reliable regulatory compliance checking. To achieve this, we follow a structured methodological pipeline, as illustrated in Figure 1. First, we collect a scale-consistent dataset of relevant 2D floor plans, focusing on the design of bathrooms in both residential and public buildings. Next, we develop a scalable and standardized data tagging and processing pipeline to prepare the data for downstream analysis. Object detection techniques are then applied to identify building elements relevant to compliance checking such as walls, toilets, and sinks. Finally, to demonstrate the viability of integrating CV-based methods into ACC workflow, selected local regulatory rules are translated into machine-readable logic and applied to the information extracted from the floor plans.

Data collection

As this study is focused on regulatory architectural spatial constraints, the floor plans must be at scale to enable the extraction of true geometric dimensions. In addition,

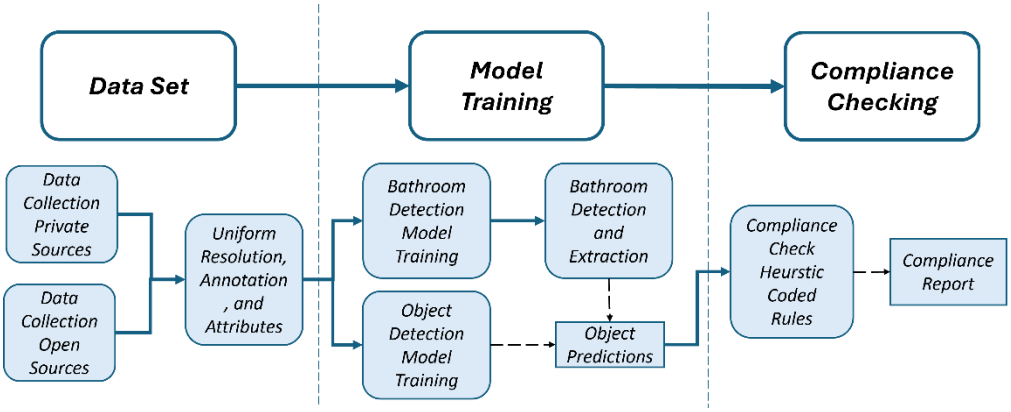


Figure 1: Proposed workflow

having focused specifically on bathrooms, any plans that did not have sanitary fixtures were deemed unusable. Several publicly available data sets were considered. The RPlan dataset (Wu et al., 2019) focuses on room division and doesn't contain sanitary fixtures, hence it was excluded. The Swiss Dwellings data set (Standfest et al., 2022) is too abstract as to train a vision-based model without significant adaptation. Lastly, the CubiCasa5k (Kalervo et al., 2019) data set was also dismissed because the rasterization processes removed any usable scale information. Once the open datasets were excluded, we also explored data sources accessible through the annotation platform used for this research, Roboflow (Alexandrova et al., 2015). Unfortunately, similar limitations persisted, including lack of consistent scale information, and absence of sanitary fixtures. Currently, our work is based on 182 distinct floorplan images collected from our partnership with private companies and public offices. Although the data set used in this study is relatively small, it is sufficient to demonstrate the feasibility of the proposed workflow. Expanding the data set is expected to further improve the results.

Data processing

Given the central role of data representation, we have developed a systematic data processing workflow which was documented in a detailed annotation guide. This guide defines the regulatory constraints addressed in the study, the data processing steps, and the annotation conventions, ensuring tagging consistency. First, each floor plan was assigned with a unique identifier and some descriptive metadata, such as bathroom type, multiuser (for public buildings) or single user (for residential buildings), and the number of bathrooms included in that plan. The goal of this step was to support data management and to establish a common language through the annotation process and team. Afterwards, all the collected floorplans were converted to a uniform resolution of one pixel per two centimeters using GIMP (The GIMP Development Team, 2025), a free image editing tool. The selected scale provides a balance between the image size and its detail, avoiding excessively large images that take a long time to process, while maintaining readability. Thus, a large, 3 meters long bathroom, easily fits within a standard, 256 pixels image. While a 10 cm wide wall occupies 5 pixels in the image and its features are clearly identifiable, as can be seen in Figure 2. The data was then uploaded to Roboflow, an online data management and annotations tool, for further processing. Roboflow was chosen for both its collaborative nature and for the added feature of nested annotations called attributes, which enabled us to tag relevant elements if they passed a compliance check. The images were annotated using 27 classes covering rooms, walls, windows, doors, and sanitary fixtures.

In addition to object classes, each of the objects in the bathrooms of the designs was annotated with an additional attribute according to the compliance rules. Each rule which was to be tested was represented using an attribute indicating whether the corresponding requirement was satisfied or violated. This enabled the annotated data to serve as training data for object detection, as well as a ground truth for downstream compliance checking.

The data was then split into three groups Train/ Validate/ Test, at 76%, 12%, 12% respectively, resulting in 140 images for training, 21 for validation and 21 for testing all in all 182 distinct floorplan images. To overcome the limitations of a small training set, the training data was augmented in the following ways: flipped horizontally and vertically, rotated 90 degrees clockwise and counter clockwise, and finally blurred by 2 pixels, resulting in 694 training images. We kept the validation and test sets without augmentation, giving us a total of 736 floorplans.

To improve detection of small objects, we decided to tile the large architectural drawings into smaller tiles of 640 pixels. To prevent missing essential elements because of tile cutoff, images exceeding 640 pixels in either dimension were tiled into 416X416 pixel images with 150-pixel overlap, as illustrated in Figure 3. The tiling process resulted in 2,410 distinct image patches used for training and evaluation.

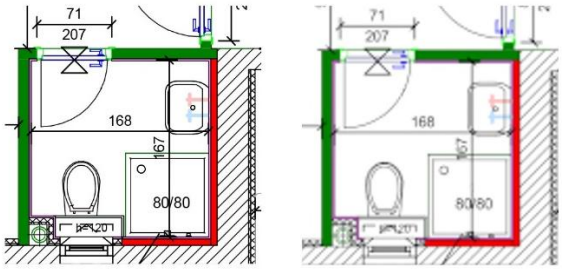


Figure 2: Results of processing with one pixel per two centimeters ratio

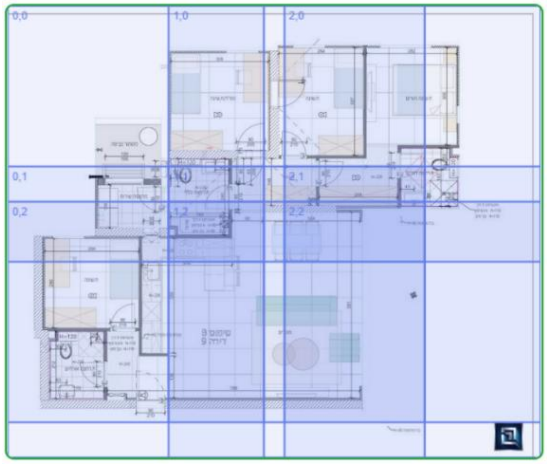


Figure 3: Tiling Example

The size and overlap of the tiling were chosen iteratively, as smaller tile sizes and overlap amounts resulted in degraded detection performance. All the annotations were processed to match the new, tiled images. Annotations corresponding to objects that were partially cut by tile boundaries were discarded if less than 70% of the object area remained visible within the tiled image. For our experiment, a subset of 11 from the 27 original classes was used. The distribution of annotated instances across

classes is provided in Table 1. Table 2 presents a breakdown of the objects involved in the regulatory compliance check within the test set.

Table 1: Description of the final data set

Class	Total	Class	Total
Bathroom	2,391	Opening	8,793
Stall	793	Window	6,803
Wall	44,309	Door Direction	8,976
Toilet	2,873	Sinks Row	204
Sink	3,139	Partition	1,172
Shower	965	Bath	983

This table also presents the proportion of compliant instances per object class, providing insight into the distribution of regulatory outcomes in the data set aside for testing. As non-compliant elements were rare in the original dataset, selected floor plans were modified to introduce additional non-compliant cases, enabling a more balanced and controlled evaluation of the compliance checks.

Table 2: Description of the test set

Class	Accepted	Rejected	Ratio
Opening	43	64	40.2%
Stall	21	20	51.2%
Sink	13	9	59.1%
Door Direction	48	32	60.0%
Toilet	40	26	60.6%
Shower	15	8	65.2%
Bathroom	23	11	67.6%

Object detection

Object detection was performed using YOLOv11 (Jocher and Qiu, 2024). YOLO was selected for both its strong detection performance and its intuitive, easy-to-use nature. YOLOv11 has demonstrated superior accuracy over other versions detecting small and medium-sized symbols (Khanam and Hussein, 2024). The architectural improvements in YOLOv11, particularly its attention module, allow the model to focus on both spatial and channel dimensions making it excellent for the repetitive nature of floorplans. As demonstrated in He et al (2024), YOLOv11 outperforms older YOLO versions and other object detectors like DeepLabv3 +.

To reflect a realistic code checking process, our process was divided between two models. The first model identifies bathroom regions within complete floorplans, enabling the exclusion of images without relevant spaces. After the bathrooms are detected, they are extracted from the full floorplan, then the second model focuses on detecting architectural elements and sanitary fixtures within extracted bathroom regions. This two-stage detection strategy was chosen to improve performance on small objects while reducing unnecessary computation in irrelevant areas.

Model training was initialized using a pre-trained Yolo11m running for 300 epochs with early stopping applied if no improvement was observed in over 30 consecutive epochs. Optimization parameters, including the optimizer and learning rate, were selected automatically by the model. The batch size was set to 16 and the image size chosen was 640x640 to fit our tiling results. This has the added advantage of not compressing the images for training, which would have caused small objects like sinks to become skewed. Validation was performed during training on a withheld validation set. Following the training, the ensembled object detection model was applied to previously unseen 2D floorplans to identify all objects relevant to compliance checking.

Compliance checking

To demonstrate the applicability of CV workflows to support ACC, this study focuses on six regulatory requirements: door opening direction, and minimal sizes of bathroom, shower, opening, sink usage area, and toilet usage area, as defined in the Israeli sanitary installations code (Ministry of the Interior, 2007).

The first requirement we checked was that all bathroom doors must swing inward, opening into the bathroom or stalls. Minimum dimensions for stalls as defined in the regulations are illustrated in Figure 4 (a).

Showers are required to have both a minimum area and minimal dimensions as described in Figure 4 (b). The most complex regulatory checks addressed in this work concern the usage areas of toilets and sinks. As per regulations, a clearance of 40 cm on each side of the fixture centerline and 75 cm in front of the fixture are required, as illustrated in Figure 3(c). And finally, all openings must be at least 80 cm. All regulatory checks were implemented using a rule-based, conditional logic operating on the geometrical information detected and extracted from the floor plans through the CV pipeline. An example of such a computational procedure used to evaluate the shower size is formalized in Algorithm 1.

Algorithm 1: Check Shower Compliance

INPUT: Detections *(D)*, Dimensions *(W_img, H_img)*

OUTPUT: ComplianceResults *(R)*

FUNCTION CheckDimensions *(w, h):*

// Check 1: Rectangular config (min 70x90 or 80x80)

IsRect =

(w + SIZE_TOL >= 70 AND h + SIZE_TOL >= 90) OR
(w + SIZE_TOL >= 90 AND h + SIZE_TOL >= 70)

```
    // Check 2: Square config (min 80x80)
    IsSquare =
(w + SIZE_TOL >= 80 AND h + SIZE_TOL >= 80)
    RETURN IsRect OR IsSquare

MAIN PROCEDURE:
    Showers = GetDetections (D, Class="Shower")
    IF Showers IS EMPTY THEN RETURN R
    FOR EACH s IN Showers:
        // Convert bounding box to calculation units

        w = s.bbox_width  * W_img
        h = s.bbox_height * H_img
        // 1. Validate Area
        CalculatedArea = (w / 100) * (h / 100) * 2
        PassArea =
(CalculatedArea + AREA_TOL >= MIN_AREA)
        // 2. Validate Dimensions (Width/Length)
        PassDims = CheckDimensions(w, h)
        // 3. Determine Result
        Status , Reason
        IF PassArea AND PassDims:
            Status , Reason
        ELSE IF PassArea AND NOT PassDims:
            Status , Reason
        // Store Result
        ADD TO R
    RETURN R
```

Since all the floor plans were scaled at 1 pixel per 2 centimeters, object detection results produced by the YOLO models could be directly converted to real-world dimensions by scaling bounding box coordinates accordingly. This allowed us to translate the object detection box sizes to centimeters, consistent with regulatory requirements. However, since predictions are also not as geometrically accurate as our tagging, a tolerance of 1.5 pixels was used in all algorithms.

Each compliance check was structured to yield a clear pass or fail outcome for every relevant element in the floor plan. For instance, in a multi-user bathroom (public buildings), each toilet's coordinates were assessed individually for the Toilet Usage Area rule, resulting in either a pass or fail for both that object and for the entire room. When an object fails a compliance test, the corresponding reason for non-compliance is reported.

Each image was evaluated against the full set of applicable regulations, and the results were recorded in a structured library indexed by image identifier and regulation type. This library was subsequently compared against the annotated ground truth.

Experiments and results

The object detection model achieved an overall F1 of 0.9059 effectively generalizing the features of the floor plans despite the limited dataset size. As for the second model, performance varied across classes, as detailed in Table 4, with higher precision for recognition of toilets, sinks, and door direction (99.16%, 96.15%, and 95.11% respectively). Lower performance was observed for visually ambiguous classes such as partitions, reflecting the greater variability in their graphical representation within the dataset.

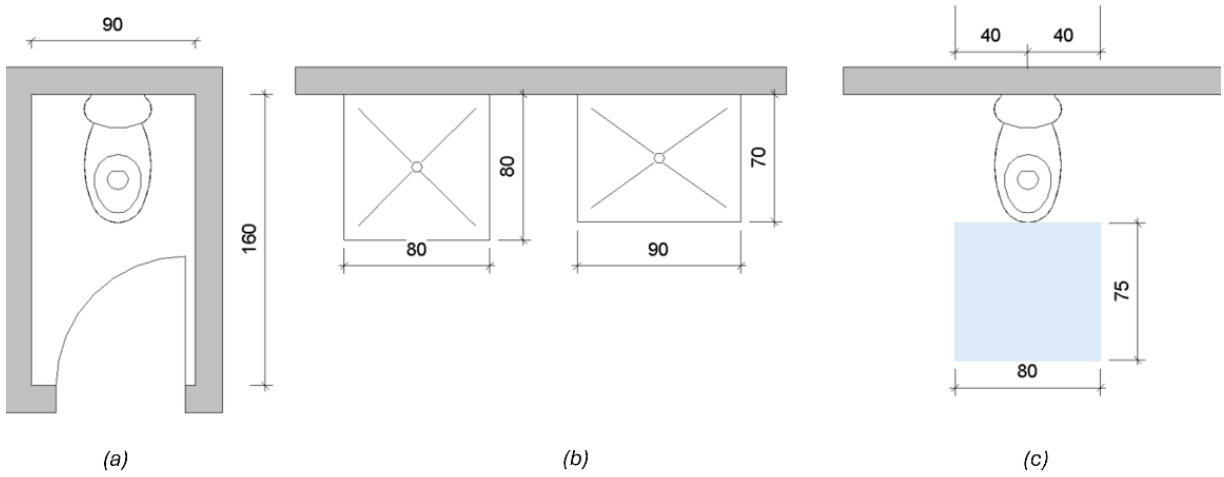


Figure 4: Requirements for (a) stall size (b) shower size (c) toilet fixture usage area

Table 4: YOLO detection Metrics

Class	Precision	Recall	F1-Score
Bathroom	0.8519	08421	0.8470
Door Direction	0.9511	0.9435	0.9473
Opening	0.8995	0.8609	0.8798
Partition	0.8259	0.3250	0.4664
Shower	0.9475	0.8889	0.9173
Sink	0.9615	0.8659	0.9112
Sinks Row	0.8900	1.0000	0.9418
Stall	0.8331	1.0000	0.9090
Toilet	0.9916	0.9041	0.9459
Wall	0.7611	0.8112	0.7854
Window	0.9190	0.9247	0.9218
OVERALL (Mean)	**0.8932**	**0.8700**	**0.9059 (mAP50)**

In order to assess the correctness of the developed regulatory checks independently of the object detection results, they were first evaluated using the ground truth annotations. All heuristic tests yielded 100% precision and recall, except for the toilet usage area which produced one false negative, resulting in 97.4% recall. This error was caused by the algorithm assigning an incorrect toilet orientation affecting the evaluation of the corresponding spatial constraint. Next, the extracted bathroom regions were processed using the trained YOLO model to predict the different elements in our floorplans. Based on these predictions, 296 compliance checks were performed.

Table 5: Compliance tests YOLO prediction results metrics

Category	Precision	Recall	F1
Door Swing	94.7%	100.0%	97.3%
Shower Size	78.3%	85.7%	81.8%
Opening Size	85.0%	100.0%	91.9%
Sink Usage Area	81.8%	75.0%	78.3%
Bathroom Size	70.6%	100.0%	82.8%
Toilet Usage Area	82.1%	74.2%	78.0%
Total	**82.2%**	**87.4%**	**84.7%**

Table 6: YOLO prediction compliance results breakdown

Category	TP	TN	FP	FN
Door Swing	18	14	1	0
Shower Size	18	6	5	3
Opening Size	17	49	3	0
Sink Usage Area	9	10	2	3
Bathroom Size	12	51	5	0
Toilet Usage Area	23	27	5	8
Overall	**97**	**157**	**21**	**14**

Of these, 97 elements were classified as compliant to the code (Approved) and 157 as not compliant (Rejected). The results of the heuristic compliance checks applied to the YOLO-based predictions are presented in Table 5. As shown in Table 6, 21 False Positives (FP) and 14 False Negatives (FN) were observed. As the compliance evaluation relies directly on the detected objects and their bounding box geometries, errors in object detection propagate to the compliance checking stage and affect the resulting classifications.

Conclusions

This study demonstrates the feasibility of using a CV-based workflow for compliance checking, providing a practical and accessible alternative to BIM-based systems. We developed an end-to-end, 2D based workflow that uses conventional architectural drawings. The workflow integrates object detection ML models with algorithmic compliance checking, without requiring structured building information or standardized naming conventions. By focusing on the regulatory requirements for bathroom design, the study provides a demonstration of how CV can support automated compliance checking. Our findings support the potential of this approach in three key areas. First is the robustness of YOLO object detection in floorplans. The YOLOv11 model achieved high precision and recall for key classes, including toilets (99.16% precision), showers (94.75% precision), and door directions (95.11% precision). This indicates that modern vision models can successfully interpret the semantic content of technical 2D drawings without relying on standardized naming conventions, even when trained on small datasets. Second is integration of CV with compliance checks. The heuristic algorithms developed for six specific regulatory clauses (e.g., door swing, fixture usage areas, and room dimensions) proved highly effective. When applied to the model-predicted

data, the system maintained strong performance, identifying compliant and non-compliant elements with an overall precision of 82.2%, and recall of 87.4%. Lastly, we established a scalable data processing pipeline, from scale normalization to attribute-rich annotation, that directly addresses the industry's reliance on 2D documentation. By utilizing the ubiquitous floor plan format, this system lowers the barrier to entry for automation and offers an immediate tool to alleviate the workload of expert reviewers in municipal and private sectors. Compliance checking is often constrained not only by the availability of BIM models, but by the explicit information needed to make those models suitable for automated evaluation. This work suggests a need to reconsider which regulatory checks truly require fully structured BIM data, and which can be reliably performed using geometric information already present in standard design outputs. In this context, CV analysis of 2D floor plans offers an alternative pathway for selected compliance checks that do not depend on rich semantics. As such, we see the proposed approach as a complementary subsystem within ACC environments, enabling partial automation of compliance checking without imposing any additional modeling constraints or information requirements. A systematic analysis is still needed to determine which regulatory checks can be reliably supported by this approach, and to evaluate its potential to reduce the burden placed on designers by additional modeling constraints and information requirements.

Limitations and future work

While the results are promising, certain limitations remain. The current dataset is relatively small (182 plans) and is mostly made up of apartments floorplans. The detection results for specific classes, such as partitions (recall of 32.5%), require improvement. Future research will focus on expanding the dataset to encompass a wider variety of architectural styles and regulatory contexts. This can be achieved by manually scaling and adding annotations to existing datasets or tagging architectural 2D plans which are readily available worldwide. An additional avenue can utilize self-supervised learning methods such as DINO (Caron et al., 2021) to speed up the manual tagging process. Further work is also needed to refine the detection of small or complex objects to include the orientation of sanitary fixtures in our YOLO predictions. As the effectiveness of the algorithmic ACC is tightly coupled with detection accuracy, the effectiveness of the entire process is expected to improve. An alternative direction for dataset expansion is to incorporate more complex regulatory scenarios by generating 2D representations, such as sections or detailed views derived from BIM models. This would enable the evaluation of more spatially complex compliance requirements. Such an approach further supports the concept of a computer vision subsystem operating within a predominantly BIM-based compliance checking workflow, where different representations are leveraged according to the information requirements of each regulatory check.

Additionally, the rule-based logic should be extended to cover a broader range of building codes beyond wet rooms as well as regulations from other countries. To scale this framework beyond manually encoded heuristics, future work will explore the integration of Large Language Models (LLMs) to automatically generate compliance checking algorithms adapted to local building codes. Using techniques such as In Context Learning (ICL) we can base new regulatory requirements on our existing heuristic methods.

Ultimately, this research serves as proof of the concept that CV can effectively bridge the gap between traditional architectural deliverables and modern automation, paving the way for more resilient and inclusive ACC systems. By leveraging LLMs to automatically translate natural language regulations into executable logic, it is possible to eliminate the reliance on human-generated checks and achieve a fully automated, adaptable compliance verification process. Such a process will be especially useful in less developed countries, where BIM adoption is still low. Another potential avenue is in retrofit work, where BIM information of the existing building stock is rarely available.

Acknowledgments

This project was funded by the Israeli Ministry of Defense (IMOD).

References

Ahmed, S., Liwicki, M., Weber, M., Dengel, A., 2012. Automatic room detection and room labeling from architectural floor plans, in: 2012 10th IAPR International Workshop on Document Analysis Systems. IEEE, pp. 339–343.

Alexandrova, S., Tatlock, Z., Cakmak, M., 2015. RoboFlow: A flow-based visual programming language for mobile manipulation tasks, in: 2015 IEEE International Conference on Robotics and Automation (ICRA). IEEE, pp. 5537–5544.

Bueno, M., Bosché, F., 2024. Pre-processing and analysis of building information models for automated geometric quality control. Autom. Constr. 165, 105557.

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A., 2021. Emerging Properties in Self-Supervised Vision Transformers. https://doi.org/10.48550/arXiv.2104.14294

Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., Miao, Y., 2021. Review of image classification algorithms based on convolutional neural networks. Remote Sens. 13, 4712.

Chowdhury, M., Hosseini, M.R., Edwards, D.J., Martek, I., Shuchi, S., 2024. Comprehensive analysis of BIM adoption: From narrow focus to holistic understanding. Autom. Constr. 160, 105301. https://doi.org/10.1016/j.autcon.2024.105301

He, L., Zhou, Y., Liu, L., Ma, J., 2024. Research and Application of YOLOv11-Based Object Segmentation in Intelligent Recognition at Construction Sites. Buildings 14, 3777. https://doi.org/10.3390/buildings14123777

Huang, H.C., Lo, S.M., Zhi, G.S., Yuen, R.K.K., 2008. Graph theory-based approach for automatic recognition of CAD data. Eng. Appl. Artif. Intell. 21, 1073–1079. https://doi.org/10.1016/j.engappai.2007.12.001

Inc, S., n.d. Solibri Office [Computer software].

Jocher, G., Chaurasia, A., Qiu, J., 2023. Ultralytics YOLOv8.

Jocher, G., Qiu, J., 2024. Ultralytics YOLO11.

Khanam, R., Hussain, M., 2024. YOLOv11: An Overview of the Key Architectural Enhancements. https://doi.org/10.48550/ARXIV.2410.17725

Lv, X., Zhao, S., Yu, X., Zhao, B., 2021. Residential floor plan recognition and reconstruction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16717–16726.

Russell, S.J., Norvig, P., 2020. Artificial Intelligence: A Modern Approach, 4th ed. Pearson.

Standfest, M., Franzen, M., Schröder, Y., Medina, L.G., Hernandez, Y.V., Buck, J.H., Tan, Y.-L., Niedzwiecka, M., Colmegna, R., 2022. Swiss Dwellings: A large dataset of apartment models including aggregated geolocation-based simulation results covering viewshed, natural light, traffic noise, centrality and geometric analysis.

The GIMP Development Team, 2025. GNU Image Manipulation Program (GIMP), Version 3.0.4. Community, Free Software (license GPLv3).

Ullah, K., Witt, E., Lill, I., 2022. The BIM-Based Building Permit Process: Factors Affecting Adoption. Buildings 12, 45. https://doi.org/10.3390/buildings12010045

Wu, W., Fu, X.-M., Tang, R., Wang, Y., Qi, Y.-H., Liu, L., 2019. Data-driven interior plan generation for residential buildings. ACM Trans. Graph. TOG 38, 1–12.

Yusuf, A.O., Opawole, A., Musa, N.A., Kadiri, D.S., Ebunoluwa, E.I., 2024. Factors influencing the organisational capabilities of the public sector for implementation of building information modelling in construction projects. Int. J. Build. Pathol. Adapt. 42, 1058–1083. https://doi.org/10.1108/IJBPA-01-2022-0020

Zahedi, F., Majrouhi Sardroud, J., Kazemi, S., 2022. Global BIM Adoption Movements and Challenges: An Extensive Literature Review, in: Proceedings of the Creative Construction E-Conference 2022. Presented at the Creative Construction e-Conference 2022, Budapest University of Technology and Economics, Online, pp. 382–395. https://doi.org/10.3311/CCC2022-048

Zeng, Z., Li, X., Yu, Y.K., Fu, C.-W., 2019. Deep floor plan recognition using a multi-task network with room-boundary-guided attention, in: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9096–9104.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

ALIGNING HETEROGENEOUS VIEWS: STAKEHOLDER TERMINOLOGY AND MULTI-LEVEL MAPPINGS IN THE AEC INDUSTRY

Jakob Martin[1], Cornelius Preidel[1], Simon Vilgertshofer[1], André Borrmann[2]
[1]University of Applied Sciences, Munich, Germany
[2]Technical University of Munich, Munich, Germany

Abstract

Construction stakeholders refer to the same asset using discipline-specific terms and abstraction levels, producing representations with inconsistent naming conventions and modeling granularity. Alignment across views often fails even with standards such as IFC unless semantic relations and transformations are made explicit. We contribute a typology of cross-view mapping requirements as a preceding decision layer for alignment. It determines the mapping type, intended semantic relation, and required transformation between views, guiding candidate generation and validation. Using representative examples from planning, construction, and operations, we show why cross-view alignment requires more than label matching alone.

Introduction

Semantic interoperability and semantic consistency remain challenges in the Architecture, Engineering, Construction, and Operations (AECO) domain, because project information is produced and consumed through many discipline-specific tools, standards, and organization-specific conventions (Ozturk, 2020; Pauwels et al., 2017). Beyond technical friction, poor data quality causes major economic losses, including rework triggered by data-related decision errors (FMI, 2021). While open standards such as Industry Foundation Classes (IFC) are well established in practice and widely adopted to support information exchange, they still fall short of enabling fully consistent and seamless integration. Persistent challenges arise from inconsistent naming conventions, incompatible property structures, and divergent semantics, particularly when project-specific extensions and local naming conventions are applied (Ozturk, 2020; Beck, 2023). As a result, project information ecosystems and exchange workflows frequently fail to provide consistent, linkable, and computationally interpretable data across stakeholder boundaries.

These recurring mismatches are closely related to ontology matching, which provides established primitives and workflows for relating heterogeneous schemas and knowledge representations, including correspondence types, confidence scoring, and hybrid matching and validation methods (Euzenat and Shvaiko, 2013).

A key reason is that stakeholders use different terminology and conceptualize and delineate assets differently, often referring to partially overlapping portions of the physical world. In this paper, *referent* denotes the underlying physical asset, space, or process, while modeled objects may differ slightly in scope across views (e.g., whether a door frame is considered part of the door) (Beck, 2023). Rather than enumerating all roles, we treat their information artifacts as view-based projections of this referent.

Figure 1 illustrates three recurring AECO views: a *design view* emphasizing geometry, types and instances, and design intent; a *checking view* restructuring information into rules, thresholds, and pass or fail outcomes; and an *operations view* focusing on maintainable assets, maintenance plans, and spare parts.

The central challenge is that these views can be internally consistent, yet difficult to link across contexts.

These view-specific representations differ in terminology, modeling intent, and information logic. Depending on the stakeholder task, the same referent may be represented as a geometric element, as the subject of rules and constraints, or as a maintainable asset with operational responsibilities. As a result, views systematically use different naming and classification conventions, delineate and group elements differently (in terms of granularity and decomposition), and encode information using different modeling constructs such as entities, attributes, relationships, or constraints, each optimized for a specific purpose rather than cross-view integration.

Across lifecycle phases, stakeholders need to reference and query the same assets consistently, for example, to ensure reliable handover, provide compliance evidence, and track the impact of design changes. Explicit mappings are therefore required to connect these heterogeneous representations. In practice, alignment is rarely just a matter of renaming labels or matching synonyms. Typical cases include entity-to-entity mappings, such as linking an IfcDoor element in a design model to the corresponding asset entry in a facility management system, such as a computer-aided facility management (CAFM) record. Another common case is an attribute-to-entity transformation, in which a fire rating expressed as a property is reinterpreted as an explicit requirement or compliance constraint. A further recurring case is many-to-one aggregation, in which several modeled doors are grouped into one operational set, such as a fire compartment boundary or a maintenance batch. These mapping types determine which semantic relations and transformations must be applied before data

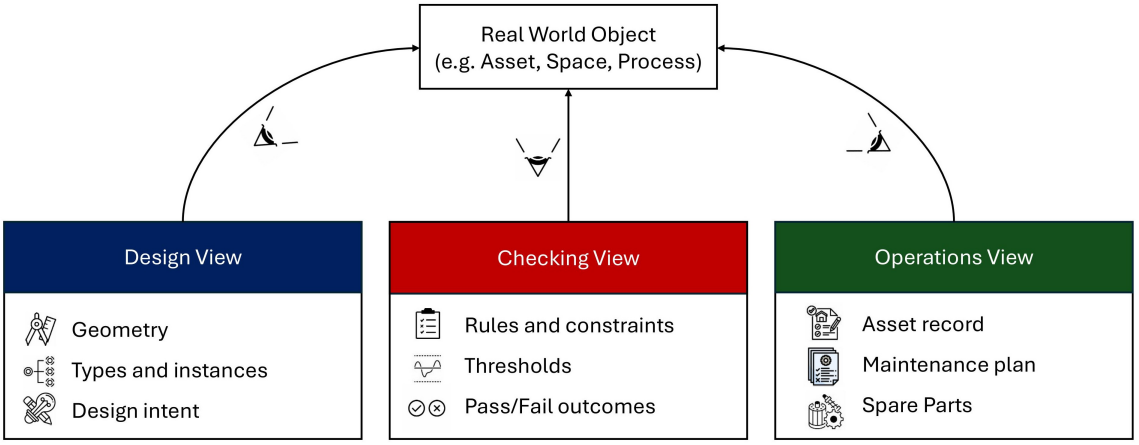


Figure 1: A shared physical referent is represented through different stakeholder views. The design view emphasizes geometry and design intent, the checking view encodes rules, thresholds, and pass/fail outcomes, and the operations view organizes information around maintainable assets. Meaning-preserving integration requires explicit semantic relations and transformations between views.

can be integrated or reasoned over across stakeholder systems.

Ontology matching approaches typically emphasize correspondences between entities or concepts. In AECO integration, link semantics are often context dependent and require interpreting modeling intent and scope differences between views; the core challenge is to determine the intended semantic relation and the required transformation for meaning-preserving integration (Beck, 2023).

This becomes evident when linking design models to downstream uses such as compliance checking and facility management handover. Although much of the required information may already be present, it often has to be restructured or made explicit to become usable in the target view. Even with IFC-based exchange, semantic ambiguity and alternative representation choices can create structural mismatches across federated models and stakeholder systems (Venugopal et al., 2015). More broadly, interoperability research in AECO/FM shows that data loss and communication problems persist across lifecycle exchanges despite the use of established standards (Ozturk, 2020).

This paper contributes a typology of cross-view mapping requirements as a preceding decision and structuring layer for cross-view alignment. It determines the mapping type, the intended semantic relation, and the required transformation between views, guiding candidate generation and validation. The goal is to reduce redundant information creation across disciplines by making explicit where information resides and how it can be referenced and transformed, so that it can be reused and delivered in the form each stakeholder expects. Rather than framing alignment as a uniform similarity-matching task, we derive mapping types that operate at different levels of granularity and semantic interpretation. Using representative examples from planning, construction, and operations, we characterize recurrent mismatch patterns and specify the semantic relations and transformations needed to resolve them, including direct entity correspondence, attribute-to-entity transformation, cluster-based mapping, granularity shift, and value or unit normalization. The contribution is twofold. First, the typology makes explicit which kind of mapping problem is present in a given integration scenario and, by doing so, clarifies the specific semantic relations and transformations that are required for meaning-preserving integration, rather than treating alignment as a uniform matching task. Second, the typology makes these distinctions actionable for AI-supported alignment workflows by specifying what evidence and mapping outputs are needed for candidate generation, validation, and reuse.

State of the Art

Prior work on ontology alignment has established how to define correspondences and transformations between heterogeneous representations (Portisch et al., 2022; Thiéblin et al., 2020). Ontology mapping and alignment research frames interoperability as the establishment of correspondences between entities in different models, enabling information to be interpreted, translated, and queried across sources (Abd Nikooie Pour et al., 2023). Such correspondences are not restricted to equivalence. They may encode subsumption or disjointness and often require transformations, for example, unit conversion, aggregation, or reinterpretation, to preserve semantics (Beck et al., 2021; Beck, 2023). In information exchange across AECO workflows, similar issues arise when stakeholders apply different boundaries and discipline-specific conventions, for example, gross versus net area definitions or properties expressed in different units, which cannot be resolved reliably through label similarity alone.

Matching approaches differ mainly in the information sources they use, for example, labels, structure, data instances, or formal semantics, and in how they combine them (Portisch et al., 2022; Abd Nikooie Pour et al., 2023). Across these approaches, candidate correspondences are typically treated as hypotheses whose usefulness depends on the application context, and many applications require

user scrutiny before results are operationalized (Portisch et al., 2022; Beck et al., 2021).

Prior work on semantic interoperability in AECO indicates that lifecycle integration, meaning reuse, comparison, and translation of project information across design, construction, and operations views often require more than identifying similar terms (Beck et al., 2021). Ontology alignment research distinguishes between discovering candidate links and specifying mappings that preserve meaning under transformation, for example, when aggregation, decomposition, unit conversion, or reinterpretation are required (Thiéblin et al., 2020; Portisch et al., 2022).

In the AECO domain, semantic interoperability has been addressed through standards, ontologies, and linked data approaches, yet recurring gaps remain across lifecycle handovers. IFC provides a shared exchange schema, but IFC-based exchanges still exhibit variability in representation choices and a frequent reliance on property-set extensions, which can introduce ambiguity and structural mismatches across stakeholders and software toolchains (Venugopal et al., 2015). Semantic web research has therefore explored representing BIM data as RDF and OWL to enable explicit linking and cross-dataset querying, including ifcOWL as a representation that allows BIM graphs to be processed with ontology-based methods (Pauwels and Terkaj, 2016; Vilgertshofer et al., 2017). Lightweight ontologies, such as the Building Topology Ontology (BOT), provide shared structural primitives for linking elements, spaces, and zones to complementary sources, such as monitoring and operations data (Rasmussen et al., 2021). Complementary to such lightweight linking ontologies, modular ontology suites such as DiCon cover richer process and workflow semantics for digital construction and integration scenarios (Zheng et al., 2021). Linked data studies further show that interoperability can be achieved through explicit, maintained linking relations between datasets, provided that these relations are made machine-interpretable and that governance mechanisms are in place (Curry et al., 2013; Vilgertshofer et al., 2017).

Recent research operationalizes semantic enrichment and BIM knowledge graphs in concrete workflows, for example, for digital twin data integration and graph-based enrichment of BIM representations (Ramonell et al., 2023; Wang et al., 2024; Bloch et al., 2022). In parallel, machine-readable information requirements and automated checking approaches, such as the Information Delivery Specification (IDS) introduced and maintained by buildingSMART International, support systematic verification of required IFC information in practice (buildingSMART International, 2026; Kremer and Beetz, 2023; Fischer et al., 2024; Akbas et al., 2025). This line of work illustrates how view-specific requirements can be formalized, but it typically assumes that the required information is already structured in the target view and does not address how heterogeneous stakeholder representations should be related when transformations or grouping decisions are needed. However, recent reviews characterize this body of work as still fragmented and largely workflow- and method-driven, with limited consolidation of reusable patterns for relating heterogeneous stakeholder representations (Bloch et al., 2022). This motivates the development of an explicit mapping typology for recurrent cross-view mismatches in AECO.

The proposed typology complements rather than replaces ontology alignment and IDS-based checking. While ontology alignment provides general correspondence types and matching strategies, and IDS specifies target-side information requirements for validation or exchange, our contribution addresses the preceding AECO-specific task of determining what kind of cross-view mapping problem is present and which semantic relation and transformation are required for meaning-preserving integration. The typology therefore serves as a decision and structuring layer before candidate generation, rule execution, or downstream validation takes place. Based on this positioning, the following methodology derives the typology from recurrent cross-view scenarios and abstracts them into a set of mapping types.

Methodology

The State of the Art highlights that alignment problems often differ in the intended semantic relation and in the required transformation, and that these differences are frequently not captured when alignment is treated as a uniform matching task. Building on these observations, we systematize recurrent mismatch patterns into mapping types that make the required relation and transformation explicit.

The typology was derived through comparative analysis of recurrent cross-view integration scenarios involving design, checking, construction, and operations representations. In this paper, three lifecycle examples are discussed in detail, covering design-to-checking, design-to-construction, and design-to-operations transitions. These examples are complemented by additional recurrent cases reflected in the broader type definitions summarized in Table 1. The analyzed artifacts include BIM element representations, checking-oriented rule or constraint representations, construction-oriented grouping views, and operations-oriented asset records. Scenarios were selected deliberately to cover recurring handover and coordination situations, clearly different modeling intents across views, and distinct semantic relations or transformation requirements.

For each case, the source and target view were compared along five dimensions: information units, identifiers, boundaries, grouping, and representation form, meaning whether information is represented as explicit entities or encoded as attributes. A mapping type was assigned when a recurring combination of semantic relation and required transformation could be observed across the analyzed examples. For example, in the design-to-checking case, the door remains the same referent, but the checking view

changes the representation context and reinterprets fire-rating information from an attribute into an explicit constraint object. This combination leads to direct entity correspondence for the door itself and attribute-to-entity transformation for the fire-rating requirement.

Accordingly, the typology is an analytical generalization derived from recurrent patterns in representative cases rather than from large-scale benchmarking, and it does not claim statistical completeness. In this paper, the most representative types are illustrated through concise lifecycle scenarios, while additional types are summarized in Table 1 to indicate the typology's broader applicability.

We refer to the source view as the representation from which information is taken, and to the target view as the representation to which data are mapped to satisfy the stated integration goal. These terms describe mapping direction only and do not assume a fixed lifecycle order.

Mapping Typology for Stakeholder Views

This section introduces a mapping typology that systematizes recurrent cross-view mismatches in AECO by the semantic relation and transformation required for meaning-preserving integration. To support practical use, Table 1 summarizes each mapping type as a recurring integration pattern. For each type, it identifies typical situations in stakeholder views that indicate the need for this mapping, and it specifies what kind of mapping result is required so that meaning is preserved in the target view.

Table 1 shows that resolving cross-view mismatches often requires structural rather than term-level alignment. Depending on the mapping type, integration requires identity linking, constraint lifting, operational grouping, granularity reconciliation, or value and unit normalization. The following scenarios illustrate three representative patterns, while granularity shift and value or unit normalization are summarized in Table 1.

Lifecycle Examples

This section shows how different stakeholder views lead to recurring mismatch patterns. It also shows how these patterns fit the typology categories defined in the Mapping Typology for Stakeholder Views section. Each example describes the source and target view, identifies the mismatch pattern, assigns a mapping type, and explains the semantic relation and transformation needed for actionable mapping across views.

Planning View

The planning view is contrasted with a compliance-based checking view to illustrate how a shared physical referent is conceptualized under different modeling intents. In the design view, the door is represented as a geometric building element, typically as an explicit object with shape, placement, and descriptive properties such as fire rating, acoustic rating, and clear opening width and height. In the compliance view, the focus is not only on the door itself but also on the rule context in which it participates. The door

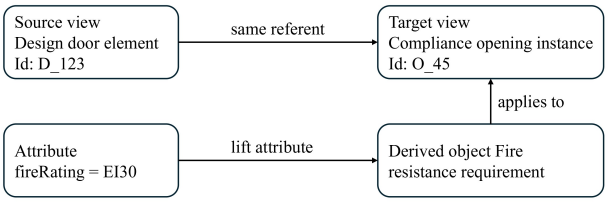


Figure 2: Planning pattern: identity link and attribute lifting.

is considered an opening that must meet egress and fire safety standards. The checking workflow requires explicit constraints that are evaluated against the model. Figure 2 summarizes the mapping output structure required to connect these two representations.

There are two mismatch patterns. First, the door entity in the design view must be linked to the door or opening instance in the checking view. This is a direct entity correspondence, since both target the same referent but use different naming and referencing conventions and delineate the element differently due to modeling conventions. Second, the fire rating stored as a property in the design view must be an explicit constraint in the compliance view. This is an attribute-to-entity transformation. The target view requires a constraint object with explicit semantics, like a required fire resistance class and an explicit evaluation target.

A correct mapping output has two linked statements. The first links the design door entity to the checking door or opening instance using an equivalence relation. The second lifts the fire rating attribute into a compliance constraint object. It then links this constraint to the door instance as its evaluation target. The transformation requires normalizing the fire rating value and linking it with the correct rule context. Label similarity may help rank candidates. However, it is insufficient to infer the intended constraint semantics or the correct target mapping scope (which elements, at what level of granularity) without structural and contextual interpretation.

Construction View

To highlight phase-specific aggregation, the construction case relates a planning-oriented model view to a construction-oriented view centered on work packaging. The construction view organizes information by work packages, assemblies, and temporary states. In the construction view, the main units are often not single model elements. Instead, they are operational units for scheduling, procurement, and site work. Doors may be treated as installation packages, floor-based batches, or trade-specific scopes. These scopes combine multiple components and tasks. Figure 3 summarizes the cluster-based mapping relation needed to link element-level design representations to construction packages.

A common mismatch is a many-to-one operational grouping across views: multiple design elements correspond to a single construction concept, so a grouping transformation

Table 1: Mapping typology for stakeholder views and the semantic relations and transformations required for resolution.

Mapping type	**Definition**	**Typical trigger in stakeholder views**	**Required semantic relation**	**Required transformation**
Direct entity correspondence	A modeled entity in one view refers to the same real-world referent as an entity in another view, but identifiers, boundaries, or context differ. This type applies when the integration goal requires cross-view identity or traceability.	Cross-phase linking where the same object is represented in design and operations, for example, a design element versus an FM asset record for the same referent.	Equivalence, or a same-referent link, between entities.	Identifier linking and normalization, optional projection of relevant attributes, and provenance capture for traceability. Label similarity can support ranking, but it cannot establish identity when identifiers and scopes differ.
Attribute-to-entity transformation	Information encoded as an attribute in the source view must be lifted into an explicit object in the target view because the target view represents it as a requirement or constraint to be evaluated or reasoned over. This type applies when the target contains an explicit constraint object that is not present in the source and must be derived from values and context.	A property is descriptive in one view but normative in another, for example, a performance property in design that must become an explicit constraint in a checking view.	Constraint on, requirement for, or governs relation linking the derived constraint object to the relevant entity.	Attribute lifting into an explicit object, value parsing and normalization, interpretation of units and thresholds, and linkage to the checked entities. Label similarity alone cannot determine the target object type and its semantics.
Cluster-based mapping	A set of elements in one view corresponds to a single concept or unit in another view, where the target concept is defined operationally or functionally. Assign this type when the target concept is defined by membership across multiple source elements and the boundary is not explicitly defined in the source view.	Operational groupings are not explicit in the source view, for example, grouping elements into a maintenance bundle, an operational zone, or a compartment-related set.	Aggregation, part of, or member of a relation linking source elements to the target cluster concept.	Grouping and aggregation based on spatial, functional, or rule-based criteria, with optional creation of a derived target object representing the cluster. Label similarity alone cannot determine membership and boundary conditions.
Granularity-shift mapping	An entity in one view corresponds to multiple entities or a composed structure in another view because stakeholders model at different levels of granularity. Use this type when integration requires decomposition or composition across a part hierarchy, or alignment between type-level and instance-level representations.	Design uses type definitions or assemblies, while construction and operations use instances, components, or maintainable parts, for example, a type definition mapped to multiple installed assets, or an assembly mapped to its components.	Part of, composed of, or an instance of relations connecting coarse and fine-grained representations.	Decomposition or composition, propagation of attributes across levels, consistency checks to preserve totals, constraints, and traceability across levels. Label similarity alone cannot infer the correct hierarchy level.
Value and unit normalization mapping	The same intended property exists in both views, but its value semantics differ due to units, reference conditions, calculation rules, rounding conventions, or code lists. Use this type when the mapping challenge is primarily at the value level rather than the entity level.	Cross-domain exchanges of quantities and performance values, for example, gross versus net area conventions, thermal or acoustic ratings expressed under different standards, or values encoded as codes in one view and numeric values in another.	Same property relation with an explicit transformation function, or equivalence under conversion.	Unit conversion, normalization to reference conditions, code list translation, optional recomputation from available parameters, and validation against allowed ranges or constraints. Label similarity alone cannot guarantee comparable values.

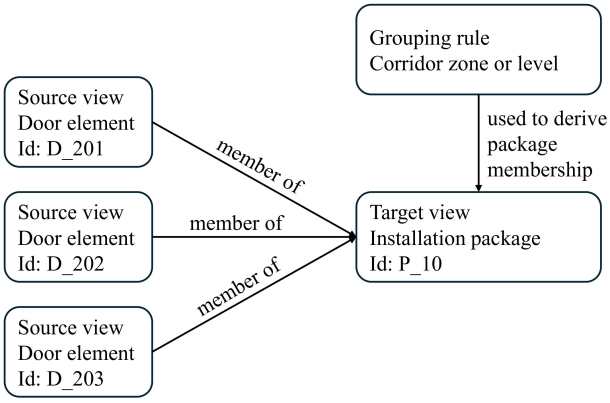


Figure 3: Construction pattern: clustering elements into an installation package.

is required. For instance, a set of doors in a corridor zone may be installed by a single subcontractor during a single scheduled activity. As a result, these doors form a single operational package in the construction view. The design view provides multiple distinct door entities, each with its own properties. In contrast, the construction view groups them as a package entity with references to its members and added attributes, such as planned start date, responsible trade, and installation status.

This mismatch is a cluster-based mapping. Here, the target concept is defined operationally, and its boundary is not clear in the design view. The transformation requires creating a group concept and assigning member relations from each door entity to the package. A correct mapping output includes a construction package object with a list of source doors and a grouping rule. The grouping rule can use spatial zone, level, or schedule logic. The mapping must keep the original door identifiers and introduce a stable identifier for the package. This allows construction reporting and later handover to reference both the group and the elements across phases.

Operations View

The operations case shows how a design representation is reinterpreted in facility management, where a door is managed as a maintainable asset linked to responsibilities and requirements such as maintenance and inspection. In the design view, the door exists as a building element instance with geometric and specification properties. In the operations view, the door becomes an asset record within an equipment register, linked to maintenance plans, spare parts, inspection intervals, warranty information, and potentially sensor or access control system identifiers. Figure 4 summarizes the mapping output structure needed to preserve traceability between the design element, the FM asset record, and operational groupings such as maintenance batches.

Two mismatch patterns often occur. First, the door entity must link to its asset record in the FM system. This is a direct entity correspondence. Both target the same physical door but use different naming and referencing conventions. They may also use different boundaries. For example, an asset record may include both the door leaf and hardware, while the design model separates these components. Second, FM workflows often create groupings that do not exist in the design. Doors may be managed as a maintenance batch, such as all doors of a given type on a floor, or as an operational set, such as a fire compartment boundary. This results in a cluster-based mapping.

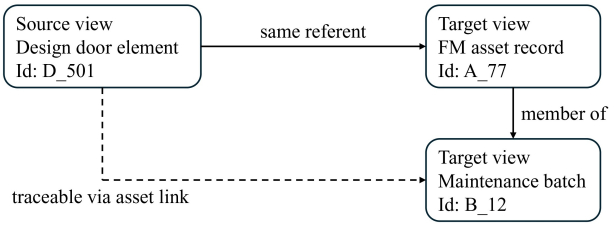


Figure 4: Operations pattern: identity link and operational grouping.

A correct mapping output must keep traceability between views. The output links each design door to its FM asset record and notes which identifiers each side uses. For maintenance batches, the output adds a grouping concept and assigns membership at the asset level. This lets FM tasks be managed at the batch level without losing links to the original elements. The transformation needs identifier reconciliation, optional aggregation rules, and preservation of essential attributes. Key attributes include location, type classification, fire rating, and manufacturer data if available.

Implications for AI-Supported Alignment

Our contribution lies in a mapping typology that provides a preceding decision and structuring layer for AI-assisted alignment workflows connecting different stakeholder views. By determining the mapping type, the intended semantic relation, and the required transformation, it constrains candidate generation and validation.

For direct entity correspondence, the main objective is to resolve identity across differing identifiers, modeling conventions, and element boundaries. For attribute-to-entity transformation, the workflow must detect when a descriptive property needs to be reinterpreted as an explicit requirement or constraint object in the target view. For cluster-based mapping, the workflow must recognize when multiple source elements collectively correspond to an operational concept and must make the grouping boundary and membership relations explicit. Granularity-shift mapping requires reconciling representations across levels of detail, and value and unit normalization mapping requires explicit conversion or normalization logic.

This process, therefore, requires type-specific training and reference data to support both automated candidate generation and subsequent validation. Lexical similarity and terminology resources are useful for candidate retrieval and ranking, but they are rarely sufficient when bound-

ary decisions or semantic transformation is involved. For direct entity correspondence, robust evidence is typically provided by persistent identifiers where available, geometric and spatial consistency, containment relations, and stable type information. For attribute-to-entity transformation, relevant cues include requirement-oriented vocabulary, normative contexts, and value patterns that encode classes, thresholds, or compliance-relevant categories. Cluster-based mapping benefits from shared context and operational metadata, for example, co-location within a zone, shared classification, or explicit grouping rules derived from scheduling logic, inspection regimes, or maintenance strategies.

Human supervision remains essential because mapping correctness is context-dependent and directly affects lifecycle tasks. Validation effort should be prioritized for cases with ambiguous identity evidence, uncertain attribute interpretation, or unclear grouping boundaries. Once validated, these cases can be stored as reference mappings. Over time, curated reference sets enable calibration of type detection, support more reliable candidate ranking, and reduce recurring manual effort in comparable scenarios.

The typology also provides a concrete basis for specifying the minimum mapping outputs that an AI-assisted workflow must produce. Each mapping should record its mapping type, identify the linked source and target information units, and make the intended semantic relation explicit. When transformation is required, the aggregation rule or attribute-lifting logic should be represented in a reproducible form. Provenance metadata and a confidence estimate should be included to support auditability and to enable revision when models, standards, or organizational conventions evolve.

Discussion

Lessons Learned

The lifecycle examples show that semantic heterogeneity across stakeholders is driven not only by terminology, but also by representational differences. Even when referring to the same referent, views vary in what is modeled as an entity versus an attribute, and in how boundaries and groupings are defined. Therefore, interoperability often requires mappings beyond term equivalence, including transformation or aggregation logic.

Direct entity correspondence requires identity resolution across inconsistent identifiers and boundary conventions. Same-referent cases do not necessarily imply strict equivalence, because different views may adopt different boundaries or decomposition logics for the same physical asset. In such cases, the typology helps distinguish direct correspondence from granularity shift, for example when one view represents a door as a single object while another distinguishes frame, leaf, and hardware. To preserve traceability, such mappings should record provenance and, where ambiguity remains, an explicit boundary assumption and review status. Attribute-to-entity transformation requires normative interpretation, turning a descriptive value into an explicit requirement or constraint representation. Cluster-based mapping requires membership and boundary decisions that are often organization-specific and hard to infer reliably.

For practice, the typology suggests classifying the mapping need first, because evidence, validation effort, and maintenance burden differ by type. For standardization, it supports guidance on stable identifiers, operational groupings, and explicit requirement representations.

Limitations

The typology is scoped to recurrent cross-view mismatch patterns illustrated through representative examples from planning, construction, and operations and does not claim exhaustive coverage of AECO interoperability situations. Geometric conversion, coordinate reference transformations, and purely syntactic schema translation are outside scope.

Because mapping needs and boundaries depend on project context, organizational conventions, and metadata quality, some cases remain ambiguous, especially where identifiers are inconsistent or decomposition differs across views. The present paper establishes the typology through analytical comparison of representative cases rather than through a large annotated benchmark. Accordingly, it does not yet report inter-rater agreement, quantitative coverage, or downstream tool-performance evaluation. Its contribution is to provide a structured basis for distinguishing recurrent semantic relations and transformation requirements in AECO view alignment.

An important next validation step is therefore to test whether different analysts consistently assign the same mapping type, semantic relation, and transformation requirement to the same case, particularly in boundary-sensitive situations involving decomposition, grouping, or partial overlap of referents. This is especially relevant because the typology is intended to structure downstream AI-supported candidate generation, validation, and traceability.

Conclusion

AECO stakeholders represent the same real-world assets, spaces, and processes through different terminologies, conventions, and abstraction levels. As a result, structural mismatches persist even with IFC and cannot be resolved reliably through label matching alone.

This paper proposed a mapping typology that characterizes recurrent cross-view mismatch patterns across planning, construction, and operations and makes explicit the semantic relations and transformations required to resolve them. By turning alignment into an explicit decision layer, the typology supports more systematic interoperability practice and provides a foundation for AI-supported candidate generation and validation.

Acknowledgments

This research was supported by the Nemetschek Innovation Foundation and benefited from collaboration with industry partners within the framework of the openDBL project. The openDBL project is funded by the European Union's Horizon Europe research and innovation program under grant agreement No. 101092161, providing valuable test cases during development.

References

Abd Nikooie Pour, M. et al. (2023). Results of the ontology alignment evaluation initiative 2023. In Proceedings of the 18th International Workshop on Ontology Matching (OM 2023), volume 3591. CEUR Workshop Proceedings.

Akbas, B., Idour, A., Isikdag, U., Hammad, A., and Bonneville, C. (2025). A holistic approach to information requirements: Integration of level of information need and information delivery specification. Journal of Information Technology in Construction (ITcon), 30:731–744.

Beck, S. F. (2023). Context-sensitive linking of heterogeneous information models from the building and the urban domain. PhD thesis, Technical University of Munich.

Beck, S. F., Abualdenien, J., and Borrmann, A. (2021). An evaluation of the strict semantics of owl:sameas in the field of bim-gis integration. In Proceedings of the 9th Linked Data in Architecture and Construction Workshop (LDAC2021), volume 3081 of CEUR Workshop Proceedings.

Bloch, T., Sacks, R., Pollmann, C., and Borrmann, A. (2022). Connecting research on semantic enrichment of bim: review of approaches, process, main principles, goals, methods and associated concepts. Journal of Information Technology in Construction (ITcon), 27:416–440.

buildingSMART International (2026). Information delivery specification (ids) project page. Accessed 2026-01-14.

Curry, E., O'Donnell, J., Corry, E., Hasan, S., Keane, M., and O'Riain, S. (2013). Linking building data in the cloud: Integrating cross-domain building data using linked data. Advanced Engineering Informatics, 27(2):206–219.

Euzenat, J. and Shvaiko, P. (2013). Ontology matching: Second edition. Springer, Berlin, Heidelberg, 2 edition.

Fischer, S., Urban, H., Schranz, C., Loibl, P., and van Berlo, L. (2024). Extending information delivery specifications for digital building permit requirements. Developments in the Built Environment, 20:100560.

FMI (2021). Harnessing the data advantage in engineering and construction.

Kremer, N. and Beetz, J. (2023). Extending information delivery specification for linking distributed model checking services. In Proceedings of the 2023 European Conference on Computing in Construction (EC3 2023) and the 40th International CIB W78 Conference, Heraklion, Crete, Greece.

Ozturk, G. B. (2020). Interoperability in building information modeling for aeco/fm industry. Automation in Construction, 113:103122.

Pauwels, P. and Terkaj, W. (2016). Express to owl for construction industry: Towards a recommendable ifcowl ontology. Automation in Construction, 63:100–133.

Pauwels, P., Zhang, S., and Lee, Y.-C. (2017). Semantic web technologies in aec industry: A literature overview. Automation in Construction, 73:145–165.

Portisch, J., Hladik, M., and Paulheim, H. (2022). Background knowledge in ontology matching: A survey. Semantic Web.

Ramonell, C., Chacón, R., and Posada, H. (2023). Knowledge graph-based data integration system for digital twins of built assets. Automation in Construction, 156:105109.

Rasmussen, M. H., Lefrançois, M., Schneider, G. F., and Pauwels, P. (2021). Bot: The building topology ontology of the w3c linked building data group. Semantic Web, 12(1):143–161.

Thiéblin, É., Haemmerlé, O., Hernandez, N., and Trojahn, C. (2020). Survey on complex ontology matching. Semantic Web, 11(4):689–727.

Venugopal, M., Eastman, C., and Teizer, J. (2015). Ontology-based analysis of the industry foundation class schema for building information model exchanges. Advanced Engineering Informatics, 29(4):940–957.

Vilgertshofer, S., Amann, J., Willenborg, B., Borrmann, A., and Kolbe, T. H. (2017). Linking bim and gis models in infrastructure by example of ifc and citygml. In Computing in Civil Engineering. TUM CMS related work, includes linked data based combined querying of IFC and CityGML.

Wang, Z., Sacks, R., Ouyang, B., Ying, H., and Borrmann, A. (2024). A framework for generic semantic enrichment of bim models. Journal of Computing in Civil Engineering, 38(1):04023038.

Zheng, Y., Törmä, S., and Seppänen, O. (2021). A shared ontology suite for digital construction workflow. Automation in Construction, 132:103930.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

Parametric VPINN Framework for Rapid Design Exploration of Thin Slabs

Ahmad Arafat[1,3], Dimosthenis Kifokeris[1,2], and Mohamad Omar Alzokani[1]
[1]Dynzatec AB, Gothenburg, Sweden
[2]Chalmers University of Technology, Gothenburg, Sweden
[3]Etteplan Sweden AB, Gothenburg, Sweden

Abstract

This paper proposes a robust Variational Physics-Informed Neural Network (VPINN) framework specifically tailored for the structural analysis of thin Kirchhoff-Love plates. We utilize Approximate Distance Functions (ADFs) for hard boundary constraints and a hybrid spectral variational formulation to overcome challenges associated with fourth-order partial differential equations, such as vanishing gradients and the need for C^1-continuous meshing. Our VPINN framework can learn a parametric design space and enable real-time, mesh-free structural analysis, eliminating the ''modeling-meshing-solving'' cycle. The results confirm that our VPINN accurately captures critical stiffness trends and global deformation modes, satisfying the precision requirements for conceptual design and topology optimization.

Introduction: Background, related literature, and research aim

Structural engineering plays a pivotal role in defining parameters of safety, material efficiency, and decision-making processes in building projects (Ketabi and Heravi, 2022). Among structural components, slabs often serve as the primary horizontal load-bearing elements. The vast majority of slabs in residential and commercial buildings are geometrically characterized as ''thin plates'' (span-to-depth ratio $L/h > 20$), where the structural response is dominated by bending, and transverse shear deformations are negligible (Timoshenko and Woinowsky-Krieger, 1959). Consequently, the Kirchhoff-Love plate theory, which neglects shear deformation, serves as the most appropriate mechanical description for these elements (Marti et al., 2022). While the Mindlin-Reissner theory is essential for thick plates ($L/h < 10$) to account for shear flexibility, its application to thin slabs introduces unnecessary computational variables and complexity (Teixeira de Freitas et al., 2025). The Kirchhoff-Love theory allows for more computationally efficient models accurately reflecting the stiffer behavior of thin slabs (Marti et al., 2022), potentially leading to more resource-efficient material usage during building production.

The Finite Element Method (FEM) is widely adopted to analyze the structural behavior of slabs. However, despite its high fidelity, FEM faces significant limitations in the context of rapid, iterative design; its reliance on mesh generation becomes a bottleneck during design optimization loops with frequent geometry changes (Park and Kang, 2024). Moreover, most commercial FEM software utilizes Mindlin-Reissner formulations during FEM (using C^0-continuous elements) to avoid the numerical difficulties associated with Kirchhoff-Love theory; implementing Kirchhoff-Love in FEM would require the displacement field to be C^1-continuous (smooth slope everywhere), necessitating complex and computationally expensive elements, such as Hermite or Argyris (Zienkiewicz and Taylor, 2000). However, the generalized utilization of Mindlin-Reissner in FEM unavoidably introduces, in the case of thin slabs, the computational inefficiencies mentioned earlier.

Recently, Physics-Informed Neural Networks (PINNs) have emerged as a promising alternative for solving Partial Differential Equations (PDEs) (Raissi et al., 2019). Unlike FEM, PINNs are mesh-free and differentiable, allowing them to learn the solution operator over a continuous parametric space. This can enable a single trained network to predict the structural response across thousands of design variations instantly, effectively replacing the time-consuming ''modeling-meshing-solving'' cycle with real-time inference (Raissi et al., 2019). However, standard PINNs relying on the strong form of the governing equations struggle with the fourth-order derivatives present in the Kirchhoff-Love bi-harmonic equation ($\nabla^4 w = q/D$); the computation of high-order derivatives via automatic differentiation often leads to numerical instability and vanishing gradients, Wang et al. (2021a).

To address this, two variational approaches that reduce the order of differentiation have gained traction: the Deep Energy Method (DEM), based on the Ritz method (minimizing potential energy) (Samaniego et al., 2020), and the Variational Physics-Informed Neural Network (VPINN), based on the Galerkin method (minimizing the residual of the variational form) (Kharazmi et al., 2021). DEM is intuitive, but recent studies suggest that its unconstrained loss formulation can exhibit instability when learning across complex geometric parameter spaces due to spectral biases in the energy landscape (Wang et al., 2021b). In contrast, the Galerkin-based VPINN framework offers a robust mechanism to project residuals onto test functions, providing improved stability for parametric problems (Kumar and Singh, 2025). In particular, Huang and Peng (2024) proposed an improved plate DEM utilizing finite differ-

ence approximations via convolutional kernels and domain decomposition with coordinate transformations for irregular Kirchhoff plates, achieving high pointwise accuracy on L-shaped geometries. However, their method was demonstrated only for single-instance configurations with structured grids, whereas the parametric setting, where a single network must generalize across a continuous family of geometries, remains largely unexplored.

Therefore, for thin slabs, leveraging VPINN can take advantage of the suitability of the Kirchhoff-Love theory while addressing the combination of computational inefficiencies discussed above; however, such a solution has received limited attention in the literature. As such, this paper proposes a robust framework for VPINN specifically tailored for the analysis of thin slabs. By utilizing the variational form of Kirchhoff-Love theory, we reduce the derivative requirements to the second order, ensuring numerical stability while maintaining mechanical consistency. We demonstrate the framework's capability to learn a parametric design space and discuss its potential to enable real-time, mesh-free structural analysis.

Kirchhoff-Love plate theory

The mechanical behavior of the slab is modeled using the Kirchhoff-Love plate theory, which provides the governing mathematical description for thin structural elements where the span-to-thickness ratio is sufficiently large ($L/h \geq 20$) (Timoshenko and Woinowsky-Krieger, 1959). This theory relies on the kinematic assumption that straight lines remain inextensible and normal to the mid-surface before and after deformation. Consequently, transverse shear strains are neglected, and the deformation is fully characterized by the transverse displacement $w(x, y)$ of the mid-surface $\Omega \subset \mathbb{R}^2$. Under the assumption of linear isotropic elasticity and small deformations, the relationship between the bending moments M and the generalized curvatures $\boldsymbol{\kappa}$ is governed by the constitutive equation $\mathrm{M} = \mathrm{D}\boldsymbol{\kappa}$ (Reddy, 2006).

The equilibrium of the plate under a distributed transverse load q is described by the strong form of the governing partial differential equation. By combining the kinematic relations, constitutive laws, and equilibrium conditions, the problem reduces to the classical bi-harmonic equation, i.e., a fourth-order partial differential equation:

$$D\nabla^4 w = D\left(\frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial y^4}\right) = q \quad \text{in } \Omega, \quad (1)$$

where ∇^4 denotes the bi-harmonic operator and D is the flexural rigidity of the plate, defined as:

$$D = \frac{Eh^3}{12(1-\nu^2)}, \quad (2)$$

dependent on the Young's modulus E, Poisson's ratio ν, and the slab thickness h.

To form a well-posed problem, Eq. (1) must be complemented by boundary conditions. In this study, we focus on clamped edges, which are representative of slabs rigidly connected to stiff vertical supports (e.g., shear walls or monolithic columns). For clamped edges (Γ_C), the boundary conditions are defined as follows:

$$\Gamma_C : \left.\begin{aligned} w &= 0 \\ \frac{\partial w}{\partial n} &= 0 \end{aligned}\right\}, \quad (3)$$

where n denotes the outward unit normal direction.

To avoid the numerical instability of computing fourth-order derivatives in PINNs (Wang et al., 2021a), we adopt the variational formulation. Multiplying Eq. (1) by a test function $v \in \mathcal{V}$ satisfying homogeneous essential boundary conditions and integrating by parts twice yields:

$$\int_\Omega D\left[\nabla^2 w \nabla^2 v - (1-\nu)\mathcal{K}(w, v)\right] d\Omega = \int_\Omega qv \, d\Omega, \quad (4)$$

where $\mathcal{K}(w, v)$ represents the Gaussian curvature interaction term:

$$\mathcal{K}(w, v) = \frac{\partial^2 w}{\partial x^2}\frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 w}{\partial y^2}\frac{\partial^2 v}{\partial x^2} - 2\frac{\partial^2 w}{\partial x \partial y}\frac{\partial^2 v}{\partial x \partial y}. \quad (5)$$

Although $\int \mathcal{K} d\Omega$ theoretically vanishes for clamped slabs, we retain the full formulation for generality. Minimizing this variational residual requires only second-order derivatives, significantly enhancing numerical stability compared to strong-form collocation methods.

Approximate Distance Functions (ADFs)

Standard PINNs enforce boundary conditions via penalty terms, which introduce competing objectives and often lead to inexact satisfaction for fourth-order problems. To overcome this, we employ Approximate Distance Functions (ADFs) (Sukumar and Srivastava, 2021) for ''hard constraint'' enforcement. In this framework, the neural network output $\hat{w}(\mathrm{x})$ is constructed not as a direct approximation of the displacement, but as a composite ansatz:

$$\hat{w}(\mathrm{x}) = \Phi(\mathrm{x}) \cdot \mathcal{N}(\mathrm{x}; \theta), \quad (6)$$

where $\mathcal{N}(\mathrm{x}; \theta)$ is the raw output of the deep neural network parameterized by weights θ, and $\Phi(\mathrm{x})$ is a geometrically constructed distance function that strictly encodes the boundary requirements.

Theory of R-Functions

The function $\Phi(\mathrm{x})$ is constructed using the theory of R-functions (real-valued logic functions), which allows for the composition of complex domain geometries from simple primitives (e.g., lines, circles) using Boolean operations. For a domain Ω formed by the intersection or union of N primitive shapes, the global distance function $\phi(\mathrm{x})$ is derived using the R-equivalence operation:

$$\phi(\mathrm{x}) = \mathcal{R}_m(\phi_1, \phi_2, \ldots, \phi_N), \quad (7)$$

where ϕ_i are the distance functions of the individual primitives. The R-equivalence operation ensures smoothness

(C^m-continuity) and gradient normalization ($|\nabla\phi| = 1$ near the boundary), preventing gradient pathologies during optimization (Biswas et al., 2004).

Application to clamped slabs

For the specific case of clamped Kirchhoff-Love plates, the boundary conditions on the clamped edge Γ_C require both the deflection and the normal derivative to vanish ($w = 0$ and $\partial w/\partial n = 0$). To satisfy both conditions simultaneously using the ADFs, the geometric ansatz in Eq. (6) is modified to use the square of the distance function:

$$\Phi(\mathrm{x}) = [\phi(\mathrm{x})]^2. \tag{8}$$

Since $\phi(\mathrm{x}) = 0$ on Γ_C, it follows that $\Phi(\mathrm{x}) = 0$. Furthermore, the derivative of the ansatz is $\nabla\Phi = 2\phi\nabla\phi$, which also vanishes on the boundary where $\phi = 0$. Consequently, the trial solution $\hat{w} = \phi^2\mathcal{N}$ automatically satisfies:

$$\hat{w}|_{\Gamma_C} = 0 \quad \text{and} \quad \left.\frac{\partial\hat{w}}{\partial n}\right|_{\Gamma_C} = 0, \tag{9}$$

regardless of the values predicted by $\mathcal{N}$. This eliminates the need for boundary loss terms, reducing optimization to a purely physical energy minimization problem.

VPINN framework for parametric slab analysis

In this section, we present the details of the VPINN framework as sketched in Fig. 2.

Geometric parametrization and input space

The structural component under investigation is a non-convex, L-shaped slab, a geometry frequently encountered in architectural layouts around elevator cores or re-entrant corners (Timoshenko and Woinowsky-Krieger, 1959). To establish a comprehensive parametric design space, the geometry is driven by two independent dimensionless parameters, $\mathcal{P} = \{R_1, R_2\}$, which control the aspect ratios of the horizontal and vertical slab segments, respectively.
The physical dimensions of the slab are derived from these parameters to ensure that the domain remains within a normalized bounding box. As depicted in Fig. 1, the total height (b_1) is normalized to unity (1.0) and the top leg width (a_2) is fixed at 0.5. The remaining dimensions are directly controlled by the parameters, such that the total width is $a_1 = b_1/R_1$ and the right leg height is $b_2 = a_2/R_2$.
The parameters R_1 and R_2 vary continuously within the range $[1.0, 1.6]$, creating a family of geometries ranging from thick, compact L-shapes to slender, extended configurations. A critical geometric feature implemented in this framework is the introduction of a circular fillet normalized radius of $r = 0.05$ at the re-entrant corner (the inner corner of the L-shape). This modification removes the geometric singularity associated with sharp re-entrant corners. As noted by Sukumar and Srivastava (2021), smoothing re-entrant corners is essential when using distance-based ansatz functions, as it ensures the gradient of the distance function remains well-defined and continuous ($\nabla\phi \in C^0$) along the entire boundary.

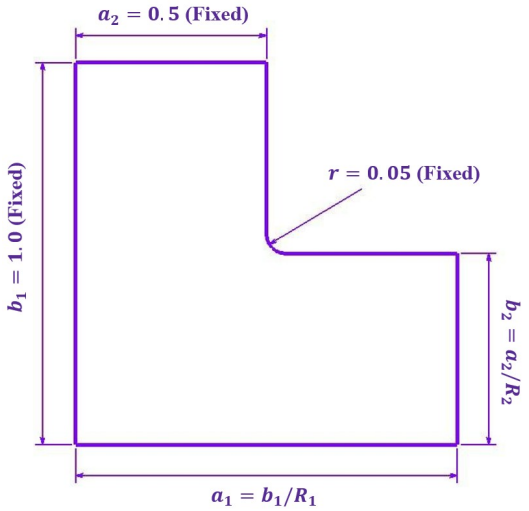


Figure 1: Parametric L-shaped slab geometry

Consequently, the neural network input space is four-dimensional, consisting of the spatial coordinates and the design parameters: $\mathrm{x}_{in} = \{x, y, R_1, R_2\}$.

Neural network architecture in the VPINN

To approximate the solution map $\mathcal{N}(\mathrm{x}_{in}; \theta)$, we employ a fully connected feed-forward neural network, commonly referred to as a Multi-Layer Perceptron (MLP), which serves as the core approximator within the VPINN framework. The architecture is designed to map the four-dimensional input vector $\mathrm{x}_{in} = \{x, y, R_1, R_2\}$ to the scalar latent deflection field w_{model}.
The network depth and width are configured to ensure sufficient expressive capacity for learning the complex solution manifold across the parametric design space. The architecture consists of 6 hidden layers, with each layer containing 512 neurons. This results in a deep over-parameterized model capable of learning the complex parametric solution manifold. To facilitate the training of this deep architecture and prevent the degradation problem (vanishing gradients), residual skip connections are implemented across the hidden layers (He et al., 2016). Mathematically, the output of the l-th layer $\mathrm{z}^{(l)}$ is given by:

$$\mathrm{z}^{(l)} = \sigma(\mathrm{W}^{(l)}\mathrm{z}^{(l-1)} + \mathrm{b}^{(l)}) + \mathrm{z}^{(l-1)}, \tag{10}$$

where W and b are the weights and biases, and σ is the activation function.
A critical feature of this implementation is the use of the Self-Scalable Tanh (Stan) activation function. Unlike standard fixed activation functions (e.g., ReLU or Tanh), the Stan function includes a single global learnable scalar parameter β, shared across all layers of the network, that allows the model to adaptively scale the slope of the activation during training. The function is defined as:

$$\sigma(z) = \tanh(z) \cdot (1 + \beta z). \tag{11}$$

As demonstrated by Gnanasambandam et al. (2023), this specific formulation provides a smooth, non-saturating response. By optimizing a single global scaling factor β, the network can automatically tune the aggregate spectral properties of the approximation without the computational overhead or optimization complexity of layer-wise adaptation. Furthermore, Weight Normalization is applied to all linear layers to decouple the magnitude of the weight vectors from their direction, further stabilizing the optimization process (Salimans and Kingma, 2016).

Ansatz construction for hard boundary constraints

A defining characteristic of the proposed framework is the exact satisfaction of boundary conditions through a geometric ansatz, rather than relying on penalty terms in the loss function. To achieve this, we implement the hard constraint formulation derived in the previous section. The final physical deflection field, $\hat{w}(\mathrm{x})$, is constructed by modulating the raw scalar output of the neural network, $w_{model}(\mathrm{x})$, with the square of the approximate distance function $\phi(\mathrm{x})$:

$$\hat{w}(\mathrm{x}) = w_{model}(\mathrm{x}) \cdot [\phi(\mathrm{x})]^2. \tag{12}$$

Since $\phi = 0$ on Γ_C, both $\hat{w}$ and $\partial\hat{w}/\partial n$ vanish automatically, confining the optimization to kinematically admissible functions and eliminating boundary penalty terms (Berg and Nyström, 2018; Sukumar and Srivastava, 2021).

Construction of test functions

In the VPINN framework, the choice of the test space $\mathcal{V}$ is critical for the accuracy and stability of the numerical integration. While traditional formulations often rely solely on polynomial bases (Kharazmi et al., 2021), we adopt a hybrid spectral approach to capture a broader range of solution frequencies. Specifically, we construct the test space by combining Legendre polynomials, Chebyshev polynomials, and trigonometric functions. This leverages the orthogonality and polynomial approximation power in Legendre and Chebyshev (Canuto et al., 2006) while incorporating the periodic spectral properties of trigonometric series to address the ''spectral bias'' inherent in fully connected neural networks (Rahaman et al., 2019), thereby enhancing the network's ability to resolve complex local features.

The Galerkin formulation requires test functions to satisfy homogeneous essential boundary conditions ($v = 0$, $\nabla v \cdot \mathrm{n} = 0$ on Γ_C) (Hughes, 2012). We enforce this via the same geometric masking used for the trial ansatz (Sukumar and Srivastava, 2021). Let $B_k(\mathrm{x})$ denote the k-th order raw basis function (selected from the combined pool of Legendre, Chebyshev, or trigonometric functions) evaluated at coordinate x. The admissible test function $v_k(\mathrm{x})$ is constructed as:

$$v_k(\mathrm{x}) = B_k(\mathrm{x}) \cdot [\phi(\mathrm{x})]^2, \tag{13}$$

where $\phi(\mathrm{x})$ is the same approximate distance functions (ADFs) used for the trial ansatz. By modulating the hybrid basis with the squared distance field, we guarantee that every test function in the discrete space $\mathcal{V}_h$ vanishes along with its normal derivative at the boundary. This allows for a consistent mesh-free implementation where the variational form residuals are computed solely via domain integration, avoiding the numerical complexity of boundary integrals.

Variational loss formulation

The VPINN minimizes the residual of the integral variational form rather than pointwise strong-form residuals. Since all boundary integrals vanish due to the hard constraints on both $\hat{w}$ and v_k, the variational residual for the k-th test function reduces to:

$$\begin{aligned}\mathcal{R}_k(\theta) = \int_\Omega D\Big[& (\hat{w}_{xx} v_{k,xx} + \hat{w}_{yy} v_{k,yy}) \\ & + \nu(\hat{w}_{xx} v_{k,yy} + \hat{w}_{yy} v_{k,xx}) \\ & + 2(1-\nu)\hat{w}_{xy} v_{k,xy} \Big] d\Omega \\ & - \int_\Omega q v_k d\Omega.\end{aligned} \tag{14}$$

To numerically evaluate these integrals, we employ Monte Carlo integration over the sampled interior points. The total loss function $\mathcal{L}(\theta)$ is constructed by aggregating the squared residuals over the finite set of N_v test functions:

$$\mathcal{L}(\theta) = \sum_{k=1}^{N_v} |\mathcal{R}_k(\theta)|^2 . \tag{15}$$

By minimizing this scalar loss, the optimizer forces our VPINN to satisfy the principle of virtual work over the entire parametric domain. This formulation is purely physics-based and unsupervised, requiring no labeled data or reference solutions (Kharazmi et al., 2021).

Results and discussion

To evaluate the predictive capability of the proposed parametric VPINN framework, we conduct a comparative analysis against high-fidelity FEM reference solutions generated using the legacy FEniCS finite element library with the Ciarlet-Raviart mixed formulation, which decomposes the biharmonic equation into two coupled second-order equations. The FEM discretization employs quadratic Lagrange elements (P_2) on unstructured triangular meshes with approximately 236,440 vertices per configuration, with mesh convergence verified to within 0.5% variation in maximum deflection. For the training phase, the spatial and parametric domains were discretized using 80,000 collocation points generated via uniform quasirandom sampling, ensuring efficient coverage of the irregular L-shaped geometry. The network parameters were optimized using the L-BFGS algorithm in a single optimization step containing approximately 300,000 internal line-search iterations, following the PhysicsNeMo single-step pattern. The optimization process exhibited robust convergence, reducing the variational residual loss from an initial magnitude

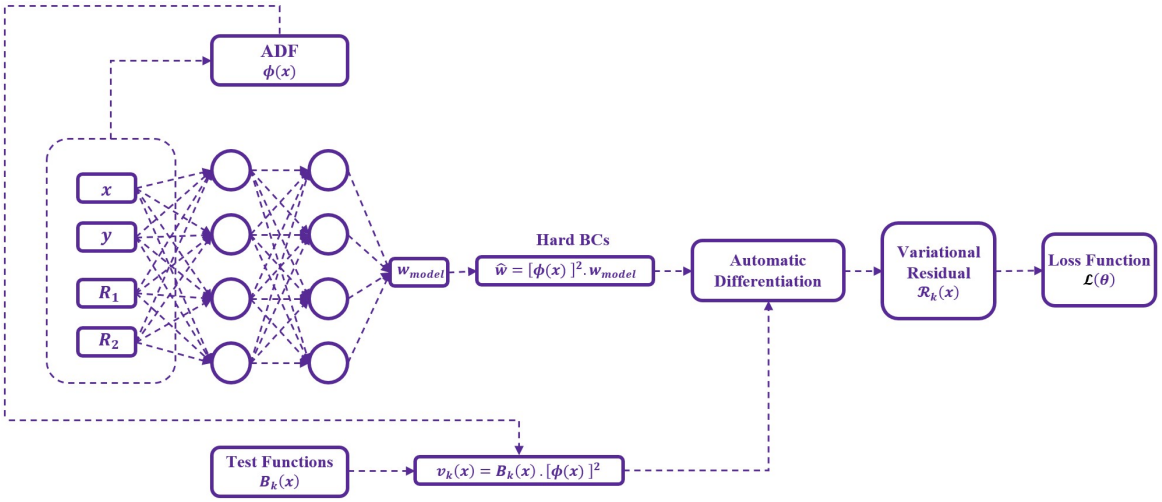


Figure 2: Schematic diagram of the proposed VPINN

of $O(10^1)$ to a final stable residual of $O(10^{-6})$, confirming that the network successfully located a distinct energy minimum.

For the numerical experiments, the physical problem is fully non-dimensionalized. The flexural rigidity is normalized to $D = 1$, the uniform transverse load is set to $q = -1$, and the Poisson's ratio is fixed at $\nu = 0.3$. Furthermore, the spatial domain size is non-dimensionalized such that the entire geometry is strictly bounded within the unit square defined by coordinates $(0, 0)$ and $(1, 1)$. The total offline training cost was approximately 4 hours on a single NVIDIA A100 (80GB) GPU, with a peak memory consumption of approximately 79 GB. In contrast, each FEM reference solution required approximately 63 seconds at mesh resolution 400 (≈236,440 vertices). The amortized computational advantage of the trained VPINN becomes favorable beyond approximately $N_{\text{break}} \approx \lceil T_{\text{train}}/T_{\text{FEM}} \rceil \approx 230$ design evaluations, a threshold readily exceeded in optimization workflows requiring hundreds to thousands of iterations. The neural network was trained once over the parametric domain, and subsequently evaluated to predict the deflection fields for varying geometric configurations without any retraining. This inference process demonstrates true real-time capability, requiring approximately 0.015 seconds per design evaluation on an NVIDIA A100 (80GB) GPU.

Generalization across geometric parameters

We present inference results for two representative geometric cases that test the network's ability to generalize across both symmetric and asymmetric configurations:

1. **Case I:** $R_1 = 1.0, R_2 = 1.0$. This case represents the baseline geometry, characterized by a fully symmetric configuration with equal leg dimensions.

2. **Case II:** $R_1 = 1.0, R_2 = 1.6$. This case represents a boundary of the parametric design space, corresponding to a highly asymmetric configuration with a compressed right leg ($b_2 = 0.3125$). It validates that the VPINN has captured the independent sensitivity of the deflection field to each geometric parameter, rather than simply learning symmetric scaling.

Figure 3 illustrates the comparison between the predicted deflection field $\hat{w}_{\text{VPINN}}$ and the reference solution $\hat{w}_{\text{FEM}}$ for these configurations. A comprehensive quantitative evaluation across all 16 grid points in the parametric space (Fig. 4) is presented in the following subsection.

Error analysis and validation

The absolute error distributions (right column of Fig. 3) reveal the exact satisfaction of boundary conditions. The error vanishes completely at the clamped edges (Γ_C), empirically confirming the theoretical validity of the ansatz construction $\hat{w} = w_\theta \cdot \phi^2$. The VPINN correctly identifies the shift in the maximum deflection point in the asymmetric Case II, demonstrating that the network has effectively learned the parametric dependence of the solution map.

To comprehensively evaluate the framework's generalization across the design space, we conduct a systematic grid evaluation over 16 configurations spanning $R_1, R_2 \in \{1.0, 1.2, 1.4, 1.6\}$. Figure 4 presents the global relative (L_2) error and maximum deflection relative ($\Delta w_{\max}$) error as heatmaps over the parametric grid. The model achieves a mean L_2 error of 7.87% (min: 5.75%, max: 11.55%) and a mean $\Delta w_{\max}$ of 5.90% (min: 0.68%, max: 11.49%) across all 16 cases. Error analysis reveals a systematic increase for larger R_1 values, corresponding to narrower bottom-leg geometries where the aspect ratio is most extreme. The outlier at $(R_1 = 1.6, R_2 = 1.2)$ represents the most elongated configuration, where the fixed-order test function basis provides reduced effective resolution. Notably, the maximum deflection error, the quantity most relevant for serviceability assessment, remains below 10% for 14 of 16 cases, confirming the model's suitability for rapid design screening. This error magnitude is consistent with serviceability tolerances in Eurocode 2 (EN 1992-1-1, $w_{\max} \leq L/250$) and the 10% threshold for successful PINN solutions established by the PINNacle bench-

mark (Hao et al., 2024).

Critically, the error distribution analysis challenges common assumptions regarding L-shaped domains. While the re-entrant corner typically dominates numerical error due to reduced regularity, our parametric filleting successfully regularizes this region. Consequently, maximum deviations shift to the outer clamped corners, where the strict enforcement of zero-slope constraints imposes high-frequency gradients.

This behavior highlights the specific interaction between the test space and the boundary constraints:

1. **Hybrid Spectral Resolution:** The use of a mixed test space combining Legendre polynomials and Chebyshev polynomials (Canuto et al., 2006) with trigonometric functions (Rahaman et al., 2019) was crucial for stabilizing the variational loss. While this hybrid basis effectively captures the global oscillatory modes of the plate, the high-frequency gradients required to strictly enforce the zero-slope condition ($\nabla w \cdot \mathrm{n} = 0$) at the sharp outer corners remain computationally demanding to resolve with high precision. The systematic error increase observed for larger R_1 values in Fig. 4 confirms that the fixed-order basis provides non-uniform resolution across geometries with extreme aspect ratios.

2. **Trade-off against Coordinate Mapping:** In classical spectral methods, achieving high-precision convergence on irregular domains with corner singularities typically requires coordinate mapping or domain decomposition to regularize the geometry (Boyd, 2001). However, applying such mappings in a parametric learning framework is computationally prohibitive. A continuous parametric mapping would require the computation of parameter-dependent Jacobians at every collocation point, significantly increasing the training cost and architectural complexity. Consequently, our Cartesian approach represents a deliberate trade-off: accepting a bounded approximation error (mean L_2 of 7.87% across 16 configurations) to enable a unified, mapping-free parametric model that enables real-time inference.

3. **Design-Centric Accuracy:** In the context of inverse design and topology optimization, the primary requirement is the accurate prediction of stiffness trends and relative performance between different geometries, rather than the recovery of the exact solution to floating-point precision. As evidenced by the systematic evaluation across 16 configurations (Fig. 4), our VPINN consistently captures the correct global deformation mode and parametric sensitivity, with a mean maximum deflection error of 5.90%. This confirms that the model has learned the parametric physics required to drive an optimization loop across the entire design space.

Conclusions

This study has developed and validated a robust parametric VPINN framework specifically tailored for the structural analysis of thin Kirchhoff-Love plates. By integrating ADFs for hard boundary constraints with a hybrid spectral variational formulation, the proposed method successfully overcomes the traditional challenges associated with fourth-order partial differential equations, such as vanishing gradients and the need for C^1-continuous meshing.

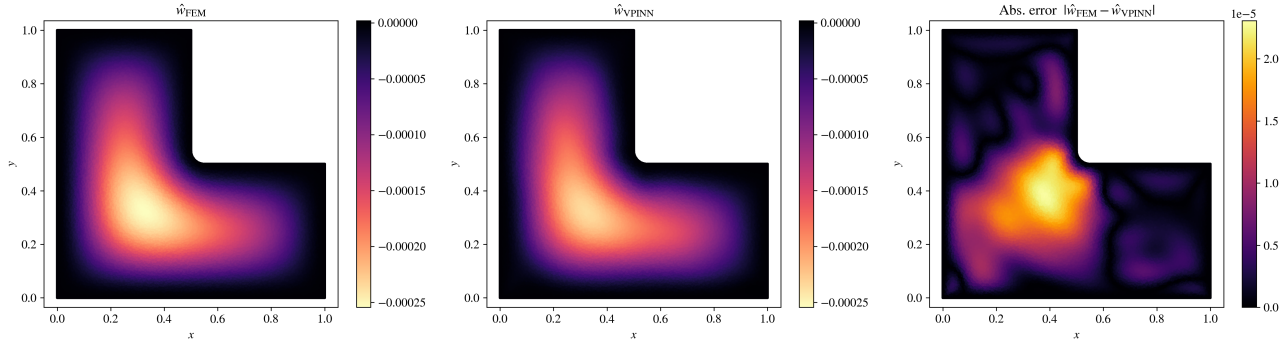
The scientific and practical implications of this work are summarized as follows:

1. **Robustness of the Variational Formulation:** The utilization of a hybrid test space (combining Legendre polynomials, Chebyshev polynomials, and trigonometric functions) proved essential for numerical stability. The framework achieved robust convergence, reducing the variational energy residual from $O(10^1)$ to $O(10^{-6})$. This confirms that limiting the derivative requirements to the second order via the variational form is a superior strategy for bending-dominated problems compared to strong-form collocation.

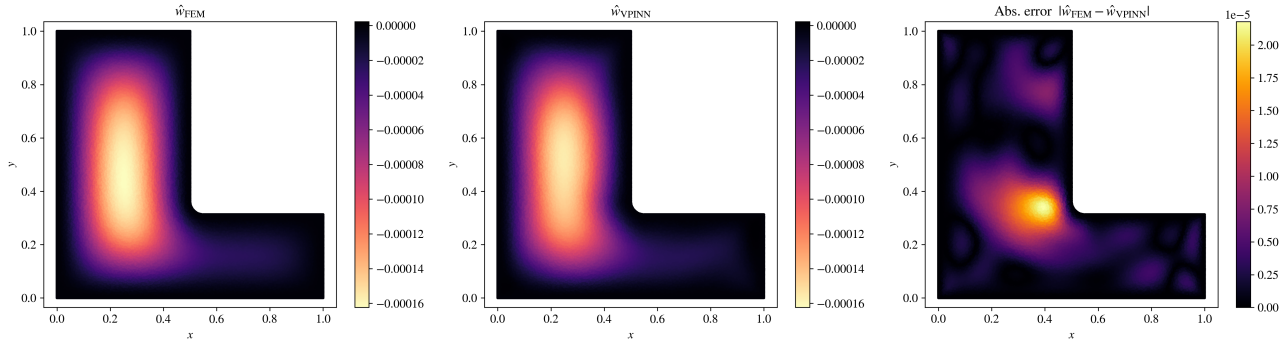
2. **Exact Boundary Adherence:** The implementation of the geometric ansatz $\hat{w} = w_\theta \cdot \phi^2$ guaranteed the exact satisfaction of clamped boundary conditions, as empirically confirmed by vanishing error at all clamped edges across the 16 evaluated configurations. This eliminates the need for hyperparameter tuning of boundary penalty terms, a common bottleneck in standard PINN architectures, and ensures kinematic admissibility regardless of the network's training state.

3. **Parametric Efficiency vs. Precision Trade-off:** Systematic evaluation across 16 configurations spanning the $R_1 \times R_2$ design space demonstrated a mean global L_2 error of 7.87% and a mean maximum deflection error of 5.90%, with 14 of 16 cases remaining below 10%. While higher than refined FEM, this represents a strategic trade-off. By avoiding computationally expensive coordinate mappings, our VPINN framework delivers a unified, mapping-free parametric model that accurately captures critical stiffness trends and global deformation modes, satisfying the precision requirements for conceptual design and topology optimization.

4. **Real-Time Design Capability:** Once trained (approximately 4 hours on a single NVIDIA A100 GPU), the VPINN enables inference of mechanical fields in approximately 0.015 seconds per design, amortizing the training investment after approximately 230 evaluations compared to FEM. This eliminates the classical ''modeling-meshing-solving'' cycle and paves the way for integrating physics-based surrogate models into real-time design workflows, allowing struc-



(a) Case I – $R_1 = 1.0, R_2 = 1.0$



(b) Case II – $R_1 = 1.0, R_2 = 1.6$

Figure 3: Comparative analysis of FEM reference (left), VPINN prediction (center), and absolute error (right)

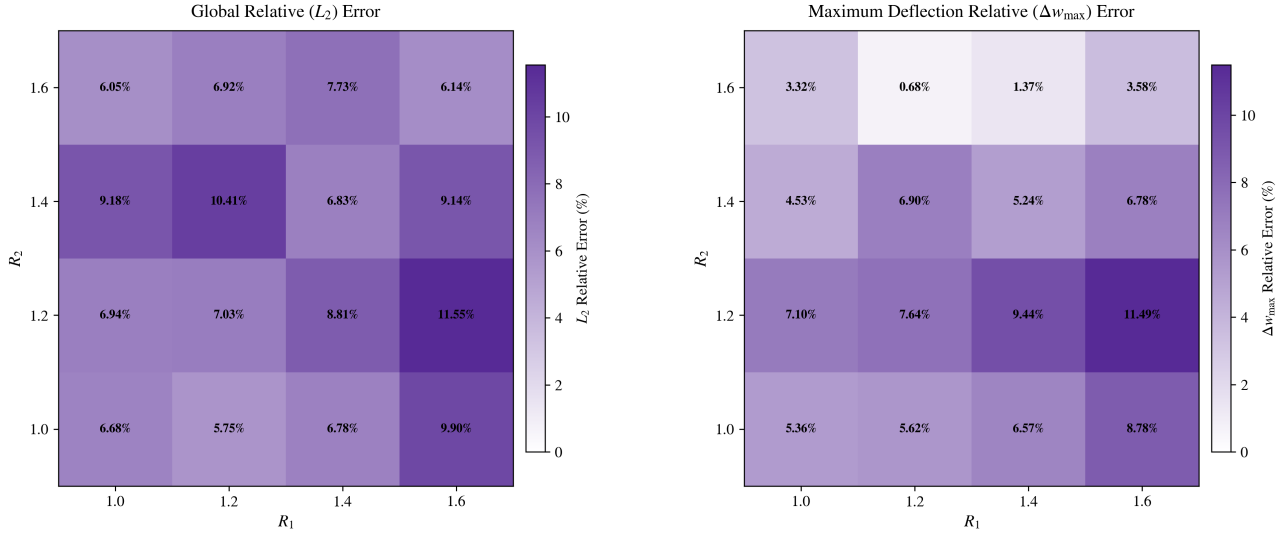


Figure 4: Parametric error distribution across the $R_1 \times R_2$ design space (16 validation cases). L_2 error: mean = 7.87%, min = 5.75%, max = 11.55%. $\Delta w_{\max}$ error: mean = 5.90%, min = 0.68%, max = 11.49%.

tural engineers to explore thousands of design iterations with immediate structural feedback.

Future work will extend this framework along three directions: (i) expanding the parametric space to include slab thickness, load distributions, and mixed boundary conditions; (ii) benchmarking against data-driven surrogate approaches to quantify the advantages of physics-based training; and (iii) incorporating reinforced concrete nonlinearities, including cracking and steel reinforcement, to bridge the gap between conceptual design and detailed verification.

References

Berg, J. and Nyström, K. (2018). A unified deep artificial neural network approach to partial differential equations in complex geometries. Neurocomputing, 317:28–41.

Biswas, A., Shapiro, V., and Tsukanov, I. (2004). Approximate distance functions. International Journal of Computational Geometry & Applications, 14(01n02):95–114.

Boyd, J. P. (2001). Chebyshev and Fourier Spectral Methods. Dover Publications, Mineola, New York, second edition.

Canuto, C., Hussaini, M. Y., Quarteroni, A., and Zang, T. A. (2006). Spectral Methods: Fundamentals in Single Domains. Springer Science & Business Media, Berlin.

Gnanasambandam, R., Shen, B., Chung, J., Yue, X., and Kong, Z. (2023). Self-scalable tanh (stan): Faster convergence and better generalization in physics-informed neural networks. IEEE Transactions on Neural Networks and Learning Systems, 34(9):5687–5701.

Hao, Z., Yao, J., Su, C., Su, H., Wang, Z., Lu, F., Xia, Z., Zhang, Y., Liu, S., Lu, L., and Zhu, J. (2024). PINNacle: A comprehensive benchmark of physics-informed neural networks for solving PDEs. In Advances in Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778.

Huang, Z. and Peng, L. (2024). An improved plate deep energy method for the bending, buckling and free vibration problems of irregular Kirchhoff plates. Engineering Structures, 301:117235.

Hughes, T. J. R. (2012). The Finite Element Method: Linear Static and Dynamic Finite Element Analysis. Dover Publications, Mineola, New York. Section 5.3: Plate Bending.

Ketabi, A. B. and Heravi, G. (2022). Developing a framework for evaluating construction project safety levels and optimal cost allocation to safety influential factors. Engineering, Construction and Architectural Management, 29(9):643–668.

Kharazmi, E., Zhang, Z., and Karniadakis, G. E. (2021). Variational physics-informed neural networks for solving partial differential equations. Computer Methods in Applied Mechanics and Engineering, 374:113535.

Kumar, V. and Singh, G. (2025). A variational physics-informed neural network framework using petrov-galerkin method for solving singularly perturbed boundary value problems. arXiv preprint arXiv:2509.12271.

Marti, P., Kaufmann, W., Seelhofer, H., and Karagiannis, D. (2022). Kirchhoff–love plate deformations reinterpreted. Journal of Engineering Mechanics, 148(6):04022026.

Park, J. and Kang, N. (2024). Bmo-gnn: Bayesian mesh optimization for graph neural networks to enhance engineering performance prediction. Journal of Computational Design and Engineering, 11(1):260–271.

Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., and Courville, A. (2019). On the spectral bias of neural networks. In Proceedings of the 36th International Conference on Machine Learning, pages 5301–5310. PMLR.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378:686–707.

Reddy, J. N. (2006). Theory and Analysis of Elastic Plates and Shells. CRC Press, Boca Raton, second edition.

Salimans, T. and Kingma, D. P. (2016). Weight normalization: A simple reparameterization to accelerate training of deep neural networks. Advances in Neural Information Processing Systems, 29.

Samaniego, E., Anitescu, C., Goswami, S., Nguyen-Thanh, V., Guo, H., Hamdaoui, K., Zhuang, X., and Rabczuk, T. (2020). An energy approach to the solution of partial differential equations in computational mechanics via machine learning: Concepts, implementation and applications. Computer Methods in Applied Mechanics and Engineering, 362:112790.

Sukumar, N. and Srivastava, A. (2021). Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks. Computer Methods in Applied Mechanics and Engineering, 389:114333.

Teixeira de Freitas, J. A., Tiago, C., and Pereira, E. M. B. R. (2025). Modelling of kirchhoff and mindlin–reissner plate bending with hybrid-trefftz stress elements. International Journal for Numerical Methods in Engineering, 126(3):e7668.

Timoshenko, S. P. and Woinowsky-Krieger, S. (1959). Theory of Plates and Shells. McGraw-Hill, New York, second edition.

Wang, S., Teng, Y., and Perdikaris, P. (2021a). Understanding and mitigating gradient pathologies in physics-informed neural networks. SIAM Journal on Scientific Computing, 43(5):A3055–A3081.

Wang, S., Wang, H., and Perdikaris, P. (2021b). Learning the solution operator of parametric partial differential equations with physics-informed deeponets. Science Advances, 7(40):eabi8605.

Zienkiewicz, O. C. and Taylor, R. L. (2000). The Finite Element Method: Solid Mechanics. Butterworth-Heinemann, Oxford.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

POTENTIAL OF LARGE LANGUAGE MODELS FOR CONSTRUCTION QUOTATION WORKFLOWS

Anja Brelih, Žiga Krvina, and Robert Klinc
University of Ljubljana, Faculty of Civil and Geodetic Engineering

Abstract

Large language models (LLMs) can automate text-intensive tasks in construction. However, preparing investigation quotations remains manual and time-consuming, with outputs varying between engineers due to non-deterministic steps like item selection and text drafting. This study proposes an LLM-assisted human-in-the-loop (HITL) workflow and evaluates several models (GPT 4.1, GPT 5, Claude Sonnet 4.5, and a fine-tuned open-source model, QwQ 32B) on 596 inquiry-quotation pairs. GPT 5 achieved the highest item-selection F1 score, while Claude 4.5 delivered comparable accuracy with the fastest processing time. This approach improves quotation efficiency and consistency, demonstrating practical benefits for construction workflows.

Introduction

Artificial intelligence (AI) is increasingly adopted in project management to support decision-making, automate repetitive tasks, and improve overall project performance (Taboada et al., 2023; Brelih et al., 2025). Recent advances in large language models (LLMs) have enabled systems that can understand and generate technical text, creating new opportunities for automating construction workflows.

Construction projects are characterised by unique products, processes, and project-specific contexts, which limit the applicability of rigid, fully predefined automation approaches (Turk, 2006). Quotation preparation is a critical but time-consuming activity for engineering companies. Preparing a quotation typically involves analysing client inquiries, selecting appropriate investigation or service items, estimating quantities, and composing a structured technical and commercial response. While parts of this workflow can be automated using conventional software, key steps remain non-deterministic and rely heavily on expert judgement. These include (1) selecting relevant investigation or service items and (2) formulating a coherent quotation text adapted to the specific inquiry.

This paper examines whether LLMs can effectively support and partially automate quotation preparation. Generic LLMs do not have access to internal company knowledge such as service catalogues, pricing structures, and domain-specific standards. To address this limitation, an open-source LLM was fine-tuned with domain-specific data and evaluated against selected commercial LLM solutions. For the use case, data from a company operating in the field of geotechnical investigations was used.

The contributions of this paper are threefold: (i) the design of an LLM-supported quotation workflow for (geotechnical) engineering practice, (ii) quantitative evaluation of multiple LLMs using accuracy- and time-based performance metrics, and (iii) assessment of the practical impact of LLM integration on quotation preparation efficiency and consistency.

The remainder of the paper is structured as follows. Section 2 presents related work, section 3 describes the quotation workflow and LLM integration concept, section 4 outlines the methodology, section 5 reports the results, section 6 discusses implications, and section 7 provides conclusions.

Related work

This section reviews relevant research on the application of AI in construction workflows, the use of LLMs for technical and domain-specific tasks, and existing AI-based approaches for cost estimation and quotation processes.

AI in construction workflows

Within construction informatics, research focuses on methods for creating, processing, and exchanging construction-specific information in both human- and machine-readable forms. AI has therefore been increasingly explored as a decision-support technology in construction management, particularly for scheduling, resource allocation, risk forecasting, and monitoring (Taboada et al., 2023).

Recent studies indicate that AI-powered systems can assist project teams by synthesising heterogeneous data sources, generating alternative scenarios, and providing early warnings of potential deviations (Mikhridinova et al., 2024). However, many existing approaches rely on predefined workflows and limited interaction capabilities, which restrict their adaptability in highly dynamic project environments.

LLMs for technical and domain-specific tasks

Large language models have demonstrated strong capabilities in understanding natural language and generating coherent technical text (Devlin et al., 2019;

Lewis et al., 2020). Their integration into AI agents enables advanced reasoning, task planning, and interaction with external tools through application programming interfaces (Ruan et al., 2023; Liu et al., 2024).

Recent surveys highlight the growing role of LLM-based agents as intelligent collaborators that can process diverse information, coordinate tasks, and support complex decision-making (Xi et al., 2025). Nevertheless, their effectiveness is constrained by the quality of training data, limited domain adaptation, and the stochastic nature of model outputs (Brelih et al, 2025). Consequently, human-in-the-loop (HITL) designs are widely recommended, particularly in safety-critical domains such as construction.

AI support for cost estimation and quotation processes

AI techniques have been explored for various cost-related tasks, including cost estimation, resource optimisation, and project staffing (Taboada et al., 2023; Mikhridinova et al., 2024). These studies demonstrate the potential of LLM-based systems to assist engineers in analysing project data and generating recommendations.

More recently, Gatto et al. (2025) propose an LLM-based methodology to automate the classification and extraction of cost-related descriptions in building projects, mapping textual specifications to a hierarchical taxonomy and populating a structured database to support downstream estimation tasks. Their work highlights how LLMs can reduce manual effort in interpreting and organising cost-relevant information, while also exposing the need for careful prompt design and validation in real-world workflows.

However, existing research primarily addresses cost prediction or high-level decision support rather than the end-to-end preparation of quotations. Little attention has been given to automating the selection of investigation or service items and the generation of quotation texts tailored to specific client inquiries.

Research gap

While prior research demonstrates the potential of AI and LLM-based agents to support construction workflows, there is a lack of empirical studies. This paper addresses this gap by evaluating multiple LLMs and proposing an LLM-supported quotation workflow tailored to domain-specific requirements for construction practice.

Quotation preparation workflow and LLM integration

In the company analysed, quotation preparation is a multi-step process initiated by a client inquiry received via email, which includes a brief project description, location, and requested services. An engineer reviews the inquiry, identifies the relevant type of geotechnical investigation, and manually selects appropriate service items from the internal service catalogue. Based on the selected items, quantities are estimated and prices are retrieved from internal pricing lists.

The engineer then prepares a quotation document that combines structured information (selected items, quantities, prices) with unstructured text describing the scope of work, assumptions, and conditions. The final quotation is reviewed and sent to the client. This workflow is largely manual and depends on individual experience, which can result in variability in both content and structure.

Although parts of the workflow follow predefined rules, several steps require expert judgement. Selecting of investigation or service items based on the inquiry content, estimating of quantities for individual items, and formulating of a coherent quotation text tailored to the specific project are non-deterministic steps and similar inquiries may result in different valid solutions depending on interpretation and experience.

LLM-supported workflow

To address these limitations, an LLM-supported workflow with human-in-the-loop (HITL) was designed. Upon receipt of a client inquiry, the inquiry text is provided to the LLM along with contextual information about available services, pricing structures, and domain-specific guidelines.

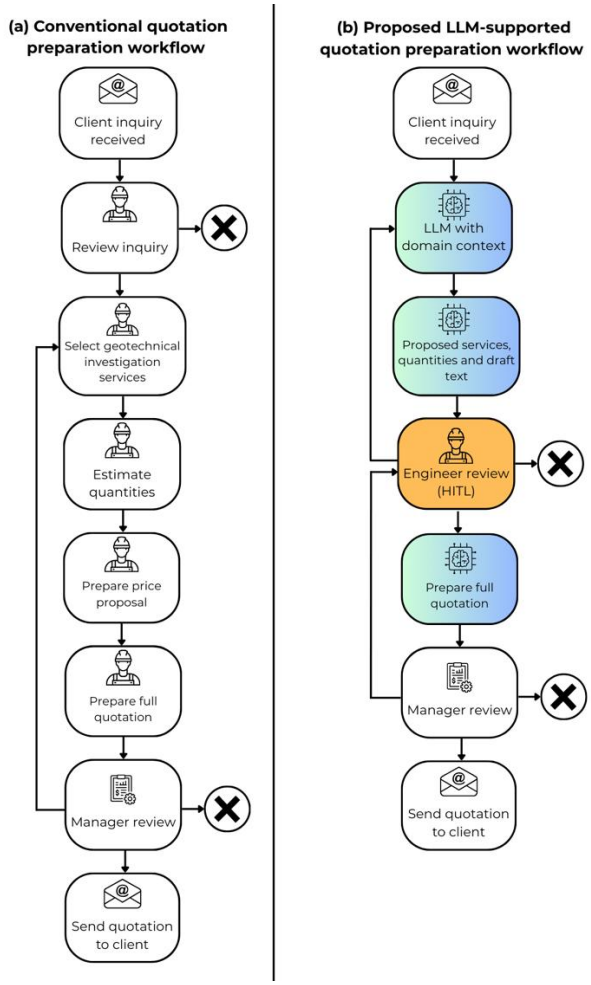


Figure 1: (a) Conventional quotation preparation workflow and (b) Proposed LLM-supported quotation preparation workflow.

The LLM generates a proposed set of service items, suggested quantities, and a draft quotation text. The engineer reviews the proposed output, makes corrections where necessary, and approves the final quotation. This HITL approach ensures that responsibility remains with the engineer while reducing the time required for routine tasks.

Quotation preparation combines structured data (service items, quantities, prices) with semi-structured and unstructured textual information (project descriptions, assumptions, and conditions). These characteristics make the process well suited for LLM, as LLMs are particularly effective at processing natural language and generating coherent technical text.

The repetitive nature of quotation preparation and the availability of historical quotations enable the creation of domain-specific training and evaluation datasets, supporting model adaptation and performance assessment.

The proposed workflow, shown in Figure 1, shifts the engineer's role from manual creation (a) to supervisory validation and refinement (b).

In this context, the HITL design is not just a safeguard but a critical architectural choice for deploying AI systems in construction workflows. The LLM serves as a cognitive assistant, while engineers retain full authority over validation, correction, and approval. This ensures that accountability and compliance remain firmly anchored in human expertise, which is essential in a highly regulated and safety-sensitive industry. HITL enables explainability, auditability, and trust, which are key requirements when AI-generated outputs may affect technical or financial decisions. Recent studies emphasized that in many industrial domains, fully autonomous AI agents may not be justified, and AI is more effective when deployed as a supervised contributor, particularly in contexts demanding trust, reliability, and regulatory compliance (He et al., 2024; Kotsiopoulos et al., 2024).

Methodology

We evaluated both commercial and open-source LLMs for supporting quotation preparation. The commercial models included GPT 4.1 and GPT 5 (OpenAI), and Claude Sonnet 4.5 (Anthropic). As an open-source baseline, we used QwQ 32B in (i) its base form and (ii) a fine-tuned version adapted to the company domain using anonymized examples from past project inquiries.

To assess the impact of supplementary project information, GPT 4.1 and GPT 5 were evaluated in two input configurations: (a) inquiry text only and (b) inquiry text plus attachments (when provided with the client inquiry). All models were prompted using a unified prompt template requiring a structured output (JSON) to ensure comparability across model families. This schema was designed to reflect real quotation workflows, and included fields such as project metadata, client inquiry content, and the desired response structure (e.g., service item list, estimated quantities, response text). Prompts were iteratively refined across models to ensure consistency. Domain knowledge was embedded into the prompts through a combination of static context injection (e.g., glossaries, service catalog excerpts, company-specific terminology) and heuristic templates derived from past quotation examples. Due to contractual and confidentiality constraints, external tool integration (e.g., retrieval-augmented generation) was not used, and no real-time application programming interface (API) calls were permitted.

Listing 1: Example prompt structure (abbreviated)

SYSTEM PROMPT:

Task: *Based on the client inquiry, generate a structured JSON listing of required geotechnical services, quantities, and a brief rationale.*

Context: *The company is based in Melbourne, Australia.*

Instructions:

1. *Review the client inquiry.*
2. *Select required work items from a predefined list.*
3. *Estimate quantities and justify selections in one sentence each.*
4. *Output a JSON array including item name, quantity, and justification.*

CLIENT INQUIRY *(excerpt):*

"…4 boreholes to 10 m or refusal. Provide retaining wall parameters and bearing capacity. Northern wall = gravity, Southern = MSE…"

EXPECTED OUTPUT FORMAT (JSON):

```
[
{"work_item": "Soil Drilling (<10m)", "quantity": 4},
{"work_item": "CBR Testing", "quantity": 3},
{"work_item": "Geotechnical Report", "quantity": 1}
]
```

The evaluated models were selected to represent both state-of-the-art commercial systems and high-capacity open-source alternatives. GPT 4.1 and GPT 5 (OpenAI) were included due to their status as leading commercial LLMs at the time of our research, with broad adoption and strong performance across technical NLP tasks. Claude Sonnet 4.5 (Anthropic) was selected based on reports of fast inference and strong reasoning in applied scenarios. QwQ 32B was chosen as a high-capacity open-source model suitable for domain fine-tuning. Its fine-tuned variant (QwQ 32B FT) was trained on domain-specific data to explore the feasibility of internal model adaptation without commercial licensing. Simpler baselines (e.g., BERT variants, rule-based systems) were not included, as the application scenario required generative capabilities and flexible outputs.

We compiled a dataset of 596 quotation inquiries collected over several years from a geotechnical

engineering firm. These inquiries vary in format, language quality, and complexity, reflecting realistic working conditions. From this dataset, we extracted a representative evaluation set of 120 inquiries, which were manually annotated with expected service items and quantities for comparison.

Fine-tuning of the open-source model

The open-source model was adapted using a two-stage, parameter-efficient fine-tuning approach:

1. Domain adaptation (DA). The model was adapted on domain texts describing geotechnical investigation service items (including professional item descriptions) to incorporate company terminology and service scope knowledge.
2. Supervised fine-tuning on input-output (inquiry-quotation) pairs (I/O). The base model, augmented with the DA adaptation, was further fine-tuned using supervised learning on curated inquiry-quotation examples. Each training example consisted of (i) an instruction plus the anonymised inquiry text as input and (ii) a strict JSON output containing the selected service items and their quantities as the learning target. This stage was designed to improve both domain-appropriate item selection and adherence to a consistent, machine-readable output structure.

The resulting fine-tuned model is referred to as QwQ 32B-FT in the remainder of the paper.

Dataset

The dataset consists of 596 real inquiry-quotation pairs collected from a geotechnical engineering company. Each pair contains (i) the anonymised client inquiry (email text and associated attachments, where applicable) and (ii) the corresponding engineer approved quotation, from which the reference service items and quantities were extracted.

The dataset was divided into 476 training pairs (80 %) and 120 test pairs (20 %). The training set was used exclusively for fine-tuning QwQ 32B-FT. The 120 pair test set was used for evaluation of all models.

Due to the signed NDA and client confidentiality, original data cannot be disclosed. However, a manually constructed mock example is provided below to illustrate the structure of a typical case:

Listing 2: Example inquiry-quotation pair

Inquiry (excerpt):

We are preparing a residential development in Werribee. Please include standard fieldwork, Atterberg limits, PSD tests, and recommendations for slab-on-ground foundations.

Output (excerpt):

[{"work_item": "Soil Drilling (<3m)", "quantity": 2},
{"work_item": "Atterberg Limits", "quantity": 2},
{"work_item": "PSD Testing", "quantity": 2},
{"work_item": "Geotechnical Report", "quantity": 1}]

Evaluation metrics

Performance was quantified using both quality and operational metrics. Service item selection is treated as a multi-label selection problem, evaluated using precision, recall, and F1 score. In addition, we measured quantity accuracy and processing time.

Quantity accuracy is the ratio of correctly determined quantities to all quantities proposed by the model for the selected items.

Processing time is the end-to-end time required to generate a complete structured output for the inquiry set. The paper reports the total time for 120 inquiries and average time per inquiry.

Results

This section reports model performance on the 120 inquiry test pairs using item selection metrics (precision, recall, F1), quantity accuracy, and processing time. Table 1 summarises the quantitative comparison across GPT 4.1, GPT 5, Claude Sonnet 4.5, QwQ 32B base, and QwQ 32B-FT. Figure 2 visualises the trade-off between effectiveness (F1) and speed (processing time).

GPT 5 with attachments achieved the highest F1 score (50%), driven by the highest recall (65 %), while Claude Sonnet 4.5 achieved a similar F1 score (48 %) with the fastest processing time (6.84 s/inquiry). The open-source baseline QwQ 32B performed substantially worse than commercial models on all quality metrics. QwQ 32B-FT improved precision (from 11 % to 21 %) and F1 score (from 8 % to 18 %) but remained far below the commercial LLMs. Quantity accuracy is low for all evaluated models (≤ 15 %), indicating that quantity determination remains a major bottleneck even when item selection improves.

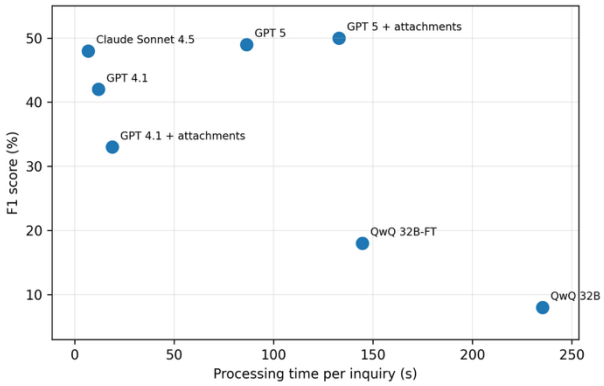


Figure 2: Scatter plot of F1 score relative to processing time.

Figure 2 shows that Claude Sonnet 4.5 offers the best balance between speed and quality (high F1 score at low latency), while GPT 5 achieves a slightly higher F1 score but with significantly longer runtime, especially when attachments are included. Figure 2 also shows that the fine-tuned version of the open-source model (QwQ 32B-FT) improved significantly in processing time as well as in F1 score, compared to the base model QwQ 32B.

Model	Precision (%)	Recall (%)	F1 score (%)	Quantity accuracy (%)	Total processing time (h)	Processing time per inquiry (s)
GPT 4.1	41	52	42	10	0,4	12,0
GPT 4.1 + attachments	38	38	33	8	0,63	18,9
GPT 5	43	63	49	13	2,88	86,4
GPT 5 + attachments	43	**65**	**50**	**15**	4,43	132,9
Claude Sonnet 4.5	**45**	55	48	11	**0,228**	**6,84**
QwQ 32B	11	7	8	1	7,84	235,2
QwQ 32B-FT	21	17	18	5	4,82	144,6

Table 1: Model performance on 120 inquiries.

Figure 3 shows that quantity accuracy remains consistently low across all evaluated models, with the best-performing configuration (GPT 5 with attachments) achieving only 15 %. Even models that perform well in service item selection and text generation do not reliably determine numerical quantities. This confirms that quantity estimation is the main bottleneck for higher levels of automation in quotation preparation and supports the use of hybrid workflows, where LLMs propose candidate items and draft text, while quantities are computed or verified using deterministic or rule-based tools.

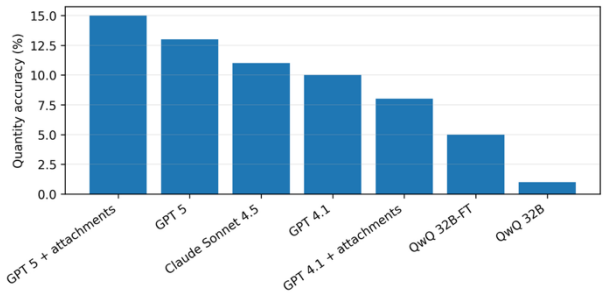


Figure 3: Quantity accuracy across evaluated model configurations.

Although all the accuracies were low, there is a notable increase in accuracy for the fine-tuned open-source model QwQ 32B-FT (4%) compared to the base model QwQ 32B (1%).

Discussion

The results indicate that commercial LLMs can meaningfully support quotation preparation by improving the accuracy of selecting geotechnical investigation service items, but they do not yet provide reliable quantity estimation without additional safeguards. Across models, the highest F1 score for item selection is approximately 50 %, meaning that even the best-performing configuration on average still misses or misidentifies a substantial portion of required items. This level of performance supports the proposed human-in-the-loop workflow: the models can accelerate drafting and reduce manual effort, but responsibility and final validation must remain with engineers.

A key practical finding is the quality-latency trade-off. GPT 5 with attachments achieves the best overall selection performance (highest F1 score and recall), but at a markedly higher processing time than other models. In contrast, Claude Sonnet 4.5 achieves an F1 score close to the best-performing GPT 5 configuration while being the fastest model in this study. For deployment in a real quotation preparation workflow, where the primary goal is to quickly provide a strong draft for engineer review, Claude Sonnet 4.5 appears to offer the most favourable operational profile.

The effect of attachments is not uniform. For GPT 5, attachments provide small improvements in recall and F1 score, suggesting that additional project context can help when the model can reliably integrate it. For GPT 4.1, however, adding attachments decreases performance sharply. This suggests that naively appending additional documents may introduce noise, cause relevant content to be diluted, or trigger truncation effects in long contexts.

The persistently low quantity accuracy across all models (≤15 %) is the most important limitation for practical adoption. This aligns with broader observations in construction-focused LLM research: generative models can produce fluent text and plausible suggestions, but numerical correctness and strict domain constraints remain challenging, especially without structured inputs and verification mechanisms. For quotation preparation, this motivates a hybrid architecture. LLMs can propose candidate items and draft explanatory text, while quantities should be computed or checked by deterministic tools (e.g. rule-based calculators) and engineer before finalisation.

The open-source model results clarify the current limits of local deployment. Although QwQ 32B-FT improved over the base QwQ 32B model, it did not reach competitive performance against commercial models. Moreover,

processing time remained high, undermining the operational benefits of automation. This does not preclude open-source adoption, but it indicates that achieving comparable performance likely requires larger models, higher-quality domain datasets, stronger retrieval augmentation, or more extensive fine-tuning, each requiring non-trivial engineering and infrastructure effort.

Finally, several limitations affect generalisability. The dataset originates from a single (geotechnical) company and a specific service catalogue, so model behaviour may differ in other contexts. The evaluation focuses on structured outputs (items and quantities), and the quality of the generated quotation text was not scored with a dedicated metric, although it is operationally important. Although the use case focuses on inquiries for geotechnical investigations, the results, and especially the workflow, are applicable across the entire construction domain.

While this study focuses on generative LLMs for their flexibility and alignment with comapny expectations, we acknowledge the absence of classical baselines such as BERT classifiers or rule-based matching systems. These approaches have been successfully applied to other construction-related tasks but were not included here due to the specific need for coherent text generation and adaptive item selection. Additionally, time and resource constraints limited the scope of model benchmarking. Future work could extend the evaluation to include retrieval-augmented generation or hybrid architectures that incorporate rule-based precision.

A qualitative review of model outputs suggests that accuracy was higher for inquiries with clear structure or phrasing patterns similar to previously encountered examples. In contrast, performance declined for inquiries that were vague, combined multiple project scopes, or included fragmented or informal language. Quantity errors mainly resulted from incomplete or ambiguous inputs. Many inquiries lacked explicit numerical requirements or used inconsistent units. These findings highlight the importance of structured input and human oversight when deploying LLMs in quotation workflows.

Limitations and Future work

This study is limited by its focus on a single company and service domain, which may affect generalisability to other construction contexts. Although models were evaluated using structured outputs, the quality of generated text was not assessed, despite its relevance in operational settings. Classical baselines and retrieval-augmented generation pipelines were excluded due to project scope and the requirement for flexible, generative outputs. Additionally, quantity estimation accuracy remains low across all models, highlighting the need for structured inputs and complementary rule-based tools. Future work could address these gaps by evaluating across multiple organisations, benchmarking against hybrid approaches, and assessing text quality and compliance in practical deployment.

Conclusions

This paper addresses the acceleration of quotation preparation using LLMs, where expert judgement is required for selecting appropriate service items and drafting quotation text. We designed an LLM-supported quotation preparation workflow and empirically evaluated several LLMs on a real dataset of inquiry-quotation pairs for geotechnical investigations, comparing commercial models (GPT 4.1, GPT 5, Claude Sonnet 4.5) with an open-source baseline (QwQ 32B) and its fine-tuned version (QwQ 32B-FT).

The results show that commercial LLMs significantly outperform open-source models in item selection, with GPT 5 (+ attachments) achieving the highest F1 score and Claude Sonnet 4.5 offering the best balance between F1 score and processing time. However, quantity accuracy is consistently low across all models, indicating that reliable quantity determination remains the main barrier to higher levels of automation.

Overall, the findings support the adoption of LLMs as drafting and decision-support tools within a human-in-the-loop process, while highlighting the need for hybrid architectures that combine LLM generation with deterministic quantity computation and validation. These findings underscore the value of HITL approaches not only as operational safeguards, but as foundational design choices that align AI support tools with the needs of high-stakes engineering domains. Future work should extend evaluation to multiple organisations, quantify text quality and compliance, and integrate rule-based or tool-assisted quantity estimation to reduce risk and improve end-to-end quotation reliability.

Acknowledgments

This research was supported by the Slovenian Research and Innovation Agency (ARIS) under the Young Researcher funding program and research program E-Construction (E-Gradbeništvo: P2-0210).

References

Brelih, A., Robnik Šikonja, M., Klinc, R. (2025) The Role of AI Agents in Construction Project Management, Tagungsband des 36. Forum Bauinformatik : 24.-26. September 2025 = Proceedings of the 36th Forum Bauinformatik 2025 / herausgegeben von Baris Özcan, Hristo Vassilev, p. pages 147–154. Available at: https://doi.org/10.18154/RWTH-CONV-254903.

Devlin, J., Chang, M-W., Lee, K., Toutanova, K. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv. Available at: https://doi.org/10.48550/arXiv.1810.04805.

Gatto, C., Mirarchi, C., Pavan, A. (2025) Towards Automation in Cost Estimation: LLM-based Methodology for Classifying and Extracting Cost Data from Tender Documents, in. 2025 European Conference on Computing in Construction. Available at: https://doi.org/10.35490/EC3.2025.370.

He, Y., Wang, E., Rong, Y., Cheng, Z., Chen, H. (2024) Security of AI Agents. arXiv. Available at: https://doi.org/10.48550/arXiv.2406.08689.

Kotsiopoulos, T., Papakostas, G., Vafeiadis, T., Dimitriadis, V., Nizamis, A., Bolzoni, A., Bellinati, D., Ioannidis, D., Votis, K., Tzovaras, D., Sarigiannidis, P. (2024) Revolutionizing defect recognition in hard metal industry through AI explainability, human-in-the-loop approaches and cognitive mechanisms, Expert Systems with Applications, 255, p. 124839. Available at: https://doi.org/10.1016/j.eswa.2024.124839.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W-T., Rocktäschel, T., Riedel, S., Kiela, D. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks, in Advances in Neural Information Processing Systems. Curran Associates, Inc., pp. 9459–9474. Available at: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

Liu, Z., Yao, W., Zhang, J., Yang, L., Liu, Z., Tan, J., Choubey, P. K., Lan, T., Wu, J., Wang, H., Heinecke, S., Xiong, C., Savarese, S. (2024) AgentLite: A Lightweight Library for Building and Advancing Task-Oriented LLM Agent System. arXiv. Available at: https://doi.org/10.48550/ARXIV.2402.15538.

Mikhridinova, N., Yurdakul, F.A., Wolff, C. (2024) Using Large Language Models for Project Staffing: Evaluation of GPT-Based Mapping of Teams to Projects, in Proceedings of the 12th IPMA Research Conference Project Management in the Age of Artificial Intelligence. 12th IPMA Research Conference Project Management in the Age of Artificial Intelligence, International Project Management Association – IPMA, IPMA USA, pp. 1–19. Available at: https://doi.org/10.56889/cykr4175.

Ruan, J., Chen, Y., Zhang, B., Xu, Z., Bao, T., Du, G., Shi, S., Mao, H., Li, Z., Zeng, X., Zhao, R. (2023) TPTU: Large Language Model-based AI Agents for Task Planning and Tool Usage. arXiv. Available at: https://doi.org/10.48550/ARXIV.2308.03427.

Taboada, I., Daneshpajouh, A., Toledo, N., de Vass, T. (2023) Artificial Intelligence Enabled Project Management: A Systematic Literature Review, Applied Sciences, 13(8), p. 5014. Available at: https://doi.org/10.3390/app13085014.

Turk, Ž. (2006) Construction informatics: Definition and ontology, Advanced Engineering Informatics, 20(2), pp. 187–199. Available at: https://doi.org/10.1016/j.aei.2005.10.002.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., Yin, Z., Dou, S., Weng, R., Cheng, W., Zhang, Q., Qin, W., Zheng, Y., Qiu, X., Huang, X., Gui, T. (2025) The rise and potential of large language model based agents: a survey, Science China Information Sciences, 68(2), p. 121101. Available at: https://doi.org/10.1007/s11432-024-4222-0.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

MULTIDIMENSIONAL WEIGHTED RANKING RECOMMENDATION SYSTEM FOR CIRCULAR MATERIAL MATCHMAKING

Noah Gerber[1], Ibrahim Al-Hazwani[2], and Pei-Yu Wu [3]
[1]ETH Zurich, Zurich, Switzerland
[2]University of Zurich, Zurich, Switzerland
[3]Delft University of Technology, Delft, The Netherlands

Abstract

Circular construction relies on digital marketplaces to overcome supply chain fragmentation, yet manual search for reclaimed components remains a critical bottleneck. This research develops a context-aware recommendation system for reusable material matchmaking. Utilizing a multidimensional weighted ranking algorithm with hard and soft constraints, the system adaptively aligns material attributes with project requirements. Validated through expert elicitation and a design simulation user study, results demonstrate high material findability and low cognitive load. However, insufficient algorithmic transparency can induce user friction. This work establishes a foundation for role-centric matchmaking, with future research focusing on similarity score cutoffs, enhanced interpretability, and API-driven cross-platform interoperability.

Introduction

Digital material marketplaces are evolving into mediatised platforms that foster a circular construction sector by enabling the sustainable reuse of building materials. Nevertheless, many existing reuse platforms operate within isolated local communities, preventing economies of scale and restricting the volume of available secondary products compared to primary materials. Consequently, the low quantity and high diversity of reclaimed batches often preclude their procurement in new construction projects. Recent literature identifies several critical barriers to platform effectiveness: (1) structural and economic obstacles such as high upfront costs; (2) operational constraints caused by the lack of spatial and temporal synchronization between deconstruction and construction phases; (3) informational gaps stemming from a lack of automated data entry and standardized performance documentation (Wilson et al., 2025; Quertamp et al., 2025); and (4) low user adoption. Notably, 70% of marketplaces prioritize auxiliary services (e.g., building inventory, logistics) over core digital infrastructure (Menny et al., 2024), resulting in underutilized databases that fail to reach the critical mass necessary to bridge the gap between supply and demand.

To address these challenges, research suggests transitioning from project-centered to role-centric matchmaking through asynchronous loops that mitigate strict synchronization requirements (Menny et al., 2024). Technological interventions such as blockchain, the InterPlanetary File System (IPFS) (Protocol Labs, 2026), and Digital Product Passports (Honic et al., 2024) are being explored to ensure traceability (Wilson et al., 2025). Furthermore, the integration of standardized circularity indicators and interoperable APIs aligned with Circular Economy standards (ISO/TC 323, 2018) facilitates cross-platform communication. Finally, participatory action research and regulatory alignment with Extended Producer Responsibility laws ensure that digital tools meet real-world operational needs and compliance requirements (Podkalicka and Fredriksson, 2023; Quertamp et al., 2025).

While these frameworks provide the necessary infrastructure for data exchange, their utility depends on early-stage intervention within the design process to maximize reuse potential before in-situ reclamation. Building upon this requirement, this study develops a Context-Aware Recommendation System (CARS) designed to facilitate the matchmaking of both in-situ components and off-site reclaimed materials. By employing a multi-phased participatory design approach with actors from Swiss marketplaces, this research seeks to: (1) identify critical decision-making factors to encode heterogeneous inputs into a unified schema; (2) integrate temporal and geospatial preferences into a context-aware engine to enhance logistical viability; and (3) provide explainable recommendations to foster user trust. To achieve this, the system encodes semantic and quantitative material records into a multidimensional vector database, facilitating similarity matching based on the mathematical proximity between dynamic user requirements and material attributes. By incorporating image data and explainability features, the system addresses the ”black box” friction common in traditional vector retrieval.

The paper demonstrates the feasibility of integrating a CARS into a user-friendly interface that delivers personalized matchmaking based on dynamic constraints. By enabling users to interactively adjust priority levels across multilingual inventories, the system generates ranked outcomes with matching scores that reveal the underlying weighted logic. This guides users to iteratively refine search parameters and explore alternative material choices when exact matches are insufficient. Preliminary findings validate that high operational speed and intuitive weighted search logic effectively bridge the gap between technical retrieval and architectural design rationale. However, the results also highlight a critical need for enhanced semantic

transparency and similarity thresholds to ensure reliability as databases scale. This research contributes a methodological framework for optimizing the quality and diversity of early-stage matchmaking in circular construction.

Related work

Relevance of context-aware recommender systems
CARS address the structural fragmentation and logistical constraints of material reuse by synchronizing spatial and temporal data. Unlike static systems, CARS incorporate dynamic attributes, such as disassembly windows and geographic proximity, to adapt results to immediate project needs (Adomavicius et al., 2011; Afzal et al., 2024). This approach mitigates the data sparsity and cold-start problems inherent in unique, transient secondary material databases by leveraging contextual constraints to estimate preferences when historical interaction data is absent (Kulkarni and Rodd, 2020). Consequently, the system models evolving professional requirements throughout the project lifecycle rather than relying on fixed interest models (Huang et al., 2024).

Distinctive nature of secondary matchmaking
Secondary material matchmaking differs from standard e-commerce due to high item heterogeneity and the absence of reciprocal supplier constraints. The process shifts from exact stock keeping unit matching to similarity matching of non-standardized components using quantitative, qualitative, and vector-based image data (Adomavicius et al., 2011). Furthermore, professional selection requires a multi-criteria evaluation framework rather than a single consumer rating (Afzal et al., 2024). Practitioners must concurrently weigh technical specifications, aesthetics, and environmental impact. This necessitates a ranking system capable of balancing diverse factors against specific professional workflows rather than simple popularity metrics.

Explainability in professional decision-support
Explainability transforms the recommendation engine from an opaque ”black box” into a transparent support framework, establishing the epistemic trust necessary for high-stakes construction decisions (Tintarev and Masthoff, 2015). By allowing users to explicitly define attribute weighting, such as prioritizing structural load over aesthetics, the system provides requirement-specific justifications for its suggestions (Lai et al., 2024). This transparency is vital for validating unique, reclaimed components. Explicitly describing how a material aligns with complex project constraints ensures that recommendations are both actionable and interpretable within professional discovery processes.

Methodological framework

This research employs a multi-phased participatory design approach to develop a context-aware matchmaking platform. Initial scoping involved consultations with experts from Swiss secondary material companies to elicit operational workflows and define technical metadata requirements. An iterative development phase followed, where stakeholders evaluated prototype design considerations and prioritized functionalities. Finally, the platform underwent empirical evaluation through a structured user study with four native German-speaking participants. This linguistic selection ensured valid interaction with the localized interface and database.

User requirements

Technical and metadata requirements
Effective matchmaking necessitates the integration of high-fidelity material attributes across three categories: fundamental dimensions and classification; performance metrics (e.g., thermal or acoustic properties and mounting specifications); and a standardized physical condition scale. Expert elicitation indicates that practitioners prioritize chronological age over manufacturer identifiers when assessing reclaimed value. Furthermore, high-quality visual documentation is critical for the remote assessment of unique components. To ensure interoperability, the search schema aligns with DIN SPEC 91484 (DIN German Institute for Standardization, 2024) and Digital Product Passport frameworks (Honic et al., 2024), facilitating a unified infrastructure for cross-platform exchange.

Role-specific functional requirements
Matchmaking architectures must accommodate the divergent workflows of specialized professional groups. Architects and designers prioritize reliability and contextual inspiration, typically matching available inventories against pre-defined design constraints. Conversely, business owners and contractors emphasize carbon savings and regulatory compliance over initial procurement costs. To ensure functional integrity, these stakeholders require integrated services such as quality assurance, professional testing, and specialized cleaning. The system addresses these needs through a dual-announcement model comprising both supply listings and search requests. This bifacial architecture facilitates a proactive procurement process tailored to the logistical demands of circular economy actors.

Operational and logistical requirements
Systemic viability depends on managing the physical constraints of the construction sector. Geospatial filters are integrated to mitigate transport costs, which frequently dictate the economic feasibility of secondary materials. To address the primary cost drivers in the reuse cycle, the system supports material demand prediction to optimize storage management. Additionally, operational data regarding deconstruction viability (e.g., ease of disassembly, fragility, and weight) is essential for sourcing materials in situ.

Psychological and soft factors
Material provenance, historical context, and environmental significance are decisive factors in the selection of reclaimed components. To manage the complexity of

non-standardized items, the system utilizes intuitive interface features designed to minimize cognitive load and user anxiety. Successful matchmaking requires a synthesis of a component's history, physical health, and spatiotemporal availability. Aligning these narrative qualities with rigorous technical constraints ensures a reliable material exchange.

Design goals

Based on the research framework and expert elicitation, five core design goals are established to transition digital marketplaces toward multidimensional decision support.

1. Multidimensional similarity matching
The system implements similarity matching to shift from rigid queries to the flexible discovery of heterogeneous components. Representing qualitative and quantitative dimensions within a vector database enables the engine to suggest functional alternatives aligned with latent project objectives rather than restricted search terms.

2. Spatiotemporal and logistic optimization
Geospatial proximity and temporal availability are integrated as dynamic factors managed through user-defined priority levels. This addresses the critical synchronization required between in-situ deconstruction windows and reconstruction schedules, ensuring logistical feasibility.

3. Role-centric recommendation logic
The engine adapts weighting parameters to the distinct priorities of stakeholders. While individual users may focus on cost, professional architects often emphasize carbon savings and technical reliability. Tailoring the ranking logic to these requirements ensures professional relevance.

4. Trust through explainability and user experience design
Transparent recommendations detail the reasoning behind matches, such as shared mounting specifications or condition ratings. This transparency, combined with streamlined interaction patterns, bridges the trust gap regarding the performance of reclaimed materials in a risk-averse sector.

5. Data interoperability and aggregation
The platform functions as a meta-aggregator to address market fragmentation. By aligning with international standards and aggregating data across multiple engines, the system provides a comprehensive regional overview, ensuring recommendations are derived from high-quality, interoperable datasets.

System overview

Material data schema

The data integration and structural representation, denoted as M, synthesizes metadata of each material item into an integrated schema.

$$M = \{\text{id}, I, Q, P, S, L, T\}$$

$$\begin{cases} I = (T, D) & \text{Title and description} \\ Q \in \{\text{new, good, poor}\} & \text{Quality interval} \\ P \in \mathbb{R}_{\geq 0} & \text{Unit price} \\ S = (w, h, d) \in (\mathbb{R}^+)^3 & \text{Volumetric Dimensions} \\ L = (\phi, \lambda) & \text{Geospatial coordinates} \\ T = (t_{start}, t_{end}) & \text{Temporal period} \end{cases} \quad (1)$$

The material schema is formalized as a multi-layered tuple. The identification layer $\{\text{id}, I\}$ integrates a unique identifier with a descriptive metadata. Economic and qualitative dimensions are represented by unit price P and material integrity Q, with the latter categorized into three discrete intervals to facilitate standardized circular recovery assessment. Finally, the system encodes physical, geospatial, and temporal constraints via S, L, and T, which define volumetric attributes (w, h, d), coordinates (ϕ, λ), and the operational validity period (t_{start}, t_{end}).

Platform architecture

The platform is engineered for data ingestion and semantic retrieval using a serverless framework. The system is partitioned into two distinct operational pipelines leveraging edge computing to minimize latency and ensure linear scalability, as illustrated in Figure 1.

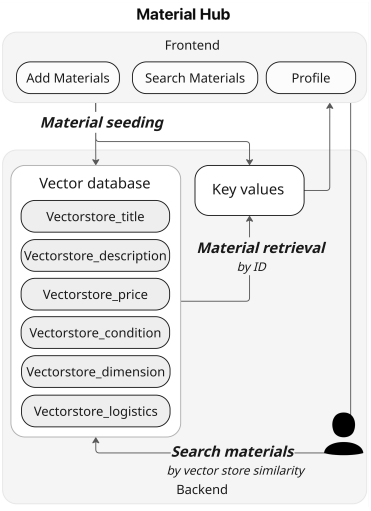


Figure 1: System archtiecture and workflow.

Material seeding pipeline

The seeding pipeline initiates with edge-based input processing via Cloudflare Workers. These functions execute data normalization and validation of attributes, including Swiss eBKP classification codes (Swiss Society of Engineers and Architects, 2020), pricing, and dimensions. Following refinement, the Gemini Embedding model (Lee et al., 2025) transforms textual descriptions and technical

parameters into high-dimensional vectors. These representations are indexed within Vectorize, a distributed vector database maintaining the latent space of all material profiles.

Material retrieval pipeline

The retrieval pipeline facilitates semantic search by translating natural language queries into vector representations. The architecture executes similarity analysis using a k-nearest neighbor (k-NN) algorithm to calculate the cosine similarity between the query vector and stored indices. By identifying elements based on spatial proximity in the latent space, the system overcomes the limitations of traditional keyword matching for non-standardized material descriptions.

Material recommendation system widget

Integrating specifications into the digital *Material Hub* (Figure 2) triggers a ranked list of recommendations based on calculated matching scores. System widgets allow for the iterative adjustment of preference weights, enabling real-time exploration of alternative candidates.

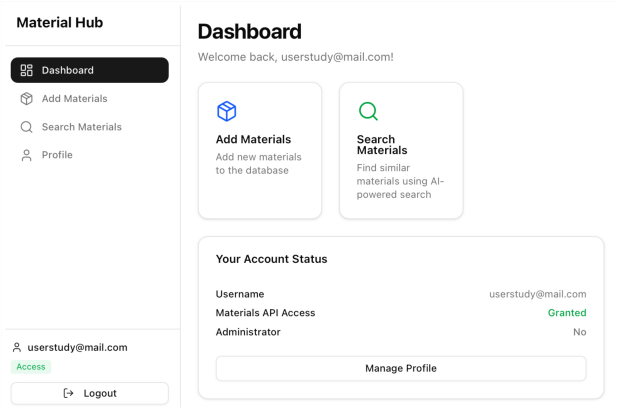


Figure 2: Material Hub dashboard prototype highlighting the integrated modules for material registration, multidimensional search, and user profile management.

Material data ingestion

The system supports dual-mode data ingestion: automated batch processing via inventory files (e.g., pre-renovation or pre-demolition audits) or manual entry, as shown in Figure 3. The automated pipeline utilizes Large Language Models (LLMs), specifically Llama 3.1 8B Instruct (Meta AI, 2024), to parse unstructured data and extract material properties. In cases of data scarcity, the system maintains integrity by imputing missing values based on category averages. Imported materials are automatically mapped to Swiss eBKP-H classification codes (Swiss Society of Engineers and Architects, 2020) to standardize heterogeneous documentation into a structured database.

Preference adjustments

To optimize discovery within multidimensional decision spaces, the system utilizes a weighted vector retrieval

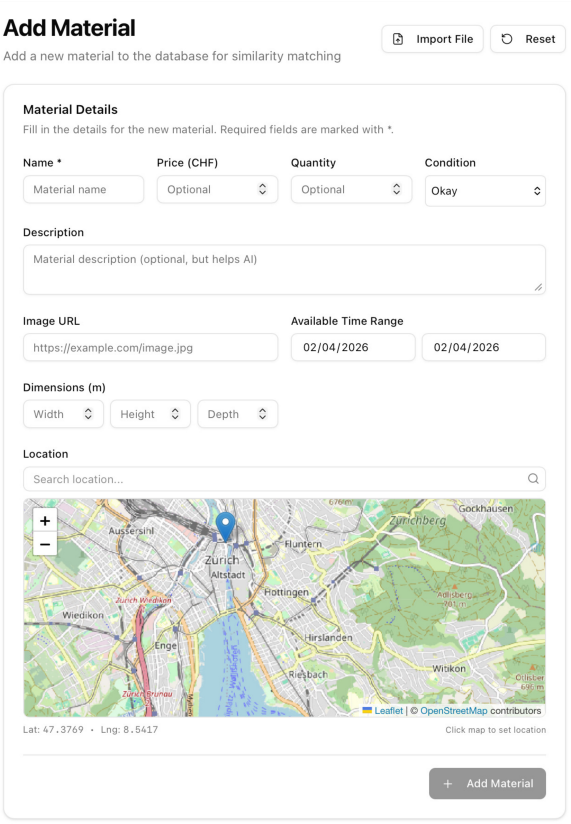


Figure 3: User interface for material data ingestion by batch processing or manual entry.

mechanism. Users calibrate the recommendation engine by assigning discrete priority levels to specific attributes, which are translated into quantitative weights. This process employs a dual-constraint logic: hard constraints serve as binary filters to exclude non-compliant materials, while default soft constraints facilitate fuzzy matching to preserve high-potential alternatives.

Geospatial logistics and proximity

The system integrates an interactive mapping interface powered by OpenStreetMap to manage geographical constraints. Users define a central project coordinate and a search radius, restricted to a maximum of 100 km. This parameter is empirically derived from expert interviews, reflecting the logistical threshold where transportation costs and environmental impacts typically become counterproductive for reclaimed materials.

Global weight distribution and ranked results

To enhance interpretability, the system visualizes the contribution of individual attributes to the final similarity scores. As illustrated in Figure 5, the global weighted attributes of the aggregated k-matched items are displayed in a stacked bar chart.

The top k items are ranked using a late-fusion integration logic that synthesizes unstructured semantic data with structured constraints, presented in modular cards in Figure 6. To provide granular transparency, each entry dis-

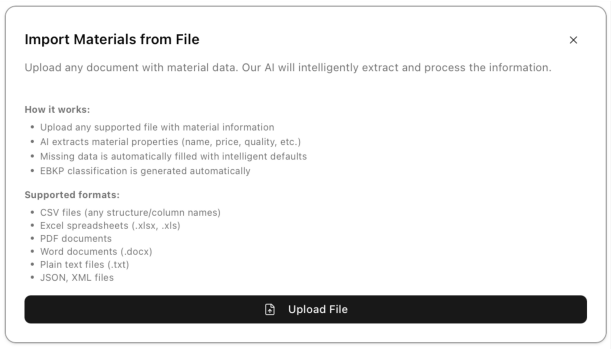


Figure 4: Material information import, extraction, and standardization.

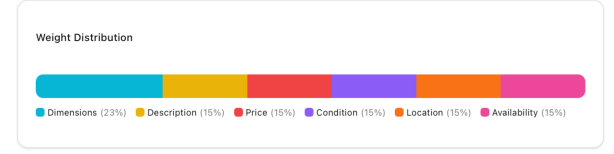


Figure 5: Global weight of attribute-level contribution to the similarity scores of k-matched items.

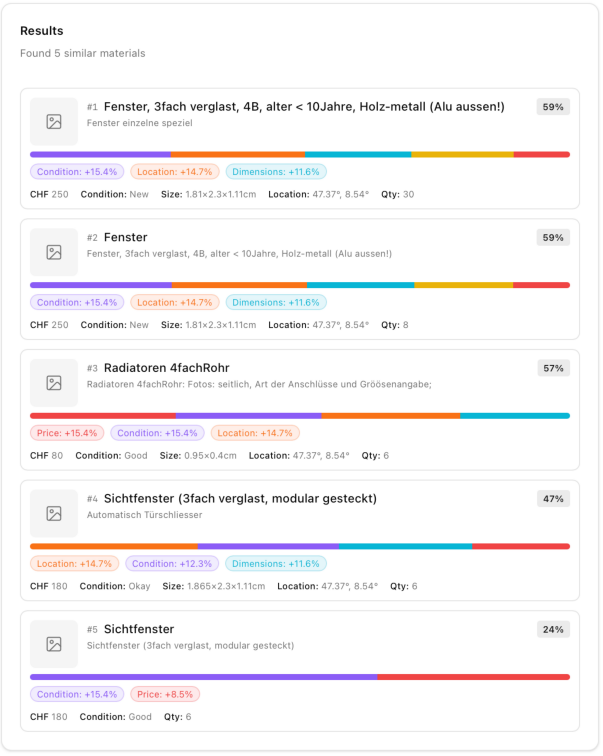


Figure 6: Ranked material candidates with localized attribute decomposition for the search query "windows."

plays a "local weight" representing the aggregate alignment of its specific attributes. We define $s_{\text{sem}}(m_i)$ as the cosine similarity between embeddings, and $s_j(m_i) \in [0,1]$ as normalized scores for J structured preference criteria (e.g., distance, quantity, and condition). The final ranking score $S_{\text{final}}(m_i)$ is produced via linear aggregation: Items are sorted in descending order of $S_{\text{final}}(m_i)$. This framework ensures the ranking remains interpretable, allowing users to discern the relative importance of latent semantic relevance versus explicit attribute requirements.

$$S_{\text{final}}(m_i) = w_{\text{sem}}\, s_{\text{sem}}(m_i) + \sum_{j=1}^{J} w_j\, s_j(m_i),$$

$$w_{\text{sem}} + \sum_{j=1}^{J} w_j = 1, \quad w_{\text{sem}}, w_j \geq 0.$$

System evaluation

This section presents a multi-faceted evaluation of the platform, assessing technical efficiency via quantitative metrics, retrieval accuracy through a hybrid ranking analysis, and practical utility via controlled user studies.

Performance assessment

Performance was evaluated based on latency (end-to-end processing duration) and scalability (performance stability relative to repository growth). Latency measurements across the retrieval pipeline, comprising embedding generation, vector similarity search, deterministic filtering, and final scoring, indicate that response times remain within thresholds for real-time interaction. Vector similarity search is the most intensive stage. However, the implementation of approximate nearest neighbor indexing significantly reduces retrieval time compared to linear scans. Embedding generation and final ranking operations contribute marginal overhead, as the latter processes a pruned candidate subset.

Incremental dataset scaling experiments confirm that the approximate nearest neighbor indexing structure maintains sub-linear growth, avoiding exhaustive comparisons. This scalability and throughput is augmented by a serverless architecture that enables concurrent operations through automatic worker instantiation. These findings suggest the architecture can support expanding marketplaces without significant performance degradation.

Retrieval quality and error analysis

Preliminary quality assessments verified the alignment of retrieved materials with semantic intent and structured constraints. A comparative analysis between purely semantic and hybrid strategies reveals that while semantic-only retrieval excels at discovering conceptually related materials with divergent terminology, it frequently identifies candidates that fail physical or logistical requirements (e.g., geographic proximity). The hybrid pipeline resolves this by integrating deterministic filtering, successfully pruning unsuitable results while preserving the discovery advantages of vector search.

Error analysis identified three primary failure modes: (1) data sparsity, where incomplete descriptions prevent high-fidelity semantic representations; (2) category overlap,

where distinct materials (e.g., panels vs. beams) share similar descriptive language; and (3) linguistic ambiguity, where vague queries produce broad, low-precision result sets. These findings underscore that structured metadata is essential for resolving the inherent ambiguities of natural language in technical domains.

User study

The user study evaluated the platform's functional performance across four dimensions: usability, retrieval efficacy, design alignment, and qualitative feedback. Participants conducted a design simulation for a temporary pavilion in Zurich, navigating a database to source technical components (e.g., glazing units, HVAC systems) under specific lifespan and safety constraints. This iterative process, as shown in Figure 7, facilitated a multidimensional analysis of system usability, the efficacy of the recommendation engine in discovery, and the alignment of results with architectural design rationales.

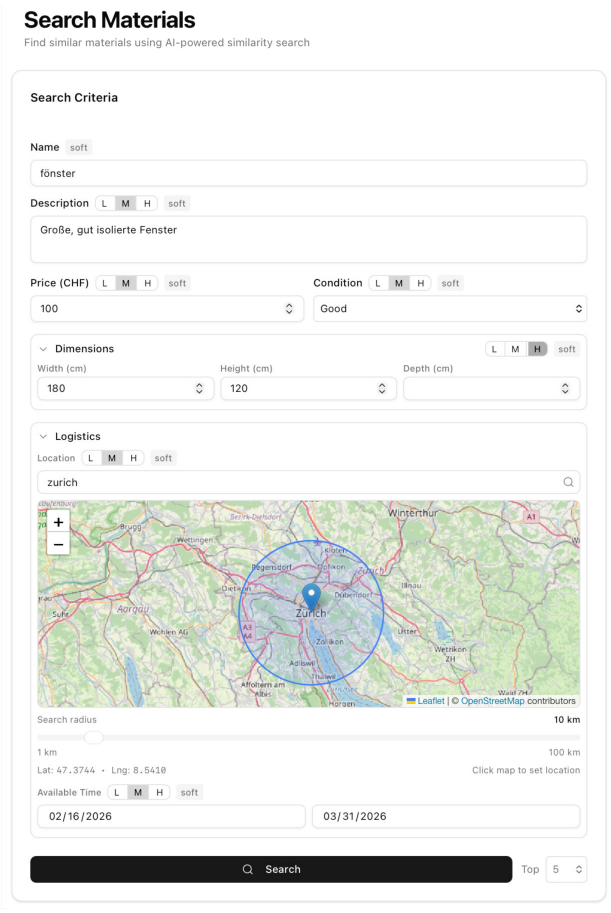


Figure 7: User interface for multidimensional vector search with constraint-based material retrieval logic.

System usability and user experience

Usability was assessed via metrics adapted from the System Usability Scale (Lewis, 2018). Participants reported low system complexity and high consistency, evidenced by high learnability scores (Median = 4.5) and low cognitive load. Technical confidence remained high (Median = 4.0), though variability (SD = 1.50) indicates that individual technical self-efficacy influences the user experience. Workflow fluidity analysis showed strong functional integration between vector search and filtering (Mean = 3.75, SD = 0.50). While 75% of participants perceived the system as highly consistent, divergent scores in intended frequency of use suggest that perceived utility is sensitive to individual professional context.

Information retrieval and discovery efficacy

The efficacy of the vector store and recommendation engine was validated by a mean search success score of 8.0 (SD = 1.15) on a 10-point scale, with no reported search failures. Participants identified "description", "name", and "size" as the most critical metadata for procurement decision-making. However, a "black-box" effect was observed regarding the ranking logic. Participants noted that weighted search results lacked sufficient descriptive interpretation, highlighting a requirement for greater algorithmic transparency and interpretability in the retrieval output.

Design rationale and requirement alignment

The simulation demonstrated the platform's ability to bridge technical retrieval with architectural application. Material selection was driven by semantic matches (type, color) and structured constraints (price, location, and age). Participants successfully identified components aligned with project requirements, demonstrating a correlation between search precision and the ability to justify circularity. However, search friction occurred when qualitative reasoning was required, suggesting that effective reuse advocacy depends on both engine precision and the user's prior technical knowledge.

Qualitative feedback and iterative improvement

Participants cited high operational speed and the weighting algorithm as primary drivers of efficacy. Interactive filters provided "positive friction", allowing users to refine discovery based on geographical requirements. Conversely, significant bottlenecks included high initial information density and the absence of "empty state" feedback. The vector engine's tendency to return loosely related items for unmatched queries, such as displaying glazing units for sanitary ware, caused user frustration.

Discussions

Theoretical positioning and contributions

A primary challenge in circular construction platforms is the semantic gap caused by non-standardized, short-text descriptions, which often triggers cold-start failures in traditional algorithms (Van Capelleveen et al., 2021). While existing systems rely on rigid feature crosses and static metadata (Tserng et al., 2025), they struggle with the linguistic ambiguity of "in-the-wild" data. Our research ad-

dresses this by implementing a late-fusion linear aggregation logic. By utilizing semantic embeddings to navigate jargon-heavy descriptions, the system identifies conceptually relevant materials that keyword-based searches exclude. This adaptive approach enables soft trade-offs where high semantic relevance compensates for minor attribute mismatches, ensuring findability despite the data sparsity identified in our error analysis.

Our evaluation confirms that algorithmic "black-box" behavior remains a significant source of user friction (Blackburn et al., 2026). Participants noted that while the ranking was successful, the rationale behind specific scores remained opaque. This validates the necessity of our proposed "glass-box" framework. By exposing ranking logic via local attribute decomposition, the system allows professional users to verify how semantic intent is balanced against physical requirements. This transparency is critical for reducing the cognitive load and decision-making friction observed during the user study, where a lack of descriptive interpretation was highlighted as a primary bottleneck.

The architecture is positioned to facilitate high-value component flows between demolition and construction, addressing the need for platforms to move beyond disciplinary silos (Blackburn et al., 2026). By adopting an API-ready framework, this work supports the platform-enabled coordination necessary to close resource loops. The multidimensional weighted ranking provides the governance logic required to manage ecosystem-wide interdependencies, shifting the focus from low-grade recycling to high-value resource retention and supporting circular construction through role-centric matchmaking.

Limitations and future work

The system evaluation revealed several technical and user experience limitations that define our future research pathway. The error analysis highlighted a critical need for a similarity score cutoff. Currently, the vector engine lacks a hard threshold, occasionally retrieving unrelated items (e.g., displaying glazing units for sanitary ware queries) when direct matches are absent. Implementing a calibrated cutoff will enable "no results found" notifications, preventing the user frustration observed during redundant filter refinement. While the weighting algorithm is technically sound, the user study identified a gap in semantic transparency. Future iterations will replace the current toggles with dynamic weight tuning and introduce semantic proximity labels. These features, alongside integrated imagery and empty-state feedback, will transform the ranking logic into a more interpretable tool, mitigating the information density issues reported by participants. To enhance data robustness, the architecture will be restructured as an API service layer to ingest structured material data from existing public marketplaces. Furthermore, the limited sample size necessitates future evaluations with larger, diverse participant groups to benchmark the hybrid system against other context-aware recommender systems and validate the findings across broader supply-and-demand scenarios.

Conclusions

This study validates that a multidimensional vector ranking system, augmented by dual-constraint logic, effectively supports the sourcing of reusable materials for architectural design. The preliminary findings indicate that while high operational speeds and weighted results facilitate discovery, the lack of transparency in algorithmic ranking and the ambiguity of weight-based toggles can obscure the system's underlying logic. Participants highlighted a "black box" friction where the rationale for specific material suggestions remained unclear, confirming that explainability is as critical as retrieval accuracy for professional adoption. To address these challenges, the research establishes that implementing similarity thresholds, clearer semantic labeling, and the progressive disclosure of filters is necessary to mitigate user frustration and build trust in the recommendation output. Future development will focus on API-driven scalability and the integration of broader marketplace datasets to solidify the platform's role as a meta-aggregator. By bridging the gap between technical material retrieval and architectural design rationale, this work provides a methodological foundation for optimizing circular material flows within the construction sector.

References

Adomavicius, G., Mobasher, B., Ricci, F., and Tuzhilin, A. (2011). Context-aware recommender systems. AI Magazine, 32(3):67–80.

Afzal, I., Yilmazel, B., and Kaleli, C. (2024). An approach for multi-context-aware multi-criteria recommender systems based on deep learning. IEEE Access, 12:99936–99948.

Blackburn, O., Ritala, P., Keränen, J., and Bocken, N. (2026). Circular economy platforms: A systematic review. Business Strategy and the Environment, 35(2):2810–2841.

DIN German Institute for Standardization (2024). Din spec 91481:2024-02 classification of recycled plastics of polyamides by data quality levels for use and (digital) trading.

Honic, M., Magalhães, P. M., and Bosch, P. V. D. (2024). From data templates to material passports and digital product passports. In A Circular Built Environment in the Digital Age, pages 79–94. Springer.

Huang, S., Yang, H., Yao, Y., Lin, X., and Tu, Y. (2024). Deep adaptive interest network: Personalized recommendation with context-aware learning. arXiv preprint arXiv:2401.00000.

ISO/TC 323 (2018). Circular economy: Standards and guidance.

Kulkarni, S. and Rodd, S. F. (2020). Context aware recommendation systems: A review of the state of the art techniques. Computer Science Review, 37:100255.

Lai, K.-H., Yang, Z.-R., Lai, P.-Y., Wang, C.-D., Guizani, M., and Chen, M. (2024). Knowledge-aware explainable reciprocal recommendation. In The Thirty-Eighth AAAI Conference on Artificial Intelligence.

Lee, J., Chen, F., Dua, S., Cer, D., Shanbhogue, M., Naim, I., Ábrego, G. H., Li, Z., Chen, K., Vera, H. S., Ren, X., Zhang, S., Salz, D., Boratko, M., Han, J., Chen, B., Huang, S., Rao, V., Suganthan, P., Han, F., Doumanoglou, A., Gupta, N., Moiseev, F., Yip, C., Jain, A., Baumgartner, S., Shahi, S., Gomez, F. P., Mariserla, S., Choi, M., Shah, P., Goenka, S., Chen, K., Xia, Y., Chen, K., Duddu, S. M. K., Chen, Y., Walker, T., Zhou, W., Ghiya, R., Gleicher, Z., Gill, K., Dong, Z., Seyedhosseini, M., Sung, Y., Hoffmann, R., and Duerig, T. (2025). Gemini embedding: Generalizable embeddings from gemini. Technical report, Google DeepMind.

Lewis, J. R. (2018). The system usability scale: Past, present, and future. International Journal of Human-Computer Interaction, 34(7):577–590.

Menny, T., Le Guirriec, S., and De Wolf, C. (2024). The butterfly matchmaking model for circular construction: Towards a digital matchmaking platform tailored to french policy. Sustainable Production and Consumption, 49:130–143.

Meta AI (2024). Llama 3.1 8B Instruct.

Podkalicka, A. and Fredriksson, M. (2023). Mediatised marketplaces: Platforms, places, and strategies for trading material goods in digital economies. Convergence, 29(5):1352–1368.

Protocol Labs (2026). IPFS - Powers the Distributed Web.

Quertamp, F., Tang, T. H. L., and Ly, V. K. (2025). Designing a digital marketplace for vietnam's plastic recycling sector: Needs assessment and strategic blueprint. Technical report, United Nations Development Programme in Viet Nam, Ha Noi.

Swiss Society of Engineers and Architects (2020). ebkp-h sn 506 511 elementbasierter baukostenplan hochbau.

Tintarev, N. and Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In Ricci, F., Rokach, L., and Shapira, B., editors, Recommender Systems Handbook, pages 353–382. Springer, 2 edition.

Tserng, H.-P., Chang, C.-C., and Tai, S.-H. (2025). The development of building materials recommendation system based on feature crosses. Advances in Civil Engineering, 1.

Van Capelleveen, G., Amrit, C., Zijm, H., Yazan, D. M., and Abdi, A. (2021). Toward building recommender systems for the circular economy: Exploring the perils of the european waste catalogue. Journal of Environmental Management, 277(111430).

Wilson, S., Adu-Duodu, K., Li, Y., Sham, R., Wang, Y., Solaiman, E., Perera, C., Ranjan, R., and Rana, O. (2025). Trustworthy second-hand marketplace for built environment. Technical report, Technical Report 111.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

TURNING CONSTRUCTION CHECKLISTS INTO MACHINE-READABLE INSPECTION SCHEMA

Till Weist, Leon Pfenning, and Prof. Sven Rogalski
Darmstadt University of Applied Sciences, Darmstadt, GER

Abstract

Construction verification currently relies on unstructured checklists, hindering integration with autonomous agents. To scalably convert these legacy instructions into machine-readable data, we propose a Hybrid Inspection Pipeline. This system uses LLMs and "Balanced Retrieval-Augmented Generation" to map text into an "Inspection Triad" aligned with IFC and AEC3PO ontologies. Achieving 92.5% semantic precision, the pipeline eliminates hallucinations using a deterministic fidelity loop. Ultimately, this framework provides the executable logic necessary for autonomous inspection agents, establishing a foundation for future construction research.

Introduction

The construction industry is characterized by a digital paradox, whereby buildings are designed in sophisticated three-dimensional (3D) environments yet are verified on-site using tools that have remained unchanged for decades (Nkosi, 2025). Contemporary digitization endeavors prioritize the capture of geometry (BIM) while the Quality Assurance (QA) process, responsible for verifying Built as Planned attributes, has exhibited a comparatively lagging adoption of digital methodologies (Toyin, 2025). Presently, the Quality Assurance (QA) sector utilizes a manual inspection process, wherein inspectors employ pen-and-paper checklists or static digital checklists to verify compliance (Sawhney & Knight, 2024). These analog methods are not only susceptible to human error but also face challenges in scaling up to meet the demands of a growing labor shortage and increasingly complex project requirements (Baiburin, 2017). As Naderi et al. (2025) point out, this reliance on human-centered instructions creates a significant gap, thereby preventing the integration of inspection tasks with software agents or autonomous robotic systems.

In order to address this discrepancy, the industry has developed standardized semantic frameworks, with the most notable examples being the Industry Foundation Classes (IFC) and the AEC3PO Ontology. While IFC provides the standardized geometry and spatial data of the building, AEC3PO extends this by providing the structural scaffold for regulatory management. It functions as a metadata container, organizing compliance requirements and linking them to legal documents, thereby creating a structured target for digital inspection (Dridi et al., 2024). However, a critical interoperability gap persists: While these target standards are well-defined, the source data—thousands of legacy inspection instructions—remains locked in unstructured text (Guo et al., 2021). Consequently, a scalable mechanism to migrate these analog checklists into a standardized digital format has yet to be developed. As a result, manual translation remains a labor-intensive and inconsistent bottleneck. As posited by Besta et al. (2025), this paper proposes a Hybrid Inspection Pipeline as a solution to the identified semantic interoperability gap. We propose a reinterpretation of the conversion of checklists from a generative interpretation task to a rigorous mapping problem. A workflow has been developed that aligns inspection texts with the IFC and AEC3PO ontologies using a Grammar-Constrained Extraction Pipeline. The utilization of Large Language Models (LLMs) in conjunction with stringent schema constraints enables the automation of the mapping of free text to standardized classes, exhibiting a high degree of reliability. This approach involves the conversion of unstructured field instructions into computable, object-oriented data.

Related Work

Early efforts in construction information extraction relied heavily on rule-based natural language processing (NLP) and syntactic pattern matching (Zhang & El-Gohary, 2016). While these methods have proven effective for highly structured regulatory documents, they have demonstrated fragility when applied to the heterogeneous natural language found in field checklists. In order to address this issue of rigidity, research has shifted towards the utilization of Deep Learning (DL) models. Approaches that utilize Long Short-Term Memory (LSTM) networks (Wang & El-Gohary, 2022) and Transformer-based architectures like BERT (Moon et al., 2022) have been shown to significantly improve entity recognition and classification tasks.

However, a critical limitation persists in these discriminative models: they excel at identifying entities (e.g., tagging a "wall"), but struggle to construct the complex, hierarchical object relationships required by strict ontologies like AEC3PO. As indicated in a recent systematic review (Sousa & Martins, 2024), while Deep Learning (DL) models have demonstrated high accuracy in document classification, their capacity to extract

detailed, actionable logic remains constrained without substantial task-specific training data.

Recently, (Chen et al., 2024) sought to address this discrepancy by integrating Large Language Models (LLMs) with Deep Learning for compliance verification. While their hybrid approach demonstrated enhanced reasoning capabilities, it primarily relied on few-shot prompting without strict output enforcement. This distinction is of critical importance: Even when optimized, probabilistic prompting cannot guarantee the syntactic validity required for downstream software agents. As emphasized by Fuchs et al. (2025) in their review of automated compliance systems, the dearth of "computable" regulatory data persists as a predominant impediment. Current State-of-the-Art (SOTA) methods frequently yield "likely" matches rather than the guaranteed valid schema instances necessary for safe robotic execution. Consequently, a gap remains for a method that treats text extraction not as probabilistic prediction, but as a constrained mapping problem.

The transition toward machine-readable regulatory frameworks has been driven by the "Linked Building Data" (LBD) movement. In the early stages of this research, the focus was on translating static building codes into computable logic. This was primarily achieved by using the ifcOWL ontology to validate Industry Foundation Classes (IFC) data (Pauwels et al., 2017; Beach et al., 2025). These approaches have been demonstrated to effectively automate the process of "Design Compliance" verification, which queries whether a digital model aligns with the stipulated code requirements (e.g., "Is the hallway > 1.2m wide?").

Recently, the AEC3PO ontology (Architecture, Engineering, and Construction Compliance Checking and Permitting Ontology) has emerged as a comprehensive standard for harmonizing the structural metadata of regulatory knowledge (Dridi et al., 2024). In contrast to rigid code-checking schemas, AEC3PO provides a flexible, top-level schema for defining compliance checks, offering classes for CheckStatements, verification methods, and evidence requirements. Nevertheless, a fundamental gap persists in the application of these ontologies. While AEC3PO provides a standardized "container" for regulatory information, its function is primarily descriptive rather than prescriptive. It delineates the relationship between a legal clause and a data requirement; however, it lacks a standardized library of executable properties necessary for automated grounding. Consequently, extant research remains predominantly oriented toward Design Verification (checking the BIM model against the Code), rather than Execution Verification (checking the physical construction against the Instruction). Recent works, such as the Ontology for Construction Quality Assurance (OCQA), have modeled inspection planning (Seiß et al., 2025). However, these works do not address the semantic interpretation of the inspection content itself. The instructions themselves, such as "Verify that all bolts are torqued to 50Nm," are expressed in unstructured natural language. Presently, there is an absence of an automated mechanism to bridge this semantic gap, which involves the mapping of raw text to specific Object-Property-Value triplets that can populate the AEC3PO container.

Methodology I: The Schema

In order to address the semantic discrepancy between unstructured inspection instructions and machine-readable frameworks, the Inspection Triad has been proposed as a solution. This framework functions as the intermediate formal representation, thereby isolating the components of a compliance check that are indispensable to the process.

Multi-Target Instructions

A fundamental barrier to digitization is the compact style of legacy checklists. From a pragmatic standpoint, human-authored items are frequently activity-centered, consolidating multiple checks around a singular verb to reduce reading time. Consider the following example from our dataset:

"Check for proper installation of Mirror, Paper Holder, Towel Rack, Shower Curtain Rod."

While efficient for a human, this structure poses a "many-to-one" problem for autonomous systems. The single instruction encompasses four distinct inspection targets. It is imperative to recognize that the control flow of an agent or digital twin is inherently bound to specific spatial coordinates. Consequently, the execution of "bundled" tasks is contingent upon their disaggregation. Our analysis of 200 industry inspection tasks (Mahoney, 2008) confirms the prevalence of this issue. We found that 61% of tasks were compound instructions, bundling multiple requirements into a single sentence. Consequently, we define Atomicity as a prerequisite for processing: a valid inspection task must contain exactly one activity mapped to exactly one target object.

Syntactic Pattern

Despite the structural messiness of Activity-Centered tasks, our analysis reveals a consistent underlying linguistic structure across the dataset. When decomposed, valid inspection instructions invariably follow a specific syntactic pattern:

$$\text{Instruction} \approx [\text{Verb}] + [\text{Object}] + ([\text{Modifiers}]/[\text{Conditions}])$$

Every executable instruction contains an action (Verb) and a reference to an entity (Object). The specificity of the instruction relies on the Modifiers (quantities, temporal cues, or reference standards). However, our analysis highlights gaps in how these slots are filled. 7.5% of tasks lacked a clear object reference (e.g., using generic terms like "units"), and 4% relied on ambiguous modifiers (e.g., "proper," "adequate") rather than measurable criteria. These gaps highlight that inspection instructions are rarely self-contained algorithms; rather, they serve as shorthand prompts for experienced professionals. This profound lack of explicit logic confirms that raw field data is unstructured.

Inspection Triad

We address the structural complexity and inherent ambiguity of legacy checklists by mapping raw linguistic patterns onto the Inspection Triad (T), a formal data structure designed for autonomous execution.

$$T=(\tau,\alpha,\kappa)$$

This structure formalizes the raw text into three necessary and sufficient components:

The Target (τ) ← [Object]: This component anchors the instruction to a single, specific building element. If the source text contains multiple objects the task must be split so that each result in a distinct Triad. Further there is a resolution constraint: if an inspection target is not identified and is ambiguous, the Triad cannot be completed. Logically, an inspection cannot be executed without a confirmed physical anchor.

The Action (α) ← [Verb]: It maps the instructional verb to a specific inspection procedure.

The Criterion (κ) ← [Modifiers]: It captures the quantitative limits or conditions. Where a criterion is not explicitly defined, the schema defaults to a binary Pass/Fail requirement anchored to the verification method of the Action (α).

While the triad represents the irreducible atomic unit of verification, the framework supports higher-order logic. Complex regulatory requirements, such as conditional "IF-THEN" sequences, are facilitated by nesting multiple triads within logical operational wrappers. This ensures that even intricate compliance logic maintains data integrity at the component level.

Ontological Integration

The Triad serves as the bridge between raw site data and standardized digital frameworks. In this architecture, the AEC3PO ontology provides the structural scaffold (the semantic "containers"), while the Industry Foundation Classes (IFC) provide the physical grounding.

The Target (τ) is operationalized as a semantic pointer to the Digital Twin. By utilizing the IFC standard to identify specific entities (e.g., IfcWall, IfcDoor), the system ensures that every instruction is spatially addressable within the Building Information Model (BIM). This "pointer-based" approach ensures the inspection data remains lightweight, referencing existing geometry rather than duplicating it.

Simultaneously, the Action (α) and Criterion (κ) are instantiated through the Domain-Specific Extension (DSE), which provides the prescriptive definitions required for automated execution.

Domain-Specific Extension (DSE)

While the AEC3PO Ontology provides the necessary semantic containers (the "Scaffold"), our initial gap analysis revealed it lacks the prescriptive definitions required for automated execution. Standard classes like aec3po:CheckStatement describe that a check exists, but not how to measure specific properties like "Flatness" or "Torque."

To bridge this semantic gap, we developed the DSE as a controlled vocabulary of 50 distinct inspection properties, formally extending the high-level AEC3PO classes into executable leaf nodes (e.g., aec3poext:VerticalityCheck).

The development of the DSE followed a rigorous Ontology Learning methodology. We compiled a corpus of 1,500 industry inspection tasks (Mahoney, 2008). To identify the salient inspection criteria, we applied Term Frequency-Inverse Document Frequency (TF-IDF) extraction. This statistical method allowed us to filter out generic stop words and isolate the domain-specific nouns and modifiers with the highest information density (e.g., "width," "spacing," "cracks"). These candidate terms were then formalized into rigid definitions through expert review, ensuring the DSE is not a theoretical construct, but a data-driven representation of actual field requirements.

Methodology II: The Pipeline

While the Inspection Triad defines the target semantic structure, the challenge remains to reliably populate this structure from unstructured legacy documents. As established in the literature review, Generative Large Language Models (LLMs) often struggle with strict schema adherence and factual consistency when operating in unconstrained environments (Ji, et al., 2023). To mitigate these stochastic risks, we propose the Hybrid Inspection Pipeline.

This architecture shifts the extraction task from a purely generative paradigm to a "Propose and Verify" workflow. As illustrated in Figure 1, the system is composed of three distinct functional layers: (A) Semantic Ingestion and Atomization, (B) The Extraction Engine, and (C) Data Formatting and Storage. The result of this process is a serialized list of inspection requirements, each containing a unique name and technical description.

Layer A: Semantic Ingestion and Atomization

The input layer serves as the normalization pre-processor. Legacy inspection data typically exists in "frozen" formats (PDFs, scanned images) where semantic hierarchy is conveyed through visual layout rather than machine-readable tags. Layer A utilizes Optical Character Recognition (OCR) to digitize these documents. A critical preprocessing step is Atomization. As identified in our analysis of the dataset, legacy instructions are compound, bundling multiple targets into a single sentence. To satisfy the "One-Target-One-Action" requirement of the Inspection Triad, we employ a Grammatical Parser combined with lightweight LLM-based validation.

This component identifies the grammatical ROOT (Action), direct object (Target), and adjectival modifiers (Criterion) using linguistic dependency rules. By applying deterministic heuristic rules, the system ensures that every extracted object is bound to a specific action. This logic allows for the decomposition of compound instructions where a single action governs multiple physical targets. If the parser identifies a verb without a corresponding object, the instruction is immediately flagged as "Incomplete" to prevent downstream processing errors.

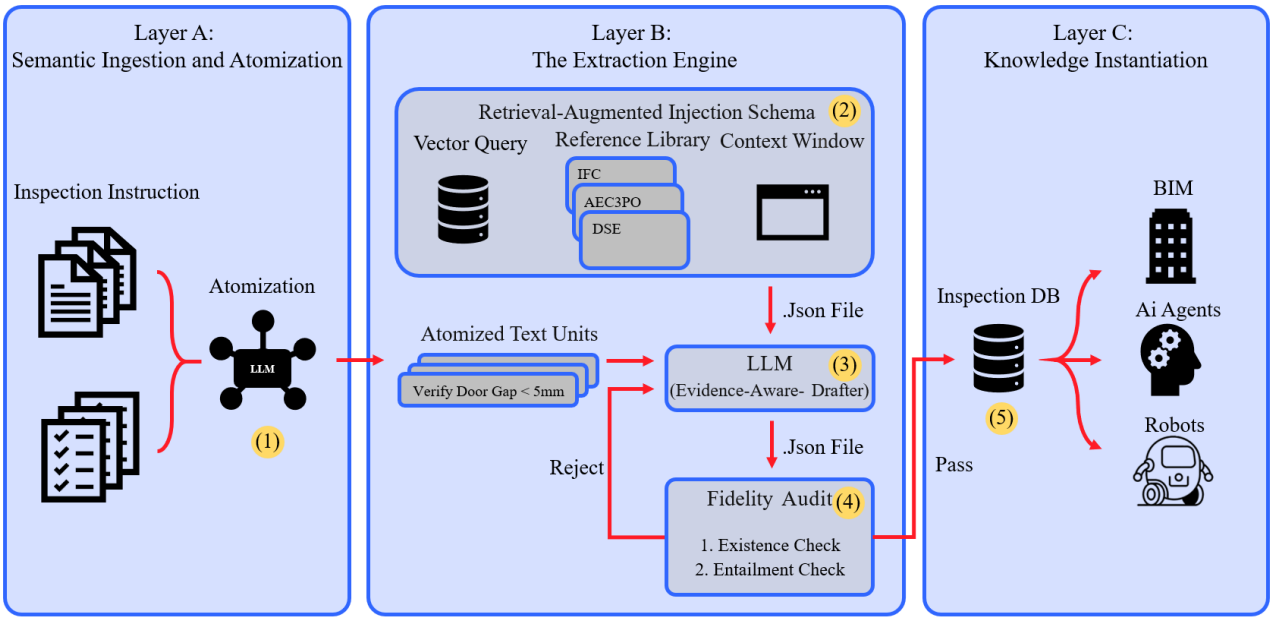


Figure 1: Architecture of the Proposed Framework

This linguistic skeleton, along with the original text, is then passed to the LLM for a final verification and refinement pass. If an incomplete action is detected, the LLM is also tasked with resolving the missing context. As shown in the data flow of Figure 1, this ensures the module decomposes compound sentences into discrete, atomic text units (1). For example, the input “Verify gap for door and window” is split into two independent units: “Verify gap for door” and “Verify gap for window.” This ensures that downstream extraction (Layer B) operates on unambiguous, single-subject inputs.

Layer B: The Extraction Engine

Layer B represents the core algorithmic contribution of this framework, acting as the semantic bridge between atomized natural language and formal ontologies. To resolve the interoperability gap, this layer implements a Retrieval-Augmented Schema Injection (RASI) strategy combined with a deterministic Fidelity Loop. The primary barrier to utilizing extensive ontologies like IFC and AEC3PO within a Large Language Model (LLM) is the risk of "Attention Dilution" and hallucination. To mitigate these risks, the system maintains a Dual-Source Reference Library consisting of two distinct vector indices: physical object definitions (IFC) and prescriptive inspection logic (DSE). During runtime, the engine performs a Balanced Retrieval (2) to counteract "Object Drowning"—a phenomenon where high-frequency physical terms (e.g., "Wall" or "Slab") statistically overshadow abstract inspection rules in the vector space, leading to incomplete data extraction. By enforcing a fixed-ratio retrieval quota (e.g., k/2 from each index), the system independently queries both the Target and Logic indices to pull an equal number of IFC classes and DSE properties. These results are interleaved and dynamically injected into the LLM’s context window, transforming the extraction from an open-ended generative task into a constrained classification task that forces the model to select from a verified ontology subset rather than fabricating non-standard properties.

Semantic ambiguity in the source text, specifically the use of vague modifiers like "properly" or "adequate," is addressed through Semantic Best-Fit Mapping. By providing the LLM with a filtered subset of technical definitions from the DSE, the workflow acts as a semantic anchor; the model performs a contextual interpolation to map subjective human language to the most semantically relevant engineering property provided in the retrieved context. In instances where no explicit metric exists, the system defaults to the most frequent domain-specific interpretation within the DSE. Even with these schema constraints, generative models may hallucinate quantitative values. To eliminate this risk, Layer B enforces a two-stage Fidelity Audit. In the first stage, the LLM is tasked with Evidence-Aware Drafting (3), requiring it to return a Citation Tuple (Value,Substring) containing the exact verbatim evidence from the source text. In the second stage, a deterministic verification module performs an Existence Check to ensure the substring exists in the original document, and a Logic Entailment Check to verify the extracted value is supported by that quote. If the audit fails, a Rejection Feedback Loop (4) triggers a re-evaluation or flags the item as "Unprocessable," ensuring that only grounded, verifiable logic is serialized into the final digital repository.

Layer C: Knowledge Instantiation

Validated Triads from Layer B are serialized into a Digital Inspection Repository (5), which functions as an active API Gateway. Instead of static files, this system aggregates atomic checks into location-based sets (e.g., "Sector A Requirements"), creating a queryable endpoint accessible to downstream consumers such as BIM Applications, autonomous robots, or AI compliance engines.

Implementation

To validate the Hybrid Inspection Pipeline, a functional prototype was developed using a modular Python 3.9 architecture. A primary design driver was component decoupling; the extraction logic is intentionally isolated from the underlying model providers to ensure the architecture remains agnostic. While the current prototype utilizes specific Google and open-source tools, the framework is designed to swap backend LLMs or vector stores as the technology landscape evolves.

System Architecture & Tech Stack

The core Extraction Engine was implemented using the google-genai SDK, specifically utilizing the Gemini 2.5 Flash model. This model was selected for its optimization for high-throughput, low-latency tasks. To ensure scientific reproducibility and deterministic extraction, the model was configured with a Temperature of 0.0.

The system utilizes a Structured Prompting strategy where the LLM is provided with a strict JSON schema. If the Fidelity Loop identifies a discrepancy, the system initiates a targeted follow-up query, providing the model with specific error feedback to refine the extraction in a "human-in-the-loop" style automated correction.

Linguistic Atomization

For the initial decomposition of compound instructions, the prototype utilizes the spaCy NLP library with the en_core_web_lg (large) model. This accuracy model was chosen to ensure precise identification of complex instructions. The implementation specifically leverages spaCy's Dependency Parser to programmatically identify:

The ROOT (the primary action verb or Action α).

The dobj (direct object or Target τ).

The amod/compound modifiers (forming the basis for Criterion κ).

This "Linguistic Skeleton" is then verified by a fast LLM pass to ensure atomicity before any ontological grounding occurs, preventing downstream processing of multi-subject tasks.

The Semantic Knowledge Base (RAG)

The source ontologies were embedded using the SentenceTransformers framework with the all-MiniLM-L6-v2 model, producing 384-dimensional vectors.

We compiled two distinct vector stores—the Target Index (IFC4) and the Logic Index (DSE). Unlike standard keyword indices, these stores contain also metadata: the IFC Index includes Entity Definitions, while the DSE entries include Data Types (e.g., Boolean, Float) and Standardized Verification Methods.

These indices are persisted locally as .json vector arrays. This ensures full data sovereignty and vendor independence.

The Balanced Retrieval Strategy

Early testing revealed a phenomenon we term the "Drowning Effect". Standard vector retrieval tends to favor concrete nouns over abstract concepts. A query like "Check the flatness of the slab" would return 20 results related to "Slab" (Object) and zero results related to "Flatness" (Rule), causing the LLM to hallucinate the property check.

To counteract this, we engineered a Balanced Retrieval Strategy. The retriever does not simply fetch the top-k results. Instead, it enforces a quota: It independently queries the Target Index and the Logic Index, retrieving the top-k/2 matches from each. These distinct result sets are then merged into the unified context window. By decoupling the retrieval of targets and properties, the framework maintains semantic parity within the prompt. This prevents the stochastic exclusion of abstract rules and guarantees that the model's context window reflects the dual-component nature of the Inspection Triad.

Schema Validation & Error Handling

To bridge the gap between probabilistic AI and the deterministic requirements of engineering data, we integrated Pydantic v2.0. We defined the InspectionTriad as a strict Python class, forcing the LLM to act as a structured data parser. The pipeline automatically validates every generated response against this class. If a ValidationError is triggered (e.g., malformed JSON or missing mandatory fields), the system initiates an automated retry loop (up to 3 attempts). This creates a "Type-Safe" boundary that prevents malformed data from propagating to the digital repository.

The Fidelity Loop

The Fidelity Loop was implemented as a deterministic Python module to eliminate silent hallucinations. It acts as a non-probabilistic auditor by performing two sequential string-matching checks on the LLM output:

Stage 1: Existence Check. The module verifies that the evidence_quote provided by the LLM exists verbatim in the source document. Any character mismatch triggers an immediate rejection.

Stage 2: Entailment Check. The module ensures that extracted quantitative values (e.g., "50Nm") or property types (e.g., "Torque") are present within the cited evidence_quote. This prevents the model from fabricating values unsupported by the evidence.

This "Trust but Verify" logic transforms the system from a generative writer into a deterministic extraction verifier. It provides the rigorous safety layer necessary for high-stakes construction compliance.

Evaluation

To validate the efficacy of the proposed Hybrid Inspection Pipeline, we conducted a systematic evaluation focused on two key performance indicators: Semantic Precision (the ability to correctly map text to the Inspection Triad) and Operational Safety (the ability to detect and reject hallucinations).

Experimental Setup

Due to the absence of a standardized, publicly available benchmark for construction inspection extraction, a novel evaluation dataset was curated for this study. The dataset comprises 200 distinct inspection instructions synthesized from a combination of real-world inspection checklists and technical field reports (cf. Mahoney, 2008; gsa.gov, 2025; carlislesyntec.com, 2021).

To ensure the evaluation was representative of real-world site conditions and to mitigate selection bias, the instructions were purposively stratified into three complexity levels. This distribution guarantees that the dataset reflects the linguistic heterogeneity of the industry rather than a simplified subset of easily parsed data.

Geometric Checks: Explicit quantitative constraints (e.g., "Verify window width is < 1200mm").

Semantic Classifications: Checks relying on specific object enumerations and subclass distinctions (e.g., "Ensure T-Beam is free of defects" vs. "Ensure Beam is...").

Ambiguous/Qualitative: Instructions lacking clear metrics or relying on subjective professional judgment (e.g., "Check for proper installation").

For each input, a human domain expert manually annotated the "Expected Inspection Triad," defining the Target IFC Class and the specific prescriptive property from the Domain-Specific Extension (DSE).

To ensure a fair comparative analysis, all experimental runs utilized the Gemini 2.5 Flash model. The temperature was set to 0.0 across all tests to ensure deterministic outputs and minimize stochastic variance, thereby isolating the impact of the retrieval architecture on the results.

While traditional Rule-Based NLP (Natural Language Processing) was initially considered, it was ultimately excluded from the final benchmarking. Legacy NLP systems are fundamentally limited by the "brittleness" of manually curated rules; they are prone to overfitting to specific vocabulary and fail to account for the high linguistic variance of site reports. Consequently, an NLP baseline would represent a "straw man" comparison rather than a contemporary performance standard. Instead, the study benchmarks the Hybrid Pipeline against two state-of-the-art LLM strategies:

Zero-Shot (Blind): The model attempts to map instructions to IFC and DSE schemas based purely on internal training data, without external library access.

Zero-Shot (Full Context): The entire IFC and DSE libraries are injected into the prompt (approx. 8,500 tokens), testing performance under maximum "Context Overload."

Hybrid Pipeline (Balanced RAG): The proposed method utilizes the Balanced Semantic Retriever to inject only the 5 most relevant IFC classes and 5 most relevant DSE properties into the prompt.

Performance Metrics

System performance was evaluated across two primary dimensions: Extraction Quality and Operational Safety. Information Retrieval performance for the Triad (Target, Action, Criterion) was quantified using the following metrics:

Precision (P): The ratio of correctly identified Triad components to the total components extracted.

Recall (R): The ratio of successfully mapped instructions to the total instructions in the dataset.

F1-Score: The harmonic mean of Precision and Recall, providing a balanced measure of system robustness:

For the safety analysis, we introduced the Hallucination Catch Rate, defined as the percentage of fabricated values (hallucinations) successfully identified and blocked by the Fidelity Loop before data entry.

Results: Semantic Precision

Quantitative results indicate that the Hybrid Pipeline significantly outperforms baseline methods in both technical accuracy and computational efficiency.

Table 1: Performance Comparison Across 200 Inspection Instructions

Method	Precision	Recall	F1-Score
Zero-Shot (Blind)	34.5%	38%	0.36
Zero-Shot (Full Context)	76%	82%	0.79
Hybrid (Balanced RAG)	92.5%	91%	0.91

The "Full Context" approach suffered from "Attention Dilution," where the model's reasoning capabilities were degraded by irrelevant ontological data, often leading to the selection of phonetically similar but semantically incorrect IFC classes. Conversely, the Hybrid Pipeline's retriever isolated the most relevant candidates, maintaining high precision. Furthermore, the Hybrid Pipeline achieved a 13x reduction in token usage compared to the Full Context method, demonstrating superior scalability for enterprise-level BIM integration.

Results: The Fidelity Loop

A primary barrier to AI adoption in AEC is the risk of "hallucinated" data. The Zero-Shot baselines exhibited a combined hallucination rate of 11.4%, where the LLM generated plausible but non-existent quantitative values (e.g., inventing a "5mm" tolerance) based on general patterns.

The Fidelity Loop achieved a 99.5% Hallucination Catch Rate. By enforcing a strict "Verbatim Evidence" requirement, the system effectively blocked fabricated data. Residual failures typically involved "partial hallucinations," where the model correctly identified a numeric value but slightly altered the unit or context (e.g., misinterpreting "5 meters" as "5mm" during the mapping phase).

Error Analysis

A qualitative review of the 7.5% error margin in the Hybrid Pipeline identified "Retriever Ambiguity" as the primary failure mode. For example, the instruction "Check the seal on the pipe" was correctly mapped to the Action SealantVerification, but the retriever proposed the Target IfcSurfaceFeature instead of IfcPipeSegment due to the close proximity of these terms in the high-dimensional vector space. Future iterations will explore the integration of cross-encoders to further refine the retrieval ranking and reduce semantic noise.

Discussion

The evaluation confirms that a constrained LLM pipeline can function as a reliable semantic bridge. However, moving this framework from a controlled prototype to a production-scale construction environment requires addressing several key operational limitations.

Automated Scalability of the Knowledge Base

The current system's reliance on a manually curated Domain-Specific Extension (DSE) remains a primary scalability bottleneck. To move beyond manual labor, future research should explore a "Loop-Back Ontology Learning" mechanism. In this framework, instructions that fail the Fidelity Audit due to missing DSE definitions are not discarded. Instead, they are sent to a "Candidate Pool" where an LLM clusters frequent failures to suggest new ontological leaf-nodes. This allows the system to evolve dynamically, expanding its technical vocabulary through a semi-automated, human-in-the-loop process.

Orchestration of Complex Regulatory Logic

While the Inspection Triad ensures high precision for atomic tasks, it currently lacks the native syntax to represent complex conditional dependencies (e.g., multi-step "if-then" requirements). To support the rich logical expressions found in safety regulations, the framework must transition from "Atoms" to "Molecules." By introducing a "Logical Orchestration Layer," multiple atomic Triads can be nested using standard Boolean operators (AND, OR, IF). This adaptation would allow the pipeline to maintain its current extraction reliability while supporting the sophisticated reasoning required for advanced compliance checking.

Transition to Active 4D BIM Integration

Currently, the pipeline functions as a one-way extraction tool rather than an active participant in the project lifecycle. To bridge the gap between static logic and physical execution, an "End-to-End Handshake" with live project data is necessary. Future work will investigate a direct integration with the project schedule (4D BIM), where the completion of a construction task (e.g., "Slab Pour") automatically triggers the extraction of relevant Triads. By querying the BIM database for specific GlobalIDs in real-time, the system can provide executable tasking for autonomous agents or mobile robots based on the actual progress of the site.

Resolution of Temporal and Qualitative Edge Cases

The system encountered specific failures when handling temporal constraints (e.g., "Verify 28 days after pour") and vague qualitative modifiers (e.g., "Adequate"). Future iterations will require a dedicated temporal logic parser to differentiate between execution timing and performance duration. Additionally, for subjective instructions, the pipeline could be adapted to output a "Confidence Score," flagging ambiguous text for human clarification rather than forcing a binary mapping. This ensures the machine-readable output remains a "Trust-but-Verify" source of truth for downstream automated systems

Conclusion

This research demonstrates that the transition from unstructured legacy checklists to machine-readable inspection schemas is achievable through a constrained, hybrid pipeline. By shifting the extraction task from open-ended generation to ontological mapping via the Inspection Triad, we provide a method for populating the AEC3PO and IFC frameworks with executable logic.

The significance of this work lies in its ability to transform static regulatory metadata into "computable" field instructions. While existing BIM-to-Field workflows prioritize geometry, the proposed framework establishes the semantic foundation necessary for autonomous agents to perform site inspections with deterministic reliability. Future development in automated ontology learning and 4D BIM integration will be essential to scale this system across the diverse and evolving landscape of global construction standards.

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research (BMBF) [Grant No. 01MK24004D] as part of the project BIM-4-CARE. The responsibility for the content of this publication lies with the authors. Darmstadt University of Applied Sciences is a member of the European University of Technology (EUt+) alliance.

References

Baiburin, A., 2017. Errors, Defects and Safety Control at Construction Stage. *Procedia Engineering,* pp. 807-813.

Beach, T., Rezgui, Y., Li, H. & Kasim, T., 2025. A rule-based semantic approach for automated regulatory compliance in the construction sector. *Expert Systems with Applications.*

Besta, M. et al., 2025. CheckEmbed: Effective Verification of LLM Solutions to Open-Ended Tasks.

carlislesyntec.com, 2021. *EPDM Rooftop Guide/Inspection Checklist.* [Online] Available at: https://www.carlislesyntec.com/en/Document-Viewer/epdm-rooftop-guideinspection-checklist/Nh-nHipBP0qjYUutMmHf5Q [Accessed: 11 01 2026].

Chen, N., Lin, X., Jiang, H. & An, Y., 2024. Automated Building Information Modeling Compliance Check

through a Large Language Model Combined with Deep Learning and Ontology. *Intelligent Monitoring and Detecting Methodologies for Building Structures.*

Dridi, A. et al., 2024. Aec3po: A Knowledge Model for Machine Executable Construction Regulations.

Fuchs, S., Fauth, J., Boden, M. & Amor, R., 2025. The Challenge of Automated Compliance Checking: A Regulatory View. *2025 European Conference on Computing in Construction.*

gsa.gov, 2025. *2025 Interim Core Building Standards (CBS).* [Online] Available at: https://www.gsa.gov/system/files/CBS%20Submittal%20Matrix%202025%20with%20hyperlinks%20%281%29_0.pdf [Accessed: 11 01 2026].

Guo, D., Onstein, E. & Rosa, A. D. L., 2021. A Semantic Approach for Automated Rule Compliance Checking in Construction Industry.. *IEEE Access,* Band 9, pp. 129648-129660.

Ji, Z., Lee, N., Frieske, R. & Yu, T., 2023. Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys,* 55(12), pp. 1 - 38.

Mahoney, W. D., 2008. *Construction Inspection Manual.* 8. Hrsg. Los Angeles: BNI Publications.

Moon, S., Chi, S. & Im, S.-B., 2022. Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT). *Automation in Construction.*

Naderi, H. et al., 2025. Autonomous Construction-Site Safety Inspection Using Mobile Robots: A Multilayer VLM-LLM Pipeline.

Nkosi, T., 2025. Zero-Paper Construction Sites: Leveraging BIM-To-Field Data, Sqlite Sync, And Real-Time Quality Assurance for Smarter Projects.. *International Research Journal of Advanced Engineering and Technology,* 2(11), pp. 1-7.

Pauwels, P., Zhang, S. & Lee, Y.-C., 2017. Semantic web technologies in AEC industry: A literature overview. *Automation in Construction.*

Sawhney, A. & Knight, A., 2024. Digitalisation in constructionreport 2024. *Royal Institution of Chartered Surveyors (R.*

Seiß, S., Zheng, Y. & Melzner, J., 2025. Ontology-based representation of quality assurance and inspection planning in construction. *Automation in Construction,* Band 177.

Sousa, L. J. d. & Martins, J. P., 2024. Automation of text document classification in the budgeting phase of the Construction process: a Systematic Literature Review. *Construction Innovation: Information Process Management.*

Toyin, J., 2025. Innovative Approaches to QA/QC in Construction: Current State. *CIB Conferences.*

Wang, X. & El-Gohary, N., 2022. Deep Learning-Based Relation Extraction from Construction Safety Regulations for Automated Field Compliance Checking. *Construction Research Congress.*

Zhang, J. & El-Gohary, N. M., 2016. Semantic NLP-Based Information Extraction from Construction Regulatory Documents for Automated Compliance Checking. *Journal of Computing in Civil Engineering,* 30(2).

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

VISION–LANGUAGE MODELS IN CONSTRUCTION: ASSESSING PROMPT ROBUSTNESS IN OPEN-VOCABULARY OBJECT DETECTION

Federica Madaschi[1], Gelare Taherian[2], Marco Lorenzo Agostino Trani[1] and Ehsan Rezazadeh Azar[2]
[1]Department of Architecture Built environment and Construction Engineering,
Polytechnic University of Milan, Milan, Italy
[2]Department of Architectural Science, Toronto Metropolitan University, Toronto, Canada

Abstract

Vision–language models provide an alternative to supervised computer vision, enabling open-set object recognition without retraining or large datasets. Recent advances like GroundingDINO enable open-vocabulary object detection using natural-language prompts, but their sensitivity to prompt formulation in the AEC domain remains underexplored. This study assesses GroundingDINO's reliability in detecting construction objects through analysis of prompt variations, including lexical variation and prompt multiplicity. A custom dataset of 48 objects was tested under eight prompts. Results show lexical variation affects detection performance, with "regular base term" achieving a 56.1% F1 score. Multiple prompting lowers performance by ~2%, highlighting AEC's effectiveness and limits.

Introduction

Computer vision has been increasingly adopted in the Architecture, Engineering, and Construction (AEC) domain to support operational and managerial tasks such as progress monitoring, safety inspections, and quality control. However, deployment faces challenges due to scene changes, object diversity, occlusions, and project-specific visuals. Conventional supervised object detection needs large, labelled datasets and a fixed set of objects, limiting scalability and adaptability across projects and phases. (Duan et al., 2022). Vision–language models (VLMs) provide the foundation for open-vocabulary object detection (OVOD), enabling detection driven by natural-language prompts rather than predefined class sets, and have shown promise in overcoming closed-set limitations in AEC applications (Adil et al., 2025).

Vision–language models (VLMs) support open-vocabulary object detection (OVOD), which uses natural-language prompts instead of fixed labels. This offers an alternative to closed-set detection in AEC (Adil et al., 2025). OVOD models such as *GroundingDINO* (Liu et al., 2023) accept textual queries and provide spatially grounded detections, allowing flexible scene querying without retraining or extensive annotation. However, the quality and relevance of the detections generated by VLMs depend heavily on the prompt's quality and formulation (Minderer et al., 2024; Nebbia and Kovashka, 2024). Prior work shows prompt design greatly influences vision-language model performance and outcomes, underscoring their sensitivity to wording and structure (Gu et al., 2023). Since prompts are key to customizing VLM behavior, their design is crucial to aligning outputs with specific AEC needs.

In safety- and management-critical workflows, sensitivity to prompt formulation may lead to missed detections or false positives, undermining trust in automated systems. While recent studies have demonstrated the feasibility of VLM-based methods for construction-related tasks, the role of textual prompt formulation has not yet been systematically examined in AEC contexts, where scene complexity and operational requirements pose additional challenges. Moreover, prompt-related reliability issues, such as inconsistent detections or hallucinated objects, have not been explicitly analyzed in complex construction scenes. To address this gap, this paper presents a structured evaluation of prompt robustness for open-vocabulary object detection in construction settings. Focusing on *GroundingDINO*, the study investigates two key factors: (1) lexical variation, examining how different linguistic representations of the same construction object affect detection performance, and (2) prompt multiplicity, assessing the impact of single versus multiple prompts on accuracy and false positives. Using a custom dataset of construction-related objects spanning indoor and outdoor scenes, the study aims to benchmark the practical effectiveness and limitations of prompt-driven OVOD for AEC operational and managerial applications.

Related Works

Recent literature on OVOD methods has drawn mainly on pre-trained vision-language models, such as CLIP (Radford et al., 2021), which enable a detector to be queried with natural-language inputs, thereby overcoming the limitations of closed-vocabulary paradigms. In this context, numerous works have adopted hand-crafted textual prompts, in which the class name is inserted into simple linguistic templates (e.g., "a photo of a {class}"), sometimes combining multiple templates and averaging their embeddings to reduce dependence on a single formulation (Minderer *et al.*, 2024). These approaches have demonstrated the effectiveness of textual prompting for querying OVOD models, but they implicitly assume a stable correspondence between the canonical class name and the visual concept. A subsequent line of research challenged this assumption by examining lexical variation in prompts, including the use of synonyms and

semantically related terms. Recent studies show that replacing the base term with synonyms or near-synonyms can significantly degrade performance, highlighting the limited semantic robustness of OVOD models relative to equivalent linguistic variations (Nebbia and Kovashka, 2024). To mitigate this sensitivity, class-embedding enrichment strategies have been proposed by aggregating multiple prompts that include both the base term and its synonyms, suggesting that prompt design is a structural component of the detection process (Gu *et al.*, 2022). At the same time, other works have extended the use of prompts beyond isolated class names, introducing descriptive and attributive nominal phrases, as in the Open-Vocabulary Attribute Detection task, where textual representations consist of name-adjective combinations or more articulated descriptions (Bravo *et al.*, 2023). Even when such formulations are not explicitly defined as "definitions", they implicitly encode information about the function, material, and context of the object, delineating a continuum from the bare class name to increasingly rich and semantically specific prompts. However, most existing studies focus on the impact of these strategies on average accuracy, without systematically analyzing how the prompt's linguistic complexity affects model stability or the risk of generating false predictions. This is the context for *GroundingDINO*, an advanced OVOD model that accepts free-form natural language sentences and returns spatially anchored bounding boxes aligned with textual queries (Liu *et al.*, 2023). *GroundingDINO* is queried using single object names, short sentences with attributes, or multiple prompts separated by punctuation. However, the literature analyzing its performance tends to keep the textual prompt fixed or generic, focusing primarily on variations in the scene or visual content, as in studies that evaluate the model's robustness to visual perturbations while keeping The present work analyses *Groundingdino*'s prompt formulation, especially in construction, building on existing experimental methods and focusing on lexical variation and multiplicity (Mütze et al., 2025).

Methodology

The methodology systematically generates prompts to evaluate the OVOD model's performance across different texts, as shown in Figure 1. After creating a custom AEC dataset, prompt formulations were organized along two dimensions. The first-dimension concerns prompt complexity, distinguishing between easy and complex prompts. Easy prompts are characterized by minimal linguistic manipulation of the target object and involve simple lexical variations, such as a base term or a synonym. In contrast, complex prompts introduce additional levels of linguistic transformation as descriptive definitions, thereby increasing interpretive complexity while still referring to the same visual target. This dimension was designed to assess the model's sensitivity to wording and semantic expressiveness through controlled lexical variations for each object class. The second-dimension addresses prompt multiplicity, differentiating between single- and multiple-prompt configurations. Single Prompt settings rely on a single lexical call for a single object present in the analyzed image, enabling the evaluation of the model's response to isolated linguistic variations. Based on the prompt type identified in the first stage, a Multiple Prompt configuration was subsequently defined, allowing comparison of OVOD performance when detecting AEC-related objects under single- and multi-prompt conditions.

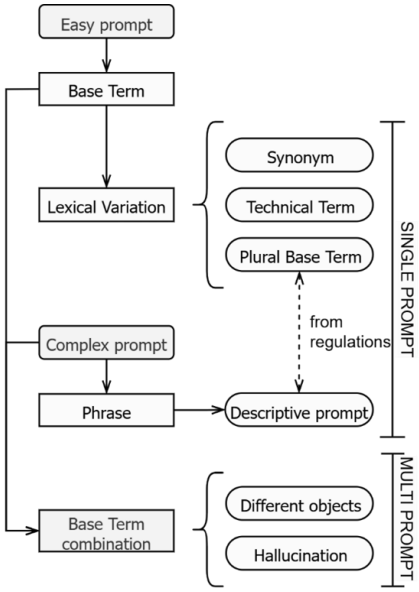


Figure 1: Overview of the prompt formulation strategy

Dataset development

A tailored image dataset was created to facilitate model assessment across various object categories and varying levels of visual complexity. Specifically, 48 construction-related objects were included and mainly organized into indoor and outdoor scenario groups. For each object, three images were selected to represent a range of realistic conditions, resulting in a custom dataset of 144 images. These depict variations in visual appearance: a normal view, where the object is the primary subject; a small-scale view, where the object appears reduced relative to the surrounding context (Lin *et al.*, 2014); and a cluttered view, in which other elements in the scene partially occlude the object (Malaikrisanachalee et al., 2025). A structured naming convention was adopted to organize the dataset. To each object a Base Term was assigned, intended as the canonical object name used for Ground Truth labelling, while a suffix indicated the visual condition (Normal, Small, Cluttered). The Base Terms were formulated using natural, everyday language, following the labelling strategy introduced by Taherian et al. (2025).

Lexical variations definition

In the single-prompt setting, lexical variation was employed to systematically evaluate the model's sensitivity to different textual references for the same visual. Each prompt includes a single linguistic

Table 1: Example of Lexical prompt strategies used in the single-prompt evaluation

Prompt Strategy	Indoor	Outdoor
Base Term [Baseline]	Split System	Construction Barrier
Synonym Term	Air Conditioner	Barricade
Technical Term	Ducted Air-Conditioner	Impact Barrier
Plural Term	Ducted Air-Conditioners	Impact Barriers
Definition Term	Encased assembly or assemblies designed primarily to provide ducted delivery of conditioned air to an enclosed space, room or zone	Energy-absorbing device installed in front of a rigid object to reduce the severity of impact of a vehicle

expression linked to one target object in the image, which helps isolate the impact of wording changes from the image's context or composition. For each object class, multiple lexical variants were created, such as a base term, synonyms, technical terms, plural forms, and descriptive definitions. Concrete examples of the prompt construction for each lexical category are reported in Table 1, which illustrates how prompts were instantiated for representative indoor and outdoor object classes. All variants were evaluated independently under identical but different visual conditions, allowing a controlled comparison of how lexical choices influence detection performance while maintaining a one-to-one correspondence between the prompt and the object.

Base Term

The Base Term labels adopted in this study are based on the labelling strategy introduced by Taherian et al. for evaluating *GroundingDINO* as an OVOD model on construction job sites under complex visual scenarios. In that work, object class labels were intentionally formulated using clear, unambiguous natural-language expressions, with each image assigned to a single target object. This approach ensured a one-to-one correspondence between the text prompt and the visual instance. Following the same principle, Base Term in this study is defined as the most direct and commonly used words for each object class, reflecting standard naming practices used by professionals in the AEC field. These terms act as the linguistic baseline for prompt formulation and show the least linguistically altered way of naming objects, against which the effects of lexical differences are later assessed. However, the work by Taherian et al. did not explicitly analyze the effects of lexical variation on model performance. Building on this gap, the present work further examines the influence of inflectional variants by evaluating whether the use of singular or plural forms of a term affects object recognition accuracy (Minderer *et al.*, 2024).

Synonym Term

Lexical variation using Synonyms Terms was analyzed to verify the model's ability to recognize the same object when indicated using semantically equivalent terms (Nebbia and Kovashka, 2024).For each object class, synonyms were assigned through a cross-dataset analysis of naming conventions in related datasets, including *COCO* (Lin *et al.*, 2014), *Objects365* (Shao *et al.*, 2019), *Open Images* (Open Images V7), and *SODA* (Duan *et al.*, 2022). For missing class, *WordNet* served as a complementary lexical reference (Miller et al., n.d.), and for remaining domain-specific classes not supported by either dataset taxonomies or *WordNet*, additional terminology was derived from online annotation repositories such as *Roboflow* to reflect common usage. This cross-sectional analysis of multiple datasets and resources not only underscored domain gaps, as several object classes in the developed dataset were not well represented in existing datasets, but also allowed for mitigating biases associated with individual taxonomies and evaluating the model's robustness to synonym-based lexical substitutions.

Technical Term and Plural Term

Lexical variation using "related technical terms" was examined, with a focus on terminology in construction sector regulations. The Technical Terms in tests came from international standards in construction and plant sectors, including the *International Organisation for Standardisation* (ISO), the *International Electrotechnical Commission* (IEC), and the *European Norms* (EN/CEN). This choice is motivated by the need to evaluate the model's behavior in the presence of specialized terminology consistent with the application domain's regulatory practice. The main regulations analyzed are listed in Table 2.

In addition, these Technical Terms were also inflected according to standard English grammatical rules to generate Plural Forms, enabling the assessment of the model's response to number variation within specialized terminology.

Descriptive Term

Finally, tests based on descriptive prompts (Descriptive Technical Term) have been established, as many open-vocabulary object detection benchmarks use textual representations based on descriptive sentences rather than individual tokens (Bianchi *et al.*, 2023). Technical regulations often provide descriptive textual definitions of objects, which have been utilized to construct descriptive technical terms. These definitions are regarded as formal and unambiguous linguistic representations of objects.

Prompt Multiplicity

In contrast to single-prompt inference, the multiple-prompt approach enables the model to detect and localize several object categories within a single image. This is achieved by constructing a prompt configuration

composed of multiple textual queries, each corresponding to a target object class. The input prompt is formatted as a sequence of object descriptors separated by delimiters, for example:

Prompt = {"object 1.", "object 2.", ..., "object n."}.

Within this setup, the prompt configuration is further extended to evaluate model hallucination. This is done by augmenting the multi-object prompt with an additional query that corresponds to an object does not present in the image. The absent object is selected from the dataset object categories based on its visual similarity to the AEC-related object(s) depicted in the scene, enabling systematic evaluation of the model's susceptibility to visually induced false detections under multi-prompt conditions. An example of prompt format is as following with further demonstration provided in Figure 5.f:

Prompt = {"object 1", "object 2", ..., "object n", "absent object"}.

Table 2: Regulations analyzed for defining the Single prompt-Technical Term

Standard	Domain
ISO 13253:2017	Ducted air conditioners & air-to-air heat pumps
ISO 6707-1:2020	Building construction
ISO 7240-1:2014	Fire safety
ISO 8421-4:1990	Fire safety
ISO 6165:2022	Earth-moving machinery
ISO 30500:2018	Non-sewered sanitation systems
ISO 18650-1	Construction machinery
ISO 45001:2018	Occupational health and safety management systems
ISO 8528-1:2018	Generating sets
ISO 7010:2019	Safety signage & fire prevention symbols
ISO 830:2024	Logistics
ISO 14122-3:2016	Machinery safety
ISO 19085-9:2024	Machinery safety
ISO/CIE 8995-1:2025	Workplace lighting
IEC 61439-2:2020	Low-voltage switchgear and controlgear assemblies
IEC 60364-1:2025	Low-voltage electrical installations
IEC 60050 (series)	Electrotechnical terminology
EN 442-2:2014	Heating systems

Experiment Setup

Based on the proposed framework, prompt elements were systematically defined and combined to create experimental test cases, enabling a controlled analysis of how different forms of lexical and compositional variation influence the OVOD model's object recognition performance.

OVOD Model

In this research, *GroundingDINO* is used as the OVOD model. *GroundingDINO* is designed to operate in scenarios where the categories to be detected are not fixed in advance but are defined dynamically through textual prompts. In this paradigm, text is the primary means for semantic conditioning of inference, making prompt formulation a key factor in the model's behavior, as discussed in works on grounded detection and phrase grounding (Li *et al.*, 2021; Liu *et al.*, 2023). The model is built upon the Transformer-based DINO detector, a member of the DETR family, and extends it by incorporating grounded language–image pre-training for open-set object detection (Carion *et al.*, 2020; Li *et al.*, 2021). In this approach, categories are not treated as separate labels but as context-aware linguistic representations generated by a Transformer-based text encoder, enabling connections between visual regions and textual descriptions, including lexical variants and free phrases (Devlin *et al.*, 2018; Li *et al.*, 2021). From an architectural perspective, *GroundingDINO* combines a multi-scale visual backbone with a pre-trained language encoder and employs a deep multimodal fusion strategy in which the textual prompt affects multiple stages of the detection process, from visual feature enhancement to query initialisation and final prediction refinement. (Liu *et al.*, 2023). Consequently, variations in prompt formulation can affect text–region alignment and prediction consistency, making the model sensitive to linguistic choices at inference time. For inference in this study, GroundingDINO was implemented using the Hugging Face Transformers via zero-shot-object-detection pipeline (Hugging Face, 2025). Given the input prompts, the model returns bounding boxes, associated labels, and confidence scores for detected regions. Each confidence score is computed by selecting the maximum logit across all candidate labels and applying a sigmoid function to obtain a normalized value between 0 and 1, reflecting the alignment between a region and its most relevant textual label. A single confidence threshold is then applied to filter the final detection outputs. In this study, a threshold of 0.25 was used to retain predictions for further evaluation (Raoofi *et al.*, 2024). Future work will investigate the impact of varying this threshold on detection performance and its interaction with prompt formulation.

Evaluation Metrics

To evaluate the performance, detection outputs were filtered using a confidence threshold of 0.25 prior to ground-truth (GT) association and metric computation. *GroundingDINO* operates on free-form input prompts and the evaluation involves the verification of region–prompt alignment. A predicted bounding box was considered a correct detection if its Intersection-over-Union (IoU) with a GT instance exceeded the threshold, and the assigned class label correctly corresponded to the ground-truth class label. IoU quantifies spatial overlap and is defined as the ratio of the intersection area $|A \cap B|$ to the union area $|A \cup B|$ between the predicted (A) and GT (B) bounding boxes.

IoU is highly sensitive to minor localization errors, where even small pixel-level shifts can reduce the overlap,

despite successful detection. Since the dataset in this study exhibits scale variation and includes particularly small objects per class, the IoU threshold was set to 0.25 ensuring that the detection of small-scale indoor components is not unfairly penalized while maintaining the focus on reliable object presence verification (Eggert *et al.*, 2017). Precision, Recall, and F1-score were computed for each object class and experimental scenario according to Eq. (1). A detection was counted as a true positive when the IoU threshold was met. GT instances without matched predictions were labelled as false negatives, while incorrectly associated predictions were treated as false positives.

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}, F1 = \frac{2\times P\times R}{P+R} \quad (1)$$

Results and Discussion

This section shows *GroundingDINO*'s evaluation results in construction scenarios. The analysis focuses on two main questions: (i) how prompt formulation impacts the reliability of OVOD; (ii) whether using multiple prompts enhances performance.

An initial overview of the results compares the Single-Prompt and Multiple-Prompt configurations for indoor and otdoor scenes, as shown in Figure 2. The Multiple Prompt setup does not significantly outperform the Single Prompt baseline: recall drops slightly from 0.752 to 0.716, while precision (0.448 vs. 0.446) and F1 score (0.561 vs. 0.550) are nearly the same or slightly lower with the multiple-prompt configuration. These results suggest that, in the present experimental setting, increasing the number of prompts does not yield a clear overall improvement in OVOD performance and may introduce additional ambiguity.

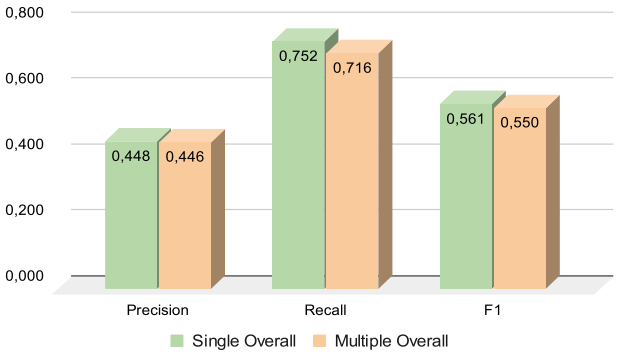


Figure 2: Overall performance comparison between Single Prompt and Multiple Prompt.

A more detailed analysis categorized by scene type indicates that the effect of prompt formulation is significantly influenced by contextual factors. In indoor environments (Figure 3), the utilization of multiple prompts results in a moderate enhancement of the F1 score, rising from 0.547 in the Single Prompt condition to 0.577 in the Multiple Prompt arrangement. This enhancement is chiefly attributable to an increase in recall (from 0.706 to 0.750), with a slight improvement observed in precision (from 0.447 to 0.469). These results suggest that, in structured indoor construction scenes with constrained geometry and reduced background clutter, prompt multiplicity can help the model retrieve a broader set of relevant objects without substantially increasing false positives.

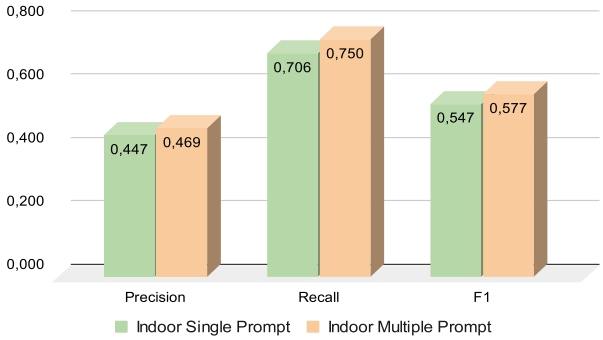


Figure 3: Performance comparison by scene type: Indoor

In contrast, outdoor construction scenes exhibit markedly different behavior (Figure 4). In this case, the Multiple Prompt configuration leads to a simultaneous reduction in recall (from 0.800 to 0.680) and precision (from 0.449 to 0.422), resulting in a more noticeable drop in F1 score from 0.575 to 0.520.

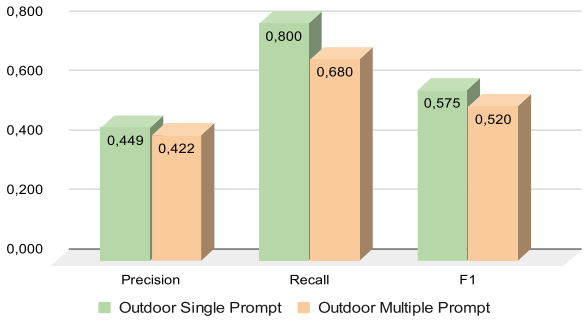


Figure 4: Performance comparison by scene type: Outdoor.

In visually complex, less structured outdoor environments, prompt multiplicity does not yield clear improvements in detecting and discriminating target objects and may be associated with less stable detection outcomes. The presence of heterogeneous materials and temporary structures can raise semantic interference between prompts, leading to both missed detections and false positives. Overall, managing multiple prompts seems more difficult in these environments. To further investigate prompt sensitivity, the following analysis focuses on lexical variation under single-prompt conditions, allowing the effect of wording to be examined independently from prompt multiplicity.

Lexical variation

Table 3 summarizes how lexical variation impacts detection performance. Because this analysis aims to assess the model's sensitivity to lexical differences rather than scene-specific features, the results are aggregated across both indoor and outdoor environments. The Base Term prompt serves as the baseline and achieves the highest overall F1 score (0.561), suggesting that normal and commonly used language for object class names

yields the most reliable alignment between textual prompts and visual representations in this experimental setting. The Synonym Term strategy yields a very similar performance (F1 = 0.555), with nearly identical precision and a slight reduction in recall. This result suggests that *GroundingDINO* exhibits a limited performance variation across closely related lexical forms, although even minor variations in wording can affect recall. A pronounced performance degradation is observed when transitioning to Technical Term prompts. In this case, both precision (0.395) and recall (0.622) decrease, resulting in a lower F1 score (0.483). This behaviour indicates that specialised and domain-specific terminology, while semantically precise from a technical or regulatory standpoint, is less effectively grounded by the model. This reflects a mismatch between the linguistic distributions encountered during large-scale pre-training and the formal vocabulary commonly used in construction standards and technical documentation. The use of Plural Term further degrades performance (F1 = 0.454), with the reduction driven primarily by a drop in precision. This finding suggests that the model is sensitive to inflectional variation and that plurality introduces ambiguity in the mapping between textual prompts and visual instances, even when the underlying semantic concept remains unchanged. This ambiguity increases false detections, thereby reducing detection reliability. The lowest performance is obtained with Definition Term prompts, which exhibit a substantial decline across all metrics (F1 = 0.314). In this configuration, the combination of descriptive modifiers and increased linguistic complexity prevents the model from identifying a stable visual reference for the target object, resulting in both poor recall and a high false-positive rate. Overall, these results suggest a trade-off between lexical variation and detection robustness. While *GroundingDINO* appears to tolerate limited lexical variation, performance tends to decrease as prompts differ from concise normal language object names and incorporate greater technical, inflectional, or descriptive complexity. This emphasizes that in the context of OVOD for construction applications, semantically accurate prompt formulations from a human perspective do not necessarily translate into improved model performance and may instead hinder reliable grounding.

Table 3: Precision, recall, and F1 scores for single-prompt lexical variation strategies.

Prompt Strategy	Precision	Recall	F1
Base Term	0.448	0.752	0.561
Synonym Term	0.451	0.720	0.555
Technical Term	0.395	0.622	0.483
Plural Term	0.347	0.657	0.454
Definition Term	0.230	0.496	0.314

Multiple prompting and Hallucination.

In this section, the analysis focuses on the effects of multiple promoting and hallucination, particularly when prompts include object categories that are not present in the scene. Table 4 reports both absolute performance metrics and relative percentage variations with respect to the single-prompt baseline, enabling a clearer comparison of the impact of hallucination and multi-object prompting configurations on detection performance.

Table 4: Precision (P), recall (R), and F1 scores for SP, MP, and Hall. configurations (Δ% relative to single prompt).

	P	ΔP (%)	R	ΔR (%)	F1	Δ F1 (%)
SP	0.448	–	0.752	–	0.561	–
MP	0.446	−0.4%	0.716	−4.8%	0.550	−2.0%
Hall.	0.447	−0.2%	0.737	−2.0%	0.557	−0.7%

Findings show that Precision remains nearly unchanged across configurations. Moving from single-prompt to multiple-prompt, precision decreases by approximately 0.4%, while the hallucination setting shows only a minor decrease of 0.2%. Recall exhibits a more noticeable change, with a 4.8% decrease for multiple-prompt and 2.0% decrease for hallucination relative to the single-prompt baseline, resulting in modest decreases in F1 score (−2.0% and −0.7%, respectively). These results suggest that hallucination primarily affects the model's ability to reliably retrieve all relevant objects, while having a limited impact on precision.

Given the qualitative examples in Figure 5, the results show that the impact of increased prompt multiplicity is class dependent. For instance, the *sprinkler* (Figure 5-a/5-b) category is not detected under single prompting and remains unaffected by multiple prompting. In contrast, some categories exhibit stable behavior across prompting strategies, such as *ladder* (Figure 5-c/5-d), which is correctly detected under both single- and multiple-prompt configurations. Other categories, however, show marked performance variations when multiple prompts are introduced, reflecting increased semantic ambiguity and class-level confusion. This effect is evident for *red & white tape* (Figure 5-d), where multiple prompting leads to an increase in false detections. Moreover, the results indicate that introducing prompts for object categories absent from the scene can induce false detections without affecting the detection of present targets. An example of this behavior occurs when the "absent duct" (Figure 5-f) category is incorrectly assigned to another object in the image while the present object still is correctly detected (Figure 5-e compared to Figure 5-f). Overall, the comparison between configurations indicates that hallucination impacts OVOD performance by adding false detections. This reveals a trade-off between target detection (recall) and detection accuracy (precision), which should be carefully balanced based on the specificity and requirements of the construction scenario under analysis.

Limitations and Future Works

This work is intended as a focused, exploratory study of prompt sensitivity in OVOD for construction scenarios. The experimental design studies different prompt configurations across visual conditions for various object



a) "Sprinkler" - Single Prompt
Failure: Target Missed detection & FPs



c) "Ladder" – Single Prompt
Success: Correct detection without FPs



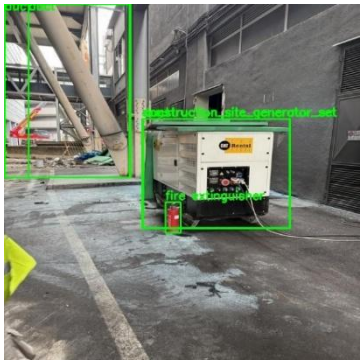
e) "CS generator"- Single Prompt
Succuss: Correct detection without FPs



b) "Sprinkler", "Vent cover", and "lighting" – Multiple Prompt
Failure: Target missed detection for "sprinkler" but correct detections for other classes.



d) "Ladder" and "Red&white tape"- Multiple Prompt
Success: Correct detection for targets but increased FPs



f) "CS generator set", "Fire extinguisher", and "Duct" – Multiple Prompt
Hallucination: FP for absent object (Duct) but correct detections for present Targets

Figure 5: Qualitative detection examples showing representative successes and failure modes under different prompting strategies

categories, yielding numerous evaluation instances. Although this has enabled structured analysis of prompt behavior, some aspects of the evaluation have been limited, given the scope and space constraints. First, while the dataset captures both indoor and outdoor construction environments with diverse object categories and varying visual conditions to reflect real-world scenarios, the number of images per object category remains limited. Second, the evaluation is conducted using a single threshold for IoU. The IoU threshold was intentionally selected to better account for small and partially occluded objects common in construction scenes. However, it limits comparability with standard object detection benchmarks. Future work will address these limitations by expanding the dataset to include a larger and more balanced set of images across object categories. Additionally, subsequent studies will evaluate performance across multiple thresholds to enable more comprehensive and standardized comparisons. Building on the current findings, future research will also further investigate and refine prompting strategies to improve robustness across diverse construction environments.

Conclusions

This study presented an exploratory investigation of prompt robustness in OVOD for construction scenarios, focusing on lexical variation and prompt multiplicity using *GroundingDINO*. The results of this study indicate patterns rather than providing conclusive evidence and show that prompt formulation plays a relevant role in detection performance, influencing both recall–precision trade-offs and the occurrence of hallucinated predictions. In particular, simple base-term prompts provide the most stable and reliable results, while increasing linguistic complexity tends to reduce detection performance. The analysis also shows that multiple-prompt configurations do not consistently improve performance and may introduce context-dependent effects. While some benefits are observed in structured indoor environments, more complex outdoor scenes are associated with reduced stability and increased false detections. Furthermore, the introduction of hallucinated object categories highlights a limitation of OVOD systems, as the model tends to produce false detections rather than null responses, affecting detection completeness and reliability. These findings suggest a gap between human semantic expectations and the model's grounding behavior, which may affect deployment in safety-critical construction applications. Overall, the study emphasizes the importance of careful prompt design in OVOD-based AEC workflows. Future work will extend the evaluation across multiple IoU thresholds and investigate more robust prompting strategies to improve reliability and reduce hallucination effects in real-world construction scenarios.

References

Adil, M., Lee, G., Gonzalez, V.A. and Mei, Q., 2025. Using vision language models for safety hazard identification in construction. *arXiv:2504.09083*.

Bianchi, L., Carrara, F., Messina, N., Gennaro, C. and Falchi, F., 2024. The devil is in the fine-grained details: Evaluating open-vocabulary object detectors for fine-grained understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 22520-22529). DOI: 10.1109/CVPR52733.2024.02125.

Bravo, M.A., Mittal, S., Ging, S. and Brox, T., 2023. Open-vocabulary attribute detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7041-7050). DOI: 10.1109/CVPR52729.2023.00680.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S., 2020, August. End-to-end object detection with transformers. In *European conference on computer vision* (pp. 213-229). Cham: Springer International Publishing. DOI: 10.1007/978-3-030-58452-8_13.

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., 2019, June. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1* (pp. 4171-4186). *arXiv:1810.04805.*

Duan, R., Deng, H., Tian, M., Deng, Y. and Lin, J., 2022. SODA: A large-scale open site object detection dataset for deep learning in construction. *Automation in Construction*, *142*, p.104499.

Eggert, C., Zecha, D., Brehm, S. and Lienhart, R., 2017, June. Improving small object proposals for company logo detection. In *Proceedings of the 2017 ACM on international conference on multimedia retrieval* (pp. 167-174). DOI: 10.1145/3078971.3078990

Gu, X., Lin, T.Y., Kuo, W. and Cui, Y., 2021. Open-vocabulary object detection via vision and language knowledge distillation. *arXiv:2104.13921*.

Gu, J., Han, Z., Chen, S., Beirami, A., He, B., Zhang, G., Liao, R., Qin, Y., Tresp, V. and Torr, P., 2023. A systematic survey of prompt engineering on vision-language foundation models. *arXiv:2307.12980*.

Hugging Face (2025) *Zero-shot object detection*, Hugging Face Transformers Documentation, version 5.4.0. Available at: https://huggingface.co/docs/transformers/v5.4.0/en/tasks/zero_shot_object_detection (Accessed: 31 March 2026).

Li, L.H., Zhang, P., Zhang, H., Yang, J., Li, C., Zhong, Y., Wang, L., Yuan, L., Zhang, L., Hwang, J.N. and Chang, K.W., 2022. Grounded language-image pre-training. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10965-10975). DOI:10.1109/CVPR52688.2022.01069.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L., 2014, September. Microsoft coco: Common objects in context. In *European conference on computer vision* (pp. 740-755). Cham: Springer International Publishing. DOI:/10.1007/978-3-319-10602-1_48.

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H. and Zhu, J., 2024, September. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *European conference on computer vision* (pp. 38-55). Cham: Springer Nature Switzerland. DOI: 10.1007/978-3-031-72970-6_3.

Malaikrisanachalee, S., Wongwai, N. and Kowcharoen, E., 2025. ESPCN-YOLO: A high-accuracy framework for personal protective equipment detection under low-light and small object conditions. *Buildings*, *15*(10), p.1609. DOI: 10.3390/BUILDINGS15101609.

Minderer, M., Gritsenko, A. and Houlsby, N., 2023. Scaling open-vocabulary object detection. *Advances in Neural Information Processing Systems*, *36*, pp.72983-73007.a*rXiv: 2306.09683*.

Mütze, A., Ilyas, S., Dörpelkus, C. and Rottmann, M., 2026. Can We Challenge Open-Vocabulary Object Detectors with Generated Content in Street Scenes?. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 740-750). DOI: 10.48550/arXiv.2506.23751.

Nebbia, G. and Kovashka, A., 2024, June. Synonym relations affect object detection learned on vision-language data. In *Findings of the Association for Computational Linguistics: NAACL 2024* (pp. 3770-3776). DOI:/10.18653/v1/2024.findings-naacl.239.

Open Images Dataset V7 (s.d.) Open Images Dataset V7 and Extensions. https://storage.googleapis.com/openimages/web/index.html (Accessed: 5 February 2026).

Raoofi, H., Sabahnia, A., Barbeau, D. and Motamedi, A., 2024. Deep learning method to detect missing welds for joist assembly line. *Applied System Innovation*, *7*(1), p.16. DOI: 10.3390/asi7010016.

Shao, S., Li, Z., Zhang, T., Peng, C., Yu, G., Zhang, X., Li, J. and Sun, J., 2019. Objects365: A large-scale, high-quality dataset for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 8430-8439). DOI: 10.1109/ICCV.2019.00852.

Taherian, G. et al. (2026), "Vision–Language Models in Construction: Evaluating Open-Vocabulary Object Detection under Complex Scenarios," Proceedings of the Canadian Society for Civil Engineering (CSCE) Annual Conference 2026, Québec City, QC, Canada, June 3–5, 2026, to be published.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

Benchmarking Deep Learning Approaches for AEC Engineering Drawing Layout Detection and Information Extraction

Tianyang Huang[1,2], Alessio Lombardi[1], Ahmed Elnagar[1], Ahmed Zalouk[2], George Paul[1], Sepehr Najjarpour[1]
Arvid Sigurdsson[1], Khalid Ismail[2], Mohamed Ragab[2], Edlira Vakaj[2]
[1]Buro Happold Ltd., London, United Kingdom
[2]Birmingham City University, Birmingham, United Kingdom

Abstract

Information Extraction (IE) from Architecture, Engineering, and Construction (AEC) drawings remains hindered by manual inefficiency, while Layout Detection—a vital "middleware" organizing graphical and textual hierarchies—is underexplored. General document layout models, optimized for text-centric content, lack validation on engineering drawings. This study constructs a custom AEC-specific layouts dataset and benchmarks five deep learning architectures. RF-DETR achieves state-of-the-art performance with an mAP_{50} of 0.949, while the Vision-Language Model Qwen3-VL attains a leading F1-score of 0.911. Conversely, models pre-trained on general document datasets suffer from "domain interference", causing performance degradation. This establishes a robust technical foundation for automated IE in AEC.

Introduction

In the Architecture, Engineering, and Construction (AEC) industry, drawings are still a common medium for information exchange. Efficient Information Extraction (IE) is essential for downstream tasks such as building refurbishment, Quality Assurance (QA), and digital twin development. In contemporary practice, however, IE from drawings remains heavily reliant on manual effort, which is labor-intensive and error-prone. Automating drawing interpretation is thus an urgent industrial priority.

Information within an engineering drawing can be systematically categorized into three hierarchical layers: the *Metadata Layer* (titleblocks), the *Layout Layer* (spatial arrangement), and the *Main Drawing Content Layer* (graphical entities) (shown in Figure 1). Serving as the vital "middleware" between these two, the *Layout Layer* defines the logical relationships and hierarchical structure between disparate elements, such as drawing areas, notes, and titleblocks.

Previous efforts have primarily focused on isolated layers of this hierarchy, such as automated metadata extraction from titleblocks (Lombardi et al., 2025) or localized symbol detection (Khan et al., 2024; Jamieson et al., 2025). While advancing specific tasks, these studies often overlook the overall engineering drawing layout. This research gap is critical because the layout provides the semantic framework necessary for holistic understanding; for instance, the meaning of a symbol often depends on its functional orientation within the layout context. Without this structural link, isolated IE loses the global logic required for complex engineering reasoning.

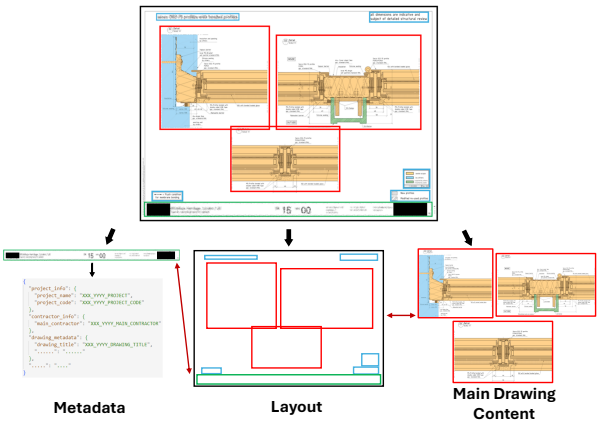


Figure 1: Hierarchical layers of information in AEC drawings

While Document AI has produced powerful Document Layout Analysis (DLA) models like LayoutLM (Xu et al., 2020, 2021; Huang et al., 2022) and DocLayout-YOLO (Zhao et al., 2024b), these are primarily optimized for general text-centric documents such as invoices or academic papers. Due to the unique characteristics of engineering drawings—such as extreme aspect ratios, high-resolution requirements, and sparse graphical density—the performance of general-purpose Document AI models in this domain has not been systematically benchmarked or validated.

To fill this research gap, we collaborated with industry experts to construct the *Facade Engineering Layout Dataset*, comprising 551 high-quality annotated drawings. Based on this dataset, we conducted a comprehensive benchmark of state-of-the-art models, including Faster R-CNN (Ren et al., 2017), YOLOv12 (Tian et al., 2025), RF-DETR (Robinson et al., 2026), Qwen3-VL (Bai et al., 2025) and DocLayout-YOLO, evaluating their efficacy in engineering drawing layout detection. Additionally, an ablation study on DocLayout-YOLO investigates whether general DLA architectural designs and pre-trained weights can be effectively transferred to the AEC domain.

The primary contributions of this research are as follows.

1. We identify the importance of Layout Detection in the AEC domain as a prerequisite for holistic drawing un-

derstanding.
2. We provide an extensive benchmark of multiple deep learning architectures, analyzing their strengths and limitations when dealing with large-scale, high-complexity engineering documents.
3. We reveal a significant "domain interference" effect through the ablation study, demonstrating that specialized architectural designs and pre-trained weights from general document domains can act as negative priors for AEC drawings.

Our findings establish a robust technical pathway for performing detailed IE and QA on engineering drawings in AEC projects.

Related Work

Information Extraction in Engineering Drawings

Engineering drawings encapsulate high-density geometric and semantic data, necessitating their transformation into structured formats for industrial digitalization. Current research predominantly bifurcates into two domains: Metadata Extraction from titleblocks and Main Drawing Content Analysis.

Metadata Extraction from Titleblocks

Early heuristic and OCR-based methods (Kashevnik et al., 2023) often lack robustness against non-standard layouts or scanning artifacts. Recent advancements have shifted toward Transformer-based end-to-end detection. Microsoft's Donut (Kim et al., 2022) bypassed traditional OCR errors by employing a joint visual encoder and language decoder. Furthermore, Lombardi et al. (2025) integrated Faster R-CNN with GPT-4o to parse titleblocks in AEC drawings into structured key-value pairs. They achieved 86% key retrieval and 95% value-matching accuracy, demonstrating Multimodal Large Language Model's potential for semi-structured AEC documents.

Main Drawing Content Analysis

In contrast to titleblocks, information within the main drawing area is dispersed and topologically complex. Zhang et al. (2023) addressed this by modeling mechanical drawings as graph structures, utilizing GraphSAGE to classify geometric contours and annotations. Khan et al. (2026) proposed a two-stage hybrid framework using YOLOv11-OBB to identify text regions (e.g., GD&T, materials, surface roughness), followed by a finetuned Donut model for structured JSON extraction. To extract main drawing content from high-resolution AEC drawings, current studies frequently employ tiling strategies—partitioning massive images into smaller patches to ensure compatibility with Vision Transformers or CNN-based detectors (Jamieson, 2024; Carrara et al., 2025).

Ontology-based floor plan analysis (DAnO)

The work of Schönfelder and König (2025) proposed the *Drawing Analysis Ontology* (DAnO) as an ontology-driven framework for automatic floor plan analysis, focusing primarily on architectural plans. It provides a domain model to represent drawing elements and their semantic relationships, and it is used to support reasoning tasks such as aggregation and consistency checking on floor-plan content. In typical DAnO use cases, each drawing sheet contains a single main plan view, and the ontology targets the interpretation of this view rather than the explicit modelling of multiple layouts within a sheet.

Layout Detection: The Missing Link

Existing AEC "layout" research is primarily restricted to architectural floor plans, focusing on semantic segmentation of object elements (e.g. walls, doors) within individual views (Heras et al., 2014). These approaches fail to address the holistic sheet-level layout of generic AEC drawings, where heterogeneous views—such as plans, sections, and elevations—coexist. To bridge this gap, we propose multi-view layout recognition as a strategic "middleware" stage that facilitates holistic data extraction through the following benefits:

Semantic Contextualization: Layout detection identifies the spatial distribution of drawing areas, which is critical as identical symbols carry distinct semantic meanings depending on their viewport (e.g., plan vs. section). Without this context, isolated symbol detection loses essential semantic associations.

Optimizing High-Resolution Processing: Unlike "blind" tiling techniques that cause semantic fragmentation by severing continuous engineering entities, prioritizing layout detection identifies information-dense "Regions of Interest" (RoI). This facilitates semantic-aware tiling, ensuring high-resolution crops maintain logical boundaries and necessary global context for downstream extraction tasks.

General Document Layout Analysis

While layout analysis remains under-explored in the AEC sector, general Document Layout Analysis (DLA) has transitioned from early heuristic rules (Namboodiri and Jain, 2007) to sophisticated deep learning paradigms. This shift was catalyzed by large-scale benchmarks such as PubLayNet (Zhong et al., 2019) and the structurally diverse DocLayNet (Pfitzmann et al., 2022); the latter has become the primary benchmark in the field due to its accurate and detailed manual labels. Currently, the field is dominated by multimodal Transformers, notably the LayoutLM series (Xu et al., 2020, 2021; Huang et al., 2022), which integrates visual and positional features for SOTA performance. To meet industrial throughput requirements, models like DocLayout-YOLO (Zhao et al., 2024b) have emerged, utilizing a G2L_CRM architecture to achieve competitive accuracy with significantly faster inference speeds. Recent research has further matured toward advanced structural understanding, including reading path reconstruction (Wang et al., 2021) and graph-based document structure analysis (Chen et al., 2025).

The Research Gap: Unique Challenges of AEC Drawings

Despite the maturation of general DLA, research specifically tailored to AEC drawings remains sparse. Models optimized for general documents cannot be directly migrated due to data scarcity and significant domain disparity. Unlike text-centric documents, AEC drawings feature massive formats, a dominance of non-textual symbols, and complex topological interdependencies that require specialized handling.

To address these challenges, this study constructs a high-quality *Facade Engineering Layout Dataset* and provides a comprehensive benchmark of SOTA models on this specialized domain. Moreover, an ablation study of DocLayout-YOLO is conducted to systematically evaluate how its specialized G2L_CRM architecture and large-scale document pre-training weights influence performance when transferred to the unique graphical landscape of AEC engineering drawings.

Methodology

In this section, we outline our methodology for benchmarking engineering drawing layout detection. We first introduce the curation of the *Facade Engineering Layout Dataset* (FELD). This is followed by our selection of candidate models and the standardized experimental framework used for evaluation.

Dataset and Ontology Overview

To support the benchmarking experiments, this study curated the FELD dataset. This dataset consists of 551 facade drawing images derived from 10 real-world construction projects, comprising nine common drawing layout elements. A comprehensive overview of the dataset is provided in Table 1 below.

Table 1: FELD Dataset Overview

Category Name	Image Counts	Annotation Counts
Titleblock	551	604
Titleblock; Revision Table	425	425
Legend	54	91
Notes	145	288
Layout; Elevation	56	60
Layout; Section; Horizontal	239	321
Layout; Section; Vertical	307	376
Layout; Section; GA	23	37
Layout; Plan; GA	21	26
All Categories	**551**	**2228**

GA: General arrangement

Prior to annotation, we first collaborated with facade domain experts to define an AEC Drawing Information Representation Ontology (ADIRO); the ontology is illustrated in Figure 2 below. This ontology is intended to delineate the spatial structure and semantic connotations of the drawings through a hierarchical logical framework. ADIRO is tailored to the needs of IE from engineering drawings, rather than directly adopting existing building ontologies or classification schemes. Existing models and ontologies (e.g., IFC, ifcOWL, Uniclass, Brick, BOT) are designed primarily for BIM authoring, lifecycle asset management, or building automation, and therefore optimize for comprehensive domain coverage and interoperability. As a result, they tend to be large, multi-purpose and conceptually heavy, with many distinctions and constraints that are irrelevant at the level of 2D drawing analysis and information extraction. For drawing IE, we instead require a schema that is tightly aligned with visual patterns in drawings, annotation workflows, and graph-based reasoning, in addition to local spatial relations. A custom ontology allows one to minimise complexity, expose only the concepts and relations that matter for IE, downstream queries, and QA checks, and iterate rapidly as new tasks emerge. Interoperability is preservable by attaching external identifiers (such as IFC types or Uniclass codes) as attributes when required, while the core representation remains simple and learnable. We included nine of the most representative and common layout categories in ADIRO (see Table 1) and established annotation guidelines.

ADIRO is closely related to DAnO (Schönfelder and König, 2025), which defines concepts such as DrawingElement, DisplayElement, and DescriptionElement, and their relationships, enabling semantic graph-based reasoning over text–symbol–geometry relations in architectural plans. In contrast, ADIRO is designed for engineering drawings where multiple layouts commonly coexist on a single sheet. To capture this, ADIRO introduces Layout as a parent concept to DrawingElement: a layout can contain drawing elements as well as other sheet content such as tables, schedules, and textual regions, thereby explicitly modelling the hierarchical structure of multi-layout sheets. While ADIRO and DAnO partially overlap at the level of basic drawing concepts, their scope and structuring differ; we therefore consider ADIRO complementary to DAnO and plan to explore explicit linking or reuse of DAnO modules in future extensions of ADIRO, once we move beyond the layout-focused core required for the present work.

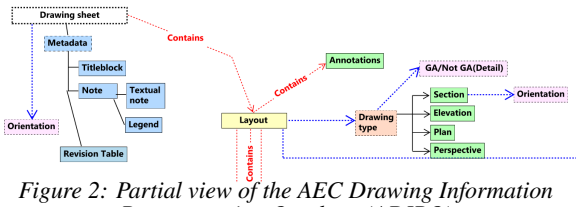


Figure 2: Partial view of the AEC Drawing Information Representation Ontology (ADIRO)

During the annotation phase, domain experts performed labeling which was cross-checked. Standard rasterization was performed to ensure consistency across vector and raster drawings, with all drawings at 300 DPI. Due to the varying dimensions of the originals, the dataset image dimensions range from 3509×2480 to 14042×9934 pixels.

Evaluated Model Architectures

To comprehensively evaluate layout detection in AEC drawings, we selected five architectures representing distinct technical paradigms:

Faster-RCNN & YOLOv12

Faster R-CNN is employed as a robust two-stage baseline to evaluate traditional anchor-based frameworks on large-scale layouts. Conversely, YOLOv12 (m) represents the state-of-the-art in one-stage detectors, selected to assess the balance between real-time inference efficiency and detection accuracy via its Area Attention mechanism.

RF-DETR

RF-DETR is selected as the latest iteration of Transformer-based detection models and currently stands as a SOTA solution in the field. By reframing detection as a set prediction problem, it provides a purely end-to-end pipeline without reliance on Non-Maximum Suppression (NMS). We utilize RF-DETR to evaluate the efficacy of DETR's global attention mechanisms in modeling the intricate topological relationships between disparate drawing zones.

Qwen3-VL

To explore the potential of state-of-the-art multimodal models in specialized domains, we introduce Qwen3-VL. This high-performance vision-language model natively supports long-text streams up to 256K tokens and utilizes an enhanced Interleaved-MROPE (Multimodal Rotary Positional Embedding) to improve spatial modeling for large-scale documents.

By benchmarking Qwen3-VL-8B-Instruct, we evaluate whether the semantic knowledge from large-scale multimodal pre-training can be effectively transferred to the specialized domain of architectural symbolic layouts.

DocLayout-YOLO

Specifically engineered for general DLA, DocLayout-YOLO is included to evaluate its domain-specific architectural optimizations and whether pre-training on general document layout datasets can be effectively transferred to AEC drawings. Based on the YOLOv10m backbone, its core innovation—Global-to-Local Contextual Modeling (G2L_CRM) — enables the model to simultaneously perceive the overall logical structure of a page and fine-grained semantic details. By comparing this model with general-purpose detectors and conducting additional ablation studies, we aim to quantify the extent to which its architectural enhancements and large-scale document pre-training contribute to performance gains when applied to AEC-specific documents.

Experimental Setup

Hardware and Environmental Configuration

All experiments were implemented using the PyTorch framework on the Ubuntu operating system. To ensure computational efficiency, Qwen3-VL-8B-Instruct was trained on an NC24ads-A100-v4 instance with MS-Swift framework (Zhao et al., 2024a), while Faster R-CNN, YOLOv12, RF-DETR, and DocLayout-YOLO were trained on an NC4as-T4-v3 instance.

Model Specifications and Initialization

To ensure a fair and comprehensive comparison, we selected specific variants of each architecture and initialized them with specialized pre-trained weights. Faster R-CNN was implemented using the ResNet-50 backbone with Feature Response Normalization (FRN). For the YOLOv12 and RF-DETR architectures, we utilized their respective medium (m) variants to balance computational cost and detection capacity. These three models were initialized with weights pre-trained on the COCO dataset. In contrast, Qwen3-VL was deployed using the 8B-Instruct version, which had undergone extensive large-scale pre-training. For DocLayout-YOLO, we utilized a version specifically pre-trained on large-scale document layout datasets, namely DocLayNet and DocSynth, to leverage its domain-specific knowledge in general DLA.

Data Partitioning and Preprocessing

Given the imbalanced distribution of labels in engineering drawings, we employed Multi-label Stratification (Sechidis et al., 2011) to partition the dataset into training, validation, and test sets with a ratio of 70:15:15. In the preprocessing stage, no data augmentation was applied. Images were resized while preserving their original aspect ratios to match the default input specifications of each model: 1344×800 pixels for RF-DETR, Faster R-CNN and Qwen3-VL, 640×640 pixels for YOLOv12 and DocLayout-YOLO.

For Qwen3-VL, the dataset was converted from the standard COCO JSON format into a specialized Grounding prompt format (JSONL). The model is trained to generate JSON-formatted outputs containing a list of objects with their respective category labels and 2D bounding boxes. To maximize detection performance, we implemented a dual-query strategy:

General Queries: The model is prompted to identify and locate all layout elements simultaneously (e.g., "Identify and locate all drawing layout elements...").

Per-category Queries: The model focuses on localizing a specific target class (e.g., "Locate the Legend region...").

Training Strategy

Training was conducted using the AdamW optimizer coupled with a Cosine Learning Rate Scheduler. A warm-up period was implemented during the first 500 iterations to stabilize the initial gradients. All architectures were initialized with pre-trained weights mentioned above and fine-tuned according to their respective default loss configurations. During model training, we implemented selective parameter freezing strategies tailored to the specific architecture of each model. Specifically, for Qwen3-VL, we employed Parameter-Efficient Fine-Tuning via LoRA (Low-Rank Adaptation) to achieve efficient adaptation.

Table 2: Overall Performance Comparison of Evaluated Models on the FELD Dataset.

	Accuracy Metrics					Efficiency Metrics			
Model	**mAP$_{50:95}$**	**mAP$_{50}$**	**P**	**R**	**F1**	**Params**	**GFLOPs**	**Time**	**FPS**
Faster R-CNN	0.779	0.930	0.829	**0.924**	0.874	41.39M	187.0	88.0ms	11.4
RF-DETR (m)	**0.890**	**0.949**	**0.985**	0.800	0.889	33.70M	100.8	44.8ms	22.3
YOLOv12 (m)	0.755	0.871	0.871	0.793	0.830	20.06M	67.1	**23.1ms**	**43.3**
Qwen3-VL-8B	0.668	0.758	0.916	0.906	**0.911**	8.81B	–	6120.3ms	0.16
DocLayout-YOLO†	0.589	0.675	0.736	0.602	0.662	19.97M	68.0	30.0ms	33.4

† Pre-trained on DocLayNet and DocSynth datasets. **P**: Precision; **R**: Recall; **F1**: F1-score; **Time**: Inference time per unit.

Table 3: Detailed performance comparison ($mAP_{50:95}$) across specific layout categories.

Model	**Title**	**Rev.**	**Notes**	**Legend**	**H-Sec**	**V-Sec**	**GA**	**Elev.**	**Plan**
Faster R-CNN	0.866	0.924	0.666	0.896	0.734	0.747	0.820	0.722	0.775
RF-DETR (m)	**0.989**	**0.932**	**0.734**	0.928	**0.917**	0.845	**0.843**	**0.849**	**0.969**
YOLOv12 (m)	0.952	0.886	0.684	0.910	0.757	0.804	0.500	0.794	0.510
Qwen3-VL-8B	0.939	0.903	0.560	**1.000**	0.717	**0.856**	0.554	0.501	0.330
DocLayout-YOLO†	0.894	0.713	0.395	0.720	0.554	0.684	0.039	0.505	0.409

Title: Titleblock; **Rev.**: Revision table; **H-Sec**: Horizontal section; **V-Sec**: Vertical section; **GA**: GA section; **Elev.**: Elevation.

† Pre-trained on DocLayNet and DocSynth.

Evaluation Metrics

Accuracy Metrics

To assess the detection performance, we utilize standard COCO-style metrics and classification statistics. The core benchmarks are the mean Average Precision (mAP), specifically mAP_{50} and $mAP_{50:95}$. The precision (P), recall (R), and F1-score are calculated at a confidence threshold of 0.25 to evaluate the balance between prediction accuracy and coverage.

Model Complexity and Efficiency

Beyond accuracy, we quantify the computational requirements and real-time performance of each model. Model complexity is measured by the number of Parameters and Floating Point Operations (FLOPs). To evaluate practical deployment feasibility, we record models' Inference Time per unit (pre-processing and post-processing time excluded) and Frames Per Second (FPS) on T4 GPU.

Results and Analysis

This section presents a comprehensive analysis of the experimental results achieved by the candidate models on the FELD dataset. We first evaluate the overall performance of each architecture from the perspectives of both accuracy and computational efficiency, followed by an in-depth exploration of performance variations across different layout categories. Furthermore, an ablation study is conducted on DocLayout-YOLO to investigate how its specialized architectural innovations for general DLA, combined with pre-trained weights from general document datasets, influence the model's performance on AEC-specific drawings.

Overall Performance Analysis

Table 2 presents the overall performance of the candidate models on the FELD dataset. Throughout this paper, the best-performing results for each metric are highlighted in bold, while the second-best are indicated with an underline.

First, RF-DETR (m) achieves the SOTA performance in detection accuracy, reaching a peak mAP_{50} of 0.949, $mAP_{50:95}$ of 0.890, and a Precision of 0.985. This result demonstrates the superiority of the global attention mechanism within the Transformer architecture when processing large-scale drawings. In contrast, the traditional Faster R-CNN exhibits significant robustness, ranking second in mAP_{50} (0.930) and $mAP_{50:95}$ (0.779) while achieving the highest Recall (0.924). This validates the continued efficacy of the Region Proposal Network in identifying regular geometric layouts prevalent in AEC drawings. However, its inference speed is limited to 11.4 FPS. Regarding YOLOv12 (m), this model remains the most efficient architecture, achieving the highest frame rate of 43.3 FPS and a leading inference time of 23.1ms while its Precision (0.871) remains competitive.

The vision-language model Qwen3-VL-8B demonstrates a unique performance profile influenced by its specific evaluation methodology. During inference, the model performs per-category queries. Following the standard convention for VLM-based grounding tasks, the detection confidence for all outputs is fixed at 1.0. This could negatively impact metrics such as mAP, which rely on sorting detections to calculate precision-recall curves.

Despite this inherent disadvantage in ranking-based metrics, the model achieves the highest F1-score (0.911)

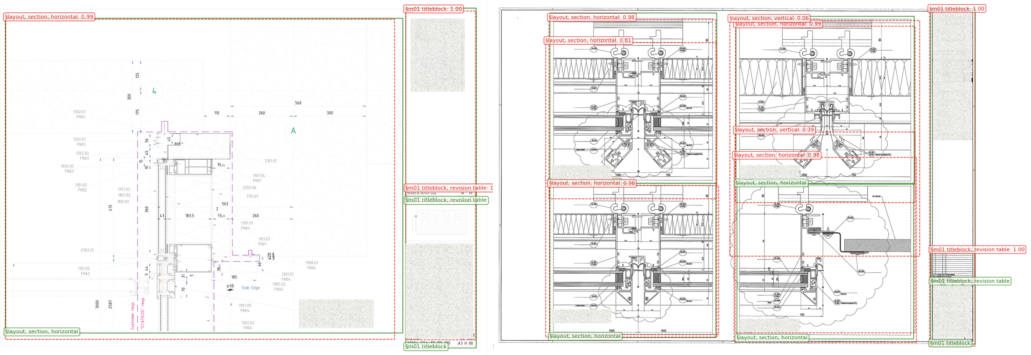


*Figure 3: Visualization of Qwen3-VL inference results. The **left** panel shows a successful layout detection; the **right** panel illustrates a failed case with redundant overlaps.*

and maintains strong performance with the second-highest Precision (0.916) and Recall (0.906). This suggests that the extensive multimodal knowledge and vision-language alignment embedded in Qwen3-VL provide superior semantic classification consistency. However, its localization precision ($mAP_{50:95}$ of 0.668) still lags behind specialized models. Moreover, its significant 8B parameter count leads to extreme latency, with an inference time of 6120.3ms per drawing, making it currently unsuitable for real-time industrial throughput.

Finally, DocLayout-YOLO, pre-trained on DocLayNet and DocSynth, fails to outperform general-purpose vision models on the FELD test set. This outcome indicates that architectural innovations and pre-trained features derived from general DLA do not effectively transfer to the AEC domain. The distinctive topological logic and the high-density, line-based graphical-symbolic logic of architectural drawings remain fundamentally different from the structural patterns found in general documents. A more granular analysis of this domain gap will be conducted in the Ablation Study section.

Category-specific Analysis

Table 3 provides a granular breakdown of the $mAP_{50:95}$ scores across nine layout categories, revealing distinct performance characteristics of the evaluated models.

First, RF-DETR (m) exhibits a dominant performance by achieving the highest precision across all categories. Its consistent lead validates the robustness of the Transformer-based architecture in managing diverse layout structures. Meanwhile, Faster R-CNN and YOLOv12 (m) alternate as the second-best performers depending on the category. Faster R-CNN maintains a competitive edge in regular geometric areas such as *GA section* (0.820), while YOLOv12 performs reliably in *Titleblock* (0.952) and *Legend* (0.910).

Second, DocLayout-YOLO† shows highly polarized results across different classes. While it achieves respectable performance in structured regions like *Titleblock* (0.894) and *Legend* (0.720), its accuracy collapses in *Notes* (0.395) and *GA section* (0.039). This disparity suggests that while its DLA-oriented architecture can recognize common document-like blocks, it struggles with the specialized graphical density and topological complexity inherent in AEC-specific sections.

Despite mAP calculation disadvantages, Qwen3-VL-8B achieved a perfect 1.000 $mAP_{50:95}$ for *Legends* and led in *Vertical Section* (0.856). While highly competitive in *Titleblocks* (0.939), its performance in *Notes* (0.560) was hampered by over-generalization, frequently misidentifying miscellaneous text as notes. This underscores the model's robust pattern recognition alongside current limitations in fine-grained textual classification.

Finally, a cross-model observation reveals that *GA section* and *Plan* generally yield the lowest scores among all categories. This trend is primarily attributed to the limited sample size for these specific classes in the current dataset, which restricts the models' ability to learn sufficient discriminative features during the training phase.

Ablation Study on DocLayout-YOLO

To further investigate the impact of architectural innovations and pre-training weights of DocLayout-YOLO on model performance within the FELD dataset, we conducted a comprehensive ablation study.

DocLayout-YOLO is a variant of the YOLOv10m baseline. Its primary structural modification involves replacing the standard C2f and C2fCIB modules in layers 4, 6, and 8 with specialized G2L_CRM modules designed for DLA. Additionally, the standard DocLayout-YOLO leverages large-scale pre-training on general document layout datasets, specifically DocLayNet and DocSynth. We designed this ablation study to specifically examine how these DLA-centric modifications and pre-training weights respond to the unique graphical nature of AEC drawings.

We evaluated three specific configurations to isolate these variables:

1. YOLOv10m (COCO pre-trained): The baseline general-purpose object detector with COCO weights.
2. DocLayout-YOLO (COCO + Random Init): Incorporates the G2L_CRM architectural changes but utilizes COCO weights for common layers and random initialization for the new modules, effectively testing the architecture without DLA-specific pre-training.
3. DocLayout-YOLO (DocLayNet&DocSynth pre-

Table 4: Ablation Study on Model Architecture and Pre-training Weights of DocLayout-YOLO

Model / Configuration	$\mathbf{mAP}_{50:95}$	$\mathbf{mAP}_{50}$	P	R	F1	Params
DocLayout-YOLO (DocLayNet&DocSynth Pre-trained)	0.589	0.675	0.736	0.602	0.662	19.97M
DocLayout-YOLO (COCO + Random Init)	<u>0.727</u>	<u>0.854</u>	<u>0.921</u>	<u>0.695</u>	<u>0.792</u>	19.97M
YOLOv10m (COCO pre-trained)	**0.772**	**0.900**	**0.954**	**0.826**	**0.885**	16.46M

trained): The full model incorporating both architectural changes and large-scale document-domain pre-training.

The results presented in Table 4 revealed a surprising trend: both the architectural innovations and the document-specific pre-training weights of DocLayout-YOLO resulted in a performance degradation on the AEC dataset compared to the baseline.

Moving from the YOLOv10m baseline to the DocLayout-YOLO architecture (COCO + Random Init) led to a decrease in $mAP_{50:95}$ from 0.772 to 0.727. This suggests that while G2L_CRM modules may excel at hierarchical text layouts, they are less effective at capturing the high-density, purely graphical symbolic logic of architectural drawings. While the most significant drop occurred when applying the DocLayNet pre-trained weights, which caused the $mAP_{50:95}$ to plummet to 0.589.

This confirms a strong "domain interference" effect, indicating that the structural priors learned from general documents like newspapers and academic papers are fundamentally incompatible with AEC drawings. The features optimized for general documents act as a negative prior, preventing the model from adapting to the complex graphical logic of architectural drawings.

Conclusion and future work

This research identifies and addresses a critical "missing link" in the automated interpretation of AEC engineering drawings: layout detection. We systematically evaluated the performance of five deep learning architectures on this task, revealing how different modeling paradigms navigate the unique structural constraints of AEC layouts.

Our experimental results demonstrate that RF-DETR (m) is the most effective technical pathway for this task, achieving a superior mAP_{50}. The success of RF-DETR highlights the necessity of global attention mechanisms in capturing the intricate spatial dependencies inherent in large-format engineering drawings. Notably, Qwen3-VL-8B also performed excellently, achieving the highest F1-score, which demonstrates the significant potential of VLMs for interpreting AEC engineering drawing. Conversely, the benchmarking of DocLayout-YOLO revealed a significant "domain gap". Our ablation study specifically proved that architectural priors and pre-trained weights optimized for text-centric documents can act as a negative prior, hindering the model's ability to adapt to the high-density graphical-symbolic logic of AEC layouts.

Despite these advancements, the FELD dataset's limited scale and facade-specific focus restrict broader generalisability. To address this, future work will enhance robustness via domain-specific augmentation and evaluate cross-domain generalisation through a dual-metric framework: quantifying "domain shift" via zero-shot testing and measuring "adaptation velocity" through few-shot transfer learning across structural and MEP sectors.

Furthermore, the inherent high-resolution nature of AEC drawings poses a significant challenge for standard deep learning pipelines; our current methodology simply relies on standard resizing, which inevitably leads to the loss of fine-grained textual features and thin geometric lines. This factor likely constrained the localization precision of RF-DETR and prevented Qwen3-VL from fully leveraging its native capacity for dynamic, native-resolution visual inputs.

Future research efforts will focus on three primary dimensions to achieve more holistic drawing understanding. First, we aim to expand the FELD dataset into a multi-domain AEC layout benchmark to enhance cross-domain generalization and address current class imbalances. Second, in this work, ADIRO is used primarily as a layout-centric schema to support consistent dataset construction and annotation, rather than as a full reasoning layer; in future work, we plan to integrate ADIRO more deeply into the pipeline by leveraging its structure in model design and training, and by using it as a foundation for ontology-based reasoning, knowledge graph construction, and potential integration with BIM models. Finally, we will leverage recognized layout boundaries as "middleware" to guide Main Drawing Content Layer extraction, implementing semantic-aware tiling to ensure that downstream symbol recognition and geometric analysis are performed within a coherent logical context.

References

Bai, S., Cai, Y., Chen, R., et al. (2025). Qwen3-VL technical report. arXiv preprint arXiv:2511.21631.

Carrara, A., Nousias, S., and Borrmann, A. (2025). Content-based classification of construction drawings. In Proceedings of the 32nd EG-ICE International Workshop on Intelligent Computing in Engineering, Glasgow, UK.

Chen, Y., Liu, R., Zheng, J., et al. (2025). Graph-based document structure analysis. In The Thirteenth International Conference on Learning Representations (ICLR).

Heras, L.-P., Ahmed, S., Liwicki, M., et al. (2014). Statistical segmentation and structural recognition for floor plan interpretation. International Journal on Document Analysis and Recognition, 17(3):221–237.

Huang, Y., Lv, T., Cui, L., et al. (2022). LayoutLMv3: Pre-training for document AI with unified text and image masking. In Proceedings of the 30th ACM International Conference on Multimedia, pages 4083–4091. ACM.

Jamieson, L. (2024). Deep Learning for Digitising Complex Engineering Drawings. Phd thesis, Robert Gordon University.

Jamieson, L., Moreno-Garcia, C. F., and Elyan, E. (2025). Towards fully automated processing and analysis of construction diagrams: AI-powered symbol detection. International Journal on Document Analysis and Recognition, 28(1):71–84.

Kashevnik, A., Shilov, N., Teslya, N., et al. (2023). An approach to engineering drawing organization: Title block detection and processing. IEEE Access, 11:15143–15155.

Khan, M. T., Chen, L., Ng, Y. H., et al. (2024). Fine-tuning vision-language model for automated engineering drawing information extraction. arXiv preprint arXiv:2411.03707.

Khan, M. T., Yong, Z., Chen, L., et al. (2026). A multi-stage hybrid framework for automated interpretation of multi-view engineering drawings using vision language model. In Proceedings of the 13th International Conference on Industrial Engineering and Applications (ICIEA).

Kim, G., Hong, T., Yim, M., et al. (2022). OCR-free document understanding transformer. In Computer Vision – ECCV 2022, volume 13688 of Lecture Notes in Computer Science, pages 498–517. Springer.

Lombardi, A., Duan, L., Elnagar, A., et al. (2025). Title block detection and information extraction for enhanced building drawings search. In 2025 European Conference on Computing in Construction (EC3).

Namboodiri, A. M. and Jain, A. K. (2007). Document Structure and Layout Analysis, pages 29–48. Springer London.

Pfitzmann, B., Auer, C., Dolfi, M., et al. (2022). DocLayNet: A large human-annotated dataset for document-layout segmentation. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 3743–3751. ACM.

Ren, S., He, K., Girshick, R., et al. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6):1137–1149.

Robinson, I., Robicheaux, P., Popov, M., et al. (2026). RF-DETR: Neural architecture search for real-time detection transformers. In The Fourteenth International Conference on Learning Representations (ICLR).

Schönfelder, P. and König, M. (2025). Ontology-based reasoning in automatic floor plan analysis. Advanced Engineering Informatics, 68:103761.

Sechidis, K., Tsoumakas, G., and Vlahavas, I. (2011). On the stratification of multi-label data. In Machine Learning and Knowledge Discovery in Databases (ECML PKDD 2011), volume 6913 of Lecture Notes in Computer Science, pages 145–158. Springer.

Tian, Y., Ye, Q., and Doermann, D. (2025). Yolov12: Attention-centric real-time object detectors. In Advances in Neural Information Processing Systems 38 (NeurIPS).

Wang, Z., Xu, Y., Cui, L., et al. (2021). LayoutReader: Pre-training of text and layout for reading order detection. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 4735–4744. Association for Computational Linguistics.

Xu, Y., Li, M., Cui, L., et al. (2020). LayoutLM: Pre-training of text and layout for document image understanding. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1192–1200. ACM.

Xu, Y., Xu, Y., Lv, T., et al. (2021). LayoutLMv2: Multimodal pre-training for visually-rich document understanding. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 2579–2591. Association for Computational Linguistics.

Zhang, W., Joseph, J., Yin, Y., et al. (2023). Component segmentation of engineering drawings using graph convolutional networks. Computers in Industry, 147:103885.

Zhao, Y., Huang, J., Hu, J., et al. (2024a). SWIFT: A scalable lightweight infrastructure for fine-tuning. arXiv preprint arXiv:2408.05517.

Zhao, Z., Kang, H., Wang, B., et al. (2024b). DocLayout-YOLO: Enhancing document layout analysis through diverse synthetic data and global-to-local adaptive perception. arXiv preprint arXiv:2410.12628.

Zhong, X., Tang, J., and Yepes, A. J. (2019). PubLayNet: Largest dataset ever for document layout analysis. In 2019 International Conference on Document Analysis and Recognition (ICDAR), pages 1015–1022.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

IDS-Constrained Text-to-Graph Querying for IFC Models

Ivan Smirnov[1], Tanya Bloch[2], Fulvio Re Cecconi[1],
Sebastiano Maltese[3], Lavinia Chiara Tagliabue[4], Silvia Meschini[4] and Shabtai Isaac[5]
[1]Politecnico di Milano, Milan, Italy
[2]Technion – Israel Institute of Technology, Haifa, Israel
[3]University of Applied Sciences and Arts of Southern Switzerland, Mendrisio, Switzerland
[4]University of Turin, Turin, Italy
[5]Ben-Gurion University of the Negev, Beersheba, Israel

Abstract

Accessing Building Information Model (BIM) data encoded in Industry Foundation Classes (IFC) remains difficult for non-expert stakeholders due to heterogeneous property sets and nested relationships. This paper presents a constrained text-to-graph querying framework that transforms IFC data into a labeled property graph, queried via Cypher, while project-specific Information Delivery Specification (IDS) requirements define admissible entity labels, relationships, and property keys. A regular-expression grammar restricts Large Language Model (LLM) generation to executable, schema-compliant queries. Evaluation shows grammar constraints eliminate schema hallucinations, achieving 100% Semantic Compliance and improved execution accuracy, with competitive performance from smaller open-weights models.

Introduction

BIM workflows increasingly presume that building information can be accessed on demand to support coordination, compliance checking, and decision-making. In practice, however, obtaining a first correct and verifiable answer from an IFC model remains costly (Hellin et al., 2025). This cost extends beyond computation to include the time and expertise required to translate an information need into a formal query, execute it on an IFC-derived representation, and validate the results against both the model content and project-specific modelling conventions. This initial problem typically involves multiple iterations of identifying relevant entities and relations, selecting appropriate properties and value types, and refining queries until defensible results are produced. The burden is most evident for information needs that exceed direct attribute lookups and require combinations of property constraints and topological reasoning. Common quality and compliance checks, such as isolating spatially contained elements or verifying classification and completeness requirements, remain technically demanding, as they require detailed knowledge of IFC structures and querying mechanisms. Prior research consistently identifies an intent-to-query gap in BIM information access and proposes abstractions and natural-language interfaces to mitigate direct interaction with raw IFC schemas (Mazairac and Beetz, 2013; Yin et al., 2023a; Hellin et al., 2025). However, query verbosity and representational overhead remain pragmatic obstacles, particularly where relevant information is accessible only through long navigation paths and reified relationship structures. IFC, as an open and vendor-neutral standard (International Organization for Standardization, 2024), encodes semantics indirectly, leading to complex multi-hop queries across relationship entities and nested value representations. Linked Data approaches, including ifcOWL and higher-level ontologies such as the Building Topology Ontology (BOT) and BimSPARQL, improve expressiveness and topology-centric querying (Beetz et al., 2009; Rasmussen et al., 2020; Zhang et al., 2018). Nevertheless, semantic variance across projects, arising from custom property sets, naming conventions, heterogeneous datatypes, and tool-specific export practices, limits query robustness and reuse (Yu et al., 2023). IDS addresses this by formalising project-specific information requirements and constraints in a machine-interpretable manner (buildingSMART International, 2024).

Recent advances in LLMs demonstrate strong capabilities in natural-language-to-structure translation (Jiang et al., 2023), yet their probabilistic nature poses reliability risks when applied to deterministic, schema-bound engineering data. Hallucinations and invalid query generation remain critical concerns (Huang et al., 2023). Building on recent BIM research that combines LLMs with graph-based grounding (Iranmanesh et al., 2025), this paper proposes a reliability-oriented framework that translates natural-language questions into constrained graph queries over IFC-derived representations, informed by IDS constraints. A key objective is to enable the framework to run privately on local machines, preserving data confidentiality and substantially reducing the energy and operational costs associated with reliance on commercial, cloud-based LLM services. By combining graph abstractions, explicit project constraints, and constrained decoding, the approach aims to improve the correctness and verifiability of results for topology- and constraint-dependent BIM checks.

Background and Related Work

Querying BIM models

BIM models integrate geometry, spatial relations, material properties, quantities, and semantics in one digital structure. In practice, this richness creates large datasets that are difficult to interrogate without specialized tools. Most exchanges rely on IFC (currently version 4) (build-

ingSMART International, 2024), which provides interoperability but also substantial complexity due to broad entity vocabularies, heterogeneous property sets, and deeply nested relations (Borrmann et al., 2018). As repositories grow, efficient BIM search becomes increasingly important (Khademi and Behan, 2017). Early efforts introduced domain-specific query languages such as BIMQL (Mazairac and Beetz, 2013) and QL4BIM (Daum and Borrmann, 2014). BIMQL focuses on class- and attribute-based filtering, whereas QL4BIM emphasizes spatial and topological queries. Both approaches improve expressiveness but still require users to understand formal query syntax and IFC internals, including indirect relationship encoding. Database-based approaches have also been explored. SQL is mature and efficient, but less suited to IFC-style hierarchical and multi-hop relations. SPARQL over ifcOWL maps IFC into RDF triples and supports linked-data querying (Zhang et al., 2018), yet this representation is often verbose and large (Pauwels et al., 2017). As an alternative, labeled property graphs (LPGs) represent relations more directly and support efficient traversal in systems such as Neo4j (Neo4j, 2026), queried with Cypher (Zhu et al., 2025). The RDF–LPG choice remains use-case dependent, but LPGs are often preferred for storage and query efficiency (Zhu et al., 2023). Prior work shows the value of LPG-based BIM querying. Ismail et al. (2017) demonstrated IFC-to-LPG transformation for graph analyses (e.g. emergency routes and model comparison), while later studies reported benefits for change tracking (Esser et al., 2022), semantic enrichment (Wang et al., 2022), and code compliance (Bloch et al., 2023). With recent LLM advances, graph representations are increasingly relevant as a grounding layer for natural-language BIM interaction and hallucination reduction.

LLM based BIM query

Natural-language BIM querying lowers the barrier for non-expert users by avoiding direct interaction with formal query syntax. However, LLMs alone do not reliably query IFC data: without grounding, they hallucinate entities, properties, and relations. Du et al. (2026) identify three recurring causes: long prompt context, limited global spatial understanding, and cumulative errors in multi-agent pipelines. The spatial reasoning gap is particularly critical for BIM querying. Accordingly, prior work introduces external constraints. Ontology-based approaches translate natural language into SPARQL over RDF/OWL models, as in Yin et al. (2023b). Tool-augmented pipelines instead couple LLM planning with deterministic BIM operations; for example, Hellin et al. (2025) use a ReAct-style workflow with Python tools over IFC, reporting good results for direct queries but weaker performance for indirect and relational questions. As also discussed by Du et al. (2026), graph representations can improve structural grounding for spatial and topological reasoning. Retrieval-augmented generation (RAG) provides another grounding strategy by injecting retrieved context at inference time (Gao et al., 2023). In BIM, this can supply IFC documentation and project-specific metadata (Iranmanesh et al., 2025). Yet text-centric RAG does not fully capture relational model structure. Graph-RAG improves this by retrieving connected nodes and edges, but still underperforms on complex multi-hop queries with multiple entity references and ambiguity (Iranmanesh et al., 2025). Formal project specifications offer a complementary constraint layer. IDS (buildingSMART International, 2024) defines expected IFC entities, properties, value domains, and vocabularies in machine-readable XML, primarily for automated compliance checking. In query generation, IDS can constrain allowable query patterns across IFC-, RDF-, or LPG-based representations, grounding LLM output in explicit project rules. This study therefore examines whether project-specific IDS constraints reduce hallucinations in LLM-mediated BIM querying, a direction that remains largely unexplored.

Appropriate use cases and study scope

From a theoretical perspective, model checking activities can be classified into three categories (Solihin and Eastman, 2015; Hjelseth and Nisbet, 2016). Simple checks involve the verification of the existence of specific property sets, the presence of assigned values, and the correctness of those values. Elaborated checks address design, spatial, or technical characteristics that are not explicitly stored as IFC properties but can be derived through relatively straightforward algorithms or mathematical relationships. A representative example is the verification of regulatory requirements governing the minimum ratio between window surface area and room floor area, as prescribed by Italian building codes. Complex checks, by contrast, focus on performance-based assessments and require advanced simulations, such as those used for evaluating building energy performance.

For simple checks, established rule-based approaches (e.g. IDS frameworks and commercial tools like Solibri) are likely to remain more efficient and reliable than LLM-based solutions. In this context, the primary advantage of LLMs lies in their usability, as they lower the technical barrier for non-specialist users. The suitability of LLMs for elaborated checks is clearer, however, concerns remain regarding computational accuracy and the risk of hallucinated outputs (Huang et al., 2023). In these cases, LLMs may be a useful interface between the complexity of rule-based checking software and end users, providing the ability to perform elaborated checks without needing to learn the software's query grammar. With respect to complex checks, neither traditional rule-based tools nor LLM-based approaches currently offer adequate capabilities.

Beyond model checking, greater value may be found in leveraging LLMs as intuitive interfaces for querying BIM data (Yin et al., 2023b). Their accessibility enables meaningful interaction with BIM models by non-technical stakeholders, including building occupants, maintenance personnel, municipal officers involved in digital permit-

ting, and senior decision-makers within organizations. This work focuses on elaborated checks, where the key challenge is reliably translating natural-language intent into verifiable, executable queries over IFC-derived data.

Research Questions

To structure the study, three research questions are formulated, each addressing a key design choice of the proposed pipeline:

RQ1 (text-to-query over direct QA): Can LLM-mediated text-to-graph querying produce correct results for IFC-based information needs where Direct QA over serialized data fails?

RQ2 (graph representation): How does representing IFC as a graph affect the correctness and robustness of answers generated from queries over the model?

RQ3 (IDS as a constraint layer): Does integrating IDS improve the semantic correctness of LLM-generated Cypher queries over IFC-derived graphs?

Together, these questions assess whether a constrained graph query workflow can provide predictable, auditable access to BIM information for non-expert stakeholders, while remaining practical for real project models.

Proposed Method

System overview

A Constrained Natural Language to Cypher (NL2Cypher) framework is proposed to ground LLMs in project-specific constraints defined by the IDS. An IFC file is transformed into a Neo4j (Neo4j, 2026) LPG and an IDS file is compiled into a formal grammar. This grammar restricts the LLM's decoding process, ensuring that generated Cypher queries are syntactically valid and semantically compliant with the project schema.

The architecture consists of three primary layers:

a) The data ingestion layer transforms the IFC model into a flattened property graph and scans the model for available vocabulary.

b) The constraint layer parses the IDS to extract enforced constraints (entities, properties, enums) and merges them with the model vocabulary to build a strictly typed schema.

c) The generation layer uses constrained decoding to guide the LLM, masking invalid tokens at each generation step based on a regular expression derived from the schema.

IFC-to-graph mapping

To enable efficient graph transit, the system simplifies the complex IFC schema into a LPG. The mapping process involves three transformation rules. First, every `IfcProduct` becomes a graph node. The node label corresponds to the IFC class (e.g., `IfcWall`). Then, the relationships hosted in `IfcPropertySet` entities linked via `IfcRelDefinesByProperties` are flattened and embedded directly onto the entity node as key-value pairs. Finally, the complex relationship entities (e.g., `IfcRelAggregates`, `IfcRelContainedInSpatialStructure`) are converted into directed edges. The relationships most commonly required for property- and topology-based queries are prioritized: CONTAINS (from `IfcRelContainedInSpatialStructure`), DECOMPOSES (from `IfcRelAggregates`), and HAS_MATERIAL (from `IfcRelAssociatesMaterial`). Relationship types that serve different querying needs—such as `IfcRelVoidsElement` and `IfcRelFillsElement` (opening and infill associations), `IfcRelConnectsPathElements` (wall connectivity), and `IfcRelSpaceBoundary` (thermal envelope geometry)—are not mapped in the current implementation. The mapping is extensible: supporting additional relationship types requires adding the corresponding extraction logic and registering new edge labels in the grammar.

IDS parsing and constraint compilation

The core innovation of this method is the translation of the IDS "semantic contract" into a generative grammar. Solely relying on the IDS can be insufficient if the model contains valid data not covered by the specification. At the same time relying only on the model allows the LLM to hallucinate on dirty data. Therefore a Hybrid schema merging two sources is constructed.

Enforced Vocabulary from IDS defines strict constraints. The system parses the IDS XML structure, extracting entities from `applicability` clauses and required properties from `requirements` clauses. It captures domain-specific constraints defined via XML Schema Definition (XSD) restrictions (e.g., `xs:enumeration`), ensuring that sensitive properties (e.g., `FireRating`) are strictly limited to their defined enums.

Available Vocabulary from IFC is extracted by scanning the model file for existing labels and properties. Conflicts are resolved by giving precedence to the IDS. Properties are categorized into two tiers: **strict** (constrained to the IDS-defined enums) and **open** (inferred from the model with data types like Boolean, Numeric or String). This ensures regulatory compliance without rendering valid non-standard model data unqueryable.

This hybrid vocabulary is then compiled into a Regular Expression (Regex) that guides the LLM's decoding process via Outlines library (dottxt-ai, 2026).

Query generation

Instead of fine-tuning the model or relying on "soft" prompt engineering, constrained decoding is utilized. The hybrid schema is translated into a regex pattern that defines the set of all possible valid Cypher queries for that specific building. For a schema with entities E and properties P, a regular expression $r = \texttt{build_cypher_regex}(E, P)$ is compiled to act as a grammar for valid Cypher. r is assembled from (i) an entity alternation group E_{alt} (e.g., `(IfcDoor|IfcWall)`), (ii) a property alternation group P_{alt} (e.g., `(FireRating|Height|Name)`), and (iii) a small set of admissible literals and operators. Property access is restricted to the bound match vari-

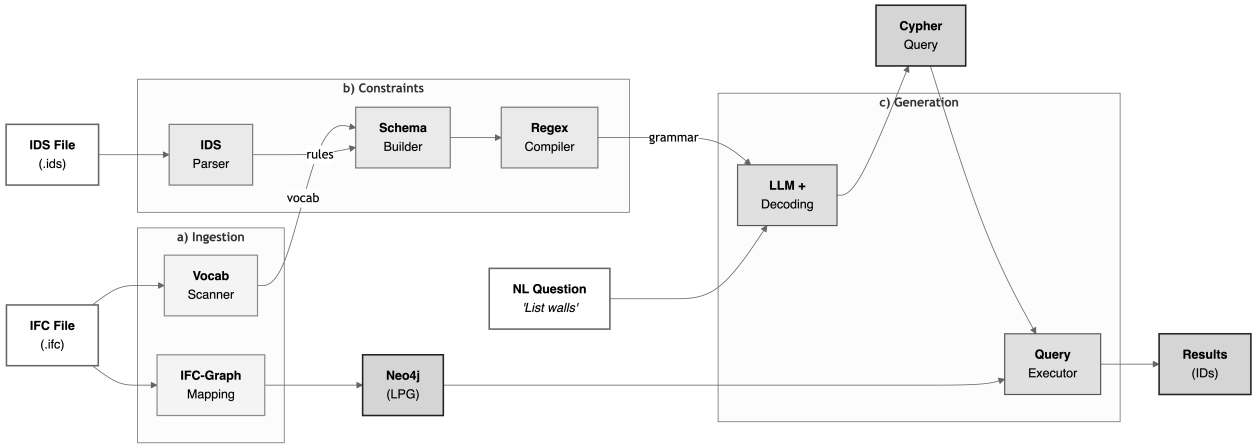


Figure 1: System overview of the NL2Cypher framework. The three layers correspond to (a) data ingestion, (b) constraint compilation, and (c) grammar-constrained query generation.

able (e.g., v.`Name`) with optional type conversions (e.g., `toInteger`(v.`FireRating`)). The grammar supports semantic string matching (e.g., $CONTAINS$) and null-safety checks ($ISNOTNULL$) which are essential for handling natural language ambiguity. The `WHERE` clause is optional and bounded to a fixed maximum number of boolean conditions to avoid overly long or incomplete queries.

$$
\begin{aligned}
r &= r_{\text{MATCH}}\, r_{\text{REL}}^{?}\, r_{\text{WHERE}}^{?}\, r_{\text{RETURN}}, \\
r_{\text{MATCH}} &:= \texttt{MATCH}\,(v : E_{\text{alt}}), \\
r_{\text{REL}} &:= -[: R_{\text{alt}}] \rightarrow (v_2 : E_{\text{alt}}), \\
r_{\text{WHERE}} &:= \texttt{WHERE}\; c_1\, \big((\texttt{AND}|\texttt{OR})\, c_2\big)^{\leq 4}, \\
c_i &:= \big(f^{?}(v.P_{\text{alt}})\ \text{op}\ \text{val}\big) \mid \big(v.P_{\text{alt}}\ \texttt{CONTAINS}\ \text{str}\big), \\
f &\in \{\texttt{toInteger}, \texttt{toFloat}, \texttt{toString}\}.
\end{aligned}
\tag{1}
$$

Note that R_{alt} represents the relationship alternation group (e.g., (CONTAINS|HAS_MATERIAL)). Furthermore, the return clause r_{RETURN} is constrained to return the node variable v. This is strictly enforced to guarantee that the execution engine retrieves the element's `GlobalId` for evaluation against the Gold Standard, preventing the generation of queries that return unidentifiable scalar values.

During inference, the constrained decoder only allows tokens that keep the partial output compatible with r (invalid next tokens are masked by setting their logits to $-\infty$). As a result, generated queries are (i) executable (high SVR - Syntactic Validity Rate) and (ii) limited to entity labels and property names present in the hybrid schema (high SCR - Schema Coverage/Compliance Rate).

Experimental Design

Baselines and datasets for demonstration and validation

To evaluate the efficacy of IDS grounding and graph-based querying, four experimental settings are compared in a 2×2 matrix of Direct Question Answering (QA) versus Text-to-Cypher, crossed with baseline versus IDS-constrained configurations (Table 1.)

Table 1: Experiment settings.

	Direct QA	Text-to-Cypher
Baseline	**S1** Naive QA **Input:** IFC text dump **Output:** JSON `GlobalIds`	**S3** Soft Cypher **Input:** graph schema **Output:** Cypher → IDs
With IDS	**S2** Grounded QA **Input:** IFC + IDS rules **Output:** JSON `GlobalIds`	**S4** Strict Cypher (proposed) **Input:** hybrid schema (IDS + model) **Output:** Cypher → IDs

Settings 1&2 simulate a RAG approach where the model reasons over raw text. To ensure comparability, the LLM is instructed to output a JSON list of `GlobalId` strings rather than the natural language. Settings 3&4 generate executable Cypher code. Setting 3 represents the current state-of-the-art in Text-to-Cypher (prompting), while Setting 4 represents our proposed method.

Dataset and Preprocessing

A sample project from Autodesk (Autodesk, 2026), exported as an IFC model, and a corresponding IDS requirement file are utilized. For the Direct QA settings (1, 2), the IFC model is serialized into a condensed JSON format. For the Graph Settings (3, 4), the model is loaded into a Neo4j (Neo4j, 2026) instance using the mapping de-

scribed above.

The IDS file defines seven specifications that constrain naming conventions, required properties, and material assignments for the principal IFC entities (Table 2). Properties such as `FireRating` are limited to enumerated values (0, 30, 60, 90 minutes), while `LoadBearing` and `IsExternal` are enforced as Boolean indicators. Material constraints require both presence (for structural elements) and specific content (concrete for load-bearing columns). These specifications form the enforced vocabulary layer in the hybrid schema.

Table 2: Summary of IDS specifications used in the evaluation.

Applicable Entity	Required Properties	Value Constraints
`IfcBuildingStorey`	`Name`	Level naming pattern
`IfcDoor`	`Name`, `IsExternal`, `FireRating`	`FireRating` ∈ {0, 30, 60, 90}
`IfcSpace`	`NetFloorArea`, `Reference`	Controlled room-type vocabulary
`IfcWall`	`LoadBearing`, `IsExternal`	Boolean
`IfcWindow`	`Name`	Component naming pattern
Structural†	`Material`	Presence required
`IfcColumn`*	`Material`	Must contain ''Concrete''

*Load-bearing only. †`IfcStair`, `IfcBeam`, `IfcSlab`, `IfcPile`, `IfcRamp`, `IfcColumn`.

Task design and use cases

The evaluation consists of 16 natural-language queries, categorized into three levels of increasing complexity:

- **Level 1 – Class retrieval** (6 queries, Easy): direct entity lookups that test the ability to identify correct IFC classes. Example: *''List all doors''* → `MATCH (n:IfcDoor) RETURN n`.
- **Level 2 – Property filtering** (5 queries, Medium/Hard): queries that add attribute constraints, Boolean checks, numeric comparisons, or null-safety tests. Examples: *''List all external walls''* → `MATCH (n:IfcWall) WHERE n.IsExternal = true RETURN n`; *''List all doors that have a fire rating defined''* → `MATCH (n:IfcDoor) WHERE n.FireRating IS NOT NULL RETURN n`.
- **Level 3 – Topology and relationships** (5 queries, Medium/Hard): multi-hop graph traversals combining relationship patterns with property constraints. Example: *''List spaces on Level 2 larger than 12 m^2''* → `MATCH (s:IfcBuildingStorey)-[:DECOMPOSES]->(n:IfcSpace) WHERE s.Name CONTAINS '02' AND n.GrossFloorArea > 12 RETURN n`.

Each query is paired with a gold-standard Cypher query whose result set defines the ground-truth `GlobalId`s used for evaluation.

Evaluation

Three metrics are employed to evaluate performance. To enable a common comparison across all settings, the primary metric is the accuracy of the returned `GlobalId`s:

- **Execution Accuracy (EA):** Jaccard similarity between the set of `GlobalId`s returned by the system and the ground-truth set.

$$EA = \frac{|R_{\text{gen}} \cap R_{\text{gold}}|}{|R_{\text{gen}} \cup R_{\text{gold}}|} \tag{2}$$

  For Direct QA (Settings 1&2): ground truth is parsed from the JSON output.
  For Text-to-Cypher (Settings 3&4): ground truth is obtained by executing the generated Cypher query against the database.
- **Syntactic Validity Rate (SVR):** for Direct QA, the percentage of responses that are valid, parseable JSON; for Cypher, the percentage of generated queries that execute without syntax errors.
- **Semantic Compliance Rate (SCR):** the percentage of generated schema elements (entity labels, property names) that exist in the hybrid schema (reported for the Cypher-generating settings).

Results and Discussion

Comparative performance

Phase 1: Cloud baselines and data access

To establish a performance ceiling and verify the necessity of structured data access (RQ1), the framework was first evaluated using state-of-the-art cloud models: Google Gemini 2.5 Flash and Gemini 3.0 Flash Preview. These models were tested on the unconstrained settings: Naive QA (S1), Grounded QA (S2), and Soft Cypher (S3).

Table 3 presents the results of the cloud-based evaluation. Both Direct QA settings (S1 and S2) achieved near-zero EA across all models and runs (Gemini 2.5 Flash: 0.09 ± 0.08; Gemini 3.0 Flash: 0.05 ± 0.06). Despite the advanced reasoning capabilities of the Gemini models, the system failed to reliably identify correct elements. The size of the IFC model chosen for tests exceeds 1M tokens both as a .ifc and as a .json dump, so the truncation of the input plays an important role in such a low performance, too. This confirms that without structured access to the underlying model data, LLMs cannot reliably retrieve specific BIM elements. The unique nature of IFC identifiers (22-character encoded strings) prevents probabilistic guessing, and schema knowledge alone (S2) proved insufficient for identifying specific instances.

In contrast, the graph-based approach (S3: Soft Cypher) demonstrated significant potential even without grammar constraints. Gemini 2.5 Flash achieved a mean EA of 0.67 ± 0.03 across three runs, successfully translating natural language into executable Cypher queries for the majority of test cases.

This disparity supports the hypothesis that the ''Text-to-Graph'' abstraction is superior to ''Direct QA'' for engineering tasks. The LLM excels at logic generation (constructing the query pattern) but struggles with direct information retrieval when denied an explicit index.

While Setting 3 (Soft Cypher) performed well, it was not flawless. Gemini 2.5 Flash achieved a mean SVR of 0.83 ± 0.07, meaning that on average roughly 17% of generated

Table 3: Phase 1 Results: Cloud-based Models (Unconstrained). Values reported as mean ± std over n independent runs.

Model	Setting	SVR	SCR	EA
Gemini 2.5 Flash	S1 (Naive)	0.54 ± 0.14	–	0.09 ± 0.08
($n = 3$)	S2 (Grounded)	0.60 ± 0.10	–	0.08 ± 0.07
	S3 (Soft Cypher)	0.83 ± 0.07	0.99 ± 0.01	0.67 ± 0.03
Gemini 3.0 Flash (Preview) ($n = 3$)	S1 (Naive)	0.44 ± 0.17	–	0.05 ± 0.06
	S2 (Grounded)	0.44 ± 0.06	–	0.02 ± 0.04
	S3 (Soft Cypher)	0.67 ± 0.03	0.99 ± 0.02	0.62 ± 0.03

queries were invalid Cypher and could not be executed. Detailed analysis of the error logs revealed two primary failure modes in the ''Soft'' setting:

1. **Hallucinated properties:** the model occasionally queried properties that did not exist in the graph (SCR < 1.0), despite the schema being provided in the system prompt.
2. **Complexity degradation:** performance dropped sharply as query complexity increased. For ''Hard'' queries (involving relationships and multi-hop traversals), the mean EA for Gemini 2.5 Flash dropped from 100% (Easy) to 22% (Hard).

These limitations in the unconstrained "Soft" approach (specifically a non-perfect SVR) motivate the need for the grammar-constrained generation (S4) explored in the subsequent phase of this study.

Phase 2: Efficacy of grammar constraints

To evaluate the proposed grammar-constrained generation (S4), Qwen2.5-Coder was utilized as a state-of-the-art open-weights model family. Unlike the prompt-based approaches in Phase 1, Setting 4 requires direct intervention in the decoding process to enforce adherence to the hybrid schema; this requirement motivates the use of local, open-weights models rather than API-based cloud services.
Table 4 summarizes the performance impact of the constraints. S3 results are reported as mean ± std over multiple independent runs ($n = 5$ for 7B, $n = 9$ for 3B); S4 results are deterministic (greedy decoding with grammar constraints) and therefore reported from a single run.

Table 4: Phase 2 Results: Open-weights models (constrained vs. unconstrained). S3 values reported as mean ± std over n runs; S4 is deterministic (single run).

Model	Setting	SVR	SCR	EA
Qwen2.5-Coder-7B Instruct	S3 ($n = 5$)	0.88 ± 0.04	0.87 ± 0.01	0.34 ± 0.03
	S4 (Strict)	**1.00**	**1.00**	**0.63**
Qwen2.5-Coder-3B Instruct	S3 ($n = 9$)	0.95 ± 0.03	0.94	0.62 ± 0.02
	S4 (Strict)	**1.00**	**1.00**	**0.64**

The effect of grammar constraints varied markedly between the two models. For the 7B model, which achieved only $EA = 0.34 \pm 0.03$ under unconstrained prompting (S3), grammar constraints nearly doubled accuracy to $EA = 0.63$, while also raising SCR from 0.87 ± 0.01 to a perfect 1.00. In this case, constrained decoding compensated for the model's weaker query-generation logic by narrowing the output space to valid schema elements.

The 3B model tells a different story. It already achieved $EA = 0.62 \pm 0.02$ ($n = 9$) under Soft Cypher—comparable to the 7B model's *constrained* result—with low variance across runs. Grammar constraints (S4) brought EA only marginally higher to 0.64, a difference well within the S3 standard deviation. The benefit of S4 for the 3B model was therefore not in accuracy but in *reliability*: SVR improved from 0.95 ± 0.03 to 1.00 and SCR from 0.94 to 1.00, guaranteeing that every generated query is syntactically valid and schema-compliant. This distinction matters in production settings, where even a 5% rate of invalid queries can erode user trust.

These results suggest that grammar constraints serve two distinct roles depending on model capability: for weaker models, they act as an accuracy amplifier by restricting the search space; for stronger models that already generate correct logic, they act as a safety net that eliminates the residual tail of schema hallucinations and syntax errors.

To further verify the effect of grammar constraints (S4) relative to the prompt-based baseline (S3), results were decomposed by query complexity: Level 1 (Class Retrieval), Level 2 (Property Constraints), and Level 3 (Topology/Relationships).

- **Level 1: Class retrieval (easy).** Under Strict Cypher (S4), both the 7B and 3B models achieved perfect performance ($EA = 1.00$, $SCR = 1.00$, $SVR = 1.00$). In the unconstrained S3 setting, the 3B model already performed strongly ($EA = 0.98 \pm 0.06$), while the 7B model showed more variability ($EA = 0.30 \pm 0.07$). The transition to S4 eliminated all remaining instances of invalid entity labeling.
- **Level 2: Property value retrieval (medium).** The 3B model showed high resilience in this category already under Soft Cypher ($EA = 0.79 \pm 0.02$). However, constraints ensured that property keys were strictly mapped to IDS- and model-derived vocabulary, maintaining $SCR = 1.00$ across both models and avoiding confusion between similarly named attributes.
- **Level 3: Topology and relationships (hard).** Multi-hop traversals remained challenging. While the 7B model under S4 constraints achieved only marginal correctness ($EA \approx 0.03$), the 3B model under S4 provided the only successful executions in this category ($EA \approx 0.13$). This suggests that by narrowing the search space to valid relationship types and property keys, grammar constraints can occasionally assist smaller models in constructing complex graph patterns that would otherwise fail under unconstrained prompting.

Overall, grammar constraints substantially improve reliability for class and property retrieval by mitigating schema hallucinations, but they cannot compensate for limited logical reasoning in complex topological queries.

Limitations and implications

The results indicate that constrained NL2Cypher generation can substantially improve the reliability of natural-language querying over IFC-derived graphs, particularly for class retrieval and property-level queries. In practice, this can make BIM information more accessible to non-expert stakeholders by reducing the need to learn specialist query languages and by limiting common failure modes such as schema hallucinations and non-executable queries.

Several limitations must be acknowledged. First, the evaluation constitutes an initial validation on a single sample project with one IDS file and 16 queries; broader replication across IFC files from real projects, with different modelling conventions, varying data quality, and larger query sets, is required before generalisability claims can be made. Second, the hybrid schema is only as effective as the underlying IDS and the extracted model vocabulary; incomplete or overly permissive specifications may still allow ambiguous queries, while overly strict specifications can reduce coverage of valid (but unspecified) information. Third, and most critically, the framework exhibits near-total failure on Level 3 topology queries ($EA \approx 0.03$–0.13), demonstrating that grammar constraints—while effective at eliminating schema-level errors—cannot compensate for the multi-step logical reasoning that complex graph traversals require. Time and cognitive effort were not measured in this study; future work should include user studies to quantify usability gains.

Fourth, IDS availability is still uneven in current practice. The proposed framework is most effective in projects where at least a partial IDS already exists. When the IDS is incomplete, the hybrid schema can still enforce the entities and properties that are specified and rely on model-derived vocabulary elsewhere. When the IDS is missing entirely, the method loses its project-specific grounding and should be interpreted as a query-assistance setup rather than a full compliance-oriented one; grammar-constrained generation remains possible over the model-only vocabulary, but without the strict property typing and enumeration controls that prevent hallucinations on ambiguous or inconsistent data. The approach therefore does not replace careful IDS authoring, but it can benefit from gradual IDS formalisation: because the grammar makes explicit which entities and properties are queryable, project teams receive immediate feedback on specification gaps, turning IDS creation from a retrospective compliance exercise into a proactive step that directly improves query reliability.

Conclusions

This paper presented an IDS-constrained text-to-graph querying framework that addresses the reliability gap when applying LLMs to BIM information retrieval. By transforming IFC data into a labeled property graph and compiling project-specific IDS requirements into a hybrid schema, the approach enables grammar-constrained decoding that eliminates schema hallucinations and guarantees syntactically valid Cypher queries. The experimental results confirm that LLMs are most effective when used to generate query logic rather than retrieve information directly, and that combining graph-based representations with formal semantic constraints improves accuracy.

These findings have practical implications for the Architecture, Engineering, Construction and Operators (AECO) industry. First, LLM-mediated BIM querying should be treated as a query generation problem with hard constraints, not a prompt engineering problem. Second, IDS can serve proactively to reduce variability in query formulation, beyond its traditional role in retrospective compliance checking. Third, by offloading syntactic precision to the constraint engine, smaller open-weights models can achieve reliability levels previously reserved for large cloud models, enabling privacy-preserving, local deployment without uploading sensitive project data to external APIs.

Future work will (i) scale experiments across additional domains and IDS complexities, (ii) investigate robustness under non-compliant or inconsistent IFC models, and (iii) extend the graph schema to better represent spatial context (e.g., second-level space boundaries) for queries about connectivity, adjacency, and energy-relevant relationships. Beyond accuracy, future evaluations should consider deployment factors such as latency and the environmental footprint. Two promising directions are interactive self-correction, where execution feedback (e.g., "query returned 0 nodes") drives iterative refinement, and hybrid graph-vector retrieval for unstructured data. Nevertheless, the results demonstrate that this NL2Cypher technique provides a practical and effective pathway toward democratizing BIM information access in a reliable way.

References

Autodesk (2026). Autodesk (company website). `https://revit.downloads.autodesk.com/download/2024RVT_RTM/Docs/InProd/racadvancedsampleproject.rvt`. Accessed: 2026-03-23.

Beetz, J., van Leeuwen, J. P., and de Vries, B. (2009). IfcOWL: A case of transforming EXPRESS schemas into OWL ontologies. In Proceedings of the 1st International Workshop on Semantic Digital Product Memories (SDPM 2009). Co-located with the 6th European Semantic Web Conference (ESWC 2009).

Bloch, T., Borrmann, A., and Pauwels, P. (2023). Graph-based learning for automated code checking–exploring the application of graph neural networks for design review. Advanced Engineering Informatics, 58:102137.

Borrmann, A., Beetz, J., Koch, C., Liebich, T., and Muhic, S. (2018). Industry foundation classes: A standardized data model for the vendor-neutral exchange of digital building models. In Borrmann, A., König, M., Koch, C.,

and Beetz, J., editors, Building Information Modeling, chapter 5. Springer, Cham.

buildingSMART International (2024). Information Delivery Specification (IDS). Standard and documentation. Accessed 2026-01-30.

Daum, S. and Borrmann, A. (2014). Processing of topological BIM queries using boundary representation based methods. Advanced Engineering Informatics, 28(4):272–286.

dottxt-ai (2026). Outlines documentation. `https://dottxt-ai.github.io/outlines/latest/`. Accessed 2026-02-05.

Du, C., Esser, S., Nousias, S., and Borrmann, A. (2026). Text2bim: Generating building models using a large language model-based multiagent framework. Journal of Computing in Civil Engineering, 40(2):04025142.

Esser, S., Vilgertshofer, S., and Borrmann, A. (2022). Graph-based version control for asynchronous BIM collaboration. Advanced Engineering Informatics, 53:101664.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, H., and Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997, 2(1).

Hellin, S., Nousias, S., and Borrmann, A. (2025). Natural language information retrieval from BIM models: An LLM-based agentic workflow approach.

Hjelseth, E. and Nisbet, N. (2016). Model checking in BIM-based projects. International Journal of 3-D Information Modeling, 5(2):1–20.

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Liu, X., Wang, W., and Feng, H. (2023). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. arXiv preprint arXiv:2311.05232.

International Organization for Standardization (2024). ISO 16739-1:2024 — industry foundation classes (IFC) for data sharing in the construction and facility management industries — part 1: Data schema. `https://www.iso.org/standard/84123.html`. Accessed 2026-02-05.

Iranmanesh, S., Saadany, H., and Vakaj, E. (2025). LLM-assisted Graph-RAG information extraction from IFC data. arXiv preprint arXiv:2504.16813.

Ismail, A., Nahar, A., and Scherer, R. (2017). Application of graph databases and graph theory concepts for advanced analysing of BIM models based on IFC standard. Proceedings of EGICE, 161.

Jiang, H. et al. (2023). StructGPT: A general framework for large language model to reason over structured data. arXiv preprint arXiv:2305.09645.

Khademi, H. and Behan, A. (2017). A review of approaches to solving the problem of BIM search: towards intelligence-assisted design. In Proceedings of CitA BIM Gathering 2017, Dublin, Ireland.

Mazairac, W. and Beetz, J. (2013). BIMQL–an open query language for building information models. Advanced Engineering Informatics, 27(4):444–456.

Neo4j (2026). Neo4j graph database platform. `https://neo4j.com/`. Accessed 2026-02-05.

Pauwels, P., Krijnen, T., Terkaj, W., and Beetz, J. (2017). Enhancing the ifcOWL ontology with an alternative representation for geometric data. Automation in Construction, 80:77–94.

Rasmussen, M. H., Hviid, C. A., and Karlshøj, J. (2020). Bot: The building topology ontology of the w3c linked building data group. Semantic Web, 11(4):535–559.

Solihin, W. and Eastman, C. (2015). Classification of rules for automated BIM rule checking development. Automation in Construction, 53:69–82.

Wang, Z., Sacks, R., and Yeung, T. (2022). Exploring graph neural networks for semantic enrichment: Room type classification. Automation in Construction, 134:104039.

Yin, M., Tang, L., Webster, C., Xu, S., Li, X., and Ying, H. (2023a). An ontology-aided, natural language-based approach for multi-constraint BIM model querying. Journal of Building Engineering, 76:107066.

Yin, Y., Heaton, J., Borrmann, A., and Beetz, J. (2023b). An ontology-based approach for natural language querying of building information models. arXiv preprint arXiv:2303.15116.

Yu, Y. et al. (2023). Trends and advances in IFC schema extensions for BIM interoperability: A systematic review. Applied Sciences, 13(23):12560.

Zhang, C., Beetz, J., and de Vries, B. (2018). BimSPARQL: Domain-specific functional SPARQL extensions for querying RDF building data. Semantic Web, 9(6):829–855.

Zhu, J., Nisbet, N., Yin, M., Wei, R., and Brilakis, I. (2025). Releasing the power of graph for building information discovery. Automation in Construction, 172:106034.

Zhu, J., Wu, P., and Lei, X. (2023). IFC-graph for facilitating building information access and query. Automation in Construction, 148:104778.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATED SCHEDULE ENRICHMENT FROM DAILY PROGRESS REPORTS: A BI-DIRECTIONAL NEURO-SYMBOLIC AI APPROACH

Sudheer Kumar Nanduri, Venkata Santosh Kumar Delhi
Indian Institute of Technology Bombay, India

Abstract

Construction progress monitoring requires integrating daily progress reports (DPRs) with baseline schedules for productivity analysis and risk management. Manual integration is unscalable due to heterogeneous formats and inconsistent terminologies. This paper presents a neuro-symbolic system that enriches schedules from the DPR files utilizing the large language models and knowledge graphs. The system learns construction terminology from schedule and DPRs, adapts confidence weighting, and validates outputs against project constraints. Tested on a real construction project with 173 DPRs and 1000 validation activities, it achieved a median confidence of 0.825 and zero constraint violations.

Introduction

Construction schedules are central to project planning, monitoring, and control (Project Management Institute, 2021), yet they are rarely updated using the execution data collected during the construction (Olawale & Sun, 2015). Intermittent and partial schedules are used in the routine project communications, losing the golden thread of workflow information across multiple digital documents (Adekunle et al., 2022). Daily progress reports (DPRs) capture detailed fieldwork plans, completed and delayed activities, and exact quantities, resources, and locations. But they typically remain disconnected from the baseline schedules.

Automating the integration of DPRs with schedules, referred to as schedule enrichment in this work, remains a challenge due to heterogeneity arising from document origins. DPRs are semi-structured documents, and the terminology used varies significantly across projects, contractors, and engineers. In contrast, schedules are more formal, structured, and governed by strict temporal and precedence constraints. Field implementations may sometimes relax the strictness, but the precedence information remains logical. Bridging these two information sources requires both semantic flexibility and logical consistency.

Existing automated approaches address part of the problem. Language models can recognize semantic similarity between activity descriptions, but cannot reliably enforce temporal or causal constraints. Rule-based systems provide logical rigor but fail when faced with vocabulary variations and incomplete ontologies. Sequential combinations of these approaches remain brittle, as validation occurs after matching decisions are made.

This paper proposes a neuro-symbolic approach that integrates neural semantic reasoning with symbolic constraint validation during inference. The system learns construction vocabulary automatically from successful alignments and adaptively balances neural and symbolic contributions as knowledge accumulates. The approach is validated on real project data and demonstrates that bi-directional neuro-symbolic integration enables scalable, interpretable, and constraint-aware schedule enrichment.

Related Work

Project monitoring and automated schedule updating are actively researched topics in the construction automation domain (Turkan et al., 2012). Existing approaches rely heavily on building information modeling (BIM) with visual data analytics (Golparvar-Fard et al., 2015) or point clouds (Park & Cho, 2022). While model-based approaches are effective in geometric progress tracking, they are less suited for capturing workflow semantics and also face adoption challenges (Saka & Chan, 2020).

These limitations motivate a shift towards knowledge-based approaches that operate directly on documents. Natural language processing (NLP)- based approaches were explored for extracting activity information from schedules (Zhao et al., 2020) and for bridging schedule activities to the Uniformat (Jung et al., 2024) standard, thereby standardizing the ontology.

Transformer-based approaches were found to be valuable for aligning schedules with field reports and for generating look-ahead planning (Amer et al., 2021, 2023). Large Language Models (LLMs) have demonstrated strong performance in understanding domain-specific vocabulary with minimal task-specific engineering. From generating schedules autonomously (Al-Sinan et al., 2024) to identifying constraints from field meeting transcripts (He et al., 2024). Although LLMs have shown generative capabilities in semantics, the ontology is the primary driver of the logic and validation, which LLMs do not explicitly model.

In contrast, graph-based methods (C. Wu et al., 2023) provide a structured mechanism for enriching existing knowledge, either from schedules or from constraints. Domain-specific knowledge graphs are graph-based

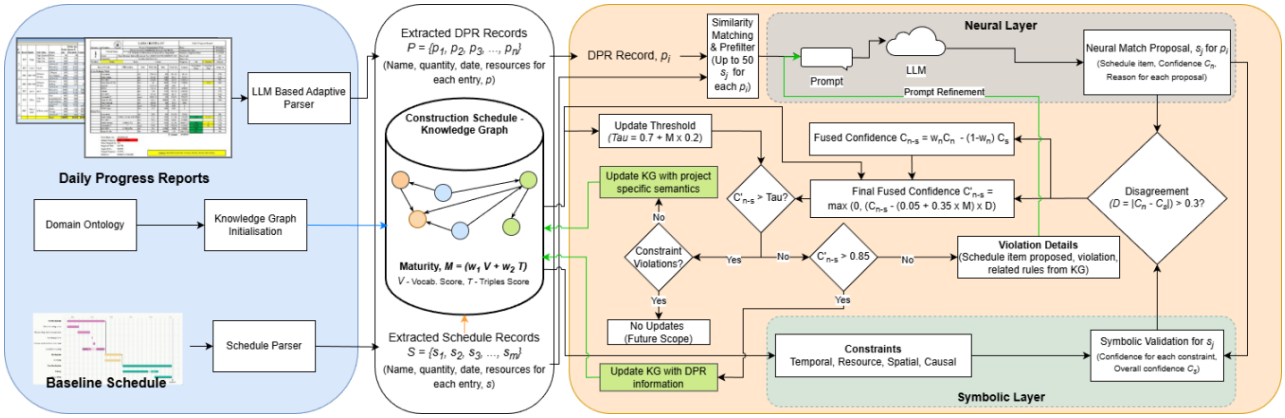


Figure 1 Overall System Architecture

approaches that simultaneously store the project information and allow for constraint application mapping very easily with a construction schedule (S. Wu et al., 2024).

Further, LLMs are shown to work well with knowledge graphs (KG) (Cai et al., 2025), even within the domain of construction automation (Li et al., 2024). However, most current efforts remain unidirectional flows that allow a neural system to extract data for a symbolic system to process, limiting the system's ability to incrementally gather knowledge from daily changes in the field.

Neuro-symbolic AI offers a more robust approach. Recent work in neuro-symbolic AI has proposed multiple integration strategies ranging from loose coupling to tightly embedded symbolic reasoning (Kautz, 2022). The current methods of NLP lack the logical rigor needed for understanding schedules and field documentation without engineering. Language model approaches demonstrate strong semantic understanding of unstructured information but lack mechanisms to enforce constraints required for schedule correctness. Rule-based and ontology-driven approaches enable constraint checking and logical consistency but depend on manually curated vocabularies and struggle with heterogeneous, informal field terminologies.

Thus, the specific research gap lies in the absence of a Type 6 (Neuro[Symbolic]) framework that embeds symbolic reasoning as a subroutine within a neural process loop. Existing research does not address automated vocabulary learning or adaptive integration, in which the system transitions the governance as domain knowledge evolves. As a result, there is currently no method that enables scalable, constraint-aware alignment of DPRs with schedules while continuously learning project-specific terminology.

Methodology

To address the identified research gap, this paper proposes a neuro-symbolic methodology for automated construction schedule enrichment from daily progress reports. The method integrates neural semantic reasoning with symbolic constraint validation in a unified inference process, enabling robust alignment between DPR data and schedule data.

System Architecture

The proposed methodology, structured and presented in Figure 1, is designed as a Bi-directional Neuro[Symbolic] (Type 6) (Kautz, 2022) architecture. This approach balances the semantic capabilities of LLMs and the validation capabilities of KGs in an iterative loop. The overall architecture has three interacting layers:

1. Data preparation layer responsible for parsing DPR files, scheduling data, and initializing the knowledge graph.
2. The Knowledge graph layer, which gets enriched through the processes, can be used to enrich the schedule after processing DPR.
3. Neuro-Symbolic matching layer responsible for candidate DPR-schedule alignment proposal, validation, and refinement.

Problem formulation

Let

$P = \{p_1, p_2, p_3, \dots, p_n\}$ denote the set of extracted DPR records,

$S = \{s, s_2, s_3, \dots, s_m\}$ denote the set of baseline schedule tasks, and

Each DPR activity p_i is associated with observed attributes, including execution date, quantity, location, and resource usage. Each scheduled task s_j is associated with planned attributes, including time windows, locations, quantities, and precedence relationships.

The objective is to determine a mapping function.

$$F : P \rightarrow S$$

that aligns the DPR activities with corresponding schedule tasks while maintaining logical consistency and enabling schedule enrichment. Constraint information used during alignment is derived from schedule-defined constraints and from observed constraints inferred from DPR execution data. The system validates proposed vocabulary and alignments against available constraints and incrementally augments the knowledge graph with newly observed data when sufficient evidence is present. These learned mappings are subsequently used to refine future alignments and to enrich the baseline schedule without manual ontology engineering.

Data Preparation Layer

Daily Progress Report Parsing

In this work, we utilised DPR files generated at the project level by the contractor's planning office and submitted to the cost consultant. All reports originate from a single organizational source and use a common spreadsheet template, reducing the heterogeneity of file type choices and broader design choices. However, different engineers introduced varying column ordering, multiple data regions per sheet, mixed data types, merged cells and colours, comments along with the data, and hidden row data, apart from progress details. In case DPRs originate from multiple sources with different formats, the same parsing pipeline described below can be applied to each source.

We followed a semi-automated approach for parser scripting. The first 3-5 DPR files are sampled for inspection to read raw cell values. Known patterns of the document structure that can be ignored are coded to allow focusing on the header, data presence, and hierarchy markers from merged cells. The extracted details are used to construct a detailed prompt for Large Language Models (LLMs) to generate code for targeted data parsing.

The LLM-generated script is run on all the DPR files for data parsing and extraction. Each parsed record is checked for activity name, quantity, date, and location hierarchy, and missing data is used to calculate a quality score for the LLM-generated script. The quality score is used to generate new code versions for improved data extraction, with details of the previous run's errors and additional directions added to the prompt.

Schedule Loading and Knowledge Graph Initialisation

The baseline schedule is parsed for task definitions, structure, temporal constraints, quantities, and relationships. Simple natural language processing steps, such as case correction, punctuation removal, and lemmatization, are applied to the extracted textual data. The scheduled tasks and their relationships are represented as entities and relations within the knowledge graph.

The knowledge graph is initialised as a formal RDF graph. Base ontology is created with domain-specific semantics comprising 16 common activity types (excavation, PCC, etc.), 12 logic-based precedence constraints (reinforcement before shuttering, etc.), and 4 base synonyms (shuttering synonym of formwork, etc.), forming around 80 triples.

For each schedule task, the graph instantiates a unique entity, mapping it to its corresponding activity type and its temporal predecessors as graph edges. One project-specific problem was that the baseline schedule is activity-based but not delineated by location, whereas DPRs present activity and location hierarchy. At the initial schedule ingestion phase, enforcing location is not useful, and so the location constraint is kept for semantic hinting to LLMs and only soft matching in the KG.

Since we do not have the project-specific knowledge at the start, the knowledge graph is set up with a maturity score to guide the transition from purely neural to neuro-symbolic. This setup allowed us to achieve better semantic performance without encountering cold-start issues. The equations below represent the maturity formula built for the purpose.

$$Maturity\ Score, M = w_1 * V + w_2 * T \quad (1)$$

$$Vocabulary\ Score, V = min\left(1.0, \frac{v_l}{v_T}\right) \quad (2)$$

$$Triples\ Score, T = min\left(1.0, \frac{t_l}{t_T}\right) \quad (3)$$

Where,

v_l = Learned mappings beyond the baseline KG initialisation (0 mappings),

v_T = Threshold mappings (200 mappings) – considered to be fully mature,

t_l = Learned triples beyond the baseline KG initialization (86 triples),

t_T = Threshold triples (10,000 triples),

w_1, w_2 = Weightage for vocabulary score and triples score. (0.7, 0.3 respectively)

The weights are empirically determined, specific to the project, based on the available data. The thresholds are considered based on domain knowledge. Roughly 200 unique activities and 10,000 triples can be considered for the construction activities.

The KG is considered immature when its score is less than 0.1, and most subsequent layers tend to use the neural system in such cases. As the KG becomes more enriched, its maturity increases, and the dependency on the symbolic component also increases.

Neuro-Symbolic Layer

This process follows each DPR activity in a multi-step alignment process.

Neural Proposal Stage

The activity description of each task pi, a list of leaf tasks (non-summary tasks), and a prompt to select the best matching task based on semantic, spatial context, and hierarchy are sent to an LLM.

One specific detail of the task list is the workflow based on the KG's maturity. If the KG is immature, the top 50 schedule tasks are utilised. If the KG is mature, then a symbolic pre-filtering stage is applied. The filters include activity type match, keyword overlap, and trade separation, which are used to filter the tasks. The schedule data lacked separation of MEP concerns, whereas the DPR provides detailed separation. So, immature KG will not know the correct MEP trade for pre-filtering by activity type, and so this part is explicitly added. If filtering is successful, the top 50 tasks are considered for similarity matching.

The LLM returns a scheduled task that matches the prompt, along with its confidence and reasoning. The confidence value is the neural system confidence (C_{neural}), ranging from 0 to 1.

Symbolic Validation Stage

The proposed schedule element is validated against four constraints in the KG.

1. Temporal constraint: The DPR task must fall within a temporal window with ±7 days over the scheduled start and end dates of the task. Additional validation regarding the completion of the prerequisite tasks before the activity's date
2. Resource constraint: The DPR task achieved cumulative quantity must fall within a 30% range of the scheduled task
3. Spatial constraint: The DPR task overlaps with activity locations in the KG
4. Causal constraint: The DPR task preceding activities must be completed. Check if earlier matched DPR activity exists.

All violations are treated as soft violations, as weather issues, resource constraints, scope changes, and other factors can occur during field execution. We assigned empirical confidence penalties (0.9 when valid, 0.3-0.4 when invalid), calibrated on a held-out subset of 50 DPR records.

To assess sensitivity, the penalties were uniformly varied by 20%. The median fused confidence ranged from 0.61 (for -20%) to 0.78 (+20%). A span of 0.17 points. The high-confidence alignment rate varied from 1.5% to 11.0%. This moderate variation suggested that exact penalty calibration is not a critical operation, and the system remains stable under reasonable parameter changes. The mean aggregated confidence across all four constraint dimensions is considered the symbolic system confidence ($C_{symbolic}$), ranging between 0 and 1.

Disagreement Resolution

The absolute difference between the neural and symbolic system confidence is considered Disagreement (D) if it exceeds 0.3. If there is a disagreement, a fused confidence score is calculated for processing.

$$D = \left|C_{neural} - C_{symbolic}\right| \quad (4)$$

$$C_{n-s} = w_{neural} * C_{neural} + (1 - w_{neural}) * C_{symbolic} \quad (5)$$

$$w_{neural} = 0.85 - (\mathrm{M} * 0.35) \quad (6)$$

Where,

D is the disagreement between neural and symbolic components,

C_{n-s} is the Fused Confidence Score of the system,

w_{neural} is the weight for the neural component, derived from the Maturity Score, M, described in (1)

C_{neural} & $C_{symbolic}$ are the neural and symbolic confidence values

The KG maturity score is inversely correlated with the symbolic system confidence. Thus, when KG is immature, the neural system is preferred for fused confidence. If the KG is mature, and there is a disagreement between the two components, the fused score is calculated with a penalty term based on the KG maturity score.

In practical terms, an immature KG keeps the workflow neural-heavy, the candidate selection remains broad, and the fused confidence is driven more by the LLM. As the KG matures, the workflow shifts towards symbolic governance through pre-filtering and stronger constraint weighting, typically producing lower but more reliable confidence scores.

$$Penalty_strength, p = 0.05 + (M * 0.15) \quad (7)$$

$$Penalty, P = Penalty_strength, p * Disagreement, D \quad (8)$$

$$C'_{n-s} = \max(0.0, (C_{n-s} - P)) \quad (9)$$

Where,

p is the penalty strength calculated over the Maturity Score S described in (1),

P is the penalty value, and

C'_{n-s} is the Final Fused Confidence Score.

The final fused confidence score is the decisive variable for the iterative refinement step.

Iterative refinement

If the fused confidence exceeds 0.85, the DPR is aligned with the schedule activity and updates the KG.

A variable threshold, τ, as detailed in (10), is used to determine whether to map the vocabulary and enrich the KG.

$$\tau = 0.7 + M * 0.2 \quad (10)$$

Where,

τ is the variable threshold calculated from the Maturity Score M mentioned in (1).

The final fused confidence score (C'_{n-s}) is compared against the threshold τ. If the confidence exceeds the threshold and there are no constraint violations, the DPR element is utilised for vocabulary mapping and for the DPR entry as a record of the progress of the actual schedule. Even if the confidence is below the threshold but greater than 0.85, the DPR entry is still recorded as a progress entry, but no vocabulary mapping will be made. If there are violations, but the confidence score is high, the constraints can be updated. But this step will require a detailed study of which constraints can be safely updated and which should be left for future work.

Both the above steps of threshold verification and updating the KG form the core of forming a $Neural -> Symbolic$ learning system. For the DPR activity p_i and the proposed schedule activity s_i, the fused confidence scores, with breakdowns, constraint violations, and other relevant details, are collected and sent back to the LLM for the next alignment suggestion, forming a $Symbolic -> Neural$ constraint enforcement system. To prevent the system from entering an infinite loop, iterative refinement stops after 3 rounds.

Experimental design

System Design Validation

To validate the model, the pipeline is run on the schedule and applied to 1000 activities from the DPRs. Six validation tests were conducted on the research claims. In

addition, 105 DPR-schedule alignments (from the ground truth dataset, stratified across all three buildings) were manually verified by a domain expert. Metrics like precision and F1 score were computed at four evaluation levels: Strict task-UID match, building-aware match for correct activity type and building, activity type-only match, and coverage, where any match is produced when one is expected.

1. Neuro-Symbolic Integration – Integration logs were used to count events by direction between neural and symbolic systems, and measured the refinement success rates by comparing confidence scores. Considered a pass if we observe > 100 events with > 70% refinement success.
2. Bidirectional Learning – Alignment confidence scores were split into early (first 50%) and late (last 50%) groups, and the mean confidence change was calculated; improvement at the late stage was considered a pass if it exceeded 1%. Also, the vocabulary counts are fit to linear regression to detect trends.
3. Knowledge Accumulation – KG statistics before and after pipeline execution were measured as the growth factor in triples relative to the initial stage. If the growth factor is > 10 and the vocabulary mappings increase by> 50, then it is considered a pass. If the growth factor increases by 50, it is regarded as strong evidence
4. Entity Resolution Effectiveness – Entity counts before and after deduplication are compared. If the ratio is more than 50%, it is considered a pass, and it is considered strong if the ratio > 70%.
5. Symbolic Inference Performance – Triples discovered by the reasoning system and logical conflicts found are measured. The validation passed if inferred triples are >1000, and conflicts < 5%.

Baseline and Ablation Studies

For baseline studies, we implemented the same three approaches from scratch and executed them on the sample.

1. Pure Neural (LLM) – Direct LLM matching without KG validation or iterative refinement. We performed 10 LLM calls, each aligning one DPR activity to the full schedule. A smaller random sample of DPR activities is chosen to isolate and assess the LLM behaviour for relative trends rather than a full-scale evaluation.
2. Pure Symbolic (Rules) – String matching (and Jaccard similarity of keyword overlap executed on all 100 activity samples.
3. Traditional NLP (TF-IDF) –Vectorization and cosine similarity matching-based approach on all 1000 activities.

For ablations, we adopted an analytical approach. Each component's contribution is used to compute the degraded confidence if it were absent.

Learning Trajectory Verification

To verify the learning trajectory, we rerun the pipeline on the dataset 10 times, with a sample size of 10 in each run. In each iteration, the activity batch is randomized. The KG maturity, average confidence, and vocabulary learned are extracted at each iteration, and the final KG serves as the starting point for the next. We computed the maturity growth between the final and first iteration, and the confidence improvement ratio. Successful learning must show at least a 1% increase in confidence and a monotonically increasing level of maturity.

Results

System Design Validation

Table 1. System Design Validation Results

Test	Evidence	Metrics
Neuro-Symbolic Integration	Moderate	775 integration events, 100% refinement
Bi-directional Learning	Weak (single pass)	-7.7% confidence change (0.838 to 0.774); 0 vocabulary at pre-populated KG; 375 vocabulary mappings acquired during processing
Knowledge Accumulation	Strong	KG grew from 86 ontology triples to 39,620 (456x); alignment-driven growth: +4660 triples from 34,600 post-schedule baselines
Entity resolution	Strong	1139 activity variants reduced to 134 canonical entities (88% compression); 760 resolution events across exact (58%), LLM (34%), and similarity (3%) methods
Symbolic inference	Strong	28,975 triples inferred; 0 logical conflicts detected

On the manually verified subset, the precision and F1 scores for activity-type match are 0.930 and 0.815; for building-aware match, 0.338 and 0.400; for Strict task-UID matches, 0.141 and 0.189; and for coverage matches, 1.000 and 0.850. Across all 1000 alignments, 769 were matched, yielding a 76.9% match rate. Most strict-UID errors were not semantic failures. 56 of 61 wrong matches still identified the correct activity type, with errors dominated by cross-building confusions (42) and granularity mismatches (14). These results indicate strong semantic recognition, with the remaining difficulty concentrated in spatial disambiguation.

The bidirectional learning row in Table 1 reflects an early vocabulary saturation. The confidence decrease reflects the KG maturity (Eq. 6), penalizing matches that lack full constraint support. Thus, the negative confidence change reflects a calibration effect rather than evidence against bidirectional learning. The broader evidence for

bidirectional learning comes from the same run's 375 vocabulary mappings acquired during DPR processing.

Details of the symbolic inference validation suggest that the symbolic reasoning primarily generated triples of transitive closure over the precedence relationship (about 73.8% of the total KG content), suggesting that implicit knowledge discovery amplifies explicit data. However, the absence of conflicts indicates the need to reexamine the precedence graph derived from the validated schedule. Further verification is needed using synthetic, contradictory data.

Baseline & Ablation Comparison

The baseline comparison across approaches is presented in Figure 2. The median confidence is higher than the mean, indicating a bimodal distribution with many high-quality matches. The pure LLM approach performs comparably but has lower mean and median values than the proposed approach. Higher confidence (10% compared to 4.3% of ours) in the LLM approach might suggest the lack of constraint validation, which is derived from the symbolic component in our system. The purely symbolic approach, understandably, fails to handle vocabulary variation. Still, the traditional NLP implementation treats pairs as minimal datasets to measure similarity and cannot be applied directly. However, this doesn't affect the overall implementation and results.

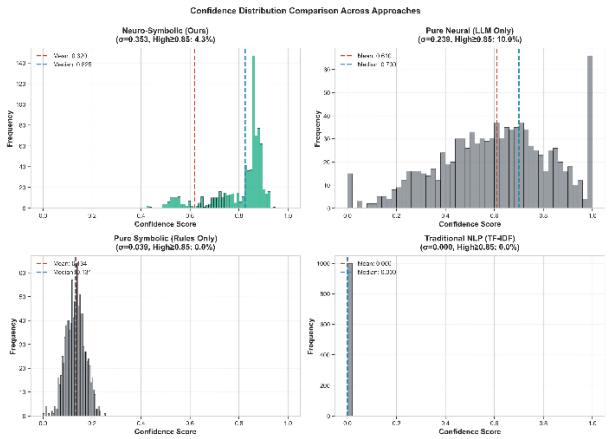


Figure 2 Baseline Comparison Results

The ablation study results are presented in Figure 3. The top contributors are KG validation, vocabulary learning, and iterative refinement, which together account for nearly 70% of the system's performance. All three are mechanisms of the neuro-symbolic system, confirming the system's synergy. Of the 769 alignments requiring refinement, KG validation successfully identified inconsistencies and is the most valuable component.

Qualitative Analysis

Representative examples from the evaluation illustrate the system's practical behaviour. A DPR entry reporting "Waterproofing" activity for Building C (Terrace Floor, Super-Structure section) was matched to the schedule task "Terrace Floor water proofing", with a fused confidence of 0.949 (neural 0.95, symbolic 0.947), indicating semantic agreement, temporal validity, spatial consistency, and precedence satisfaction. In a contrasting case, the DPR entry "Sunk Area Filling" (for building B) was matched to "PCC" with a fused confidence of 0.758 (neural 0.90, symbolic 0.0), resulting in a disagreement and flagging for expert review. 188 such cases were identified of the 1000 processed alignments, demonstrating the system's ability to surface uncertain matches rather than silently accepting them. The KG also learned 375 vocabulary mappings, including synonyms such as shuttering for formwork, typographical errors like conctere for concrete, and field-specific terms like UGWT Raft, mapped to the canonical Underground Water Tank activity type, which persisted in the KG without requiring further LLM calls.

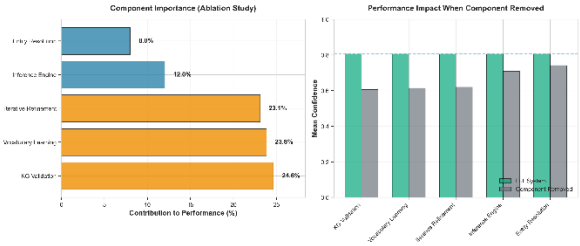


Figure 3 Ablation Study Results

Learning Trajectory Verification

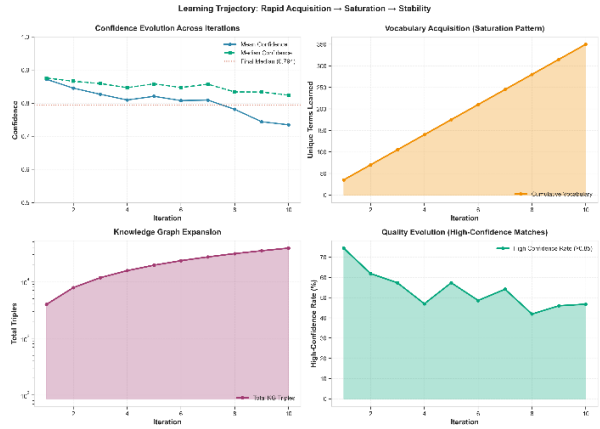


Figure 4 Learning Trajectory Verification

The learning trajectory results over 10 iterations are presented in Figure 4. Overall, the KG maturity shows a +68% increase, with 7 terms and 44 KG triples learned. The confidence decreases by 1.7%, with intermittent dips, possibly due to adaptation to edge cases. The initial iterations also show high values, suggesting that the initial sending with the schedule and the first iteration itself rapidly accumulates valuable project-specific knowledge.

Figure 5 shows that most KG growth comes from schedule-derived structural scaffolding, while DPR alignment provides a smaller but semantically important set of learned mappings and observations.

Discussion

The results demonstrate that the proposed neuro-symbolic system can appropriately align daily progress reports with baseline schedules and regulate confidence in logically

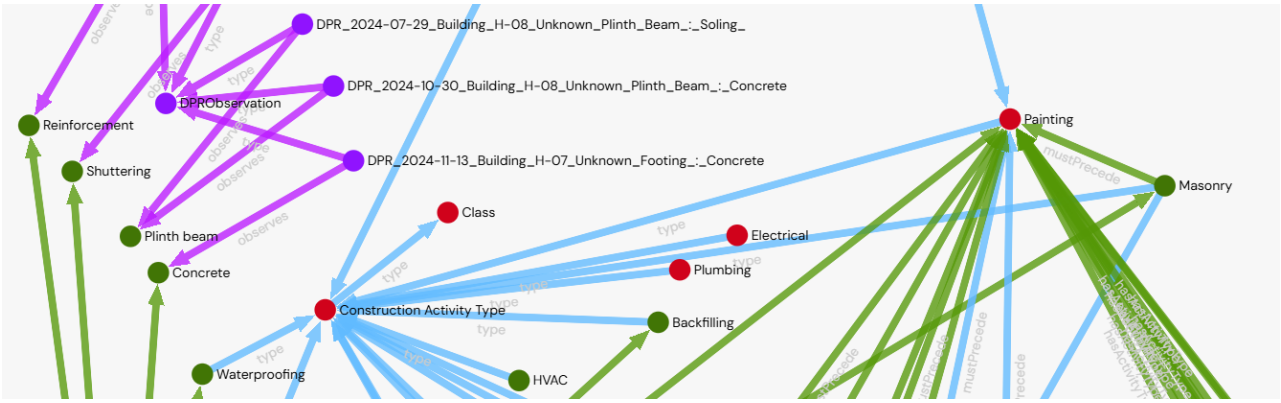


Figure 5 Knowledge Graph Evolution Across Pipeline Stages. Baseline nodes and edges are in red; schedule parsing-based nodes and edges are in green, DPR parsing-based nodes and edges are in violet.

invalid matches. This result is reflected in consistently higher median confidence values and the absence of constraint violations, even when operating on the DPR files. Of the 1000 activities, 769 alignment matches passed all constraint validation, and the remaining 231 activities were left unmatched, achieving zero logical conflicts in the inferred knowledge graph. Relative to the LLM baseline, the main gain is better confidence calibration rather than higher peak scores.

A key observation from the experiments is that system performance is driven by the ability to calibrate confidence. The LLM baseline occasionally produced higher peak confidence values, but these were not supported by symbolic validation. In contrast, the proposed approach is more robust, limiting the extreme confidence, since incorrect high-confidence updates can distort the KG for downstream applications.

The presence of 775 recorded interaction events with successful refinements suggests that embedding symbolic reasoning within the inference loop enables systematic reconciliation rather than post-hoc filtering. This behavior directly supports the effectiveness of a Type 6 neuro-symbolic integration as discussed earlier. The ablation study further reinforces this observation, showing that knowledge graph validation, vocabulary learning, and iterative refinement mechanisms account for nearly 70% of the system's performance contribution.

Error analysis shows that most strict UID failures were cross building or granularity confusions rather than semantic misunderstandings, indicating that the main remaining limitation is spatial disambiguation. Knowledge accumulation was substantial with KG growth dominated by precedence closure and 375 vocabulary mappings acquired during DPR processing. Validation on a single project remains a limitation, although the dataset spans 173 DPR files (i.e., 173 days of data) and 1000 activities, and the initial ontology is domain-general. The current validation does not stress-test contradictory schedules, leaving multi-project and synthetic conflict evaluation as future research directions. Overall, the findings indicate that the value of the proposed approach lies not in outperforming the neural baselines on isolated similarity metrics but in delivering a scalable, interpretable, and constraint-aware mechanism for schedule enrichment under realistic field conditions.

Conclusion and Future Work

This paper presented a bi-directional neuro-symbolic framework for automated construction schedule enrichment using daily progress reports. By embedding symbolic constraint reasoning within a large language model-driven inference loop, the proposed system enables semantically flexible, logically governed alignment between heterogeneous field documentation and formal schedules. Validation on a real construction project demonstrated that the framework achieves robust alignment performance, characterized by high median confidence, zero accepted constraint violations, and substantial growth of project-specific knowledge without manual ontology engineering.

Comparative baseline and ablation studies confirmed that the system's effectiveness arises from neuro-symbolic synergy rather than isolated neural or rule-based components. Symbolic validation and iterative refinement were shown to actively govern inference, preventing logically invalid schedule updates while allowing the language model to adapt to informal terminology and evolving field context. The learning trajectory analysis further indicated that the system transitions from rapid vocabulary acquisition to stable, conservative alignment behavior as domain knowledge matures, a desirable property for operational deployment.

Several directions for future work naturally follow from this study. First, while the current validation focused on a single project, extending the framework to multi-project datasets would enable systematic evaluation of generalizability across different schedule structures and documentation practices. Second, stress-testing the system with synthetic or adversarial contradictions, such as intentionally inconsistent precedence or delayed prerequisite activities, would provide deeper insight into its conflict-detection and resolution capabilities. Third, the current symbolic constraints rely on empirically defined thresholds. Adaptive or data-driven thresholds could further improve the robustness under varying data quality conditions. Finally, tighter integration with BIM may enable more explicit reasoning about spatial

dependencies, thereby enhancing the system's ability to interpret fine-grained execution workflows.

Overall, the results suggest that the neuro-symbolic integration offers a promising and practical pathway towards scalable, interpretable, and optionally safe automation of construction schedule management. By balancing semantic adaptability with explicit domain constraints, the proposed framework establishes a foundation for future research on intelligent, learning-driven planning systems in construction.

Acknowledgements

The authors would like to thank Mr. Divyansh Gupta for his extensive support in data collection and coordination with the project field team. The authors also gratefully acknowledge the Indian Institute of Technology Bombay for facilitating access to project data and supporting this research.

References

Adekunle, P., Aigabvboa, C., Thwala, W., Akinradewo, O., & Oke, A. (2022). Challenges confronting construction information management. *Frontiers in Built Environment*, *8*, 1075674. doi: 10.3389/fbuil.2022.1075674

Al-Sinan, M. A., Bubshait, A. A., & Aljaroudi, Z. (2024). Generation of Construction Scheduling through Machine Learning and BIM: A Blueprint. *Buildings*, *14*(4), 934. doi: 10.3390/buildings14040934

Amer, F., Jung, Y., & Golparvar-Fard, M. (2021). Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction. *Automation in Construction*, *132*, 103929. doi: 10.1016/j.autcon.2021.103929

Amer, F., Jung, Y., & Golparvar-Fard, M. (2023). Construction schedule augmentation with implicit dependency constraints and automated generation of lookahead plan revisions. *Automation in Construction*, *152*, 104896. doi: 10.1016/j.autcon.2023.104896

Cai, L., Yu, C., Kang, Y., Fu, Y., Zhang, H., & Zhao, Y. (2025). Practices, opportunities and challenges in the fusion of knowledge graphs and large language models. *Frontiers in Computer Science*, *7*, 1590632. doi: 10.3389/fcomp.2025.1590632

Golparvar-Fard, M., Peña-Mora, F., & Savarese, S. (2015). Automated Progress Monitoring Using Unordered Daily Construction Photographs and IFC-Based Building Information Models. *Journal of Computing in Civil Engineering*, *29*(1), 04014025. doi: 10.1061/(ASCE)CP.1943-5487.0000205

He, C., Yu, B., Liu, M., Guo, L., Tian, L., & Huang, J. (2024). Utilizing Large Language Models to Illustrate Constraints for Construction Planning. *Buildings*, *14*(8), 2511. doi: 10.3390/buildings14082511

Kautz, H. A. (2022). The third AI summer: AAAI Robert S. Engelmore Memorial Lecture. *AI Magazine*, *43*(1), 105–125. doi: 10.1002/aaai.12036

Li, H., Yang, R., Xu, S., Xiao, Y., & Zhao, H. (2024). Intelligent Checking Method for Construction Schemes via Fusion of Knowledge Graph and Large Language Models. *Buildings*, *14*(8), 2502. doi: 10.3390/buildings14082502

Olawale, Y., & Sun, M. (2015). Construction project control in the UK: Current practice, existing problems and recommendations for future improvement. *International Journal of Project Management*, *33*(3), 623–637. doi: 10.1016/j.ijproman.2014.10.003

Park, J., & Cho, Y. K. (2022). Point Cloud Information Modeling: Deep Learning–Based Automated Information Modeling Framework for Point Cloud Data. *Journal of Construction Engineering and Management*, *148*(2), 04021191. doi: 10.1061/(ASCE)CO.1943-7862.0002227

Project Management Institute. (2021). *A Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Seventh Edition and The Standard for Project Management*. Project Management Institute.

Saka, A. B., & Chan, D. W. M. (2020). Adoption and implementation of building information modelling (BIM) in small and medium-sized enterprises (SMEs): A review and conceptualization. *Engineering, Construction and Architectural Management*, *28*(7), 1829–1862. doi: 10.1108/ECAM-06-2019-0332

Turkan, Y., Bosche, F., Haas, C. T., & Haas, R. (2012). Automated progress tracking using 4D schedule and 3D sensing technologies. *Automation in Construction, Planning Future Cities-Selected Papers from the 2010 eCAADe Conference*, *22*, 414–421. doi: 10.1016/j.autcon.2011.10.003

Wu, C., Li, X., Jiang, R., Guo, Y., Wang, J., & Yang, Z. (2023). Graph-based deep learning model for knowledge base completion in constraint management of construction projects. *Computer-Aided Civil and Infrastructure Engineering*, *38*(6), 702–719. doi: 10.1111/mice.12904

Wu, S., Yao, X., Liu, S., Liang, H., & Liu, Z. (2024). Domain Knowledge Graph Construction Methods of Construction Schedule in Steel Structure Projects. *2024 16th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, 39–44. doi: 10.1109/IHMSC62065.2024.00017

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

ASSESSING THE VIABILITY OF LLM AGENTS FOR GENERATING REUSABLE COMPLIANCE CHECKING FUNCTIONS

Stefan Fuchs[1,2], Sylvain Hellin[1,2], and André Borrmann[1,2]
[1]Chair of Computing in Civil and Building Engineering, Technical University of Munich, Munich, Germany
[2]TUM Georg Nemetschek Institute, Technical University of Munich, Munich, Germany

Abstract

The complexity of building codes and variations in Building Information Modelling (BIM) practices have long challenged Automated Compliance Checking (ACC). The success of Large Language Models (LLMs) in code generation opens new opportunities for this domain. Unlike previous work relying on machine-readable rules or LLMs using hard-coded tools that limit scalability, this paper examines whether LLMs, using the Code-Act pattern, can generate scalable compliance logic for regulations up to Solihin and Eastman's complexity class 3. We show that iterative refinement with a strong verifier evolves reusable checking functions for consistently modelled BIM data, while geometric reasoning requiring auxiliary constructions remains challenging.[1]

Introduction

The digitalization of building permit processes promises faster, more consistent reviews of construction projects. Accelerating these processes is particularly relevant given the pressing demand for housing in many regions, where lengthy approval cycles contribute to construction delays, and increased costs and uncertainty (Bloch and Fauth, 2023). At the center of this effort lies ACC, which refers to the automatic verification of building designs against applicable codes and standards. If realized effectively, ACC offers a dual benefit: architects can pre-check their building models prior to submission, leading to higher-quality designs, while authorities can review submissions faster and more thoroughly.

Eastman et al. (2009) proposed a widely adopted four-step process for automated rule-based checking: (1) rule interpretation, which involves translating regulatory text into a machine-readable format; (2) building model preparation, including semantic enrichment to make information that is implicitly present in the model explicitly available; (3) rule execution, encompassing spatial checks, simple property checks, and more complex calculations; and (4) reporting of compliance results.

The ACC process has seen substantial evolution over time. For example, for rule interpretation, early approaches relied on hard-coded implementations, resulting in opaque black-box systems that were difficult to modify or extend (Solihin, 2004). Subsequent work shifted toward machine-readable intermediate representations, such as the RASE (Requirement, Applicability, Selection, and Exception) semantic markup methodology (Hjelseth and Nisbet, 2011). Such representations offer greater transparency but prove difficult to author and often fail to capture the full complexity of regulatory language. Natural language processing (NLP) techniques were applied to assist this translation, initially, through rule-based methods that required considerable manual effort and were difficult to scale (Zhang and El-Gohary, 2017). Machine learning and deep learning approaches improved scalability across different types of regulatory texts but introduced a dependency on high-quality, labeled training data (Fuchs et al., 2024). LLMs have further reduced data requirements and enabled human-in-the-loop workflows, for example, to apply RASE markup utilising YAML-based representations (Al-Turki et al., 2024). More recent work has explored feeding regulations directly into LLMs for ACC, bypassing intermediate representations entirely (Iversen and Huang, 2026). However, most LLM agent-based approaches required manually authored tool functions that the LLM can invoke, a process that reintroduces significant manual effort. Allowing LLMs to generate the checking code themselves addresses this limitation, but the resulting functions have generally not been reusable across different building models (Madireddy et al., 2025). Achieving reusability would make these functions independently deployable as transparent, reproducible compliance checks, shareable as plain code across practitioners and tools.

In this paper, we assess whether LLM agents can generate reusable compliance checking functions for building regulations with complexity up to Class 3 in the classification by Solihin and Eastman (2015), which require extended data structures such as geometric helper elements. The LLM agent uses the Code-Act architecture (Wang et al., 2024), where the agent interleave reasoning with code generation and execution. By employing a strong verifier, i.e., ground-truth compliance results for multiple Industry Foundation Classes (IFC) models against which generated functions are iteratively tested, we investigate whether this process enables the development of verifiable, deterministic, and reusable compliance checking functions. Accord-

[1]Source code and data available at: `https://github.com/stefan-1992/ACC-function-generation`

ingly, we address the following research questions (RQs):

- **RQ1:** To what extent can LLM agents generate reusable compliance checking functions that generalize across building models?
- **RQ2:** What difficulties arise when scaling across diverse building models?
- **RQ3:** Up to what regulation complexity class can LLM agents reliably generate checking functions?

These questions are investigated through a systematic training and evaluation framework applied to 16 compliance rules across 12 building models. The generated functions are evaluated for correctness and generalization, with a detailed analysis of failure modes and the impact of regulation complexity.

Related work

The challenge of translating building regulations into executable checks has been addressed through two contrasting paradigms. Deterministic, machine-readable representations offer transparency and reproducibility but require substantial manual authoring effort and struggle to capture the full complexity of regulatory language. For example, Solibri, one of the most used commercial tools, provides significant functionality for general model checking and compliance checking. Existing rule templates can be parametrised, but checks beyond the available templates require custom Java development through Solibri's API. The inner workings of built-in rules remain a black box, and the proprietary format makes keeping rules up-to-date with regulatory amendments difficult (Solibri Inc., 2026). Moreover, many regulations cannot easily be transformed into parameterised rule templates or other machine-executable formats, particularly when they involve complex spatial relationships or context-dependent conditions (Hjelseth and Nisbet, 2011). Information availability poses an additional challenge: approaches such as Singapore's IFC-SG standard (Government of Singapore, 2026) specify precise information requirements, validated through Information Delivery Specifications (IDS), but this shifts the burden onto the modeller. Semantic enrichment can partially alleviate this burden by automatically deriving missing information from the model (Bloch, 2022), yet it remains an open research problem. LLM-based approaches offer a more dynamic alternative that can work directly with natural language regulations and heterogeneous model data, yet they often lack the determinism and reusability needed for production use.

Agent-based approaches to automated compliance checking

Several recent studies have explored LLM-based agents for ACC. Ying and Sacks (2024) proposed a framework with different LLM techniques, such as Retrieval-Augmented Generation (RAG), ReAct (reasoning–action) agents, and tool usage, which cover the full ACC pipeline: retrieving relevant regulations from PDF documents, executing compliance checks through switching between reasoning and action phases with predefined tool functions, and resolving non-compliance issues. While this approach demonstrates the feasibility of end-to-end agent-based checking, it relied on manually authored tool functions, and the compliance checking was only tested on one building requirement.

Zheng et al. (2026) and Iversen and Huang (2026) showed that LLMs can identify which predefined functions to orchestrate for a given rule check, but in Iversen and Huang (2026) only a limited number of functions were actually implemented, restricting the approach's practical applicability. While Zheng et al. (2026) used LLMs for function implementation in a case study, they did not systematically evaluate the LLMs' capability for this task. Shi et al. (2025) explored fine-tuning LLMs specifically for automated code compliance, indicating that domain adaptation can improve checking accuracy. However, they did not generate fully executable functions but high-level function calls, which removes the LLM's ability to receive feedback from the rule execution and adjust the generated functions accordingly. Madireddy et al. (2025) advanced the concept by having LLMs generate Python scripts for compliance checking; however, the scripts required manual transfer to Revit and human-in-the-loop verification, limiting automation. Furthermore, several rule-specific hints were required, implying extensive manual analysis.

BIM question answering as a foundation

A related field is BIM question answering, where LLM agents generate code to extract information from building models. Zheng and Fischer (2023) introduced BIMS-GPT, a virtual assistant that dynamically generates queries against BIM data. However, this approach is limited to information directly available in the BIM database. Hellin et al. (2025) developed a multi-agent system for natural language-based BIM question answering. The agents can invoke multiple IfcOpenShell-based tools to explore the BIM model until it has enough information gathered to answer the question. Nithyanantham et al. (2025) proposed MCP4IFC (Model Context Protocol), providing tools to query, edit, and create IFC elements, including dynamic code generation when the existing toolkit is insufficient. Li et al. (2025) demonstrated an LLM agent capable of retrieving data from multiple sources and generating code for spatial queries on building models. These systems address similar technical challenges to ACC, i.e., interpreting domain-specific queries and invoking functions or generating executable code, but typically evaluate results on individual models, where the logic can be adapted to the model at hand. Furthermore, the geometric complexity of such queries is typically lower than in ACC.

Our work extends tool generation to compliance checking functions that must generalize across multiple building models, necessitating a more scalable evaluation framework. Crucially, LLMs are used only during function generation, not at inference time: the resulting checking functions are deterministic, reusable, and execute without LLM

involvement, combining the authoring flexibility of LLMs with the auditability of conventional rule logic.

Methodology

We adopt a machine learning-inspired training strategy in which LLM agents iteratively write, refine, and validate compliance checking functions operating on BIM models in IFC format (Figure 1), which are then tested on unseen models. This strategy is inspired by the evolutionary code optimization strategy of AlphaEvolve (Novikov et al., 2025), where a strong verifier guides iterative improvement. At the center of the approach lies the function creation, which draws on the Code-Act pattern to iteratively generate and refine checking functions.

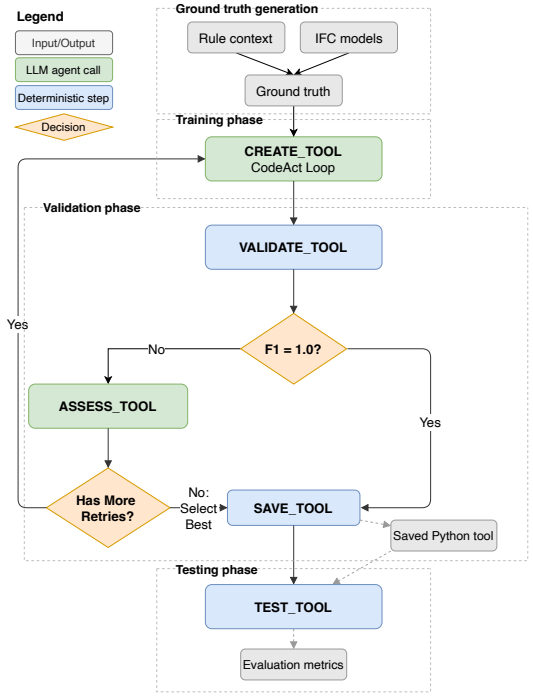


Figure 1: Overview of the agent pipeline for generating compliance checking functions.

Ground truth generation

While in AlphaEvolve, a test function was used to evaluate the generated code, our framework relies on ground-truth compliance results for each building model. These can be produced manually, through coded reference implementations, or using existing rule-checking tools, as long as each element receives a definitive pass or fail label.

The dataset should comprise at least three building models to support separate training, validation, and test splits, with violations present across most or all models, and a mix of compliant and non-compliant elements. Having more BIM models with rule violations increases generalisability of the generated functions, as the agent must handle diverse geometric and semantic configurations rather than overfitting to a single case. We use the Globally Unique Identifier (GUID) to identify and compare violating building elements, producing a deterministic binary signal per element that remains stable throughout the experiment.

Training phase

The training phase begins by selecting a *primary* training BIM model, selected as the model containing the highest number of non-compliant issues, to provide the most informative training model as context for testing the developed functions. The code-generation agent (CREATE_TOOL) then receives the natural-language rule text, including any relevant thresholds and exceptions, along with the expected answers for each training BIM model (Figure 2). It synthesizes a candidate checking function in Python, using `IfcOpenShell` for IFC access and `trimesh`/`shapely` for geometric operations where needed.

Rule: This rule checks that doors and windows in the model are located in the same floor as the wall they are related to. The rule also checks that the model does not contain any orphan doors or windows (a door or window without a relation to any wall).

Parameters: Check Orphan Doors and Windows: True

Question: What doors or windows in the current IFC model are not located in the same floor as the wall they are related to, or are orphan doors/windows without a relation to any wall? Return a list of IFC GUIDs of all elements that violate this rule.

Ground Truth:

`Model: 146` – pass (no violations, return `[]`)

`Model: duplex (primary)` – 2 violations:

Required GUIDs: `[1hOS...lSFK, 1hOS...lSDm]`

Issue 1: *''Door.1.7 And Wall.0.10'': Door components are included in Level 2, but Wall components are included in Level 1. Door components should be included in the same floor as the component they belong to.*

Issue 2: *''Door.1.8 And Wall.0.13'': [...]*

Figure 2: Condensed example of the agent prompt for the ''doors and windows'' rule.

During the code-generation phase, visualised in Figure 3, the agent operates in a CodeAct loop of up to $n_{\text{max_iter}}$ iterations. The agent's task is to iteratively develop a function that identifies the same violations as the ground truth for all training models. At each iteration, it returns either a *CodeAction* or the final *ACC Function*. The *CodeAction* contains reasoning traces and Python code that is executed in a Python sandbox where imports and runtime tools are pre-loaded. Multiple BIM models are used for training to encourage generalization from the outset, ensuring that the function does not overfit to a single building's characteristics. The agent commonly developed strategies to first test functions on only the primary model, before including test cases for all training models. The agent can decide at any point to return the final implementation of the *ACC Function*, which terminates the loop. On the last permitted iteration (i.e., $n_{\text{max_iter}}$), the agent is prompted to return its best implementation. Before proceeding to validation, a pre-check executes the function against all training models to verify the absence of runtime errors; if it crashes, the agent re-enters the loop with the error log appended (up to three consecutive pre-check failures were permitted).

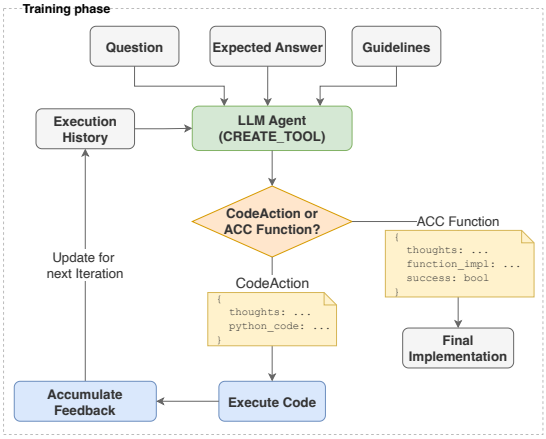


Figure 3: Overview of the code act training loop for generating compliance checking functions.

The agent's prompt specifies a function contract; the generated function must accept a single IFC file path and return a list of non-compliant element GUIDs. Beyond the geometric libraries, the agent has access to `query_ifcopenshell_docs` as runtime tool, which retrieves up-to-date IfcOpenShell documentation and code examples for a given topic.

Validation phase

Once the agent has converged on a function during training, the function is executed against a separate set of validation building models. A dedicated assessment agent interprets the training and validation results. This agent receives the tool's source code, its execution logs, aggregated performance metrics (true-positive, false-positive, and false-negative counts), and a comparison of expected versus predicted element counts. We use an LLM because the diagnosis requires correlating source code logic with per-model error distributions, which involves too many interacting factors for simple heuristic rules. The agent determines the likely cause of failure from five diagnostic categories (overfitting, missing generalisation, implementation bug, model difference, or unknown) and outputs a concrete, GUID-free improvement hint for the code-generation agent. Critically, the assessment agent is prompted not to disclose ground-truth results from the validation set to the code-generation agent; only this high-level diagnostic feedback is provided, discouraging information leakage. Exposing validation GUIDs would allow the code-generation agent to reverse-engineer element-specific properties, or worse, hard-code the GUIDs, undermining generalisation beyond the training distribution.

The assessment result is checked against a retry budget (default: 2 retries). If retries remain and the diagnosis suggests a fixable error, the improvement hint is injected into the next CREATE_TOOL prompt. If the budget is exhausted or the error is deemed negligible, the best-performing implementation is saved. This mechanism bounds computational cost while still allowing targeted refinement.

Testing phase

Functions that pass validation are evaluated on a held-out test set of building models, measuring how well the generated functions transfer to unseen buildings. The separate test set ensures that neither the agent's prompts nor the tooling are optimised against it. We apply this process to 16 compliance rules spanning multiple complexity classes as detailed in the next section.

Experimental setup

We evaluate our approach using 12 building models, divided equally into training, validation, and test models (4 models each), as shown in Table 1. All models are freely available: three originate from the GNI BIM Dataset - 2025_BIMfundamentals (Wang et al., 2026), and the remaining models are sourced from IFCBench (Hellin et al., 2025). The data split ensures that each compliance rule exhibits non-compliant elements in the training set, with test and then validation sets prioritised where possible.

Table 1: Building models used in each data split.

Training	Validation	Test
146*	106*	4351
AC20	172*	Digital Hub
Dental Clinic	FZK House	S. MacAlister
Duplex	Smiley West	WBDG Office

*From GNI BIM Dataset (Wang et al., 2026).

Ground-truth compliance labels are produced using Solibri (Solibri Inc., 2026). We executed predefined rule sets against each IFC model using Solibri's autorun. From the resulting BIM Collaboration Format (BCF) reports, we identified issues with single (or well separable) constraints and defined how to automatically locate them and distinguish the GUIDs that violate the rule from contextual reference elements. Where rules depend on semantic classifications (e.g., space usage types), we applied them through importable Solibri pattern rules. To keep the agent on equal footing with the verifier, we provided it with the `classify_spaces` tool, which assigns semantic labels (e.g., circulation, habitable) to IFC spaces by matching names against the predefined patterns.

Compliance rules

We select 16 compliance rules spanning the complexity classes defined by Solihin and Eastman (2015), summarized in Table 2. These range from simple property checks (Class 1) through rules requiring derived attributes or geometric relationships among multiple objects (Class 2) to rules requiring extended data structures and geometric helper elements (Class 3). In practice, Class 1 and 2 rules may behave as higher-complexity checks when the required properties or relations are absent from the IFC model and must be derived geometrically.

Table 2: Compliance rules grouped by Solihin & Eastman complexity class, with non-compliant element counts across data splits.

Rule	Class	Ruleset	Description	Train	Val	Test
Slab thickness	1	BIM Validation	Slab thickness within 0.30 m–1.0 m range	14	–	3
Riser height	1	ADA/ABA - 504.2	Risers between 100 mm and 180 mm high	5	37	3
Tread length	1	ADA/ABA - 504.2	Treads at least 280 mm deep	2	45	2
Non-uniform risers/treads	2	ADA/ABA - 504.2	Uniform riser heights and tread depths per flight	2	21	–
Stair–slab connection	2	ADA/ABA - 504.2	All stairs connected to slabs	3	–	2
Doors and windows	2	BIM Validation	Doors/windows on same floor as host wall; no orphans	7	8	18
Large spaces ≥ 2 doors	2	Adv. Space Check	Spaces ≥ 50 m^2 have at least 2 doors	22	9	25
Same-storey elevation	2	Gen. Space Check	Spaces in same storey share bottom elevation	180	110	124
Space validation (inside)	2	Gen. Space Check	No components incorrectly inside spaces	17	7	17
Space validation (intersect)	2	Gen. Space Check	Spaces do not intersect with slabs/walls	24	55	22
Circular space	3	ADA/ABA - 304.3.1	Wheelchair turning space ≥ 1525 mm diameter	2	22	2
Clear floor space	3	ADA/ABA - 305.3	Clear floor/ground ≥ 760 × 1220 mm	18	23	1
Two doors in series	3	ADA/ABA - 404.2.5	≥ 1220 mm between hinged/pivoted doors in series	8	–	2
Clearance front of doors	3	BIM Validation	Sufficient clearance on both sides of doors	28	30	11
Slabs guarded against falling	3	BIM Validation	Slab edges protected by barriers or acceptable drop	34	11	8
Unallocated areas	3	BIM Validation	Floor area must be assigned to a space ≤ 0.50 m^2	6	–	23

Evaluation metrics

Evaluation is performed at the building element level using unique element identifiers (i.e., GUIDs). For each model, we compare the set of GUIDs returned by the generated ACC function against the ground-truth set. For violations involving multiple elements, each GUID is counted once. We compute precision, recall, and F1 score per model. The *average* F1 score ($F1_{\text{avg}}$) is the arithmetic mean of the per-model F1 scores (macro-average over models). As a special case, if a model is fully compliant (no non-conforming elements) and the function correctly identifies it as such, this yields Precision, Recall, and F1 score of 1.0. As an alternative, the *aggregated* F1 score ($F1_{\text{agg}}$) is computed by pooling all true-positive, false-positive, and false-negative counts across models before calculating precision, recall, and F1 score (micro-average over models). This provides a more balanced measure, as fully compliant models receive proportionally less weight compared to $F1_{\text{avg}}$. In Table 3, per-class and overall averages are macro-averaged over the rules within each class and over all 16 rules, respectively.

Agent configuration

The agents are powered by GLM-4.7 by Z.ai[2], chosen for its competitive performance on SWE-Bench Verified[3] and favourable performance-to-cost ratio. The function creation agent is permitted up to 15 internal coding iterations per attempt ($n_{\text{max_iter}} = 15$). The validation loop allows up to 2 feedback-driven retries before the best implementation is saved. The agent has access to `IfcOpenShell` for IFC model traversal, and `shapely` and `trimesh` for 2-D and 3-D geometric operations, respectively. The libraries are preloaded and briefly described within the prompt. Documentation retrieval is currently only supported for IfcOpenShell; extending this to cover `trimesh` and `shapely` is left for future work.

[2]https://z.ai/blog/glm-4.7

[3]https://openai.com/index/introducing-swe-bench-verified/

Results

Table 3 summarises the per-rule performance across the three data splits. We use the aggregated F1 score ($F1_{\text{agg}}$) as the primary evaluation metric. Overall, the LLM agent shows moderate capability for implementing ACC functions, with $F1_{\text{agg}}$ values of 0.55 on training, 0.54 on validation, and 0.40 on test. The near-identical training and validation scores suggest that the validation feedback loop generalises to the validation models. The subsequent drop on the unseen test set, however, indicates that many generated functions still overfit to assumptions about how geometry, properties, and relationships are represented in the training and validation IFC models. Performance varies by complexity class. Class 1 rules achieved near-perfect validation scores ($F1_{\text{agg}} = 0.96$), yet dropped sharply on the test set (0.38). Class 2 rules achieved the strongest test results ($F1_{\text{agg}} = 0.57$). The strong Class 2 average is driven by three rules that generalised fully to the test set, whereas the weak Class 3 average (0.21) reflects the continued difficulty of rules that require auxiliary geometric constructions or richer rule logic.

Rules that achieved perfect or near-perfect training scores showed varying degrees of degradation on test, from none (*doors and windows*, *same-storey elevation*, and *stair–slab connection*: $1.00 \rightarrow 1.00$) to complete failure (*slab thickness*: $1.00 \rightarrow 0.00$). The validation feedback loop improved some rules: for *tread length*, the validation $F1_{\text{agg}}$ reached 0.89, and the function still achieved 0.67 on test. For more complex rules, however, the feedback loop could not compensate for missing geometric detail, incomplete IFC relationships, or ambiguous rule semantics. The gap between $F1_{\text{avg}}$ and $F1_{\text{agg}}$ on the test set (0.59 vs. 0.40) suggests that performance remains uneven across models, with several rules receiving inflated average scores because some test models are fully compliant.

Table 4 shows the execution metrics for each rule. Total duration ranges from 4.3 minutes (slab thickness) to 56.9 minutes (clear floor space). Token usage spans 116 k

Table 3: Per-rule $F1$ scores across data splits. $F1_{\text{avg}}$: macro-average over models; $F1_{\text{agg}}$, P_{agg}, R_{agg}: micro-average over models (pooled counts). Class and overall rows: macro-average over rules.

		Training				Validation				Test			
Rule	**Cl.**	$F1_{\text{avg}}$	P_{agg}	R_{agg}	$F1_{\text{agg}}$	$F1_{\text{avg}}$	P_{agg}	R_{agg}	$F1_{\text{agg}}$	$F1_{\text{avg}}$	P_{agg}	R_{agg}	$F1_{\text{agg}}$
Slab thickness	1	1.00	1.00	1.00	1.00	1.00†	1.00†	1.00†	1.00†	0.75	0.00	0.00	0.00
Riser height	1	0.75	1.00	0.67	0.80	1.00	1.00	1.00	1.00	0.54	0.30	1.00	0.46
Tread length	1	0.75	0.33	1.00	0.50	0.50	1.00	0.80	0.89	0.92	0.50	1.00	0.67
Class 1 avg.	1	*0.83*	*0.78*	*0.89*	*0.77*	*0.83*	*1.00*	*0.93*	*0.96*	*0.74*	*0.27*	*0.67*	*0.38*
Non-uniform risers/treads	2	1.00	1.00	1.00	1.00	0.50	0.00	0.00	0.00	0.75†	0.00†	0.00†	0.00†
Stair–slab connection	2	1.00	1.00	1.00	1.00	1.00†	1.00†	1.00†	1.00†	1.00	1.00	1.00	1.00
Doors and windows	2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Large spaces ≥ 2 doors	2	0.62	0.64	0.32	0.42	0.75	1.00	0.89	0.94	0.14	0.80	0.16	0.27
Same-storey elevation	2	1.00	1.00	1.00	1.00	0.75	1.00	0.24	0.38	1.00	1.00	1.00	1.00
Space validation (inside)	2	0.57	0.43	0.18	0.25	0.50	0.00	0.00	0.00	0.40	0.60	0.18	0.27
Space validation (intersect)	2	0.20	0.21	0.88	0.34	0.25	0.28	0.35	0.31	0.50	0.30	1.00	0.46
Class 2 avg.	2	*0.77*	*0.75*	*0.77*	*0.72*	*0.68*	*0.61*	*0.50*	*0.52*	*0.68*	*0.67*	*0.62*	*0.57*
Circular space	3	0.75	0.40	1.00	0.57	0.75	1.00	0.05	0.09	1.00	1.00	1.00	1.00
Clear floor space	3	0.50	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.50	0.00	0.00	0.00
Two doors in series	3	0.50	0.00	0.00	0.00	1.00†	1.00†	1.00†	1.00†	0.75	0.00	0.00	0.00
Clearance front of doors	3	0.25	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Slabs guarded against falling	3	0.45	0.51	0.87	0.64	0.23	0.04	0.40	0.08	0.23	0.21	0.43	0.29
Unallocated areas	3	0.35	0.20	0.33	0.25	1.00†	1.00†	1.00†	1.00†	0.00	0.00	0.00	0.00
Class 3 avg.	3	*0.47*	*0.18*	*0.37*	*0.24*	*0.62*	*0.51*	*0.41*	*0.36*	*0.41*	*0.20*	*0.24*	*0.21*
Overall avg.	1-3	**0.67**	**0.55**	**0.64**	**0.55**	**0.69**	**0.65**	**0.54**	**0.54**	**0.59**	**0.42**	**0.49**	**0.40**

† All models were compliant; the scores reflect false-positive control only.

(slab thickness) to 2.4 M (clearance front of doors), with geometry-heavy Class 2 and 3 rules consistently consuming the most time and tokens. *Doors and windows* converged quickly and achieved perfect scores across all three splits, suggesting that the agent performs well when the rule can be expressed through relatively direct relationship checks. By contrast, *clear floor space* (56.9 min), *space validation (intersect)* (45.4 min), and *clearance front of doors* (45.3 min) were among the longest-running rules, reflecting the cost of repeated geometric exploration without strong generalisation.

Discussion

The results indicate that reusable compliance checking functions can be generated when IFC semantics are consistent, but generalisation remains fragile across heterogeneous models. The perfect test performance of *doors and windows*, *same-storey elevation*, and *stair–slab connection* shows the approach is viable when the relevant semantics are explicitly represented and stable across models. The broader decline from training to test suggests that many functions encode assumptions about properties, relationships, or geometry that do not transfer reliably. Manual error analysis of the test set explains this gap. Geometry-based heuristics were the dominant failure mode, affecting seven rules: *clearance front of doors*, *slabs guarded against falling*, *space validation (inside)*, *space validation (intersect)*, *non-uniform risers/treads*, *clear floor space*, and *unallocated areas*. The agent often produced plausible but coarse approximations based on bounding boxes, convex hulls, perimeter sampling, or random point placement. Such heuristics succeed on training models but may fail on unseen IFC geometries, where Solibri uses richer geometric reasoning and internal tolerances.

Relationship traversal is a second major source of error. The clearest examples are *large spaces ≥ 2 doors* and *two doors in series*, where the generated functions rely on `IfcRelSpaceBoundary` to associate spaces, walls, and doors. When these relationships are incomplete or absent, the functions fail to identify the relevant door pairs or counts. Property lookup is similarly brittle: *slab thickness* fails because the relevant property is stored under the German label `Dicke`, while *riser height* is affected by inconsistent property interpretation, for example when `Pset_StairCommon.RiserHeight` reports 0.55 m while the Revit-derived maximum riser height is 0.18 m. These examples show that the core challenge is not only code generation, but also learning robust strategies for handling heterogeneous IFC conventions.

Many of these issues could be reduced by increasing the number and diversity of training and validation models. A larger dataset would expose the agent to more variants of relationship structures, property naming conventions, unit systems, and geometric edge cases, making it less likely to overfit to a single export implementation. Realising that benefit would require longer exploration budgets, since the agent would need more iterations to inspect a broader range of failure cases. At the same time, scaling to a broader model set would require automating Solibri object classification, which is currently provided through manually prepared pattern rules. Without such automation, the approach does not scale cleanly to new models

Table 4: Execution metrics per compliance rule: retry count, coding iterations per attempt, and time/token breakdown.

		Iterations per attempt			Agent (min)		Execution (min)			
Rule	**Retries**	**Att_0**	**Att_1**	**Att_2**	**CREATE**	**ASSESS**	**Tr+Val**	**Test**	**Total (min)**	**Tokens**
Slab thickness	0	9	–	–	3.1	0.0	0.8	0.4	4.3	116,497
Riser height	0	14	–	–	11.4	0.0	0.7	0.4	12.5	328,637
Tread length	2	4	6	7	6.5	1.0	2.2	0.4	10.1	163,662
Non-uniform risers/treads	2	9	15	12	14.8	0.2	1.5	0.4	16.9	373,004
Stair–slab connection	2	13	15	7	26.4	2.3	2.4	0.4	31.5	511,586
Doors and windows	0	15	–	–	9.4	0.0	0.8	0.4	10.5	430,186
Large spaces ≥ 2 doors	2	13	15	15	15.6	0.9	2.3	0.4	19.2	920,618
Same-storey elevation	2	12	15	15	17.6	0.4	2.3	0.4	20.7	1,465,624
Space validation (inside)	2	2	15	7	9.8	1.1	3.9	0.6	15.4	579,809
Space validation (intersect)	2	13	15	15	38.4	0.9	5.4	0.6	45.4	1,435,136
Circular space	2	6	11	13	37.4	3.2	2.4	0.4	43.4	697,468
Clear floor space	2	15	12	15	48.2	5.9	2.4	0.4	56.9	1,131,448
Two doors in series	0	15	–	–	10.4	0.0	1.3	0.6	12.3	249,662
Clearance front of doors	2	15	15	15	33.4	1.0	10.5	0.4	45.3	2,378,279
Slabs guarded against falling	2	15	15	15	30.8	0.4	4.2	2.1	37.5	747,190
Unallocated areas	1	15	15	–	11.6	0.3	2.0	0.5	14.3	1,033,179

because both the ground-truth generation and the agent's auxiliary classification tool depend on handcrafted semantic mappings.

Geometric reasoning poses a barrier that current LLMs cannot overcome reliably. The weak Class 3 scores confirm this. The circular space result demonstrates that geometry per se is not the barrier: because fitting a circle inside a polygon requires no model-specific assumptions, the agent achieved a perfect test score ($F1_{\text{agg}} = 1.00$) despite the check's geometric nature. In contrast, rules that require helper elements such as door swing arcs, clearance zones, or barrier envelopes remain difficult, and the resulting implementations often fall back on heuristics that are too coarse to match Solibri consistently.

Another difficulty is that several Solibri checks combine multiple sub-rules within a single template. We partly addressed this by passing rule parameters to the agent, but their exact operational meaning often remains opaque from the prompt alone. This is visible in *space validation (inside)*, where Solibri flags both, components inside spaces and cases where a space does not properly touch the slab or surface below itself. Access to Solibri rule documentation via RAG could help the agent interpret these composite checks and checking parameters more faithfully and reduce the need to infer hidden sub-rules from sparse examples. Without visual feedback on intermediate geometry, the agent cannot verify whether helper elements such as clearance zones are correctly positioned.

We note several limitations. The 12-model evaluation limits statistical robustness. The GUID-level evaluation counts each element once, regardless of its involvement in multiple violations of the same type, potentially overestimating quality for rules that affect many similar elements. The ground truth was established through Solibri rule checks; differences in rule interpretation between the ground truth and the generated ACC functions cannot be excluded. A future dataset should include checks from multiple sources and different methods for generating (or manually creating) the ground-truth violations. Semantic labels for spaces and doors had to be provided via the `classify_spaces` tool, applying pre-defined name-matching patterns. Simple name-matching is effective in controlled settings but produces false positives when naming differs across building types, and introduces a manual dependency that does not scale to new models. Automating object classification is therefore important for scaling the evaluation pipeline and giving the agent access to consistent semantic categories. Integrating more dynamic components, such as semantic similarity or auxiliary agents for edge cases, might improve generalisation but would also weaken the deterministic character of the generated functions, which is central to their reusability across design tools.

Conclusions

This study investigated whether LLM-based agents using the Code-Act paradigm can generate reusable compliance checking functions from building regulations. The results show that such functions are achievable for a subset of Class 1 and Class 2 rules when IFC representations are consistent and the required relationships or properties are explicitly available. However, checks that depend on geometry-heavy heuristics, incomplete relationship traversal, non-standard property lookup, or composite Solibri rule semantics remain difficult, especially for Class 3 rules.

Several directions for future work emerge from these findings. First, larger and more diverse datasets are needed so that the agent can cope with variation in IFC exports, property names, unit conventions, and relationship completeness. Therefore, Solibri object classification should be automated so that semantic labels can be generated consistently at scale rather than supplied through manual pattern rules. Additionally, a curriculum-based generation strategy, in which agents reuse and compose previously veri-

fied functions, might reduce the search space for complex rules. Finally, the generated functions can be integrated into design tools and further enable extended use cases, including reinforcement learning-based design adaptations.

Acknowledgments

We thank the TUM Leonhard Obermeyer Center for financially supporting this research.

References

Al-Turki, D., Hettiarachchi, H., Gaber, M. M., Abdelsamea, M. M., Basurra, S., Iranmanesh, S., Saadany, H., and Vakaj, E. (2024). Human-in-the-loop learning with LLMs for efficient RASE tagging in building compliance regulations. IEEE Access, 12:1–1.

Bloch, T. (2022). Connecting research on semantic enrichment of BIM - review of approaches, methods and possible applications. J. Inf. Technol. Constr., 27(20):416–440.

Bloch, T. and Fauth, J. (2023). The unbalanced research on digitalization and automation of the building permitting process. Adv. Eng. Inform., 58(102188):102188.

Eastman, C., Lee, J.-M., Jeong, Y.-S., and Lee, J.-K. (2009). Automatic rule-based checking of building designs. Autom. Constr., 18(8):1011–1033.

Fuchs, S., Dimyadi, J., Witbrock, M., and Amor, R. (2024). Intermediate representations to improve the semantic parsing of building regulations. Advanced Engineering Informatics, 62:102735.

Government of Singapore (2026). CORENET X Code of Practice. 3.1 edition.

Hellin, S., Nousias, S., and Borrmann, A. (2025). Natural Language Information Retrieval from BIM Models: An LLM-Based Agentic Workflow Approach. In Proc. of the European Conference on Computing in Construction (EC3).

Hjelseth, E. and Nisbet, N. (2011). Capturing normative constraints by use of the semantic mark-up RASE methodology. Proceedings of CIB W78-W102 conference.

Iversen, O. and Huang, L. (2026). Leveraging large language models for BIM-based automated compliance checking. Automation in Construction, 182:106707.

Li, A., Wong, P. K.-Y., Tao, X., Ma, J., and Cheng, J. C. P. (2025). An interactive system for 3D spatial relationship query by integrating tree-based element indexing and LLM-based agent. Adv. Eng. Inform., 66(103375):103375.

Madireddy, S., Gao, L., Din, Z. U., Kim, K., Senouci, A., Han, Z., and Zhang, Y. (2025). Large Language Model-driven code compliance checking in Building Information Modeling. Electronics (Basel), 14(11):2146.

Nithyanantham, B. K., Sesterhenn, T., Nedungadi, A., Garijo, S. P., Zenkner, J., Bartelt, C., and Lüdtke, S. (2025). MCP4IFC: IFC-Based Building Design Using Large Language Models. arXiv:2511.05533 [cs].

Novikov, A., Vũ, N., Eisenberger, M., Dupont, E., Huang, P.-S., Wagner, A. Z., Shirobokov, S., Kozlovskii, B., Ruiz, F. J. R., Mehrabian, A., Kumar, M. P., See, A., Chaudhuri, S., Holland, G., Davies, A., Nowozin, S., Kohli, P., and Balog, M. (2025). AlphaEvolve: A coding agent for scientific and algorithmic discovery. arXiv:2506.13131 [cs].

Shi, J. W. L., Solihin, W., and Yeoh, J. K. W. (2025). Fine-tuning a large language model for automated code compliance of building regulations. Adv. Eng. Inform., 68:103676.

Solibri Inc. (2026). Solibri Office.

Solihin, W. (2004). Lessons learned from experience of code-checking implementation in Singapore. In Building SMART Conference, Singapore.

Solihin, W. and Eastman, C. (2015). Classification of rules for automated BIM rule checking development. Autom. Constr., 53:69–82.

Wang, X., Chen, Y., Yuan, L., Zhang, Y., Li, Y., Peng, H., and Ji, H. (2024). Executable Code Actions Elicit Better LLM Agents. arXiv:2402.01030 [cs].

Wang, Z., Fuchs, S., Wu, J., Esser, S., Wrabel, T., and Borrmann, A. (2026). GNI BIM Dataset. Technical University of Munich, Georg Nemetschek Institute. Zenodo. `https://doi.org/10.5281/zenodo.19722012`.

Ying, H. and Sacks, R. (2024). From Automatic to Autonomous: a Large Language Model-driven Approach for Generic Building Compliance Checking. In Proceedings of the 41st International Conference of CIB W78.

Zhang, J. and El-Gohary, N. M. (2017). Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. Autom. Constr., 73:45–57.

Zheng, J. and Fischer, M. (2023). Dynamic prompt-based virtual assistant framework for BIM information search. Autom. Constr., 155:105067.

Zheng, Z., Han, J., Chen, K.-Y., Cao, X.-Y., Lu, X.-Z., and Lin, J.-R. (2026). Translating regulatory clauses into executable codes for building design checking via large language model driven function matching and composing. Eng. Appl. Artif. Intell., 163(112823):112823.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

A MULTI-AGENT RAG-ENABLED KNOWLEDGE SYSTEM FOR NATIONAL-SCALE WHOLE-LIFE CARBON ASSESSMENT OF RESIDENTIAL RETROFITS

Sobia Bano[1], Usman Ali[1], Conor Shaw[2], Tobias Maile[3], Fatemeh Golpayegani[4] and James O'Donnell[5]
[1]School of Mechanical and Materials Engineering, University College Dublin, Ireland
[2]Department of Sustainability Science, Lappeenranta-Lahti University of Technology, Finland
[3]Technical University of Applied Sciences Augsburg, Germany
[4]School of Computer Science, University College Dublin, Ireland
[5]School of Engineering, University of Limerick, Limerick, Ireland

Abstract

National residential retrofit is essential for meeting climate targets, making Whole-Life Carbon Assessment (WLCA) critical for balancing operational and embodied emissions. However, fragmented datasets and incomplete material inventories limit its application at scale. This study proposes a multi-agent, Retrieval-Augmented Generation (RAG)–enabled Knowledge-Based System (KBS) for national-scale WLCA of residential retrofit scenarios. Domain-specific LLM agents retrieve, interpret, and explain carbon metrics with full data provenance, improving transparency and reducing manual effort. Case studies on Irish semi-detached dwellings achieve 88% feasibility, 81% nDCG, 70% Precision, and 79% Mean Reciprocal Rank, confirming the system's potential for evidence-based national decarbonisation planning.

Introduction

Global population growth and the associated demand for housing and infrastructure continue to increase the built environment's climate impact. Recent demographic projections estimate the world population will approach ∼9.7–9.8 billion by 2050 (INED, 2024). At the same time, international and regional policy frameworks are tightening requirements on building energy and carbon performance. In the European Union, the recast Energy Performance of Buildings Directive (EPBD) has strengthened the pathway toward a zero-emission building stock and explicitly advances whole-life thinking, including requirements to calculate and disclose life-cycle global warming potential for new buildings according to a phased timeline (European Commission, 2025). These developments increase the need for decision-support tools that quantify carbon impacts consistently and transparently across large residential portfolios.

Residential retrofit programmes are widely recognised as an effective near-term strategy for reducing operational emissions in existing housing stocks, yet they also introduce embodied emissions that can materially influence net carbon outcomes over the assessment horizon. National-scale and case-study analyses consistently identify this operational–embodied trade-off, showing that retrofits optimised for operational performance may increase whole-life impacts when embodied effects are excluded (McGinley et al., 2022). Whole-life carbon assessments across diverse residential contexts further indicate that an exclusive focus on operational reductions can shift emissions toward materials and construction activities as retrofit depth and deployment scale increase (Hegarty and Kinnane, 2023). Simulation-based LCA studies of residential retrofits similarly demonstrate that conclusions regarding optimal retrofit strategies are highly sensitive to methodological assumptions, including system boundaries, replacement cycles, grid decarbonisation pathways, and uncertainty treatment (Aghamolaei et al., 2025). The broader retrofit LCA literature highlights persistent challenges in translating existing guidance into routine practice, with analyst judgement and modelling choices continuing to drive substantial variability in reported outcomes (Hurst and O'Donovan, 2025).

National-scale complexity has been addressed through the use of building archetypes, building-stock models, and parametric simulation pipelines. Reviews indicate that archetype-based approaches enable scalable analysis but remain sensitive to data quality, representativeness, and calibration decisions (Shen et al., 2024). Bottom-up building-stock frameworks capture heterogeneity and support targeted strategies when detailed inputs are available, although fragmented and incomplete datasets remain a major limitation (Penaka et al., 2024). Parametric life-cycle approaches enable systematic exploration of retrofit options under uncertainty, but their application at scale depends on robust inventories, consistent life-cycle boundaries, and transparent assumptions (Arbulu et al., 2024).

Recent advances in large language models (LLMs) and Retrieval-Augmented Generation (RAG) frameworks present an alternative pathway for scaling whole-life carbon assessment while maintaining traceability. RAG approaches ground model outputs in retrieved evidence, reducing hallucination risk (Lewis et al., 2020), while multi-agent RAG architectures decompose complex analytical workflows into specialised agents, improving robustness and interpretability (Xie et al., 2025). These developments are increasingly discussed within sustainability assessment research as mechanisms for improving scalability and reducing expert workload, alongside continued emphasis on transparency, auditability, and governance.

This study introduces a novel national-scale WLCA framework for residential retrofits that integrates parametric simulation with a retrieval-augmented, multi-agent knowledge-based system to address challenges posed by

fragmented data, limited transparency, and poor scalability. The core contribution lies in the tight coupling of ISO-consistent operational and embodied carbon accounting with a domain-specific multi-agent RAG architecture, where dedicated Operational, Embodied Carbon, and Whole-Life Carbon Agents retrieve, validate, and synthesize results directly from an authoritative knowledge base, ensuring full numeric grounding and traceable provenance. The framework further advances current practice through an orchestrated interface and interactive dashboard that present complex WLCA results in a clear, explainable form suitable for stakeholders and policy decision-making. This design enables scalable exploration of retrofit pathways, transparent evaluation of operational–embodied trade-offs, and reproducible decision support for residential decarbonization planning at regional and national scales.

The paper follows a structured format, with Section 2 presenting the research methodology. Section 3 examines a comprehensive Irish case study, providing real-world context for the study's findings. Finally, Section 4 concludes the paper and outlines potential avenues for future research.

Methodology

This study proposes a scalable framework for whole-life carbon assessment of national residential building stocks by integrating data-driven archetype modeling, parametric simulation, and standardized life-cycle assessment. Representative dwelling archetypes are constructed from geometric, material, and operational data, with key parameters systematically varied to generate realistic retrofit scenarios and evaluate both operational and embodied carbon impacts. All simulation outputs and life-cycle results are stored in a structured, version-controlled knowledge base that preserves numerical consistency, provenance, and traceability, while a domain-specific, multi-agent retrieval-augmented generation layer enables auditable, constraint-aware querying and transparent trade-off analysis. Together, the framework translates high-dimensional WLCA evidence into explainable, policy-relevant insights and supports robust, reproducible evaluation of residential retrofit strategies at scale.

Data acquisition

The framework establishes an integrated data infrastructure that can be populated from nationally available statistics, open databases, or project-specific surveys. Geometric inputs describe spatial and envelope characteristics such as envelope U-values (W/m^{2}K), floor area (m^2), glazing ratio (WWR in %), and orientation (θ in $^\circ$), which govern heat transfer and solar gains. Non-geometric data capture operational conditions including occupancy schedules, internal gains, ventilation behaviour, and system efficiencies (Ali et al., 2019).

Embodied-carbon assessment relies on material and product-level information drawn from recognized life-cycle inventories and environmental product declarations. Typical sources include international or national databases such as ICE, ecoinvent, CIBSE TM65, or equivalent regional inventories; however, the framework allows substitution with any dataset providing transparent emission factors and system boundaries (Chartered Institution of Building Services Engineers (CIBSE), 2021; Circular Ecology, 2025; Ecoinvent Association, 2025). Climate data consisting of hourly temperature, solar irradiance, and wind profiles are incorporated to enable climate-sensitive operational modelling.

Archetype development

The methodology employs building archetypes representing dominant residential typologies to enable national-scale analysis without modelling individual dwellings. Each archetype encapsulates typical geometry, construction, and operational characteristics derived from national housing statistics and empirical studies. Baseline specifications include envelope assemblies, material composition, thermal performance (U-values in W/m^{2}K), building services, and occupant behaviour assumptions across age bands and dwelling forms. These attributes are formalized within the *ArchetypeData* table of the WLCA Knowledge Base System (KBS), where stable identifiers enable consistent linkage to operational simulations, embodied material records, and scenario states across the pipeline (Ali et al., 2019).

Archetypes are implemented in validated dynamic building simulation environments, such as EnergyPlus or IES VE, or equivalent tools capable of hourly energy modelling (Erdogan, 2019). Simulation outputs are benchmarked against national or measured datasets to ensure realistic performance. Once validated, archetypes function as flexible templates for systematic variation of materials, systems, and operating conditions in retrofit and whole-life carbon analyses.

Operational and embodied parametric feature selection

Parametric feature selection defines the multidimensional design space explored within the WLCA framework. Parameters are identified through literature review, data availability, and preliminary sensitivity testing, rather than expert judgement alone. Selected variables demonstrably influence energy demand, thermal behaviour, or material-related impacts and are registered in the WLCA Parametric KBS with metadata specifying domain relevance, allowable ranges, default values, and dataset versions.

Parameters are categorized into geometric, operational, and material groups. Geometric variables influence envelope exposure and solar gains; operational parameters capture use patterns and system performance; material parameters affect heat transfer and embodied carbon intensity. Together, these parameters bound the parametric space feeding simulation, LCA calculation, and retrieval. Centralized storage enables the *Interface Agent*

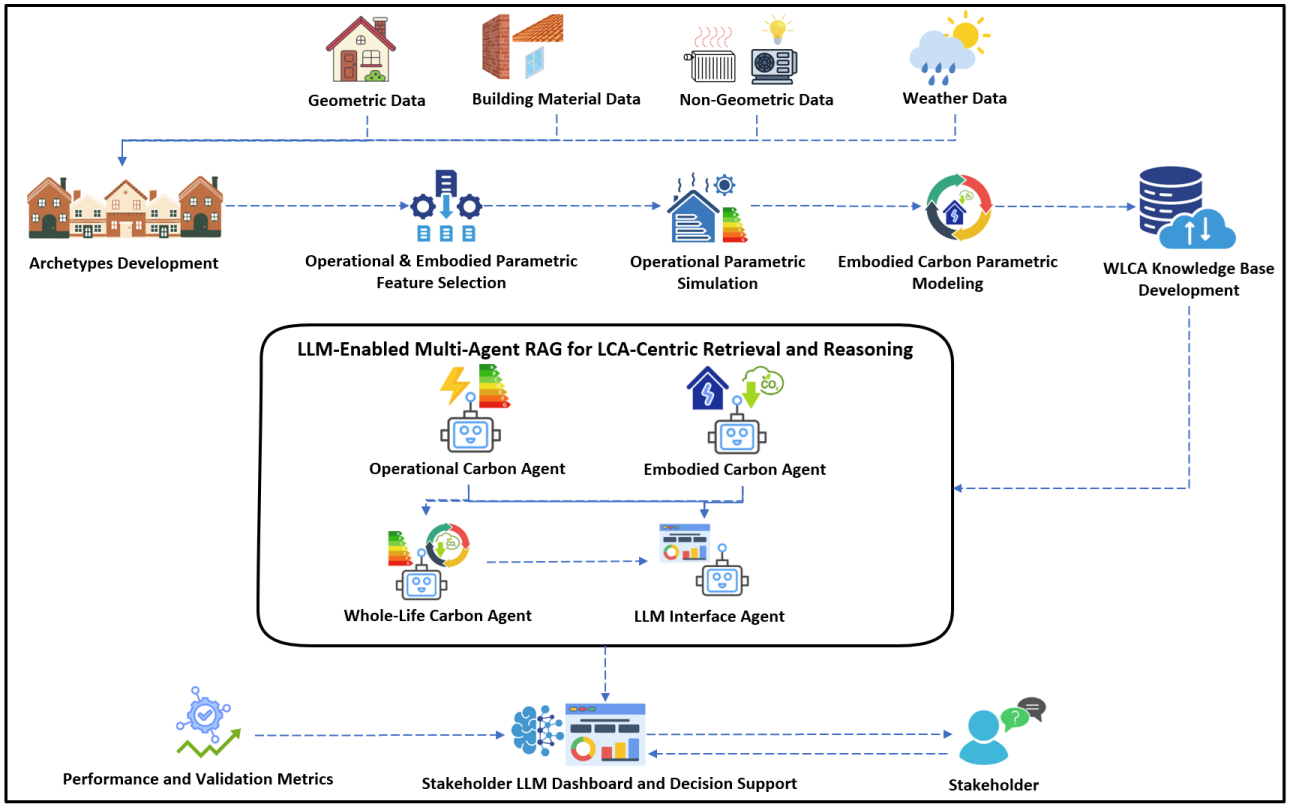


Figure 1: Methodology for Multi-Agent RAG-Enabled Knowledge System for national-Scale Whole-Life Carbon Assessment of Residential Retrofits.

and *Query Interpretation* process to consistently translate natural-language queries into structured retrieval constraints.

Operational parametric simulation

Operational simulation evaluates the effects of variations in form, systems, and building type on energy demand through an ensemble of scenarios generated by perturbing selected parameters. High-throughput execution is achieved using parametric automation tools such as jEPlus, Python-based workflows, or equivalent batch-processing environments (Zhang and Korolija, 2010). Each simulation is associated with a specific archetype identifier, parameter set, climate file, and scenario state, with results stored in the *OperationalData* table of the KBS.

To manage computational cost in high-dimensional spaces, statistical sampling methods such as Simple Random Sampling, Latin Hypercube Sampling, or Sobol sequences are employed. Simulation outputs include annual energy use intensity, heating and cooling loads, system consumption, and comfort metrics. These results form the primary evidence used by the *Operational Agent* within the RAG architecture. Post-processing normalizes outputs and structures them for integration with WLCA calculations, producing a reusable, queryable synthetic dataset that supports consistent reuse without re-running simulations.

Embodied carbon parametric modeling

For each parametric configuration, a cradle-to-grave life-cycle assessment is conducted in accordance with ISO 14040/14044 standards (ISO, 2006). The assessment follows the A–C life-cycle structure, covering product and construction stages (A1–A5), operational energy use (B6), and end-of-life processes (C1–C4). Lifecycle boundaries and assessment horizons are aligned within the WLCA Parametric KBS to ensure consistent integration of embodied and operational impacts by the *Whole-Life Carbon Agent*.

Material emission factors are sourced from established inventories such as ICE, ecoinvent, and CIBSE TM65, or equivalent national datasets. Where project-specific data are unavailable, transparent proxy selection is permitted based on material function and composition, with uncertainty and provenance explicitly recorded. All material quantities, emission factors, and uncertainty annotations are persisted in the *EmbodiedData* table, enabling full auditability during retrieval.

An automated calculation pipeline links simulation outputs, material quantities, and emission factors to compute embodied and operational impacts across large scenario sets. Results are expressed as total life-cycle emissions ($kgCO_2e$) and normalized by floor area to enable

cross-scenario comparison. This integrated dataset underpins the knowledge base and is consumed by the *Whole-Life Carbon Agent* to compute total WLCA values, savings, and baseline–post-retrofit comparisons within the auditable RAG pipeline.

WLCA knowledge base development

The Whole-Life Carbon Assessment Knowledge Base System (KBS) functions as the authoritative, version-controlled data backbone of the proposed multi-agent, retrieval-augmented framework. All user queries, agent reasoning, and dashboard outputs resolve directly to records stored in the KBS, which persists not only final whole-life carbon indicators but also the intermediate parameters, scenario states, and modeling artifacts required to reconstruct any reported value. This design ensures that WLCA results remain traceable, auditable, and reproducible across software environments and jurisdictions, while preserving numerical consistency and provenance throughout the assessment workflow.

At the data layer, the KBS is organized into normalized tables that store archetype descriptors in *ArchetypeData*, operational simulation outputs in *OperationalData*, embodied material inventories in *EmbodiedData*, and higher-level assessment parameters in the WLCA Parametric KBS, all indexed by explicit scenario and state identifiers. Baseline and post-retrofit conditions are encoded as distinct scenario states, enabling consistent before–after comparisons and trade-off analysis without duplicating archetype definitions. To support retrieval-augmented generation, the KBS exposes both structured access for deterministic constraint filtering and semantic representations for contextual retrieval, allowing the system to combine semantic relevance with strict numeric grounding and ensuring that all returned WLCA outputs are anchored in authoritative, auditable records.

LLM-enabled multi-agent RAG for LCA-centric retrieval and reasoning

The proposed system implements a domain-specific, multi-agent retrieval-augmented generation (RAG) layer that mediates between a stakeholder-facing dashboard and an underlying WLCA Knowledge Base System (KBS) (Table 1). The architecture is explicitly designed for research-grade life cycle assessment, where numerical correctness, scenario consistency, and full provenance traceability are mandatory. Rather than allowing a large language model to generate estimates directly, the system decomposes query handling, data retrieval, validation, domain reasoning, and output persistence into explicit agents and deterministic process steps.

To support transparency and interpretability, the operation of each architectural component is explained using the same representative retrofit query that is traced end-to-end in Table 1. The example considers a dwelling constructed in 1965, originally heated by a boiler, with the user seeking to upgrade to a heat pump, achieving a BER A2 rating while minimizing embodied carbon impacts.

User Query and Interface Agent The process begins when the user submits a natural-language query via the dashboard, such as: *"I have a house built in 1965 with a boiler and would like to upgrade to A2 rating with a heat pump and lower embodied carbon."* As shown in the first row of Table 1, this raw query is persisted verbatim in the *UserQuery* table, ensuring that the original user intent is preserved for audit and traceability.

The Interface Agent then performs query normalization. Informal expressions such as "house built in 1965," "boiler," and "A2 rating" are mapped to standardized domain concepts, including the 1950–1966 age band, a baseline boiler heating system, and a post-retrofit BER A2 performance target. The resulting normalized query string (Table 1, second row) removes linguistic variability while preserving semantic intent and provides a stable input for deterministic interpretation and retrieval.

Query Interpretation and Constraint Encoding. The normalized query is translated into a structured representation of intent and constraints. As illustrated in Table 1, this step extracts the retrofit action (boiler to heat pump), the target post-retrofit energy rating (BER A2 or better), the optimization objective (minimize embodied carbon), and the building context (constructed in 1965, corresponding to the 1950–1966 age band). These elements are encoded in a *RetrievalConstraints* record, which governs all subsequent retrieval and filtering operations and ensures that the user's intent is expressed in a fully specified and reproducible form.

Semantic Retrieval and Candidate Set Generation. Guided by the encoded constraints, the RAG layer initiates semantic retrieval over the WLCA KBS. This step identifies archetypes, retrofit scenarios, and related records that are contextually similar to the interpreted query. The results are stored as a ranked *CandidateSet* (Table 1), in which multiple mid-century dwelling archetypes with low-carbon retrofit characteristics are initially considered. This stage prioritizes recall and ensures that potentially relevant records are not prematurely excluded.

Structured Filtering and Compatibility Enforcement. The candidate records are then subjected to structured filtering using the explicitly encoded constraints. The candidates are retained only if they support a heat pump retrofit and achieve a post-retrofit BER A2 or better (Table 1). This step enforces methodological and scenario compatibility, eliminating records that are semantically similar but technically or contextually invalid for the query.

Join, Re-ranking, and Retrieved Record Bundling. Filtered records are joined across the *ArchetypeData*, *OperationalData*, *EmbodiedData*, and WLCA parametric tables to align operational simulations, material inventories, and assessment assumptions under a common archetype and scenario definition. The joined records are then re-ranked according to the optimization objective in this example, lowest embodied carbon per square metre, yielding

Table 1: End-to-end processing of a natural language retrofit query using the proposed retrieval-augmented, agent-based whole life carbon assessment framework.

Process Step	Input / Context	Example Output for the Query
1. User Query Capture	Raw user input	*"I have a house built in 1965 with a boiler and would like to upgrade to A2 rating with a heat pump and lower embodied carbon."*
2. Interface Agent (Normalization)	Raw query text	Normalized query string used for interpretation and retrieval: *"house built 1965 boiler upgrade heatpump a2 rating lower embodied carbon"*
3. Query Interpretation	Normalized query	Extracted intent and constraints: – Retrofit from boiler to heat pump – Target post-retrofit energy rating: BER A2 or better – Optimization objective: minimize embodied carbon – Building context: constructed in 1965 (Age Band D: 1950–1966)
4. Semantic Retrieval	Interpreted query and parametric knowledge base	Candidate archetypes ranked by semantic similarity to the query, e.g.: 1) 1950–1966 dwelling with heat pump retrofit 2) Post-war detached dwelling with insulation upgrade 3) Retrofit archetype with low-carbon heating
5. Structured Filtering	Candidate set and extracted constraints	Records filtered to retain only those that: – Achieve BER A2 or better after retrofit – Are compatible with heat pump systems
6. Join and Re-ranking	Filtered candidate set and optimization objective	Remaining records re-ranked by lowest embodied carbon per square metre, yielding a top-K shortlist
7. Grounding and Validation	Top-ranked records	Selection of the best-matching archetype with full provenance: – Document identifier – Dataset version – Fields used for energy and carbon calculations
8. Operational Carbon Agent	Grounded archetype	Before/after operational performance estimates, e.g.: – BER before retrofit: C2 – BER after retrofit: A2 – Primary energy intensity reduced from 210 to 75 kWh/m^2/yr
9. Embodied Carbon Agent	Grounded archetype	Embodied carbon breakdown by components (walls, windows, roof, floor, heat pump) and total embodied carbon per m^2
10. Whole Life Carbon Agent	Grounded archetype	Combined 60-year whole life carbon result including operational and embodied emissions, e.g.: – Operational: 540 kgCO_2e/m^2 – Embodied: 390 kgCO_2e/m^2 – Total WLCA: 930 kgCO_2e/m^2
11. Auditable Output Storage	Agent outputs and provenance	Persisted auditable record containing retrieval preview, operational results, embodied results, WLCA results, and dataset version

a bounded Top-K shortlist. This shortlist is materialized as a *RetrievedRecordBundle*, representing the best-supported interpretation of the user query.

Grounding and Validation. Before numerical reasoning occurs, the retrieved records undergo grounding and validation. This step verifies that the selected archetype is fully parameterized, that all required operational and embodied data are present, and that dataset versions and source documents are explicitly identified. Only records that pass these checks are exposed to downstream agents,

ensuring that all reported results are fully grounded and reproducible.

Operational Carbon Agent. The Operational Carbon Agent consumes the grounded records to report on pre- and post-operational performance. In the example, this includes changes in BER rating and primary energy intensity, with reductions from a pre-retrofit C2 level to a post-retrofit A2 level, and a corresponding reduction in energy use intensity. All values are linked to underlying simulation outputs stored in the *OperationalData* table.

Embodied Carbon Agent. In parallel, the Embodied Carbon Agent computes embodied emissions associated with the retrofit measures. For the running example, this includes a component-level breakdown of embodied carbon contributions from walls, windows, roof, floor, and the heat pump system itself, aggregated into a total embodied carbon intensity per square metre. Factor sources and dataset versions are explicitly documented.

Whole-Life Carbon Agent. The Whole-Life Carbon Agent integrates operational and embodied results over a 60-year assessment horizon to compute total whole-life carbon impacts. The combined operational and embodied emissions enable direct comparison between baseline and retrofit scenarios and the quantification of net carbon savings (Table 1).

Auditable Output Storage. Finally, all agent outputs and associated provenance metadata are consolidated into a persistent *AuditableOutputs* record. This record includes a retrieval preview, operational results, embodied results, whole-life carbon results, and dataset version identifiers. Persisting these outputs ensures that all reported values can be reproduced by re-running the same pipeline against the same WLCA KBS snapshot.

Stakeholder LLM dashboard and decision support

The stakeholder-facing LLM dashboard is the primary interface to the multi-agent RAG system and is tightly integrated with its architecture. It performs no independent computation, acting instead as a thin interaction layer that submits user queries, presents grounded outputs, and exposes audit information generated by the backend. Each interaction corresponds to a *UserQuery* record and follows the same retrieval, validation, and agent-reasoning workflow.

Using natural language, users can explore archetypes, compare retrofit strategies, and analyze embodied–operational trade-offs across assessment horizons. Queries are routed through the *Interface Agent*, which coordinates retrieval and agent execution and returns structured results with narrative explanations. Quantitative outputs are displayed as tables and charts, while an embedded audit view provides dataset versions, document identifiers, factor sources, and scenario states. When evidence is incomplete, uncertainty flags and provenance gaps are shown instead of unsupported values.

The dashboard supports exporting results in CSV or JSON formats and generating auto-formatted reports with embedded audit metadata. Because outputs are derived from persisted *AuditableOutputs* records, all exported artifacts are reproducible outside the system. Portability across regions and jurisdictions is achieved by substituting local archetypes, material inventories, and climate datasets within the KBS while preserving schemas, retrieval logic, and audit guarantees, allowing the dashboard and methodology to be reimplemented in new contexts without modifying the core pipeline.

Performance and validation metrics

The evaluation framework assesses retrieval quality, answer faithfulness, and numerical correctness within the LCA-centric multi-agent RAG system to ensure transparency and methodological compliance. Retrieval performance is measured using standard information-retrieval metrics, including Recall@k, Precision@k, Mean Reciprocal Rank (MRR), and normalized Discounted Cumulative Gain (nDCG), computed on a labeled query set linking stakeholder questions to authoritative KBS records. High Recall@k and nDCG indicate that relevant WLCA evidence is retrieved and correctly ranked, while MRR captures how quickly usable evidence appears, which is particularly important in multi-agent workflows. In addition, feasibility-oriented checks evaluate whether at least one constraint-satisfying retrofit option is retrieved within top-k results, reflecting the decision-search nature of stakeholder queries (Sharma, 2025).

Faithfulness and numerical correctness are enforced through strict grounding of all reported values to versioned KBS records, with citation coverage, provenance completeness, and contradiction checks used to audit responses. WLCA outputs are further validated against ISO 14040/14044 system boundaries and accounting conventions, with spot checks ensuring that reported whole-life totals are consistent with the sum of operational and embodied components within defined tolerances (ISO, 2006). All evaluations are tied to dataset version tags and audit metadata, enabling reproducibility across regions by substituting local archetypes, material inventories, and climate data while preserving retrieval logic and guardrails.

Case study and results

This case study demonstrates how a proposed whole-life carbon (WLC) methodology can be applied to the Irish residential sector using nationally available datasets, regulatory conventions, and realistic retrofit pathways. Focusing on semi-detached dwellings as a representative housing typology, it shows an end-to-end workflow from stock characterization and archetype definition to dynamic operational energy simulation and embodied-carbon accounting (Table1). The national Energy Performance Certificate (EPC) database underpins the analysis by providing standardized dwelling attributes, which are translated into physically interpretable archetypes aligned with Ireland's Dwelling Energy Assessment Procedure (DEAP). These

Table 2: Summary of dashboard results for retrofit query on Irish semi-detached dwellings (roof built 2000).

Attribute	Result
Query constraints	I have a house built in 1965 with a boiler and would like to upgrade to A2 rating with a heat pump and lower embodied carbon.
BER (Pre → Post)	C2 → A2
Primary EUI	177.1 → 49.7 (kWh/m^2/yr)
WLCA 60y total	1620.8 → 211.6 (kgCO$_2$e/m^2)
WLCA savings	1409.2 (kgCO$_2$e/m^2)
Operational carbon 60y	1554.2 → 108.1 (kgCO$_2$e/m^2)
Embodied carbon 60y	66.6 → 103.5 (kgCO$_2$e/m^2)
Post-retrofit carbon split	Operational 51% / Embodied 49%

archetypes capture age-dependent construction practices, baseline thermal performance, and window characteristics, forming a consistent basis for simulating pre- and post-retrofit operational energy demand under controlled climate and occupancy assumptions.

Operational and embodied carbon impacts are evaluated through a parametric simulation and whole-life carbon assessment pipeline that systematically explores retrofit scenarios, including envelope upgrades and low-carbon heating systems. Dynamic simulation outputs are combined with Irish and international embodied-carbon factors to account for initial construction, retrofit measures, component replacements, and end-of-life processes over the assessment horizon. Results are stored in a structured knowledge base that supports auditable, reproducible analysis and retrieval-augmented generation, enabling stakeholders to query and compare retrofit options by operational, embodied, or total life-cycle carbon. Together, the case study illustrates how an integrated, data-driven framework can support transparent whole-life carbon assessment and decision-making for residential decarbonization in the Irish context.

The framework enables stakeholders to enter constraint-based queries through the dashboard to analyse existing buildings and evaluate retrofit outcomes. The system identifies a transition from a current BER C2 dwelling to a post-retrofit A2 option with heatpump upgrade, reducing WLCA_{60y} from 1620.8 to 211.6 kgCO$_2$e/m^2, corresponding to a saving of 1409.2 kgCO$_2$e/m^2 (approximately 86%). In parallel, Primary EUI decreases from 177.1 to 49.7 kWh/m^2/yr (a reduction of 128 kWh/m^2/yr, or 72%), while the post-retrofit carbon contribution shifts to 51% operational and 49% embodied, demonstrating transparent quantification of operational–embodied trade-offs (Table 2).

The study is evaluated using the multi-axis protocol described in the methodology, combining retrieval effectiveness, decision relevance, numerical faithfulness, and LCA correctness. Assessment is conducted on a stratified 20% sample of the 1.6 million parametric configurations (approximately 320,000 cases) using 30 stakeholder-style queries representing typical retrofit constraints. Each query yields on average 118 feasible designs after structured filtering, confirming that the design space remains highly multi-modal even under BER and envelope requirements, while the ten lowest-WLCA options form the reference set for decision-oriented metrics.

Table 3: Summary of retrieval performance across feasibility, ranking quality, and decision relevance metrics.

Method	Metrics at $K = \{1,3,5,10\}$
Full multi-agent system	Hit: 0.78/0.82/0.85/0.88 nDCG: 0.81/0.76/0.74/0.71 Precision: 0.74/0.72/0.71/0.70 Recall: 0.70/0.71/0.70/0.69 MRR: 0.74/0.76/0.77/0.79
Hybrid candidates	Hit: 0.61/0.69/0.73/0.79 nDCG: 0.69/0.67/0.65/0.63 Precision: 0.58/0.56/0.55/0.53 Recall: 0.57/0.55/0.54/0.52 MRR: 0.58/0.61/0.63/0.66
Semantic-only retrieval	Hit: 0.21/0.33/0.38/0.46 nDCG: 0.41/0.38/0.35/0.32 Precision: 0.24/0.21/0.20/0.19 Recall: 0.23/0.20/0.19/0.18 MRR: 0.24/0.27/0.29/0.31
Random (20% pool)	Hit: 0.10/0.11/0.18/0.32 nDCG: 0.16/0.17/0.17/0.18 Precision: 0.11/0.11/0.10/0.13 Recall: 0.11/0.10/0.10/0.12 MRR: 0.11/0.12/0.14/0.17

Feasibility discovery shows that the multi-agent framework reliably identifies valid retrofit options under strict constraints, achieving HitFeasible@1 of 0.78 and HitFeasible@10 of 0.88, far exceeding semantic-only retrieval (0.46 at $K = 10$) and random sampling (0.32). Hybrid candidate filtering improves feasibility (0.79 at $K = 10$) but remains inferior to the full WLCA-aware reranking, underscoring the value of objective-driven ordering in sparse solution spaces (Table 3).

Decision relevance and ranking quality further demonstrate the system's effectiveness. The framework attains PrecisionTop10@10 and RecallTop10@10 of 0.70, with strong early-rank performance (Precision@1 and MRR@1 of 0.74), indicating that optimal low-carbon designs typically appear at the top of the list. Ranking alignment with ideal WLCA ordering is high, with nDCG@1 of 0.81 and nDCG@10 of 0.71, while hybrid and semantic-only methods show substantially weaker performance.

Faithfulness checks confirm that all reported results remain numerically grounded and auditable, with citation coverage of 0.98, a numeric mismatch rate of 0.01, and universal compliance with WLCA boundary arithmetic. Together, these results demonstrate that the framework can navigate large parametric retrofit spaces while preserving LCA rigor, transparency, and interactive usability, making it suitable for stakeholder-facing decision support and transferable to other national building stocks.

Conclusion and future work

This study presents a scalable framework for estimating whole-life carbon emissions across national residential building stocks by integrating parametric simulation, standardized WLCA, and a multi-agent retrieval layer. Heterogeneous inputs, including archetype definitions, operational simulation outputs, material inventories, and life-cycle calculations are consolidated within a structured knowledge base to support transparent, query-driven analysis. Application to the Irish residential sector demonstrates that the framework can explore more than one million retrofit configurations while preserving numeric grounding, data provenance, and interactive performance, with structured filtering and WLCA-aware re-ranking significantly improving the identification of feasible low-carbon options over semantic retrieval alone.

The framework delivers three key contributions: it unifies operational and embodied carbon assessment within a single parametric pipeline to enable consistent before and after retrofit evaluation; it provides versioned, auditable records that support reproducible national-scale WLCA without embedding assumptions in the language layer; and it enables natural-language access to complex WLCA evidence through multi-agent orchestration with enforced numeric guardrails. Future work will extend the approach to additional building typologies and integrate dynamic occupancy, monitored energy data, and spatial datasets to enhance realism and support neighbourhood- and stock-scale decarbonization planning.

Acknowledgments

This publication has emanated from research supported partly by a grant from the Sustainable Energy Authority of Ireland (SEAI) under the Research, Development and Demonstration Funding Programme Grant No. 22/RDD/822. The opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the SEAI. This research is also part of the IEA EBC - Annex 91 – Open BIM for Energy Efficient Buildings.

References

(ISO 14044:2006). Environmental management: Life cycle assessment framework requirements and guidelines.

Aghamolaei, R. et al. (2025). Integrated life cycle assessment of residential retrofit strategies: Balancing operational and embodied carbon, lessons from an irish housing case study. Sustainability, 17(18):8173. Accessed: 2026-01-08.

Ali, U., Shamsi, M. H., Hoare, C., Mangina, E., and O'Donnell, J. (2019). A data-driven approach for multi-scale building archetypes development. Energy and buildings, 202:109364.

Arbulu, M., Oregi, X., Etxepare, L., et al. (2024). Parametric simulation tool for the enviro-economic evaluation of energy renovation strategies in residential buildings with life cycle thinking: Pararenovate-lct. Energy and Buildings, 312:114182.

Chartered Institution of Building Services Engineers (CIBSE) (2021). Embodied carbon in building services: A calculation methodology (tm65). Accessed: 2025-06-26.

Circular Ecology (2025). Embodied carbon footprint database – the ice database. `https://circularecology.com/embodied-carbon-footprint-database.html`. Accessed: 2025-10-23.

Ecoinvent Association (2025). Ecoinvent database. `https://ecoinvent.org/database/`. Accessed: 2025-10-23.

Erdogan, E. (2019). Comparative studies of building performance simulation tools (energyplus, ies-ve, ida-ice).

European Commission (2025). Global warming potential of buildings. Accessed: 2026-01-08.

Hegarty, R. O. and Kinnane, O. (2023). A whole life carbon analysis of the irish residential sector - past, present and future. Energy and Climate Change, 4:100101.

Hurst, L. J. and O'Donovan, T. S. (2025). Defining new approaches for retrofit life cycle analysis to improve design outcomes. Energy and Buildings, page 116844.

INED (2024). 2024: the united nations publishes new world population projections. Accessed: 2026-01-08.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. In Advances in Neural Information Processing Systems (NeurIPS).

McGinley, O., Moran, P., and Goggins, J. (2022). An assessment of the key performance indicators (kpis) of energy efficient retrofits to existing residential buildings. Energies, 15(1):334.

Penaka, S. R., Feng, K., Olofsson, T., Rebbling, A., and Lu, W. (2024). Improved energy retrofit decision making through enhanced bottom-up building stock modelling. Energy and Buildings, 318:114492.

Sharma, C. (2025). Retrieval-augmented generation: A comprehensive survey of architectures, enhancements, and robustness frontiers. arXiv preprint arXiv:2506.00054.

Shen, P. et al. (2024). Archetype building energy modeling approaches and applications: A review. Renewable and Sustainable Energy Reviews, 199:114478.

Xie, Y., Jiang, B., Mallick, T., et al. (2025). Marsha: multi-agent rag system for hazard adaptation. npj Climate Action, 4(1).

Zhang, Y. and Korolija, I. (2010). Performing complex parametric simulations with jeplus. Shanghai Jiao Tong University.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

DO BUILDINGS REALLY NEED COMPREHENSIVE SENSOR NETWORKS? A GRAPH THEORY-BASED FRAMEWORK FOR PHYSICAL CONSISTENCY MODELING

Yangzhe Chen[1], Qinglu Zeng[1], Nan Hu [1], Thomas Ohlson Timoudas[2], and Qian Wang[1]
[1]Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Stockholm 10044, Sweden
[2]RISE Research Institutes of Sweden, Division Digital Systems, Computer Science, Isafjordsgatan 22, Kista 164 40, Sweden

Abstract

To enable building energy flexibility, sensor networks play a critical role in building energy management. However, comprehensive sensor network deployment is hindered by high Capital Expenditure (CAPEX) and retrofitting complexities. This study proposes a graph theory-based framework to address these limitations by exploiting the physical consistency inherent in building systems. By modeling the system topology using a bipartite graph, the framework identifies unknown variables given existing measurements via physical laws (e.g., thermodynamics). The proposed framework can also be used as virtual sensing for anomaly detection and sensor network design for both existing and new buildings.

Introduction

The global energy landscape is undergoing a shift, moving the system from centralized toward decentralized with high penetration of renewable energy sources. As a major participant in the energy system, the building sector is transitioning from a passive consumer role to an active prosumer, which provides energy flexibility via Demand-Side Management (DSM) (IEA, 2019; European Environment Agency, 2024). In this context, sensor networks play a central role in monitoring physical properties in the energy system and transmitting the real-time information to Building Management System (BMS) for further optimal control (Bae et al., 2021).

While the benefits of sensor networks are well-established, the practical deployment of comprehensive sensor networks faces significant barriers. Specifically, the high Capital Expenditure (CAPEX) associated with purchasing hardware and the complexity of retrofitting wiring infrastructure into the existing BMS hinder the practical deployment of sensors in buildings. These challenges occur in both design and renovation phases.

To address these limitations, there is a need to move beyond brute-force sensor deployment and instead exploit the potential of non-intrusive virtual sensing (Aguilar and R-Moreno, 2025). Although virtual sensing could be set up manually by experts, that is time-consuming, especially for identifying complex relationships between physical sensors and virtual ones. Building energy systems possess inherent physical consistency, for instance, the thermodynamic relationships between temperature, mass flow rate, and heat transfer. Thus, this indicates that not every variable requires a dedicated physical sensor. If the system topology is well understood, missing variables can be estimated through physical laws.

This study addresses the gaps by proposing a graph theory-based framework designed to represent the information flow in the building energy system by leveraging the physical consistency. By modeling the system topology as a bipartite graph, the correlation between each measurement point and physical laws is identified. This approach effectively creates a layer of virtual sensing, reducing the reliance on physical hardware without compromising the measurements required for DSM.

Previous studies

Sensor networks are necessary for building energy systems to collect data, detect faults and assist control (Koo et al., 2022). To ensure operational continuity, traditional approaches rely heavily on physical redundancy to enhance fault-tolerant ability. However, excessive physical redundancy often leads to high deployment costs and resource waste (Filios et al., 2024). Under the premise of meeting the system requirements, optimizing sensor configuration and virtual sensing are becoming the research focus. On the one hand, the number of sensors can be decreased by optimizing sensor configuration to reduce the deployment costs. For detecting and reducing faults, hardware redundancy is needed by using two same sensors in the same place. This strategy leads to a waste of sensors, so Panayiotis M. Papadopoulos et al. (Papadopoulos et al., 2022) designed a local fault diagnosis agent to estimate the state of Air Handling Units (AHUs) and compare the difference between the estimated value and the measured value. When the system is in the normal operation state, the difference will be constrained in an adaptive threshold which formed an Analytical Redundancy Relation (ARR). J. Aguilar et al. (Aguilar and R-Moreno, 2025) evaluated the sensor redundancy by structural observability analysis and optimized the sensor configuration by using Genetic Algorithm (GA). However, their focus is only on minimal sensor placement for fault detection. On the other hand, for the area lacking physical sensors because of high deployment costs or technical limits, virtual sensors can be constructed using physical-mathematical correlations or data-driven models (Wang and Li, 2022). Jiteng Li et al. (Li et al., 2024) established a framework of sensor off-

set definition by using redundant sensors, which can be a theoretical basis for virtual in-situ calibration method. Jabeom Koo et al. (Koo and Yoon, 2022) proposed an in-situ virtual sensing method that enables automatic virtualization and replacement when physical sensor bias exceeds approximately 1.0 °C. By integrating data-driven modeling with calibration techniques, the approach significantly enhances sensing reliability, reducing the initial average RMSE from 0.88 °C to 0.30 °C and eliminating 44.0% of bias errors and 32.9% of random errors.

Stylianos Karatzas et al. (Karatzas et al., 2024) used the LSTM algorithm to build a virtual sensing of framework for Indoor Air Quality (IAQ) index through data from two sensors for capturing pollutants. With this framework, this model can predict IAQ without the need of new sensors and its mean absolute percentage error is around 15%. Anupama Kowli et al. (Kowli et al., 2023) proposed a data-driven virtual sensing framework that utilizes linear regression to map the spatial distribution of indoor thermal variables, achieving estimation accuracies within 1 ℃ for temperature and 3% for relative humidity. By matching the performance of physical sensors like the BME280 over 95% of the time while minimizing hardware redundancy, this approach provides a cost-effective solution for high-fidelity indoor environmental monitoring. Yejin Hong et al. (Hong et al., 2021) proposed a physical-relation-driven assistance virtual sensor based on an autoencoder, which uses physical law and equipment performance curves to make up for the lack of key data. The accuracy for temperature, pressure, and control signals of the system under this strategy is all over 98%, effectively replacing the physical sensor.

While the previous studies have investigated the sensor configuration and virtual sensing techniques in different systems, they did not address the inherent physical consistency in the sensor networks and utilize it in a generalized approach. The detailed gaps are as follows:

- Sensor redundancy is well-established in control theory; it is traditionally categorized into hardware redundancy (multiple sensors) and analytical redundancy (model-based estimation). However, the potential of analytical redundancy for investigating the redundancy of sensor networks in complex building topologies remains underexplored.
- The majority of them approached the challenges by developing a complete high-fidelity model. This approach requires an extensive amount of data to train and calibrate the model, and it cannot be generalized.
- The inherent physical consistency has not been leveraged for sensor network related studies in the building energy system.

Contribution

Given the previous graph-based approaches (Papadopoulos et al. (2022); Aguilar and R-Moreno (2025)), this study proposes a generalized framework that focuses on the inherent physical consistency mapping in building energy systems. In contrast, the proposed framework is not oriented toward fault diagnosis or sensor placement optimization, but toward systematically mapping the physical consistency of an existing sensor network as a whole. By abstracting the building energy system as a bipartite graph, the governing physical laws are bridged with the sensor points. The proposed framework is then applied to an office building in Scandinavian climate.

Methodology

This section first introduces the problem formulation. In addition, the proposed framework is explained, and that includes physical system abstraction, graph construction, and observability propagation. The overall methodology is shown in Fig 1. The core of the proposed framework is an iterative propagation algorithm designed to map the information trajectory within the building energy system. Its primary function is to determine the maximal subset of system state variables that can be analytically deduced from a limited set of physical measurements. This process effectively converts the static topological model into a dynamic evaluation of system observability.

Problem formulation

In this study, the **physical consistency** of sensor networks in building energy systems is defined as the analytical interdependence of distributed physical measurements governed by fundamental physical laws. Specifically, the measurements are mathematically coupled at system components (e.g., Heat Exchanger (HEX), fans). For example, in a district heating substation, the thermal energy extracted from the primary side equals the energy delivered to the secondary side through the heat exchanger, with constant efficiency losses. This principle implies that a sensor network is not a set of independent measurements, but an interconnected system in which each measurement directly links with the related variables.

Likewise, **physical redundancy** is defined here as the ability to deduce a system variable (v_0) using a set of interdependent measurements ($\{v_1, v_2, ...v_n\}$) connected via a governing physical law (f):

$$v_0 = f^{-1}(v_1, v_2, ..., v_n) \tag{1}$$

where, v_i is calculated by inverting the governing physical law (f^{-1}). Accordingly, if the variables $\{v_1, ..., v_n\}$ are measured, v_j can potentially become a physical redundant sensor.

The **observability** is defined as the structural capacity to uniquely determine the value of a state variable given a specific network topology and measurement points. A variable is considered structurally observable if there exists physical consistency linking it to the set of known physical sensors and components. Therefore, the system's observability is binary for each sensor point: a variable is observable if it belongs to a closed set generated by

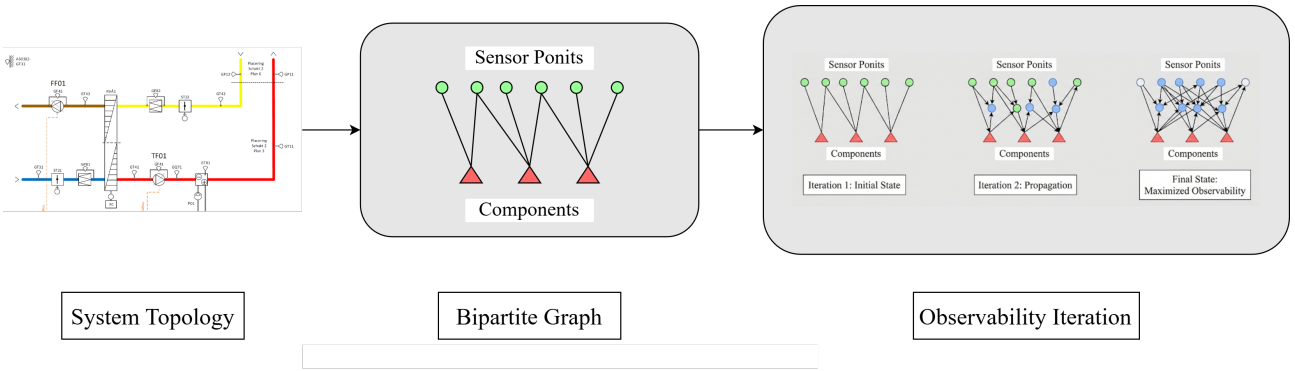


Figure 1: The overall methodology.

the propagation of known values through solvable physical laws.

Physical system abstraction

A building energy system can be decomposed into node sets of physical components (e.g., heat exchangers, pumps) and measurement variables (e.g., temperature, mass flow rate). The behavior of each component is governed by physical laws, such as mass and energy balance.

Let $\mathcal{V}$ be the set of measurement variables and C be the set of physical components. A variable $v_j \in \mathcal{V}$ represents a potentially measurable physical property, and $c_i \in C$ represents a component in the system, subject to the corresponding physical law involving a subset of variables. It is worth noting that the identification of both sets can be done by computer rather than manual work.

Maybe we should also highlight that this framework can be run by a computer to identify and set up potential virtual sensors?

Bipartite graph construction

The system can be modeled as a Bipartite Graph $G = (\mathcal{V}, C, \mathcal{E})$, where the two sets of vertices are the measurement variables $\mathcal{V}$ and the physical components C, respectively.

An undirected edge $e_{ij} \in \mathcal{E}$ exists between a variable v_j and a component c_i if and only if v_j is included as the subset variable in the physical law defined by c_i. Moreover, the $\mathcal{E}$ indicates the physical consistency of the target system.

The connectivity of the graph can also be represented algebraically by an incidence matrix $\mathrm{M} \in \{0,1\}^{|C|\times|\mathcal{V}|}$, defined as:

$$\mathrm{M}_{ij} = \begin{cases} 1, & \text{if } v_j \text{ is involved in } c_i \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where, $|\cdot|$ denotes the cardinality of the set. Additionally, another state vector S is defined, representing the observability status of each measurement variable:

$$\mathrm{S}_j = \begin{cases} 1 & \text{if variable } v_j \text{ is known (either measured} \\ & \text{or deduced)} \\ 0 & \text{if variable } v_j \text{ is unknown} \end{cases} \tag{3}$$

Observability propagation

Initialization. To initialize the problem, the global variable set $\mathcal{V}$ is split into two mutually exclusive subsets based on their initial observability status, as shown in Equation 4, and a variable may flip membership from unknown to measured through the iterative process below. The framework develops an iterative scheme to successively deduce unknown variables to increase the set of measured variables, and the details are explained as follows.

$$\mathcal{V} = \mathcal{V}_{\text{measured}} \cup \mathcal{V}_{\text{unknown}} \tag{4}$$

where:

- $\mathcal{V}_{\text{measured}}$ represents the subset of variables directly measured by sensors (e.g., legacy BMS sensors).
- $\mathcal{V}_{\text{unknown}}$ represents the subset of unknown variables. It can be unmeasured physical variables or the measurement to be diagnosed.

Iterative Propagation. The iteration logic is grounded in the solvability of the governing physical laws. At each iteration, the algorithm evaluates the unknown degree (d_i) for every physical component c_i at iteration k:

$$d_i^{(k)} = \sum_{j=1}^{|\mathcal{V}|} M_{ij} \cdot (1 - S_j^{(k)}) \tag{5}$$

where, M_{ij} denotes if the variable v_i is involved in the components c_j; $(1 - S_j^{(k)})$ evaluates to 1 if the variable v_i is unknown ($S_j = 0$) and 0 if it is known ($S_j = 1$). Consequently, d_i only sums the unmeasured variables connected to c_i. The propagation adheres to the following:

1. **Solvable condition** ($d_i = 1$): If a component c_j is connected to exactly one unknown variable v_{unknown} (i.e., all other involved variables are members of the measured set), the component becomes solvable. The variable v_{unknown} is then moved from $\mathcal{V}_{\text{unknown}}$ to the $\mathcal{V}_{\text{measured}}$.
2. **Redundant Condition** ($d_i = 0$): If all variables associated with the component c_i are known, the equation

ceases to provide new information and instead serves as a consistency check, providing analytical redundancy useful for fault detection.

3. **Unsolvable Condition** ($d_i > 1$): If a component contains multiple unknowns, it remains unsolvable at the current iteration k. However, it may become solvable in subsequent iterations as information propagates from neighboring components.

Convergence Criteria. The iterative process repeats, updating the state vector $\mathrm{S}^{(k)} \rightarrow \mathrm{S}^{(k+1)}$, until the system reaches the redundant condition ($d_i^k = 0$), and that is, $\mathcal{V} = \mathcal{V}_{\text{measured}}$.

Use case

In order to demonstrate the implementation of the proposed framework, it is applied to a Mechanical Ventilation with Heat Recovery (MVHR) in an office building in Scandinavian climate as a use case. The schematic of the system is shown in Figure 2. The exhaust air (yellow line) passes through the rotary HEX to recover waste heat and preheat the inlet air. Then the inlet is further heated/cooled at the heating battery to a specific set point temperature, supplied by district heating/cooling via Shaunt Group (SHG).

Table 1: The tag list of sensors in the target MVHR system.

Tag	Measurement
GT	Temperature sensor [°C]
GF	Mass flow sensor [l/s]
GP	Pressure sensor [Pa]
GQ	Smoke detector [ppm]
ST	Damper position [%]
SV	Valve position [%]

In order to show the whole workflow of the proposed framework, several assumptions are made:

- This study considers only temperature and mass flow rate sensors.
- The temperature and mass flow rate losses in the duct/pipes are not considered.
- The rotary HEX efficiency is assumed to be a constant.

Graph construction

This section explains the process of constructing the bipartite graph.

Measurement Variables. Seven temperature measurement variables are included in the variable set $\mathcal{V}$: GT42 (T_1), GT43 (T_2), GT31 (T_3), GT41 (T_4), GT82 (T_5), GT81 (T_6), and GT11 (T_7). Notably, GT81 measures both the supply and return water temperatures of the SHG. In addition, two mass flow rate sensors are considered: GF42 ($\dot{m}_1$) and GF41 ($\dot{m}_2$). All variables listed above belong to the subset $\mathcal{V}_{\text{measured}}$, as they are directly measured. In contrast, no sensor is available for the mass flow rate of the hydraulic system at the heating battery. Thus, this variable is denoted by $\dot{m}_3$ and classified as $\mathcal{V}_{\text{unknown}}$.

Physical Components. In this MVHR system, the two main physical components are the rotary HEX (c_1) and the heating battery (c_1). For the HEX, the governing physical law is energy balance between the exhaust and supply air:

$$\dot{m}_2 c_{p,a} \cdot (T_4 - T_3) = \eta_{\text{hex}} \cdot \dot{m}_1 c_{p,a} \cdot (T_1 - T_2) \quad (6)$$

where, $c_{p,a}$ represents the specific heat capacity of air; η_{hex} denotes the efficiency for HEX, and is assumed to be a constant and can be estimated from the historical data. As for the heating battery, the governing equation of that is the energy balance between air and water:

$$\dot{m}_2 c_{p,a} \cdot (T_7 - T_4) = \eta_{\text{hb}} \cdot \dot{m}_3 c_{p,w} \cdot (T_5 - T_6) \quad (7)$$

where, $c_{p,w}$ is the specific heat capacity of water; η_{hb} is the efficiency, and can also be calculated from historical data.

Bipartite Graph. Based on the defined variable set and physical component set, the system is mapped to a bipartite graph $G = (\mathcal{V}, \mathcal{C}, \mathcal{E})$. The variable node set $\mathcal{V}$ comprises the $n = 10$ state variables identified above (7 temperature nodes and 3 mass flow nodes). The component set contains $m = 2$ nodes, representing the specific energy balance equations for the rotary HEX (c_1) and the heating battery (c_2). The edge set $\mathcal{E}$ is determined by the incidence of variables in the governing equations. The graph is visualized in Figure 3.

- Component c_1 (Equation 6) is connected to the subset $\{v_{\dot{m}_1}, v_{\dot{m}_2}, v_{T_1}, v_{T_2}, v_{T_3}, v_{T_4}\}$.
- Component c_2 (Equation 7) is connected to the subset $\{v_{\dot{m}_2}, v_{\dot{m}_3}, v_{T_4}, v_{T_5}, v_{T_6}, v_{T_7}\}$.

Incidence Matrix. The connectivity of the bipartite graph is formalized by the incidence matrix $\mathrm{M} \in \{0,1\}^{2\times 10}$. The rows correspond to the component set C, and the columns correspond to the variable set $\mathcal{V}$. The matrix is defined as:

$$\mathrm{M} = \begin{array}{c} \\ c_1 \\ c_2 \end{array} \begin{array}{c} \begin{array}{cccccccccc} \dot{m}_1 & \dot{m}_2 & \dot{m}_3 & T_1 & T_2 & T_3 & T_4 & T_5 & T_6 & T_7 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \end{array} \quad (8)$$

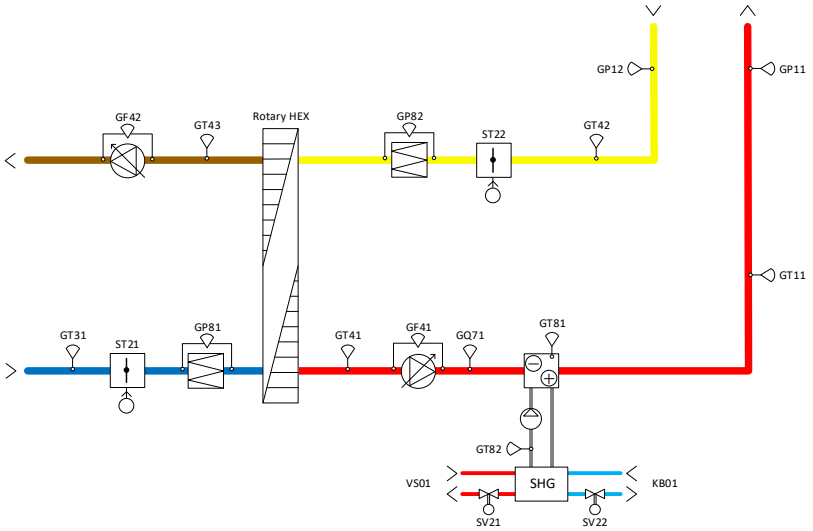


Figure 2: The system schematics of the Mechanical Ventilation with Heat Recovery system. The heat recovery from exhaust air happens at the rotary Heat Exchanger. The inlet air is further heated or cooled at the heating battery, supplied by district heating/cooling via the Shunt Group.

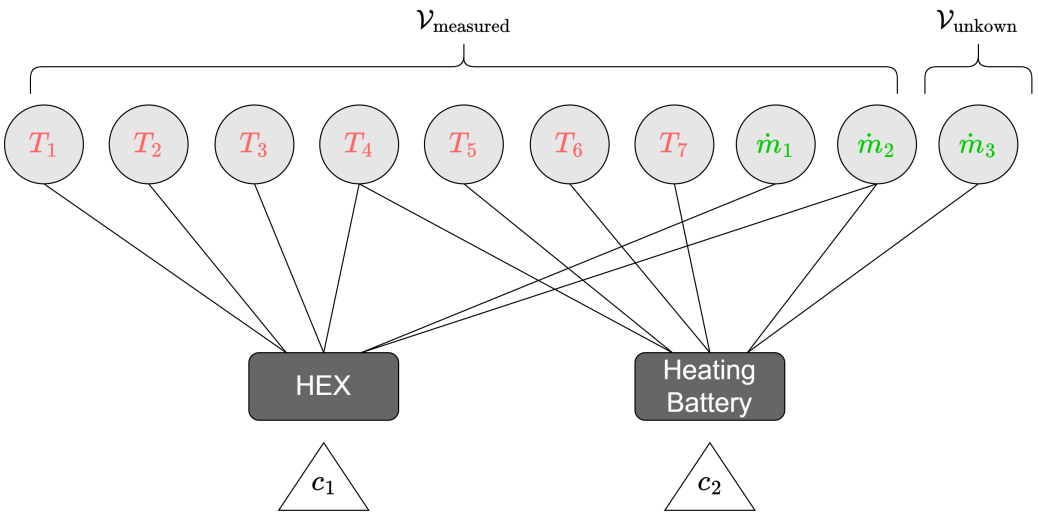


Figure 3: The constructed graph for the use case.

Observability propagation analysis

In this use case, the only unknown variable is the hydraulic mass flow rate, $\dot{m}_3$.

Step 1: Initialization ($k = 0$). The algorithm begins by populating the initial state vector $\mathrm{S}^{(0)}$ based on the sensor tag list (Table 1).

- The measured subset is initialized as $\mathcal{V}_{\text{measured}} = \{v_{\dot{m}_1}, v_{\dot{m}_2}, v_{T_1}, v_{T_2}, v_{T_3}, v_{T_4}, v_{T_5}, v_{T_6}, v_{T_7}\}$. Corresponding entries in $\mathrm{S}^{(0)}$ are set to 1.
- The unknown subset is initialized as $\mathcal{V}_{\text{unknown}} = \{v_{\dot{m}_3}\}$. The entry for $\dot{m}_3$ in $\mathrm{S}^{(0)}$ is set to 0.

Step 2: Iteration 1 ($k = 1$). The algorithm scans the constraint set $C = \{c_1, c_2\}$ to calculate the unknown degree $d_i^{(0)}$ for each component. For the rotary HEX (c_1), the incident variables are $\{v_{\dot{m}_1}, v_{\dot{m}_2}, v_{T_1}, v_{T_2}, v_{T_3}, v_{T_4}\}$. Since all these variables belong to $\mathcal{V}_{\text{measured}}$, the unknown degree is calculated as:

$$d_1^{(0)} = \sum_{c_1} (1 - 1) = 0 \tag{9}$$

According to the propagation logic, $d_1^{(0)} = 0$ satisfies the redundant condition. Consequently, c_1 does not infer new state variables.

For the heating battery (c_2), the incident variables are $\{\dot{m}_2, \dot{m}_3, T_4, T_5, T_6, T_7\}$. Among these, only $\dot{m}_3$ is in

$\mathcal{V}_{\text{unknown}}$. The unknown degree is:

$$d_2^{(0)} = (1-1)M_{c_2,v_{\dot{m}_2}} + (1-0)M_{c_2,v_{\dot{m}_3}} + \sum_{T \in c_2}(1-1) = 1 \tag{10}$$

Since $d_2^{(0)} = 1$, the solvable condition is met, c_2 is able to estimate the unknown variables.

Step 3: Resolution and Update. As triggered by the solvable condition, the algorithm algebraically inverts the governing equation of c_2 (Equation 7) to express the unknown variable $\dot{m}_3$ in terms of the measured set:

$$\dot{m}_3 = \frac{\dot{m}_2 c_{p,a}(T_7 - T_4)}{\eta_{\text{hb}} c_{p,w}(T_5 - T_6)} \tag{11}$$

In this context, $\dot{m}_3$ can be set up as a virtual sensor. The state vector is updated such that $\mathrm{S}_{\dot{m}_3}^{(1)} \rightarrow 1$, and the variable can be moved to $\mathcal{V}_{\text{measured}}$ for further propagation if needed.

Validation

In order to validate the approach, the temperature sensor GT41 (T_4) is assumed to be missing. It is used as a virtual sensor, and the prediction is compared with real measurement.

Figure 4 shows the validation results. The overall Root Mean Square Error (RMSE) of the validation results is 1.47 °C. It is worth noting that the RMSE during working hours (8:00 to 17:00) is only 0.23 °C. This can be explained by the different control strategies during working and non-working hours. During non-working hours, the building does not need ventilation, therefore, the rotary wheel is not actuated.

Discussion

The proposed graph-based framework offers a systematic approach to exploiting physical redundancy in building sensor networks. By visualizing the correlation between component topology and measurement variables, the methodology extends from complex system modeling towards modeling the inherent physical consistency in the system itself. This section discusses the potential application of the proposed methodology, along with the limitations and future works.

Potential applications

During the design phase, one potential application of the proposed framework is serving as a quantitative design tool to minimize CAPEX. Traditionally, sensor placement often follows rule-of-thumb and professional experience, which could lead to redundancy and unnecessary investments, especially in highly complex Heating, Ventilation, and Air Conditioning (HVAC) systems. By applying observability propagation during the design phase, the process of identifying a minimum set of sensors needed could be automated, thus reducing costs.

The framework could play a critical role in the retrofitting phase. As the building sector becomes proactive, a robust sensor network is required for future flexibility services. However, retrofitting robust sensor networks into old buildings is often hindered by the complex rewiring. As demonstrated in the case study, the deployment of virtual sensors, such as the inferred hydraulic mass flow rate, allows facility managers to upgrade legacy systems by using non-intrusive measurements, effectively bypassing the need for disruptive physical interventions.

Besides, the framework provides a robust foundation for fault detection. Instead of relying on static thresholds, the system can continuously cross-validate sensor readings by treating the target measurement as an unknown variable. A significant residual between the measured and virtual values serves as a high-confidence indicator of anomalies.

Limitations and future works

Although the proposed framework has shown potential strengths in different applications, there are still some assumptions that simplify the modeling. First, in the framework, the components are governed by theoretical physical laws; however, the uncertainties in the real practice are not considered. Moreover, some sensors are installed for safety reasons (e.g., sensors to prevent freezing), and this aspect needs to be taken into account when designing or calculating the redundancy of a sensor network. In addition, another limitation is that the energy loss is simplified to a parameter, which is not always constant in real practice. Therefore, stochastic modeling of the physical governing equations could be considered for future studies.

Future works could also focus on the coupling of semantic models (e.g., Knowledge Graphs and ontologies) with the framework for automating the development. One promising avenue for automating graph construction is the integration with established semantic building data models, such as Brick Schema. Brick Schema provides a standardized ontology for representing building energy systems, equipment, and sensor relationships using RDF-based knowledge graphs, where entities' interconnections are explicitly encoded as typed nodes and edges. This structured representation maps naturally onto the graph formulation proposed in this study: Brick's equipment classes (e.g., *brick:Heat_Exchanger*, *brick:Air_Handler*) correspond directly to the component set $\mathcal{C}$, while its point classes (e.g., *brick:Temperature_Sensor* and *brick:Flow_Sensor*) populate the variable set $\mathcal{V}$, and the *brick:feeds* or *brick:isPartOf* relationships define the edge set $\mathcal{E}$. Therefore, to ensure the scalability, the key focus for future studies is to identify entities' interconnections regardless of complexity of the system, from existing semantic models. Once the graph is constructed, the observability propagation algorithm could be integrated with BMS as a software toolbox.

Conclusion

This study proposed a graph theory-based framework to systematically exploit the inherent physical consistency in building energy systems, aiming to reduce reliance on

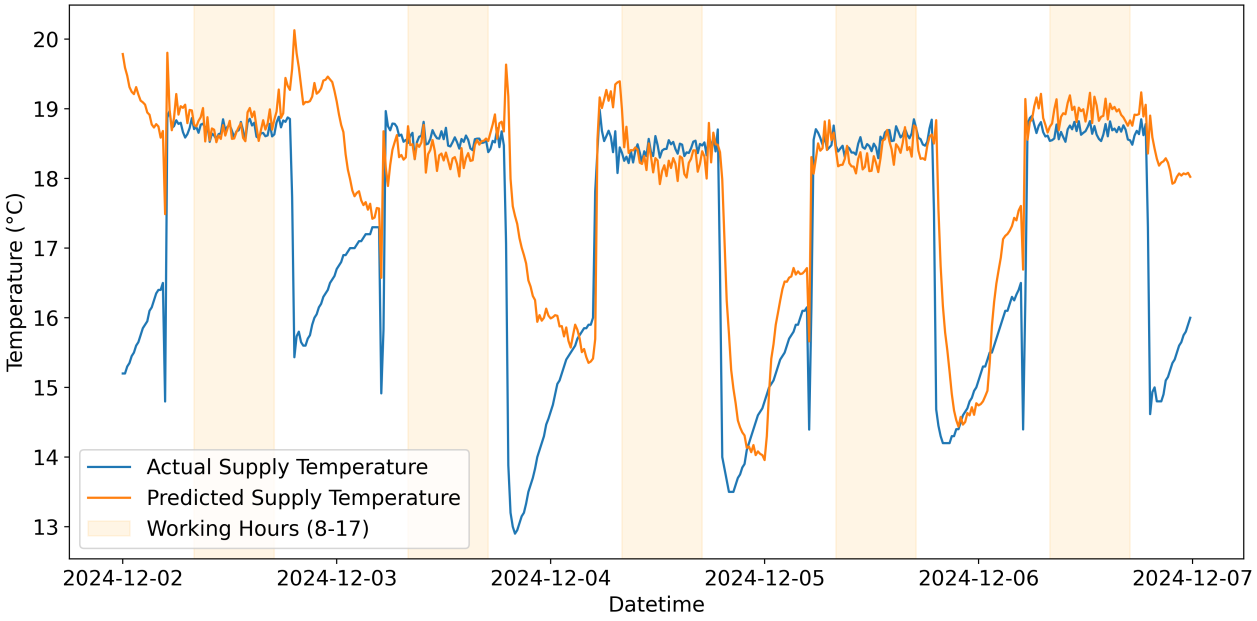


Figure 4: The validation results for 1 week.

comprehensive physical sensor networks. The main findings are as follows:

- By abstracting system topology as a bipartite graph, the framework bridges governing physical laws with measurement variables, transforming sensor networks from isolated measurement points into a physically consistent information system.
- The framework enables analytical identification of observable and redundant variables through an iterative observability propagation process. Compared to data-intensive or high-fidelity modeling approaches, the proposed framework is topology-driven, generalizable, and requires minimal calibration, making it suitable for scaling up to different Heating, Ventilation, and Air Conditioning (HVAC) systems.
- The framework provides a foundation for sensor network optimization and analytical redundancy throughout the building's lifecycle, including cost-effective sensor network design/retrofitting, and robust fault detection in building energy systems.

Acknowledgments

This work was supported by the Swedish Energy Agency under Grant number P2024-01185.

References

Aguilar, J. and R-Moreno, M. (2025). An optimal sensor configuration method using genetic algorithms for a fault-tolerant energy management for smart buildings. Computing, 107(9):188.

Bae, Y., Bhattacharya, S., Cui, B., Lee, S., Li, Y., Zhang, L., Im, P., Adetola, V., Vrabie, D., Leach, M., and Kuruganti, T. (2021). Sensor impacts on building and hvac controls: A critical review for building energy performance. Advances in Applied Energy, 4:100068.

European Environment Agency (2024). European climate risk assessment. Technical Report EEA Report 01/2024, European Environment Agency, Copenhagen, Denmark. Accessed: 2025-07-04.

Filios, G., Nikoletseas, S., and Stivaros, I. (2024). IAQ monitoring system optimizing data-driven sensor placement. In 2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), pages 408–415. IEEE.

Hong, Y., Yoon, S., Kim, Y.-S., and Jang, H. (2021). System-level virtual sensing method in building energy systems using autoencoder: Under the limited sensors and operational datasets. Applied Energy, 301:117458.

IEA (2019). Energy in buildings and communities programme Annex 67 energy flexible buildings summary report. Technical report, International Energy Agency.

Karatzas, S., Merino, J., Puchkova, A., Mountzouris, C., Protopsaltis, G., Gialelis, J., and Parlikad, A. K. (2024). A virtual sensing approach to enhancing personalized strategies for indoor environmental quality and residential energy management. Building and Environment, 261:111684.

Koo, J. and Yoon, S. (2022). In-situ sensor virtualization and calibration in building systems. Applied Energy, 325:119864.

Koo, J., Yoon, S., and Kim, J. (2022). Virtual in situ cali-

bration for operational backup virtual sensors in building energy systems. Energies, 15(4):1394.

Kowli, A., Rani, V., and Sanap, M. (2023). Data-driven virtual sensing for spatial distribution of temperature and humidity. Journal of Building Engineering, 67:105726.

Li, J., Wang, P., Li, Y., Rezgui, Y., Yoon, S., and Zhao, T. (2024). Analysis of sensor offset characteristics in building energy systems based on redundant sensors: A case study on variable air volume system. Energy and Buildings, 306:113957.

Papadopoulos, P. M., Lymperopoulos, G., Polycarpou, M. M., and Ioannou, P. (2022). Distributed diagnosis of sensor and actuator faults in air handling units in multi-zone buildings: a model-based approach. Energy and Buildings, 256:111709.

Wang, D. and Li, X. (2022). A novel virtual sensor modeling method based on deep learning and its application in heating, ventilation, and air-conditioning system. Energies, 15(15):5743.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

SEMANTIC MAPPING OF REQUEST FOR INFORMATION (RFI) TO BIM ELEMENTS WITH LARGE LANGUAGE MODELS

Limon Paul Joy, Farzaneh Zarei, and Mazdak Nik-Bakht
Department of Building, Civil & Environmental Engineering, Concordia University, Montréal, Québec, Canada

Abstract

Requests for Information (RFIs) remain a critical yet poorly integrated component of construction workflows, typically managed as unstructured text disconnected from Building Information Modeling (BIM) data. Existing studies primarily focus on reducing RFI frequency or response time, with limited focus on linking RFI content to specific BIM elements. This study presents a data-driven framework that combines large language model (LLM)-based intent extraction with BIM metadata reasoning through a staged matching pipeline. By separating semantic cue extraction from rule-based element selection, the framework enhances interpretability and robustness. Validation on experimental and real-world datasets demonstrates reliable element and family-level retrieval.

Introduction

The Request for Information (RFI) is a formal communication procedure to seek information, clarification, or resolution of uncertainties arising in construction. As an essential communication tool, RFI addresses discrepancies, prevents misunderstandings and reduces the risk of delays, claims and cost overruns. Typically, RFIs are initiated by contractors or subcontractors and directed towards designers or consultants to clarify ambiguities in designs, specifications, or constructability issues. The most common reasons for issuing an RFI are unclear specifications, design element contradictions and unforeseen field conditions (Hanna *et al.,* 2012). Despite this significance, RFIs are widely recognized as an inefficient process in project information flow and design coordination. Mohamed *et al.,* (1999) estimated a mean time of 17-person hours to process a single RFI. The cumulative impact of RFI related delays can be substantial at the project level. Total RFI response delays in a project can account for up to 10% of the overall project duration (Papajohn *et al.*, 2018). One of the primary reasons for this inefficiency is the RFI process being largely manual and document centric. Designers are often required to search through multiple sets of drawings, specifications, and revisions to identify the issue raised in an RFI, assess its implications, and issue a response. This fragmented and manual workflow contributes to elongated response times and a lack of traceability between RFIs and the affected building elements.

On the other hand, building information modeling (BIM) has reshaped the construction industry by introducing a digital framework for integrated project management. BIM is a digital representation of the physical and functional characteristics of a facility, which acts as a shared knowledge base for forming reliable decisions throughout its lifecycle, starting from the conceptual stage to its demolition (*NBIMS-USTM V3 – National Institute of Building Sciences*, 2026). Its significance in the construction industry lies within its ability to integrate data across disciplines within a collaborative environment. Furthermore, it provides a visual and spatial context that enhances the understanding of complex design information, enabling stakeholders to interpret the design more effectively than through documentation alone.

Despite their importance in construction practice, the interaction between RFIs and BIM remains limited. While BIM provides structured and spatially rich project information, RFIs are largely text-based and context-poor. Existing research has focused on improving RFIs within BIM-enabled environments, with limited attention to their systematic integration. This creates a noticeable gap in methods for semantically linking RFIs to BIM elements to support faster interpretation, more accurate responses, and improved traceability. Addressing this gap is increasingly relevant with the growth of digital project data and advances in information extraction and semantic analysis.

This research proposes an approach to link RFIs with BIM models to enhance the efficiency and effectiveness of the RFI process. It introduces a data-driven digital framework that utilizes large language models (LLMs) to semantically interpret unstructured RFI text and link it to structured BIM elements, enabling more efficient and context-aware RFI handling. The role of LLM is limited to extracting structural cues for keyword extension to guide the matching, while the core matching is driven by rule-based BIM reasoning. By establishing a direct connection between textual RFI content and BIM-based building elements, the proposed approach seeks to reduce manual effort, shorten response times, and improve information consistency across project stakeholders.

Background review

BIM use cases in RFI

Building Information Modeling (BIM) is one of the most widely adopted digitalization tools for enhancing collaboration in the Request for Information (RFI) process, as well as for identifying design and coordination issues that typically generate a high volume of RFIs. Previous studies have demonstrated that improved coordination within BIM models can significantly reduce the number of RFIs by minimizing design ambiguities and information gaps (Bhat *et al.,* 2017). Similar results were also found by Dan *et al.,* 2023,where the usage of BIM introduced better interdisciplinary design coordination, effective clash detection, and enhanced communication among project stakeholders, leading to reduced RFI frequency. Furthermore, studies investigating the combined implementation of BIM and lean construction principles reported reductions in RFI numbers of up to 70%, highlighting the potential of integrating digital modeling with process optimization strategies (Bhat et al., 2018). Beyond reducing RFI volume, BIM supported workflows also contribute to faster response times, improved design decisions, and more proactive issue resolution, thereby improving overall project efficiency. Morales *et al.,* (2022) identified reactive and proactive uses of BIM application to mitigate common issues raised in RFIs through 3D coordination, specialty design, design review and as-built modelling. BIM uses have also proved to have better financial management regarding RFI related costs. Giel *et al.,* (2013) quantified the return on investment of BIM by comparing BIM-based and non-BIM projects of similar scale, illustrating that BIM enabled projects resulted in reduced RFIs, improved coordination, fewer change orders, and faster response times, leading to significant cost improvement.

However, although substantial research has been conducted on BIM enabled environments and their impacts on the RFI process, the majority of existing studies have primarily focused on three key performance aspects of RFIs: cost, frequency, and response time. These studies have largely evaluated BIM as a coordination and visualization tool to reduce the overall burden of RFIs, rather than examining how RFI itself can be integrated within digital project environments. One of the few studies which focused on the linkage was conducted by Mao *et al.,* 2007 for generating structured RFIs by linking them to structured models such as IFC, enabling easier navigation and retrieval of relevant information for construction professionals. However, despite these early contributions, existing approaches remain largely conceptual and limited in scope, with minimal emphasis on automation. They do not adequately address automated interpretation of unstructured RFI text or its systematic linkage to BIM elements, leaving the process reliant on manual effort. The absence of explicit RFI–BIM connections limits traceability and hinders automated impact assessment.

Integration of LLM within BIM design environment

Recent advances in large language models (LLMs) have enabled new forms of interaction with BIM by enabling semantic interpretation of natural language and facilitating more intuitive interaction with complex model data. This capability has motivated growing research interest in integrating LLMs with BIM to improve accessibility, automation, and decision support. Existing research has predominantly applied LLMs to natural-language BIM querying, conversational model interaction, and parametric design generation ((Zheng *et al.,* 2023; Fernandes *et al.,* 2024; Jang *et al.,* 2024; Ko *et al.,* 2025). While these approaches demonstrate clear benefits in terms of usability, they remain largely confined to design centric use cases and do not address construction communication connections. More recent work has extended LLM applications toward regulatory and compliance-related BIM workflows. LLM-based frameworks have been proposed for automated interpretation of building codes (Yang *et al.,* 2024; Madireddy *et al.,* 2025) and autonomous compliance checking, demonstrating the potential of LLM agents to retrieve BIM data, execute rule checks, and support model revisions through natural-language interaction. Parallel to these developments, prior research has explored the application of natural language processing techniques to analyze RFI documents, focusing on extracting themes, classifying issues, and identifying key entities from unstructured text (Afzal *et al.,* 2023; Afzal *et al.,* 2024). These studies demonstrate the potential of data-driven approaches to analyze RFI datasets through automated topic discovery, issue classification, and entity extraction. In particular, unsupervised topic modelling identified recurring construction issues, while deep learning-based pipelines have been used to classify RFIs by project phase and issue type and extract key entities out of them.

Despite these advances, existing research primarily focuses on design generation, model querying, visualization, and compliance checking. Limited attention has been given to construction communication processes, and in particular, automated linkage of unstructured construction documents and texts to specific BIM elements. This gap limits traceability and automation in construction communication workflows and demonstrates the need for model centric approaches to explicitly connect construction documentation, such as RFIs, with BIM elements.

Methodology

This study proposes a data-driven framework linking unstructured RFIs to corresponding BIM elements, implemented through a Revit-based prototype. The framework integrates three components: (i) extraction of BIM element metadata, (ii) RFI preparation and intent extraction using an LLM, and (iii) a multi-stage matching and scoring pipeline combining deterministic rules with embedding-based semantic similarity to identify the best-matched elements. The LLM is used solely for intent extraction; while matching and final selection are governed by explicit rules and BIM metadata. The staged

pipeline was created based on an initial review of the RFI dataset, which indicated its unstructured and variable nature, requiring a stepwise refinement approach. This staged method evaluates high confidence cues first and progressively narrows the candidate through contextual and semantic filtering, with integrated shortlisting to maintain computational efficiency for large BIM models. *Figure 1* depicts the overall architecture of the proposed prototype. Overall, the prototype operates across three environments. BIM element extraction, visualization, and selection are performed within Revit. Structured cues extraction from RFI text is handled by a GPT-based language model. The multistage matching, scoring, and data management are executed in an external Python-based environment that coordinates interactions between the BIM model and the language model. Finally, a human-in-the-loop interface enables users to review and validate the final element selection.

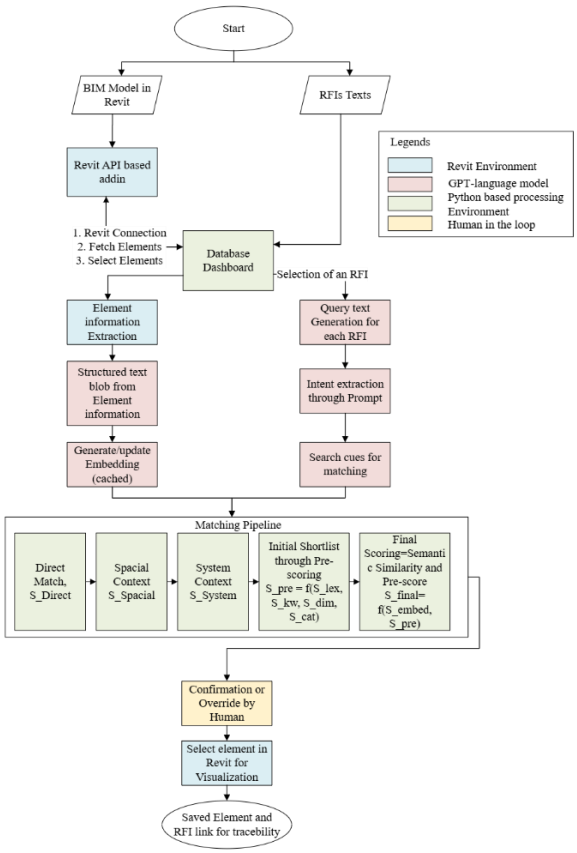


Figure 1: The overall architecture of the proposed prototype

Extraction of BIM element data from Revit model

Autodesk Revit was used as the BIM environment due to its widespread industry adoption. BIM element extraction is a critical step, enabling the transfer of structured geometric and semantic information to external components for semantic analysis and element matching.

BIM model data retrieval

To integrate the BIM model with the LLM-based matching framework, a custom C# Revit add-in was developed using the Revit API. The add-in acts as an HTTP bridge, exposing model data through endpoints for external retrieval, selection, and visualization of elements. This approach was adapted due to Revit API limitations, as API calls must run on the main thread, where computationally intensive tasks such as language processing and embedding generation would otherwise block the interface and degrade performance (*Help | Deployment Options | Autodesk*, 2026; *Help | Revit API Developers Guide | Autodesk*, 2026). The bridge supports three primary operations: Revit connection verifier, element data extraction and selection and highlighting elements.

Element metadata extraction

Using the extraction endpoint, BIM elements are retrieved in JSON format. For each element, a comprehensive set of attributes is collected to reflect how elements are typically referenced in RFIs and coordination documents. *Table 1* includes the parameters extracted from the element metadata.

Table 1: Extracted BIM element metadata parameters.

Category	Extracted attributes
Identifiers	UniqueId, Element Id
Semantic attributes	Category, Family, Instance
Contextual attributes	Level, Room, Space
Tags and Labels	Mark, Type Mark
System attributes	System of MEP elements
Spatial proxies	Nearest grid/axis reference

Structured textual representation and embedding

For text-based comparison, each BIM element's JSON data is converted into a structured textual representation (text blob), organizing key attributes into a normalized format. These text blobs support both (i) lexical/keyword-based matching and (ii) embedding-based similarity computation. Text embedding refers to the transformation of textual data into numerical vector representations that capture semantic meaning. BIM element embeddings were generated using *text-embedding-3-small* and *text-embedding-3-large* for performance comparison and are cached in the database. The embeddings are only recomputed when BIM metadata changes.

RFI data preparation and intent extraction

For semantic mapping of RFI documentation with structured BIM, RFI data is standardized and preprocessed into a consistent textual representation, making it suitable for analysis. Subsequently, structured intent cues are extracted from the normalized RFI text using LLM models.

RFI data preparation

RFIs are imported to the dashboard in a tabular format containing RFI identifier, subject/discipline, question, answer and status. For each RFI, these fields are combined into a single query text to ensure that the contextual information recorded throughout the RFI is incorporated into the matching process. Text normalization is applied to remove accents, standardize

casing, and normalize whitespace. This preprocessing of the text mitigates inconsistencies in style, abbreviations and formatting found in real-life practice.

Extract intent cues from RFI through LLM

For extracting relevant semantic cues for BIM element matching from RFI texts, two different large language models (GPT 4.1 and GPT-4.1-mini) were used to extract structural intent cues. Structured intent cues are predefined, machine-readable attributes (e.g., primary object, keywords, location) extracted from RFI text to guide downstream matching and reasoning. The normalized RFI text is fed into the model using a predefined prompt that specifies the exact structure and content of the expected output. The prompt instructs the model to return a strictly formatted JSON object containing only intent-related cues commonly expressed in RFIs, including discipline, primary object description, keywords and synonyms, spatial context (e.g., levels, grid references), and any explicitly mentioned dimensions or quantities. The model is explicitly restricted from inferring BIM element identifiers, Revit categories, or specific element instances. *Figure 2* illustrates the prompt used for intent extraction.

```
You are a construction manager.
Extract structured intent from the RFI text below.
Return ONLY valid JSON with these fields:
- action: one of [move, install, modify, remove, verify, document, clarify, other]
- primary_object: a short canonical English noun phrase (e.g., "pendant light", "skylight", "door", "circuit breaker")
- keywords: 5-12 keywords/phrases relevant to matching the correct element (include synonyms; include abbreviations)
- location: location info if mentioned (e.g., level, zone, direction, grid/axis), or empty string
- dimensions: dimensions or quantities if mentioned (e.g., 80 mm), or empty string
- discipline: one of [architectural, structural, mechanical, electrical, plumbing, general, unknown]
- confidence: number between 0 and 1
Rules:
- Do NOT guess BIM element IDs or Revit categories.
- Do NOT suggest specific element instances.
- Return JSON only. No extra text.
RFI TEXT: {RFI_TEXT}
```

Figure 2: Prompt used for intent extraction from RFI text.

In this process, the LLM does not participate in element selection; instead, extracted intent cues guide matching by expanding keywords, activating discipline-specific filters, and identifying spatial or dimensional constraints. Intent influences scoring only when supported by corresponding BIM metadata, and since the LLM cannot introduce new elements or identifiers, hallucinations have no direct impact. This design preserves transparency and limits hallucination effects through rule-based decision-making.

Multistage matching and scoring pipeline

The proposed framework employs a five-stage scoring and matching pipeline to select the best matching candidate. Early stages of this pipeline rely on explicit, high confidence cues while later stages incorporate semantic scoring. This staged design improves interpretability, computational efficiency, and robustness.

Direct identifier matching

The first stage evaluates if the RFI text explicitly mentions BIM identifiers such as ElementId, UniqueId, Mark or Type Mark. These identifiers are matched exactly against the cached BIM metadata. When the match is detected, a direct matching score S_{direct} is assigned.

$$S_{direct} = \begin{cases} 1.00 \; if \; ElementID \; match \\ 0.80 \; if \; UniqueId \; match \\ 0.60 \; if \; Mark \; match \\ 0.55 \; if \; Type \; Mark \; match \\ 0 \; if \; no \; match \end{cases} \quad (1)$$

If multiple identifiers are matched for the same element, their scores are accumulated. Because explicit identifiers provide the highest confidence linkage, if one or more candidates are identified at this stage, the pipeline terminates early and downstream stages are skipped.

Spatial context match

If the RFI contains explicit spatial references such as level, room or grid/axis identifiers, the spatial context matching is used. These references are extracted from the RFI and compared against corresponding BIM element attributes, including levels, associated room or space, and nearest axis/grid references derived from the model. Elements inconsistent with the referenced spatial context are penalized or filtered.

To compute axis/grid distance, each element represented by a point, p (location or centroid) and its distance to axis/grid, g is computed in XY plane as the minimum distance to its segment,

$$d(p, g) = \; min_{s \in g} \, dist_{XY} \, (p, s) \quad (2)$$

Here, s represents a segment of the axis/grid curve. Based on this, the nearest numeric and alphabetic axis/grid are identified and used for the matching.

Finally, A spatial context matching score, $S_{spatial}$ is computed as,

$$S_{spatial} = \begin{cases} +0.25 \; if \; match \\ -0.25 \; if \;\; conflicts \\ 0 \; if \; no \; match \end{cases} \quad (3)$$

To preserve recall, if the filter eliminates all the candidates, the unfiltered candidate set is retained.

System context match

For mechanical, electrical, and plumbing (MEP) related RFIs, system identifier provides an important functional constraint. When system related keywords are present in the RFI, alignment between the element system name and the extracted intent is evaluated. A system context matching score, S_{system} is computed as,

$$S_{system} = \begin{cases} +0.15 \; if \; match \\ -0.10 \; if \;\; system \; is \; missing \\ 0 \; if \; otherwise \end{cases} \quad (4)$$

This stage reduces cross-disciplinary mismatches that may arise from semantic similarity alone.

Prescoring for shortlist

Remaining candidates from the past three stages are ranked based on a pre-scoring process which combines multiple scoring.

First, lexical similarity within the RFI query text, Q and element text blob, E is computed as,

$$S_{lex} = \frac{|Token \; (Q) \cap Token \; (E)|}{|Token \; (Q)|} \quad (5)$$

On the contrary, intent cues derived from RFI texts are grounded against element metadata. The keyword boost score, S_{Kw} is computed as,

$$S_{kw} = \min(0.35, 0.05 x Nos\ of\ matched\ keywords \quad (6)$$

Furthermore, a dimensional match is computed if the numeric values mentioned in the RFI is a match the element metadata. The score,

$$S_{dim} = \begin{cases} +0.10\ if\ dimention\ match \\ 0\ if\ no\ match \end{cases} \quad (7)$$

Additionally, sometimes, material names exhibit strong lexical and semantic similarity with RFI text. This induces false positives giving materials as the best match. To avoid this issue and as the scope was to connect physical BIM design elements, material-level entities are explicitly penalized. The score, S_{cat},

$$S_{cat} = \begin{cases} -0.35\ if\ category\ \{\text{Materials, Analytical Surfaces}\} \\ -0.20\ if\ category = Mass \\ 0\ if\ otherwise \end{cases} \quad (8)$$

Finally, an overall prescore is calculated as a weighted sum of all the scores, as S_{pre},

$$S_{pre} = w_1 \times S_{lex} + S_{spacial} + S_{system} + S_{kw} + S_{dim} + S_{cat} \quad (9)$$

Here, the lexical similarity score was assigned a weight $w_1 = 0.35$ to reflect its role as a supporting signal rather than a dominant factor within the scoring pipeline. Lexical overlap is effective when RFIs explicitly mention element names, families, or parameters, but is sensitive to variations in terminology and phrasing. A moderate weight supports early ranking without overshadowing contextual, spatial, and semantic cues, while only top-ranked candidates are passed to the final stage, improving efficiency and accuracy.

Embedding based semantic similarity

After initial shortlisting through the prescoring, semantic similarity is computed using embedding vectors generated by text-embedding-3-small model. Both the RFI query text, Q and the element text blob, E are embedded into a shared vector space. Semantic similarity, S_{embed}, is computed using cosine similarity,

$$S_{embed} = \frac{V_Q . V_E}{|V_Q| . |V_E|} \quad (10)$$

Final ranking and user validation

The final matching score S_{final} is computed by incorporating embedding similarity with lexical similarity as well as remaining evidence-based scorings, while avoiding double counting of the lexical score. Here, a weight is used to prioritize the semantic score instead of the lexical score, as in some instances, the lexical score can overpower due to keyword matches in the element data. In this study, the weight w_2, was counted as 0.75, which can be increased or decreased depending on the user's requirement in the prototype dashboard to put more importance on either the semantic score or the lexical score. However, the weights and scorings were empirically chosen and no sensitivity analysis was conducted.

$$S_{final} = w_2 \times S_{embed} + (1 - w_2) \times S_{lex} + (S_{pre} - (1 - w_2) x\ S_{lex}) \quad (11)$$

Through the candidate final scoring, candidate elements are listed accordingly in an interactive interface. The next step includes user validation, where the user may accept the top-ranked element or select the appropriate one from the candidate list as the best. This step includes a human-in-the-loop validation mechanism that allows the user to review and confirm the prototype's suggestion. Upon confirmation, the selected element is selected in Revit and the confirmed RFI element linkage, along with all score components, is stored to support evaluation and traceability.

Evaluation of RFI-to-BIM Mapping

To determine the feasibility of this prototype, two BIM environments were considered: (i) an experimental Revit architectural Model with manually created RFIs based on the design, (ii) A real project already being completed with a set of real project RFIs.

Experimental BIM model with generated RFIs

For the first environment, an experimental architectural model was used to assess the baseline performance. Ten RFIs were manually created based on the model content to represent typical RFIs. These RFIs intentionally exclude explicit element identifiers (e.g., ElementId or UniqueId) and rely on descriptive cues commonly found in practice, such as element type, approximate location *(e.g.,* west-side skylight*),* or dimensional information *(e.g.,* timber column 228x502 mm*).* This setup ensured matching based on semantic and contextual interpretation rather than direct identifier matching. As the RFIs were created directly from the known model content, a ground-truth correspondence between RFIs and BIM elements could be explicitly defined. Figure 3 illustrates the experimental model, prototype dashboard with intents extracted from the RFI, best results found through the matching pipeline and the selection in Revit through the prototype environment. For quantitative evaluation of the performance, successful linkage was defined using two criteria: 1. Exact element match, where the top-ranked candidate corresponds to the intended BIM element, and 2. Partial match where the correct family or category is identified despite incorrect instance level selection.

The results demonstrate that the prototype successfully identified and ranked relevant BIM elements for the majority of RFIs. Out of 10 RFIs, 8 were successfully linked to the correct BIM elements, while 2 were unsuccessful. In one failed case, the correct family was identified but not the exact instance, and in the other, no element was matched. Treating exact matches as true positives yields a Top-1 accuracy of 80%, increasing to 90% when partial family or category matches are included, indicating effective narrowing of the candidate search space indicating effective narrowing of the candidate search space. From a retrieval perspective, each RFI was treated as a single query, yielding a precision and recall of 0.80. However, due to the limited sample size these metrics demonstrate feasibility rather than

generalizable performance. Analysis of the two unsuccessful cases shows reliance on room-level spatial references, where the absence of instance-level room associations prevented spatial grounding. This highlights the dependency on BIM metadata quality rather than limitations in semantic matching. Overall, the staged pipeline effectively prioritizes lexical and contextual cues, enabling accurate ranking when sufficient metadata is available.

Real project model with issued RFIs

The second environment involved a real construction project with an existing BIM model and a set of real RFIs, issued during the project lifecycle. The project had a total of 524 RFIs in disciplines from architectural, structural, building envelope, electrical, plumbing, mechanical and civil works. Unlike the experimental case, these RFIs exhibited greater linguistic variability, incomplete references, informal phrasing, and cross-disciplinary content, reflecting the complexity of real-world construction communication.

As the disciplines were modeled in separate Revit files, three were selected for implementation: architectural, structural, and building envelope, generating a total of 316 RFIs. In practice, most RFIs were communicated via emails with limited textual detail, with key context provided in separate marked drawing attachments. As this study considers only RFI text, a limited subset contained sufficient semantic information for meaningful linkage to BIM elements. After carefully screening all available RFIs, a subset of 74 RFIs was obtained where the affected design elements could be reasonably inferred. These RFIs were selected based on sufficient textual and contextual information for meaningful linkage. In most cases, identification relied on indirect spatial references (e.g., grid/axis and level information) rather than explicit element identifiers, with only a limited number of RFIs including detailed attributes such as tags, marks, or precise location descriptors that enabled direct mapping. The analysis was conducted under two language model configurations: (i) GPT-4.1 with text-embedding-3-large and (ii) GPT-4.1-mini with text-embedding-3-small, to compare the performance of higher-capacity and lightweight models. Additionally, both configurations were evaluated under two conditions: with intent extraction from RFI text and without intent extraction.

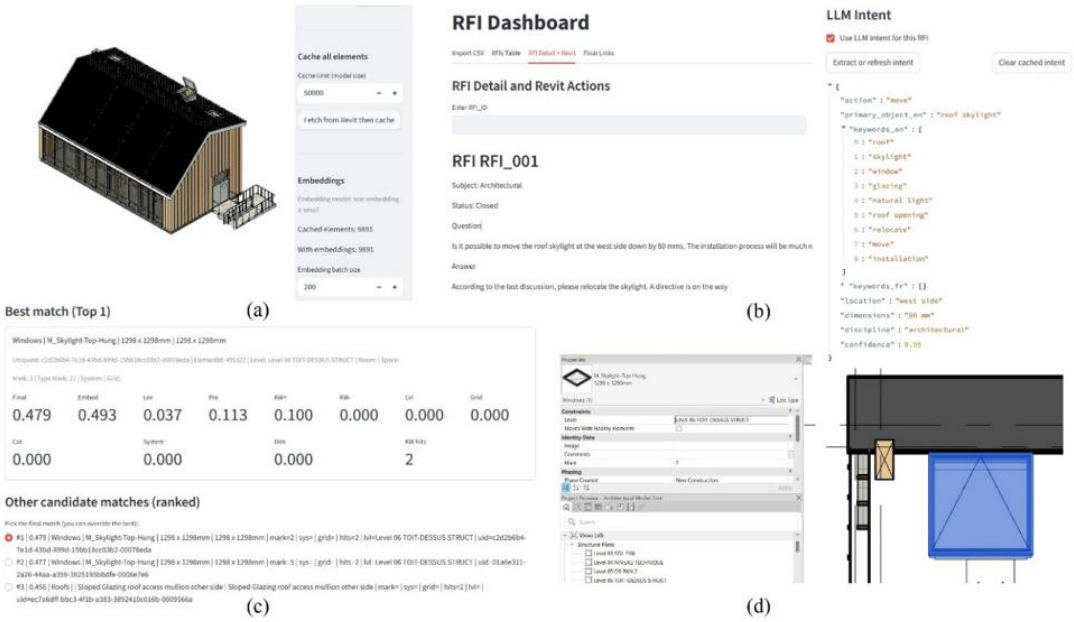


Figure 3: (a) Experimental BIM model, (b) Prototype dashboard with extracted intent from RFI text, (c)Ranked candidate list through the matching pipeline with best match element highlighted separately, (d)Best element selected in Revit

This experimental design enables a systematic evaluation of whether increased model capacity leads to measurable improvements in handling ambiguous and unstructured RFIs.

The results indicate that the framework is particularly effective in narrowing down candidate elements at the category or family level, rather than consistently achieving exact element-level matching. Exact element-level matching was achieved in a limited number of cases, with 4 RFIs correctly linked under configurations with intent extraction and 3 without across both LLM models. These cases typically involved RFIs with explicit identifiers such as element marks, tags, or clear location references. For example, the RFI *"The door supplier from Groupe Brosseau requests clarification on the door pair C-4EX6…?"*, was accurately matched by all models to the element instance due to the presence of specific mark and level information. Although exact matching was limited, the models performed well at category or family-level identification. Of the evaluated RFIs, 19 were correctly predicted by all configurations, 52 by at least one model, and only 1 was not correctly predicted by any model. This demonstrates effective narrowing of the search space by linking RFIs to semantically and functionally similar BIM elements, consistent with real scenarios where RFIs rely on indirect spatial cues. RFIs such as *"Here is a question*

for the architect. What is the window type for R01? According to the elevation and the tab..” were correctly linked by all models due to matching identifiers in both text and BIM metadata. By contrast, a failure case was observed when both the RFI text and BIM metadata lacked sufficient alignment of detail. For instance, the RFI describing parapet conditions based on multiple drawing references and attachments, the absence of explicit identifiers and insufficient corresponding BIM metadata prevented all models from achieving both element-level and category-level matches. This highlights the dependency of the approach on the availability of consistent textual and model-based information. Comparison across models shows that intent extraction does not significantly improve performance. For GPT-4.1, accuracy remains unchanged at 60.27% with and without intent. By contrast, GPT-4.1-mini shows a slight decrease when intent is used with 56.16%, compared to 58.9% accuracy with intent extraction not being used. This suggests that performance is primarily driven by rule-based matching and BIM metadata rather than LLM-derived cues. Additionally, GPT-4.1 shows more stable performance, while GPT-4.1-mini is slightly sensitive to intent inclusion, indicating greater consistency in higher-capacity models. The results indicate that although exact element-level matching is challenging due to ambiguous RFIs, the framework effectively retrieves relevant candidate elements. The high number of RFIs correctly predicted by at least one configuration demonstrates robustness, while the few unmatched cases indicate strong baseline performance even with limited context. Although the prototype shows promising feasibility by narrowing candidates to relevant categories or families, further validation with larger real-world datasets is required.

Discussion

The study demonstrates that linking unstructured RFI text to BIM elements is feasible under appropriate data and modeling conditions. The prototype contextualizes RFI information within the BIM environment by restricting the LLM to intent extraction and relying on rule-based matching with BIM metadata, ensuring transparency and minimizing hallucinations. Analysis indicates that the primary challenge in automated RFI–BIM linkage is not semantic interpretation, but the misalignment between how RFIs are authored and how BIM models represent design information. In the experimental setting, RFIs with clear element, spatial, or dimensional cues were accurately linked, whereas in the real project, references to repeated elements or design intent led mostly to category- or family-level matches with few exact matches. This observation highlights a key limitation of the current framework, which assumes a primary linkage between an RFI and a single BIM element. While this supports structured evaluation, it does not fully reflect real-world practices, where RFIs often relate to multiple elements or assemblies. Accordingly, extending the framework to support multi-element associations is an important direction for future work. A second major limitation is that many RFIs rely on marked drawing attachments for context, making textual content alone insufficient for interpretation. As the current framework operates primarily on text and does not incorporate such attachments, integrating multimodal data sources, including annotated drawings and visual inputs, is a key direction for future research to enhance contextual understanding and matching accuracy. Another limitation of this study is the dependency on the quality and completeness of BIM metadata. The framework relies in part on BIM metadata for rule-based matching, and although supplemented by semantic and lexical similarity measures, its performance may be affected in cases of incomplete or inconsistent model information. A further limitation is that matching weights were empirically selected without sensitivity analysis, which is identified as a future study.

Overall, the findings indicate that RFI BIM linkage is most effective with explicit and well-defined BIM representations. A key limitation of this study is its reliance on a relatively limited dataset, which may affect generalizability. The results underscore the importance of high-quality BIM data, where consistent categorization, complete parameters, and reliable spatial context enhance RFI grounding. In practice, family- or category-level retrieval helps narrow large models to relevant elements for efficient review. Rather than fully automating instance-level matching, the value lies in enhancing human interpretation through improved traceability and context. Future work will address multi-element linking, integration of drawing or multimodal inputs, and validation on larger, more complex real-world datasets.

Conclusion

This study presents a data-driven framework for semantically linking unstructured RFI text to BIM elements using LLM-based intent extraction and BIM metadata-driven reasoning. In the matching pipeline, the LLM is used for intent extraction, while core element matching and decision-making are governed by rule-based BIM reasoning. By combining intent extraction with a rule-based pipeline, the prototype enables context-aware linkage while limiting AI hallucination. Validation with experimental and real-world data demonstrates feasibility and its potential to improve RFI traceability and spatial understanding in BIM environments. The findings suggest that effective integration of LLMs in BIM workflows requires constraining their role to semantic interpretation, while relying on domain knowledge and BIM reasoning for decision-making. Although limitations related to BIM data completeness, vague RFIs, and single-element linkage remain, they reflect real industry conditions and guide future research. Overall, this work provides a foundational step toward more structured, transparent, and traceable RFI–BIM integration.

Acknowledgments

This research was supported by the NSERC Alliance program under grant ALLRP 585937-23, with additional support from the industry partner, CREO Solutions.

References

Afzal, M. et al. (2024) Improving Request for Information (RFI) Processing in Construction Projects using Natural Language Processing (NLP) Techniques. University of Technology Sydney.

Afzal, M., Wai Wong, J.K. and Fard Fini, A.A. (2023) "Unlocking Insights: Analysing Construction Issues in Request for Information (RFI) Documents with Text Mining and Visualisation," IEEE International Conference on Automation Science and Engineering. IEEE Computer Society. Available at: https://doi.org/10.1109/CASE56687.2023.10260517.

Bhat, A.S., Poirier, E.A. and Staub-French, S. (2017) "Investigating the potential of BIM to address project delivery issues," 6th CSCE-CRC International Construction Specialty Conference 2017 - Held as Part of the Canadian Society for Civil Engineering Annual Conference and General Meeting 2017, 2, pp. 955–964.

Bhat, V., Trivedi, J.S. and Dave, B. (2018) "Improving design coordination with Lean and BIM, an Indian case study," IGLC 2018 - Proceedings of the 26th Annual Conference of the International Group for Lean Construction: Evolving Lean Construction Towards Mature Production Management Across Cultures and Frontiers, pp. 1206–1216. Available at: https://doi.org/10.24928/2018/0479.

Dan, W. et al. (2023) "Potentials of Building Information Modelling (BIM) in Managing Variations," Journal of Advanced Research in Applied Sciences and Engineering Technology, 30(2), pp. 439–456. Available at: https://doi.org/10.37934/araset.32.2.439456.

Fernandes, D. et al. (2024) "A GPT-Powered Assistant for Real-Time Interaction with Building Information Models," Buildings 2024, Vol. 14, Page 2499, 14(8), p. 2499. Available at: https://doi.org/10.3390/buildings14082499.

Giel, B.K. and Issa, R.R.A. (2013) "Return on Investment Analysis of Using Building Information Modeling in Construction," Journal of Computing in Civil Engineering, 27(5). Available at: https://doi.org/10.1061/(ASCE)CP.

Hanna, A.S., Tadt, E.J. and Whited, G.C. (2012) "Request for Information: Benchmarks and Metrics for Major Highway Projects," Journal of Construction Engineering and Management, 138(12), pp. 1347–1352. Available at: https://doi.org/10.1061/(ASCE)CO.1943-7862.0000554.

Help | Deployment Options | Autodesk (2026). Available at: https://help.autodesk.com/view/RVT/2026/ENU/?guid=Revit_API_Revit_API_Developers_Guide_Introduction_Getting_Started_Using_the_Autodesk_Revit_API_Deployment_Options_html (Accessed: March 30, 2026).

Help | Revit API Developers Guide | Autodesk (2026). Available at: https://help.autodesk.com/view/RVT/2026/ENU/?guid=Revit_API_Revit_API_Developers_Guide_html (Accessed: March 30, 2026).

Jang, S. et al. (2024) "Automated detailing of exterior walls using NADIA: Natural-language-based architectural detailing through interaction with AI," Advanced Engineering Informatics, 61(9), p. 102532. Available at: https://doi.org/10.1016/j.aei.2024.102532.

Ko, J., Ajibefun, J. and Yan, W. (2025) "Generative AI-powered parametric modeling and BIM for architectural design and visualization," Proceedings of the Design Society, 5, pp. 1943–1952. Available at: https://doi.org/10.1017/pds.2025.10208.

Madireddy, S. et al. (2025) "Large Language Model-Driven Code Compliance Checking in Building Information Modeling," Electronics 2025, Vol. 14, Page 2146, 14(11), p. 2146. Available at: https://doi.org/10.3390/electronics14112146.

Mao, W., Zhu, Y. and Ahmad, I. (2007) "Applying metadata models to unstructured content of construction documents: A view-based approach," Automation in Construction, 16(2), pp. 242–252. Available at: https://doi.org/10.1016/J.AUTCON.2006.05.005.

Mohamed, S., Tilley, P.A. and Tucker, S.N. (1999) "Quantifying the Time and Cost Associated with the Request for Information (RFI) Process in Construction," Construction Innovation: information, process, management [Preprint].

Morales, F. et al. (2022) "Potential Application of BIM in RFI in Building Projects," Buildings 2022, Vol. 12, Page 145, 12(2), p. 145. Available at: https://doi.org/10.3390/BUILDINGS12020145.

NBIMS-USTM V3 – National Institute of Building Sciences (2026). Available at: https://nibs.org/nbims/v3 (Accessed: March 31, 2026).

Papajohn, D. et al. (2018) "Exploring Potential Delays Associated with Requests for Information in CM/GC Highway Construction," Construction Research Congress 2018: Infrastructure and Facility Management - Selected Papers from the Construction Research Congress 2018, pp. 640–649.

Yang, L. and Zhang, Z. (2024) "Using RAG, LLMs, and langchain to create a query application on open repositories," Proceedings of the ACM/IEEE Joint Conference on Digital Libraries [Preprint]. Available at: https://doi.org/10.1145/3677389.3702488.

Zheng, J. and Fischer, M. (2023) "BIM-GPT: a Prompt-Based Virtual Assistant Framework for BIM Information Retrieval." Available at: http://arxiv.org/abs/2304.09333 (Accessed: February 3, 2026).

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

EXPLAINABLE AI FOR REQUIREMENT TRACEABILITY IN CONSTRUCTION SYSTEMS ENGINEERING

Timon Kuijper, Ranjith K. Soman, Tong Wang, and Sander van Nederveen
Delft University of Technology, Delft, the Netherlands

Abstract

Requirement verification in integrated contract models, such as Design and Construct, remains a manual, document-centric, and error-prone activity. While technical verification through Automated Code Compliance has advanced significantly, process requirements are typically verified through human inspection of textual documentation. This research addresses this gap by developing a novel artefact architecture that integrates general-purpose Large Language Models, the RASE mark-up language, and Knowledge Graphs to automate traceability between contractual obligations and contractor deliverables. Tested through iterative Design Science Research, the final prototype achieved a precision of 1.000 and accuracy of 0.968, offering a transparent, "white-box" framework for contractual auditability.

Introduction

Requirements traceability – the ability to track relationships between requirements and project deliverables – is essential in systems engineering and contract management for validating that project outcomes meet the specified requirements from start to finish (SixSigma, 2025). In large construction projects, the contractor must demonstrate that the delivered work complies with the Employer's requirements. Ensuring every requirement is satisfied (and providing evidence for it) is critical to project success in terms of cost, time, and quality (Hoehne, 2014). Traditionally, traceability of process-oriented requirements in construction is maintained manually through document management and version control systems (Guru, 2025). This manual approach is time-consuming, error-prone, and difficult to scale, especially given the growing volume of digital project documents and the involvement of many stakeholders. As a result, full end-to-end traceability is often impractical in current practice, and important links between requirements and evidence may be missed or lost.

Recent advances in Artificial Intelligence (AI), particularly Large Language Models (LLMs), offer new opportunities to automate requirements traceability by automatically recognizing and maintaining semantic links between textual artifacts. Research on AI-based traceability to date has focused largely on automated code compliance checking for technical requirements and on contract analysis tools, but there is a notable gap in applying AI to link process requirements with verification evidence in construction projects (Karim-Jallow et al., 2014; Ismail et al., 2017; Zheng et al., 2025). No prior work, to the authors' knowledge, has systematically explored a zero-shot LLM approach for automatically verifying process requirements in design-and-construct contracts by linking them to evidence in contractor documents. Moreover, the black-box nature of powerful LLMs raises concerns about transparency and explainability of AI-generated traceability links.

This paper addresses the above gap by presenting an AI-assisted requirement traceability approach that leverages a general-purpose LLM combined with a knowledge graph to automatically link contractual process requirements to corresponding evidence in project documents. The approach is designed to improve the efficiency and accuracy of traceability while also enhancing transparency (through a formal knowledge graph representation of the trace links). We developed and evaluated a prototype tool following a Design Science Research methodology, with iterative refinement and expert feedback. In a case study, the prototype achieved high accuracy in identifying valid requirement-evidence links, indicating that such an AI-assisted traceability system can significantly reduce manual effort and error in construction compliance verification.

The remainder of the paper is structured as follows. Background section reviews related literature on requirements traceability and AI applications in construction, identifying the research gap. Methods section explains the Design Science Research method used to develop and evaluate the solution. Implementation section describes the AI-assisted traceability tool, including its architecture and components. The following section presents the results of evaluating the tool's performance and usability. The final section discusses the findings, implications for research and practice, limitations, and future work

Literature review

Requirements traceability in construction

Requirements traceability refers to the ability to link requirements to corresponding artifacts that demonstrate their implementation or verification (Madni and Sievers, 2018). In construction projects, particularly within a Systems Engineering framework, traceability ensures

that contractual requirements are systematically verified against design documents, management plans, and other deliverables (SixSigma, 2025). Maintaining these links provides transparency, supports impact analysis, and reduces the risk of overlooked obligations.

In practice, traceability in construction remains largely manual. Engineers typically rely on traceability matrices, spreadsheets, or document annotations to map requirements to evidence (Guru, 2025). This process is time-consuming and prone to omissions, especially in large Design & Construct projects involving extensive documentation (Karim-Jallow et al., 2014). Process-oriented requirements, which describe how work must be conducted rather than what must be built, are particularly challenging to verify because evidence is often embedded in lengthy textual documents rather than structured models. As a result, complete and continuously updated traceability is difficult to achieve in current industry practice.

AI approaches to compliance and traceability

Recent advances in natural language processing and Large Language Models (LLMs) have stimulated research into automated traceability and compliance checking. In software engineering, information retrieval and machine learning techniques have been used to recover trace links between requirements and code or design artifacts (Rempel and Mäder, 2016). More recently, retrieval-augmented generation approaches using LLMs have demonstrated improved trace link recovery performance (Hey et al., 2025). In the architecture, engineering and construction (AEC) domain, AI applications have primarily focused on automated technical code compliance checking using rule-based systems, ontologies, or BIM-integrated workflows (Ismail et al., 2017; Yang et al., 2023). Parallel research has explored contract analysis using knowledge graph-enhanced LLMs (Zheng et al., 2025). While these studies show the potential of combining semantic representations with LLM reasoning, they do not directly address automated verification of contractual *process* requirements through evidence linking.

The RASE (Requirement-Applicability-Selection-Exception) methodology provides a structured way to extract normative statements from regulatory and contractual texts (Hjelseth and Nisbet, 2011). RASE has been shown to support automated processing of construction regulations by isolating enforceable requirements. In parallel, knowledge graph representations have been proposed to model requirements and compliance relationships in a transparent and queryable manner (Wang and El-Gohary, 2023). These approaches suggest that combining semantic structuring techniques with AI-driven text understanding could enable more robust verification workflows.

Research gap

Despite progress in automated compliance checking and LLM-supported document analysis, limited research has addressed the automated verification of contractual process requirements in construction projects. Existing approaches either focus on technical design compliance, require significant rule engineering, or rely on supervised training data. The use of a zero-shot LLM approach, combined with a knowledge graph to produce auditable requirement–evidence mappings for process requirements, remains largely unexplored.

This paper addresses this gap by developing and evaluating an AI-assisted verification tool that automatically links contractual process requirements to supporting evidence in contractor documentation, using a hybrid LLM and knowledge graph architecture.

Method

This study followed the Design Science Research Methodology (DSRM) (Peffers et al., 2007) to develop and evaluate the AI-assisted traceability tool. Design Science Research (DSR) is a problem-solving paradigm that iteratively designs, builds, and evaluates artifacts (in this case, a software tool) to address identified problems, thereby contributing new knowledge. The research process was structured according to the DSRM steps proposed by Peffers *et al.* (Peffers et al., 2007):

(1) Problem Identification and Motivation: Through preliminary literature review and expert interviews, we established that manual requirement traceability in large construction projects is inefficient and error-prone, and that no adequate automated solution exists. The need for an AI-based approach was justified by the increasing volume of documentation and the potential of LLMs to interpret complex texts. We defined the core problem as: "How can AI support automated traceability between contractual process requirements and contractor documentation in construction projects?"

(2) Define Objectives for the Solution: Based on the problem analysis, we set functional and non-functional objectives for the artifact. Key functional objectives included: automatically identifying requirement statements in contract documents; extracting evidence of compliance from contractor documents; and establishing trace links between them. Non-functional objectives included a high degree of accuracy and recall in link identification (targeting at least 90 % accuracy and 90% recall to be comparable with manual quality), high precision (to minimize false positives, as an incorrect assertion of compliance is unacceptable), and usability/transparency for end-users. These objectives were operationalized into measurable success criteria (e.g., performance metrics and a usability score). We also aimed for the solution to be zero-shot, meaning it should work without project-specific training, relying on prompt engineering and general LLM capabilities, to facilitate adoption in new projects.

(3) Design and Development: We designed an artifact architecture that integrates a general-purpose LLM with a knowledge graph. The design and implementation details are described in following section. Importantly,

we adopted an iterative development approach (typical in DSR): the artifact went through multiple design iterations, each adding improvements or new features to overcome shortcomings identified in the previous iteration. For example, after an initial prototype (Iteration 1) showed low recall, we refined the prompt and graph constraints to ensure all requirements would be captured (Iteration 2). Further iterations introduced a confidence scoring mechanism and automated consistency checks. Throughout this process, design decisions were informed by both literature (e.g., using RASE mark-up and SHACL constraints as derived from prior work) and practitioner input.

(4) Demonstration: We demonstrated the artifact using a real-world case study. The case involved a Design and Construct infrastructure project in the Netherlands. The input documents included a Project Requirements document (a contract appendix listing process requirements from the Employer) and a Project Management Plan (PMP) submitted by the contractor as evidence of how requirements are addressed. The tool was used to automatically generate traceability links between the requirements and the PMP. Domain experts (project engineers) observed the demonstration to validate that the identified links correspond to actual compliance evidence.

(5) Evaluation: We evaluated the artifact's performance and utility using both quantitative and qualitative methods. Quantitatively, we measured classical classification performance metrics for the trace link predictions: True Positives (TP), False Positives (FP), True Negatives (TN), False Negatives (FN), accuracy, precision, recall, and F_1 score (harmonic mean of precision and recall). These metrics were computed by comparing the AI-generated links against a ground truth traceability set established via manual verification of the case documents. We applied this evaluation after each major iteration of the artifact. Qualitatively, we gathered feedback on usability and usefulness through a focus group and a survey. After the final prototype was demonstrated, $N = 5$ construction engineering professionals used the tool on sample queries and then filled out a System Usability Scale (SUS) questionnaire. The SUS provides a standardized score out of 100 for perceived usability. We also collected open-ended feedback about what users liked and what could be improved. This multi-faceted evaluation approach follows guidance from DSR literature that emphasizes not only technical performance of artifacts but also their fit with user needs and contexts (Venable et al., 2016).

(6) Communication: Finally, we document the problem, artifact, and findings in this paper and in a companion thesis report, to disseminate the knowledge to both academic and practitioner audiences. The iterative DSR approach ensured that the research was grounded in solving a relevant practical problem while also contributing general insights (e.g., how LLMs can be combined with knowledge graphs for traceability).

Implementation of the AI-assisted traceability tool

System overview

The developed artifact is a prototype tool that supports verification of contractual process requirements by automatically producing an auditable *requirement–evidence mapping* (See Figure 1). The system combines (i) a Large Language Model (LLM) for natural-language understanding and evidence extraction and (ii) a knowledge graph (KG) to store requirement entities, evidence snippets, and their links in a structured and queryable form. The KG acts as a traceability backbone: each verification decision is represented as a link between a requirement node and one or more evidence nodes, including provenance metadata (e.g., document and section identifiers).

LLM-driven requirement extraction and evidence identification

The tool operates in two stages. First, it extracts requirement statements from the contractual requirements document. Second, for each extracted requirement, it searches for supporting evidence in the contractor document(s) (e.g., project management plan) and returns a quoted excerpt plus a document location reference. Prompting is designed to be conservative: the LLM is instructed to only propose a link when it can point to explicit textual evidence; otherwise, it returns ''no evidence found''. This design choice prioritizes avoiding false positives (incorrectly asserting compliance) over maximizing coverage.

Knowledge graph representation

All outputs are stored as RDF-like triples in a knowledge graph with a lightweight ontology. Core classes include `Requirement` and `Evidence`, with a predicate such as `hasEvidence` linking a requirement to an evidence excerpt (See Figure 2). Evidence nodes store provenance fields (document name, chapter/section identifiers) to support auditability and facilitate queries such as: ''Which requirements have no evidence?'' or ''Which document sections satisfy the most requirements?''

Design iterations (build–evaluate cycles)

The artifact was developed using a Design Science Research (DSR) approach with multiple **design iterations**. In this paper, an **iteration** is defined as one complete *build–evaluate cycle*: (1) implement a concrete change to the artifact (e.g., prompt structure, document chunking strategy, evidence provenance requirements, or KG validation rules), and (2) re-run the same case study evaluation against a fixed ground truth to quantify the impact on link-quality metrics (accuracy, precision, recall, and F_1). Each iteration was motivated by observed error patterns (false negatives/false positives) in the previous iteration and aimed to improve verification completeness while maintaining conservative, audit-friendly outputs. Table 1 summarizes the iterations and their key design changes.

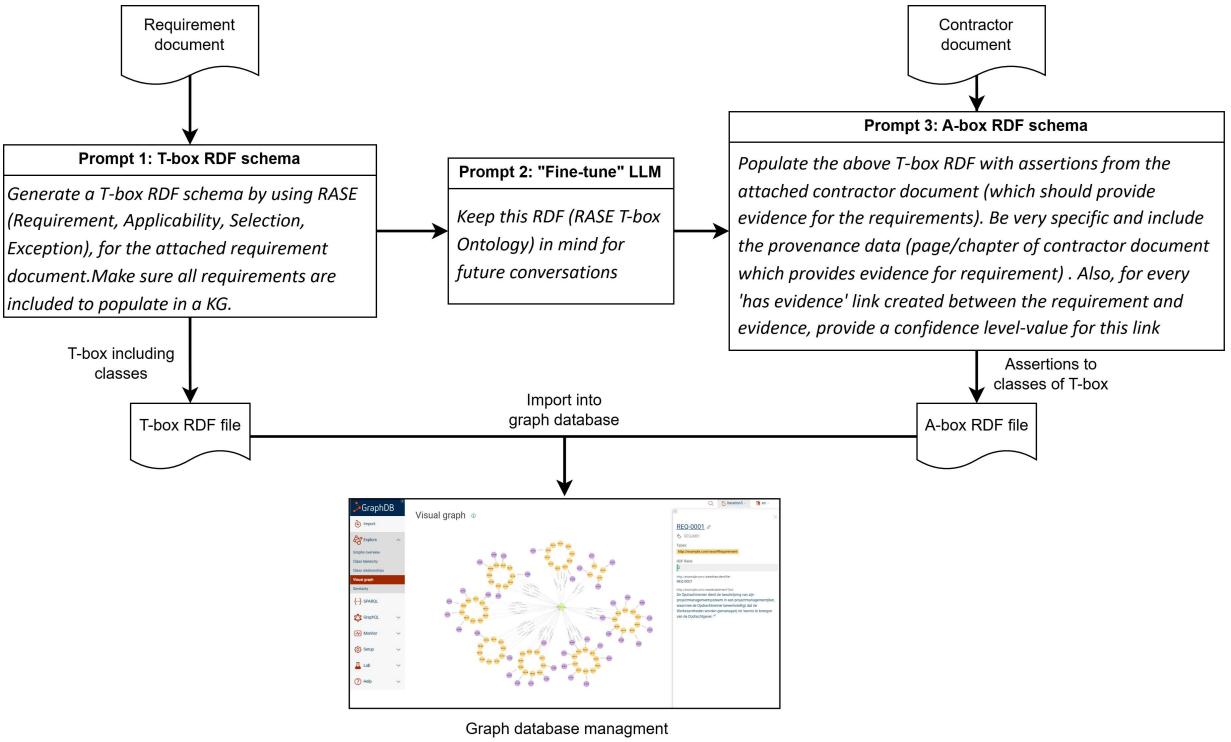


Figure 1: AI-assisted requirement traceability workflow combining T-box generation, LLM-based A-box population, and graph database integration.

Results

We evaluated the performance of the AI-assisted traceability tool on the case study project. The quantitative results demonstrate that through iterative refinement, the artifact achieved a high level of accuracy in identifying valid traceability links between requirements and evidence. Table 2 summarizes the classification performance metrics over four iterations of the artifact development. In the first iteration, the tool's recall was very low (around 17%), indicating it missed the majority of actual requirement-evidence links; accuracy was also low (17.2%). However, precision was perfect (100%) even in that iteration, meaning that although the tool found few links, those it did find were correct (no false positives). This high precision can be attributed to the conservative approach in the LLM prompting as described earlier. The low recall and accuracy in Iteration 1 highlighted the need to capture more links (reduce false negatives). By Iteration 2, after improving the prompt to ensure no requirements were skipped, recall jumped to 67.9% and accuracy to 73.4%. Further enhancements in Iteration 3 (such as better handling of partial matches and refining evidence extraction) yielded about 73.6% recall. Finally, Iteration 4 introduced a confidence scoring mechanism and additional knowledge graph constraints to catch missing links; the tool then achieved 96.6% recall and 96.8% accuracy, essentially meeting the target of $> 90\%$ on both. Precision remained 100% throughout all iterations, meaning the system did not produce any incorrect links in our test dataset. The F_1 score, which balances precision and recall, improved dramatically from 29.3% in Iteration 1 to 98.3% in Iteration 4, indicating a well-balanced and robust performance by the end.

In absolute terms, for the final iteration, out of 64 total requirements in the test contract, the tool correctly linked 57 requirements with relevant evidence found in the PMP (True Positives) and correctly identified that 3 requirements had no evidence in that document (True Negatives, which in this context means those requirements were indeed not addressed in the PMP and likely require evidence from other documents or were out of scope). There were 2 False Negatives, i.e., requirements for which the tool failed to find the existing evidence in the PMP. Notably, there were 0 False Positives in any iteration, so the precision is 1.0 (the tool never claimed an evidence link where there was none). By comparing with manual verification, we found that most of the false negatives in early iterations were due to the LLM not recognizing evidence phrased very differently from the requirement wording. After ontology enhancements and additional prompt context, only 2 requirements remained unlinked (false negatives) in the final iteration. Those cases were ones where the evidence was implicit or scattered in the document, indicating a limit to the LLM's ability to infer compliance without explicit statements (a topic for future improvement). Beyond the binary performance metrics, the tool provided qualitative insights. For example, as mentioned, it identified inconsistencies in the contractor document's chapter numbering while extracting evidence. In one case, the LLM correctly pointed to a section titled "Risk management"

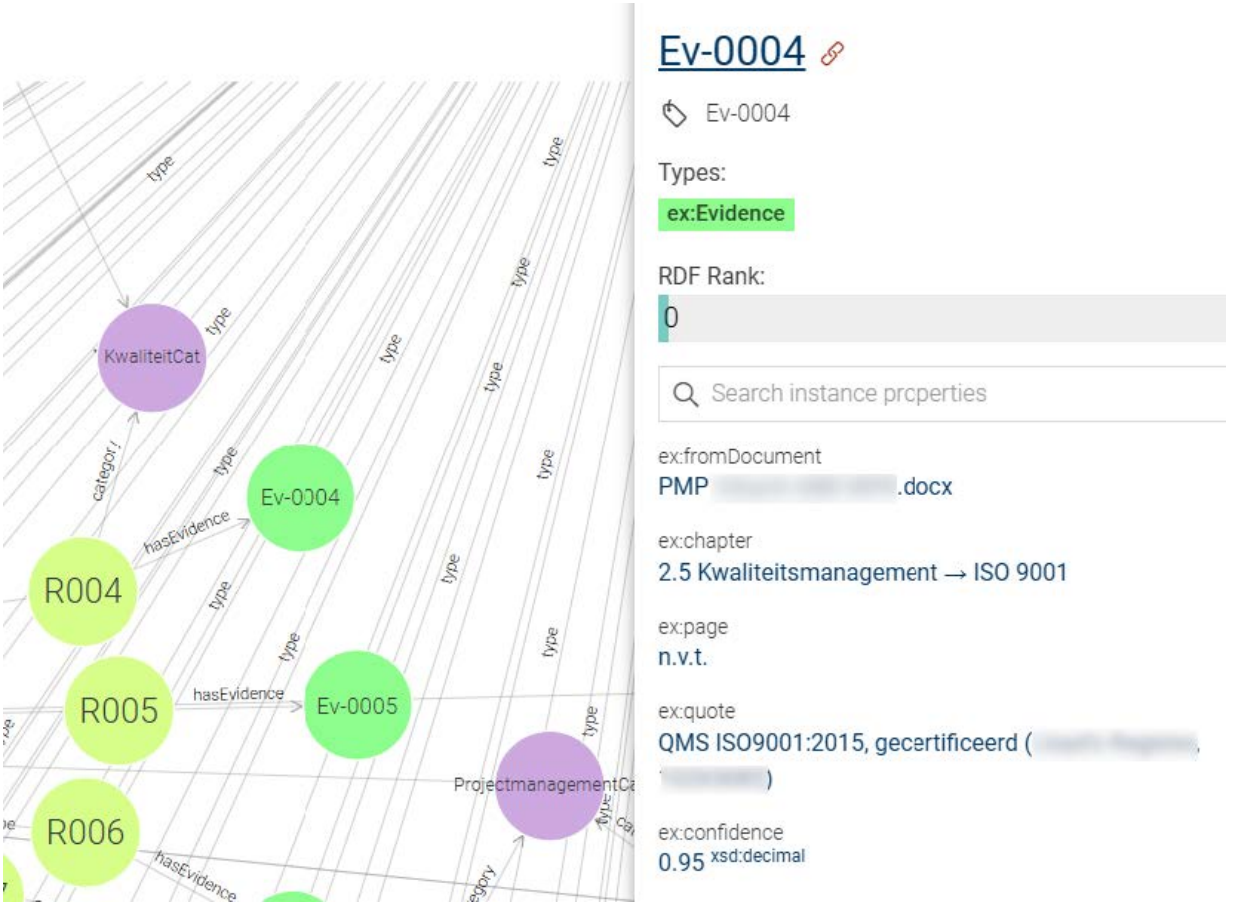


Figure 2: Knowledge Graph Representation

as evidence for a requirement on risk coordination, but flagged it under chapter 4 whereas the contractor's table of contents had it under chapter 3. This led to discovering a documentation error by the contractor. Such findings suggest the AI tool not only checks compliance but can also help improve document quality and consistency. We also evaluated the usability and acceptance of the tool via the focus group and SUS questionnaire. The average SUS score given by the five participants was **78.5**. According to standard SUS interpretations (Lewis, 2018; Sauro, 2018), this score falls in the 80th percentile, corresponding to a rating above "good" (approximately a "B+" on a grading scale). An SUS above 68 is generally considered above average; thus, our result indicates that users found the prototype reasonably easy to use. They particularly appreciated the ability to click a requirement and immediately see the evidence text, which one participant noted "saves a lot of time compared to flipping through PDFs manually." Some participants were initially skeptical about AI "reading" their documents, but after seeing the accuracy of results, they expressed increased trust. The focus group discussion revealed that users valued the knowledge graph's transparency — it reassured them that the AI wasn't making arbitrary guesses but actually linking to specific document sections. However, the participants also pointed out areas for improvement. In a few instances, the evidence text returned by the tool was quite lengthy; users suggested highlighting the exact sentences that matter. They also desired a clearer indication of confidence for each link (even though the system internally had confidence scores, the front-end only showed the link or no link). This feedback aligns with our own observation that a confidence threshold could be used to only show high-confidence links by default, with lower-confidence suggestions flagged for human review. Overall, the results demonstrate that the AI-assisted traceability tool can dramatically improve the traceability process: it achieved near-perfect precision and very high recall in linking requirements to evidence in a complex real document, and it did so in a matter of minutes (the end-to-end runtime for analyzing 64 requirements against a 150-page PMP was under 10 minutes on an LLM API, which is negligible compared to days of manual work). The high usability score further suggests that with a bit more refinement, practitioners would be willing to adopt such a tool in their workflow.

Discussion

The successful demonstration of the prototype confirms that leveraging LLMs for requirement traceability in construction is not only feasible but highly effective under the tested conditions. In this section, we reflect on the impli-

Table 1: Design iterations of the AI-assisted requirement verification tool.

Iter.	Implemented change	Design focus
1	Baseline pipeline combining LLM-based requirement extraction, evidence identification, and knowledge graph storage of requirement–evidence links.	Establish end-to-end feasibility of automated verification workflow.
2	Revised prompt structure to enforce complete requirement coverage and systematic section-wise evidence scanning.	Improve completeness of requirement handling and structured evidence search.
3	Refined evidence prompts to require explicit textual quoting and provenance references; improved document chunking strategy.	Strengthen grounding of links and standardize evidence extraction format.
4	Integrated additional knowledge graph validation constraints and final prompt refinements for handling implicit references and edge cases.	Increase structural robustness and consistency of requirement–evidence mapping.

Table 2: Traceability link classification performance per iteration (case study evaluation)

Metric	Iteration 1	Iteration 2	Iteration 3	Iteration 4
Accuracy	0.172	0.734	0.781	0.968
Precision	1.000	1.000	1.000	1.000
Recall	0.172	0.679	0.736	0.966
F_1-score	0.293	0.809	0.848	0.983

cations of these findings, discuss the contribution of this work in context, and outline limitations and future work.

Implications and contributions

The research provides evidence that an AI-assisted approach can significantly reduce the manual effort required for verifying process compliance in construction projects. By automatically generating traceability links, the tool addresses a pain point for contractors, who typically must allocate considerable time to compile compliance matrices for each requirement. The near-100% precision means that users can trust the links that are present, focusing their attention on verifying missing links (false negatives) rather than double-checking every provided link for correctness. In practice, this flips the verification process from a manual search for evidence to an AI-assisted review of suggested evidence, potentially saving time and catching errors that humans might overlook due to fatigue or oversight. One of the experts in our focus group mentioned that the tool "would allow junior engineers to do in hours what used to take an experienced engineer days," hinting at productivity gains and knowledge transfer benefits. From a scientific perspective, this work contributes a novel integration of technologies: it combines an advanced natural language understanding model (Open AI GPT-5) with semantic web techniques (RASE markup, RDF knowledge graph, SHACL constraints) to achieve a level of performance that neither approach could easily attain alone. Pure LLM-based solutions might achieve good recall but suffer from hallucinations, while pure rule-based/ontology solutions struggle with the ambiguity of natural language. Our hybrid approach demonstrates a synergistic effect: the knowledge graph and constraints guide the LLM and provide structure, while the LLM provides the flexible understanding needed to interpret diverse requirement and evidence texts. This approach is generalizable beyond the specific case; it could be applied to other domains where requirements (or regulations) need to be traced to evidence (e.g., safety certification, quality assurance in manufacturing, or even legal compliance in finance). Thus, a broader contribution is showing how zero-shot LLM capabilities can be harnessed in a controlled, explainable way for traceability tasks. Furthermore, this research adds to the Design Science knowledge base by illustrating the application of DSRM in the construction IT context. We followed a structured iterative design-evaluate cycle and have documented the evolution of the artifact. The improvement across iterations (Table 2) serves as a case example of how setting concrete objectives (like recall > 90%) and continuously

evaluating against them leads to a more effective artifact. It also underscores the importance of evaluation at multiple levels: our evaluation was not just technical (metrics) but also involved end-user feedback, which is crucial for any AI tool to be practically viable. The high SUS score in the final evaluation is as important a result as the high F_1-score; both technical performance and user acceptance must align for successful adoption.

Limitations

Despite the encouraging results, there are several limitations to acknowledge. First, the evaluation was conducted on a single case project (albeit a fairly large and representative one). While the results show the tool can work extremely well on that case, projects can vary. Different contract structures, requirement phrasing styles (especially if not in English, whereas our LLM was primarily English-trained), or evidence document types might pose new challenges. For instance, some projects might have requirements scattered across multiple documents or require linking not just one contractor report but a chain of documents (like design drawings or test reports). Our current prototype handles one requirements document and one evidence document; scaling to multiple heterogeneous sources will require further development. Second, the reliance on a third-party LLM (OpenAI's GPT 5) introduces practical issues such as cost and data privacy. Running GPT-5 on large documents can be expensive, though we tried to optimize prompts to stay within reasonable token limits. More critically, sending potentially sensitive project documents to an external API can be problematic for companies concerned with confidentiality. This limitation could be mitigated by on-premise LLM solutions or fine-tuned smaller models in the future. There is ongoing research to distill or adapt large models for specific tasks which could be explored to reduce dependency on external APIs. Another limitation is that, while precision was perfect in our tests, we cannot guarantee it will always remain so as we generalize. We tuned the system to be conservative to avoid false positives; this worked for the case data, but a very differently structured document might confuse the LLM to produce an incorrect link. If a false positive (hallucination) does occur, it could mislead users into thinking a requirement is satisfied when it is not, which is a serious risk. Therefore, deploying this in a real project would require thorough validation and perhaps maintaining a human in the loop who at least spot-checks the evidence for each requirement. Additionally, our method of splitting the evidence document by sections and sequentially querying the LLM might miss evidence that is, say, a combination of two sections or implied across the document. More sophisticated retrieval (like using vector embeddings to fetch the top relevant passages for a requirement) could enhance the evidence search phase. Finally, regarding the evaluation method: our ground truth traceability links were established via manual review by the researchers (and double-checked with the help of the domain experts). There is some degree of subjectivity in determining what counts as "evidence" for a requirement, especially for process-oriented requirements that might not have a binary satisfied/not satisfied state. We treated the contractor's own verification statements in the PMP as the evidence of compliance. It is possible that in some cases the contractor's documentation might claim compliance where an external auditor might disagree. In our metrics, we assumed the contractor's documentation combined with our judgment is the truth. A more rigorous validation in the future could involve independent auditors verifying the links as well.

Future work

Building on the prototype's success and the limitations identified, several avenues for future work are evident. One important next step is to generalize and test the tool on additional projects. We plan to apply the approach to other case studies, possibly in different domains (for example, rail or highway projects, or even outside construction such as aerospace systems engineering) to evaluate how well the LLM handles different writing styles and terminologies. This could be complemented by fine-tuning the LLM on a corpus of requirement-evidence pairs (if such data can be accumulated) to see if a fine-tuned model could achieve similar results more efficiently. Another future improvement is enhancing the explainability of the AI decisions. Integrating techniques from explainable AI (XAI) could allow the system to provide a justification for each trace link, beyond just the raw text. For example, the tool could highlight which keywords or phrases in the requirement matched with the evidence, or even output a natural language justification like "Requirement X is satisfied in section Y of the PMP where it states '...' which corresponds to the required process." Such features would further increase user trust and understanding of the AI's reasoning. Introducing a confidence threshold for suggestions is also worth exploring. We observed that by iteration 4, the few missed links could potentially be discovered by allowing the AI to be more "creative" (in iteration 4 it might have still erred on the side of caution). If we allow a slightly lower precision in exchange for 100% recall, we could then flag the uncertain links for human review. A hybrid approach could be: the tool marks some links as "high confidence" (auto-verified) and others as "low confidence – needs human check." This way, no potential evidence is missed, and the human verifier's effort is focused only on the questionable cases. From a technical standpoint, integrating a vector-based retrieval step could improve evidence discovery. Instead of or in addition to scanning document sections sequentially, we can use embeddings to find the most semantically similar passages to each requirement and present those to the LLM. This retrieval augmented generation approach (Hey et al., 2025) could help when evidence is not explicitly a near copy of requirement wording. Finally, expanding the knowledge graph to capture more context could be valuable. Currently, our KG links requirements to evidence found in one document. In

reality, evidence of compliance could span multiple documents (e.g., a test report plus a certification). The ontology could be extended to represent complex verification chains (for instance, a requirement is satisfied by a procedure which is evidenced by a test, which is documented in a report). This aligns with the broader vision of Model-Based Systems Engineering (MBSE) where every requirement can be traced to model elements, test cases, and results. Our work is a step in that direction by automatically populating a knowledge base; future work might integrate with MBSE tools or requirement management databases to pull in more types of trace links.

Conclusion

This research demonstrates a practical AI-assisted solution to requirement traceability in construction projects. By combining the strengths of LLMs in understanding unstructured text with the rigor of knowledge graphs for information organization and validation, we achieved a tool that can automate a large portion of the traceability process with high accuracy and user acceptance. The Design Science approach ensured the artifact was iteratively improved to meet the desired performance. The resulting prototype not only contributes a new method for managing requirements in construction but also provides a blueprint for tackling similar problems in other domains where requirements verification is crucial. As AI technologies continue to evolve, we expect such hybrid human-AI systems to become an integral part of the systems engineering toolkit, augmenting human capabilities in managing complexity and ensuring compliance in large-scale projects.

References

Guru (2025). Documentversiebeheer: Een uitgebreide gids voor projectmanagers. `https://www.getguru.com/nl/reference/document-version-control`.

Hey, T., Fuchss, D., Keim, J., and Koziolek, A. (2025). Requirements traceability link recovery via retrieval-augmented generation. In Proc. REFSQ (Requirements Engineering: Foundation for Software Quality), pages 381–397.

Hjelseth, E. and Nisbet, N. (2011). Capturing normative constraints by use of the semantic mark-up rase methodology. In Proc. CIB W78-W102 Conference, pages 1–10.

Hoehne, O. M. (2014). On motivating people to implement systems engineering: Getting from the necessary to the impossible. INCOSE International Symposium, 24(1):362–377.

Ismail, A. S., Ali, K. N., and Iahad, N. (2017). A review on bim-based automated code compliance checking system. In Proc. International Conference on Research and Innovation in Information Systems (ICRIIS), pages 1–6.

Karim-Jallow, A., Demian, P., Baldwin, A., and Anumba, C. (2014). An empirical study of the complexity of requirements management in construction projects. In Engineering, Construction and Architectural Management (Proc. ICE), volume 21, pages 505–531.

Lewis, J. R. (2018). The system usability scale: Past, present, and future. International Journal of Human-Computer Interaction, 34(7):577–590.

Madni, A. M. and Sievers, M. (2018). Model-based systems engineering: Motivation, current status, and research opportunities. Systems Engineering, 21(3):172–190.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3):45–77.

Rempel, P. and Mäder, P. (2016). Preventing defects: The impact of requirements traceability completeness on software quality. IEEE Transactions on Software Engineering, 43(8):777–797.

Sauro, J. (2018). 5 ways to interpret a sus score. `https://measuringu.com/interpret-sus-score/`.

SixSigma (2025). Matrix voor traceerbaarheid van vereisten: Een complete gids voor projectsucces.

Venable, J., Pries-Heje, J., and Baskerville, R. (2016). Feds: A framework for evaluation in design science research. European Journal of Information Systems, 25(1):77–89.

Wang, X. and El-Gohary, N. (2023). Deep learning-based relation extraction and knowledge graph-based representation of construction safety requirements. Automation in Construction, 147:104696.

Yang, M., Zhao, Q., Zhu, L., Meng, H., Chen, K., Li, Z., and Hei, X. (2023). Semi-automatic representation of design code based on knowledge graph for automated compliance checking. Computers in Industry, 150:103945.

Zheng, C., Wong, S., Su, X., Tang, Y., Nawaz, A., and Kassem, M. (2025). Automating construction contract review using knowledge graph-enhanced large language models. Automation in Construction, 175:106179.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

ENABLING CROSS-STUDY COMPARISON: A FRAMEWORK FOR AUTOMATED BIM-QA EVALUATION

Sylvain Hellin[1,2], Stefan Fuchs[1,2], Stavros Nousias[1,2], André Borrmann[1,2]
[1]Chair of Computing in Civil and Building Engineering, Technical University of Munich, Germany
[2]TUM Georg Nemetschek Institute, Technical University of Munich, Munich, Germany

Abstract

Driven by LLM advances, Question Answering (QA) systems for Building Information Models (BIM) have proliferated, yet no standardized methods exist for cross-study comparison. We present an automated evaluation pipeline employing LLM judges to assess BIM-QA answers across five quality criteria and validate it through inter-rater agreement analysis with two domain experts and two LLM judges on 74 IFC-Bench question-answer pairs. LLM judges achieve high inter-rater reliability (Krippendorff's α = 0.70–1.00), exceeding human experts (α = 0.32–0.57), and show good agreement with expert consensus (α = 0.48–0.85), positioning the pipeline as a suitable solution for automated, reproducible BIM-QA benchmarking.

Introduction

Building Information Models (BIMs) store comprehensive building data, yet accessing this information requires specialized software and domain expertise. This creates a barrier for many project stakeholders, such as facility managers, owners, or contractors, who need timely access to building information but lack the technical skills to navigate complex BIM software. Natural language Question Answering (QA) systems offer a promising solution by enabling stakeholders to query building models using natural language (Zheng and Fischer, 2023; Hellin et al., 2025). As these systems transition from research prototypes toward practical deployment, a critical question arises: how can we systematically evaluate their answer quality?

Traditional exact-match metrics fail when answers vary in phrasing, and automated metrics like BLEU and ROUGE correlate poorly with human preference when multiple valid responses exist (Liu et al., 2023). This limitation reveals a deeper challenge: the BIM-QA field currently lacks standardized evaluation methodology. Research groups typically develop custom benchmarks and evaluate on non-shared datasets, making cross-study comparison difficult and hindering cumulative scientific progress.

While IFC-Bench provides a shared benchmark dataset (Hellin et al., 2025), a consistent evaluation methodology remains missing. A validated, reproducible evaluation method would enable meaningful performance comparison across systems and research groups. For non-deterministic LLM-based systems, such standardized evaluation is essential before practitioners can trust them for professional use.

This paper addresses this gap by introducing an automated LLM-based evaluation pipeline for BIM-QA systems. The pipeline implements a multi-dimensional evaluation framework comprising a four-category question taxonomy and five binary quality criteria (Hellin et al., 2026), detailed in the Evaluation Framework section. While multi-dimensional evaluation offers richer assessment than binary accuracy, its practical utility depends on reliable and automated application across evaluators. Specifically, this validation targets inter-rater *reliability*: whether different evaluators reach the same judgments when assessing the same answers against the same ground truth. Assessing evaluation *validity* for novel questions without established ground truth remains a separate, open direction. To validate the pipeline and the underlying evaluation framework, we investigate the following research questions:

- **RQ1:** What discriminative power and redundancy do the evaluation criteria exhibit across BIM-QA systems of varying quality?
- **RQ2:** How reliably can the proposed framework be applied across different evaluators?
- **RQ3:** How well do LLM judges approximate human expert evaluation?

The experimental results yield three key findings:

- *Criteria differ in discriminative power and redundancy (RQ1).* *Transparency* and *Completeness* best differentiate BIM-QA systems (56–57 percentage points (pp) cross-system range). *Relevance* shows ceiling effects (38pp range) and high correlation with *Completeness* (ρ = 0.76–0.81), suggesting potential consolidation to four criteria.
- *LLM judges achieve higher inter-rater reliability than human experts (RQ2).* Percentage agreement exceeds 77% across all evaluator combinations and criteria. However, chance-corrected reliability (α) reveals a clear split: LLM-LLM agreement is high (α = 0.70–1.00), while human-human agreement remains lower (α = 0.32–0.57).
- *Automated evaluation enables standardized cross-study comparison (RQ3).* Post-hoc consensus analysis confirms that LLMs align with calibrated expert judgment better than uncalibrated experts agreed with each other. This alignment holds across two model families (GLM-

4.7 and Gemini), positioning LLM-as-judge as a practical solution to the field's fragmentation problem.

Related work

BIM question answering systems

The evolution from BIM Natural Language Query (NLQ) systems to generative BIM-QA systems has transformed evaluation requirements. Traditional BIM-NLQ systems translate natural language into structured queries (e.g., BIMQL, SPARQL), enabling evaluation through query execution correctness: the generated query either matches the expected query or it does not (Yin et al., 2023; Guo et al., 2020). In this paradigm, evaluation remains tractable: correct queries produce correct results.

Recent BIM-QA systems go beyond query translation, responding in natural language directly and introducing evaluation challenges absent in query translation approaches. Prior work presented an LLM-based agentic workflow achieving 79.8% accuracy on IFC-Bench, but acknowledged that binary accuracy fails to capture nuanced answer quality: a system might provide correct values without disclosing calculation methods, leaving engineers unable to verify results (Hellin et al., 2025). This transition from query generation to answer generation requires different evaluation approaches, as binary accuracy cannot capture the nuanced quality dimensions that practitioners need. Multi-dimensional frameworks address this gap by decomposing answer quality into distinct, measurable criteria, motivating their application to BIM-QA.

Multi-dimensional QA evaluation

Beyond BIM, the inadequacy of traditional metrics for evaluating generative QA has driven the development of multi-dimensional frameworks. When evaluating long-form QA answers, domain experts consistently prioritize factuality and completeness (Xu et al., 2023). Traditional reference-based metrics such as BLEU and ROUGE, which rely on surface-level lexical overlap between generated and reference answers, fail to capture these semantic quality dimensions (Liu et al., 2023). The RAGAS framework (Es et al., 2024) operationalized evaluation through faithfulness (grounding in retrieved context), answer relevance, and context precision, while other work distinguished *correctness* from *faithfulness* (Adlakha et al., 2024). Notably, these frameworks lack explicit transparency criteria, specifically disclosure of methods and sources, identified as essential for high-stakes technical decisions where practitioners must verify results (Kell et al., 2024). Despite these advances, no validated and automated evaluation method currently exists for BIM-QA systems or AEC-related QA systems in general.

LLM-as-judge for evaluation

Applying multi-dimensional frameworks at scale faces a practical constraint: human expert evaluation is time-consuming and resource-intensive. The LLM-as-judge paradigm offers a potential solution by enabling consistent, automated evaluation at scale. Zheng et al. (2023) demonstrated that GPT-4 achieves over 80% agreement with human preferences on MT-Bench, a multi-turn chatbot evaluation benchmark, while the G-Eval framework (Liu et al., 2023) incorporated chain-of-thought reasoning for improved alignment with human judgments.

However, systematic biases may limit reliability. Studies have identified position bias, verbosity bias, and self-enhancement bias (Zheng et al., 2023), while other work demonstrated that swapping response order can flip evaluation outcomes, revealing sensitivity to presentation rather than content quality (Wang et al., 2023).

Critically, LLM-as-judge has not been systematically tested for reliability in technical domains like BIM-QA, where answers require verification against structured building data. Its applicability to such domain-specific evaluation remains an open question.

Inter-rater agreement for evaluation validation

Validating evaluation frameworks requires quantifying agreement between evaluators. Krippendorff's α is recommended for computational linguistics tasks: unlike Cohen's κ, it handles missing data and multiple raters without assuming rater-specific marginal distributions (Artstein and Poesio, 2008). Following established guidelines, $\alpha > 0.8$ indicates good reliability and $0.67 < \alpha < 0.8$ allows tentative conclusions (Krippendorff, 2004). Since α can be misleadingly low with imbalanced class distributions, percentage agreement provides a complementary view of rater concordance.

These four research threads converge on three unexplored questions: what evaluation criteria should be selected for granular multi-dimensional evaluation of BIM-QA systems, how reliably can different judges apply these criteria, and to what extent can LLM-as-judge replicate human expert judgments?

Evaluation framework

The evaluation framework comprises two components, both introduced in a prior study (Hellin et al., 2026). First, questions are categorized by data retrieval complexity; then answers are assessed using category-informed criteria. The question's category influences how the *Faithfulness* criterion is assessed, as detailed in the criteria definitions. We briefly describe both components below.

Question taxonomy

Most QA taxonomies categorize questions by syntactic structure or reasoning complexity (Rogers et al., 2021). This approach works well for text-based corpora (e.g., documents or web pages) where answers are stated explicitly, but BIM models contain heterogeneous information: some values are stored as explicit properties, others must be computed through aggregation or geometric operations, and some questions address data not present in the model at all. Prior work in BIM compliance checking proposed categorizing rules by data retrieval complexity (Solihin

and Eastman, 2015). Building on this, Hellin et al. (2026) adapted this classification to BIM-QA, defining four categories based on data retrieval complexity:

Category 1 (Direct Retrieval): Queries answerable through direct access to explicitly stored BIM properties without computation. *Example:* ''What is the fire rating of door D-101?'' The value is stored as an explicit property.

Category 2 (Aggregation): Queries requiring simple, deterministic operations (counting, summing, averaging) on accessible data. *Example:* ''How many windows are on the second floor?'' This requires filtering and counting, not stored directly.

Category 3 (Geometric/Spatial): Queries requiring complex computation beyond simple aggregation, where multiple valid approaches (and potentially multiple valid answers) may exist. *Example:* ''What is the gross floor area?'' This requires complex geometric computation, and different standards (ISO 9836, BOMA) may yield different valid results.

Category 4 (Incomplete Information): Queries addressing information unavailable in the BIM model, potentially answerable through explicit assumptions. *Example:* ''What is the energy consumption of the building?'' This cannot be answered deterministically and requires assumptions about occupancy, climate, and systems not modeled to provide a reasonable estimation.

Note on Category 4: This category diverges from (Solihin and Eastman, 2015)'s original Category 4 (complex simulations requiring external software). In BIM compliance checking, a BIM model containing all necessary information is considered a prerequisite, without which the process is meaningless. In contrast, for BIM-QA, users do not know in advance whether the information is available in the BIM model. Even if not definitive, an estimation or partial answer with properly disclosed assumptions may still be valuable.

Since the category depends on information accessibility, the same question might be categorized differently depending on the model being queried. For example, a floor area question could be Category 1 if the value is stored as a space property, or Category 3 if it must be computed geometrically.

This taxonomy informs how evaluation criteria, particularly *Faithfulness*, are applied, as described below.

Evaluation criteria

Answer quality is evaluated using binary judgments (Yes or No) across five criteria defined in Hellin et al. (2026). Binary judgments facilitate clear decision boundaries and reliable inter-rater agreement measurement. The N/A option applies when a criterion cannot meaningfully be assessed, either because the system abstained or because the question type does not require that criterion:

Criterion 0 (Abstention): Whether the system provided an answer or declined. If abstained, Criteria 1–4 are marked N/A since no content exists to evaluate. This criterion captures the coverage-reliability tradeoff: systems that appropriately abstain on uncertain queries avoid errors at the cost of reduced coverage, while systems that never abstain may overconfidently provide unreliable answers.

Criterion 1 (Faithfulness): Whether all claims are grounded in acceptable sources for the question's category. Acceptable grounding depends on question type:

- **Category 1:** BIM model data only: values must be directly retrieved from explicit properties
- **Category 2:** BIM model data plus deterministic computations (counting, summing, averaging)
- **Category 3:** BIM data plus complex geometric computations derived from model
- **Category 4:** BIM data plus computations plus *explicitly stated* assumptions

This category-dependent approach maintains alignment with RAG evaluation literature for Categories 1–2, while extending acceptable grounding to accommodate geometric derivations (Category 3) and incomplete information handling (Category 4). N/A only if system abstained.

Criterion 2 (Completeness): Whether the answer includes all relevant facts. N/A if the question expects only a single value with no ancillary information (e.g., ''What is the building height?''). This criterion is equivalent to recall for questions asking about a list of specific facts or items, but has a broader application.

Criterion 3 (Transparency): Whether the answer explicitly identifies sources or methods. Users must be able to verify information sources for the system to be useful in practice (Kell et al., 2024). *Example:* For ''What is the gross floor area?'', a transparent answer would state ''The gross floor area is 2,450 m², calculated by summing the areas of all IfcSpace entities with PredefinedType='FLOOR'.'' An opaque answer stating only ''2,450 m²'' would fail this criterion. This criterion is also essential for faithfulness assessment in Categories 3 and 4, since the system must explicitly reveal what assumptions were made. Transparency further enables post-hoc verification: users can trace reported values back to their source in the BIM model and confirm intermediate reasoning, complementing answer-level assessment. N/A only if system abstained.

Criterion 4 (Relevance): Whether the answer directly addresses the question asked. *Example of irrelevance:* For ''What is the fire rating of door D-101?'', an answer describing the door's dimensions, material, and manufacturer without mentioning fire rating would fail this criterion: the information provided, while accurate, does not address what was asked. N/A only if system abstained.

Validation methodology

We conducted inter-rater reliability analysis with LLM judges alongside human experts, an approach that enables simultaneous assessment of framework reliability and automated evaluation feasibility.

Dataset and evaluated system

We used 74 question-answer pairs from IFC-Bench (Hellin et al., 2025). Questions were distributed across four categories: Category 1 (13), Category 2 (39), Category 3 (5), and Category 4 (17).

The evaluated BIM-QA system is an LLM-based agentic workflow that processes natural language queries against IFC-encoded building models (Hellin et al., 2025). The system (1) receives a natural language question and the path to a BIM model, (2) dynamically generates and executes Python code to query the BIM model, (3) iteratively refines queries based on intermediate results, and (4) synthesizes a final response in natural language.

Evaluators

Human Evaluators: Two BIM domain experts, both with academic backgrounds in the AEC domain and over five years of practical experience with BIM workflows, independently evaluated all 74 responses. Both followed documented evaluation guidelines.

LLM Judges: To assess whether LLM-as-judge findings generalize across models, we employed two LLM judges from different providers:

- GLM-4.7, an open-weight ~400B-parameter model by Zhipu AI and successor to GLM-4.5 (GLM-4.5 Team et al., 2025), accessed via the Zhipu AI platform (z.ai)
- Gemini 3 Pro[1], accessed via Google AI Studio API

Both LLMs received identical structured prompts containing: (1) the question, question category, ground truth answer, and system-generated answer; (2) criterion definitions with category-dependent *Faithfulness* rules specifying acceptable grounding sources; (3) output format requiring Yes, No or N/A ratings with mandatory justification. The justification field is intended as a human-inspectable audit trail of the judge's decision rather than as a chain-of-thought elicitation mechanism. The *Faithfulness* criterion is operationalized with category-specific numerical accuracy rules: for Categories 1–2 (deterministic retrieval and aggregation), numerical values must match the ground truth within 2% for continuous measurements and exactly for discrete counts; for Categories 3–4, where multiple valid approaches may exist, the focus shifts to whether the methodology is explicitly disclosed and values are plausible. Full evaluation prompts, scripts, and rating data are available in the project repository (see below).

Evaluation Inputs: Both human and LLM judges received identical inputs for each question: the question text, question category, ground truth answer, and system-generated answer, along with criterion definitions and evaluation guidelines. This design isolates evaluation consistency from verification capability: all evaluators assessed system answers against the same reference standard. Ground truth answers were validated against the BIM models prior to the study to ensure dataset reliability. Evaluation prompts, scripts, and rating data are available at `https://github.com/sylvainHellin/ec3-2026-hellin-fuchs-nousias-borrmann`.

Human Evaluation Protocol: Both human evaluators worked independently and were blind to each other's ratings throughout the study. Evaluators reviewed criterion definitions and evaluation guidelines before beginning but did not complete a formal calibration session. This absence was intentional: in practice, different research groups cannot calibrate with each other before evaluating BIM-QA systems, so the framework must be self-contained enough to produce consistent results without joint training. Evaluators had access to BIM models for reference if needed, though evaluation primarily involved comparing system answers against provided ground truth. Human evaluators averaged approximately three minutes per question-answer pair.

Statistical methods

We measure inter-rater agreement using Krippendorff's α, which handles the N/A responses that arise when criteria do not apply to certain question types. We report both α and percentage agreement, as high percentage agreement can occur with low α when class distributions are highly imbalanced (ceiling effects).

Correlation between criteria is measured using Spearman's ρ to identify redundancy and independence patterns.

Consensus evaluation

After completing the blind evaluation, both human evaluators jointly reviewed all 20 questions where their ratings disagreed on at least one criterion. For each disagreement, they discussed their reasoning against the evaluation guidelines and independently decided whether to revise their original rating. This process resolved 18 of 20 disagreements. The two remaining disagreements, Q316 (whether omitting a window name makes an answer incomplete) and Q472 (how to classify a response where the system provided partial data but stated it could not answer, a ''soft abstention'' case), reflect genuine definitional boundaries not fully specified by the current guidelines. The consolidated expert opinion serves as a calibrated human baseline for assessing LLM surrogate suitability, while all original blind evaluation metrics remain unchanged.

Cross-system discriminative analysis

In addition to the answers from the agentic QA system described above, which were evaluated by both human and LLM judges, we generated two additional answer sets using different QA systems. These additional answers were evaluated by GLM-4.7 only. Comparing criterion scores across these three BIM-QA systems reveals which criteria best differentiate systems of varying capability.

- **Agentic QA**: The LLM-based agentic workflow from prior work (Hellin et al., 2025) using GLM-4.7, which dynamically explores BIM models via Python

[1] `https://deepmind.google/technologies/gemini/pro/`

code execution.

- **Static Baseline**: Same LLM with static context: a fixed summary of model elements provided without dynamic exploration capability.
- **Lightweight Agentic**: Same agentic architecture using Gemini 2.5 Flash Lite, a smaller model.

For each criterion, we computed the cross-system range (maximum minus minimum ''Yes'' rate) as a measure of discriminative power.

Results

Cross-system discriminative power

Table 1 presents criterion scores across three systems. *Transparency* (57pp range), *Completeness* (56pp), and *Faithfulness* (50pp) best discriminate between systems; *Abstention* (41pp) and *Relevance* (38pp) show weakest discrimination.

Table 1: Criterion scores (% Yes) across BIM-QA systems

Criterion	Agentic	Static	Light	Range
Abstention†	15%	52%	56%	41pp
Faithfulness	72%	28%	22%	50pp
Completeness	73%	17%	24%	56pp
Transparency	75%	31%	18%	57pp
Relevance	76%	42%	38%	38pp

†*For Abstention, lower is better (fewer abstentions = more answers attempted).*

Figure 1 shows the full rating distributions across systems, revealing that the agentic system's advantage holds across all criteria.

Correlation analysis

Correlation analysis reveals that *Completeness* and *Relevance* are highly correlated for both LLM judges (GLM: ρ = 0.81, Gemini: ρ = 0.76) and moderately correlated for human evaluators (H1: ρ = 0.40, H2: ρ = 0.65). This suggests potential criterion redundancy: a complete answer is almost always relevant. *Faithfulness* shows the lowest median correlations with other criteria (ρ = 0.26–0.65 across raters), though with considerable variability between evaluators, suggesting it captures a largely distinct quality dimension.

Both LLM judges exhibit stronger inter-criteria correlations than human evaluators, with positive ratings on one criterion predicting positive ratings on others. Full inter-criteria correlation matrices for each evaluator are available in the project repository.

Inter-rater agreement

Krippendorff's α ranges from 0.51 to 0.76 across criteria for the four-rater analysis (both human experts and both LLM judges), indicating moderate to substantial agreement (Figure 2, top). LLM-LLM agreement (α = 0.70–1.00) consistently exceeds human-human agreement (α = 0.32–0.57), with perfect agreement on *Abstention* and *Relevance*.

Since α can be misleadingly low when class distributions are highly imbalanced, we complement it with raw percentage agreement (Figure 2, bottom). Percentage agreement across all four evaluators exceeds 77% for all criteria, with *Abstention* (94.6%) and *Transparency* (93.2%) achieving highest concordance. LLM-LLM agreement reaches 100% for *Abstention* and *Relevance*. The combination of high percentage agreement with moderate α values reflects ceiling effects: when most responses receive positive ratings, expected chance agreement is high, limiting α's discriminative power.

Surrogate validation

The post-hoc consensus process (see Consensus Evaluation) yielded near-perfect human agreement (α = 0.88–1.00), confirming that the original blind disagreements were interpretation-driven rather than reflecting fundamental framework deficiencies. With this calibrated baseline established, we assess LLM surrogate suitability by comparing consensus-LLM agreement against original human-human agreement.

Both LLM judges align with expert consensus at α = 0.48–0.85, though their per-criterion profiles differ: GLM shows stronger *Faithfulness* alignment (H_{cons}-LLM_1: *Abstention* 0.85, *Faithfulness* 0.69, *Completeness* 0.48, *Transparency* 0.63, *Relevance* 0.78), while Gemini aligns better on *Transparency* (H_{cons}-LLM_2: *Abstention* 0.85, *Faithfulness* 0.48, *Completeness* 0.56, *Transparency* 0.73, *Relevance* 0.78). Both exceed the original uncalibrated human-human agreement (α = 0.32–0.57) for every criterion, as visible in Figure 2 (rightmost columns, top; rightmost bars, bottom).

This pattern supports the surrogate argument: LLMs not only agree with each other, but align with calibrated expert consensus at levels exceeding uncalibrated human-human agreement. For practical benchmarking, where individual calibration across research groups is infeasible, LLM judges offer a more reliable approximation of expert evaluation.

Discussion

Our results reveal a notable pattern: LLM judges achieve higher reliability than human experts, and generally align with each individual human at levels comparable to or exceeding human-human agreement.

The four-rater agreement (α = 0.51–0.76) falls within the ''tentative conclusions'' range, below the $\alpha > 0.8$ threshold for ''good reliability'' (Krippendorff, 2004). Our results therefore provide preliminary evidence rather than full validation. A full validation would require larger samples, particularly for Category 3 (n=5), and consistently higher agreement. The findings nonetheless establish a foundation: the framework produces measurable inter-rater agreement, and category-dependent patterns offer actionable guidance.

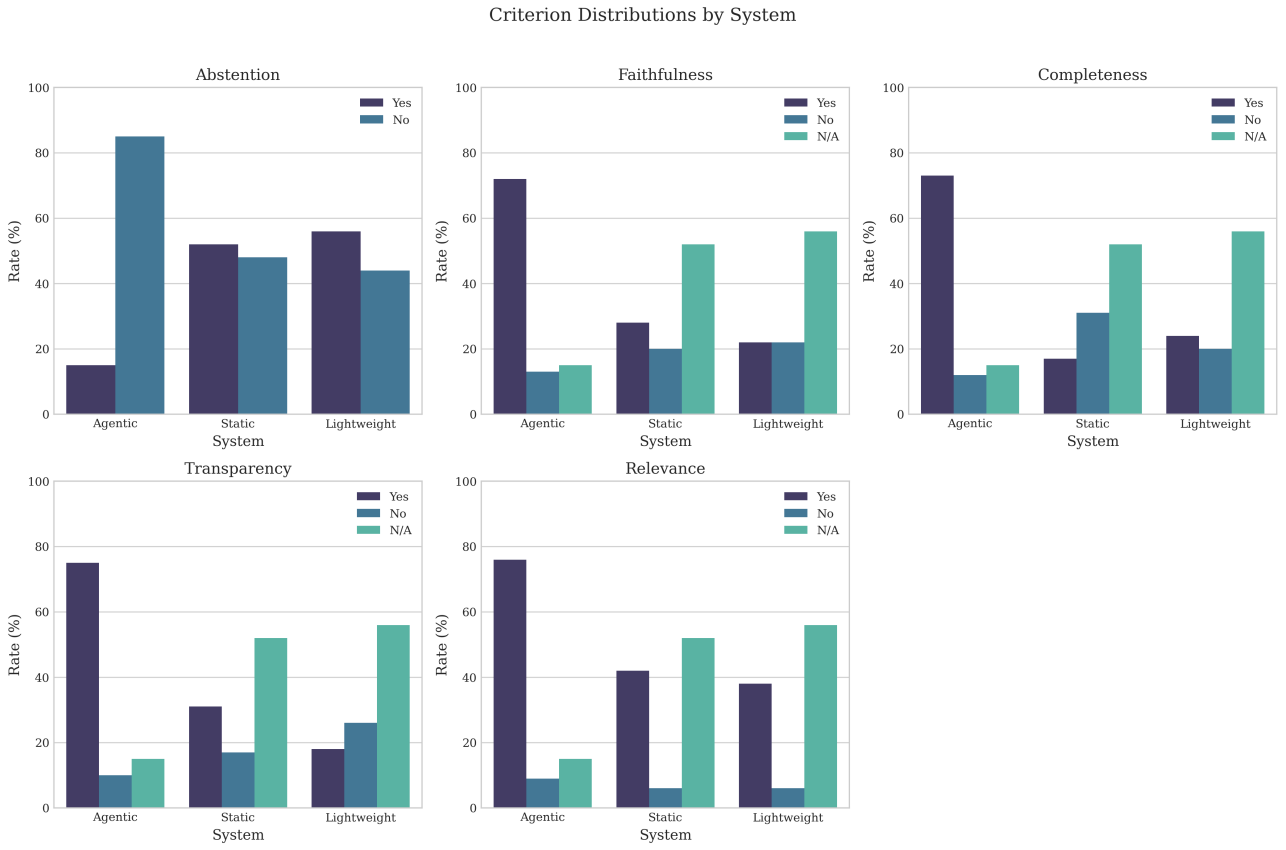


Figure 1: Criterion rating distributions across three BIM-QA systems. The agentic system achieves highest positive rates across all criteria. Higher abstention rates in Static and Lightweight systems produce corresponding N/A rates for other criteria.

Understanding evaluator differences

Why do LLM judges achieve such high reliability (α = 0.70–1.00) while human experts show lower agreement (α = 0.32–0.57)? We hypothesize three contributing factors: (1) *Criterion application consistency*: LLMs apply criteria exactly as specified, while human evaluators may interpret criteria through tacit domain expectations not captured in written guidelines; (2) *Engagement and attention*: evaluating 74 pairs against detailed criteria is repetitive work where human attention may waver, while LLMs process each instance uniformly; (3) *Criterion interpretation*: for *Completeness*, evaluators differed on whether supplementary context was required; for *Faithfulness*, they applied different verification strategies. The absence of a formal calibration session likely amplified these effects. While the criteria themselves, or the briefing provided to evaluators, may carry some inherent ambiguity, this is unlikely to be the whole story: as the post-hoc consensus analysis confirms, brief interpretive alignment is sufficient for the same raters to reach near-perfect agreement on the same criteria.

For example, Human 1's *Faithfulness* ratings exhibited higher correlation with LLM judges (α = 0.73 for H1-GLM vs. α = 0.42 for H1-H2), suggesting systematic differences in evaluation approach. Consider Question 153: ''What is the total length of exterior walls?'' The system answered ''877.47 meters,'' citing its methodology (summing the Length property for walls where IsExternal=True); the ground truth was 267.45 meters. Human 1 rated *Faithfulness* as No, recognizing the numerical discrepancy. Human 2 rated Yes, focusing on whether the methodology was grounded in BIM data. Both LLM judges also rated Yes. This illustrates a fundamental interpretation difference: does *Faithfulness* assess factual correctness, or grounded methodology? The higher LLM reliability on *Faithfulness* may reflect more uniform application of the category-specific numerical accuracy rules (see Validation Methodology).

Practical implications for BIM-QA evaluation

Our results provide preliminary evidence for LLM-as-judge suitability in domain-specific technical evaluation, extending prior findings from general-domain tasks (Zheng et al., 2023). The high LLM-LLM agreement (α = 1.00 for *Abstention* and *Relevance*, 0.89 for *Transparency*, 0.77 for *Completeness*) suggests that automated evaluation offers strong reproducibility for most criteria, with *Faithfulness* (α = 0.70), the most semantically nuanced criterion, showing the lowest but still acceptable reliability.

Two complementary findings position LLM-as-judge as a practical solution for BIM-QA evaluation. First, LLMs achieve high inter-rater reliability across model families. Second, LLMs align with calibrated expert consensus better than uncalibrated humans agreed with each other. To-

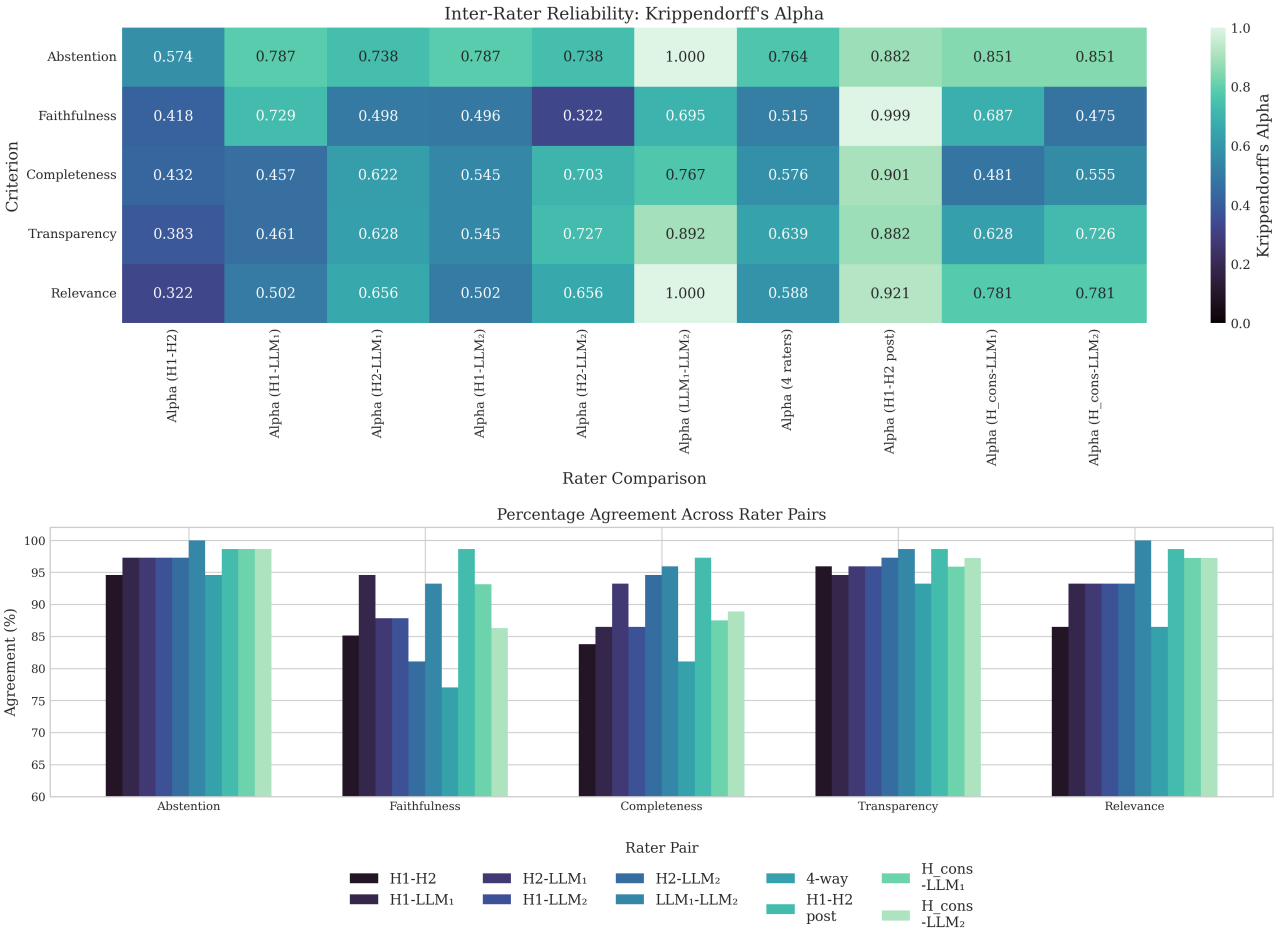


Figure 2: Inter-rater agreement analysis. ***Top:*** *Krippendorff's α heatmap showing agreement between rater pairs (H1, H2: human experts; LLM_1: GLM-4.7; LLM_2: Gemini 3 Pro) across criteria. Higher values (darker) indicate stronger agreement. Human-LLM agreement generally equals or exceeds human-human agreement. Rightmost columns show post-hoc consensus results (see Surrogate Validation).* ***Bottom:*** *Percentage agreement across rater pairs by criterion. LLM-LLM agreement consistently exceeds or matches human-human agreement, with perfect agreement on Abstention and Relevance. Rightmost bars show post-hoc consensus results.*

gether, these findings address the fragmentation problem: different research groups can compare BIM-QA systems using automated evaluation that not only produces reproducible results, but approximates the judgment of calibrated domain experts.

The correlation analysis also suggests framework simplification: given *Relevance*'s high correlation with *Completeness* (ρ = 0.76–0.81), ceiling effects (100% LLM-LLM agreement), and weak discriminative power (38pp range), we recommend consolidating *Relevance* with *Completeness* in future iterations, reducing the framework from five criteria to four.

Limitations and framework refinements

Several limitations constrain interpretation. First, both LLM judges are frontier models; findings may not generalize to smaller models. Second, all questions target IFC-based building models, limiting generalization to other BIM formats or structured-data domains. We hypothesize that LLM judges may benefit from IFC's well-documented presence in pretraining data; generalization to proprietary schemas remains to be tested. Third, the absence of calibration was intentional (to test framework self-sufficiency); the post-hoc consensus analysis revealed two definitional gaps (soft abstention handling, completeness boundaries) that enhanced guidelines should address. Fourth, with only two human evaluators, we cannot estimate population-level variance in human agreement. A replication study with balanced category sampling, the four retained criteria following *Relevance* consolidation, and a larger calibrated rater pool is needed to move from preliminary evidence to full validation. Fifth, while three systems were evaluated for discriminative analysis, inter-rater analysis used only one system's outputs. The agreement claim concerns the evaluation protocol rather than any specific system, and should hold across BIM-QA systems given consistent criteria operationalization and equal information access. Sixth, both human evaluators are co-authors of the study, which may introduce interpretation bias despite independent evaluation protocols and blinding to each other's ratings. Finally, all judges evaluated the system's reported answer against the ground truth rather than independently re-deriving it from the BIM model; intermediate reasoning verification is a property of the

system under test (supported by *Transparency* at the user level) and lies outside the protocol's scope. The study therefore assessed evaluation *reliability* (agreement across raters given the same ground truth) rather than evaluation *validity* (correctness of judgments for novel questions without ground truth); the latter remains an important direction for future work.

Conclusions

This paper introduced an automated LLM-based evaluation pipeline for BIM-QA systems and validated the underlying multi-dimensional evaluation framework through inter-rater agreement analysis with two human experts and two LLM judges. The validation reveals two complementary findings. First, LLM judges achieve high inter-rater reliability (α = 0.70–1.00), exceeding uncalibrated human agreement (α = 0.32–0.57) across all criteria. Second, a post-hoc consensus analysis demonstrates that LLMs align with calibrated expert consensus (α = 0.48–0.85) at levels exceeding the original uncalibrated human-human agreement, confirming that LLMs approximate expert judgment, not merely agree with each other. Together, these findings position LLM-as-judge as a suitable surrogate for scalable, reproducible BIM-QA benchmarking, addressing the field's fragmentation problem.

Cross-system analysis confirms discriminative validity, while correlation and ceiling effects in *Relevance* suggest consolidation with *Completeness*. Two disagreements that survived consensus discussion highlight definitional boundaries (soft abstention handling and completeness thresholds) that future guideline refinements should address.

Acknowledgments

The authors gratefully acknowledge the support of the Leonhard Obermeyer Center and the TUM Georg Nemetschek Institute.

References

Adlakha, V., BehnamGhader, P., Lu, X. H., Meade, N., and Reddy, S. (2024). Evaluating Correctness and Faithfulness of Instruction-Following Models for Question Answering. Transactions of the Association for Computational Linguistics, 12:681–699.

Artstein, R. and Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. Computational Linguistics, 34(4):555–596.

Es, S., James, J., Espinosa Anke, L., and Schockaert, S. (2024). RAGAs: Automated Evaluation of Retrieval Augmented Generation. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics .

GLM-4.5 Team, Zeng, A., Lv, X., Zheng, Q., Hou, Z., Chen, B. et al. (2025). GLM-4.5: Agentic, Reasoning, and Coding (ARC) Foundation Models.

Guo, D., Onstein, E., and Rosa, A. D. L. (2020). An Approach of Automatic SPARQL Generation for BIM Data Extraction. Applied Sciences, 10(24):8794.

Hellin, S., Nousias, S., and Borrmann, A. (2025). Natural Language Information Retrieval from BIM Models: An LLM-Based Agentic Workflow Approach. In Proceedings of the 2025 European Conference on Computing in Construction.

Hellin, S., Nousias, S., and Borrmann, A. (2026). A systematic evaluation framework for ai-driven bim question answering systems. In Proc. of the TUM GNI International Symposium on Artificial Intelligence for the Built World.

Kell, G., Roberts, A., Umansky, S., Qian, L., Ferrari, D., Soboczenski, F. et al. (2024). Question answering systems for health professionals at the point of care—a systematic review. Journal of the American Medical Informatics Association, 31(4):1009–1024.

Krippendorff, K. (2004). Reliability in Content Analysis.: Some Common Misconceptions and Recommendations. Human Communication Research, 30(3):411–433.

Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R., and Zhu, C. (2023). G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment.

Rogers, A., Gardner, M., and Augenstein, I. (2021). QA Dataset Explosion: A Taxonomy of NLP Resources for Question Answering and Reading Comprehension.

Solihin, W. and Eastman, C. (2015). Classification of rules for automated BIM rule checking development. Automation in Construction, 53:69–82.

Wang, P., Li, L., Chen, L., Cai, Z., Zhu, D., Lin, B. et al. (2023). Large Language Models are not Fair Evaluators.

Xu, F., Song, Y., Iyyer, M., and Choi, E. (2023). A Critical Evaluation of Evaluations for Long-form Question Answering. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics.

Yin, M., Tang, L., Webster, C., Li, J., Li, H., Wu, Z. et al. (2023). Two-stage Text-to-BIMQL semantic parsing for building information model extraction using graph neural networks. Automation in Construction, 152:104902.

Zheng, J. and Fischer, M. (2023). Dynamic prompt-based virtual assistant framework for BIM information search. Automation in Construction, 155:105067.

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y. et al. (2023). Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

COMPARISON OF SEGMENTATION MODELS FOR POTHOLE DETECTION USING UAV IMAGERY

Varshney Muskan[1], Singh Aryan Raj[1], Garg Sahil[2]
[1,2]Department of Civil Engineering, Indian Institute of Technology Delhi (IIT Delhi), New Delhi, India

Abstract

Potholes are a significant cause of vehicle damage and road safety accidents. Existing inspection practices rely on manual surveys using inspection vehicles or vehicle-mounted cameras, which are time-consuming, non-scalable, and often provide incomplete visibility. This paper presents a comparative study of instance-segmentation-based deep learning models for pothole detection using UAV imagery, enabling wider and more consistent coverage. A custom dataset was developed and augmented using geometric transformations. YOLOv8 and transformer-based RF-DETR models were evaluated using precision, recall, and mAP@50. The best model achieved mAP@50 of 66.3%, highlighting challenges in detecting shallow potholes under limited training data.

Introduction

Potholes are a common form of pavement distress that develops due to repeated traffic loading, environmental exposure, and material fatigue, and they pose long-term risks to road safety, vehicle integrity, and transportation efficiency (Ibragimov et al. 2022). Even shallow potholes can progressively worsen over time, contributing to increased vehicle operating costs and a higher likelihood of traffic accidents when maintenance interventions are delayed (Ali et al. 2023).

Effective monitoring of road surface conditions is therefore essential for sustainable transportation infrastructure management. Conventional inspection practices typically rely on manual surveys conducted using inspection vehicles or visual assessments performed by trained personnel (Anand et al. 2025). Although widely used, such approaches are labor-intensive, subjective, and difficult to scale across extensive road networks. In addition, vehicle-mounted inspection systems often provide limited visibility of the road surface due to restricted viewing angles, occlusions, and perspective distortion, which can hinder reliable detection of subtle or spatially distributed defects (Yu et al. 2024).

Advances in computer vision and deep learning have enabled automated approaches for pavement distress detection using image and video data, offering more objective and repeatable alternatives to manual inspection (Safyari et al. 2024). Convolutional neural networks have demonstrated strong capability in learning discriminative visual features for pothole detection, significantly outperforming traditional image processing methods under diverse imaging conditions (Jakubec et al. 2023). Object detection frameworks such as YOLO have been widely explored due to their computational efficiency and suitability for large-scale deployment, while instance segmentation techniques provide pixel-level delineation of defect boundaries that supports more detailed geometric characterization (Li et al. 2024).

Despite methodological progress, most pothole detection studies continue to rely on ground-based or vehicle-mounted imaging platforms, where consistent surface coverage is difficult to achieve (Chen et al. 2024). Unmanned aerial vehicles (UAVs) offer a complementary sensing perspective by enabling wide-area, top-down observation of road surfaces with reduced occlusion and more uniform imaging geometry. Recent work has demonstrated the effectiveness of UAV imagery for pavement crack detection and surface distress assessment, highlighting its growing relevance for scalable road condition monitoring applications (Ellenberg et al. 2015).

In parallel with advances in sensing platforms, transformer-based vision models have emerged as an alternative to conventional convolutional architectures. By leveraging self-attention mechanisms, these models can capture long-range contextual relationships and improve generalization in visually complex scenes (Carion et al. 2020). Recent developments in lightweight and deformable transformer variants have further improved convergence behavior and performance under limited training data conditions (Zhu et al. 2020). However, the application of transformer-based instance segmentation models to pothole detection using UAV imagery remains insufficiently explored in existing literature. Recent studies have further shown that improved YOLOv8-based architectures are particularly effective for small object detection in UAV aerial imagery, making them suitable for detecting subtle pavement defects from aerial perspectives (Ni et al. 2024).

Addressing this gap requires systematic evaluation of segmentation-based approaches that can accurately localize potholes while accounting for the challenges associated with aerial imagery, subtle surface defects,

and limited annotated data. Such analysis is necessary to better understand model behavior and to support the development of reliable, scalable road surface monitoring systems.

Dataset and experimental setup

Data collection and annotation

A total of 196 road surface images were collected using a quad-copter unmanned aerial vehicle (UAV) equipped with an RGB camera. The UAV was operated to capture views of pavement surfaces, enabling wide-area coverage and consistent imaging geometry. The data acquisition strategy was designed to capture realistic road surface conditions under varying illumination levels, surface textures, and environmental influences. As a result, the dataset includes visual challenges such as shadows, debris, worn asphalt regions, uneven lighting, and surface irregularities that commonly affect road condition assessment.

All images were manually annotated using polygon-level masks to precisely delineate pothole regions. The annotation process was performed using the Roboflow Annotate tool, enabling accurate boundary-level labeling required for instance segmentation tasks. Shallow potholes with weak visual contrast posed particular challenges during annotation due to gradual depth transitions, shadows, and surface stains, leading to unavoidable annotation uncertainty. Despite these challenges, polygon-based annotations were preferred over bounding boxes, as they provide a more precise spatial representation of pothole boundaries and better support pixel-level segmentation, geometric analysis, and potential severity estimation in future extensions of this work.

Data augmentation

To address dataset scarcity and improve model generalization, the original dataset was expanded from 196 to 475 images using targeted geometric data augmentation techniques. The augmentation strategy focused on preserving the structural characteristics of potholes while introducing controlled variability in spatial orientation. The applied augmentations included horizontal flipping, small-angle rotations within a range of ±10°, and shear transformations limited to ±10°.

To prevent unrealistic geometric distortions and maintain annotation fidelity, no more than two augmentation operations were applied to any single image during training. This constraint ensured that the augmented samples remained representative of real-world aerial road scenes while avoiding excessive deformation of pothole boundaries.

The selected augmentation techniques were intended to enhance robustness to variations in UAV viewpoint, flight orientation, and minor perspective changes commonly encountered during aerial image acquisition. By increasing the effective diversity of the training data, the augmentation process contributed to improved detection and segmentation performance under diverse visual and geometric conditions.

Data augmentation was performed exclusively on the training set after dataset splitting, ensuring that the validation and test sets remained unaltered. This approach prevents data leakage and supports an unbiased evaluation of model performance

Models and training strategy

This study evaluates seven instance segmentation model configurations, comprising convolutional neural network–based and transformer-based architectures, to examine the influence of data augmentation, weight initialization, and model design on pothole detection from aerial imagery. All experiments were conducted under a consistent training environment to ensure comparability across configurations.

Four convolutional neural network–based configurations were developed using the YOLOv8 instance segmentation framework. The first configuration corresponds to a baseline YOLOv8 model trained without any data augmentation. A second YOLOv8 configuration incorporated additional training samples while maintaining the same augmentation-free setting. The third YOLOv8 configuration applied horizontal flipping as a single geometric augmentation to increase viewpoint invariance. The fourth YOLOv8 configuration employed a combination of horizontal flipping with mild shear and rotation transformations, enabling greater robustness to camera orientation and road alignment variations.

Three transformer-based configurations were implemented using the RF-DETR instance segmentation framework. All RF-DETR models were initialized using large-scale pretrained weights. The first RF-DETR configuration applied horizontal flipping combined with small-angle rotation. The second RF-DETR configuration used horizontal flipping together with shear and rotation transformations. The third RF-DETR configuration employed horizontal flipping with normal rotation only. These augmentation strategies were selected to balance increased data diversity with preservation of realistic road geometry.

Across all configurations, no more than two geometric augmentations were applied to any individual image during training in order to avoid unrealistic distortions of pothole boundaries. This constraint ensured that augmented samples remained representative of real-world aerial road scenes.

All models were trained within the Roboflow cloud-based training environment using available optimization settings and GPU acceleration. Maintaining a unified training platform and consistent preprocessing pipeline ensured that observed performance differences could be attributed to model architecture and augmentation strategy rather than implementation variability.

This structured experimental setup enables a controlled comparison between convolutional and transformer-based instance segmentation models, as well as an

assessment of the role of specific geometric augmentations in aerial pothole detection. Quantitative performance outcomes are presented separately in the Results section.

Training setup

The final dataset comprised 475 images, including both original and augmented samples. The original dataset was partitioned into training, validation, and test subsets using an approximate 70:20:10 split to ensure reliable performance evaluation and to mitigate overfitting. The training subset was used for model learning, while the validation and test sets were reserved for hyperparameter tuning and final performance assessment, respectively.

Two input resolutions, 640 × 640 and 432 × 432, were systematically explored during the training process. Both YOLOv8 and RF-DETR models were trained using each of these resolutions. It was observed that YOLOv8 converged most effectively at an input resolution of 640 × 640, whereas RF-DETR achieved the most stable and optimal convergence at 432 × 432. Consequently, these resolutions were adopted for the final experiments (640×640 for YOLOv8, 432×432 for RF-DETR). Model weights were initialized using pretrained checkpoints instead of random weights, enabling effective transfer learning and faster convergence, particularly under limited data availability.

Table 1: Summary of training configurations and hyperparameters

Model Name	Input Resolution	Augmentation Strategy	Weight Initialization
Road Hazards 2	640 × 640	None	Public Checkpoints (MS COCO)
Road Hazards 3	640 × 640	Horizontal Flip	Road Hazards 2
Road Hazards 6	640 × 640	Horizontal Flip	Public Checkpoints (MS COCO)
Road Hazards 7	640 × 640	Horizontal Flip + Shear Rotation ±10° horizontal, vertical	Road Hazards 6
Road Hazards 10	432 × 432	Horizontal Flip + Rotation ±10°	Objects365 Pretrained
Road Hazards 12	432 × 432	Horizontal Flip+ Shear Rotation ±10° horizontal, vertical	Objects365 Pretrained
Road Hazards 13	432 × 432	Horizontal Flip+ Rotation ±10°	Objects365 Pretrained

In addition to the configurations summarized in Table 1, RF-DETR models were trained using a batch size of 16, 100 epochs, a learning rate of 1×10^{-4}, and the AdamW optimizer. For YOLOv8, training was performed using the available hyperparameter settings provided by the Roboflow platform. These settings were not manually modified and were kept consistent across experiments to ensure fair comparison.

Evaluation metrics and loss functions

Model performance was evaluated using standard instance segmentation metrics commonly adopted in object detection and segmentation tasks. Mean Average Precision at an Intersection over Union threshold of 0.50 (mAP@50) was used as the primary evaluation metric, as it provides a balanced measure of detection accuracy under moderate localization constraints. In addition, precision and recall were reported to analyze false-positive and false-negative behavior, which is particularly important in safety-critical applications such as road condition monitoring.

During training, multiple loss components were continuously monitored to assess learning dynamics and segmentation quality. This included a Dice loss to evaluate the quality and overlap of predicted segmentation masks, classification loss to measure label prediction accuracy, and bounding box localization loss to quantify spatial regression performance. The evolution of both loss values and evaluation metrics across training epochs was analyzed to assess convergence behavior, detect potential overfitting, and guide model selection.

Results and discussion

The quantitative performance of all trained models is summarized in Table 1.

Table 2: Performance comparison of pothole detection models

Model Name	Type	mAP@50	Precision	Recall
Road Hazards 2	YOLOv8 (Baseline)	45.9%	74.1%	40.0%
Road Hazards 3	YOLOv8 (Extra data)	48.0%	56.5%	45.5%
Road Hazards 6	YOLOv8 (Augmented)	59.6%	74.8%	54.5%
Road Hazards 7	YOLOv8 (Double Aug)	56.0%	59.9%	60.5%
Road Hazards 10	RF-DETR (Transformer)	66.3%	87.5%	55.0%
Road Hazards 12	RF-DETR (Transformer)	65.9%	75.0%	59.0%
Road Hazards 13	RF-DETR (Transformer)	65.6%	78.0%	60.0%

The experimental results demonstrate the progressive improvement in pothole detection and segmentation performance achieved through model architecture

selection and data augmentation strategies. Comparative evaluation across convolutional and transformer-based models highlights both the benefits and limitations of current instance segmentation approaches when applied to aerial road surface imagery.

The initial YOLOv8 baseline models exhibited limited generalization capability. Early configurations achieved mAP@50 values in the range of approximately 46–48%, indicating difficulty in distinguishing shallow potholes from visually similar background artifacts such as shadows, stains, and surface irregularities. Precision was moderate, while recall remained comparatively low, reflecting a tendency to miss subtle, low-contrast pothole instances. These observations suggest that purely convolutional architectures struggle to capture sufficient contextual information when pothole boundaries are weak or ambiguous.

With the introduction of targeted data augmentation and incremental dataset expansion, YOLOv8 performance improved noticeably. Augmented variants achieved mAP@50 values in the range of approximately 56–60%, accompanied by an increase in recall to around 60%. This improvement highlights the importance of data diversity in enhancing robustness to variations in viewpoint, surface texture, and local appearance. Nevertheless, despite these gains, the convolutional models continued to face challenges in accurately delineating pothole boundaries under low-contrast conditions, indicating an architectural limitation rather than a data-only constraint.

The transformer-based RF-DETR models delivered the strongest overall performance across all evaluated metrics. The best-performing configuration achieved a mAP@50 of 66.3%, with precision reaching up to 87.5% and recall stabilizing in the range of 55–60%. Other RF-DETR variants produced comparable results, with mAP@50 values consistently above 65%. The transition from YOLOv8 to RF-DETR resulted in the largest single performance improvement, underscoring the effectiveness of global self-attention mechanisms and strong pretrained representations for capturing long-range contextual relationships in aerial imagery.

Figure 1 illustrates the training dynamics and loss behavior of the best-performing RF-DETR model. The mAP@50 curve stabilizes within the range of approximately 0.63–0.67 during the first 8–10 epochs, indicating rapid convergence driven by effective transfer learning. Minor oscillations observed after convergence suggest sensitivity to limited training data rather than training instability or overfitting. Dice loss remains relatively high and fluctuating, reflecting the inherent difficulty of producing consistent segmentation masks for shallow potholes with weak or gradual visual boundaries. Classification loss and bounding box localization loss remain generally stable, with occasional spikes attributable to small object sizes and annotation uncertainty.

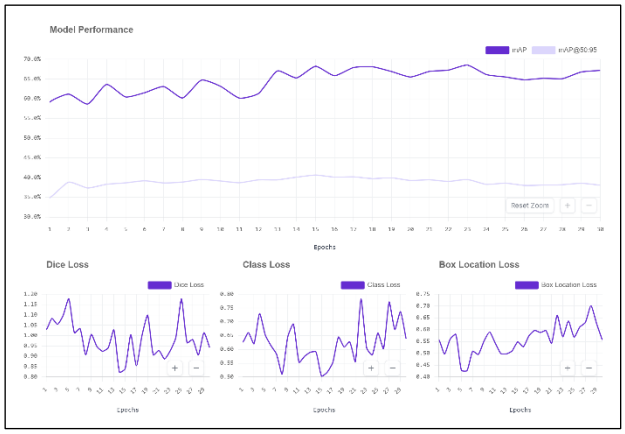


Figure 1: Performance and loss metrics of model Road Hazards 13

Qualitative evaluation further supports the quantitative findings. As illustrated in Figure 2, both YOLOv8 and RF-DETR models successfully detect potholes that are well illuminated and exhibit clear texture contrast relative to the surrounding road surface. Correct detections with high confidence scores demonstrate the capability of the models under favorable visual conditions. However, false positives are commonly triggered by dark stains, shadows, and worn surface patches that visually resemble pothole depressions. Missed detections primarily occur for extremely shallow potholes characterized by gradual depth transitions and weak contrast, particularly near road edges where surface wear blends into surrounding textures.



Figure 2: Single image showing all 3 cases: 1. Correct predictions (88%, 74%), 2. False positives (48%), and 3. False Negatives (below 88% mask)

Compared to YOLOv8, RF-DETR models exhibit a noticeable reduction in shadow-induced false positives and generate more spatially coherent segmentation masks, even in visually cluttered scenes. This behavior confirms the advantage of transformer-based architectures in leveraging broader contextual information, enabling more effective differentiation between true pothole regions and visually similar non-pothole artifacts.

Overall, the results indicate that while transformer-based instance segmentation models provide clear performance advantages over convolutional baselines, detection accuracy remains constrained by dataset scale, annotation ambiguity, and the subtle visual nature of shallow potholes. These findings emphasize that further improvements will likely require larger and more diverse

datasets, enhanced annotation strategies, and the incorporation of complementary geometric or contextual cues beyond RGB imagery alone.

Limitations

Despite the encouraging results obtained in this study, several limitations must be acknowledged. First, the dataset size remains relatively small, consisting of 196 original images, which restricts the diversity of visual scenarios available during training. Although geometric augmentation was employed to partially alleviate this constraint, the limited number of unique real-world samples may still impact generalization to unseen road conditions.

Second, all annotations were produced by a single annotator, introducing unavoidable subjectivity in boundary delineation, particularly for shallow potholes with weak visual contrast. Ambiguous transitions between damaged and undamaged road surfaces make precise annotation challenging and can lead to inconsistencies that affect pixel-level evaluation metrics.

Third, the applied data augmentation strategy was limited to basic geometric transformations. More complex augmentations that simulate real-world imaging conditions, such as motion blur, glare, weather effects, and illumination variation, were not incorporated. The absence of such augmentations may reduce robustness under highly variable environmental conditions commonly encountered in long-term field deployments.

Conclusion

This work presented the development of a custom instance-segmentation dataset for pothole detection and a systematic evaluation of deep learning models trained under realistic data constraints using aerial imagery. A total of 196 original road surface images were manually annotated and expanded through targeted geometric augmentation. Multiple model configurations were trained and evaluated within a unified cloud-based training environment to ensure consistent experimentation and reproducibility.

Baseline convolutional models based on YOLOv8 achieved mAP@50 values in the range of approximately 45–60% as dataset size and augmentation strategies were progressively improved. While these results demonstrate the feasibility of CNN-based approaches for pothole detection, they also revealed limitations in generalization, particularly when detecting shallow potholes with weak visual contrast and ambiguous boundaries. Transitioning to the transformer-based RF-DETR architecture resulted in the strongest overall performance, achieving mAP@50 values of approximately 65–66% and peak precision of up to 87.5%. These improvements confirm the advantage of strong pretraining and global attention mechanisms when operating under limited data availability and visually complex backgrounds.

The primary performance bottlenecks identified in this study were the limited scale of the annotated dataset and unavoidable annotation uncertainty associated with shallow, low-contrast pavement defects. Despite these constraints, the trained models exhibited stable convergence behavior and reliable qualitative performance, particularly in reducing shadow-induced false positives that commonly affect appearance-based detection methods. Such robustness is critical for practical deployment, where minimizing false alarms directly impacts the efficiency of inspection and maintenance workflows.

Overall, this study establishes a viable prototype for aerial pothole detection using instance segmentation and demonstrates the effectiveness of transformer-based architectures for challenging real-world road surface monitoring scenarios. Beyond detection, the generated segmentation masks provide a strong technical foundation for future extensions such as area-based pothole severity estimation, temporal damage progression analysis, and maintenance prioritization. With further dataset expansion, improved annotation strategies, and the integration of complementary geometric or sensing cues, the proposed approach has strong potential to support scalable and data-driven road infrastructure monitoring systems.

Future work

Future work will focus on improving robustness and scalability of the proposed pothole detection system. Expanding the dataset to include more diverse imaging conditions and adopting multi-annotator consensus labeling are expected to enhance segmentation reliability for shallow and ambiguous defects.

Beyond detection, integrating segmentation masks with geometric mapping and depth estimation techniques will enable physically meaningful severity metrics, supporting maintenance prioritization. Model optimization for efficient edge deployment and further improvements through transformer scaling and semi-supervised learning remain promising directions.

Acknowledgement

This project is funded by SPARC.

References.

Ali, F., Khan, Z.H., Khattak, K.S. and Gulliver, T.A., 2023. Evaluating the effect of road surface potholes using a microscopic traffic model. Applied Sciences, 13(15), p.8677.

Anand, R., Priyanka, S. and Goud, P.G., 2025. Assessment of Road Surface Conditions Using Machine Learning. Journal of Transportation Engineering and Its Applications ISSN: 3107-4499 (Online), 10(1).

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A. and Zagoruyko, S., 2020, August. End-to-end object detection with transformers. In European conference on computer vision (pp. 213–229). Cham: Springer International Publishing.

Chen, S., Laefer, D.F., Zeng, X., Truong-Hong, L. and Mangina, E., 2024. Volumetric pothole detection from UAV-based imagery. Journal of Surveying Engineering, 150(2), p.05024001.

Ellenberg, A., Branco, L., Krick, A., Bartoli, I. and Kontsos, A., 2015. Use of unmanned aerial vehicle for quantitative infrastructure evaluation. Journal of Infrastructure Systems, 21(3), p.04014054.

Ibragimov, E., Lee, H.J., Lee, J.J. and Kim, N., 2022. Automated pavement distress detection using region based convolutional neural networks. International Journal of Pavement Engineering, 23(6), pp.1981–1992.

Jakubec, M., Lieskovská, E., Bučko, B. and Zábovská, K., 2023. Comparison of CNN-based models for pothole detection in real-world adverse conditions: Overview and evaluation. Applied Sciences, 13(9), p.5810.

Li, Y., Yin, C., Lei, Y., Zhang, J. and Yan, Y., 2024. Rdd-yolo: Road damage detection algorithm based on improved you only look once version 8. Applied Sciences, 14(8), p.3360.

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T. and Omata, H., 2018. Road damage detection and classification using deep neural networks with smartphone images. Computer-Aided Civil and Infrastructure Engineering, 33(12), pp.1127–1141.

Ni, J., Zhu, S., Tang, G., Ke, C. and Wang, T., 2024. A small-object detection model based on improved YOLOv8s for UAV image scenarios. Remote Sensing, 16(13), p.2465.

Safyari, Y., Mahdianpari, M. and Shiri, H., 2024. A review of vision-based pothole detection methods using computer vision and machine learning. Sensors, 24(17), p.5652.

Yu, J., Jiang, J., Fichera, S., Paoletti, P., Layzell, L., Mehta, D. and Luo, S., 2024. Road surface defect detection—from image-based to non-image-based: a survey. IEEE Transactions on Intelligent Transportation Systems, 25(9), pp.10581–10603.

Zhu, X., Su, W., Lu, L., Li, B., Wang, X. and Dai, J., 2020. Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159.

Loshchilov, I., & Hutter, F. (2019). Decoupled weight decay regularization. International Conference on Learning Representations (ICLR)

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

SEEING IN THE DARK: LOW-LIGHT IMAGE ENHANCEMENT FOR MODULAR CONSTRUCTION

Yuanyang Qi, Jie Hong, Tingtian Li, and Xiao Li
CI^3 Lab, Department of Civil Engineering, The University of Hong Kong, HK SAR, China

Abstract

Modular construction uses off-site modules for on-site assembly; computer vision supports progress monitoring, inspection, and assembly. These tasks may occur in poorly lit spaces, where low light degrades image quality, while prior work has not addressed low-light enhancement for this setting. We propose RetinexRefiner, a two-stage framework that first performs illumination-based enhancement and then refines the result to suppress residual artefacts such as colour cast and noise. Trained end-to-end, it achieves improved visibility on night construction images, reaching 28.07 dB PSNR and 0.83 SSIM, and is applicable to construction sites and similar dim environments.

Introduction

Modular construction transforms the industry by fabricating volumetric units off-site and assembling them on-site (Ferdous et al., 2019; Zheng et al., 2020), with benefits in quality control, productivity, and safety (Tsz Wai et al., 2023). The on-site phase centres on module handling: monitoring progress (e.g. verifying module statuses) and ensuring safety (e.g. detecting proximity between modules and other objects) (Zheng et al., 2020; Ekanayake et al., 2021). Computer vision supports these activities through progress monitoring, inspection, and assembly (Zheng et al., 2020; Fang et al., 2020; Ekanayake et al., 2021; Qi et al., 2025). The accuracy of such vision tasks depends on image quality (Liu et al., 2023; Li et al., 2021). In practice, the work may take place in poorly lit indoor spaces, shaded areas, or at night (Chauhan and Seppänen, 2023; Ekanayake et al., 2021), where low light degrades images and can cause automated checks to miss defects or misestimate progress (Chauhan and Seppänen, 2023; Ekanayake et al., 2021; Li et al., 2021). Reliable imaging in such conditions is therefore important both for safety (e.g. detecting obstacles and workers) and for progress verification (e.g. documenting module placement and assembly state); improving low-light image quality can directly support these vision-based tasks. Prior work on modular construction has focused on module detection and progress monitoring under normal lighting, and most existing vision systems for construction assume well-lit scenes rather than dim or nocturnal conditions (Zheng et al., 2020; Fang et al., 2020; Ekanayake et al., 2021; Qi et al., 2025); low-light image enhancement for this setting has not been addressed. It is thus critical to explore low-light enhancement methods that can improve visibility for monitoring, inspection, and assembly in modular construction.

Low-light image enhancement aims to improve visibility and restore corruptions (noise, artefacts, colour distortion) hidden in the dark or introduced when brightening underexposed photos. Histogram equalisation and gamma correction directly amplify low visibility but barely consider illumination, often causing noise and colour distortion. Retinex-based methods (Land, 1977) decompose an image into reflectance and illumination; LIME (Guo et al., 2016) refines illumination with structure priors but assumes corruption-free inputs, leading to amplified noise in real low-light scenes. DUAL (Zhang et al., 2019a) performs dual illumination estimation for exposure correction. Deep learning methods have advanced the field (Li et al., 2021). RetinexNet (Wei et al., 2018) introduces a deep Retinex decomposition network and the LOL dataset; KinD (Zhang et al., 2019b) decouples illumination and reflectance for degradation removal. Both require multi-stage training. RUAS (Liu et al., 2021) combines Retinex unrolling with neural architecture search for a lightweight enhancement network. RetinexFormer (Cai et al., 2023) proposes a one-stage Transformer with illumination-guided attention blocks (IGAB) that outperform prior Retinex-based CNNs on multiple benchmarks. Other lines of work include zero-reference and physical-prior methods (Guo et al., 2020; Wang et al., 2024), diffusion-based and dual-branch pipelines (Feng et al., 2024), codebook retrieval and flow-based enhancement (Zhou et al., 2024; Wang et al., 2022), zero-shot neural implicit enhancement (Chobola et al., 2024), and learning-based pipelines for raw or extreme dark imaging (Chen et al., 2018; Jiang et al., 2025); recent work also explores new colour spaces for low-light enhancement (Yan et al., 2025). For construction and monitoring, we focus on Retinex-based paired learning so that we have an explicit illumination map and a clear reference (ground truth) for evaluation. Retinex-based methods follow the imaging relation that observed intensity is the product of reflectance and illumination; they offer interpretable decomposition, explicit illumination maps, and a natural fit for paired supervised learning. We build on RetinexFormer because it achieves strong performance

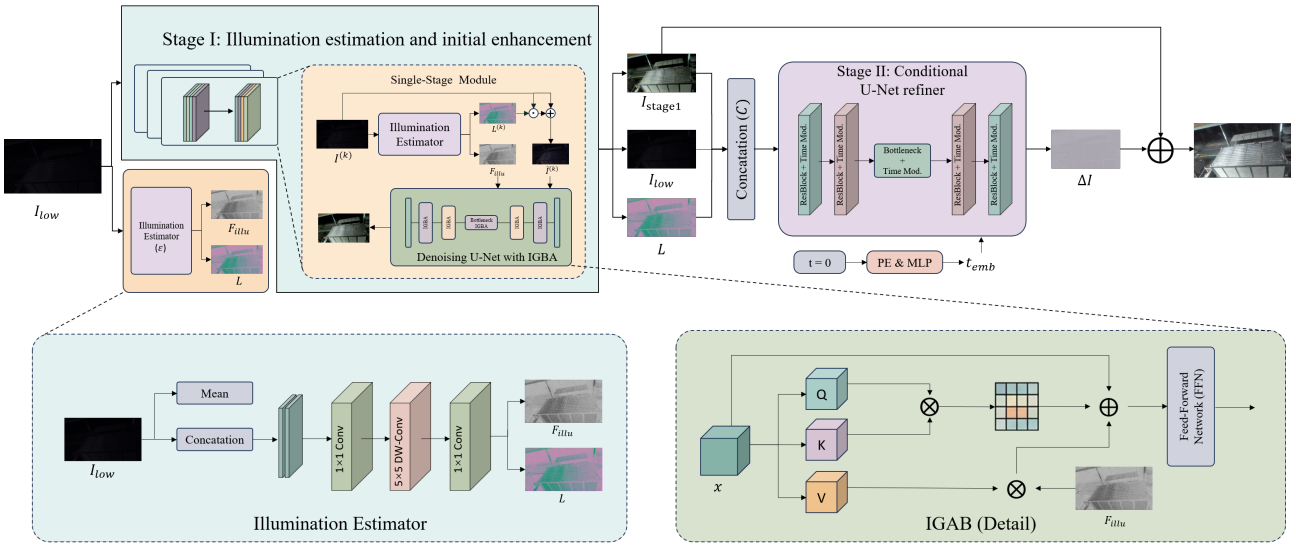


Figure 1: RetinexRefiner two-stage pipeline. Low-light input I_{low} is fed to Stage I (illumination estimation and initial enhancement), which outputs an illumination map L and an initial enhanced image I_{stage1}. Stage II (conditional U-Net refiner) takes $[I_{\text{low}}; I_{\text{stage1}}; L]$ and predicts a residual; the final output is $I_{\text{final}} = I_{\text{stage1}} + \text{Refiner}(\cdot)$. Both stages are trained end-to-end.

among Retinex-based methods and provides an explicit illumination map, which we use to condition the refiner. However, like other one-stage methods, it can leave residual artefacts in very dark regions, such as colour cast, noise, and under- or over-correction.

To address this, we propose RetinexRefiner, a two-stage pipeline that extends RetinexFormer with a refinement stage to correct such artefacts. The pipeline is trained end-to-end on paired normal-light and low-light construction images. Our main contributions can be summarised as follows:

- We address low-light image enhancement for modular construction, where on-site monitoring and assembly may occur in poorly lit or indoor conditions and prior work has not focused on this setting.
- We propose RetinexRefiner, a two-stage pipeline: stage one performs Retinex-based illumination estimation and initial enhancement; stage two uses a conditional U-Net refiner on the low-light image, stage-one output and illumination map to correct residual colour cast and noise, with both stages trained end-to-end on paired images.
- Quantitative and qualitative experiments on night construction images show improved visibility and competitive PSNR/SSIM, outperforming the stage-one baseline and comparison methods.

The paper is organised as follows. Section 2 presents the methodology (two-stage pipeline, Retinex formulation, stage I and stage II). Section 3 describes the experiments, results and ablations. Section 4 discusses limitations; Section 5 concludes and outlines future work.

Methodology

Pipeline overview

Figure 1 summarises the RetinexRefiner pipeline. Given a low-light image I_{low}, the first stage estimates illumination and yields an illumination map L and an initial enhanced image I_{stage1}. The second stage takes I_{low}, I_{stage1}, and L as input and predicts a residual. The final output is

$$I_{\text{final}} = I_{\text{stage1}} + \text{Refiner}(\cdot), \tag{1}$$

where I_{low} is the low-light input image, I_{stage1} is the stage-one enhanced image, L is the illumination map from stage one, I_{final} is the final enhanced output, and $\text{Refiner}(\cdot)$ denotes the conditional U-Net applied to the concatenation $[I_{\text{low}}; I_{\text{stage1}}; L]$. Both stages are differentiable and trained end-to-end.

Retinex-inspired formulation

We follow the Retinex assumption: an image is the element-wise product of reflectance and illumination,

$$I = R \odot L, \qquad R, L \in \mathbb{R}^{3\times H\times W}, \tag{2}$$

where I is the observed image, R is the reflectance, L is the illumination, $\odot$ denotes element-wise multiplication, and H, W are the image height and width. Stage one yields $I_{\text{stage1}} = \mathscr{F}_{\text{RF}}(I_{\text{low}})$ and $L = \mathscr{E}(I_{\text{low}})$. Stage two predicts a residual and adds it:

$$\begin{aligned} \Delta I &= \mathscr{R}([I_{\text{low}}; I_{\text{stage1}}; L], t = 0), \\ I_{\text{final}} &= I_{\text{stage1}} + \Delta I. \end{aligned} \tag{3}$$

where $[\cdot;\cdot;\cdot]$ denotes channel concatenation, $\mathscr{R}$ is the conditional U-Net refiner, ΔI is the predicted residual, t is the time step (fixed at 0 in our setup), and I_{final}, I_{stage1} are the final and stage-one enhanced images. The first stage performs illumination estimation and initial enhancement;

we extend the pipeline with the refiner for the second stage.

Stage I: Illumination estimation and initial enhancement

Stage one follows the one-stage Retinex-based framework of Cai et al. (2023), with three single-stage modules in sequence. At the k-th stage ($k = 1,2,3$), the illumination estimator forms $X_{\text{in}} = [I^{(k)}; \text{mean}_c(I^{(k)})]$ and outputs F_{illu} and $L^{(k)}$:

$$\begin{aligned} F_{\text{illu}}, L^{(k)} &= \mathscr{E}_{\text{stage}}(X_{\text{in}}) \\ &= \phi_2 \circ \phi_{\text{dw}} \circ \phi_1(X_{\text{in}}), \end{aligned} \tag{4}$$

where X_{in} is the concatenation $[I^{(k)}; \text{mean}_c(I^{(k)})]$ with mean_c the channel-wise mean, $I^{(k)}$ is the image at the k-th sub-stage ($I^{(1)} = I_{\text{low}}$), F_{illu} is the illumination feature, $L^{(k)}$ is the light-up map, $\mathscr{E}_{\text{stage}}$ is the illumination estimator, and ϕ_1, ϕ_{dw}, ϕ_2 are 1×1 conv, depth-wise conv, and 1×1 conv respectively. The input is modulated and fed to the denoising U-Net with IGAB:

$$\tilde{I}^{(k)} = I^{(k)} \odot L^{(k)} + I^{(k)}, \quad I^{(k+1)} = \mathscr{D}(\tilde{I}^{(k)}, F_{\text{illu}}) + \tilde{I}^{(k)}, \tag{5}$$

where $\tilde{I}^{(k)}$ is the lit-up image, $I^{(k+1)}$ is the output of the k-th sub-stage, and $\mathscr{D}$ is the denoising U-Net with illumination-guided attention blocks (IGAB). Each IGAB uses

$$\mathbf{A} = \text{softmax}(\mathbf{K}^\top \mathbf{Q} / \sqrt{d}), \quad \text{Attn} = \mathbf{A}(\mathbf{V} \odot F_{\text{illu}}), \tag{6}$$

and

$$\begin{aligned} \text{IGAB}(x, F_{\text{illu}}) &= W_O \text{Attn} + \text{PE}(F_{\text{illu}}) + \text{FFN}(x), \\ \text{FFN}(x) &= W_2 \, \text{GELU}(\text{DWConv}(W_1 x)), \end{aligned} \tag{7}$$

where $\mathbf{A}$ is the attention weight matrix, Attn is the attention output, $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ are the query, key, and value from the input feature x, x is the input feature to the block, d is the head dimension, W_O, W_1, W_2 are learnable projection matrices, PE is positional encoding, FFN is the feed-forward network, GELU is the Gaussian Error Linear Unit activation, DWConv is depth-wise convolution, and $\odot$ is element-wise multiplication. Thus

$$\begin{aligned} I^{(1)} &= \mathscr{F}_1(I_{\text{low}}), \\ I^{(2)} &= \mathscr{F}_2(I^{(1)}), \\ I_{\text{stage1}} = I^{(4)} &= \mathscr{F}_3(I^{(2)}). \end{aligned} \tag{8}$$

where $\mathscr{F}_k$ denotes the k-th single-stage module ($k = 1,2,3$), and $I^{(4)}$ denotes the output after the third sub-stage (i.e. I_{stage1}). We train stage one jointly with stage two.

Stage II: Conditional U-Net refiner

Stage two takes the condition $C = [I_{\text{low}}; I_{\text{stage1}}; L] \in \mathbb{R}^{9 \times H \times W}$ and fixed $t = 0$. The time embedding is sinusoidal:

$$\mathbf{t}_{\text{emb}} = \text{MLP}(\text{PE}(t)), \tag{9}$$

where t is the time step (fixed at 0), PE is sinusoidal positional encoding, MLP is a multilayer perceptron, and $\mathbf{t}_{\text{emb}}$ is the time embedding vector. Each residual block is modulated by $\mathbf{t}_{\text{emb}}$:

$$\begin{aligned} h &= \sigma(\text{Conv}_1(x) + \text{proj}(\mathbf{t}_{\text{emb}})), \\ \text{ResBlock}(x, \mathbf{t}_{\text{emb}}) &= \sigma(\text{Conv}_2(h)) + \text{skip}(x), \end{aligned} \tag{10}$$

where x is the input to the block, h is the intermediate feature after the first conv and modulation, σ is the SiLU activation, Conv_1, Conv_2 are convolutional layers, proj projects $\mathbf{t}_{\text{emb}}$ to the feature dimension, and $\text{skip}(x)$ is the identity shortcut. The U-Net computes $x_0 = \text{Conv}_{3\times3}(C)$, then encoder ResBlocks with pooling, bottleneck, decoder ResBlocks with skip connections, and

$$\Delta I = \text{Conv}_{3\times3}(u_2), \qquad I_{\text{final}} = I_{\text{stage1}} + \Delta I, \tag{11}$$

where x_0 is the initial feature map from the 3×3 convolution on C, $C = [I_{\text{low}}; I_{\text{stage1}}; L]$ is the 9-channel condition, ΔI is the predicted residual, and u_2 is the decoder feature after the second upsampling block. This single-step refinement improves detail and reduces artefacts. Stage one already handles most of the illumination correction and denoising; the remaining errors are often local and residual (e.g. colour cast, residual noise in dark regions). The refiner, conditioned on I_{low}, I_{stage1}, and L, corrects these with a lightweight U-Net rather than enlarging the first-stage model. Both stages are trained end-to-end on paired normal/low-light images; the training objective and data generation are described in the Experiments section. Algorithm 1 summarises the inference procedure.

Experiments

Datasets

We use a paired dataset: normal-light images (reference, from a construction or site image set) and synthetically darkened low-light images. In line with common practice in supervised low-light enhancement when large-scale real low-light/normal-light pairs are scarce (Wei et al., 2018), we synthesise the low-light side from normal-light references rather than using night-time captures. Low-light images are generated with the darkening curve $y = b \cdot (a \cdot x)^\gamma$, where x is the normal-light pixel value, y is the darkened value, a, b, γ are coefficients sampled per channel (e.g. $a \in [0.38, 0.52]$, $b \in [0.22, 0.38]$, $\gamma \in [1.6, 2.0]$). Because coefficients are randomised per image and channel, some synthetic low-light inputs are intentionally very dark, covering challenging under-exposure. Optionally we add signal-dependent noise variance $\sigma_s^2 \propto L$ (with L the local illumination) and signal-independent noise variance σ_c^2, yielding $z = \text{clip}(y + n_s + n_c)$, where z is the final low-light value, and n_s, n_c are zero-mean Gaussian noise. We refer to the reference set as the normal dataset and the darkened set as the night dataset; the dataset comprises 200 image pairs. We randomly split these pairs into 160 pairs for training and 40 pairs for held-out testing (ratio 4:1). Geometric augmentation (flips, rotation) and crops of 128×128 are used during training.

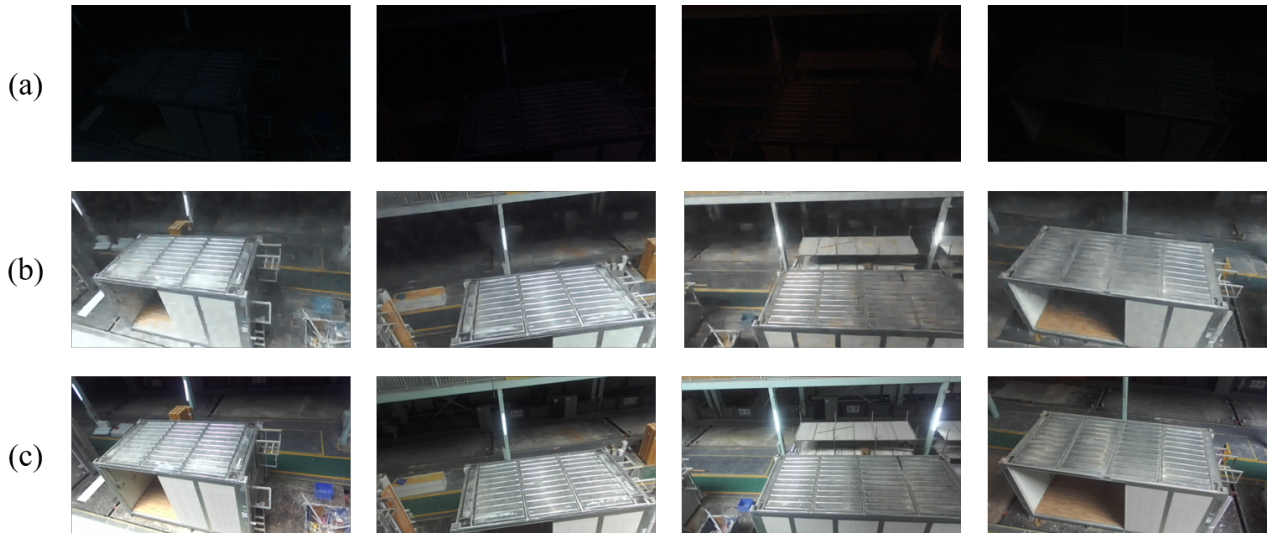


Figure 2: Multiple examples: (a) low-light input, (b) RetinexRefiner output, (c) ground truth. Each column is one sample from the construction dataset.

Implementations

We implement RetinexRefiner in PyTorch. Stage one (illumination estimation and initial enhancement) uses the RetinexFormer backbone (Cai et al., 2023) unchanged; stage two is a conditional U-Net that takes the 9-channel concatenation $[I_{\text{low}}; I_{\text{stage1}}; L]$ and predicts a residual. The refiner has an encoder-decoder structure with skip connections: it downsamples twice (average pooling), processes the bottleneck, then upsamples with bilinear interpolation and concatenates skip features before predicting the residual. The residual blocks are modulated by a sinusoidal time embedding (we fix $t = 0$ at inference). Both stages are trained end-to-end. The training objective is L1 loss between the output and the ground-truth normal-light image; we do not use a separate loss on stage one. We train for 60,000 iterations with Adam (learning rate 1×10^{-4}, betas 0.9 and 0.999), cosine-annealing restart (three periods of 5000 iterations, minimum learning rates 10^{-4} and 10^{-5}), batch size 4, and mixup (beta 1.2) with optional identity mapping. Gradient clipping is applied. Training and model settings are summarised in Table 1.

Table 1: Training and model settings for RetinexRefiner: number of iterations, batch size, crop size, optimizer (Adam with learning rate 1×10^{-4}), and loss (L1 to ground-truth normal-light image).

Setting	Value
Iterations	60k
Batch size	4
Patch size	128×128
Optimiser	Adam (1×10^{-4})
Loss	L1

Evaluation metrics

We use peak signal-to-noise ratio (PSNR, in dB) and structural similarity (SSIM) as evaluation metrics. PSNR measures the ratio between the maximum possible signal and the reconstruction error; higher is better. SSIM captures luminance, contrast, and structure similarity between the enhanced image and the reference; it lies in $[0, 1]$ and higher is better. The normal-light images serve as the reference (ground truth) for both metrics. These are standard full-reference metrics for paired low-light enhancement when references are available.

Quantitative results

We compare with traditional Retinex-based and deep learning methods: LIME (Guo et al., 2016), DUAL (Zhang et al., 2019a), RetinexNet (Wei et al., 2018), KinD (Zhang et al., 2019b), and RetinexFormer (Cai et al., 2023) (stage one only). Table 2 reports PSNR and SSIM on the held-out test set of the construction night dataset. LIME and DUAL, as unsupervised methods that do not optimise for reference matching, yield the lowest scores (PSNR $\approx$ 9.15 dB, SSIM $\approx$ 0.05). Among learning-based methods, RetinexNet reaches 21.50 dB PSNR and 0.69 SSIM; KinD attains 18.16 dB PSNR with higher SSIM (0.82). RetinexFormer (stage one) reaches 27.47 dB and 0.82 SSIM. RetinexFormer is already a strong baseline; further gains are often incremental when stacking a second stage. RetinexRefiner improves over this baseline to 28.07 dB PSNR and 0.83 SSIM (about +0.6 dB and +0.007 SSIM), showing that the conditional U-Net refiner still adds consistent gain. The refiner corrects residual colour cast, noise, and detail loss that remain after stage one. Stage one alone can leave artefacts in very dark regions; the refiner, conditioned on I_{low}, I_{stage1}, and L, both preserves well-enhanced regions and fixes under-corrected ones with modest extra parameters and one extra forward pass.

Algorithm 1 RetinexRefiner inference. Stage I (illumination estimation and initial enhancement) computes the illumination map and runs three sub-stages to obtain I_{stage1}; Stage II builds the 9-channel condition, runs the refiner U-Net, and adds the residual to obtain I_{final}.

Require: Low-light image $I_{\text{low}} \in \mathbb{R}^{3\times H\times W}$
Ensure: Enhanced image $I_{\text{final}} \in \mathbb{R}^{3\times H\times W}$

Stage I: Illumination estimation and initial enhancement
$F_{\text{illu}}, L \leftarrow \mathscr{E}(I_{\text{low}})$ ▷ Illumination map and features
$I^{(1)} \leftarrow \mathscr{F}_1(I_{\text{low}})$ ▷ First sub-stage
$I^{(2)} \leftarrow \mathscr{F}_2(I^{(1)})$ ▷ Second stage (IGAB with F_{illu})
$I_{\text{stage1}} \leftarrow \mathscr{F}_3(I^{(2)})$ ▷ Third stage; initial enhanced image
Stage II: Conditional U-Net refinement
$C \leftarrow [I_{\text{low}}; I_{\text{stage1}}; L]$ ▷ 9-channel condition
$\mathbf{t}_{\text{emb}} \leftarrow \text{MLP}(\text{PE}(t=0))$ ▷ Sinusoidal time embedding
$x_0 \leftarrow \text{Conv}_{3\times3}(C)$ ▷ Initial feature map
Encode: $x_0 \to x_1 \to x_2$ with ResBlocks + pooling, save skip features
Bottleneck: process x_2 with ResBlocks modulated by $\mathbf{t}_{\text{emb}}$ ▷ $h = \sigma(\text{Conv}_1(x) + \text{proj}(\mathbf{t}_{\text{emb}}))$, out $= \sigma(\text{Conv}_2(h)) + \text{skip}(x)$
Decode: upsample, concatenate skips, ResBlocks; obtain u_2 ▷ decoder feature after second upsampling
$\Delta I \leftarrow \text{Conv}_{3\times3}(u_2)$; $I_{\text{final}} \leftarrow I_{\text{stage1}} + \Delta I$ ▷ output equation
return I_{final}

Qualitative results

Figure 2 shows multiple examples from the construction dataset: each column is one sample, with (a) low-light input, (b) RetinexRefiner output, and (c) ground truth (normal-light references are real photographs; low-light inputs are synthetic). The enhanced results recover visibility and detail and are close to the reference. Figure 3 compares methods on the same low-light scene: the top row gives the input and results of LIME, DUAL, and KinD; the bottom row gives the ground truth and results of RetinexNet, RetinexFormer, and the proposed RetinexRefiner. LIME and DUAL exhibit strong colour cast; KinD and RetinexNet improve brightness but remain hazier or less faithful to the reference. The proposed RetinexRefiner is closest to the ground truth in brightness, colour, and detail. In construction and site imagery, text, equipment, and structural details become more visible for monitoring and inspection. These comparisons complement the PSNR/SSIM tables by illustrating perceptual differences that are not fully reflected by those scores alone.

Ablation study

We conduct ablations on the test set of the construction night dataset (Table 3). Training with darkening only yields 16.33 dB PSNR and 0.20 SSIM; adding noise during data synthesis raises performance to 28.07 dB and 0.83 SSIM, showing that noise augmentation improves robustness to real low-light conditions. For the refiner inputs, the full condition $[I_{\text{low}}; I_{\text{stage1}}; L]$ reaches 28.07 dB and 0.83 SSIM. Omitting the illumination map L gives 27.92 dB and 0.83 SSIM; omitting I_{low} gives 27.92 dB and 0.82 SSIM. Both variants are slightly worse than the full model, indicating that L helps the refiner correct residual artefacts and that I_{low} provides useful low-light context. These results validate the design of the conditional refiner.

Table 2: Quantitative comparison (PSNR / SSIM) on the test set of the construction night dataset.

Method	PSNR (dB)	SSIM
LIME (Guo et al., 2016)	9.1515	0.0499
DUAL (Zhang et al., 2019a)	9.1495	0.0522
KinD (Zhang et al., 2019b)	18.1600	0.8158
RetinexNet (Wei et al., 2018)	21.5028	0.6931
RetinexFormer (stage one) (Cai et al., 2023)	27.4707	0.8201
RetinexRefiner (ours)	**28.0747**	**0.8265**

Table 3: Ablation study (test set) on the construction night dataset.

Variant	PSNR (dB)	SSIM
Darkening only	16.3250	0.2033
Darkening + noise	**28.0747**	**0.8265**
Refiner w/o L	27.9220	0.8271
Refiner w/o I_{low}	27.9226	0.8239
Full $[I_{\text{low}}; I_{\text{stage1}}; L]$	**28.0747**	**0.8265**

Discussion

RetinexRefiner contributes to low-light enhancement for construction and built-environment imagery in two ways. First, in terms of application, prior low-light work has not targeted construction or modular assembly; we focus on night construction and similar dim environments for monitoring, inspection, assembly and related vision tasks, where RetinexRefiner can improve visibility over the compared methods. Second, RetinexRefiner provides a two-stage low-light pipeline for construction and built-environment imagery. The main novelty is a lightweight conditional refiner after a Retinex-based backbone: unlike LIME (Guo et al., 2016) and DUAL (Zhang et al., 2019a), we use end-to-end trained stages on paired images; unlike RetinexNet (Wei et al., 2018) and KinD (Zhang et al., 2019b), we train both stages jointly with a single refiner; relative to RetinexFormer (Cai et al., 2023) as stage one, the refiner conditioned on I_{low}, I_{stage1}, and L fixes residual colour cast and noise in dark regions. We position this as a practical refinement module for construction-site imagery under the paired protocol used in this study.

While this study proposes a practical design with measur-

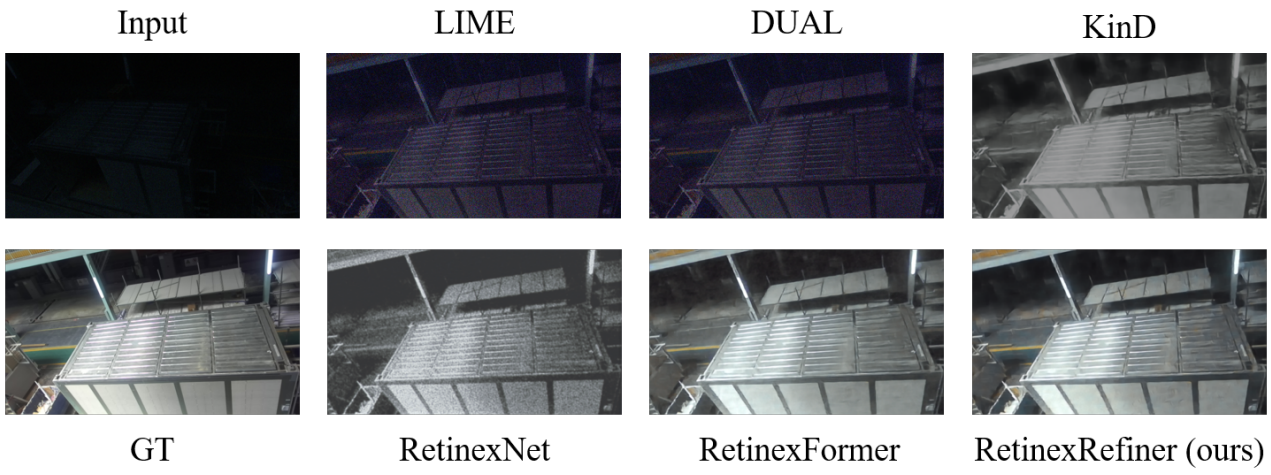


Figure 3: Comparison of methods on one low-light image (construction scene). The low-light input (top left) is a synthetic degradation of the ground-truth normal-light image (bottom left). Top row: input, LIME, DUAL, KinD. Bottom row: ground truth, RetinexNet, RetinexFormer, and the proposed RetinexRefiner.

able gains, it has the following limitations. First, the low-light image in each training/test pair is synthetically generated (darkening and optional noise) from a real normal-light construction or site photograph used as the reference; we do not use pixel-aligned real night-time field captures of the same scene paired with normal-light ground truth. The setup is therefore supervised on synthetic low-light with real references, not unpaired real night-time enhancement. Real camera response and sensor noise may differ from our synthetic low-light branch, so performance may degrade under domain shift; future work includes collecting real paired field imagery where feasible and exploring unpaired enhancement methods. Second, we report only full-reference metrics (PSNR and SSIM); they may not fully reflect perceptual issues such as colour shifts, haloing, or over-smoothing. We do not report no-reference image quality metrics or subjective user studies in this work; these are important directions for future evaluation. Third, the impact of low-light enhancement on downstream tasks such as object detection or semantic segmentation has not been measured—future work will validate this impact. Fourth, we currently compare Retinex-style and classical baselines under a paired protocol; broader cross-family comparisons will be addressed in future work.

Conclusion

In conclusion, we have presented RetinexRefiner, a two-stage Retinex-based method for low-light image enhancement in construction and built-environment contexts. Unlike multi-stage methods that train stages separately or one-stage methods that leave residual artefacts, our pipeline reuses RetinexFormer for illumination estimation and initial enhancement, then applies a lightweight conditional U-Net refiner conditioned on the low-light image, the stage-one output, and the illumination map to predict a residual that improves detail and reduces colour cast and noise; the two stages are trained end-to-end on paired images. Experiments on the construction night dataset show that RetinexRefiner reaches 28.07 dB PSNR and 0.83 SSIM, outperforming the stage-one baseline and comparison methods (LIME, RetinexNet, KinD, RetinexFormer), demonstrating its effectiveness for low-light imagery relevant to modular construction and similar dim environments. The approach is broadly applicable to construction sites, indoor sites, and other scenarios where reliable imaging in the dark is needed.

Future work will focus on: (1) exploring unpaired low-light enhancement methods and using real paired data from construction or field imagery to improve robustness under domain shift; (2) validating the impact of enhancement on downstream tasks such as object detection and semantic segmentation, and reporting task-level metrics (e.g. detection accuracy on enhanced images); (3) incorporating no-reference image quality metrics and subjective user studies to assess perceptual quality beyond PSNR and SSIM.

Acknowledgments

This research was supported by grants from the Innovation and Technology Fund, Hong Kong (Reference No. PRP-048-24LI), and by the Research Institute for Sustainable Urban Development (RISUD), The Hong Kong Polytechnic University (Project No. P0052942). We thank China State Construction Hailong Technology Company Limited for providing the data.

References

Cai, Y., Bian, H., Lin, J., Wang, H., Timofte, R., and Zhang, Y. (2023). Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 12504–12513. https://doi.org/10.1109/ICCV51070.2023.01149.

Chauhan, I. and Seppänen, O. (2023). Automatic indoor

construction progress monitoring: challenges and solutions. In Proceedings of the European Conference on Computing in Construction, volume 4, pages 0–0. European Council on Computing in Construction. https://doi.org/10.35490/EC3.2023.225.

Chen, C., Chen, Q., Xu, J., and Koltun, V. (2018). Learning to see in the dark. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3291–3300. https://doi.org/10.1109/CVPR.2018.00347.

Chobola, T., Liu, Y., Zhang, H., Schnabel, J. A., and Peng, T. (2024). Fast context-based low-light image enhancement via neural implicit representations. In European Conference on Computer Vision, pages 413–430. Springer. https://doi.org/10.1007/978-3-031-73016-0_24.

Ekanayake, B., Wong, J. K.-W., Fini, A. A. F., and Smith, P. (2021). Computer vision-based interior construction progress monitoring: A literature review and future research directions. Automation in Construction, 127:103705. https://doi.org/10.1016/j.autcon.2021.103705.

Fang, W., Ding, L., Love, P. E., Luo, H., Li, H., Peña-Mora, F., Zhong, B., and Zhou, C. (2020). Computer vision applications in construction safety assurance. Automation in Construction, 110:103013. https://doi.org/10.1016/j.autcon.2019.103013.

Feng, Y., Hou, S., Lin, H., Zhu, Y., Wu, P., Dong, W., Sun, J., Yan, Q., and Zhang, Y. (2024). Diff-flight: Integrating content and detail for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6143–6152. https://doi.org/10.1109/CVPRW63382.2024.00619.

Ferdous, W., Bai, Y., Ngo, T. D., Manalo, A., and Mendis, P. (2019). New advancements, challenges and opportunities of multi-storey modular buildings–a state-of-the-art review. Engineering Structures, 183:883–893. https://doi.org/10.1016/j.engstruct.2019.01.061.

Guo, C., Li, C., Guo, J., Loy, C. C., Hou, J., Kwong, S., and Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 1780–1789. https://doi.org/10.1109/CVPR42600.2020.00185.

Guo, X., Li, Y., and Ling, H. (2016). Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2):982–993. https://doi.org/10.1109/TIP.2016.2639450.

Jiang, H., Guan, B., Liu, Z., Liu, X., Yu, J., Liu, Z., Han, S., and Liu, S. (2025). Learning to see in the extremely dark. arXiv preprint arXiv:2506.21132. https://doi.org/10.48550/arXiv.2506.21132.

Land, E. H. (1977). The retinex theory of color vision. Scientific American, 237(6):108–129. https://doi.org/10.1038/scientificamerican1277-108.

Li, C., Guo, C., Han, L., Jiang, J., Cheng, M.-M., Gu, J., and Loy, C. C. (2021). Low-light image and video enhancement using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12):9396–9416. https://doi.org/10.1109/TPAMI.2021.3126387.

Liu, H., Jin, F., Zeng, H., Pu, H., and Fan, B. (2023). Image enhancement guided object detection in visually degraded scenes. IEEE Transactions on Neural Networks and Learning Systems, 35(10):14164–14177. https://doi.org/10.1109/TNNLS.2023.3274926.

Liu, R., Ma, L., Zhang, J., Fan, X., and Luo, Z. (2021). Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10561–10570. https://doi.org/10.1109/CVPR46437.2021.01042.

Qi, Y., Li, X., and Jiang, R. (2025). Vision-based segmentation, measurement, and pose estimation in modular integrated construction. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, volume 42, pages 1097–1104. IAARC Publications. https://doi.org/10.22260/ISARC2025/0142.

Tsz Wai, C., Wai Yi, P., Ibrahim Olanrewaju, O., Abdelmageed, S., Hussein, M., Tariq, S., and Zayed, T. (2023). A critical analysis of benefits and challenges of implementing modular integrated construction. International Journal of Construction Management, 23(4):656–668. https://doi.org/10.1080/15623599.2021.1907525.

Wang, W., Yang, H., Fu, J., and Liu, J. (2024). Zero-reference low-light enhancement via physical quadruple priors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 26057–26066. https://doi.org/10.1109/CVPR52733.2024.02462.

Wang, Y., Wan, R., Yang, W., Li, H., Chau, L.-P., and Kot, A. (2022). Low-light image enhancement with normalizing flow. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 2604–2612. https://doi.org/10.1609/aaai.v36i3.20162.

Wei, C., Wang, W., Yang, W., and Liu, J. (2018). Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560. https://doi.org/10.48550/arXiv.1808.04560.

Yan, Q., Feng, Y., Zhang, C., Pang, G., Shi, K., Wu, P., Dong, W., Sun, J., and Zhang, Y. (2025). Hvi: A new color space for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5678–5687. https://doi.org/10.1109/CVPR52734.2025.00533.

Zhang, Q., Nie, Y., and Zheng, W.-S. (2019a). Dual illumination estimation for robust exposure correction. In Computer Graphics Forum, volume 38, pages 243–252. Wiley Online Library. https://doi.org/10.1111/cgf.13833.

Zhang, Y., Zhang, J., and Guo, X. (2019b). Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, pages 1632–1640. https://doi.org/10.1145/3343031.3350926.

Zheng, Z., Zhang, Z., and Pan, W. (2020). Virtual prototyping-and transfer learning-enabled module detection for modular integrated construction. Automation in Construction, 120:103387. https://doi.org/10.1016/j.autcon.2020.103387.

Zhou, H., Dong, W., Liu, X., Liu, S., Min, X., Zhai, G., and Chen, J. (2024). Glare: Low light image enhancement via generative latent feature based codebook retrieval. In European Conference on Computer Vision, pages 36–54. Springer. https://doi.org/10.1007/978-3-031-73195-2_3.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

FROM GENERAL SPACE PLANNER TO LEARNING-BASED GENERATION: A SYSTEMATIC MAPPING OF REPRESENTATION, OBJECTIVES, AND CONTROL

Suhyung Jang[1,2] and André Borrmann[1,2]
[1]Georg Nemetschek Institute, Technical University of Munich, Munich, Germany
[1]Chair of Computing in Civil and Building Engineering, School of Design and Engineering, Technical University of Munich, Munich, Germany

Abstract

Recent advances in learning-based artificial intelligence (AI) show that shared representations (e.g., pixelized images and tokenized corpora) enable multitask foundation model training, whereas digital building layout representations have diverged. Grounded in Eastman's General Space Planner, this study systematically maps 59 studies over a half-century (1971–2025) to trace how building layout representations evolved with expanding objectives and emerging control mechanisms. Publications surged in the final half-decade (52.5%), with representations shifting from simplified to free-formed representations as control mechanisms and objectives diversify. The trends suggest a bottleneck, as consistent transfer of learning-based building layout generation is difficult without a shared representation.

Introduction

Automated building layout planning has been studied for more than five decades, beginning with early computational approaches in the 1960s and evolving through successive waves of constraint-based reasoning (Liggett, 2000), optimization, and, more recently, pattern-based generation (Weber et al., 2022). Despite such sustained effort, automatically generated building layouts have often been limited to single-floor rectilinear primitives. Although recent image-generation approaches have expanded the variety of generated forms, they still face difficulties integrating with current computer-aided design (CAD) or building information modeling (BIM) workflows (Jang et al., 2025b).

A building layout is not merely a set of space boundaries. Spaces, together with adjacency, connectivity, containment, and other relations must be considered. In addition, layout representations should reflect the hierarchy of building components (e.g., levels, rooms, and building elements). These requirements become more pronounced as layout planning shifts from conforming to connectivity and planned proportions conditions to satisfying multiple and sometimes competing design objectives such as code compliance, target cost, environmental performance, and user preferences.

Digital building layout representation, therefore, must represent geometry and topology simultaneously and often must combine multiple encodings to do so (Jang et al., 2025a). As a result, the field has progressed through diverse and sometimes incompatible choices of representation, alongside diversified objective formulations, emerging representations and control techniques. This has made it difficult to establish a stable representational basis for comparing methods or transferring models across building layout tasks. By contrast, in vision and language, shared representations provide stable minimal units for composition spaces, allowing learning-based artificial intelligence (AI) models to learn common structures and generalize across datasets and downstream tasks (LeCun et al., 2015).

For raster images, inputs are consistently defined by a two-dimensional pixel grid with channel values, and for text, contemporary transformer pipelines convert language to token sequences whose semantics are embedded into latent spaces to model context. These shared conventions have been central to multitask learning and foundation models. Building layout representations, however, have diverged over time as objectives and control mechanisms expanded, leaving unclear which representational components are minimally required for transferable learning-based generation across diverse objectives and contexts.

To address this issue, this paper analyzes the automated building layout planning studies conducted over the last half-century to clarify how building layouts have been represented, what has been optimized, and how layout generation has been controlled. To make longitudinal comparisons possible, we use Charles Eastman's General Space Planner (GSP) as a foundational reference (Eastman, 1973), and reorganize its core elements into a three-part analytical framework consisting of representation, objective, and control for decade-by-decade trend mapping (Figure 1).

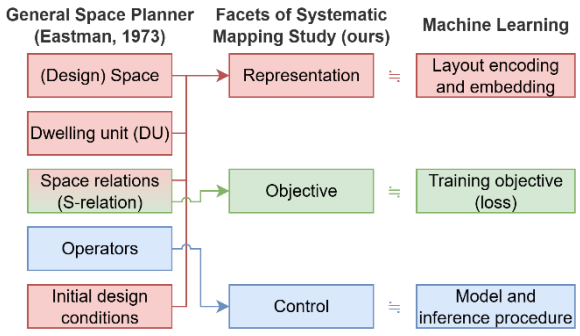


Figure 1: Three analytical facets of the systematic mapping study

Representation captures how layouts are encoded in terms of spatial units, boundaries, and relational structures. Objectives capture what requirements are targeted and how they are formalized through metrics, constraints, or performance criteria. Control captures how planning proceeds through mechanisms used to generate, modify,

and converge on candidate layouts, including search, optimization, constraint solving, learning-based procedures, or interactive refinement. In generalized learning-based building layout generation, these three facets correspond to the representational space in which layouts are encoded, the objective function that defines desirable solutions, and the procedural mechanism through which candidate layouts are generated and refined. This framing is designed to provide a common analytical basis for comparing classical and recent learning-based approaches within a single coding scheme. The goal of this study is to systematically identify persistent elements as central features in building layout planning over multiple decades, as well as to elucidate emerging requirements and methodological advancements associated with the increasing complexity of building layout planning. The analysis is directed toward establishing prerequisites for learning-based AI models that support transferable and extensible general space planning.

The remainder of this paper is organized as follows. The second section describes the systematic mapping study protocol and coding procedure. The third, fourth, and fifth sections present decade-by-decade analyses of trends in representation, objectives, and control. The sixth section discusses current challenges and future prospects, and the seventh section concludes the paper.

Research method

Systematic mapping studies aim to structure a research area by classifying and counting contributions across a predefined scheme, providing an overview of what has been studied and where the literature has appeared. We adopted the updated systematic mapping study guidelines (Petersen et al., 2015), and followed a protocol consisting of planning, conducting, and reporting activities.

Study identification

Space layout planning has co-evolved with broader layout optimization research in manufacturing and industrial engineering, which makes keyword-only retrieval prone to substantial noise. To keep the map focused on building layout planning, a discipline-scoped sampling strategy was adopted for study identification. This study first identified studies from representative journals in the field by discipline. Journal names were collected from Scimago Journal Rankings (SJR), gathering all journals under the "Engineering" subject area, specifically "Building and Construction," and "Architecture" subject categories. Then we developed a Python code using pybliometrics.scopus, a library that supports application programming interface (API)-based searching, to collect the papers from the Scopus database within the relevant disciplines. We queried Scopus using title-level layout terms, namely "space planning," "space layout," "building layout," "architectural layout," "room layout," "space plan," "room plan," "floor plan," and "floorplan," combined with building-related filters in titles, abstracts, and keywords (e.g., "building," "architectural," "dwelling unit," "room," and "housing"). In total, 212 records were retrieved within the target time window (1970-2025).

Study selection was conducted in two stages through title and abstract screening, followed by a full-text screening. We excluded (1) review papers, (2) design theory papers without layout generation, (3) layout analysis papers without automated planning/generation, and (4) studies focusing on urban or site-scale layouts. This screening resulted in 35 included articles.

To improve coverage of earlier and transitional work, we additionally identified candidate studies from two prior secondary sources covering the search window, with one source targeting studies before 2000 (Liggett, 2000) and another targeting studies between 2000 and 2025 (Xu et al., 2025). After removing duplicates and records that could not be retrieved in full text, 17 additional articles were included. Additionally, seven conference papers from the computer vision and machine learning domain were included after snowballing. In total, 59 records were included in the final map.

We then organized the 59 included studies by decade for analysis. The final corpus contains 3 studies from the 1970s, 2 from the 1980s, 6 from the 1990s, 5 from the 2000s, 12 from the 2010s, and 31 from 2020–2025, meaning that publications from the most recent five years account for 52.54% of the screened corpus.

Data extraction and classification

For systematic mapping, we constructed a facet-based classification scheme grounded in GSP's core concepts and aligned the facets with learning-based AI components to support cross-paradigm comparison.

GSP defines a layout planning problem using the following concepts: Space, Dwelling Unit (DU), Space Relations (S-relations), Operators, and an initial design condition. In GSP, S-relations specify constraints to be satisfied, and operators specify how DUs are manipulated within Space. However, across fifty years of research, the role of "relations" has expanded: relations are used in two ways: (1) to encode relational structure (e.g., adjacency / connectivity graphs) and (2) to define what must be satisfied or optimized (e.g., requirements, objectives, performance criteria). To avoid conflating these roles, we separate relation information into two coding targets: relational encoding under *representation*, and relation-derived requirements under *objectives*.

The classification is organized into three primary facets: *representation*, *objective*, and *control*:

- *Representation* captures how the design space, units, and relations are encoded. Because building layouts are often represented using mixed forms (e.g., rectangular vector boundaries together with adjacency graphs), we record (R1) the primary representation and (R2) one or more supporting representations. We also map the representation dimension (R3), where 2.5D augments planar layouts with limited height information (e.g., extrusion or vertical elements for simulation) and 3D models spaces as full volumetric entities.

- *Objective* captures what requirements the method targets and how they are formalized. We record (O1) objective structure (e.g., constraint satisfaction, single-objective, multi-objective, non-optimization generation, learned objective) and (O2) one or more objective metric families (e.g., connectivity, cost, preferences, performance).
- *Control* captures how candidate layouts are generated, modified, and converged. We record (C1) the primary control mechanism, (C2) supporting mechanisms used for enforcement or refinement, and (C3) the interaction mode.

Table 1: Sub-facets and categories of the three analytical facets of systematic mapping study.

Facet	Sub-facet	Categories
Representation (R)	Primary representation (R1)	R1∈{Vector (Rectangle), Vector (Coordinate), Graph, Grid, Voxel, Image}
	Supporting representation (R2)	R2⊆{Vector (Rectangle), Vector (Coordinate), Graph, Grid, Voxel, Image}
	Dimension (R3)	R3∈{2D, 2.5D, 3D}
Objectives (O)	Structure (O1)	O1∈{Constraint-satisfaction, Single-objective, Multi-objective, Learned-objective, Non-objective-oriented}
	Metric (O2)	O2⊆{Adjacency/Connectivity, Area/Size, Circulation/Travel, Distance/Proximity, Orientation/View, Thermal/Energy, Lighting/Daylight, Code/Regulation, Cost, Structural/Engineering performance, Preference/User, Constructability}
Control (C)	Primary mechanism (C1)	C1∈{Operation-based, Case-based, Parameter-based, Learned-based (single-stage), Learned-based (multi-stage)}
	Supporting mechanism (C2)	C2⊆{Operation-based, Case-based, Parameter-based, Learned-based (single-stage), Learned-based (multi-stage)}
	Interactive control (C3)	C3∈{Automatic generation, Interactive generation}

The classification scheme was developed iteratively through an initial category draft based on domain knowledge and subsequent refinement by consolidating and renaming categories based on patterns observed in the collected studies (Table 1).

Representation (R)

Representation describes how a building layout is encoded as a computational object, including the spatial primitives (e.g., rooms, boundaries, cells) and the relational structures used to express adjacency, connectivity, and hierarchy when applicable. Because layout planning requires both geometric reasoning (shape, size, coordinates) and topological reasoning (relations and constraints), many studies rely on heterogeneous representational stacks rather than a single universal format. Consistent with early space-planning work that combined relational encodings with explicit geometric units (Eastman, 1973; Grason, 1971; Pfefferkorn, 1972), we map representation at three levels: (1) a primary representation (R1) that serves as the main modeling unit for generation or optimization, (2) a supporting representation (R2) used to encode constraints, semantics, or downstream structures, and (3) the representation dimension (R3), which captures whether the layout is modeled in 2D, 2.5D, or 3D. Here, planar layout generation with extrusion but without explicit vertical relations is classified as 2.5D, while methods that represent vertical relations are classified as 3D.

Primary representation (R1)

Figure 2 summarizes a shift from early emphasis on coarse digital layout encodings toward higher geometric freedom via grid- and image-based formats.

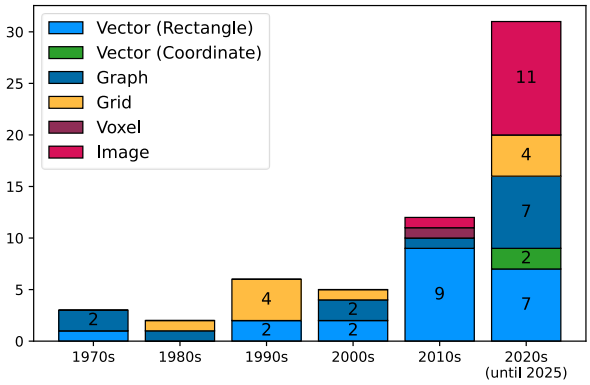


Figure 2: Period-wise distribution of primary representation formats (R1).

In the 1970–80s (n=5), graph dominates (60.0%), with vector-rectangle and grid each at 20.0%, reflecting topology-oriented formulations with simple geometric instantiation (Roth et al., 1982). Grid-based representation marks an early move toward freer boundary expression (Liggett and Mitchell, 1981). In the 1990s (n=6), grid-based representations become the majority primary representation (66.7%). Several studies in this period proposed operator-based grid methods that expand space boundaries to satisfy constraints and produce an initial draft, which is then refined by a human designer toward a less constrained layout (Jo and Gero, 1995). In the 2000s (n=5), primary encodings split between graph and vector-rectangle (40.0% each), with grid at 20.0%, indicating coexistence of topological representation and explicit geometric units. The 2010s (n=12) are dominated by vector-rectangle (75.0%) while image and voxel begin to appear (8.3% each), reflecting diversification through both 2D and 3D encodings (Dino, 2016; Hua, 2016; Marson and Musse, 2010). In the 2020s (until 2025; n=31), image becomes the largest primary category (35.5%), with graph still substantial (22.6%). Overall, the trend indicates increasing reliance on representations that support flexible boundaries and forms, while graph persists as a stable carrier of topological representation.

Supporting representation (R2)

Figure 3 reports supporting representations used alongside R1. Since use of R2 is optional, the total number reflects coded occurrences rather than the number of papers per period. In the 1970–80s (n=4), vector-rectangle dominates (75.0%) with graph at 25.0%, consistent with setups where topological reasoning is primary but explicit rectangular geometry is used to instantiate spatial units. In later work, an explicit separation between topology and geometry is made more systematic; for example, Medjdoub and Yannou (2000) generate a graph describing topological configuration first and then place rectangular space units to realize geometry. In the 2010s (11 coded occurrences), graph becomes the dominant supporting representation (81.8%), with vector-coordinate and vector-rectangle each at 9.1%, suggesting more frequent use of graph as a feasibility/semantic scaffold around diverse primary encodings. In the 2020s (until 2025; n=29), graph remains the most frequent supporting representation (51.7%), followed by image (24.1%) and vector-rectangle (20.7%), with grid at 3.4%. This aligns with recent pipelines that combine image generation with explicit relational structure; for instance, Zheng and Petzold (2023) show graph-guided generation of topology-reflected layout images by using the graph together with image representations. Across periods, even as boundary representation becomes freer (grid/image/vector-coordinate), graph-based support remains central for preserving or recovering adjacency relations and enforcing topological constraints.

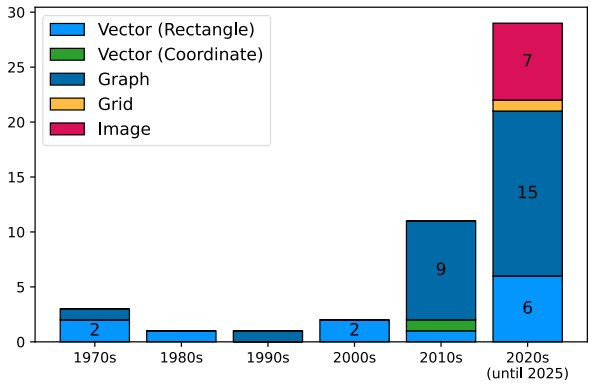


Figure 3: Period-wise distribution of supporting representation formats (R2).

Representation dimension (R3)

Figure 4 indicates that most studies remain focused on 2D representation, with limited adoption of 2.5D or 3D layout generation. In the 1970–80s (n=5), 1990s (n=6), and 2000s (n=5), all coded cases are 2D (100%). In the 2010s (n=12), 2D remains the majority (75.0%), while 2.5D appears (16.7%) and explicit 3D appears in a small minority (8.3%). The only 3D case uses voxel-based representation (Dino, 2016). In the 2020s (until 2025; n=31), 2D remains dominant (90.3%) with 2.5D at 9.7%. Most 2.5D cases arise when planar layouts are extruded or augmented with vertical elements (e.g., walls/windows) to support downstream simulation goals. For example, Rodrigues et al. (2014) extruded generated floor plans to form an energy-performance simulation loop for automated layout generation. Overall, fully automated 3D building layout generation remains underexplored, even though many studies note vertical variation in real buildings as a key limitation and future direction.

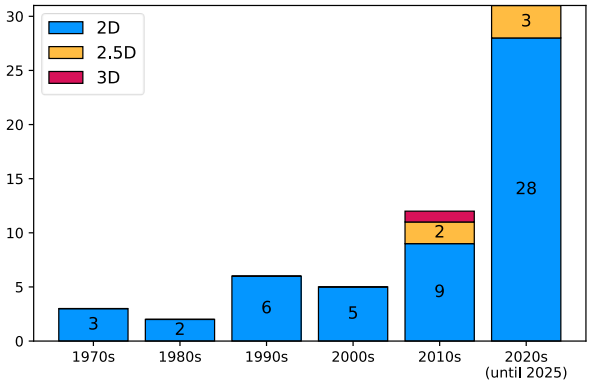


Figure 4: Period-wise distribution of representation dimension (R3).

Objectives (O)

Objectives capture what layout planning methods attempt to achieve and how requirements are formalized into computational targets. To remain compatible across rule-based optimization and learning-based pipelines, we separate objectives into objective structure (O1), which describes the formal type of target, and objective metrics (O2), which describe the substantive design criteria being operationalized.

Objective structure (O1)

Figure 5 shows a transition from constraint-satisfaction-based generation to multi-criteria optimization and, more recently, learned loss formulations. In the 1970–80s (n=5), objective structure is dominated by constraint satisfaction (80.0%), with the remainder framed as single-objective optimization (20.0%), reflecting early feasibility and rule-compliance formulations. From the 1990s (n=6), multi-objective optimization becomes dominant (66.7%), while constraint satisfaction remains substantial (33.3%), consistent with the broader use of evolutionary and annealing-style optimization for competing criteria.

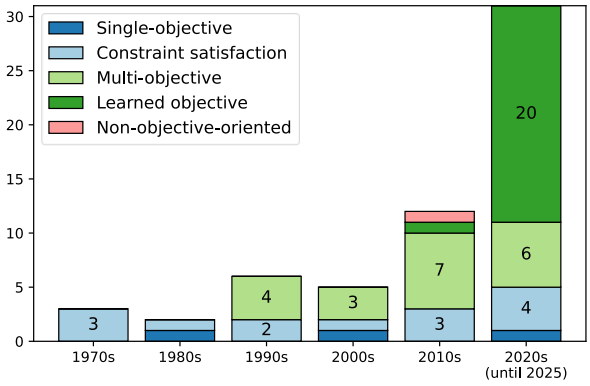


Figure 5: Period-wise distribution of objective structure (O1).

In the 2000s (n=5), multi-objective remains the majority (60.0%), with constraint satisfaction and single objective

each at 20.0%. In the 2010s (n=12), multi-objective continues to lead (58.3%), while constraint satisfaction accounts for 25.0%, and both learned objective (i.e., loss minimization) and non-optimization generation appear (8.3% each), indicating early entry of learning-based formulations. In the 2020s (until 2025; n=31), learned objective (loss minimization) becomes dominant (64.5%), while multi-objective persists (19.4%) and constraint satisfaction remains present (12.9%), with single objective at 3.2%. Overall, as learning-based AI becomes increasingly central to automated building layout generation, objective attainment shifts toward learning optimized patterns from accumulated data, which in turn elevates the importance of representations that can robustly encode and map design-relevant structure (e.g., geometry and topology) into machine-comprehensive formats.

Objective metric (O2)

Figure 6 indicates continuity in core building layout factors and divergence in the objective metrics over time. Across all periods, adjacency/connectivity and area/size are the most persistent and frequently utilized objectives, indicating that they are minimal requirements for functional building layout planning.

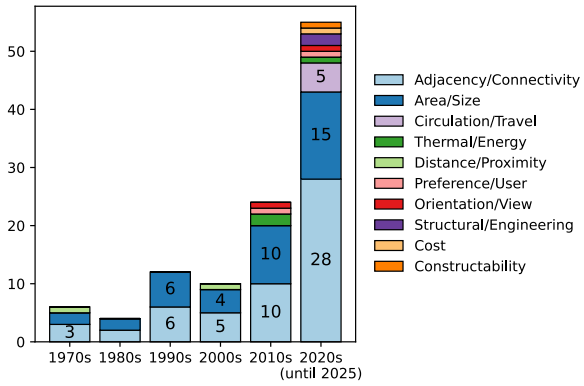


Figure 6: Period-wise distribution of objective metric categories (O2).

In early periods, between 1970s to the 2000s, are mainly concentrated on adjacency/connectivity, area/size, and distance/proximity conformance. Through the 2010s, additional categories appear, including thermal/energy, preference/user, and orientation/view. Between 2020-2025, the coded metric families expand further, adding circulation/travel, structural/engineering, cost, and constructability. These expanding objective scopes are represented by several examples. Rodrigues et al. (2014) exemplifies thermal- and energy-driven optimization by generating plan layouts under topological constraints and searching for alternatives that improve simulated thermal/energy performance. Likewise, Almashaqbeh and El-Rayes (2021) is representative of work optimizing cost and constructability, particularly in modular layout contexts. This diversification of more implementation-specific objectives reflects increasing interest beyond adjacency and area constraints satisfaction.

Control (C)

Control captures how building layout generation proceeds (i.e., the mechanisms used to explore, modify, and converge on candidate solutions) in each study. This facet includes classical strategies such as operator-based search and optimization (e.g., rule-based geometric operations followed by evaluation), case-based reuse (e.g., cased-based reasoning), parameter-based exploration of a defined design space (e.g., parametric modeling and design parameter exploration), and recent learning-based generation (e.g., generative adversarial networks, diffusion models, and transformer models). We code control into primary control mechanism (C1), supporting mechanism (C2), and interaction type (C3).

Primary control mechanism (C1)

Figure 7 shows a transition from operation-centric generation to learned generation in the 2020s, while non-learned mechanisms remain present. In the 1970–80s (n=5), primary control is entirely operation-based (100.0%). In the 1990s (n=6), operation-based control remains dominant (83.3%), with case-based approaches appearing at 16.7%. In the 2000s (n=5), operation-based and parameter-based control each account for 40.0%, with case-based at 20.0%, indicating diversification of classical control strategies. In the 2010s (n=13), operation-based control returns to a strong majority (69.2%), with parameter-based at 15.4%, and learned multi-stage and case-based each at 7.7%. In the 2020s (until 2025; n=31), learned single-stage becomes the largest category (41.9%), followed by learned multi-stage (32.2%), while operation-based (n=5) and parameter-based (n=3) remain at 16.1% and 9.7%. This split suggests that multi-stage learned pipelines may represent a modern reconfiguration of earlier operation-centric control, where learning provides the initial synthesis and operation-style steps reappear as gradual refinement to enforce feasibility and relational consistency.

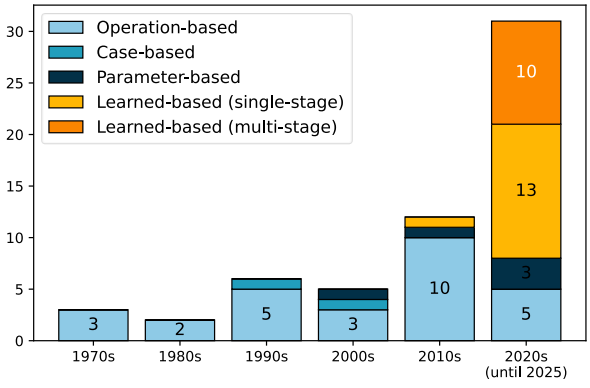


Figure 7: Period-wise distribution of primary control mechanisms (C1).

Supporting control mechanism (C2)

Figure 8 depicts that the use of C2 is sparsely reported across the mapped studies, so interpretation is restricted to cases where a supporting mechanism is explicitly described. In the 1990s, the observed supporting mechanisms are entirely parameter-based (3 occurrences).

In the 2000s, operation-based supporting control appears only once (1 occurrence). In the 2010s, parameter-based supporting control is more common (3 occurrences) while operation-based refinement again appears only in an isolated instance (1 occurrence). In the 2020s (until 2025 in this coding), the observed supporting mechanisms are parameter-based (2 occurrences), with no operation-based supporting refinement recorded.

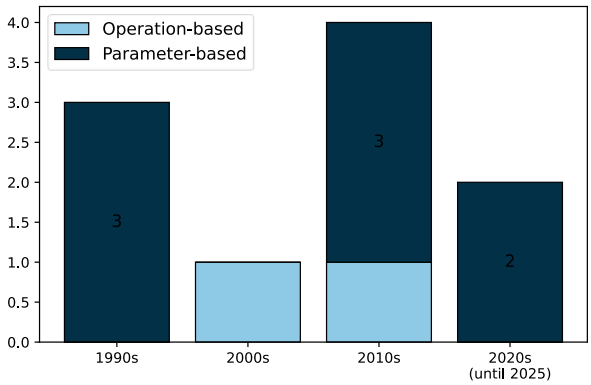


Figure 8: Period-wise distribution of supporting control mechanisms (C2).

Taken together, C2 functions as a robustness layer that stabilizes outcomes on top of the primary control mechanism. Parameter-based supporting control is consistent with restricting or re-parameterizing the design space to reduce sensitivity and guide convergence, whereas operation-based supporting control aligns with hybrid workflows where predefined operators are applied as cleanup or feasibility enforcement after an initial non-operator step. This pattern is consistent with pipelines where a layout is first proposed by a higher-level procedure and then adjusted through operators until constraints and geometric realizability are satisfied. For example, Rosenman (Rosenman, 2000) describes an evolutionary approach in which candidate plans are iteratively transformed and then modified with operators until constraints are met. Merrell et al. (2010) similarly illustrate a learned-to-operator pattern, where a probabilistic model supports program/adjacency generation and axis-aligned split/trim operators instantiate and refine rectangular layout, using supporting steps to ensure feasibility and geometric realization.

Interactive control (C3)

Across the field, interaction is typically limited to setting initial requirements (e.g., constraints, program, adjacency) while generation proceeds automatically, and interactive refinement remains comparatively underrepresented as a core research setting (Figure 9). Recent work begins to incorporate explicit interactive control modes: Shekhawat et al. (2021) propose a system where a user edits an adjacency graph via a graphical user interface (GUI) and the system generates a grid-based plan satisfying the edited relations; Zeng et al. (2024) similarly assume a designer-provided connectivity/adjacency graph and generate a plan layout conditioned on that structure. These examples indicate a shift toward interactive injection of topology or constraints in the 2020s, but the dominant workflow remains fully automated generation.

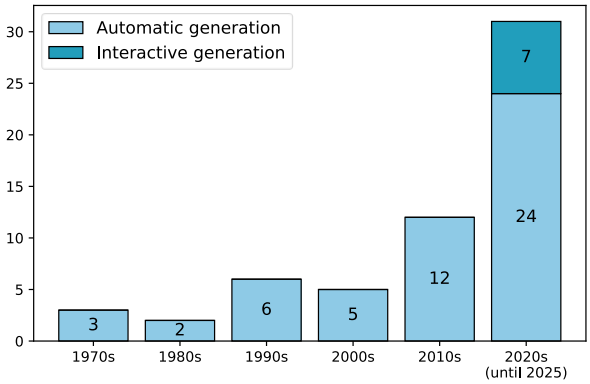


Figure 9: Period-wise distribution of interactive control types (C3).

Discussion

Over the last half-century, representation choices in building layout planning have diversified alongside increasing objective complexity and more sophisticated control mechanisms. At the same time, several primitives remain consistently central, most notably spatial proportions and adjacency relationships. The most recent half-decade marks a quantitative and qualitative inflection driven by learning-based AI. The rise of image-based primary representations and loss-minimization objectives, and learning-based control have become more prominent, increasing geometric freedom while further elevating the importance of representation and inductive bias in model learning.

Yet the trends observed in this study suggest that the field is moving toward divergence rather than convergence, unlike the trajectory that enabled foundation models in the natural language and vision domains. The confusion matrix in Figure 10 highlights this diverging landscape of representations, with the primary representation on the y-axis and the secondary representation on the x-axis. Early work in the 1970s primarily combined graph-based topology with rectangular vector units. In the 1980s–1990s, grid-based representations became more prominent, offering a robust design space that can express diverse shapes while preserving relative spatial relations. From the 2000s into the 2010s, hybridization became more systematic to represent richer context: rectangular primitives often remain the main modeling unit, while graphs increasingly serve as supporting scaffolds to maintain topological consistency.

In the 2020s, representation combinations diversify sharply around image-oriented generation, with graphs and vectors frequently retained as supporting encodings, and occasional pipelines that invert the direction by using images to support graph-based procedures. Overall, literature increasingly pursues higher geometric freedom while continuing to preserve explicit topological structure. This divergence may become more consequential as the field expands beyond predominantly two-dimensional layouts. As the R3 (Dimension) analysis indicates, most studies still operate in 2D, with only limited attention to

2.5D and 3D. However, real buildings are often multi-floor, and vertical spatial relationships must be accounted for, including inter-floor circulation, stacked adjacencies, and the continuity of cores and shafts. This may partly explain why 3D studies remain limited: compared with 2D layouts, 3D settings make it harder to define shared spatial units, build reusable datasets, and evaluate results under common criteria. One possible direction is to combine graph-based topology with more flexible geometric decoders, so that explicit relations can be preserved without overly constraining 3D form.

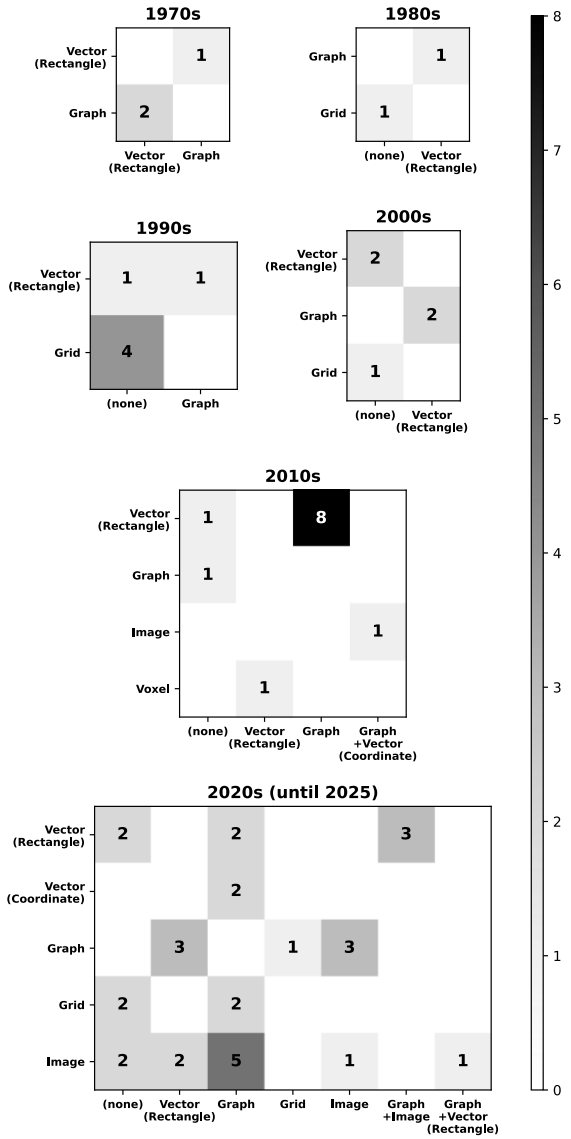


Figure 10: Combination of representation methods across periods.

If this divergence continues, the field may accumulate separate representational tracks for 2D, 2.5D, and 3D, each with different assumptions about basic units and primary relations. That would make comparison, reuse, and cumulative development more difficult. It would also weaken the conditions for domain-specific foundation model training, which usually benefits from a shared representation with stable primitives. A more convergent direction could support dataset sharing, common evaluation, and pretraining across layout tasks and dimensional settings. For practice, this means that multi-floor layout systems need representations that support not only generation, but also inspection and revision. For future benchmarks, evaluation should move beyond single-floor generation and include vertical connectivity, cross-floor consistency, and topology-preserving edits. With the expected shift toward three-dimensional building layout generation, the next stage of the field will depend in part on how this balance between geometric freedom and explicit structure is resolved.

Conclusions

This paper presented a systematic mapping study of automated building layout planning over more than five decades (1971 to 2025). Using Eastman's General Space Planner as a stable conceptual reference, we coded 59 studies along three facets: representation, objectives, and control, to enable decade-by-decade comparison across both classical and learning-based paradigms. The resulting map provides a longitudinal account of what has remained stable in layout planning, what has diversified, and where bottlenecks arise as learning-based methods become dominant.

Several findings are consistent across the full-time window. Adjacency or connectivity and area or size persist as the most central objectives, suggesting that relational coherence and proportional allocation remain minimal requirements across paradigms. As objective sets expand beyond feasibility toward broader criteria such as performance, cost, and constructability, representation and control choices co-evolve, with studies increasingly adopting richer representational stacks and more complex control strategies.

The map also indicates a recent inflection in the last half-decade, which accounts for more than half of the corpus. Learning-based generation becomes a dominant driver, coinciding with the rise of image-based primary representations and learned loss minimization. While these shifts support greater geometric freedom, they do not eliminate the need for explicit structure: graph-based encodings remain central, often as supporting representations for preserving or recovering topology and enabling feasibility and constraint enforcement. Overall, the field is trending toward representational divergence rather than convergence, unlike domains that benefited from shared conventions for foundation-model training. This divergence becomes more consequential under dimensional expansion, since most studies remain 2D despite the multi-floor nature of real buildings and the need to represent vertical relations. Continued fragmentation across 2D, 2.5D, and 3D will further raise the cost of reuse and comparison, whereas full convergence may improve comparability but risks impractical complexity when attempting to preserve both geometric freedom and explicit topology within a single formalism.

Rather than prescribing a single universal representation for building layouts, this study highlights a recurring

pattern in the literature. The pattern shows that geometric freedom is increasingly pursued through flexible encodings, while explicit topological structure is repeatedly retained. Looking ahead, especially toward three-dimensional and multi-floor layout generation, clearer reporting of representational stacks and their mappings, together with evaluation procedures that remain valid across representational and dimensional regimes, will be important for cumulative progress.

Acknowledgments

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (Grant RS-2025-00522484).

References

Almashaqbeh, M., El-Rayes, K., 2021. Optimizing the modularization of floor plans in modular construction projects. J. Build. Eng. 39, 102316. https://doi.org/10.1016/j.jobe.2021.102316

Dino, I.G., 2016. An evolutionary approach for 3D architectural space layout design exploration. Autom. Constr. 69, 131–150. https://doi.org/10.1016/j.autcon.2016.05.020

Eastman, C.M., 1973. Automated space planning. Artif. Intell. 4, 41–64. https://doi.org/10.1016/0004-3702(73)90008-8

Grason, J., 1971. An approach to computerized space planning using graph theory, in: Proceedings of the 8th Design Automation Workshop, DAC '71. Association for Computing Machinery, New York, NY, USA, pp. 170–178. https://doi.org/10.1145/800158.805070

Hua, H., 2016. Irregular architectural layout synthesis with graphical inputs. Autom. Constr. 72, 388–396. https://doi.org/10.1016/j.autcon.2016.09.009

Jang, S., Borrmann, A., Lee, G., 2025a. Automated Natural Language Building Descriptor For Building Information Models, in: Proceedings of 2025 European Conference on Computing in Construction (EC3).

Jang, S., Roh, H., Lee, G., 2025b. Generative AI in architectural design: Application, data, and evaluation methods. Autom. Constr. 174, 106174. https://doi.org/10.1016/j.autcon.2025.106174

Jo, J.H., Gero, J.S., 1995. A Genetic Search Approach to Space Layout Planning. Archit. Sci. Rev. 38, 37–46. https://doi.org/10.1080/00038628.1995.9696774

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Liggett, R.S., 2000. Automated facilities layout: past, present and future. Autom. Constr. 9, 197–215.

Liggett, R.S., Mitchell, W.J., 1981. Optimal space planning in practice. Comput.-Aided Des. 13, 277–288. https://doi.org/10.1016/0010-4485(81)90317-1

Marson, F., Musse, S.R., 2010. Automatic Real-Time Generation of Floor Plans Based on Squarified Treemaps Algorithm. Int. J. Comput. Games Technol. 2010, 1–10. https://doi.org/10.1155/2010/624817

Medjdoub, B., Yannou, B., 2000. Separating topology and geometry in space planning. Comput.-Aided Des. 32, 39–61. https://doi.org/10.1016/S0010-4485(99)00084-6

Merrell, P., Schkufza, E., Koltun, V., 2010. Computer-generated residential building layouts. ACM Trans. Graph. 29, 1–12. https://doi.org/10.1145/1882261.1866203

Petersen, K., Vakkalanka, S., Kuzniarz, L., 2015. Guidelines for conducting systematic mapping studies in software engineering: An update. Inf. Softw. Technol. 64, 1–18. https://doi.org/10.1016/j.infsof.2015.03.007

Pfefferkorn, C.E., 1972. The Design Problem Solver, A System for Designing Equipment or Furniture Layouts.

Rodrigues, E., Gaspar, A.R., Gomes, Á., 2014. Automated approach for design generation and thermal assessment of alternative floor plans. Energy Build. 81, 170–181. https://doi.org/10.1016/j.enbuild.2014.06.016

Rosenman, M., 2000. Case-based evolutionary design. Artif. Intell. Eng. Des. Anal. Manuf. 14, 17–29. https://doi.org/10.1017/S0890060400141022

Roth, J., Hashimshony, R., Wachman, A., 1982. Turning a graph into a rectangular floor plan. Build. Environ. 17, 163–173. https://doi.org/10.1016/0360-1323(82)90037-3

Shekhawat, K., Upasani, N., Bisht, S., Jain, R.N., 2021. A tool for computer-generated dimensioned floorplans based on given adjacencies. Autom. Constr. 127, 103718. https://doi.org/10.1016/j.autcon.2021.103718

Weber, R.E., Mueller, C., Reinhart, C., 2022. Automated floorplan generation in architectural design: A review of methods and applications. Autom. Constr. 140, 104385. https://doi.org/10.1016/j.autcon.2022.104385

Xu, Z., Jha, N., Mehadi, S., Mandal, M., 2025. Automatic floor plan analysis: Datasets, methods, and applications (2000–2025). Autom. Constr. 178, 106378. https://doi.org/10.1016/j.autcon.2025.106378

Zeng, P., Gao, W., Yin, J., Xu, P., Lu, S., 2024. Residential floor plans: Multi-conditional automatic generation using diffusion models. Autom. Constr. 162, 105374. https://doi.org/10.1016/j.autcon.2024.105374

Zheng, Z., Petzold, F., 2023. Neural-guided room layout generation with bubble diagram constraints. Autom. Constr. 154, 104962. https://doi.org/10.1016/j.autcon.2023.104962

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

A NEURO-SYMBOLIC AI PIPELINE FOR GENERATING SHACL-BASED DIGITAL BUILDING REGULATIONS

Alex Donkers and Ekaterina Petrova
Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract

Automated compliance checking requires the transformation of natural language building regulations into machine-interpretable code. Current Large Language Model (LLM) approaches for generating digital rules often suffer from hallucinations and structural inconsistencies. This paper proposes a neuro-symbolic AI pipeline that grounds LLM reasoning in symbolic logic and domain ontologies to generate SHACL shapes. Evaluation against Dutch construction regulations demonstrates that the pipeline successfully generates functional SHACL shapes for relational, mathematical, and conditional constraints. This work contributes a scalable framework for translating complex legislative code into executable, interoperable digital rule sets for the AECO industry.

Introduction

Digital transformation of the Architecture, Engineering, Construction, and Operations (AECO) industry is increasingly dependent on digitization of building regulations. Traditionally, regulations are authored as natural language documents, which are inherently ambiguous and require manual interpretation by experts. Not only is this process time-consuming and prone to human error, it also limits innovations in BIM-based design workflows. To bridge the gap between human-readable regulations and machine-interpretable code, the research community has explored various ways of digitizing building regulations. One of the emerging ideas is the use of the Shapes Constraint Language (SHACL) to validate Linked Building Data (Nuyts et al., 2024).

However, manual creation of SHACL shapes is a highly specialized task that requires deep knowledge of both semantic web technologies and the specific domain of the regulations, such as fire safety. Recent advancements in Natural Language Processing (NLP), specifically Large Language Models (LLMs), offer a potential solution for automating this process. These LLMs introduce risks when applied to the regulatory domain, including generating terms that do not match with the target ontology, and the appearance of structural inconsistencies in the digital rule representation. Such challenges prevent digital rules from being operationally executable and negatively affect the human trust in these rules.

To overcome these issues, this paper proposes a neuro-symbolic AI pipeline to generate SHACL-based digital building regulations. The aim is to find how symbolic logic representations, such as ontologies and SHACL templates, can be used to enhance reasoning of the LLM. After exploring related work in Section 2, Section 3 explains this neuro-symbolic AI approach in more detail. The approach is tested with Dutch construction regulations in Section 4, after which Section 5 and 6 discuss and conclude this work.

Related work

Digital building codes

Digitization of building codes is a long-standing research topic, and systems to assist in (automated) compliance checking (ACC) have been around for over 50 years (Amor and Dimyadi, 2021). A shared understanding of requirements for such digital building codes is presented by Noardo et al. (2022). These include that normative text should be interpretable by humans and machines, and that we should aim for interoperability through shared dictionaries. Early ideas on the use of Artificial Intelligence (AI) to assist ACC exist since roughly 1990 (Niwa, 1989). Much work in digitization of rules was driven by developments in Building Information Modeling (BIM). In recent years, developments in Natural Language Processing and Large Language Models drove research in rule digitization (Fuchs, 2021).

Ontologies for digital building codes

Ontologies can be used to enhance the interoperability of digital building regulations. As described by Amor and Dimyadi (2021), compliance checking happens on the intersection of building information and normative knowledge, and therefore ontologies representing both sides of the intersection are needed. The research community has increasingly adopted the Linked Building Data (LBD) concept, and the W3C LBD community group presented ontologies that can represent topological building information (Building Topology Ontology, BOT, by Rasmussen et al. (2021)), building elements (Built Element Ontology, BEO, by Ramonell Cazador (2026)), properties (Ontology for Property Management, by Rasmussen et al. (2018)), and geometry (Ontology for Managing Geometry, OMG, Wagner et al. (2019)). Next to these general building ontologies, a wide variety of domain ontologies exists that enable modeling specific domain knowledge. Ontologies representing (AECO-specific) normative knowledge are

more scarce (Werbrouck et al., 2026).

NLP and the generation of digital building codes

The transformation of natural language regulations into machine-interpretable formats has progressed from manual rule interpretation to computer-aided interpretation. A clear example is the semantic tagging of rules with the RASE (Requirement, Applicability, Selection, Exception) method, that moves from manual tagging to human-in-the-loop LLM-based tagging (Al-Turki et al., 2024). The transition is driven by using NLP and domain ontologies to map regulatory terms to semantic concepts, frequently targeting IFC-based compliance (Zhang and El-Gohary, 2016). Subsequent efforts expanded these NLP techniques to extract spatial terms for reasoning (Li et al., 2016) and used models like BERT to transform fire safety regulations into SPARQL queries via semantic labeling and keyword matching (Zheng et al., 2022). Recent research in rule interpretation highly focused on the application of LLMs. Fuchs et al. (2024) applied ChatGPT to transform natural language to LegalRuleML, while Chiappini et al. used ChatGPT to convert text to SPARQL. Currently, the field is moving toward neuro-symbolic AI, which combines neural AI approaches with formal symbolic logic. While early work relied on structured spreadsheets (Costa et al., 2024), current work uses ontologies to improve the reasoning of LLMs (Donkers and Petrova, 2025).

Methodology

System architecture

Figure 1 shows the system architecture for the SHACL converter. It contains a column with input data (the regulatory statement, a domain ontology, W3C LBD ontologies, and SHACL shape instructions), a middle block that prepares input for the LLM (represented by the colored arrows), and the last column with the LLM functionalities. The following subsections discuss the internals of each block in detail.

Linguistic pre-processing

The process begins with the extraction of features from a regulatory statement. The spaCy NLP framework is used to identify noun chunks, which are base phrases with a noun as their head. These noun chunks are later used by the LLM to correctly map terms in the regulations to terms in the ontology. Using noun chunks instead of single tokens in the sentence overcomes incorrect mapping of nouns to terms in the ontology. For example, "A portable fire extinguisher" has universal part-of-speech (POS) tags `DET-ADJ-NOUN-NOUN`, however, "fire" and "extinguisher" should not be mapped to ontology terms in isolation, as "fire" is a compound modifier of "extinguisher". Individual tokens are also extracted, where specific POS tags are excluded to not clutter the ontology term selection. The `nl_core_news_lg` model is used, given that this paper converts Dutch regulations.

Ontologies

To prevent the LLM from hallucinating terms, domain ontologies are used that contain classes that should be used by the LLM. This paper makes use of the Fire-BIM Building Ontology, which is an extensive collection of terms related to the fire safety domain. To enable bridging from these terms to the domain of building information modeling, a subset of the W3C Linked Building Data ontologies (Building Topology Ontology (BOT), Built Element Ontology (BEO), Ontology for Property Management (OPM), Ontology for Geometry Management (OMG)) is used in this work. Listing 1 shows how these ontologies could be linked to make linked building data statements about fire compartments and the spaces they contain.

```
inst:FireCompartment_1 a fbo:FireCompartment ;
    bot:hasZone inst:Space_1, inst:Space_2 ;
    omg:hasGeometry inst:Geometry_1 .
inst:Space_1 a fbo:Space .
inst:Space_2 a fbo:Space .
```

Listing 1: RDF snippet of linking a fire compartment to the spaces it contains

Optimizing ontology representation for LLMs - Encoding rdfs:label and rdfs:comment into vectors - Cosine similarity to measure the distance between the noun chunks and the embedded ontology terms - Selection: n=30

One of the major limitations of LLMs is the amount of tokens that can be used in the prompt, as larger token numbers are more costly and tend to worsen the LLM results. Token counts of the used ontologies (retrieved programmatically using the `google-genai` SDK's native `count_tokens` method to ensure alignment with the model's tokenizer) are presented in Table 1. Striking are the high token counts of BOT (caused by long rdfs:comments in multiple languages) and FBO and BEO (caused by an extensive amount of classes).

Table 1: Ontology token count in Gemini's tokenizer

Ontology	Tokens
fbo.ttl	96.384
bot.ttl	32.042
opm.ttl	4.207
omg.ttl	4.024
beo.ttl	51.900
Total	**188.557**

To limit the high token counts, two methods are applied. For ontologies with `rdfs:comments` in multiple languages, all terms are queried using `FILTER(LANG(?comment)='nl'))` to only select the Dutch comments. For ontologies with a large number of classes and properties, a cosine similarity algorithm is ran between the terms and noun chunks selected during the linguistic pre-processing, and the `rdfs:labels` and

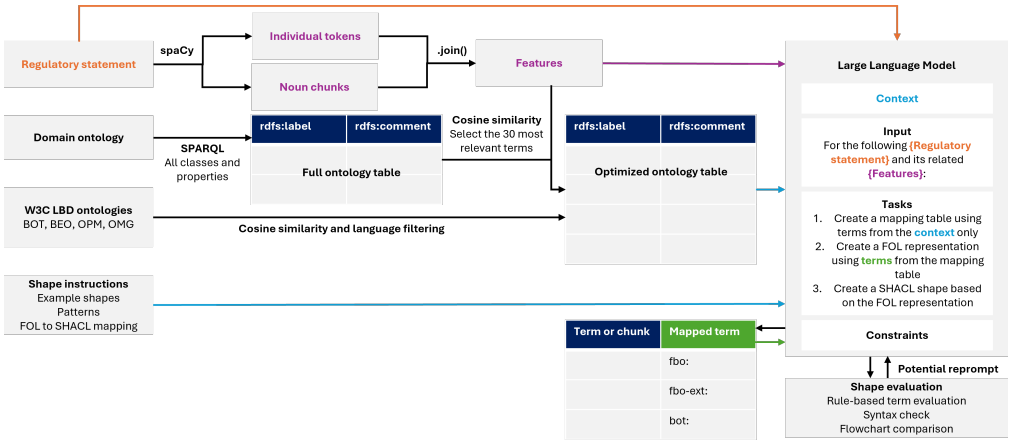


Figure 1: A neuro-symbolic pipeline for generating SHACL shapes

`rdfs:comments` of all terms in the ontology. First, both the regulatory terms and the ontology terms are embedded using the `sentence-transformers/all-mpnet-base-v2` model with Mean Pooling. Then, `sklearn.metrics.pairwise.cosine_similarity` is used to compute the cosine similarity of each entity in the regulatory terms list with each term from the ontologies. For each regulatory term, the top 10 terms are selected, after which these terms are combined and filtered for duplicates. "10" is a rather arbitrary number, that should be evaluated case-by-case, depending on the complexity of regulations and the ontology.

Digital representations of regulations

While SHACL has shown to be able to digitally represent regulations (Donkers and Petrova, 2024), it is a constraint language, not a logic language. Laws are naturally structured as logical implications, where:

$$\forall x(\text{Antecedent}(x) \rightarrow \text{Consequent}(x))$$

First-Order Logic mapping This work aims to preserve this natural structure of the laws (isomorphic modeling) using First-Order Logic (FOL) mapping. Some benefits are that 1) human experts are able to verify the mapping, 2) FOL is implementation agnostic and could be used to generate digital building codes in other formats, and 3) FOL explicitly splits the antecedent and the consequent. The latter if of great importance, as LLMs tend to mix these up when generating SHACL in a zero-shot prompting architecture. The FOL antecedent and consequent can be mapped to other rule markup languages, such as SHACL (`sh:SPARQLTarget` and `sh:PropertyShape`), RASE (Applicability/Selection and Requirement/Exception), IDS (Applicability and Requirements), SWRL (Rule body and Rule head), N3, (LHS and RHS).

There is room for discussion on how to model exceptions in FOL, as one could either model an exception as a negation $(A \wedge \neg E) \rightarrow R$ (making it part of the antecedent) or as a disjunction to the requirement $A \rightarrow (R \vee E)$ (making it part of the consequent). In this paper, we assume exceptions to be modeled using the latter method, as this relates best to how exceptions are written in Dutch regulations, and thus how a rule interpreter should interpret the rule.

SHACL Patterns While FOL serves as an intermediate representation to ensure logical isomorphism with the legal text, SHACL enables making the digital rule operationable. SHACL shapes can be used to validate Linked Building Data graphs for compliance with regulations, and provide means to generate a validation report once the data graph violates the rules dictated in the shapes graph.

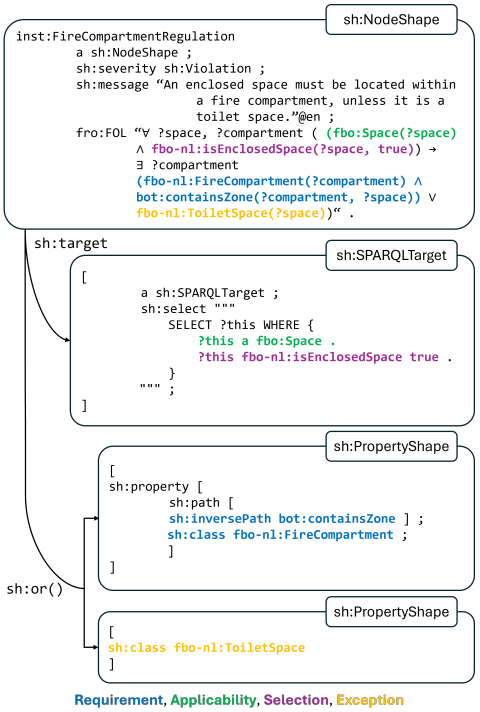


Figure 2: FOL & SHACL representation of a regulation with RASE tags

Figure 2 shows how a text-based regulation can be repersented in FOL, and how that FOL would logically convert to a SHACL shape, where the antecedent is represented in a `sh:SPARQLTarget` and the consequent in

sh:PropertyShapes. The figure also shows how this shape would map to RASE.

While SHACL is a W3C standard, it has a lot of structural ambiguity, meaning rules can be represented using different SHACL representations. These representations do cause SHACL validation engines to behave differently, e.g. by changing how exceptions are modeled, or by different use of logical operaters. An LLM without guidance typically hallucinates a hybrid syntax, making it difficult to evaluate the generated SHACL shape and to predict its constraint behavior. To guide the LLM in generating predictable shapes, a set of standardized SHACL shapes and patterns are prepared. These templates serve to constrain the LLMs output to specific modeling idioms, such as property constraints, inverse relationships, cardinality constraints, topological containment, and logical operators when modeling multiple requirements. This eliminates the structural inconsistencies that occur in zero-shot SHACL generation.

Large language model prompt strategy

A Large Language Model (gemini-2.5-flash) was used via Google's genai Python library. The study applies a structured prompting strategy following In-Context Learning and Constrained Decoding principles. The LLM is provided with explicit Inductive Bias through predefined SHACL templates, minimizing the structural inconsistency of the SHACL generation. Model configurations include temperature=0.0 and max_output_tokens=2048. The prompt contains of 1) Semantic context, 2) Input data, 3) Tasks, 4) Constraints, 5) SHACL generation rules. Those individual items are discussed in the following paragraphs. Listing 2 shows the prompt that was used in Python.

```
You are a Specialist in Semantic Web and Building Regulations.

### SEMANTIC CONTEXT
- Domain ontology: {fbo_str}
- Building ontologies: {w3c_str}
- SHACL templates: {shacl_templates}

### INPUT DATA
Regulatory Text: {regulatory_text}
Extracted Features: {', '.join(features)}

### TASKS
1. MAPPING TABLE: Link statement concepts to context URIs.
2. FIRST-ORDER LOGIC (FOL): Formalize the requirement logic.
3. SHACL SHAPE: Generate a Turtle-formatted constraint shape.

### CONSTRAINTS
- Only use URIs from the provided context.
- If you cannot map a term to the terms from the ontologies in the context, create a new term in the fbo-ext: namespace
- For properties, check directionality. bot:containsZone means that X contains Y, not Y contains X. bot:containsElement means that X contains Y, not Y contains X.
- Output the SHACL shape within code blocks.
- Classes always start with Uppercase, properties lowercase.
- In the FOL representation, preferably use the terms from the mapping table with correct prefixes.

### SHACL generation
- the →'' of the FOL logically splits the FOL into 'applicability' (left of →'') and 'requirements' (right of →''). 'applicability' should be modeled as sh:SPARQLtarget.
- The Rule that needs to be checked (REQUIREMENT) is always part of the consequent, and thus right of the →'', and thus should be in the property node in SHACL.
```

Listing 2: The LLM prompt in Python

Context injection The first part of the prompt involves semantic priming, where the LLM is provided with a domain ontology, a set of building-related W3C LBD ontologies, and SHACL templates. The ontologies are loaded as string-based representations instead of the original Turtle files, based on the optimization method in Section 'Optimizing ontology representation for LLMs'. This significantly reduces the token count of ontologies, e.g. this method optimizes the BOT tokens from 32042 to 2107.

After injecting the semantic context, the input data for the LLM in inserted, being the regulatory text as a string, and a list with features. This feature list is a combination of the noun chunks and the individual tokens selected by the NLP algorithm.

Multi-stage prompting The context is followed by a set of tasks following a multi-stage prompting method, to enable managing the high cognitive load associated with mapping context logic to SHACL. The tasks use a Chain-of-Thought reasoning to decouple the generation of a mapping table, the generation of an FOL representation, and the generation of a SHACL shape. By first mapping features to terms from the ontology in a mapping table, the prompt aims to reduce the number of hallucinated terms. The generation of FOL is then using the mapping table, reusing the mapped terms from the multiple ontologies. The FOL is then converted to a SHACL shape, using the constraints and templates loaded in the prompt. The mapping table, the FOL, and the SHACL shape are all returned by the LLM, enabling human validation.

Evaluation

The generated SHACL shapes are finally evaluated using automated and human-in-the-loop methods. A RegEx-based Python function evaluates the existance of certain syntactical elements in the SHACL shape, such as sh:SPARQLTarget and sh:NodeShape. It also evaluates the existence of the domain ontology prefixes. Secondly, a set of human-in-the-loop evaluation methods were developed. First, human experts can select a list of terms that should be in the returned SHACL shape. A RegEx search looks for the term, allowing for prefixed and unprefixed terms. If not available, the LLM is reprompted, with the previously generated SHACL shape and an error message stating that the term is missing. Secondly, a rule-based method to convert SHACL shapes to the Mermaid

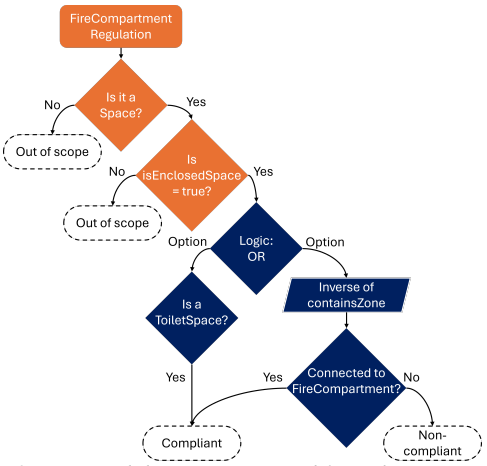


Figure 3: Mermaid diagram generated from the SHACL shape in Figure 2

diagramming language is developed, which enables presenting the SHACL shape as a human-readable flowchart. As Mermaid is code-based, it is possible to include the URIs of terms in the decision tree, making it interactive. This way, a human expert can click on the diagram to open definitions of terms. The antecedent and consequent are colored orange and blue, respectively, to visually help the expert. Finally, two data graphs are modeled for each SHACL shape; one deliberately violating the shape, and one that should pass the check. Since the shapes make use of `sh:SPARQLTarget`, which is a SHACL Advanced Feature, not all SHACL validators are able to run these shapes. The Python-based `pySHACL` engine supports SHACL Advanced Features, which is why an evaluation script is built using this library.

Dataset description To evaluate the developed method, regulations from the Dutch construction law (*'Besluit bouwwerken leefomgeving'*) were converted to SHACL shapes. Steutel (2026) developed a method to convert Dutch fire safety regulations to SHACL. The same set of regulations is manually filtered for regulations that are described by the domain ontology, after which that subset is used in this work. The regulations are categorized using the same categories proposed by Nuyts et al. (2024) to get a sense of the strengths and weaknesses of this approach for each category.

Results

The converter is tested on articles or combinations of articles, as presented in 2. Categories are numbered as follows: 1. information availability, 2. value, 3. relational, 4. mathematical, 5. conditional.

All categories described by Nuyts et al. (2024) are tested and for each of the categories, the converter managed to generate a functional SHACL shape. To test category 1, an Information Requirements rule was modeled to check whether Abox building information was correctly modeled to be checked by rule 4.50(2d). Listing 3 shows the generated SHACL shape. The generator correctly distinguished the antecedent logic (placed in the `sh:SPARQLTarget` and the consequent logic (placed in the `sh:PropertyShape`). While `sh:SPARQLTarget` is not strictly necessary for this specific rule (`sh:target fbo-nl:TechnicalRoom` would suffice), the goal is to generate a set of SHACL shapes that has a stable structure across the set, and future shapes will likely need this modeling method.

Table 2: Article Mapping Results

Article	Category	Mapping	SHACL
4.50(1)	3	●	●
4.50(1)+(2a)	3, 5	●	●
4.50(1)+(2a)+(2b)+(2d)	2, 3, 5	●	●
4.50(1)+4.50(4)	3, 5	●	●
4.50(5)	3	●	●
4.50(1)+4.50(5)+4.50(8)	3, 5	●	○
4.51(2)	3, 4	●	●
4.50(2d) IR	1	●	●

```
fbo-nl:TechnicalRoomShape a sh:NodeShape ;
    sh:target [
        a sh:SPARQLTarget ;
        sh:select """
            SELECT ?this WHERE {
                ?this a fbo-nl:TechnicalRoom .
            }
        """ ;
    ] ;
    sh:property [
        sh:path fbo:hasUsableArea ;
        sh:message "A technical room must have a usable area."@en ;
        sh:minCount 1 ;
    ] .
```

Listing 3: An information requirement rule checking whether Abox building information is suitable for a compliance check

Almost all evaluated articles are categorized as category 3. relational. The SHACL shape generated from article 4.50(1) is shown in Listing 4. Noticable is the different URI representation of the `sh:NodeShape`, which is something that the LLM is not trained for. Listing 4 shows how the LLM correctly interpreted the inverse logic of `bot:containsZone`. The use of `sh:qualifiedValueShape` is somewhat redundant, but does influence the results of the compliance check.

```
<http://www.firebim.org/shapes/enclosed-space-in-fire-compartment> a sh:NodeShape ;
    sh:target [
        a sh:SPARQLTarget ;
        sh:select """"""
            SELECT ?enclosedSpace WHERE {
                ?enclosedSpace a bot:Space .
                ?enclosedSpace fbo-nl:isEnclosedSpace true .
            }
        """""" ;
    ] ;
    sh:property [
        sh:path [ sh:inversePath bot:containsZone ] ;
        sh:class fbo-nl:FireCompartment ;
        sh:qualifiedValueShape [
            a sh:NodeShape ;
            sh:targetClass fbo-nl:FireCompartment ;
```

```
        ] ;
        sh:message ""An enclosed space must be
     contained within a fire compartment."" ;
    ] ."
```

Listing 4: A generated SHACL shape for a relational regulation

Category 2 represents regulations with value checks. Listing 5 shows how the maximum usable area of a technical room and a minimum nominal power of combustion appliances in that technical room are modeled in SHACL by the converter. The units of these values are not yet modeled by the LLM; these are already given in the ontology and can be checked by a separate information requirements check. However, the converter could in the future be extended to include units, e.g. using an extra SHACL template.

```
[ a sh:NodeShape ;
    sh:class fbo-nl:TechnicalRoom ;
    sh:and (
        [ sh:property [ sh:path fbo:hasUsableArea ; sh
     :datatype xsd:decimal ; sh:maxInclusive "50"^^xsd:
     decimal ] ]
        [ sh:not ( [ a sh:NodeShape ;
            sh:property [ sh:path fbo-nl:
     hasCombustionAppliance ; sh:class fbo-nl:
     CombustionAppliance ;
                sh:node [ sh:property [ sh:path fbo-nl
     :hasNominalPower ; sh:datatype xsd:decimal ; sh:
     minInclusive "130"^^xsd:decimal ] ]
            ] ] )
        ]
    )
] .
```

Listing 5: A generated SHACL shape for a value check

Mathematical regulations (category 4) are represented by article 4.51(2). Listing 6 shows 2 of such mathematical functions. The first half of the listing presents a `sh:PropertyShape` that checks whether a fire compartment contains maximum four caravans or ancillary functions. The listing shows that the LLM modeled ancillary functions in the `fbo-ext:` namespace, following the instructions in Listing 2. This is an indication to the ontology engineer that the current domain ontology does not suffice in representing this regulation. The second part of the listing checks whether the total area of these caravans does not exceed 1000 m^2, for which it uses a `sh:SPARQLConstraint` that first queries the area of each individual caravan, then sums these areas, and finally checks whether these areas don't exceed 1000 m^2.

```
fbo-ext:
     FireCompartmentMaxWagonsAndAncillaryFunctionsShape
      a sh:NodeShape ;
  sh:targetClass fbo-nl:FireCompartment ;
  sh:property [
    sh:path bot:containsZone ;
    sh:qualifiedValueShape [
      a sh:NodeShape ;
      sh:or (
        [ sh:class fbo-nl:ResidentialFunctionCaravan ]
        [ sh:class fbo-ext:AncillaryFunction ]
      ) ;
    ] ;
    sh:qualifiedMaxCount 4 ;
    sh:message "Max 4 caravan or ancillary functions."
     @en ;
  ] ;

  sh:sparql [
    a sh:SPARQLConstraint ;
    sh:message "Total area must not exceed 1000 m2."
     @en ;
    sh:select """
      SELECT ?this (SUM(?area) AS ?totalArea)
      WHERE {
        ?this bot:containsZone ?zone .
        ?zone a ?zoneType ; fbo:hasUsableArea ?area .
        FILTER (?zoneType IN (fbo-nl:
     ResidentialFunctionCaravan,
          fbo-ext:AncillaryFunction))
      }
      GROUP BY ?this HAVING (SUM(?area) > 1000)
    """ ;
  ] .
```

Listing 6: A generated SHACL shape for mathematical regulations

Finally, conditional constraints (category 5) can be modeled using the logical operators in SHACL. Figure 2 shows how multiple conditions (in this case exceptions) are modeled by the LLM.

Article 4.50(1)+4.50(5)+4.50(8) did not correctly convert to SHACL. Listing 7 shows the input to the converter. While the mapping from features to terms in the ontology is correctly executed, the generated SHACL shape shows mistakes. A likely cause is the fact that the exception in statement 8 relates to the function of a building, whereas the requirements in statements 1 and 5 relate to a space. Although the LLM generated a syntactically correct SHACL shape, it assumed that `fbo-nl:LightIndustrialFunctionForCultivation` is a class that can be assigned to a space, while this is meant to be a function type of a building. It will therefore in practice not match with Abox building information.

```
1. An enclosed space is located in a fire compartment.
5. A non-enclosed functional area is located in a fire
     compartment.
8. The first and fifth statement do not apply to a
    light industrial function for the cultivation,
    growing, or storage of crops or comparable
    products, with a permanent fire load no greater
    than 150 MJ/m2, determined in accordance with NEN
    6090.
```

Listing 7: A translation of article 4.50(1)+4.50(5)+4.50(8) of the Dutch construction law

```
[ sh:and (
  [ sh:property [
    sh:path rdf:type ;
    sh:hasValue fbo-nl:
     LightIndustrialFunctionForCultivation ]
  ]
  [ sh:property [
    sh:path fbo-nl:hasPermanentFireLoad ;
    sh:maxInclusive 150 ]
  ]
) ]
```

Listing 8: The exception logic of 4.50(8)

Nested exceptions The converter is designed to specify complex exceptions as separate `sh:Nodeshapes`. Listing 5 shows how a complex exception (article 4.50(2d))to the regulation in Figure 2 is modeled by the converter. This rather atomic modeling principle allows a more flexible use of SHACL shapes, allowing end users to query and apply such complex exceptions separately (Donkers and Petrova, 2025), or allowing other SHACL shapes to reuse such exceptions.

Human-in-the-loop validation The generated Mermaid diagrams for each SHACL shape enable human-in-the-loop validation. We validated the generated SHACL shapes using these Mermaid diagrams. Although this method enables non-SHACL-experts to validate SHACL shapes, such methods should still be used with caution. For example, the incorrect use of light industrial functions for cultivation in Listing 8 did result in a syntactically correct SHACL shape, and thus formed a Mermaid diagram. While this Mermaid diagram is arguably easier to read than the SHACL shape, such minor errors are still hard to spot by the human eye.

Validating a data graph To test the functionality of the SHACL shapes in practical use cases, data graphs were created for each SHACL shape, and they were subsequently tested in a `pySHACL`-based script. Listing 9 shows the data graph that was modeled to violate the SHACL shape of article 4.51(2) (which is shown in Listing 6). It violates the rule twice: 1. the fire compartment contains too many caravans or ancillary functions, and 2. the total area of these functions is larger than 1000 m^2.

```
inst:InvalidCompartment a fbo-nl:FireCompartment ;
    bot:containsZone inst:C1, inst:C2, inst:C3, inst:
     C4, inst:A1 .
inst:C1 a fbo-nl:ResidentialFunctionCaravan ;
    fbo:hasUsableArea 250 .
inst:C2 a fbo-nl:ResidentialFunctionCaravan ;
    fbo:hasUsableArea 250 .
inst:C3 a fbo-nl:ResidentialFunctionCaravan ;
    fbo:hasUsableArea 250 .
inst:C4 a fbo-nl:ResidentialFunctionCaravan ;
    fbo:hasUsableArea 250 .
inst:A1 a fbo-ext:AncillaryFunction ;
    fbo:hasUsableArea 200 .
```

Listing 9: Data graph that violates article 4.51(2)

The validation report by pySHACL is printed in Listing 10, and correctly shows the two violations of Article 4.51(2).

```
Conforms: False
Results (2):
Constraint Violation in
     QualifiedValueShapeConstraintComponent:
  Severity: sh:Violation
  Source Shape: *The max count sh:PropertyShape*
  Focus Node: ex:InvalidCompartment
  Result Path: bot:containsZone
  Message: A fire compartment may contain a maximum of
      4 ResidentialFunctionCaravan or AncillaryFunction
     .
Constraint Violation in SPARQLConstraintComponent:
  Severity: sh:Violation
  Source Shape: fbo-ext:
     FireCompartmentMaxWagonsAndAncillaryFunctionsShape
  Focus Node: ex:InvalidCompartment
  Value Node: ex:InvalidCompartment
  Source Constraint: *The sh:SPARQLConstraint*
  Message: The total usable area of
     ResidentialFunctionCaravan and AncillaryFunction
     within a fire compartment must not exceed 1000 m2.
```

Listing 10: Validation report generated by pySHACL

Discussion

The results demonstrate that grounding LLMs in domain-specific ontologies via a neuro-symbolic AI pipeline enables the generation of SHACL shapes out of natural language regulations. However, a few critical challenges remain unanswered.

The current injection of ontology terms uses a table with the `rdfs:label` and `rdfs:comment`. This loses part of the context of each term, e.g. how classes are associated to other classes, and with what properties. This knowledge is normally strongly represented in ontologies, and future work should answer whether injecting these relationships into the context of the LLM improves the results.

The converter takes individual regulatory statements, or manually inserted combinations of statements. It cannot deal with external sources or references to other articles unless they are explicitly included in the regulatory statement variable. Combining this neuro-symbolic pipeline with the regulation graph as presented by Donkers and Petrova (2025) would enable retrieving regulatory statements from a structured RDF database that helps navigating to the correct articles, tables, figures, and so forth.

While the token optimization method succesfully reduced the number of tokens of each ontology, the n=10 retrieval parameter for ontology terms could cause difficulties for complex regulations. More research is necessary to determine what ontology optimization strategy fits to different scenarios. The converter showed difficulties when converting regulations with a large number of requirements and exceptions, especially with multiple levels of nesting.

The choice of SHACL as a regulatory language poses difficulties in practice. Most SHACL validators traverse all parts of a `sh:and`-statement, even if the first statement fails, causing computational overhead. The nested SHACL approach also has difficulties with 'unknowns': SHACL is a binary that assumes that values that are not in the data graph do not exist. Especially in early design stages, some exceptions in the regulations might be unknown (e.g. whether a building has a sprinkler system), and SHACL shapes natively encounter difficulties with distinguishing between violations and data incompleteness. How to solve such practical challenges remains future research, and results of such research should then be reflected in this pipeline.

Conclusion

This paper presented a neuro-symbolic AI pipeline that leverages the strengths of Large Language Models grounded in symbolic logic and domain ontologies to automate the generation of SHACL-based digital building regulations. By using First-Order Logic as an intermediate representation and applying a structured prompting strategy with predefined SHACL templates, the approach significantly reduces the structural inconsistencies and hallucinations typically associated with zero-shot LLM outputs. The evaluation against Dutch construction regulations demonstrated that the pipeline can effectively handle various rule categories, including relational, mathematical, and conditional constraints, while providing human-interpretable validation via Mermaid diagrams.

However, the transition from individual regulatory state-

ments to a fully automated compliance ecosystem requires addressing several complexities. Future research will focus on integrating this pipeline with comprehensive regulation graphs. While the current method processes isolated articles, a graph-based approach would allow the system to navigate and resolve cross-references between different articles, annexes, and external guidelines. Furthermore, extending neuro-symbolic reasoning to interpret non-textual regulatory elements such as tables, formulas, and diagrams is essential for a holistic digitization process. By linking the generative power of LLMs with structured regulation databases, we can move toward a scalable framework capable of translating vast bodies of legislative code into executable, interoperable, and trustworthy digital rule sets for the AECO industry.

Acknowledgements

This paper is funded by the Eureka ITEA4 FireBIM project (22003) and the Netherlands Enterprise Agency.

References

Al-Turki, D., Hettiarachchi, H., Medhat Gaber, M., Abdelsamea, M. M., Basurra, S., Iranmanesh, S., Saadany, H., and Vakaj, E. (2024). Human-in-the-loop learning with llms for efficient rase tagging in building compliance regulations. IEEE Access, 12:185291–185306.

Amor, R. and Dimyadi, J. (2021). The promise of automated compliance checking. Developments in the Built Environment, 5:100039.

Costa, G., Vakaj, E., Beach, T., Lavikka, R., Lefrançois, M., Zimmermann, A., Mecharnia, T., Alexiev, V., Dridi, A., Hettiarachchi, H., and Keberle, N. (2024). Formalization of building codes and regulations in knowledge graphs. In Noardo, F. and Fauth, J., editors, Digital Building Permit Conference 2024, pages 142–146, Barcelona, Spain.

Donkers, A. and Petrova, E. (2024). Converting fire safety regulations to shacl shapes using natural language processing. In NLP4KGC 2024 : Natural Language Processing for Knowledge Graph Creation 2024, CEUR Workshop Proceedings. CEUR-WS.org.

Donkers, A. and Petrova, E. (2025). Retrieval of digital regulations using neuro-symbolic ai with graphrag. In Proceedings of the Digital Building Permit Conference 2025, Singapore. Springer Nature.

Fuchs, S. (2021). Natural language processing for building code interpretation: systematic literature review report. August, 21:2023.

Fuchs, S., Witbrock, M., Dimyadi, J., and Amor, R. (2024). Using large language models for the interpretation of building regulations.

Li, S., Cai, H., and Kamat, V. R. (2016). Integrating natural language processing and spatial reasoning for utility compliance checking. Journal of Construction Engineering and Management, 142(12):04016074.

Niwa, K. (1989). Use of artificial intelligence confirming compliance to building codes during process of making architectural drawings. Future Gener. Comput. Syst., 5(1):129–130.

Noardo, F., Guler, D., Fauth, J., Malacarne, G., Mastrolembo Ventura, S., Azenha, M., Olsson, P.-O., and Senger, L. (2022). Unveiling the actual progress of digital building permit: Getting awareness through a critical state of the art review. Building and Environment, 213:108854.

Nuyts, E., Bonduel, M., and Verstraeten, R. (2024). Comparative analysis of approaches for automated compliance checking of construction data. Advanced Engineering Informatics, 60:102443.

Ramonell Cazador, C. (2026). Built Element Ontology (BEO). `https://w3id.org/beo\#`. Revision 1.0.

Rasmussen, M., Lefrançois, M., Schneider, G., and Pauwels, P. (2021). Bot: The building topology ontology of the w3c linked building data group. Semantic Web, 12(1):143–161.

Rasmussen, M. H., Lefrançois, M., Bonduel, M., Hviid, C. A., and Karlshøj, J. (2018). OPM: An ontology for describing properties that evolve over time. In Proceedings of the 6th Linked Data in Architecture and Construction Workshop, London, United Kingdom.

Steutel, B. T. R. (2026). A framework for fire safety: Processing regulatory texts for automated compliance checking using natural language processing and semantic web techniques. Master's thesis, Eindhoven University of Technology.

Wagner, A., Bonduel, M., Pauwels, P., and Uwe, R. (2019). Relating geometry descriptions to its derivatives on the web. In Proceedings of the 2019 European Conference for Computing in Construction, pages 304–313. European Council on Computing in Construction (EC3).

Werbrouck, J., Hagedorn, P., Donkers, A., and Fauth, J. (2026). Semantic Web Technologies for Building Permitting.

Zhang, J. and El-Gohary, N. M. (2016). Semantic nlp-based information extraction from construction regulatory documents for automated compliance checking. Journal of Computing in Civil Engineering, 30(2):04015014.

Zheng, Z., Zhou, Y.-C., Lu, X.-Z., and Lin, J.-R. (2022). Knowledge-informed semantic alignment and rule interpretation for automated compliance checking. Automation in Construction, 142:104524.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

An LLM-Driven Agent-Based Framework for Cross-Regional Construction Policy Evaluation in the Greater Bay Area

Wentao Zhu, Xiao Li, Juan Wang, and Mingtian Xiong
The University of Hong Kong, Hong Kong SAR, China

Abstract

Cross-jurisdiction construction policies in the Guangdong–Hong Kong–Macao Greater Bay Area (GBA) aim to reduce institutional barriers and promote collaboration, yet their impacts are difficult to evaluate ex ante. This paper proposes an LLM-driven agent-based simulation framework integrating a stage-gated project lifecycle, heterogeneous stakeholder agents, and policy-parameterized institutional rules. Policy provisions are encoded into eligibility, cost–risk, and transition mechanisms, enabling traceable policy-to-outcome analysis. Validation is conducted through stylized-facts consistency, reproducing concentration growth and distance-decay collaboration patterns. The framework provides a reproducible testbed for analyzing cross-regional construction policies and their effects on collaboration networks and industry dynamics.

Introduction

The Guangdong–Hong Kong–Macao Greater Bay Area (GBA) is a "9+2" city-cluster initiative intended to strengthen regional connectivity and coordination. In construction, collaboration is increasingly cross-regional but remains shaped by differentiated city functions and institutional arrangements (Hui et al., 2020). To reduce these frictions, Mainland GBA jurisdictions have introduced record-filing and mutual-recognition arrangements, including pilot measures in Guangdong, Qianhai, and Hengqin (Guangdong Provincial Department of Housing and Urban-Rural Development, 2020). A central policy question is whether such measures can materially expand cross-border collaboration and reshape industry outcomes.

Ex ante assessment remains difficult. Macro approaches such as CGE and system dynamics provide transparent aggregate policy mappings, but are less suited to tracing heterogeneous behavior, repeated interaction, and institutionally differentiated responses (Shi et al., 2019; Shin et al., 2014). Agent-based modeling (ABM) addresses this limitation by representing heterogeneous actors and emergent system outcomes (Bonabeau, 2002). In construction, ABM has been used to study supply chains, labor productivity, and safety interactions (Min and Björnsson, 2008; Watkins et al., 2009; Zhang et al., 2019). However, most construction ABM still relies on manually specified heuristics or thresholds (Khodabandelu and Park, 2021). This becomes restrictive when decisions depend jointly on project stage, workload, partner history, policy friction, and compliance requirements, because rule sets grow quickly and remain brittle under newly combined conditions (Li et al., 2024; Wijermans et al., 2023).

Large language models (LLMs) provide a practical way to enrich ABM with more flexible and context-sensitive decision-making. Prior work shows that LLM agents can operate within structured environments under external constraints (Li et al., 2024; Dang et al., 2025), while memory-based LLM agents can support history-dependent adaptation (Park et al., 2023). This is useful in cross-jurisdiction construction settings, where firms must respond simultaneously to institutional heterogeneity, prior interactions, and changing project conditions. Table 1 summarizes the contrast among alternative approaches.

Two practical gaps remain. First, construction projects are process- and constraint-intensive: actions must satisfy stage gates, role dependencies, and compliance chains, so purely dialogue-oriented agents may not yield executable trajectories. Second, cross-jurisdiction policies act through computable mechanisms—entry friction, compliance burden, and scope constraints—so without explicit parameterization, simulations cannot trace effects from policy provisions to behavior and then to network and performance outcomes.

To address these gaps, we propose an LLM-driven ABM framework for evaluating cross-regional collaboration policies in the GBA construction sector. The framework combines an executable four-stage project lifecycle (INIT → SCHEME → AWARD → DELIVER), seven stakeholder roles operating through stage-feasible action sets, and a policy parameter vector embedded into gates and waiting times, cost–risk functions, and event/transition rules.

For validity, we adopt a stylized-facts consistency strategy. Rather than matching exact values, we test whether the simulation reproduces two robust empirical regularities of construction markets: (i) concentration with head-firm dominance and (ii) distance decay with local bias in cross-city collaboration. We then use the platform for counterfactual analysis of network stratification, diffusion paths, and directional asymmetries under alternative policy scenarios.

Methodology

Framework overview

We develop a micro–macro coupled agent-based simulation framework to characterize cross-regional collabora-

Table 1: Comparison of alternative simulation approaches for policy modeling

Model	Micro-level process-aware behavior	Policy heterogeneity	History / memory adaptation	Flexible under new contexts	Low manual rule burden
CGE / System Dynamics	✗	△	✗	✗	✓
Rule-based Agents	✓	△	△	✗	✗
LLM-driven Agents	✓	✓	✓	✓	△

Note: ✓ indicates strong support for the corresponding capability; △ indicates partial support, usually requiring simplified representations or additional manual assumptions; ✗ indicates that the capability is weakly supported or not explicitly represented.

Table 2: Fixed four-stage project lifecycle

Stage (s)	Purpose	Actions	Hard gates $g_s(\mathrm{x})$	Main outputs O_s
INIT	Formalize constraints and form the minimal team	OWN sets budget_ceiling; and schedule_target/risk_preference; QS derives cost_ceiling and cost_structure; ARC produces functional_brief; ENG produces feasibility_precheck.	brief_ready ∧ cost_ceiling_set ∧ team_ready (ARC,QS,ENG)	team_roster; budget_ceiling; functional_brief; risk_precheck.
SCHEME	Freeze scheme and close the cost loop	ARC generates scheme_package; ENG provides system_strategy; SUP provides price_band; and lead_time/availability; QS produces scheme_estimate; and budget_gap; OWN executes freeze_or_iterate.	scheme_frozen ∧ estimate ≤ budget_ceiling	frozen_scheme; tender_boundary; scheme_estimate; material_risk_signal.
AWARD	Mobilization gates: contract, safety, procurement	QS prepares BQ and contract_terms; CS submits bid (incl. risk premium); OWN awards and signs; SAFE outputs safety_plan; SUP confirms critical procurement terms.	contract_signed ∧ safety_plan_ready ∧ critical_supply_secured	signed_contract; safety_plan; procurement_terms; pre_start_delay.
DELIVER	Execute and close loops; handover and final account	CS executes construction; SAFE runs inspect→correct→reinspect; SUP triggers supply events; ARC/ENG initiate changes, QS prices, OWN approves/rejects; then commissioning, final account, and handover.	handover_complete ∧ final_account_closed	Outcome distributions; change_orders; compliance_log; supply_incidents; handover_docs.

tion in the GBA construction sector and evaluate policy impacts (Fig. 1). The model uses an executable, stage-gated project lifecycle as the process backbone, within which heterogeneous stakeholder decisions and interactions are embedded under hard gate conditions, producing auditable project trajectories and a time-varying collaboration network. A lightweight macro environment captures demand, capacity utilization, price pressure, and regulatory intensity, thereby coupling multi-project concurrency with micro execution outcomes. Cross-jurisdiction access and practice rules are encoded as a policy vector $\boldsymbol{\theta}(p)$ and embedded into (i) eligibility and waiting times, (ii) cost–risk functions, and (iii) event/transition mechanisms, yielding a traceable chain from policy provisions to behavioral change and system outputs, including network evolution, concentration trends, and distance-decay signatures.

Project lifecycle backbone: four-stage state machine

Each project follows a fixed four-stage state machine (INIT → SCHEME → AWARD → DELIVER), with transitions governed by hard gates defined on project states (cost, schedule, compliance, and supply). Agents may execute only stage-feasible actions, ensuring executability and direct logging of responsibilities and collaboration ties.

INIT sets budget, schedule, and team formation conditions. **SCHEME** freezes the scheme and closes the cost loop. **AWARD** finalizes mobilization prerequisites, including contract, safety, and procurement. **DELIVER** simulates execution through supply, inspection–correction, and change-management loops, propagating impacts to time–cost outcomes and network evolution.

LLM-driven agents and action execution

The framework includes seven heterogeneous, LLM-driven agent types: owner (OWN), architect (ARC), engineer (ENG), quantity surveyor (QS), contractor (CS),

Table 3: Policy provisions, quantified model parameters, and the initial baseline state

Region / regime	Entry friction	Compliance burden	Scope constraints	τ_p	π_p	σ_p	c_p	s_p
Initial state (2015 pre-policy baseline)	Long and uncertain cross-border approval/recognition procedures.	Higher recurring administrative burden and lower renewal predictability.	Narrow effective cross-border practice scope.	120	0.40	0.45	0.55	0.25
Guangdong provincial pilot (9 Mainland cities)	Online filing; registration/publication within **10 working days**; other procedural stages not uniformly specified.	Validity aligned with **Hong Kong registration validity**; change filing within **30 days**; renewal/annual reporting not explicitly specified.	Broad service coverage within the 9-city area, subject to qualification/capability.	20	0.71	0.20	0.25	0.85
Qianhai (Shen Qianhai Gui [2023] No. 6)	Preliminary review within **5 working days**; public notice ≥**5 working days**; objection handling **30 working days** (extendable).	Validity typically **5 years**; renewal within **90 days before** expiry; change filing within **90 days**.	Moderate scope constrained by practice-category filing and qualification mapping.	10	0.82	0.20	0.35	0.60
Hengqin (Bureau Doc. No. 6/2024)	Fully online; supplementation within **10 working days**; public notice ≥**5 working days**; objection review exists.	Validity aligned with **Hong Kong/Macao registration validity**; renewal within **90 days before** expiry; change filing within **30 days**; annual report by **Dec. 31**.	Broad scope covering construction and multiple professional services, subject to filing scope.	10	0.80	0.25	0.45	0.80
Hong Kong (AP/RSE/RGE registration)	Committee referral/scheduling within **2 months**; outcome notification within **3 months** after review/interview.	Validity **5 years**; renewal window from **4 months** to **28 days** before expiry; restoration within **2 years**.	Statutory scope strictly bounded by registration class.	150	0.85	0.10	0.30	0.20
Macao (urban architecture / engineering / planning qualification)	Hard entry gate requiring ≥**2 years** full-time or ≥**5 years** part-time internship before registration.	Renewal follows local rules; validity reported as **until the following January**.	Statutory scope bounded by recognized qualification and registration domain.	730	0.35	0.10	0.40	0.15

Note: τ_p is entry waiting time (days) and is set equal to the stated procedural time in the policy text, summing sequential stages where applicable. π_p is coded from normalized processing time, number of procedural stages, and whether the procedure contains unspecified or open-ended review components. σ_p is coded from the existence of extensions, objections, or discretionary review. c_p is coded from renewal frequency, annual reporting, and change-filing requirements. s_p is coded as the normalized breadth of recognized service categories and responsibility interfaces. All coded indices are scaled to $[0, 1]$. The initial state represents the pre-policy cross-border practice environment at the start of the simulation in 2015.

Table 4: 2015 empirical anchors and simulation initialization values of macro variables

	D_t	K_t	u_t	P_t	M_t	R_t
2015 real-data anchor	RMB 40.81 tn new-start investment	168,222 firms 51.15 m workers	12.43 bn m^2 / 51.15 m workers	98.2	98.9	442 accidents
Simulation init-value	40.81	168,222; 51.15	243	98.2	98.9	1.00

Note: tn = trillion; bn = billion; m = million. R_t is a constructed regulatory-intensity index, with the 2015 level set to 1.00.

safety/compliance entity (SAFE), and supplier (SUP). Each agent follows a role-consistent decision loop but acts only through the stage-feasible action set induced by the project state machine. The framework is therefore hybrid rather than purely generative: the LLM serves as a bounded behavioral policy module for context-sensitive action selection, while feasibility checks, state transitions, accounting updates, and regulatory enforcement are handled by the rule-based simulator. Agent calibration is implemented at the simulator level rather than through fine-tuning the LLM itself. Role populations, capacities, behavioral priors, and memory-update settings are initialized from empirical distributions or bounded assumptions and adjusted so that the baseline simulation reproduces key stylized collaboration patterns.

Memory mechanism. Each agent maintains two complementary memory stores. *Episodic memory* records short-horizon observations and time-stamped experience traces, including stage changes, gate outcomes, workload pressure, market or policy updates, and realized interaction outcomes such as invitations, bids, approvals, delays, violations, change orders, and partner-specific collaboration experiences. *Semantic memory* stores slower-changing summaries distilled from repeated episodes, including partner reliability, expected approval difficulty, coordination friction, effective bidding conditions, and self-assessed capability across jurisdictions.

Retrieval. At each decision step, the agent retrieves a compact subset of memories using recency–relevance scoring, with optional importance weighting for salient events.

Consolidation. Memory updating combines event-triggered and periodic summarization. Event-triggered consolidation follows high-impact milestones such as bid success or failure, project success or failure, major inspection or penalty events, and major policy changes. Periodic consolidation compresses recent episodic traces into higher-level summaries, allowing agents to revise beliefs about partners, jurisdictions, risks, and opportunities over time. After execution, realized outcomes are first written to episodic memory and may then update semantic memory through these mechanisms.

Workload modes. Agents operate in two modes: (i) *in-*

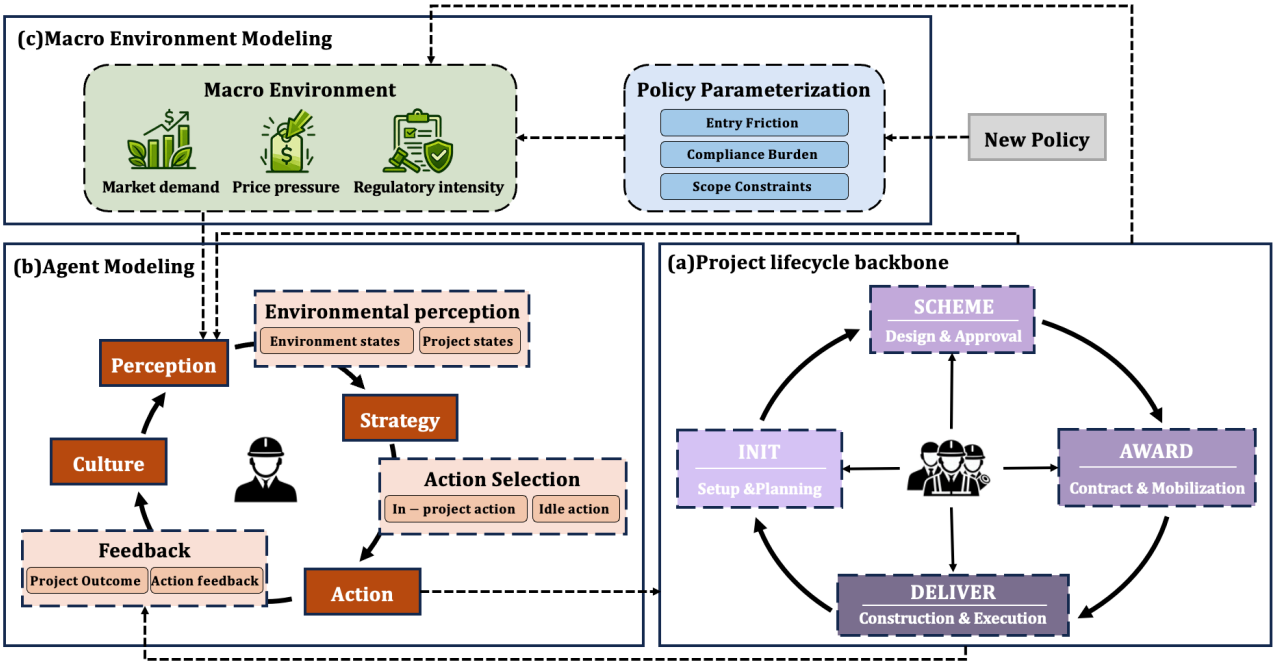


Figure 1: Overall micro–macro coupled agent-based simulation framework.

project, emphasizing resource lock-in, stage-dependent responsibilities, and reactive coordination; and (ii) *idle/out-of-project*, emphasizing opportunity search, teaming, eligibility maintenance, and bidding stance adjustment. When utilization exceeds capacity, agents enter a resource-constrained regime and apply interpretable load-shedding rules until capacity is released by project completion. This mechanism is important for reproducing concentration-upward dynamics (Fig. 3) through cumulative advantage under constrained capacity.

Decision pipeline. At each time step, the simulator builds the agent's decision context from three elements: **Perception**, **Culture**, and **Strategy**. **Perception** includes the observable state and retrieved memories, such as stage and gate status, workload and capacity, project states, macro signals, and partner reputation/availability. **Culture** captures stable value dispositions, such as baseline risk tolerance and cross-border inclination. **Strategy** captures short-horizon operating stances, such as expansion versus stability or profit-first versus scale-first. The simulator packages this context, together with role instruction and the admissible action set, into a structured prompt. The LLM selects one feasible action and returns it in a machine-parsable schema with a brief rationale. The simulator then executes the action and updates states, workloads, and memories through deterministic or stochastic transition rules.

Prompt structure and example. Each decision call is instantiated as a structured prompt with four components: (1) a *system layer* defining the agent's role, objectives, constraints, and behavioral style; (2) a *state layer* summarizing the current project stage, gate status, workload mode, key project indicators, and macro conditions; (3) a *memory layer* containing the retrieved episodic and semantic memories; and (4) an *action layer* listing the admissible actions and requiring one action with a brief rationale in a predefined output schema. Fig. 2 illustrates an example prompt template and decision interface for an LLM-driven agent.

[System Role]
You are a Contractor (CS) agent in a construction market simulation.
You must select exactly one feasible action.
Your action choice must comply with:
- {Contractor Rules} , {Project State Machine} , {Jurisdiction Policy Constraints}, {Workload Mode Constraints}

[Agent Profile]
Role: Contractor, Stable traits: {Risk Tolerance}, {Cross-border Inclination}
Current strategy stance: {Strategy Stance}, Home market: {Home Market}

[Current Decision Context]
Project stage: {Project Stage}, Gate status: {Gate Status}
Workload mode: {Workload Mode}, Utilization status: {Utilization Status}
Capacity condition: {Capacity Condition}, Observable state: {Current Project State}
Macro conditions: {Current Macro Conditions}

[Retrieved Memory]
Recent observations: {Recent World Observations}, Episodic memory: {Retrieved Episodic Memory}
Semantic memory: {Retrieved Semantic Memory}

[Admissible Actions]
{Admissible Actions}

[Output Schema]
{ "selected_action": "...", "target": "...", "rationale": "...", "confidence": 0.00}

[Fallback Rule]
If the available information is incomplete or conflicting, choose the most conservative feasible action consistent with {Contractor Rules} and current capacity constraints.

Figure 2: Example partially instantiated prompt template for a contractor agent.

Macro environment coupling

To couple micro execution with industry conditions, the macro state at time t is $z_t = \{D_t, \mathrm{K}_t, \mathrm{u}_t, P_t, M_t, R_t\}$, where D_t denotes market opportunity intensity, K_t role capacities, u_t utilization, P_t the general price level, M_t

the material price level, and R_t regulatory intensity. New projects are initiated endogenously as owners decide whether to launch projects under current macro conditions. Macro variables affect the simulation through three channels: *cost* (P_t, M_t), *delay/efficiency* (u_t), and *events* (M_t, u_t, R_t). Because the simulation starts in 2015, initial values are anchored to official data (Table 4) and then evolve through statistical inputs and endogenous feedback.
We use two parsimonious feedbacks:

$$P_{t+1} = P_t \left(1 + \kappa(\bar{u}_t - \bar{u}^\star)\right), \tag{1}$$

$$R_{t+1} = R_t + \eta \left(\mathrm{ViolationRate}_t - \rho^\star\right). \tag{2}$$

Policy parameterization and traceable embedding

For scenario evaluation, each policy regime p is represented by

$$\boldsymbol{\theta}(p) = \{\tau_p, \pi_p, \sigma_p, c_p, s_p\},$$

where τ_p denotes entry waiting time, π_p approval success probability, σ_p procedural uncertainty, c_p compliance burden, and s_p scope breadth. Table 3 reports both policy provisions and coded parameter values.
When a regional policy scenario is activated, the corresponding $\boldsymbol{\theta}(p)$ replaces the initial baseline state for that region. Specifically, τ_p is set from stated processing time; π_p is coded from normalized duration, procedural clarity, and review complexity:

$$\pi_p = \alpha_0 + \alpha_1 q_p - \alpha_2 \tilde{\tau}_p - \alpha_3 x_p,$$

procedural uncertainty is coded as

$$\sigma_p = \beta_1 u_p + \beta_2 x_p,$$

compliance burden is coded as

$$c_p = \gamma_1 \tilde{v}_p + \gamma_2 a_p + \gamma_3 f_p,$$

and scope breadth is coded as

$$s_p = \frac{m_p}{m_{\max}}.$$

These parameters enter the model through three channels: **(1) gates and waiting times,** where τ_p delays eligibility and cross-border readiness; **(2) cost and risk functions,** where π_p, σ_p, and c_p affect entry cost, compliance cost, and bid risk premia; and **(3) event and transition rules,** where c_p affects inspection, renewal, penalty, and disruption events, while s_p constrains the feasible cross-border action set.
This design yields a traceable mapping from *policy provisions → coded parameters → agent behavior → simulation outputs*.

Experiments and Analysis

The simulation is conducted in the GBA. The model includes seven heterogeneous agent types, with 100 agents per type (700 in total), allocated across cities according to observed firm distributions. Cross-jurisdiction policies are encoded as a parameter vector affecting entry friction, compliance burden, and scope constraints. The simulation spans 2015–2024 and generates multi-project concurrency and cross-city collaboration networks under a time-varying macro environment.
At the start of each run, agent populations, city allocation, macro initial values, and the pre-policy baseline are initialized using the calibration settings described earlier. A policy scenario is then activated by replacing the baseline policy state with the corresponding policy vector for the affected region. At each simulation step, owner agents decide whether to initiate new projects, while candidate partners make bounded LLM-driven decisions subject to workload, memory, policy constraints, and stage-feasible action sets. Feasible projects then advance through the lifecycle state machine, and repeated project-level interactions generate evolving collaboration networks and aggregate outputs for validation and policy analysis.

Simulation validity via stylized-facts consistency checks

We assess platform validity through *stylized-facts consistency*. Rather than targeting exact numerical matching, we test whether the model reproduces two robust empirical regularities of the construction industry: (i) market concentration with head-firm dominance and (ii) distance decay with local bias in cross-city collaboration.

Macro structure: concentration and head-firm dominance. Empirical evidence indicates rising concentration in the construction industry. For China, official statistics show that top-tier firms expanded their market share to 68.0% of output in 2021 (**?**), while the combined share of newly signed contracts held by the seven major central construction SOEs (CR7) reached 41.89% in 2023 (**?**). Consistent with these patterns, the simulation produces a clear concentration-upward trajectory (Fig. 3). We use the annual Top-10 share (CR10) as a regional proxy for GBA concentration and compare it with the national CR7 series. Both rise persistently, while the simulated CR10 exhibits a stable–accelerating–plateauing pattern driven by cumulative advantage under capacity, regulatory, and competitive constraints.

Spatial structure: distance decay and local bias. A second robust regularity is distance decay with strong local bias: cross-city collaboration becomes less frequent as distance increases, while same-city or nearby interactions remain dense. Similar patterns have been documented in large-scale procurement markets (Herz and Varela-Irimia, 2020). We conduct an analogous contract-level validation using simulated collaboration records. In Fig. 4, each point represents one collaboration contract, with inter-city distance on the x-axis and contract value on the y-axis. The scatter shows two stylized features: contracts become markedly sparser at longer distances, and contract values are heavy-tailed, with many small contracts and few large

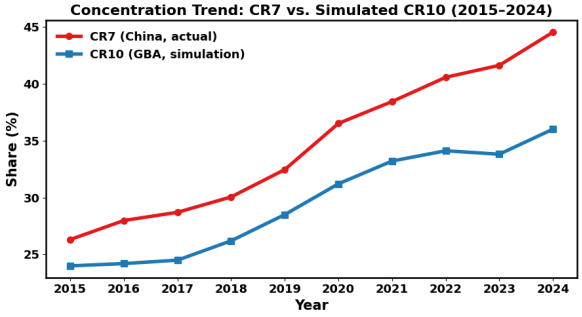


Figure 3: Concentration trend: national CR7 (China, actual) vs. simulated CR10 (GBA).

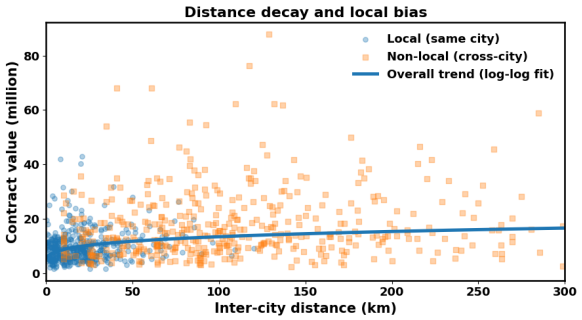


Figure 4: Distance decay and local bias in simulated collaboration contracts.

ones. Together, these results indicate that the platform reproduces representative industry regularities and therefore provides a credible baseline for counterfactual experiments.

Ablation study and stability. To assess the contribution of the proposed LLM-agent design, we compare three settings: (i) the full model with memory and intra-role heterogeneity, (ii) a variant without the memory module, and (iii) a variant without intra-role heterogeneity. Because the empirical benchmark is the national CR7 series whereas the simulation output is the GBA CR10 series, we compare normalized trajectories indexed to 2015 values. Similarity is evaluated using RMSE and trajectory correlation (ρ), and each setting is repeated three times. Table 5 reports mean values and relative variation across runs. The results show a stable ranking: the full model performs best, removing memory weakens performance, and removing intra-role heterogeneity weakens it further.

Cross-Regional Policy Analysis

After validating the platform against stylized facts, we use it to characterize the structure and city-level heterogeneity of cross-regional collaboration in the GBA, focusing on whether institutional differences yield interpretable spatial stratification and diffusion patterns.

Mainland-city collaboration network The simulated network among the nine Mainland GBA cities evolves into a clearer *core–semi-core–periphery* structure rather than expanding uniformly. Fig. 5 reports collaboration

Table 5: Ablation and stability results on concentration dynamics using normalized trajectories

Model setting	RMSE	Trajectory correlation (ρ)
Full model (memory + heterogeneity)	5.44 ± 28.47%	0.9913 ± 20.93%
Without memory module	9.62 ± 23.82%	0.9734 ± 16.39%
Without intra-role heterogeneity	11.44 ± 29.15%	0.9649 ± 23.68%

Note: The empirical benchmark is the national CR7 series, while the simulation output is the GBA CR10 series. Both series are indexed to their 2015 values before computing RMSE and trajectory correlation. Reported values are means over three repeated runs; percentages indicate relative variation across runs.

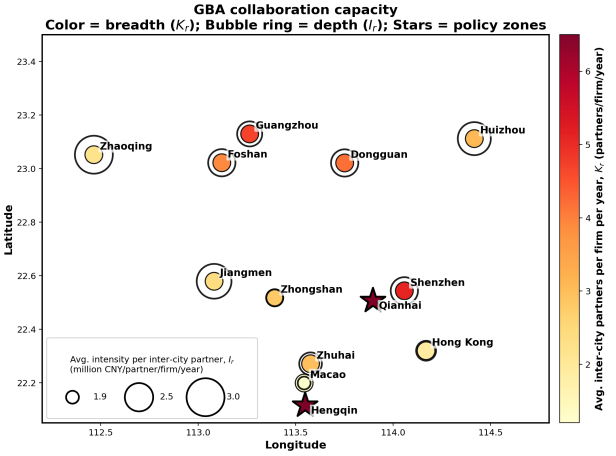


Figure 5: GBA collaboration capacity at the city level.

breadth K_r and collaboration depth I_r. Three patterns emerge: core hubs such as Guangzhou and Shenzhen exhibit high breadth; manufacturing and supply-chain cities such as Foshan and Dongguan form clustered collaboration anchored by stable division-of-labor ties; and peripheral cities such as Jiangmen, Zhaoqing, and Huizhou more often show lower breadth but higher depth, indicating fewer yet deeper channels.

Qianhai and Hengqin Qianhai and Hengqin function as *institutional gateways*. Their main role is to increase the arrival probability and diffusion speed of cross-border collaboration into the Mainland network by lowering institutional friction and improving procedural predictability. Two time-series signatures support this interpretation: (i) cross-border share rises earlier within the policy zones than in other cities (Fig. 6); and (ii) cross-border projects in other cities initially show a higher share linked to policy zones, which later declines as experience diffuses and projects shift from gateway-driven entry to broader localization (Fig. 7).

Hong Kong/Macao collaboration characteristics When institutional differences are explicitly modeled, Hong Kong/Macao shape the GBA collaboration landscape asymmetrically in both direction and intensity. Cross-border mobility becomes directional, with net inflow from Hong Kong/Macao to the Mainland—especially

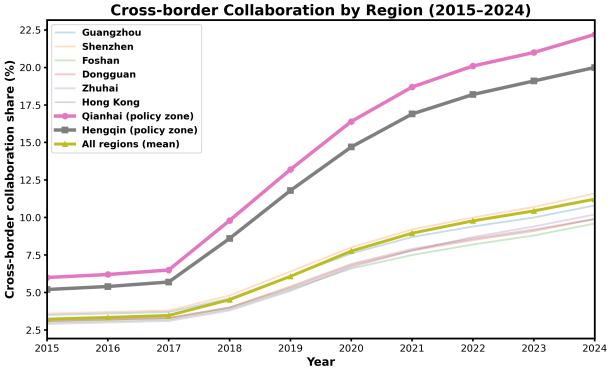


Figure 6: Cross-border collaboration by region (2015–2024).

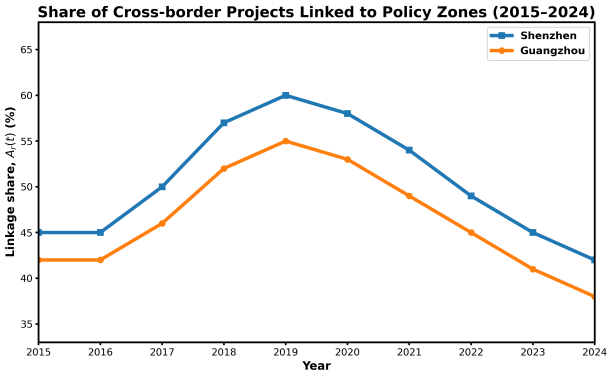


Figure 7: Share of cross-border projects linked to policy zones (2015–2024).

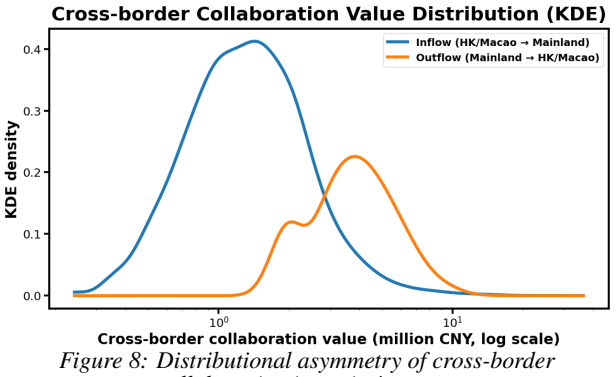


Figure 8: Distributional asymmetry of cross-border collaboration intensity/amount.

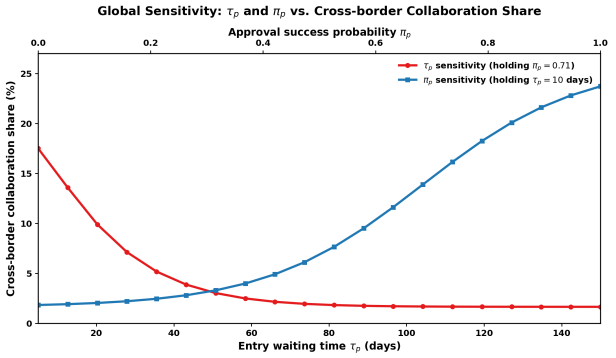


Figure 9: Global sensitivity of cross-border collaboration share.

via Qianhai and Hengqin—because Mainland-to-Hong Kong entry faces higher institutional costs and tighter constraints. Collaboration intensity also differs by direction: inflow is more concentrated in high-end professional and information/consulting services, whereas outflow is more constrained and tends to concentrate in price/efficiency-advantaged services and supply-chain cooperation.

Sensitivity analysis of key policy parameters

Global sensitivity of entry waiting time and approval success We first isolate entry waiting time τ_p and approval success probability π_p, holding other policy dimensions fixed, and record the resulting cross-border collaboration share (Fig. 9). The responses are nonlinear but smooth. For π_p, the curve flattens at both ends, indicating diminishing marginal effects. For τ_p, collaboration declines rapidly beyond the low-delay range and then approaches a floor. The model is therefore directionally sensitive but not dependent on knife-edge calibration.

Robustness of the Qianhai gateway effect under alternative friction gaps We further test whether Qianhai's gateway role remains robust when its friction advantage over the Guangdong baseline varies from 1% to 30%. For each setting, we compute the share of cross-border projects linked to Qianhai. As shown in Fig. 10, even a modest advantage increases Qianhai-linked projects, while marginal gains diminish as the advantage widens. The conclusion remains stable across the tested range: lower institutional friction systematically strengthens Qianhai's gateway role.

Conclusions

This paper proposes an LLM-driven agent-based simulation framework for evaluating cross-regional construction policies in the GBA. The framework combines an executable stage-gated project lifecycle, role-consistent agent action sets, and a lightweight macro environment, while encoding cross-jurisdiction practice rules as a policy vector embedded into gates, cost–risk functions, and event/transition rules. Stylized-facts consistency checks show that the platform reproduces two robust market regularities—increasing concentration with head-firm dominance and distance decay with strong local bias—thereby providing a credible baseline for counterfactual policy experiments. We further show how institutional differences and policy-zone mechanisms generate interpretable signatures in network stratification, diffusion paths, and directional asymmetries of cross-border collaboration intensity.

Beyond the GBA case, the framework is extensible and conditionally adaptable. While the core process backbone can be retained, applying it to other contexts would require context-specific adjustment of agent configuration parameters, macro-level parameters, and policy-vector settings to reflect local institutional arrangements, market conditions, and project structures. The framework should there-

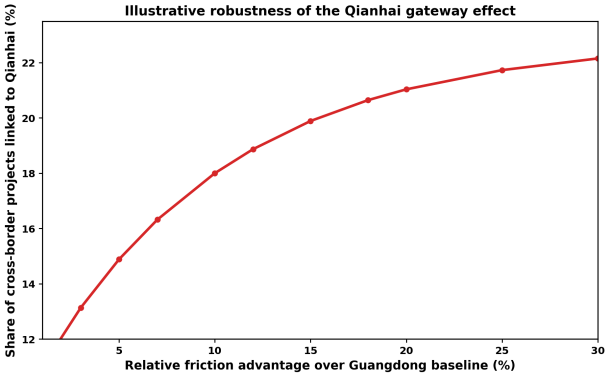


Figure 10: Qianhai gateway effect under alternative friction advantages.

fore be understood as a reproducible methodological template for comparative policy analysis, rather than a directly transferable model.

Acknowledgments

The work described in this paper was substantially supported by a grant from the Natural Science Foundation of China / Research Grants Council of the Hong Kong SAR of China (RGC Project No. CRS_PolyU505/23), and a grant from the Strategic Public Policy Research (SPPR) Funding Scheme (Project No. S2023.A6.023.23S).

References

Bonabeau, E. (2002). Agent-based modeling: methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences of the United States of America, 99(suppl. 3):7280–7287.

Dang, P., Zhu, J., Li, W., Xie, Y., and Zhang, H. (2025). Large-language-model-driven agents for fire evacuation simulation in a cellular automata environment. Safety Science, 191:106935.

Guangdong Provincial Department of Housing and Urban-Rural Development (2020). Interim administrative measures for the pilot program for hong kong engineering construction consultancy enterprises and professionals to establish practice and provide services in the mainland cities of the guangdong–hong kong–macao greater bay area. Guangdong Provincial Department of Housing and Urban-Rural Development.

Herz, B. and Varela-Irimia, X.-L. (2020). Border effects in european public procurement. Journal of Economic Geography, 20(6):1359–1405.

Hui, E. C. M., Li, X., Chen, T., and Lang, W. (2020). Deciphering the spatial structure of china's megacity region: a new bay area—the guangdong–hong kong–macao greater bay area in the making. Cities, 105:102168.

Khodabandelu, A. and Park, J. (2021). Agent-based modeling and simulation in construction. Automation in Construction, 131:103882.

Li, N., Gao, C., Li, M., Li, Y., and Liao, Q. (2024). Econagent: large language model-empowered agents for simulating macroeconomic activities. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, pages 15523–15536. Association for Computational Linguistics.

Min, J. U. and Björnsson, H. C. (2008). Agent-based construction supply chain simulator (cs2) for measuring the value of real-time information sharing in construction. Journal of Management in Engineering, 24(4):245–254.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). Generative agents: interactive simulacra of human behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, pages 1–22. Association for Computing Machinery.

Shi, Q., Ren, H., Cai, W., and Gao, J. (2019). How to set the proper level of carbon tax in the context of chinese construction sector? a cge analysis. Journal of Cleaner Production, 240:117955.

Shin, M., Lee, H.-S., Park, M., Moon, M., and Han, S. (2014). A system dynamics approach for modeling construction workers' safety attitudes and behaviors. Accident Analysis & Prevention, 68:95–105.

Watkins, M., Mukherjee, A., Onder, N., and Mattila, K. G. (2009). Using agent-based modeling to study construction labor productivity as an emergent property of individual and crew interactions. Journal of Construction Engineering and Management, 135(7):657–667.

Wijermans, N., Scholz, G., Chappin, É., Heppenstall, A., Filatova, T., Polhill, J. G., Semeniuk, C., and Stöppler, F. (2023). Agent decision-making: the elephant in the room—enabling the justification of decision model fit in social-ecological models. Environmental Modelling & Software, 170:105850.

Zhang, P., Li, N., Jiang, Z., Fang, D., and Anumba, C. J. (2019). An agent-based modeling approach for understanding the effect of worker-management interactions on construction workers' safety-related behaviors. Automation in Construction, 97:29–43.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

VLM-AUGMENTED RAG FRAMEWORK WITH ONTOLOGY-CONSTRAINED LLM REASONING FOR EXISTING BUILDING MATERIAL DOCUMENTATION

Julia Kaltenegger[1], Stavros Nousias[2], Ekaterina Petrova[1], André Borrmann[2] and Pieter Pauwels[1]
[1]Information Systems Built Environment, TU Eindhoven, Eindhoven, Netherlands
[2]Chair of Computing in Civil and Building Engineering, TU Munich, Munich, Germany

Abstract

Documenting information about materials in existing buildings is essential for circular construction and the reuse of elements. While multimodal Vision-Language Models (VLMs) can classify materials from images, they often lack the contextual semantic depth required for technical documentation. This article introduces a VLM-augmented Retrieval-Augmented Generation (RAG) framework that utilises ontology-constrained LLM reasoning to bridge this gap. By integrating domain-specific text retrieval and formal ontologies, the method enables the identification of building element layers and material types. The method is validated using a reference dataset of Dutch residential buildings and demonstrated by enriching semantic graphs with predicted material data.

Introduction

The built environment contributes significantly to resource depletion and waste generation. Addressing these impacts requires prioritising lifecycle material efficiency (reducing waste at the source), reducing the embodied carbon, and enabling reusability and assembly-disassembly models (European Commission, 2020). Consequently, systematic documentation of the existing building stock is essential to support material assessment and reuse, although such documentation remains challenging at scale. Building engineers typically assess materials in existing buildings through on-site inspections, manually documenting building characteristics, element details, and materials using images and text-based data sources (Smeyers and Mertens, 2022). As existing buildings usually lack Building Information Models (BIM) and diverse technical documentation, information is often retrieved from building codes, construction regulations, and archetype descriptions to infer probable construction types (Byers et al., 2024). Archetypes define building types as reference systems characterised by construction period, thermal and energy performance, average quantities of floor area, building-element surface area, and element thickness, thereby supporting top-down building and material stock modelling (Pei et al., 2024; Rijksdienst voor Ondernemend Nederland, 2022). For example, Dutch apartment buildings from the 1980s typically use prefabricated or cast-in-place concrete with low insulation, whereas single-family homes from the same period are predominantly constructed with insulated brick cavity walls. Therefore, the material composition of buildings varies between construction periods and building types.

With the advancement of digital technologies and Artificial Intelligence, existing buildings and their materials are increasingly documented using information-extraction methods such as image recognition, text-based reasoning, and semantic graph representations (Zeng et al., 2024; Pei et al., 2024). As such, computer vision for detecting and classifying building features, facades and materials has moved beyond solely using traditional deep neural networks (Gordon et al., 2023; Marin-Garcia et al., 2023) and has explored the use of Vision-Language Models (VLMs). By jointly learning image and text embeddings, VLMs enable prompt-based descriptive image classification, allowing, for example, the estimation of building age and energy efficiency from images and text (Zeng et al., 2024; Pan et al., 2025). Reasoning tasks based on other text data have shown success when using Large Language Models (LLM) alongside VLMs (Cooper et al., 2025). In this regard, Retrieval-Augmented Generation (RAG) is typically used to integrate domain-specific text corpora, providing natural-language rationales that support LLM-based reasoning (Oche et al., 2025). Methods for generating rationales include, among others, semantically aware sentence parsing (Chen et al., 2019) and studying textual similarities through text matching aligned with ontology definitions (Harispe et al., 2022). Semantic graphs using ontologies support context-based reasoning while integrating and contextualising heterogeneous, ambiguous data (Luo et al., 2023; Zheng et al., 2023). Research in the building materials domain has investigated a hybrid approach that combines text-based and ontology-based similarity matching to enrich material properties relevant to BIM data (Forth et al., 2025). These capabilities, in turn, support VLM-based image classification and LLM-based reasoning (Li et al., 2026; Chen et al., 2023).

This article recognises the potential of the interplay between VLM-based image classification, LLM inferencing, and ontologies to improve the formal description of visual content and enable further interpretation through linkable vocabularies. The aim of this article is to support the documentation of existing building stock using a semantic graph as the backbone, along with image and textual data enrichments. Therefore, this study investigates a VLM-augmented RAG framework for ontology-constrained LLM reasoning to enable the recognition, identification and inference of material information in existing buildings across archetypes . The proposed method

relies on (1) a VLM-based prediction using on-site images of buildings and structuring the corresponding predicted residential building types and material categories in a semantic graph; (2) a VLM-augmented RAG process that incorporates domain-specific knowledge, including construction regulations and archetype definitions; and (3) an ontology-constrained LLM-based reasoning that incorporates the RAG insights and parses the predicted information as semantic graph statements (Fig. 1). Together, these components enable a structured, semantically consistent representation of materials in existing buildings, thereby supporting reuse-oriented assessment.

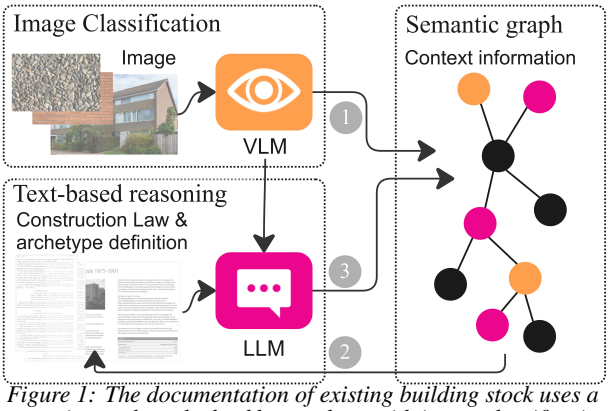


Figure 1: The documentation of existing building stock uses a semantic graph as the backbone, along with image classification and text-based reasoning.

State-of-the-Art

Related works are discussed on the topic of Vision Language Models (VLMs) and semantic-aware text processing in the context of LLM reasoning.

Vision Language models

Computer vision is the fundamental driver for understanding and accessing information from images. Methods such as image classification, semantic segmentation, and object detection are common practices in deep learning, often relying on Convolutional Neural Networks (CNNs). In Architecture, Engineering and Construction (AEC), computer vision for building material recognition supports, e.g., brick inventories by analysing replacement requirements and accounting for deterioration due to efflorescence (Marin-Garcia et al., 2023). Further, automated detection of building elements for deconstruction planning and material reuse has also been studied (Gordon et al., 2023).

Advances in computer vision have led to the development of Vision Language Models (VLMs), which jointly learn from image and text modalities. These generative models embed visual and textual information into a shared representation space, allowing image classification beyond predefined categories. VLMs are typically pre-trained on large-scale image–text pairs using transformer-based architectures. VLMs are based on vision transformers (ViT), using foundation models that transfer knowledge from seen to unseen classes by exploiting cosine similarity between vectors in the shared image–text embedding space (Hassija et al., 2025). Zero-shot VLMs enable object localisation and recognition without labelled target examples by using vision–text transformer architectures with prompt-based image inputs. In contrast, learning with few-shot VLMs predicts classes from a small number of labelled samples by guiding an LLM with limited task examples.

Application-specific studies have explored the use of a zero-shot classification for building-related tasks. For example, Zeng et al. (2024) employ GPT-4 to classify building age from images of buildings in London and predict the age epochs using a structured prompting strategy. Similarly, Pan et al. (2025) utilise GPT-4 as a pre-trained model to infer building energy efficiency levels from descriptive visual attributes, without prior task-specific training on labelled examples. In both cases, custom datasets comprising building images and corresponding labels (e.g., age or energy efficiency) are constructed to provide a baseline for validating the VLM's predictions. As a result, large VLMs can identify previously unseen objects in images using text prompts provided by an operator (Radford et al., 2021). Furthermore, the zero-shot model OWL-Vit (Minderer et al., 2022) is a simple open-vocabulary object-detection model with vision transformers trained on a variety of image-text pairs. It can query an image using one or more text prompts to detect target objects.

Semantically aware text processing for LLM reasoning on top of vision models

VLMs have proven successful in object recognition and classification tasks. However, research shows that using LLMs alongside VLMs can improve VLM performance on tasks that require reasoning from other text data (Cooper et al., 2025). Therefore, text corpora are created that store domain-specific, unstructured datasets, which can be used as natural-language rationale for LLM-based inferencing. The RAG framework is commonly applied in this context as it enables the integration of external data sources into a structured text corpus. Retrieval queries are used to guide the LLM reasoning, thus reducing hallucination (Oche et al., 2025). Methods for generating rationales, as part of the RAG framework, follow prompt engineering, such as Chain of Thought (CoT) (Wei et al., 2022) and semantically aware sentence parsing (Chen et al., 2019).

Semantically aware techniques measure semantic similarity through text matching and ontological alignment to capture conceptual relationships beyond lexical similarity (Harispe et al., 2022). Semantic awareness through text matching has been studied by Zhang et al. (2024). The authors use a RAG framework to enhance LLM-based reasoning on top of VLM image classification. The RAG incorporates context-specific information from an architecture knowledge base, thereby guiding the end user according to the logic of the original document. Semantic awareness through ontology alignments is enabled by Semantic Web technologies. Semantic graphs extend beyond structuring and organisation of domain knowledge

and enable context-based reasoning while ensuring the integration of heterogeneous and ambiguous data (Luo et al., 2023). Recent studies employ semantic graphs to support VLM- and LLM-based inference on construction data. For example, LLM-based question answering enables users to query semantic building data in natural language. Information from semantic graphs is accessed using SPARQL-based RAG queries (Li et al., 2026).

In addition, semantic graph entities and relations have been used as input for zero-shot image classification (Chen et al., 2023). Integrating image and text in semantic graphs is studied in the context of multimodal semantic graphs, which link entities that represent semantically similar concepts (Peng et al., 2023). As such, knowledge transfer strategies (Chen et al., 2023) can enable inference of semantic relationships between detected image classes or allow for semantic graph-based object detection (Chen et al., 2024). For instance, picture units that represent several entities such as buildings, trees, and cars, can be detected as classes and linked to the respective semantic entity in a graph database (Peng et al., 2023). Another approach is semantically aware object detection, proposed by Fang et al. (2017). There, the semantic similarity between two concepts is evaluated to determine the likelihood that they co-occur within the same image. Furthermore, ontological entities can be aligned with image-detected objects, thereby enriching image content with semantically related concepts in the semantic graph (Zeng et al., 2024; Zheng et al., 2023).

Method

In this article, images of buildings, taken on-site, and textual data on building archetypes and construction regulations documents are used to extract information about building element layers and possible materials. The paper proposes an image-text information generation process using a VLM-augmented RAG framework for ontology-constrained LLM reasoning, as shown in Fig. 2. A semantic graph is deployed as the backbone to integrate image-text data. The graph uses the Archetype ontology (at:)[1] and the Building Material Performance (bmp:) ontology[2], which describe the semantics of buildings, their elements, possible layer sets, and associated materials. The method is tested and validated on a ground-truth dataset comprising 75 Dutch residential buildings. The proposed method includes:

1. A zero-shot VLM model to classify building images by residential types and material categories.
2. A RAG process that integrates construction regulation and archetype definition.
3. A LLM-based reasoning process that uses the RAG results, Contrastive Language-Image Pre-training (CLIP) predictions, constrained to the ontology vocabulary. The results of the LLM reasoning are then parsed as Resource Description Framework (RDF) triples to enrich the semantic graph.

Zero-shot VLM model for image classification

Contrastive Language-Image Pre-training (CLIP) is used as it is a powerful foundation for zero-shot and few-shot classification. The CLIP framework utilises a dual-encoder architecture to bridge visual and textual data. Unlike traditional classifiers, CLIP is pre-trained on a massive dataset of image-text pairs to learn a shared vector space where related visual and textual concepts are mapped closely together. This article uses 'openai/clip-vit-base-patch32'[3], with an underlying architecture relying on the ViT-G/14 to encode images and a Transformer to encode text. CLIP predicts the residential building type y and the material category z with the function f that uses the input image x and visual prompts P_v to describe context and possible prediction classes in natural language, as shown in Eq. 1.

$$(y,z) = f(x;P_v)(x \in X) \tag{1}$$

The classification tasks for residential building types and material categories follow the formal semantics of the reused ontologies, the AT ontology[1] and the BMP ontology[2]. Prompts are used to introduce the ontology terms and definitions in natural language. Parsing the ontology solely as an RDF file, without descriptions, would result in VLM hallucinations. Therefore, definitions for building types 'Apartment House', 'Single Family House', and 'Terrace House' and for material categories, including 'Brick', 'Concrete', 'Glass', 'Natural stone', and 'Wood', are provided. Although the material categories can provide an estimate of the material origin, the classification of layer-specific material types depends on building-specific contexts, such as construction methods that vary by building type, construction year and country.

Retrieval augmentation process

The RAG pipeline comprises three steps: (2.1) building the text corpus; (2.2) constructing a retrieval query that uses the CLIP predictions and metadata of a building (construction year and country) as input; and (2.3) using textual similarity between the text corpus and the query to extract the most relevant text that the LLM can later use to support the reasoning process (Fig. 3). The text corpus includes building archetype descriptions (Rijksdienst voor Ondernemend Nederland, 2022) and building construction regulations, including building permit codes from 1904 (Schelling and van Gijn, 1904) and 1927 (Peerbolte and der Kaa, 1927). The original data sources are provided in PDF format as embedded bitmaps. Therefore, the optical character recognition (OCR) converter *pytesseract* is used to convert text to text format. The OCR extracts text sections based on the logic of the original document, depicted in Fig. 4. For example, the extracted text can

[1] https://w3id.org/at

[2] https://w3id.org/bmp

[3] https://github.com/openai/CLIP

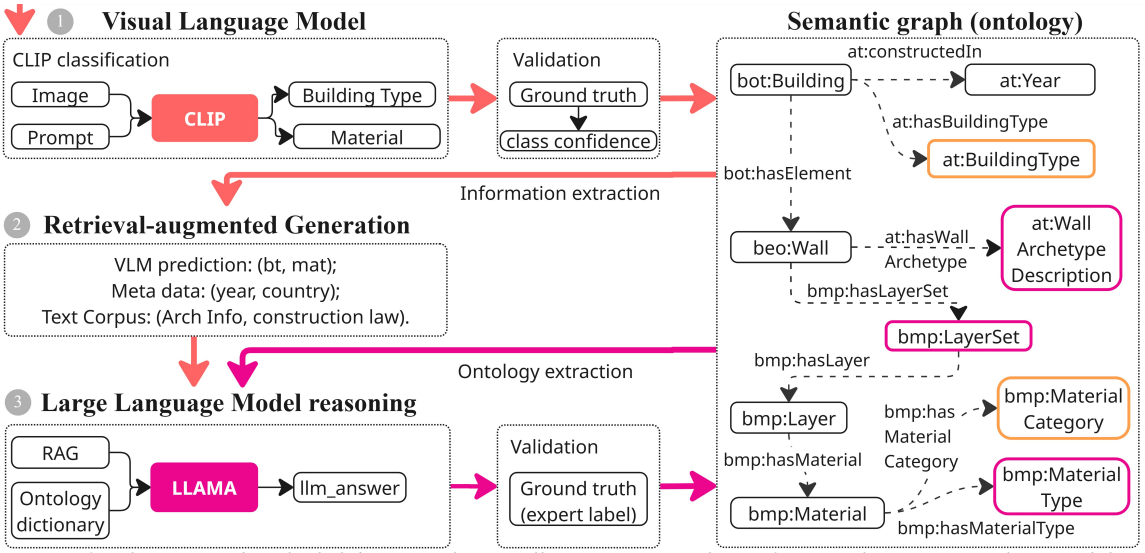


Figure 2: The proposed method elaborating the overall VLM-augmented RAG framework (in orange colour) and the ontology-constrained LLM reasoning (in pink colour).

be structured as a CSV table with columns for 'Building Type', 'Construction Year', 'Archetype Description', and the associated 'Wall Archetype Construction'. Each row of the CSV is extracted and converted to a vector as part of C, the 'Corpus embedding'. The text corpus is processed with the *SentenceTransformer* Python library to provide text embeddings.

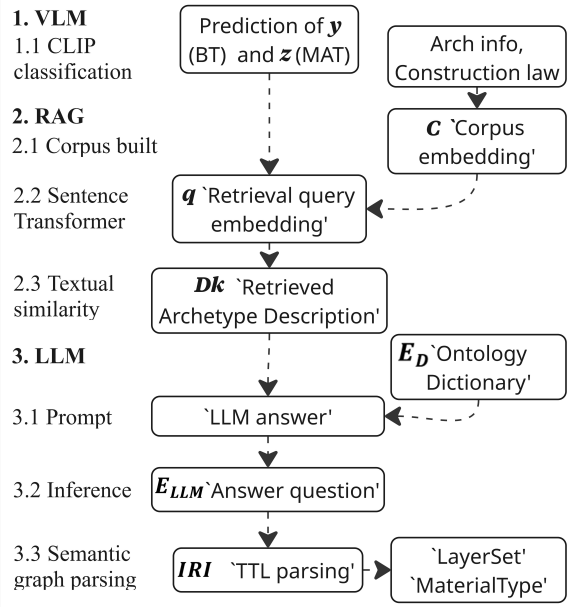


Figure 3: Sequential process of functions that enable VLM informed RAG for ontology-constrained LLM process

During inference, the function 'Retrieval query embedding' builds a query q that fetches the CLIP prediction results for building type y and material category z, and the metadata m (construction year and country) as input, and converts the results to a vector (Eq. 2). The function 'Retrieve archetype description' then extracts the embedded corpus C and the embedded query q, and computes cosine similarity between the two, denoted as Dk (Eq. 3). Dk represents the top-k largest elements of the input tensor (entries that best match the retrieval query with the highest similarity score). The cosine similarity between two vectors a and b is defined in Eq. 4.

$$q = g(y, z, m) \tag{2}$$

$$D_k = Retrieve(C, q) \tag{3}$$

$$S_C(a,b) = \frac{a * b}{||a|| ||b||} \tag{4}$$

The RAG pipeline provides the LLM with the building context information. By applying textual semantic similarity matching against the ontology definitions, the system identifies the most probable entities and integrates them into the semantic graph (Fig. 3).

LLM reasoning and ontology constraining

In the third part of the procedure (Fig.3), the LLM reasoning process, comprises three steps: (3.1) LLM answer template and prompt design; (3.2) LLM model inference of the 'LayerSet' and the 'Material Type'; and (3.3) parsing the LLM results in RDF Triples (TTL). First, the function 'LLM answer' is based on an LLM answering template in JSON notation and is constructed as the LLM prompt input. The prompt input relies on the CLIP prediction results and confidence for the 'building type' and 'material category', the metadata 'construction year' and 'country', the 'retrieved corpus text' from the RAG results and ontology embeddings 'ontology guidance'. The latter contains descriptions of ontology classes (bmp:MaterialType and bmp:LayerSet) represented as vectors. Further, the prompt describes the structure of the TTL file; an excerpt of the input prompt is shown in Listing 1.

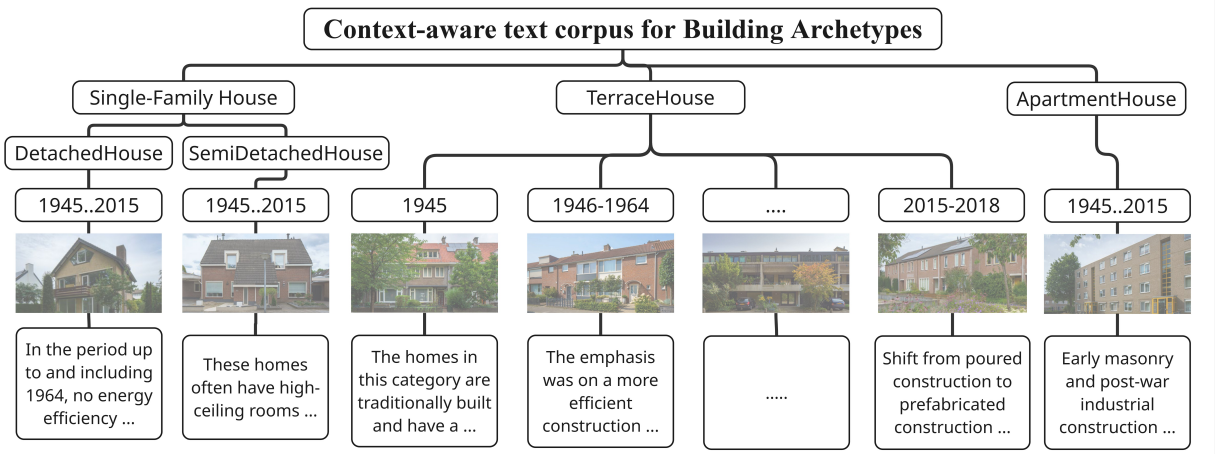


Figure 4: Archetype information structure as preparation for the RAG corpus embedding

Listing 1: LLM Input Prompt for TTL Description

```
# INPUT EVIDENCE:
Metadata Year: {year}
Metadata Region: {region}
Visual Prediction (Type): {bt_pred} (conf: {bt_conf:.2f})
Visual Prediction (Material Category): {mat_cat_pred} (conf: {mat_cat_conf:.2f})
# ARCHETYPE CORPUS EVIDENCE:
{retrieved_corpus_text}
# ONTOLOGY
{ontology_guidance}
# TASK:
# Generate RDF/TTL output
# for the building using these rules and
# proposed ttl layout:
inst:building01
  at:BuildingType [at:hasValue
    {building_types}; {bt_confidence}];
  at:ConstructedIn {year} ;
  at:Country {country} .
inst:wall01
  rdf:type beo:Wall ;
  bmp:hasLayerSet{context['ontology']
    ['wall_layer_option']} ;
  bmp:hasOuterLayer inst:Layer01 ;
  bmp:WallArchetypeDescription
    {retrieve_archdescription} .
inst:Layer01
  rdf:type bmp:Layer ;
  bmp:hasMaterialCategory [at:hasValue
    {material_cat}; {mat_confidence}]; ;
  bmp:hasMaterialType {context['ontology']
    ['material_type']}
```

Next, the 'Answer question' function relies on the context information and the prompt; the LLM model infers the appropriate 'LayerSet' and, if applicable, the 'Material Type' for building archetypes. In the experiments, the LLM model 'llama3:8b' has been used; however, the methods can be applied with any LLM. The 'LayerSet' and 'MaterialType' are selected based on the highest cosine similarity between the retrieved text-based 'Wall description' and the ontology embedding. The final ontological grounding is achieved through a maximum-likelihood selection over the embedding space interpreted by Eq. 5. In the function 'TTL parsing', the IRI is the assigned OWL syntax based on cosine similarity between the embedded LLM results E_{LLM} and the embedded ontology class description E_D. D is the ontology description as a subset of the ontology classes O_E. A fallback logic is implemented, which, if the evidence (the provided text corpus) does not match an ontology-based IRI, causes the LLM to provide a descriptive string instead of generating a new ontology entity. As a result, the LLM writes a semantic graph in Terse RDF Triple Language (TTL) format that assigns the 'LayerSet' with the associated 'MaterialType' to the observed component. The full code and datasets are available on GitHub [3].

$$IRI = \underset{D \in \mathcal{O}\mathcal{E}}{\operatorname{argmax}} \left(\frac{E_{LLM} \cdot E_D}{\|E_{LLM}\| \cdot \|E_D\|} \right) \tag{5}$$

Validation corpus

For validation, a dataset is constructed that contains images and corresponding labels describing 75 Dutch residential buildings by country, construction year, building type, material types, and definitions of wall-layer sets. The data set is manually labelled in collaboration with domain experts; see an example entry in Listing 2. The validation process validates the zero-shot VLM predictions and second, the LLM reasoning using Precision, Recall and F1-Score.

Listing 2: JSON validation dataset

```
"buildingImage":"1965-SFH.png",
"materialImage":"1965-SFH-mat.jpg",
"yearFrom":"1965",
"yearTo":"1974",
"Country":"Netherlands",
"BuildingType":"Single Family House",
"FacadeDescription":"Outer layer is sand
    limestone, in-between layer is insulated
    with eps, inside layer is brick masonry. ",
"materialCategory": "Brick",
"materialType": "Sand limestone",
"LayerSet": "Cavity Wall"
```

Results and validation

The results discuss zero-shot VLM classification, LLM reasoning on the validation dataset, and finally demonstrate

[3] https://github.com/JulKaltenegger/VLM-RAG-LLM-reasoning

the proposed methods on a use case.

Zero-shot VLM classification validation

The zero-shot VLM classification validates the CLIP-based predictions for building type and material category. For each image in the validation dataset, the CLIP generates top-3 predictions. Both classes are directly compared against the corresponding labels in the JSON validation dataset. The 'Single-family Housing' class exhibits the lowest stand-alone CLIP performance, largely due to its typological similarity to 'Terrace Houses'. In terms of material classification, CLIP achieves high precision (0.98) for 'Brick' but fails on 'Concrete' with a precision of only 0.33. This indicates a high rate of false positives, with the model correctly identifying concrete only one out of three times. This type of hallucination is likely caused by visual similarities with cement stucco or large-format stone elements. The confusion matrix for the CLIP prediction for building types is represented in Fig. 5. The precision, recall and the F1 score are presented in Tables 1 and 2.

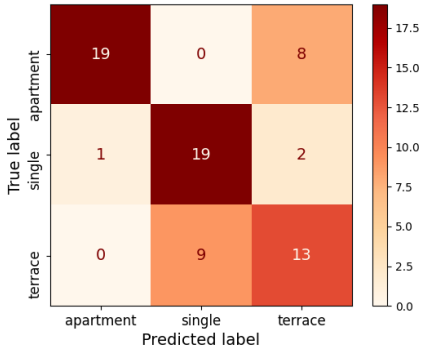


Figure 5: Confusion Matrix Clip Building Type classification

Table 1: Building type classification (CLIP-only)

Class	P	R	F1	Sup.
Apartment house	0.96	0.81	0.88	27
Single family house	0.53	0.83	0.65	12
Terrace house	0.63	0.55	0.59	22
Accuracy		0.72		61
Macro avg	0.70	0.73	0.70	61
Weighted avg	0.75	0.72	0.73	61

Table 2: Material category classification (CLIP-only)

Class	P	R	F1	Sup.
Brick	0.98	0.81	0.88	52
Concrete	0.33	0.86	0.48	7
Accuracy		0.79		61
Macro avg	0.44	0.55	0.45	61
Weighted avg	0.87	0.79	0.81	61

Detecting building type solely from facade images is often ambiguous, motivating the integration of contextual metadata and ontology-constrained reasoning in subsequent pipeline stages.

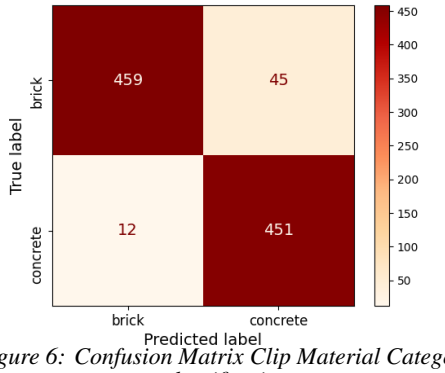


Figure 6: Confusion Matrix Clip Material Category classification

LLM reasoning validation

The VLM-augmented RAG and ontology-constrained LLM pipeline is validated on the validation dataset. Fig. 7 shows the validation performance for each image. This includes the CLIP predictions and context data (construction year and country), which are used to extract archetype descriptions from the text corpus. These retrieved descriptions are then used by the LLM to infer wall construction characteristics and to generate RDF/Turtle statements constrained by the ontology. The resulting RDF triples are parsed and compared against the JSON validation dataset. The cosine similarity is computed between the inferred information (the 'LayerSet' and the 'MaterialType') and the labelled JSON dataset, resulting in average similarity and confidence scores of 0.45 and 0.56, respectively.

Demonstration use case

The use case demonstrates how the proposed method can help construction engineers with large-scale building documentation, particularly in detecting wall layer sets and associated material definitions. A Dutch residential building, a terraced house from 1965, is selected to illustrate how insights from the reasoning process can enrich a semantic graph with entities and their associated confidence values (Fig. 8). A semantic graph is used to describe a building instance inst:Building001, and its associated elements, layer sets and materials. Fist, the semantic graph is enriched by the detected building type and material category, derived from the VLM model. The CLIP classification results with a confidence of 0.87 for 'Building type' at:TerraceHouse, and 0.98 for 'Material Category' bmp:Brick of the observed wall. Furthermore, the LLM's reasoning extracts the archetype description for each building element, in this case, the wall archetype. Furthermore, the LLM reasons that the 'WallLayerSet' and the outer layer is assigned by the 'MaterialType', with a confidence of 0.59 for bmp:CavityWall and 0.34 for bmp:ClayUnit.

This enrichment process relies on text-based documentation of building archetypes and ontology classes, including descriptive annotations. The confidence of the textual semantic matching depends on the degree of align-

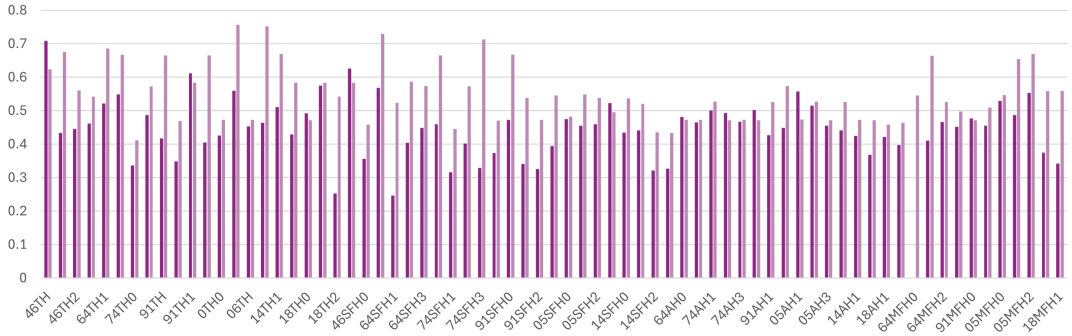


Figure 7: Validation of the complete pipeline with the validation dataset (N=75) showing the semantic similarity (dark purple) and the confidence score (light purple)

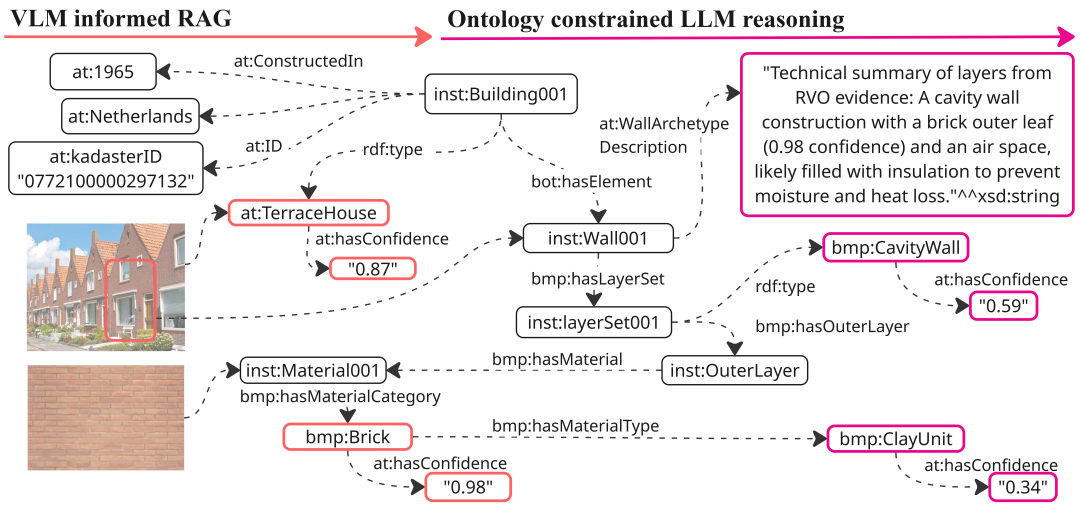


Figure 8: Results representing the enriched semantic graph with the deducted information from the VLM-augmented RAG framework (in orange colour) and the ontology-constrained LLM reasoning (in pink colour)

ment between these sources. The greater the alignment of sources in their descriptive annotations, the greater the likelihood of identifying meaningful semantic textual similarity and correctly assigning ontology classes. A key limitation of the proposed method is that the archetype descriptions require partial manual adjustment and verification to ensure terminological consistency with the ontology, which directly influences the cosine similarity scores in the embedding-based matching process. Therefore, a human-in-the-loop approach is strongly recommended, allowing building engineering experts to iteratively validate and refine LLM-generated results to ensure domain accuracy and methodological robustness.

Conclusion

Effective material-related reuse decisions in existing buildings depend on robust and well-structured documentation, which remains a challenge in current practice. Information retrieval technologies, including VLMs, LLMs, and semantic graphs, have emerged as viable approaches to modelling material information. In this context, the article explores a VLM-augmented RAG framework and ontology-constrained LLM reasoning to integrate unstructured text from building documents. The results show that the proposed VLM-informed RAG pipeline successfully extracts building-archetype text data, whereas ontology-constrained LLM reasoning yields only moderate performance. Therefore, future work will investigate advanced ontology embedding methods to fine-tune LLM reasoning.

References

Byers, B. S., Raghu, D., Olumo, A., De Wolf, C., and Haas, C. (2024). From research to practice: A review on technologies for addressing the information gap for building material reuse in circular construction. Sustainable Production and Consumption, 45:177–191.

Chen, J., Geng, Y., Chen, Z., Pan, J. Z., He, Y., Zhang, W., Horrocks, I., and Chen, H. (2023). Zero-shot and few-shot learning with knowledge graphs: A comprehensive survey. Proceedings of the IEEE, 111(6):653–685.

Chen, X., Liang, C., Yu, A. W., Zhou, D., Song, D., and Le, Q. V. (2019). Neural symbolic reader: Scalable integration of distributed and symbolic representations for reading comprehension. In International Conference on Learning Representations.

Chen, Z., Zhang, Y., Fang, Y., Geng, Y., Guo, L., Chen, X., Li, Q., Zhang, W., Chen, J., Zhu, Y., et al. (2024). Knowledge graphs meet multi-modal learning: A com-

prehensive survey. arXiv preprint arXiv:2402.05391.

Cooper, A., Kato, K., Shih, C.-H., Yamane, H., Vinken, K., Takemoto, K., Sunagawa, T., Yeh, H.-W., Yamanaka, J., Mason, I., et al. (2025). Rethinking vlms and llms for image classification. Scientific Reports, 15(1):19692.

European Commission (2020). Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions: A new circular economy action plan — for a cleaner and more competitive europe. Communication, COM(2020) 98 final CELEX: 52020DC0098, European Commission, Brussels.

Fang, Y., Kuan, K., Lin, J., Tan, C., and Chandrasekhar, V. (2017). Object detection meets knowledge graphs. International Joint Conferences on Artificial Intelligence.

Forth, K., Kaltenegger, J., Petrova, E., and De Wolf, C. (2025). Bim-based material property enrichment using text-based and ontology-based semantic similarity matching.

Gordon, M., Batallé, A., De Wolf, C., Sollazzo, A., Dubor, A., and Wang, T. (2023). Automating building element detection for deconstruction planning and material reuse: a case study. Automation in Construction, 146:104697.

Harispe, S., Ranwez, S., Montmain, J., et al. (2022). Semantic similarity from natural language and ontology analysis. Springer Nature.

Hassija, V., Palanisamy, B., Chatterjee, A., Mandal, A., Chakraborty, D., Pandey, A., Chalapathi, G., and Kumar, D. (2025). Transformers for vision: A survey on innovative methods for computer vision. IEEE Access.

Li, M., Hu, Z., Mohebi, P., Li, S., and Wang, Z. (2026). Enhancing llm-based building data query with chain-of-thought, retrieval-augmented generation, and fine-tuning. Automation in Construction, 182:106738.

Luo, L., Li, Y.-F., Haffari, G., and Pan, S. (2023). Reasoning on graphs: Faithful and interpretable large language model reasoning. arXiv preprint arXiv:2310.01061.

Marin-Garcia, D., Bienvenido-Huertas, D., Carretero-Ayuso, M. J., and Della Torre, S. (2023). Deep learning model for automated detection of efflorescence and its possible treatment in images of brick facades. Automation in Construction, 145:104658.

Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., Mahendran, A., Arnab, A., Dehghani, M., Shen, Z., et al. (2022). Simple open-vocabulary object detection. In European conference on computer vision, pages 728–755. Springer.

Oche, A. J., Folashade, A. G., Ghosal, T., and Biswas, A. (2025). A systematic review of key retrieval-augmented generation (rag) systems: Progress, gaps, and future directions. arXiv preprint arXiv:2507.18910.

Pan, Y., Lu, L., Wang, M., Reja, V. K., Noichl, F., Bormann, A., and Brilakis, I. (2025). Zero-shot learning with vision-language models for estimating building energy efficiency from street view images. In Computing in Civil Engineering 2025: Resilient, Robotic, and Educational Systems, pages 254–262. n.a.

Peerbolte, L. and der Kaa, V. (1927). Leidraad bij het samenstellen of herzien van bouwverordeningen. N. Samsom, Alphen aan den Rijn.

Pei, W., Biljecki, F., and Stouffs, R. (2024). Techniques and tools for integrating building material stock analysis and life cycle assessment at the urban scale: A systematic literature review. Building and Environment, 262:111741.

Peng, J., Hu, X., Huang, W., and Yang, J. (2023). What is a multi-modal knowledge graph: A survey. Big Data Research, 32:100380.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In International conference on machine learning, pages 8748–8763. PmLR.

Rijksdienst voor Ondernemend Nederland (2022). Voorbeeldwoningen 2022 bestaande bouw. Technical report, Rijksdienst voor Ondernemend Nederland (RVO).

Schelling, B. and van Gijn, A. (1904). Leidraad bij het samenstellen van eene Verordening, ter uitvoering van artikel 1 de Woningwet. n.p., 3 edition.

Smeyers, T. and Mertens, M. (2022). Reuse toolkit: The reclamation audit. Technical report, Facilitating the Circulation of Reclaimed Building Elements.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., Zhou, D., et al. (2022). Chain-of-thought prompting elicits reasoning in large language models. Advances in neural information processing systems, 35:24824–24837.

Zeng, Z., Goo, J. M., Wang, X., Chi, B., Wang, M., and Boehm, J. (2024). Zero-shot building age classification from facade image using gpt-4. arXiv preprint arXiv:2404.09921.

Zhang, J., Xiang, R., Kuang, Z., Wang, B., and Li, Y. (2024). Archgpt: harnessing large language models for supporting renovation and conservation of traditional architectural heritage. Heritage Science, 12(1):220.

Zheng, Y., Khalid Masood, M., Seppänen, O., Törmä, S., and Aikala, A. (2023). Ontology-based semantic construction image interpretation. Buildings, 13(11):2812.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

MULTI-CLASS PREDICTION OF OCCUPATIONAL ACCIDENT TYPES USING XGBOOST IN PIPELINE CONSTRUCTION

Ahmet Eren Ünlü[1], Recep Özkan[2], and Onur Behzat Tokdemir[2]
[1]Istanbul Technical University, İstanbul, Türkiye
[2]Turkish-German University, İstanbul, Türkiye
[3]Artificial Intelligence and Data Science Research Center, Istanbul Technical University, İstanbul, Türkiye

Abstract

In this study, the XGBoost algorithm was chosen to predict accident types using data from incidents that occurred in pipeline construction projects, and a multi-class classification approach was presented. A model was trained on 1,184 real accident records obtained from pipeline construction projects. This data includes information such as the time of the accident and the type of treatment required. Classification results showed an overall accuracy of 83%, a macro-average F1 score of 0.82, and a weighted F1 score of 0.81. Some classes showed high recall rates, while recall was low in categories with limited data. The results demonstrate that XGBoost can be a suitable model for predicting accident types using historical incident datasets and can contribute to risk analysis in construction projects through artificial intelligence.

Introduction

Construction sites have the highest number of fatal work injuries of any private industry sector, accounting for approximately one in five occupational deaths (BLS, 2024). Therefore, the construction industry poses a significant risk to both human safety and property (Mostofi & Toğan, 2022). According to the International Labour Organization (ILO, 2023), construction workers in economically developed countries are three to four times more likely to experience fatal workplace accidents compared to those in other sectors. Globally, an estimated 2.93 million workers die annually from occupational accidents and work-related diseases, while an additional 395 million suffer from non-fatal injuries (ILO, 2023). The economic impact of these incidents is substantial, estimated at approximately 4% of global GDP, which is equivalent to nearly $3 trillion in direct and indirect costs (ILO, 2023). While construction work accounts for only about 7% of global employment, it is responsible for a disproportionate 30–40% of fatal occupational injuries (ILO, 2023). Beyond macroeconomic costs, workplace accidents directly impair corporate performance by reducing productivity, operating profit, and sales growth at the firm level (Dong Koo Kim & Park, 2021). In recent years, the growing global demand for transporting energy, water, and chemicals has accelerated the development of pipeline construction projects. However, this expansion has been accompanied by a significant rise in workplace accidents within the sector (Mammadov et al., 2023). Pipeline construction activities are categorized among the highest-risk subsectors, characterized by unique challenges such as exposure to pressurized systems, flammable substances, and geographically remote working environments (Mammadov et al., 2023). Despite these risks, data-driven approaches to accident prediction in pipeline construction remain limited, highlighting the critical need for targeted research in this domain.

Most current studies focus on estimating accident severity or risk levels rather than identifying specific accident types that may occur. However, accident type prediction offers a proactive approach to safety management, as it allows for the prediction of potential events that may occur on-site and the implementation of preventive measures before accidents happen. Therefore, accident type prediction can support more effective safety planning, training, and resource allocation, especially in high-risk environments.

Pipeline construction activities are categorized among the highest-risk subsectors, characterized by unique challenges such as exposure to pressurized systems, flammable substances, and geographically remote working environments (Mammadov et al., 2023). These types of hazards differentiate pipeline construction from other constructions. Despite these distinctive risks, data-driven approaches specifically focused on accident type prediction in pipeline construction remain limited. Therefore, the motivation for this study is to fill this gap by applying a multi-class classification model using XGBoost for accident types occurring in pipeline construction.

Literature Review

In the construction industry, the use of machine learning (ML) and artificial intelligence (AI)-based models as estimation and prediction methods is increasing (Doğan et al., 2022; Koc et al., 2024). Given the persistently high accident rates in the sector, the application of data-driven predictive models has become a practical necessity not only to reduce accident frequency but also to uncover complex causal relationships that traditional statistical methods fail to capture (Alkaissy et al., 2023).

Machine learning can be particularly beneficial because it excels with complex and large datasets (Wang et al.,

2022). As Sarkar and Maiti (2020) mention, tree-based ML models are increasingly used to improve accident prediction and support safety planning. XGBoost is a popular ensemble learning algorithm known for its speed, high predictive accuracy, and ability to handle large or complex datasets (Asselman et al., 2023). For example, Pang et al. (2024) developed an early-warning model for traffic accidents using XGBoost and achieved significant prediction accuracy.

Similarly, Mammadov et al. (2023) predicted accident outcomes in pipeline construction using 11 different ML algorithms across 12 sub-datasets created by considering different event types, outcomes, timings, and causes, demonstrating the value of ML in pipeline safety management. Unlike the present study, that work focused on predicting accident outcomes rather than accident types, with deep learning outperforming other algorithms in most configurations. The study proposed a prescriptive prevention model and provides a methodological basis relevant to the current work. Davoudi Kakhki et al. (2019) compared the performance of SVMs, boosted trees, and Naïve Bayes on over 33,000 compensation claims in the agribusiness sector. Their models achieved accuracy rates between 92% and 98% in predicting injury severity, demonstrating ML's effectiveness in modeling occupational injury outcomes. Similarly, Ayhan and Tokdemir (2019) used a hybrid model integrating artificial neural networks and fuzzy logic to predict construction accident outcomes with up to 84% accuracy. Although their focus was on outcome severity rather than accident types, their work bears methodological relevance to the present study. Recent innovations have turned toward graph-based ML models. Mostofi and Togan (2023) proposed a multi-head graph attention network (GAT) for construction safety prediction, achieving up to 87.1% accuracy. Building on earlier work in which a graph convolutional network achieved 94% accuracy in predicting accident severity outcomes (Mostofi et al., 2022), their approach improves the interpretability of accidents by connecting them, thereby providing a more detailed understanding of workplace accident risks.

In another large-scale study, Choi et al. (2020) analysed over 137,000 accident records from Korea, finding that Random Forest outperformed other algorithms, particularly when class imbalance was addressed with random oversampling, an approach conceptually aligned with the SMOTE (Synthetic Minority Oversampling Technique) technique used in the present study. Luo et al. (2023) achieved 82% accuracy in predicting accident severity using a Random Forest model trained with entropy-based text mining and SMOTE, identifying key risk factors including lack of supervision and inadequate training. More recently, Alsulamy et al. (2026) demonstrated that by including the nature of the incident (NOI) as an explanatory variable, the XGBoost model achieved 89% accuracy in predicting accident severity (SOI).

Yoo et al. (2024) developed a multi-class classification model for nine accident types in Korean construction sites using SMOTE and machine learning algorithms, achieving strong predictive performance and demonstrating that accident type classification is a viable and practically relevant prediction target. Alkaissy et al. (2023) further showed that ensemble methods including XGBoost can effectively classify injury types across large-scale construction accident datasets, highlighting the suitability of boosting algorithms for multi-class safety prediction tasks. Beyond conventional ML approaches, Park et al. (2026) compared ML and large language model (LLM) approaches for accident prediction, concluding that ML-based models remain highly effective for structured classification tasks even as newer language-based methods emerge.

While most of the mentioned studies focused on predicting the severity or consequences of accidents, this research aims to predict accident types using the XGBoost algorithm on a dataset of 1,184 pipeline construction accident records. Pipeline construction projects involve distinctive, dynamic, and complex working conditions that differ significantly from other construction types. Activities involving oil, gas, and chemicals introduce critical OHS risks such as fire, explosion, and chemical exposure, while factors such as limited worker experience, extensive geographic scope, and the high number of stakeholders further elevate accident risks (Mammadov et al., 2023). Despite this, pipeline construction remains a less-researched domain in the safety prediction literature, and this study aims to address this gap. Furthermore, this research incorporates a synthetic oversampling technique to mitigate class imbalance and its negative impacts on prediction performance.

Methodology

XGBoost was chosen in this study due to its scalability, high predictive accuracy, and ability to handle large and complex datasets. It employs a novel sparsity-aware algorithm for sparse data and directly handles missing values, making it robust in real-world applications. The algorithm builds an ensemble of decision trees in an additive manner, incorporating L1 and L2 regularization terms to prevent overfitting and enhance generalization performance. Additionally, XGBoost provides built-in feature importance scores, enabling interpretable insights into the key predictors of accident types (Chen & Guestrin, 2016). It builds models sequentially. Each new tree attempts to learn from the previous tree's errors to improve its results. The general form of the objective function in the t-iteration is defined as follows:

$$\sum_{i=1}^{n} l(y_i, \hat{y}_i{}^{(t-1)} + f_t(x_i)) + \Omega(f_i) \qquad (1)$$

where l is the loss function measuring the difference between the predicted and true labels, ft is the function (decision tree) added at iteration t, and Ω(ft) is the regularization term that penalizes model complexity to avoid overfitting. Unlike traditional gradient boosting methods, XGBoost incorporates second-order derivatives of the loss function to approximate gradients more accurately and directly incorporates regularization into the objective, which enhances generalization. Additionally, techniques such as column subsampling, sparsity-aware learning, and optimized handling of missing values contribute to its superior performance and computational efficiency. The workflow of the study is shown in Figure 1.

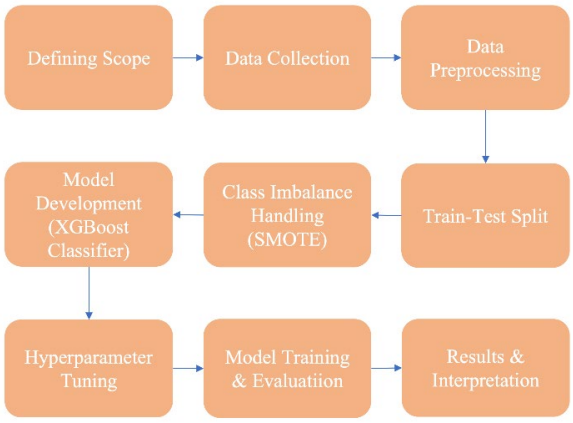


Figure 1: The workflow of the study.

Defining Scope and Data Collection

This study aims to develop a model to predict the types of accidents that may occur during pipeline construction. While results in the literature are generally predicted based on the causes of the accident, this study directly considers the type of accident as the target variable. To create the model, the necessary data were first determined, and then the data collection process was initiated. The data was collected by anonymizing it from a pipeline construction project spanning Asia to Europe, involving numerous countries and stakeholders. The organization of over 5000 unstructured accident reports resulted in 1184 structured accident records. The present study reuses this dataset with a distinct focus, rather than predicting accident outcomes, the current work utilizes root causes as input features to perform multi-class classification for predicting accident types using XGBoost. Each record is labelled with one of six predefined accident types (classes 0-5) and includes various categorical and temporal characteristics such as the accident outcome, primary causes, immediate causes, and time and date of occurrence. The six accident types used as classification targets are derived from event/exposure categories coded according to the OIICS (Occupational Injury and Illness Classification System) (United States. Bureau of Labor Statistics. (2007). The categories reflect the distribution of accident types recorded in pipeline construction projects. The classes and descriptions of the accident types are shown in Table 1.

Table 1: Class Types and Description.

Class	Description
0	Bodily reaction and exertion (Injuries resulting from physical stress, overexertion, heavy lifting, or repetitive motions)
1	Contacts with objects and equipment (Incidents where a worker is struck by, caught in, or crushed by moving machinery, equipment, or falling materials)
2	Exposure to harmful substances or environments (Incidents involving exposure to toxic chemicals)
3	Falls (Slips and trips on the same level, or falls from a higher level)
4	Fires and explosions (Accidents caused by the ignition of fiery gases)
5	Transportation accidents (Accidents involving vehicles in a construction zone)

Data Preprocessing

The data used included time information such as the hour (0–23), month (1–12), and day of the week (0: Monday – 6: Sunday). In addition, the accident's outcome (e.g., injury status), root causes (RC), and immediate causes (ICF) were also included in the model. These causes were coded according to the World Health Organization (WHO, 2001) classification system. All data was converted into numerical forms. Categorical variables and textual expressions were numerically encoded to facilitate processing by the algorithm. Missing data was removed to prevent any adverse impact on the model's results. The detailed descriptions and data types of the input features used in the proposed model are summarized in Table 2.

Table 2: Input Features.

Feature Name	Description	Data Type
Time_block	Time of the accident categorized into 4 blocks (Morning, Afternoon, Evening, Night)	Categorical
Season	Season of the accident (Winter, Spring, Summer, Autumn)	Categorical
Weekday	Day of the week (0: Monday - 6: Sunday)	Categorical
F/Injury /ADPD/NM	The immediate outcome/severity of the accident	Categorical
ICF1, ICF2	Immediate causes of the accident (based on WHO classification)	Categorical
RC1, RC2	Root causes of the accident (based on WHO classification)	Categorical

Handling Class Imbalance with SMOTE

During the analysis, it was observed that the dataset was unbalanced, meaning the available accident data was not evenly distributed, and therefore some types were underrepresented. To prevent data leakage, SMOTE (Synthetic Minority Oversampling Technique) was applied exclusively to the training subset (80%) after the train-test split. Specifically, the minority classes, such as Fires and Explosions (Class 4) and Transportation Accidents (Class 5), were oversampled to match the instance count of the majority class. Since the minority classes contained very few instances, k_neighbors were

set to 1 to ensure SMOTE could be applied without error, as a larger neighborhood size would exceed the available sample count in these classes, thereby ensuring the XGBoost model learns unbiased decision boundaries.

Hyperparameter Tuning

To obtain the results, the hyperparameters were first adjusted. One key parameter is the maximum depth, which determines how deep each decision tree can go. Deeper trees can learn more complex patterns, but they can also lead to overfitting. The learning rate determines how much each tree contributes to the overall prediction. A lower learning rate increases learning time but generally yields more consistent and accurate results. The N_estimators determine how many trees the model will build. Generally, using more trees improves performance but also increases training time. Finally, the subsampling parameter determines the proportion of training data used to build each tree. Using less than the full data adds randomness to the training process and reduces overfitting. Another important hyperparameter is `colsample_bytree`, which specifies the proportion of randomly selected features per tree. Using only a portion of the features allows the model to generalize better and reduces the likelihood of overfitting.

Several key XGBoost hyperparameters were selected to improve model performance. The maximum tree depth was adjusted to balance complexity and overfitting. A low learning rate was used to ensure reliable model learning. The number of estimators was appropriately chosen to increase accuracy without significantly increasing training time.

In addition, subsample and colsample_bytree settings were used to introduce randomness and improve generalization.

Model training was initially performed with the default hyperparameters, followed by a limited grid search to optimize them. The best model performance was obtained with the following parameter configuration: maximum tree depth 6, learning rate 0.1, N_estimators 200, subsampling rate 0.8, and column sampling rate per tree 0.8. The hyperparameter configuration established for the XGBoost model to balance complexity and prevent overfitting is presented in Table 3.

Table 3: Hyperparameters.

Hyperparameter	Value	Description
Max_depth	6	Maximum depth of a tree. Balances model complexity and overfitting.
Learning_rate	0.1	Step size shrinkage used in update to prevent overfitting.
N_estimators	200	Number of gradient boosted trees.
Subsample	0.8	Subsample ratio of the training instances.
Colsample_bytree	0.8	Subsample ratio of columns when constructing each tree.
Objective	Multi: softmax	Learning objective for multi-class classification.

Model Training and Evaluation

The final XGBoost model was trained on the preprocessed dataset following SMOTE-based augmentation. The model was trained with an 80%-20% training-to-test split. The model's performance was evaluated using metrics such as accuracy, recall, F1 score, and precision, which indicate how well the model predicted each type of accident. All analyses were conducted in Google Colab using Python. The following libraries were used: pandas and numpy for data processing, scikit-learn for model evaluation and train-test splitting, imbalanced-learn for SMOTE implementation, and xgboost for model training and prediction. Visualizations were generated using matplotlib and seaborn.

Results and Discussion

Precision, recall, F1-score, and overall accuracy metrics were used to obtain and evaluate the results. Precision measures how accurate the model's positive predictions are. Recall indicates how well the model captures all real samples of a class. The F1 score is the harmonic mean of precision and recall. It provides a balanced measure when both false positives and false negatives are significant. It is particularly useful when a single measure is needed that considers both precision and recall.

In accident type prediction using the XGBoost model, True Positive (TP) refers to correctly predicted accident types, False Positive (FP) refers to accident types that did not occur but were incorrectly predicted, and False Negative (FN) refers to actual accident types that the model could not predict; these terms are necessary to calculate recall, precision, and the F1 score to evaluate classification performance. The equations for calculating these metrics are given in Equations 2, 3 and 4.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3}$$

$$\text{F1-Score} = \frac{2 \text{ x Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}} \tag{4}$$

This study categorized 1184 pipeline construction accidents into 6 main types. We measured the XGBoost Classification algorithm's performance, which reached 83% overall accuracy. The macro average F1 score was 0.82, and the weighted F1 score was 0.81. The classification performance of the model across all classes is graphically represented in Figure 2.

In addition to the overall evaluation, the results obtained for each accident class were examined separately. The class in which the model was most successful in classification was Class 4, which refers to Fires and Explosions. In this class, the model achieved a precision of 0.96, a recall of 1.00, and an F1 score of 0.98. Fires and explosions are among the high-risk categories in pipeline

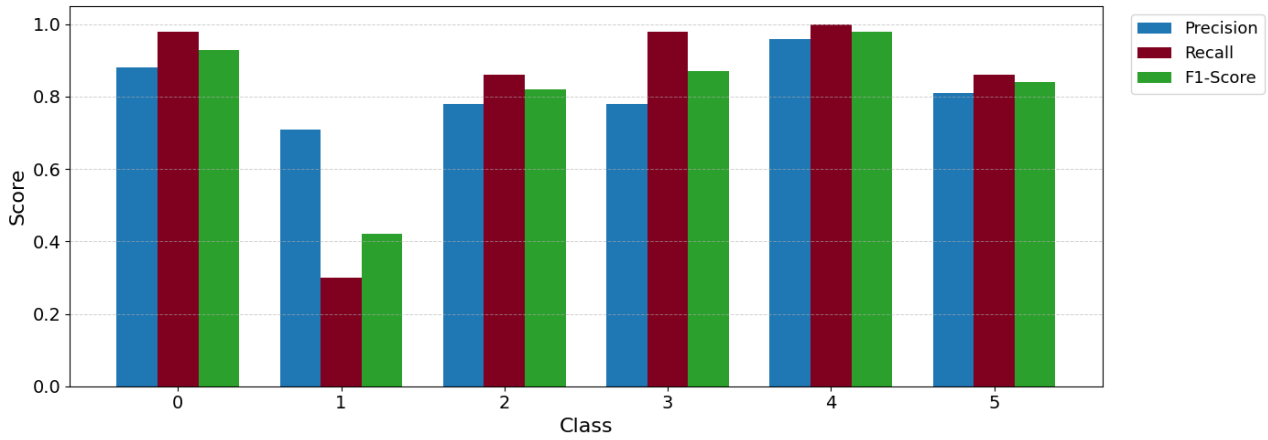


Figure 2: Class-wise classification performance of the XGBoost model

construction, and the model's successful predictive performance in this class is a noteworthy contribution to site safety.

Similarly, the recall value of 0.98 obtained for Class 3 (Falls), one of the most common and dangerous types of accidents in the construction industry, demonstrates that the model has an extremely low accident-risk skipping rate and offers a reliable monitoring mechanism. In this class, the precision was 0.78, and the F1 score was 0.87. Compared to binary classifications, the increase in the number of classes the model can choose in multi-class classifications brings with it the risk of misclassification. Therefore, achieving these metric values among 6 classes is a remarkable achievement.

When examining the results for Class 0 (Bodily reaction and exertion), the precision value was 0.88, while the recall and F1 score metrics were 0.98 and 0.93, respectively. The results show that the model can offer a reliable monitoring mechanism in this classification as well. The recall value of 0.86 and the F1 score of 0.84 obtained for Class 5 (Transportation accidents) show that the model can also distinguish these types of accidents with high accuracy.

Despite these successful results, the model's most significant weakness is its recall of 0.30 and F1 score of 0.42 in the Class 1 (Contact with Objects and Equipment) category. An examination of the predictions revealed that actual Class 1 instances were frequently misclassified as Class 3 (Falls). This misclassification largely stems from the fact that these accident types often share similar underlying causes or occur simultaneously on-site, which blurs the decision boundaries for the algorithm. Furthermore, since the SMOTE technique was primarily targeted at extreme minority categories (such as Classes 4 and 5) to match the majority class, the inherent data limitations of Class 1 remained. Combined with its lack of distinct predictive features, this prevented the model from effectively learning the unique patterns required to separate Class 1 from more dominant, similar classes like Class 3.

These results demonstrate that models like XGBoost can be used as a powerful monitoring tool for pipeline construction safety. Still, their performance may be inadequate for accidents with characteristics similar to other classes, such as Class 1. In future studies, working with more and diverse datasets in pipeline construction and testing the model for construction accidents across different areas will be meaningful for evaluating the model's success.

Compared to traditional safety analysis methods that typically rely on descriptive statistics or severity-based assessments, the proposed approach allows safety managers to understand accident patterns more proactively by identifying the likely type of incident before it occurs, helping project owners complete the project with less risk.

One of the most important strengths of this study is its focus on accident type prediction rather than accident severity prediction. While severity prediction is useful for post-incident assessment, predicting accident types provides more actionable information for safety planning, training programs, and preventive strategies. Specifically, identifying the most likely accident categories allows safety managers to allocate resources more effectively and design targeted interventions for high-risk activities.

Alongside these strengths, the study also has some limitations. The use of SMOTE to address class imbalance adds synthetic samples to the training data. While SMOTE was applied only to the training subset to prevent data leakage, the generated samples may not fully represent real-world accident scenarios and could introduce bias to the model. Although the study includes accident types from different construction sites, it focuses on a defined set of accident categories. While these categories reflect real-world classifications, future research could examine a broader or more detailed set of accident types to enhance the model's comprehensiveness.

Finally, although the model demonstrates strong classification performance, its practical application in real construction environments requires integration with real-time data collection systems and validation using live project data. Future work should focus on testing the model in operational environments and integrating it with

digital construction technologies such as BIM and IoT-based monitoring systems.

Overall, this work lays a foundation for the development of proactive, data-driven safety management tools in pipeline construction and highlights the potential of machine learning approaches to improve occupational safety outcomes.

Conclusion

This study investigated the success of the XGBoost model, trained on 1184 real-world accident data from pipeline construction sites, in predicting accident types using a multi-class classification method. Given the potential for both material and non-material damage from accidents in the construction sector, the study focused on predicting these types of accidents. Accident types were numerically translated from text and categorized into six classes. The model was trained on the accident data. Despite the relatively small dataset, the model achieved good prediction performance with 83% accuracy and a macro F1 score of 0.82. These findings validate the potential of machine learning approaches, such as XGBoost, for occupational safety analysis in the construction industry. The proposed model can be particularly beneficial for safety managers, site supervisors, and project owners by enabling automated classification of incident reports and helping uncover complex relationships between root causes and specific accident types, thereby strengthening post-incident analysis and future safety planning. Future studies could validate the model using larger, more diverse datasets in the pipeline field. It is also possible to make predictions on similar datasets using different ML algorithms. Furthermore, applying this model across different sub-sectors of the construction industry could yield interesting results. The current results provide a reliable reference point for ongoing developments in AI-assisted accident prevention.

Acknowledgments

Recep Özkan gratefully acknowledges the support received from the Scientific and Technological Research Council of Turkey (TÜBİTAK) through the 2211-[A] National PhD Scholarship Program.

References

Alkaissy, Maryam & Arashpour, Associate Prof. Mehrdad & Golafshani, Emad & Hosseini, M. Reza & Khanmohammadi, Sadegh & Bai, Yu & Feng, Haibo. (2023). Enhancing construction safety: Machine learning-based classification of injury types. Safety Science. 162. https://doi.org/106102. 10.1016/j.ssci.2023.106102.

Alsulamy, S., Alshayeb, M., Inam, I., & Ahmed, A. (2026). Machine learning applications for predicting safety incidents in construction industry. Scientific Reports. https://doi.org/10.1038/s41598-025-34763-0

Asselman, A., Khaldi, M., & Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 31(6), 3360-3379. https://doi.org/10.1080/10494820.2021.1928235.

Ayhan, B. U., & Tokdemir, O. B. (2019). Predicting the outcome of construction incidents. Safety science, 113, 91-104. https://doi.org/10.1016/j.ssci.2018.11.001.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794). https://doi.org/10.1145/2939672.2939785.

Choi, J., Gu, B., Chin, S., & Lee, J. S. (2020). Machine learning predictive model based on national data for fatal accidents of construction workers. Automation in Construction, 110, 102974. https://doi.org/10.1016/j.autcon.2019.102974.

Doğan, N. B., Ayhan, B. U., Kazar, G., Saygili, M., Ayözen, Y. E., & Tokdemir, O. B. (2022). Predicting the cost outcome of construction quality problems using case-based reasoning (CBR). Buildings, 12(11), 1946. https://doi.org/10.3390/buildings12111946.

International Labour Organization. (2023). *Nearly 3 million people die of work-related accidents and diseases.* https://www.ilo.org/resource/news/nearly-3-million-people-die-work-related-accidents-and-diseases

Kakhki, F. D., Freeman, S. A., & Mosher, G. A. (2019). Evaluating machine learning performance in predicting injury severity in agribusiness industries. Safety science, 117, 257-262. https://doi.org/10.1016/j.ssci.2019.04.026.

Kim, D. K., & Park, S. (2021). An analysis of the effects of occupational accidents on corporate management performance. Safety science, 138, 105228. https://doi.org/10.1016/j.ssci.2021.105228.

Koc, K., Budayan, C., Ekmekcioğlu, Ö., & Tokdemir, O. B. (2024). Predicting cost impacts of nonconformances in construction projects using interpretable machine learning. Journal of Construction Engineering and Management, 150(1), 04023143. https://doi.org/10.1061/JCEMD4.COENG-13857.

Luo, X., Li, X., Goh, Y. M., Song, X., & Liu, Q. (2023). Application of machine learning technology for occupational accident severity prediction in the case of construction collapse accidents. Safety science, 163, 106138. https://doi.org/10.1016/j.ssci.2023.106138.

Mammadov, A., Kazar, G., Koc, K., & Tokdemir, O. B. (2023). Predicting accident outcomes in cross-border pipeline construction projects using machine learning algorithms. Arabian Journal for Science and Engineering, 48(10), 13771-13789. https://doi.org/10.1007/s13369-023-07964-w

Mostofi, F., & Toğan, V. (2023). Construction safety predictions with multi-head attention graph and sparse

accident networks. Automation in Construction, 156, 105102. https://doi.org/10.1016/j.autcon.2023.105102

Mostofi, F., Toğan, V., Ayözen, Y. E., & Tokdemir, O. B. (2022). Construction safety risk model with construction accident network: A graph convolutional network approach. Sustainability, 14(23), 15906. https://doi.org/10.3390/su142315906

Pang, Y., Wang, X., Tang, Y., Guobin, G., Chen, J., Yang, B., ... & Wen, X. (2024, November). A traffic accident early warning model based on XGBOOST. In Fourth International Conference on Advanced Algorithms and Neural Networks (AANN 2024) (Vol. 13416, pp. 494-499). SPIE. https://doi.org/10.1117/12.3049832.

Park, S., Jeong, K., Kim, J., Kim, H., Yoon, S., Kim, J., & Jeon, J. (2026). Comparative evaluation of machine learning and LLM-based approaches for construction accident prediction. Journal of Asian Architecture and Building Engineering, 1-16. https://doi.org/10.1080/13467581.2026.2622236

Sarkar, S., & Maiti, J. (2020). Machine learning in occupational accident analysis: A review using science mapping approach with citation network analysis. Safety science, 131, 104900. https://doi.org/10.1016/j.ssci.2020.104900.

United States. Bureau of Labor Statistics. (2007). *Occupational injury and illness classification manual*. Bureau of Labor Statistics.

U.S. Bureau of Labor Statistics. (2024). *Census of Fatal Occupational Injuries Summary, 2023*. U.S. Department of Labor. https://www.bls.gov/news.release/cfoi.nr0.htm

Wang, C. C., Kuo, P. H., & Chen, G. Y. (2022). Machine learning prediction of turning precision using optimized XGBoost model. Applied Sciences, 12(15), 7739. https://doi.org/10.3390/app12157739.

World Health Organization. (2001). *International Classification of Functioning, Disability and Health (ICF)*. World Health Organization. https://iris.who.int/handle/10665/42407

Yoo, J. W., Park, J., & Park, H. (2024). Enhancing safety of construction workers in Korea: An integrated text mining and machine learning framework for predicting accident types. International journal of injury control and safety promotion, 31(2), 203-215. https://doi.org/10.1080/17457300.2023.2300424

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

PHYSICS-CONSTRAINED REINFORCEMENT LEARNING FOR INDIVIDUALIZED AND BIOMECHANICALLY FEASIBLE TASK-MOTION OPTIMIZATION

Kangrui Ren[1], Ali Golabchi[2], and Gaang Lee[1]
[1]University of Alberta, Canada
[2]Department of Mechanical Engineering, University of Alberta, Canada

Abstract

The construction industry faces high musculoskeletal disorder rates due to repetitive, awkward task-postures. Current manual corrections and static optimizations often ignore individual anthropometry and whole-body biomechanical interdependencies, leading to physically unfeasible results. This study explores Reinforcement Learning (RL) for tailored posture optimization by developing a pipeline that overcomes technical barriers like reward instability and high computational costs. The technique predicts joint-space residuals via a distilled, risk-aware RL policy within a physics simulator. Testing in construction-like scenarios showed significantly reduced mean risk and high-risk spikes while maintaining task fidelity. These findings demonstrate that RL can provide personalized, executable ergonomic interventions.

Introduction

The construction industry continues to experience a high prevalence of work-related musculoskeletal disorders (WMSDs), imposing a substantial burden on both worker health and operational productivity (Santos et al., 2025). These disorders are rarely the result of a single event; rather, they accumulate through sustained or repetitive exposure to risk factors such as awkward postures, forceful exertions, and repetitive motions during task performance (Hulshof et al., 2021). Since these contributing factors typically emerge from the complex interaction between poor work design and a worker's habitual postural patterns, effective mitigation requires not only work-redesign interventions but also, critically, systematic correction of workers' working postures during task performance.

Current industrial practices for assessing and correcting task postures rely heavily on manual processes. Assessment is typically conducted using self-reports or observational tools such as the Rapid Upper Limb Assessment (RULA) (McAtamney and Corlett, 1993) or the Rapid Entire Body Assessment (REBA) (Hignett and McAtamney, 2000). In practice, correction strategies are often limited to verbal, on-the-spot instructions delivered by safety managers. While intuitive, these local corrections often fail to account for whole-body biomechanical interdependencies; correcting a local issue (e.g., lowering an elbow) may inadvertently create or worsen risk elsewhere (e.g., increasing lower back torque) due to inherent human biomechanical coupling and task constraints (Belbeck et al., 2014; Nadon et al., 2018; Rören et al., 2024). Because these interventions depend on generic ergonomic guidelines, they often fail to consider individual anthropometric differences and habitual motion patterns, thereby making it difficult to provide posture corrections that are truly personalized and readily adoptable (Ji et al., 2023; Chen and Zhang, 2025). These limitations underscore a need for a systematic technique to generate biomechanically coherent, coordinated whole-body posture corrections that are customized to the individual's specific anthropometric characteristics within the given task context.

To derive whole-body-coordinated posture corrections, researchers have applied mathematical optimization, such as trajectory optimization (optimal control) (Dafarra et al., 2022), and inverse optimization (Westermann et al., 2020). However, these existing approaches often simplify whole-body risk as an additive sum of per-joint or angle-based scores in their equations, which limits their ability to account for the coupled constraints across the body (González-Alonso et al., 2025). By failing to enforce these coupled constraints, such approaches overlook critical factors such as load transfer, support stability, and contact maintenance (Zheng and Yumane, 2015), often resulting in "optimizations" that are non-actionable or physically impossible. Furthermore, on the optimization algorithm side, current approaches often rely on static, frame-based reasoning that does not account for individual anthropometric variabilities or the temporal flow of motion (Kee, 2022; Ciccarelli et al., 2026), prioritizing a raw "risk" score reduction without considering feasibility, motion smoothness, or workflow continuity.

Reinforcement learning (RL) offers a promising alternative to overcome these limitations by addressing the complexities of human motion that static optimization misses. Unlike frame-based techniques, RL is uniquely positioned to learn full-body coordinated actions, allowing it to discover posture adjustments that remain biomechanically feasible at the global level. What frame-based techniques miss is temporal coupling: feasibility constraints (stability, contact evolution, inter-joint coordination) are trajectory-level, not per-frame. It can also condition learned policies on individual user attributes, enabling the generation of personalized motion strategies rather than generic templates. This is because RL optimizes expected cumulative reward over time, not a single-step objective: the agent

is trained on long-horizon rollouts where feasibility violations and risk spikes occur downstream, enabling credit assignments that link later outcomes back to earlier posture choices. With a state representation that includes task context and user attributes, the learned policy becomes both temporally aware (trajectory-level) and personalized (user-conditioned) in a single decision-making loop. Crucially, RL handles sequential, history-dependent decision-making (Mnih et al., 2015; Schulman et al., 2015), which allows for the optimization of entire motion trajectories rather than isolated frames. By encoding multiple objectives within the reward function, RL agents can learn balanced posture strategies that reflect necessary real-world trade-offs between ergonomic risk, task efficiency, and motion smoothness.

Despite this theoretical promise, RL has not yet been sufficiently tested or validated as a technique for individually tailored task-posture optimization due to several technical barriers. A primary challenge is risk signal reliability and reward stability; MSD risk signals and video-based skeletal data can be noisy or non-smooth, often causing unstable learning processes or the development of "shortcut" behaviors that game the reward system without truly reducing risk (Ren, 2025). Furthermore, ensuring executability and safety remains difficult, as safety is often treated via soft penalties in RL; consequently, policies may violate strict joint limits, balance requirements, or contact constraints, resulting in failure during physics-based execution. The temporal nature of the problem also introduces long-horizon credit assignment issues, where risk reduction depends on early preparatory actions (e.g., foot placement) that are difficult to reinforce due to accumulated drift over long subtasks. Finally, the sample and computational costs of physics-based long-horizon RL might be overwhelming, requiring extensive rollouts that make training and tuning prohibitively expensive.

To address these gaps, this study aims to explore the potential of RL for individually tailored task-posture optimization by developing and testing the feasibility of a pipeline that effectively mitigates the technical barriers.

Proposed Posture Optimization Technique

Figure 1 provides an overview of the proposed RL-based individually tailored and biomechanically feasible posture optimization pipeline.

Problem Formulation

Individualized task-posture correction is formulated as a long-horizon, safety-constrained, multi-objective RL problem executed in a physics simulator. Rather than synthesizing motion from scratch, the policy predicts bounded joint-space residuals relative to a demonstration trajectory. At each control step, the action is applied as a joint-wise offset to the current demonstration pose target (i.e., the reference pose used as the tracking setpoint) and is clipped using conservative per-joint limits, ensuring that the corrections remain local and physically trackable by the simulated body. The corrected target is executed in a contact- and stability-aware simulator, and the post-contact state is evaluated to generate a risk signal for reward shaping.

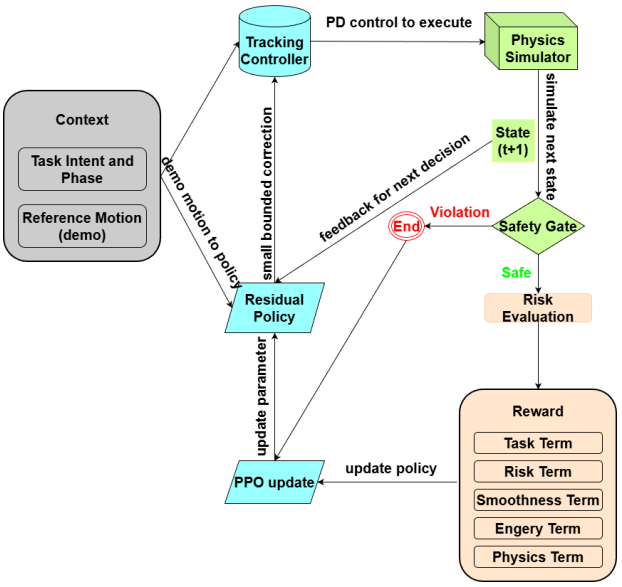


Figure 1: The proposed RL-based individually tailored and biomechanically feasible posture optimization pipeline.

To support distribution-aware risk shaping, the framework models both expected risk and the probability of high-risk events using a lightweight predictor distilled from a richer teacher model, making it practical for on-policy training. Safety is enforced through hard terminations triggered by simulator- and biomechanics-based checks (e.g., joint-limit violations, balance or contact failures, or spinal overload). Any such violation immediately terminates the episode and incurs a terminal penalty, thereby restricting exploration to physically executable behaviors. Finally, to discourage unrealistic reward shortcuts, policy updates are regularized toward a pretrained motion diffusion prior through a KL-divergence regularizer, encouraging outputs to remain within the manifold of human motion.

State Representation

A standardized kinematic encoding with explicit temporal context is used for state representation. All kinematic quantities are expressed in a consistent world frame and normalized using dataset-level statistics, while velocities are scaled by the control timestep to keep magnitudes comparable across terms. Each per-frame state s_t includes: (i) the root pose (position and orientation) and root linear/angular velocities; (ii) per-joint local rotations and angular velocities; and (iii) task-context features, including the goal pose, a phase ID (represented as a one-hot stage token), and a short reference snippet from the demonstration motion. The phase ID provides explicit stage information for indexing time-varying reference targets and disambiguating similar poses across different phases, which helps stabilize long-horizon imitation-style control (Peng

et al., 2018). Goal features are encoded as transforms relative to the pelvis (translation + 6D rotation). The reference snippet serves as a local motion anchor (e.g., a short horizon of target joint poses/velocities) to preserve task intent, while the policy is restricted to bounded corrections rather than free-form resynthesis.

Joint rotations are represented using the continuous 6D rotation encoding, defined by the first two columns of the rotation matrix, which avoids the singularities and discontinuities associated with Euler angles during learning. To provide short-term temporal context for contact and stability transitions and delayed biomechanical responses, the policy is conditioned on a stacked history window, where t indexes control timestep, H denotes the number of stacked frames, and s_k is the per-frame state at timestep k

$$\tilde{s}_t \triangleq [s_{t-H+1}, s_{t-H+2}, \ldots, s_t], \qquad H = 32. \tag{1}$$

With a 30 Hz control rate, $H = 32$ provides approximately 1.07 s of context while allowing the policy to remain feed-forward for stable on-policy learning. When personalization is enabled, the observation is augmented with subject descriptors (e.g., anthropometric measurements, range of motion indicators, and strength proxies), enabling capability-aware corrections rather than one-size-fits-all adjustments.

Action Representation

Actions are defined as continuous joint-space residuals (pose increments) bounded by joint-specific limits to enforce conservative, anatomically plausible corrections. Bounding is implemented through element-wise clipping (or tanh squashing) within predefined per-joint ranges. At each control step, the correction is added to the demonstration pose target and tracked by a stabilized PD controller in the physics simulator. Tracking is performed in joint space under torque limits and damping regularization to reduce overshoot and improve stability during contact switching. These bounds prevent implausible pose jumps and confine the corrections to a local feasible regime.

Given (s_t, a_t), where a_t is the pose correction applied at control timestep t, the simulator executes the correction-adjusted tracking command and advances the system:

$$s_{t+1} = F(s_t, a_t), \tag{2}$$

where $F(\cdot)$ denotes one physics rollout step, including contact resolution and internal integration substeps. After the transition, the risk student is queried using the updated history window $\tilde{s}_{t+1}$ (which includes the realized s_{t+1}). Evaluating risk after contact resolution and stabilization ties the reward signal to physically realized outcomes rather than to a pre-contact kinematic estimate and yields more consistent credit assignment when small corrections trigger compensation.

Risk-Awareness via Student–Teacher Distillation

The diffusion-based teacher builds upon the prior multimodal dynamic MSD-risk assessment framework, which fuses (i) skeleton dynamics, (ii) rule-based ergonomic indicators such as REBA-related cues and L5/S1-related loading proxies, (iii) visual-context features, and (iv) subject-related factors into a distributional risk representation. In the present pipeline, this upstream teacher provides supervisory targets for the mean risk and the probability of high-risk exceedance used in RL reward shaping. The upstream MSD risk evaluator represents uncertainty through a diffusion-based head and outputs a distribution over a scalar risk intensity u. Here, u denotes a normalized MSD-risk score derived from the upstream multimodal risk model. To make this risk signal numerically stable and comparable across motions and subjects during RL training, the raw teacher-side risk outputs are mapped to $[0, 1]$ using a monotonic calibration function fitted on the teacher training set. This mapping preserves the relative ordering of risk levels while placing them on a common scale for reward computation. Mean risk alone can obscure brief but hazardous spikes; therefore, the reward also penalizes the probability of exceeding a high-risk threshold. A tail event is defined using a fixed high-risk threshold $u_{\mathrm{hi}} = 0.8$, and the exceedance probability $\rho_t = \Pr(u \geq u_{\mathrm{hi}})$ is computed. This tail probability is explicitly penalized to discourage postures that retain non-negligible probability mass in the high-risk regime, thereby inducing risk-averse behavior. The threshold is selected to represent an upper-risk regime under the calibrated scale while keeping tail events sufficiently frequent for stable learning.

Under physics-based execution, contact events and safety terminations can abruptly end rollouts and amplify reward noise over long horizons. Proximal Policy Optimization (PPO) is adopted as the learning backbone because it provides stable, reliable optimization for long-horizon continuous control in such settings. However, PPO is on-policy and requires frequent rollout collection with step-wise reward evaluation, making the per-step cost of computing the risk signal a practical bottleneck. Estimating the tail probability via diffusion sampling at every environment step would be computationally prohibitive in the online PPO loop. This cost is removed from online training via teacher–student distillation: offline, a diffusion teacher produces distributional supervision targets, while online, a lightweight student predicts only the risk statistics required by the RL objective:

$$(\hat{\mu}_t, \hat{\rho}_t) = \mathrm{RiskStudent}(\tau_{t+1}). \tag{3}$$

This is sufficient because the RL objective depends on the expected risk level and the probability of entering the high-risk regime, rather than the full risk distribution. Here, $\hat{\mu}_t$ approximates the teacher mean risk and $\hat{\rho}_t$ approximates the teacher tail probability. The student is trained offline by MSE regression to teacher-derived targets, where the teacher's μ_t and ρ_t are estimated from diffusion samples per window (e.g., $S = 8$). During RL rollouts, diffusion sampling is not performed; the diffusion teacher is used only for evaluation and ablation studies.

Multi-Objective Reward & Safety Constraints

A unified multi-objective reward is optimized under hard safety enforcement:

$$r_t = \sum_{\star \in \Omega} w_\star(t) R_\star(t) - \lambda c_t, \qquad \Omega = \{\text{task, risk, smooth, energy, physics}\}. \tag{4}$$

The weights follow a fixed, deterministic curriculum, and λ is a safety multiplier. The curriculum emphasizes task fidelity and smoothness early in training (to preserve intent and stability) and gradually shifts weight toward risk reduction once reliable tracking is achieved. The indicator $c_t \in \{0, 1\}$ denotes a safety violation; when $c_t = 1$, the episode terminates immediately with an explicit terminal penalty. In implementation, the safety penalty is applied as a terminal cost with magnitude $P_{\text{term}} = 100$ (i.e., $-\lambda c_t$ is incurred only at termination).

The risk objective penalizes both the expected risk and the high-risk exceedance probability:

$$R_{\text{risk}}(t) = -\hat{\mu}_t - \alpha_{\text{tail}} \hat{\rho}_t. \tag{5}$$

where $\hat{\mu}_t$ is the predicted mean risk and $\hat{\rho}_t$ is the tail probability under the calibrated risk distribution.

The remaining objectives promote task fidelity and motion quality. The task term penalizes deviations from task intent and reference execution (e.g., goal-reaching errors, phase-dependent pose constraints, and trajectory deviation from the demonstration). The smoothness term discourages abrupt velocity and acceleration changes to suppress jerk-like corrections. The energy term penalizes actuation effort (e.g., the squared torque norm). The physics-realism term penalizes simulator-flagged artifacts such as stance-foot slip and excessive foot skating. Each term is normalized to a consistent scale before weighting to keep objectives comparable.

Safety termination is triggered by joint-limit exceedance, L5/S1 overload (excessive compressive loading at the L5/S1 junction, commonly used as a proxy for low-back injury risk), or balance/contact failure. Balance/contact failure is detected using simulator-resolved events such as falls, loss of required support contacts, or the center of mass leaving the support polygon. The L5/S1 load used for termination is computed by the physics/biomechanics engine, while the learned risk model is used only for reward shaping.

Training Loop

Training alternates between physics rollouts and policy updates. Offline, a diffusion-based risk teacher is trained and distilled into a lightweight risk student, and a separate motion diffusion model is trained as a realism prior. Online, each PPO iteration collects N episodes in the physics simulator using the stacked history input τ_t. After each transition, the risk student is queried on the updated window τ_{t+1} (which includes the realized s_{t+1}) and outputs $(\hat{\mu}_{t+1}, \hat{\rho}_{t+1})$ for distribution-aware reward shaping; hard violations trigger immediate termination with the terminal penalty. For brevity, the post-transition estimates $\text{RiskStudent}(\tau_{t+1})$ are denoted as $(\hat{\mu}_{t+1}, \hat{\rho}_{t+1})$. The policy is then updated using PPO for continuous control, with an additional KL regularizer between the policy distribution $\pi_\theta(a_t \mid \tau_t)$ and the prior distribution $\pi_{\text{prior}}(a_t \mid \tau_t)$, encouraging corrections to remain close to natural motion while still enabling risk reduction. The value function is updated by regression. Throughout training, a fixed deterministic curriculum schedules the objective weights $w_\star(t)$ to improve reproducibility and prevent weight oscillations under noisy on-policy returns.

Tests of the Proposed Technique

The framework was evaluated in physics-based, construction-like scenes along three criteria: (i) MSD risk reduction, measured in a distribution-aware manner (both the average risk level and the frequency of high-risk spikes via exceedance probability), (ii) task and motion preservation, measured by task-fidelity errors and motion-quality statistics relative to the reference motion, and (iii) safety and feasibility, measured by violation rates and rollout success/failure.

Test Setup

All evaluations were performed on a desktop with an NVIDIA RTX 5090 (32 GB VRAM), 64 GB RAM, and an Intel Core Ultra 7 265K CPU. Policies were trained in PyTorch using PPO, with rollouts executed in Isaac Lab (NVIDIA Omniverse Isaac Sim / PhysX). The simulator used $\Delta t_{\text{phys}} = 1/120$ s with decimation $d = 4$, yielding a 30 Hz control rate ($\Delta t_{\text{ctrl}} = 1/30$ s) and an episode horizon of $H = 240$ control steps (8 s). The default Isaac Lab HumanoidAMP configuration ($n_q = 28$ actuated DOFs) was used with joint limits enabled; reference targets were tracked via stabilized PD control with torque limits, using the default HumanoidAMP gains fixed across methods for fair comparison.

Results were averaged over 5 training seeds. Each seed was evaluated under 40 randomized initializations (200 runs total); metrics were averaged within each seed and reported as mean ± standard deviation (SD) across seeds. PPO updates used 4096 environment steps per iteration collected from parallel rollouts; episodes could terminate early due to hard-safety constraints, so sampling continued until the step budget was reached. Standard continuous-control PPO hyperparameters were used ($\gamma = 0.99$, $\lambda_{\text{GAE}} = 0.95$, $\epsilon = 0.2$, learning rate 3×10^{-4}, 10 epochs per update, minibatch size 256) to avoid tuning to a specific scene configuration. The risk student was queried at each timestep on the post-transition window τ_{t+1} (including the realized s_{t+1}) for reward shaping, with negligible overhead relative to physics simulation.

Test Data

Reference motion clips were drawn from AMASS (Mahmood et al., 2019), a large unified corpus of MoCap sequences represented in a common model parameterization (e.g., SMPL), which provides large-scale MoCap sequences in a standardized body-model parameterization (e.g., SMPL) for consistent kinematic processing and retargeting. Action semantics and temporal segmentation were obtained from BABEL (Punnakkal et al., 2021), which supplies sequence- and frame-level action labels aligned with AMASS motions. Using BABEL labels, clips relevant to construction-like activities were selected (e.g., turning locomotion, bending/stooping, crouching, reaching, and overhead arm motion). Sequences were resampled to 30 Hz, segmented into 8 s clips (240 frames), and retargeted to the simulator humanoid via offline constrained inverse kinematics with joint-limit enforcement (joint mapping, scale normalization, and root alignment). Root motion was normalized (heading alignment and floor-height correction) to enable consistent reuse across scene layouts and randomized initializations.

Test Configuration

Episodes were initialized from reference clips and executed in parameterized construction-like layouts (e.g., corridor-to-corner navigation and low-clearance passage traversal) that require physically feasible locomotion and posture adaptation. Task objectives were specified by scene geometry and goal conditions, such as reaching a target region, progressing along a prescribed path, and satisfying clearance constraints. Scene parameters were randomized across episodes (e.g., corridor width, turn angle, and target location) to assess robustness under layout variation. To induce MSD-relevant mechanical demand without object handling, an equivalent virtual payload was applied during rollouts (e.g., a constant downward load at the hand frame), producing trunk moments and L5/S1 loading during stooping and reaching.

Representative examples of whole-body postures relevant to these task settings are shown in Figure 2. The three poses are included to visually clarify the representative posture demands and whole-body configurations considered in this study.

Baselines and Metrics

The technique is compared against three representative baselines: (i) reference tracking only (RTO), (ii) PPO without risk shaping (SPO), and (iii) mean-risk-only PPO (tail term removed). To examine whether subject-specific inputs contribute to individualized optimization, we additionally include a descriptor-ablated variant of the proposed method, in which subject descriptors are removed while all other components and training settings are kept unchanged. This comparison is used as a targeted individualization check rather than as a separate baseline family. Reported metrics include distributional risk (mean risk and exceedance probability), safety/feasibility (success rate), task fidelity (goal-reaching error and trajectory deviation), motion quality (smoothness and correction magnitude), and effort (mean squared torque, $\|\tau\|_2^2$). Results are reported for (1) overall effectiveness under physics-based execution (risk reduction with task fidelity and feasibility), (2) risk aversion by removing the tail term (change in high-risk exceedance), and (3) an individualization check based on the descriptor-ablated comparison.



Figure 2: Representative whole-body posture categories relevant to task-motion optimization.

Results and Discussions

As shown in Table 1, the proposed technique achieved the best overall trade-off under physics-based execution. It reduced mean risk from 10.88 (RTO) and 9.93 (SPO) to 6.15, and lowered tail probability from 0.25 to 0.14, while maintaining strong task completion (success 0.85) and the lowest task error (0.19). Corrections remained bounded (residual magnitude 0.19) and smoothness stayed comparable to the baselines (1.03 vs. 1.00–1.10), supporting local, physically plausible posture adaptation rather than rewriting the reference motion.

Table 1: Overall effectiveness

Metrics	RTO	SPO	PPO	Proposed
Mean risk	10.88 ± 0.27	9.93 ± 0.33	8.32 ± 0.26	**6.15 ± 0.23**
Tail prob.	0.25 ± 0.01	0.22 ± 0.02	0.20 ± 0.02	**0.14 ± 0.01**
Task error	0.38 ± 0.01	0.29 ± 0.01	0.23 ± 0.01	**0.19 ± 0.01**
Success	0.62 ± 0.02	0.82 ± 0.02	0.84 ± 0.02	**0.85 ± 0.01**
Res. mag.	0.00	0.15 ± 0.01	0.21 ± 0.02	**0.19 ± 0.01**
Smoothness	1.00 ± 0.03	1.07 ± 0.05	1.10 ± 0.04	**1.03 ± 0.03**

The reduction in risk (without compromising success rates) demonstrates that the hard safety constraints and physics-based simulation act as a critical filter, ensuring that all suggested posture corrections remain within the user's biomechanical limits. Unlike the SPO baseline, which fails to prioritize health, or the PPO baseline, which lacks robustness against extreme risk, the proposed method identifies a "safe manifold" for human motion.

Furthermore, the low residual magnitude (0.19) and high task fidelity (0.19 error) confirm that the technique achieves optimization by actively incorporating the user's habitual motion patterns. This suggests that the pipeline

does not impose a generic, "one-size-fits-all" posture; instead, it provides a personalized refinement that preserves the worker's unique motor profile while systematically hedging against both average and tail-end MSD risks. Consequently, the framework proves to be a viable bridge between complex biomechanical theory and practical, actionable interventions in labor-intensive environments.

Beyond improving average ergonomic scores, the necessity of this distribution-aware approach is most clearly evidenced by its performance in extreme scenarios. As shown in Table 2, the proposed technique significantly reduced rare-but-severe outcomes that standard RL methods often overlook. While the SPO and PPO baselines exhibited high-risk episode frequencies of 0.34 and 0.29, respectively, the proposed framework lowered this incidence to 0.12, flattening the tail of the risk distribution.

Table 2: Tail-sensitive outcomes for risk-aversion ablation

Method	P95 risk	High-risk episodes
SPO	12.18 ± 0.42	0.34 ± 0.05
PPO	11.64 ± 0.47	0.29 ± 0.04
Proposed	**9.63 ± 0.27**	**0.12 ± 0.02**

The reduction of P95 risk to 9.63 (compared to 12.18 for SPO and 11.64 for PPO) further confirms that the pipeline does not just lower the mean burden but proactively avoids the most hazardous postural configurations. This demonstrates that by penalizing tail-event probabilities through the distilled risk student, the system provides a level of safety reliability that is essential for real-world deployment, where injury prevention must be absolute. These findings ultimately validate that the proposed RL-based pipeline effectively hedges against the technical barriers of risk-reward instability and safety enforcement in complex task environments.

Table 3: Subject-conditioning ablation

Method	Mean risk	Tail prob.	Success
W/o Subject	9.92 ± 0.41	0.23 ± 0.02	0.81 ± 0.04
Proposed	**6.15 ± 0.23**	**0.14 ± 0.01**	**0.85 ± 0.01**

As shown in Table 3, the descriptor-ablated variant produced a substantially higher mean risk (9.92 ± 0.41) than the full subject-conditioned policy (6.15 ± 0.23), while task success remained broadly comparable (0.81 ± 0.04 vs. 0.85 ± 0.01). The relatively small change in success is likely because the evaluated task settings are not highly failure-prone, so the main effect of subject conditioning is expressed more clearly through risk reduction than through task completion rate. This result suggests that subject-specific descriptors contribute materially to the optimization outcome rather than acting as inactive auxiliary inputs. The pronounced increase in risk after removing individual information indicates that the learned correction strategy is not purely generic but is meaningfully shaped by subject-conditioned features. These findings provide additional evidence for the individualized character of the proposed optimization framework.

Conclusions

An RL-based pipeline is presented for individualized posture correction under physics-based execution and hard safety constraints. Rather than synthesizing motion from scratch, the policy outputs bounded joint-space residuals that produce local, actionable corrections while preserving task intent. A distilled risk student provides mean risk and high-risk exceedance probability for distribution-aware reward shaping, enabling explicit suppression of tail events beyond mean-only optimization.

The tests demonstrate that this design effectively reduces MSD risk while maintaining high task success and physical feasibility under the given task settings. The added subject-conditioning ablation further shows that removing individual descriptors substantially increases mean risk without yielding a meaningful improvement in task success, indicating that subject-specific inputs contribute materially to the learned correction strategy and supporting the individualized nature of the proposed framework. Ablations further indicate that hard safety enforcement and tail-risk shaping are key to stabilizing long-horizon rollouts under biomechanical constraints.

These findings indicate that integrating temporal context with distribution-aware reward shaping allows RL to generate corrections that are both ergonomically superior and physically executable. Ultimately, this research provides a scalable pathway for transitioning from generic ergonomic guidelines to personalized, biomechanically coherent digital interventions that can reduce the long-term burden of MSDs in labor-intensive industries. Future work will extend the framework to more contact-rich activities and integrate higher-fidelity musculoskeletal models to better support real-world ergonomic interpretation.

Acknowledgments

This work was supported by the New Frontiers in Research Fund (NFRF), grant number NFRFE-2023-00174. The authors wish to thank their industry partners for their help with data collection, as well as the anonymous subjects who participated in the data collection.

References

Belbeck, A., Cudlip, A. C., and Dickerson, C. R. (2014). Assessing the interplay between the shoulders and low back during manual patient handling techniques in a nursing setting. International Journal of Occupational Safety and Ergonomics, 20(1):127–137.

Chen, Y.-L. and Zhang, L.-P. (2025). Postural variability in sitting: Comparing comfortable, habitual, and correct strategies across chairs. Applied Sciences, 15(13):7239.

Ciccarelli, M., Germani, M., Marinelli, F., Papetti, A., and Pizzuti, A. (2026). An ergonomic zone polyhedral representation-based mathematical program to prevent work-related musculoskeletal risks. Expert Systems with Applications, 302:130579.

Dafarra, S., Romualdi, G., and Pucci, D. (2022). Dynamic complementarity conditions and whole-body trajectory optimization for humanoid robot locomotion. IEEE Transactions on Robotics, 38(6):3414–3433.

González-Alonso, J., Martín-Tapia, P., González-Ortega, D., Antón-Rodríguez, M., Díaz-Pernas, F. J., and Martínez-Zarzuela, M. (2025). Me-ward: A multimodal ergonomic analysis tool for musculoskeletal risk assessment from inertial and video data in working places. Expert Systems with Applications, 278:127212.

Hignett, S. and McAtamney, L. (2000). Rapid entire body assessment (reba). Applied ergonomics, 31(2):201–205.

Hulshof, C. T., Pega, F., Neupane, S., Colosio, C., Daams, J. G., Kc, P., Kuijer, P. P., Mandic-Rajcevic, S., Masci, F., van der Molen, H. F., et al. (2021). The effect of occupational exposure to ergonomic risk factors on osteoarthritis of hip or knee and selected other musculoskeletal diseases: A systematic review and meta-analysis from the who/ilo joint estimates of the work-related burden of disease and injury. Environment International, 150:106349.

Ji, X., Littman, A., Hettiarachchige, R. O., and Piovesan, D. (2023). The effect of key anthropometric and biomechanics variables affecting the lower back forces of healthcare workers. Sensors, 23(2):658.

Kee, D. (2022). Systematic comparison of owas, rula, and reba based on a literature review. International journal of environmental research and public health, 19(1):595.

Mahmood, N., Ghorbani, N., Troje, N. F., Pons-Moll, G., and Black, M. J. (2019). Amass: Archive of motion capture as surface shapes. In Proceedings of the IEEE/CVF international conference on computer vision, pages 5442–5451.

McAtamney, L. and Corlett, E. N. (1993). Rula: a survey method for the investigation of work-related upper limb disorders. Applied ergonomics, 24(2):91–99.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. nature, 518(7540):529–533.

Nadon, A. L., Cudlip, A. C., and Dickerson, C. R. (2018). Joint moment loading interplay between the shoulders and the low back during patient handling in nurses. Occupational Ergonomics, 13(1_suppl):81–90.

Peng, X. B., Abbeel, P., Levine, S., and Van de Panne, M. (2018). Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. ACM Transactions On Graphics (TOG), 37(4):1–14.

Punnakkal, A. R., Chandrasekaran, A., Athanasiou, N., Quiros-Ramirez, A., and Black, M. J. (2021). Babel: Bodies, action and behavior with english labels. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 722–731.

Ren, K. (2025). Feasibility of an integrated multi-model approach for dynamic musculoskeletal disorder risk assessment.

Rören, A., Ogiez, R., Gajny, L., Blasco, A., Bouvier, F. M., Feydy, A., Rannou, F., Lefèvre-Colau, M.-M., and Roby-Brami, A. (2024). Arm elevation involves changes in the whole spine: an exploratory study using eos imaging. BMC Musculoskeletal Disorders, 25(1):993.

Santos, W., Lorente, A., Rojas, C., Isidoro, R., Dias, A., Mariscal, G., Zabady, A. H., and Lorente, R. (2025). A systematic review and meta-analysis on the prevalence and demographic risk factors of work-related musculoskeletal disorders in construction workers. Frontiers in Public Health, 13:1651921.

Schulman, J., Levine, S., Abbeel, P., Jordan, M., and Moritz, P. (2015). Trust region policy optimization. In International conference on machine learning, pages 1889–1897. PMLR.

Westermann, K., Lin, J. F.-S., and Kulić, D. (2020). Inverse optimal control with time-varying objectives: application to human jumping movement analysis. Scientific reports, 10(1):11174.

Zheng, Y. and Yumane, K. (2015). Human motion tracking control with strict contact force constraints for floating-base humanoid robots. US Patent 9,120,227.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

Text-to-AR for Home Remodeling: An End-to-End Pipeline from Natural Language to AR-Compatible 3D Assets

Vamsi Sai Kalasapudi[1], Bharadwaj R. K. Mantha[2,3], Hassan Odeh[4] and Alekhya Seelam[1]
[1]College of Computing, Engineering and Construction, University of North Florida, Florida, USA
[2]Dept. of Civil and Environmental Engineering, University of Sharjah, Sharjah, UAE
[3]Dept. of Civil Engineering, Indian Institute of Technology Madras (IITM), Chennai, India
[4]Freelance AI Consultant, UAE

Abstract

Home remodeling often relies on sketches and informal discussions, causing miscommunication, rework, and added cost. Existing AR tools are typically proprietary and limited to predefined 3D asset libraries. This study develops and validates an executable end-to-end pipeline that converts free-form homeowner text into AR-compatible 3D remodeling visualizations. A sofa case study evaluates generated assets across image resolution, input type, and quality condition using SSIM and dimensional metrics. Results show visually coherent assets suitable for early-stage communication, while highlighting depth and scale limitations for future environment-aware AR remodeling tools.

Introduction

The home remodeling market has expanded significantly in the post-pandemic period, exceeding $880 billion globally in 2022 and projected to surpass $1.1 trillion by 2027, (ReportLinker, 2023). The United States alone accounted for $567 billion in 2022 (Joint Center for Housing Studies of Harvard University, 2023). This growth has accelerated digital and remote planning workflows, with increasing reliance on immersive tools such as VR and AR, now accessible via consumer devices (Brown, 2020). Augmented Reality (AR) overlays digital elements onto real-world environments, enabling users to visualize furnishings and design changes within their own spaces before implementation (Azuma, 1997; IKEA, 2017; Inc., 2018).

Consumer applications such as IKEA Place and Houzz's "View in My Room 3D" demonstrate the growing adoption of AR in residential design. These tools improve decision-making and purchasing likelihood, with Houzz reporting significantly higher engagement and conversion rates among AR users (IKEA, 2017; Inc., 2018). However, these systems rely on proprietary, pre-modeled catalogs, limiting scalability, customization, and the ability to handle free-form user inputs (e.g., "two-seater U-shaped red sofa") (Boopathy et al., 2024).

AR is also being explored in construction workflows, such as the Augmented Carpentry project, which overlays fabrication guidance directly onto materials, demonstrating integration of digital design with physical execution (Coxworth, 2025). This AR-assisted production method adeptly merges human skill with computational precision, actualizing digital designs on the workpiece. These advancements indicate a wider trend of employing AR to seamlessly incorporate design information into actual construction processes, extending beyond mere screen visualization. Despite these advances, remodeling remains communication-intensive, particularly for homeowners with limited technical expertise. Reliance on informal sketches and verbal descriptions often leads to misinterpretation, delays, and increased costs (for Housing Studies of Harvard University, 2014). With only basic visualization support, homeowners often with little to no construction experience, frequently rely on hand-drawn sketches from imagination and informal discussions at home to explain their ideas. These ad hoc methods can make it difficult to convince the remodeler of the intended design, increasing the likelihood of miscommunication, design inaccuracies, and inefficient implementations. Many existing tools assume user familiarity with design principles or rely on fixed 3D asset libraries, limiting creativity and personalization. Even AI-based systems typically provide predefined options rather than enabling translation of nuanced, real-world inspirations into customized designs. Furthermore, end-to-end pipelines converting free-form input into AR visualizations remain largely unavailable (Höllein et al., 2023).

To address these gaps, this study presents a first-step toward automating interior remodeling visualization by converting free-form user descriptions into AR-compatible 3D models. The goal is to reduce communication barriers between non-expert homeowners and professional stakeholders including contractors, sub contractors, suppliers, interior designers, and architects. The proposed system enables users to describe desired designs in natural language and generates corresponding AR visualizations that are projected onto the real world environments. It integrates natural language processing (NLP), text-to-image generation, multi-view synthesis, 3D reconstruction, and AR deployment into a unified pipeline, generating content dynamically rather than relying on pre-built assets. The aims of this research are to (1) streamline and accelerate early-stage ideation by producing realistic previews from textual or image descriptions, (2) reduce the cost and specialized expertise typically required to create custom 3D remodeling visualizations, and (3) demonstrate the feasibility of end-to-end automation from text or image input to AR output.

Literature Review

Advances in generative AI, computer vision, and AR have expanded automated content creation in interior design. This review synthesizes progress in text-to-image, multi-view synthesis, and text-to-3D reconstruction, highlighting limitations in achieving end-to-end AR workflows for homeowner use. Diffusion-based models such as Stable Diffusion (Rombach et al., 2022) and Imagen (Saharia et al., 2022) generate high-quality images from text, with domain adaptations enabling interior design visualization (Chen et al., 2023). However, outputs remain 2D and unsuitable for AR interaction.

Multi-view synthesis methods such as SceneScape (Fridman et al., 2023) and FlexGen (Zhang and Lee, 2023b) improve geometric consistency across viewpoints. However, they remain computationally intensive and prone to inconsistencies, with limited support for iterative refinement. 3D reconstruction methods such as NeRF (Mildenhall et al., 2020b), Text2Room (Höllein et al., 2023), DreamFusion (Poole et al., 2022), and Magic3D (Lin et al., 2023) enable text-driven 3D generation. However, outputs typically require extensive post-processing (e.g., mesh cleanup, texturing), limiting usability for non-experts. A key limitation is AR integration, as generated 3D assets often require manual conversion for real-time deployment. Existing research pipelines rarely support direct AR-ready outputs, limiting scalability (Poole et al., 2022; Lin et al., 2023).

The proposed pipeline addresses this gap by integrating text input, multi-view synthesis, 3D reconstruction, and direct export to AR-compatible formats (e.g., USDZ), enabling iterative and scalable remodeling visualization. Table 1 summarizes how representative recent studies cover the major components required for a text-to-AR workflow. Prior work shows strong progress in individual stages such as text-to-image generation, multi-view synthesis, and text-to-3D reconstruction. However, most approaches do not provide a complete path to AR-ready deployment and often offer limited support for iterative refinement that matches real remodeling conversations.

Table 1: Comparative analysis of recent text-to-AR methodologies

Study	Text-to-Image	Multi-view Synthesis	3D Reconstruction	AR Integration	Iterative Refinement
(Chen et al., 2023)	Yes	No	No	No	Yes
(Fridman et al., 2023)	Yes	Yes	Yes	No	Limited
(Höllein et al., 2023)	Yes	Yes	Yes	No	No
DreamFusion (Poole et al., 2022)	No	No	Yes	No	No
Magic3D (Lin et al., 2023)	No	No	Yes	No	No
Our Proposed Pipeline	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

Reliable 3D Reconstruction as a Practical Requirement

Reliable 3D reconstruction is essential in remodeling, as decisions depend on spatial accuracy and scale rather than visual plausibility alone. Single-image reconstruction methods suffer from missing geometry due to occlusions (Mildenhall et al., 2020a; Gkioxari and Malik, 2019), while traditional multi-view approaches require extensive image capture (Seitz and Curless, 2021; Schönberger and Frahm, 2016). Generative multi-view synthesis offers a promising alternative by producing consistent viewpoints from text, reducing data requirements while improving reconstruction fidelity (Ho and Salimans, 2023; Zhang and Lee, 2023a). These observations motivate the need for workflows that couple multi-view consistency with efficient reconstruction and deployment constraints, aligning with the introduction's goal of advancing toward automated transformation from free-form user intent into AR-compatible 3D assets for remodeling visualization.

Methodology

This study presents a first-step exploration toward an end-to-end text-to-AR workflow for home remodeling visualization. The goal is not to introduce new generative or reconstruction models, but to establish an executable baseline and clarify what is practically required to translate a homeowner's free-form natural language intent into an AR-ready 3D asset that can be viewed within the real home environment. The contribution of this work is therefore the integration of established components into a deployable workflow with inspectable intermediate outputs, which supports intent alignment and helps identify where failures and quality limitations arise in practice. In this context, the pipeline is designed to reduce the homeowner–professional communication gap by turning ambiguous verbal descriptions into shared, inspectable visual artifacts that can be iteratively refined during early-stage remodeling discussions. This methodology aims to establish an end-to-end workflow that transforms free-form natural language descriptions into AR-ready 3D assets for early-stage home remodeling visualization. It further seeks to characterize how practical input conditions and key pipeline settings influence output quality, particularly reconstruction robustness and dimensional plausibility. Finally, the workflow introduces intermediate quality checks to limit error propagation and documents where human refinement remains necessary to resolve intent ambiguity and support clearer homeowner–professional communication.

Pipeline Overview

The workflow consists of four interrelated phases that progressively convert language into deployable AR content: (1) Text-to-2D Synthesis, (2) Image Enhancement and Multiview Generation, (3) 2D-to-3D Reconstruction, and (4) 3D-to-AR Integration. Each phase produces a tangible intermediate output (candidate images, processed multiview sets, reconstructed meshes, AR-compatible assets) that can be inspected and improved before moving forward, which supports faster alignment between user intent and the generated visualization. This staged design is used as a practical first step because it produces inspectable intermediate outputs and an explicit mesh for AR deployment, which is important for maintaining geometric consistency and dimensionally plausible visualization in remodeling contexts.



Figure 1: Overview of the proposed Text-to-AR pipeline. The shaded components indicate the phases evaluated in the case study.

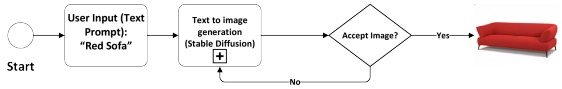


Figure 2: Phase 1: Text-to-2D synthesis workflow.

Phase 1: Text-to-2D Synthesis

Phase 1 translates a user's natural language description (e.g., "modern red two-seater sofa") into an initial set of candidate images that serve as the visual starting point for later multiview generation and reconstruction. Stable Diffusion is used as the text-to-image engine, and multiple candidates are produced per prompt by varying seed values and guidance settings to capture reasonable visual interpretations of the same intent. In this study, a Stable Diffusion model (v1.5) is used to generate candidate images, with multiple outputs produced by varying random seed values to support iterative selection. To directly support remodeling conversations, this phase is treated as an iterative intent-clarification step. Rather than assuming the first output is correct, users can select the closest candidate and refine the description to better match their target appearance. This human-in-the-loop selection is intentionally positioned as a lightweight communication mechanism: it reduces ambiguity early, before downstream 3D reconstruction amplifies errors or inconsistencies.

The current implementation does not explicitly encode metric dimensions in text prompts, as diffusion generation operates in a scale agnostic space. However, preliminary tests with dimension aware prompts (e.g., approximate size or capacity) showed improved proportional consistency.

Phase 2: Image Enhancement and Multiview Generation

After Phase 1, the selected image is conditioned to better support downstream 3D reconstruction. This step focuses on improving visual clarity and isolating the object so that later multiview synthesis and reconstruction are less sensitive to background clutter and image artifacts. Multiview synthesis is implemented using Zero123++ (sudo-ai/zero123plus-v1.2) with an Euler Ancestral scheduler and 75 inference steps, generating six consistent viewpoints per input. First, denoising and super-resolution are applied to improve sharpness and reduce noise, using ESRGAN to enhance spatial detail when needed. Next, background removal is performed using the Rembg toolbox to foreground the target object and produce cleaner object boundaries for subsequent processing. Background removal is performed using Rembg, and super resolution is applied using Real-ESRGAN (RealESRGAN_x4plus, 4× upscaling) to improve image quality prior to multiview

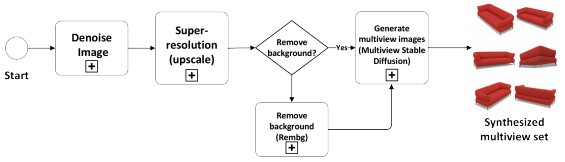


Figure 3: Phase 2: Image enhancement and multiview generation workflow.

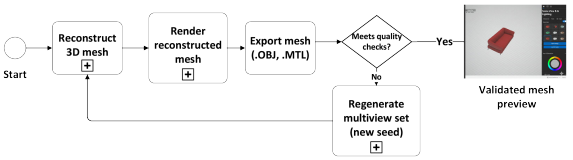


Figure 4: Phase 3: 2D-to-3D reconstruction workflow.

generation.

Following enhancement and segmentation, multiview synthesis is performed using Multiview Stable Diffusion to generate additional viewpoints of the same object. These synthetic views approximate photographs captured from different camera angles and provide the multi-perspective observations required for sparse-view 3D reconstruction. The overall sequence of enhancement, isolation, and multiview generation is summarized in Figure 3.

Phase 3: 2D-to-3D Reconstruction

Phase 3 converts the enhanced multiview image set into a 3D mesh representation. The processed views from Phase 2 are provided as input to a sparse-view reconstruction workflow (Instant Mesh), which generates a textured mesh by estimating object geometry from limited viewpoints. Reconstruction is performed using InstantMesh (instant_mesh_large.ckpt) with default configuration settings, operating on six synthesized views arranged in a 3×2 grid. Because reconstruction quality is sensitive to view consistency and segmentation quality, the resulting mesh is inspected using basic quality checks focused on geometric completeness and surface stability. When reconstruction artifacts are observed, the pipeline returns to Phase 2 to regenerate the multiview set with adjusted seeds, producing a new view collection for reconstruction. The reconstruction loop and decision point are illustrated in Figure 4. Final validated meshes are exported as `.OBJ` and `.MTL` files, preserving both geometry and associated material information for AR deployment.

The reconstruction process operates in a normalized coordinate space, and the metric scale is not inherently preserved. A configurable scale parameter is applied post reconstruction, allowing the generated mesh to be resized relative to the target scene.

Phase 4: 3D-to-AR Integration

Phase 4 prepares the reconstructed mesh for deployment in an AR environment. During AR integration, reconstructed meshes can be interactively rescaled to align with user defined spatial constraints and real-world placement requirements. The exported `.OBJ` and `.MTL` assets are con-

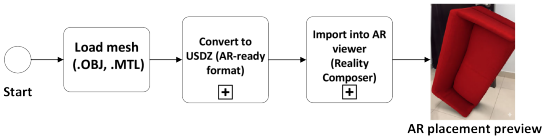


Figure 5: Phase 4: 3D-to-AR integration workflow.

verted into the USDZ format using Aspose 3D, enabling compatibility with common mobile AR workflows. After conversion, the asset is calibrated to ensure correct scale and orientation for placement in real indoor spaces. The calibrated USDZ model is then imported into Apple Reality Composer to support spatial anchoring and interactive visualization, including consistent placement within the physical environment and realistic rendering behavior. This integration sequence is summarized in Figure 5, and it produces an AR-ready asset that can be deployed on consumer-grade mobile devices for remodeling visualization.

This methodology establishes a practical baseline workflow for translating free-form user intent into AR-ready remodeling visualizations. Rather than proposing new foundation models, the contribution of this section is the structured integration of existing components into a reproducible sequence with clear intermediate outputs, enabling inspection and iteration before AR deployment. Evaluation procedures, parameter sensitivity observations, and quality outcomes are reported in the subsequent case study to assess visual plausibility, reconstruction robustness, and AR deployability in consumer-grade environments.

Unless otherwise specified, all models are used with default pretrained weights and configurations, with stochastic variation controlled through seed initialization.

Case Study: Results and Discussion

To evaluate the practical effectiveness of the proposed text-to-AR workflow, this study conducts a case study that examines how reliably visually coherent and dimensionally plausible 3D assets can be produced from limited two-dimensional inputs. The proposed methodology is fully executable, and the complete end-to-end pipeline was implemented and run as presented, from text prompting through AR-ready deployment. This evaluation is essential for home remodeling use cases where the objective is not fabrication-level precision, but the ability to translate homeowner intent into a shared visual artifact that supports clearer communication during early-stage design discussions. Accordingly, the case study assesses the pipeline's performance under varied input conditions, identifies key bottlenecks, and establishes baseline benchmarks that inform future improvements.

Experiment Overview

The primary objective of this experiment is to evaluate the robustness of the pipeline's 2D-to-3D reconstruction stage under diverse input conditions relevant to remodeling scenarios. Within the scope of this paper, the case study focuses on the most failure-prone and decision-critical portion of the workflow, Phases 2 and 3 (image preprocessing and multiview generation, and 2D-to-3D reconstruction), since these steps largely determine whether the final AR asset is usable for remodeling visualization. Sofa objects were selected because they are common in interior design tasks and present geometric challenges (curvature, occlusions, and complex silhouettes) that stress reconstruction quality in ways that are meaningful for AR-based space planning.

The dataset comprises sofa images organized into three categories namely a) Front-view images: frontal perspectives that preserve symmetry and provide clear visibility of primary features. b) Side-view images: lateral perspec-



Figure 6: Complete workflow visualization for four representative sofas. Each row shows: (a) original input image, (b) one representative view from the generated multiview set, and (c) final AR placement in a real room environment.

tives that introduce occlusion and asymmetry challenges; c) Synthetic images: controlled images collected for systematic evaluation under consistent conditions.

Real images were captured in front and side views to approximate practical acquisition conditions. In total, 12 distinct sofa models were evaluated to ensure variation in shape, style, and visual complexity. Synthetic sofa images were additionally collected from an online store to support controlled comparisons across consistent viewpoints and visual conditions.

To evaluate robustness under practical acquisition conditions, we applied controlled augmentations that vary image scale and quality. Each sofa image was resized to three resolutions (256, 512, and 1024 pixels) to assess sensitivity to input scale. For each resolution, two quality conditions were created. High-quality samples preserved sharpness and minimal noise to approximate favorable capture conditions. Low-quality samples were intentionally degraded using Gaussian blur (standard deviation = 2.0) with additional controlled noise to simulate common imperfections such as slight defocus and sensor noise, consistent with prior imaging literature (Roboflow, 2023; Gonzalez and Woods, 2023; Shorten and Khoshgoftaar, 2019). Controlled lighting variations were also introduced to reflect typical indoor capture variability. Figure 6 illustrates a representative sample processed through the complete workflow, showing the input image, preprocessing and background removal, multiview synthesis output, and the resulting reconstructed 3D mesh prepared for AR visualization.

All augmented inputs were processed through the full reconstruction workflow, including multiview synthesis, mesh reconstruction, and final rendering. This design enabled systematic comparison of pipeline performance across viewpoint type, resolution, and quality condition, providing an interpretable baseline for assessing reconstruction stability and identifying failure modes relevant to AR remodeling visualization.

Evaluation Metrics

To assess the proposed 2D-to-3D reconstruction workflow in a manner aligned with home remodeling visualization, we evaluate both (1) perceptual visual fidelity, which supports visually plausible AR previews for homeowner–professional alignment, and (2) geometric dimensional accuracy, which supports spatial plausibility during placement and scale interpretation in AR. Accordingly, we report one perceptual metric (SSIM) and two complementary dimension-based error metrics (NAE and NRMSE).

Perceptual visual fidelity (SSIM): The Structural Similarity Index Measure (SSIM), introduced by Wang et al. (Wang et al., 2004), evaluates image similarity using luminance, contrast, and structural components and is widely used in image synthesis and super-resolution research due to its correlation with human perception (Ledig et al., 2017; Zhang et al., 2018). In this study, SSIM is computed by comparing rendered projections of the reconstructed 3D mesh against the original input image and against its background-removed counterpart. The latter comparison reduces background influence and better isolates the reconstruction quality of the foreground object. SSIM values range from 0 to 1, where higher values indicate greater perceptual similarity.

Geometric dimensional accuracy (NAE and NRMSE): To evaluate dimensional plausibility, we adopt two metrics commonly used in object reconstruction and modeling studies (Tatarchenko et al., 2017; Wu et al., 2016). First, Normalized Absolute Error (NAE) measures the relative deviation between predicted and ground-truth dimensions (width, depth, and height):

$$E_{dim} = \left| \frac{D_{pred} - D_{true}}{D_{true}} \right| \tag{1}$$

where D_{pred} is the reconstructed dimension and D_{true} is the ground-truth measurement. This normalized form enables consistent comparison across objects of different sizes. Second, we use Normalized Root Mean Squared Error (NRMSE) as a scale-independent aggregate measure of dimensional accuracy:

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(D_{pred}^{(i)} - D_{true}^{(i)}\right)^2}}{\bar{D_{true}}} \tag{2}$$

where N denotes the number of evaluated samples and $\bar{D_{true}}$ denotes the mean of the ground-truth dimensions across the dataset. Lower NAE and NRMSE values indicate improved dimensional agreement.

Together, SSIM, NAE, and NRMSE provide a concise and complementary evaluation of visual plausibility and dimensional deviations. Results are reported across viewpoint type, resolution, and quality condition to characterize reconstruction robustness under varied input scenarios.

Results

Following data acquisition and augmentation, all samples were processed through the complete evaluation workflow. SSIM was computed to quantify perceptual fidelity by comparing rendered projections of the reconstructed meshes against the original input images and their background-removed counterparts. Geometric accuracy was assessed by comparing reconstructed object dimensions (width, depth, and height) against ground truth measurements across the 12 sofa models. Results are reported across image resolution (256, 512, and 1024 pixels), input type (front-view, side-view, and synthetic), and quality condition (high vs. low) to capture performance variability under realistic acquisition conditions. These results provide a baseline characterization of the tradeoff between perceptual plausibility and dimensional plausibility for AR-ready assets generated from limited inputs, which is critical for determining when such outputs are usable for early-stage remodeling communication.

Across all samples, the average SSIM between rendered meshes and original input images was 0.556, indicating

that the reconstructed outputs preserved key structural characteristics despite texture and mesh granularity limitations. When SSIM was computed against background-removed inputs, the average increased slightly to 0.564, suggesting that background complexity can depress perceptual similarity scores and that foreground isolation provides a clearer assessment of reconstruction quality. These SSIM values are consistent with ranges reported in prior single-input-view 3D reconstruction studies (Richardson et al., 2016; Wu et al., 2016). Dimensional accuracy was evaluated using NAE and NRMSE. Across the dataset, NAE values for width, depth, and height typically ranged from 20% to 60%, with width generally exhibiting greater stability. Table 2 reports the consolidated, scale-independent NRMSE values for each dimension. The results indicate that proportional structure is often preserved, while depth remains the most challenging dimension to recover reliably due to occlusion and limited visual cues.

Table 2: NRMSE by dimension across all evaluated sofa models.

Dimension	NRMSE
Width	30.98%
Depth	39.40%
Height	19.16%

Table 3: Per-sofa reconstruction quality across input conditions. SSIM is reported against original inputs. NAE is reported per dimension (W = width, D = depth, H = height).

Sofa	Input Type	Resolution (px)	SSIM (Orig.)	NAE_W (%)	NAE_D (%)	NAE_H (%)
S1	Front	512	0.325	30.0	70.0	2.6
S2	Side (Synthetic)	512	0.652	6.7	17.7	5.3
S3	Front	512	0.443	37.5	15.6	7.0
S4	Side	512	0.493	5.6	1.0	3.8
Average			**0.478**	**20.0**	**26.1**	**4.7**

To illustrate variability across representative cases and input conditions, Table 3 reports SSIM and dimensional errors for four sofa models at 512×512 resolution. The results show that height reconstruction is highly accurate (4.7% average error), while depth remains the most challenging dimension (26.1% average error), particularly for front-view inputs where occluded rear geometry must be inferred.

The results highlight the relationship between perceptual fidelity, geometric accuracy, and practical usability for AR remodeling visualization. Key observations are summarized below.

Visual plausibility is consistently maintained: The average SSIM values of 0.556 (against original inputs) and 0.564 (against background-removed inputs) indicate that reconstructed meshes preserve the primary structural features of the sofas. The slight improvement after background removal suggests that background complexity can suppress perceptual similarity scores and that foreground isolation provides a clearer assessment of reconstruction quality.

Depth remains the dominant source of dimensional error: Dimension-based evaluation shows larger discrepancies in depth compared to width and height. This outcome is expected when reconstruction begins from a single input view and relies on synthesized viewpoints, as occluded regions provide limited cues for recovering true depth.

Input view and resolution influence reconstruction stability: Reconstructions from front-view inputs are generally more stable than those from side views, where asymmetry and occlusion introduce additional ambiguity. Higher resolutions tend to improve SSIM and modestly reduce dimensional errors, although gains diminish at higher resolutions.

Implications for AR remodeling use: The pipeline produces visually coherent assets that support early-stage ideation, AR-based exploration, and communication of intent. However, the observed dimensional inaccuracies indicate that the current workflow is not intended for high-precision tasks such as fabrication or construction detailing.

Overall trade-off: The results demonstrate a clear trade-off between visual plausibility and geometric accuracy. The pipeline is effective for generating AR-ready previews for communication, while future improvements should prioritize scale awareness, depth reliability, and constraint-based reconstruction.

Conclusions and Future Research Directions

This paper implements an executable first-step text-to-AR workflow aimed at reducing the homeowner–professional communication gap in early-stage home remodeling by translating free-form user intent into an AR-ready 3D visualization. The primary contribution is the practical integration of established components (text-to-image generation, image preprocessing and multiview synthesis, sparse-view reconstruction, and USDZ-based AR deployment) into a reproducible pipeline with inspectable intermediate outputs that support iterative refinement and reduce reliance on manual 3D modeling expertise. While the full workflow is implemented end to end, the quantitative evaluation is intentionally focused on the most failure-prone and decision-critical stages that govern reconstruction quality and dimensional plausibility.

A sofa case study was selected as a conservative test object due to curved geometry and frequent occlusions, which stress multiview consistency and depth recovery more than common rectilinear interior elements. The results show that the implemented workflow can produce visually coherent assets suitable for AR ideation and communication, with average SSIM values of 0.556 against original inputs and 0.564 against background-removed inputs. However, geometric accuracy remains limited, particularly for depth, indicating that the current workflow is best positioned as an intent-alignment and early visualization tool rather than a substitute for precision measurement or fabrication-ready modeling. Overall, these findings support the feasibility of producing AR-ready previews from limited inputs and clarify the primary technical limitations that must be addressed in future work to improve dimensional plausibility and environment-aware placement.

Future work will focus on four directions: (1) environment-aware capture and spatial reasoning (e.g.,

leveraging mobile camera/LiDAR) to support scale- and fit-aware placement and basic clearance checking; (2) room-scale, multi-object scene generation with stronger geometric constraints and priors (standard dimensions and optional user measurements) to reduce depth ambiguity; (3) user-centered AR evaluation to quantify usability, perceived realism, and the impact on homeowner–professional decision-making; and (4) benchmarking direct text-to-3D methods within the same AR deployment constraints to compare dimensional plausibility and post-processing requirements against the staged workflow used in this study. Future work will also expand evaluation beyond sofas by collecting and labeling diverse furniture and interior asset datasets across multiple viewpoints to more rigorously assess generalizability and robustness. In addition, we are exploring user-captured imagery as input by integrating grounded detection and segmentation methods (e.g., Grounding DINO and Grounded SAM) to isolate real objects and evaluate projection fidelity in AR under practical conditions. Future work will also focus on incorporating explicit scale control into the pipeline, including user-defined dimensional inputs to guide scaling, automated dimension estimation from images, and scale-aware normalization prior to evaluation. Providing users with structured prompt templates that include dimensional cues is also expected to improve the generation of dimensionally plausible AR assets. Finally, we are investigating more controllable and open model alternatives to enable stronger user control over key design attributes such as dimensions and color.

References

Azuma, R. T. (1997). A survey of augmented reality. Presence: Teleoperators and Virtual Environments, 6(4):355–385.

Boopathy, K., Kalasapudi, V., Mantha, B. R., and Zou, Y. (2024). A survey-based investigation to understand the influence of an augmented reality-based application on the home remodeling market. In Construction Research Congress 2024, pages 426–435.

Brown, E. Q. (2020). How COVID-19 has changed home design. `https://www.architecturaldigest.com/story/covid-19-home-design`. Accessed: 2025-03-13.

Chen, Y., Liu, H., Zhang, X., and Lin, D. (2023). Diffusion models for text-driven interior design. ACM Transactions on Graphics (TOG), 42(2):1–15.

Coxworth, B. (2025). Augmented carpentry system brings furniture design into the real world. `https://newatlas.com/vr/augmented-carpentry-system/`. Accessed: 2025-04-06.

for Housing Studies of Harvard University, J. C. (2014). Emerging trends in the remodeling market. Technical report, Harvard University.

Fridman, A., Li, Y., Wang, C., and Funkhouser, T. (2023). Scenescape: Multi-view diffusion models for generating consistent 3d scenes from text. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 45(6):7689–7701.

Gkioxari, G. and Malik, J. (2019). Mesh R-CNN. ICCV.

Gonzalez, R. C. and Woods, R. E. (2023). Digital Image Processing. Pearson, 4th edition.

Ho, J. and Salimans, T. (2023). Diffusion models for 3d generation. Advances in Neural Information Processing Systems.

Höllein, L., Yang, Y., and Kopf, J. (2023). Text2room: Room-scale 3d reconstruction from text descriptions. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 5673–5682.

Höllein, L., Yang, Y., and Kopf, J. (2023). Text2room: Room-scale 3d reconstruction from text descriptions. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 5673–5682.

IKEA (2017). Ikea place. `https://www.ikea.com/us/en/customer-service/mobile-apps/`. Accessed: 2025-04-14.

Inc., H. (2018). Houzz introduces view in my room 3d. `https://www.houzz.com/press/590/Houzz-Introduces-View-in-My-Room-3D`. Accessed: 2025-04-01.

Joint Center for Housing Studies of Harvard University (2023). Improving America's housing 2023. `https://www.jchs.harvard.edu/improving-americas-housing-2023`. Accessed: 2025-03-16.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., and Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 4681–4690.

Lin, C.-H., Gao, J., Tang, L., Lee, Y.-C., Xu, J., Lai, W.-C., and Anandkumar, A. (2023). Magic3d: High-resolution text-to-3d content creation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 15234–15244.

Mildenhall, B., Srinivasan, P. P., and Tancik, M. (2020a). NeRF: Representing scenes as neural radiance fields for view synthesis. ECCV.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. (2020b). Nerf: Representing scenes as neural radiance fields for view synthesis. Proceedings of the European Conference on Computer Vision (ECCV), pages 405–421.

Poole, B., Jain, A., Barron, J. T., and Mildenhall, B. (2022). Dreamfusion: Text-to-3d using 2d diffusion. Advances in Neural Information Processing Systems (NeurIPS).

ReportLinker (2023). Home improvement services market - global outlook and forecast 2023-2027. `https://www.reportlinker.com/p06417077/`. Accessed: 2025-04-16.

Richardson, E., Sela, M., Or-El, R., and Kimmel, R. (2016). Learning detailed face reconstruction from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5553–5562.

Roboflow (2023). Data augmentation techniques for computer vision. Accessed: 2025-06-12.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 10684–10695.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., Ghasemipour, K., Ayan, B. K., Mahdavi, S. S., Lopes, R. G., Salimans, T., and Ho, J. (2022). Photorealistic text-to-image diffusion models with deep language understanding. Advances in Neural Information Processing Systems (NeurIPS).

Schönberger, J. L. and Frahm, J.-M. (2016). COLMAP: A general-purpose structure-from-motion pipeline. 3DV.

Seitz, S. M. and Curless, B. (2021). Multi-view stereo: A tutorial. Foundations and Trends in Computer Graphics and Vision.

Shorten, C. and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1):60.

Tatarchenko, M., Dosovitskiy, A., and Brox, T. (2017). Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 2088–2096.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4):600–612.

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., and Xiao, J. (2016). Learning 3d shape representations from 2d images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1919–1928.

Zhang, K. and Lee, Y. (2023a). Flexgen: Controllable multi-view image synthesis. CVPR.

Zhang, K. and Lee, Y. (2023b). Flexgen: Controllable multi-view image synthesis using diffusion models. Proceedings of the IEEE/CVF CVPR, pages 1254–1263.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., and Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. CVPR, pages 586–595.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A ZERO-SHOT APPROACH TO DECOMPOSING GRAPHICAL RULES IN ACCESSIBILITY REGULATIONS

Yonghan Kim[1], Ghang Lee[1,2], and André Borrmann[2]
[1]Technical University of Munich, Munich, Germany
[2]Yonsei University, Seoul, South Korea

Abstract

Composite graphical rules hinder interpretation by multimodal large language models (MLLMs), and existing annotation-isolation methods require labeled training data. To address this, we propose a training-free approach using the Segment Anything Model 3 (SAM3) to automatically isolate annotations, remove cross-rule visual clutter, and reconstruct individual rules. Using manually decomposed rules as ground truth, we compared our method against prompt-based generative baselines. Results indicate that while baselines produce visually cleaner diagrams, MLLM-as-a-judge evaluations demonstrate our SAM3-based approach more reliably preserves the original regulatory semantics.

Introduction

Normative knowledge in regulatory documents appears in multiple forms, including natural language text, tables, illustrations, calculation methods, and commentaries (Amor and Dimyadi, 2021). Among these, illustrations play a critical role in conveying requirements that are difficult to express using text alone. In this work, we define illustrated provisions that explicitly encode declarative regulatory requirements as *graphical rules*.

However, graphical rules are often presented as composite diagrams that contain multiple regulatory requirements within a single illustration. Such composite representations can hinder rule interpretation, particularly when multimodal large language models (MLLMs) are used to analyze regulatory documents. Decomposing composite diagrams into atomic, rule-level units is therefore necessary for improving the interpretability and downstream usability of graphical regulatory knowledge.

Existing approaches, however, have important limitations. First, annotation-isolation methods typically rely on supervised learning with labeled datasets, which are costly and difficult to generalize to new regulatory domains (Zhang et al., 2023; Maupou et al., 2024). Second, prompt-based generative approaches can modify diagrams without a trained annotation-isolation model, relying instead on image regeneration, which may fail to preserve the original geometry and introduce semantic distortion (Wu et al., 2023).

To address these challenges, this paper proposes a zero-shot framework for decomposing composite graphical rules and pairing them with corresponding textual provisions. The proposed method leverages the zero-shot segmentation capability of SAM3 (Carion et al., 2025) to isolate annotation elements without labeled training data. These elements are grouped by spatial proximity, and a base image is reconstructed by removing annotations via inpainting. Each annotation group is then overlaid onto the base image to produce rule-level graphical units. Finally, the reconstructed graphical rules are paired with textual provisions using semantic similarity.

Related Works

LLM and MLLM for Rule Interpretation

Recent studies have explored the use of LLMs as reasoning engines for automated code compliance checking (ACC), enabling direct interpretation of regulatory text without explicit rule formalization (N. Chen et al., 2024; Ying and Sacks, 2024; Madireddy et al., 2025; Li, 2026; Iversen and Huang, 2026).

Extending this paradigm, recent work has investigated multimodal large language models (MLLMs) to incorporate graphical regulatory information alongside text (Kim et al., 2025). However, empirical findings show that MLLM performance degrades as visual complexity increases (Hou et al., 2025), particularly when multiple regulatory requirements are encoded within a single diagram.

This limitation is especially pronounced in tasks requiring precise spatial grounding. While MLLMs can often recognize numerical annotations, they frequently fail to associate them with the correct geometric entities or spatial context (Khan et al., 2025; Picard et al., 2025). These observations highlight the need for preprocessing strategies that reduce visual complexity and disentangle composite graphical rules prior to model interpretation.

Annotation isolation on a technical drawing

Annotation isolation aims to separate annotation elements (e.g., text, dimensions, symbols) from geometric content so that each component can be analyzed independently. Because regulatory semantics are primarily conveyed through annotations, this step is essential for interpreting graphical rules.

Early approaches relied on rule-based segmentation using drafting standards and geometric constraints (Dori and Velkovitch, 1998; Lin and Ting, 1997). While effective

under controlled conditions, these methods depend on strong assumptions about drawing regularity and are difficult to generalize across heterogeneous regulatory documents.

More recent approaches formulate annotation isolation as a supervised learning problem, using deep neural networks to detect and segment annotation elements (Zhang et al., 2023; Maupou et al., 2024). Although these methods improve robustness, they require large, labeled datasets and are therefore limited in scalability across domains.

Generative and Prompt-based Diagram Manipulation

Recent generative approaches enable selective modification of diagrams using natural language prompts (Jeong et al., 2025; Goswami et al., 2025; Yang et al., 2025). While these methods can simplify visual content, they rely on image regeneration rather than explicit decomposition. As a result, they may fail to preserve the original geometry and can introduce semantic distortions. This limitation is critical in regulatory contexts, where even minor visual changes may alter the intended meaning of a rule.

Positioning of This Work

In contrast to prior studies, this work focuses on the zero-shot decomposition of composite graphical rules in regulatory documents. Rather than introducing new algorithms, the contribution lies in framing graphical rule decomposition as a zero-shot problem, motivated by the high variability of regulatory drawings, which makes large-scale labeled training impractical.

Unlike annotation-isolation approaches that focus on segmenting individual elements, the proposed method reconstructs rule-level graphical units and pairs them with corresponding textual provisions. Furthermore, in contrast to generative diagram editing methods, it preserves the original diagram structure through decomposition rather than image synthesis.

By leveraging a general-purpose segmentation model without labeled training data, the method enables rule-level understanding while maintaining fidelity to the original illustration.

Proposed Method

The proposed method aims to decompose composite graphical rules into atomic rule-level graphical units and pair them with corresponding textual rules. To achieve this without labeled training data, the framework leverages the zero-shot segmentation capability of SAM3 (Carion et al., 2025), together with geometric and semantic post-processing steps.

As illustrated in Figure 2, the pipeline consists of four main stages. First, graphical and textual provisions are extracted from regulatory documents. Second, composite graphical rules are decomposed through zero-shot annotation segmentation. Third, textual provisions are decomposed into atomic rules. Finally, graphical and textual rules are paired based on semantic similarity.

Extraction of graphical rules and textual provisions

The input to the pipeline is a regulatory document in PDF format. Graphical rules and textual provisions are extracted using PyMuPDF (McKie, 2024), which provides structured access to page-level elements.

Graphical rules are identified by detecting raster images and vector graphics within PDF pages. Each detected graphical region is treated as a candidate composite graphical rule, and its bounding box and page index are preserved to maintain contextual consistency.

Textual provisions are extracted from the same pages and segmented at the paragraph level. To associate text with graphical rules, text blocks are filtered based on layout proximity, without relying on predefined templates or rule identifiers.

Graphical Rule Decomposition

Each composite graphical rule is processed independently to isolate annotation elements. Zero-shot annotation segmentation is performed using SAM3, guided by a fixed set of annotation-related prompts (e.g., arrows, dimension lines, and text labels).

To improve detection of small elements such as arrowheads and numeric annotations, Slicing Aided Hyper Inference (SAHI) (Akyon et al., 2022) is applied. SAM3 generates multiple candidate masks, which are filtered using geometric and visual heuristics (e.g., area, aspect ratio, and stroke density).

The extracted annotation elements are then grouped into rule-level units. Since individual regulatory requirements are typically represented by clusters of annotations, masks are grouped based on spatial proximity using distance-based clustering.A base image is constructed by removing annotation elements via image inpainting. A classical, non-learning-based inpainting method is used to preserve geometric

Textual rule decomposition

Textual provisions associated with a graphical rule often describe multiple regulatory requirements within a single paragraph. To align textual granularity with rule-level graphical units, each textual provision is decomposed into atomic textual rules using a two-stage, instruction-only prompt chaining strategy.

In the first stage, *requirement identification*, the language model is instructed to identify all independent regulatory requirements contained within the input provision. The system prompt explicitly guides the model to separate requirements based on distinct compliance conditions and to report the total number of identified requirements together with brief summaries. This stage focuses on structural decomposition, without modifying the original regulatory intent.

In the second stage, *rule normalization*, each identified requirement is reformulated into a standalone textual rule. The system prompt enforces normalization constraints, requiring that (1) each rule is understandable in isolation, (2) regulatory language using “shall” is consistently applied, (3) only one requirement is expressed per rule,

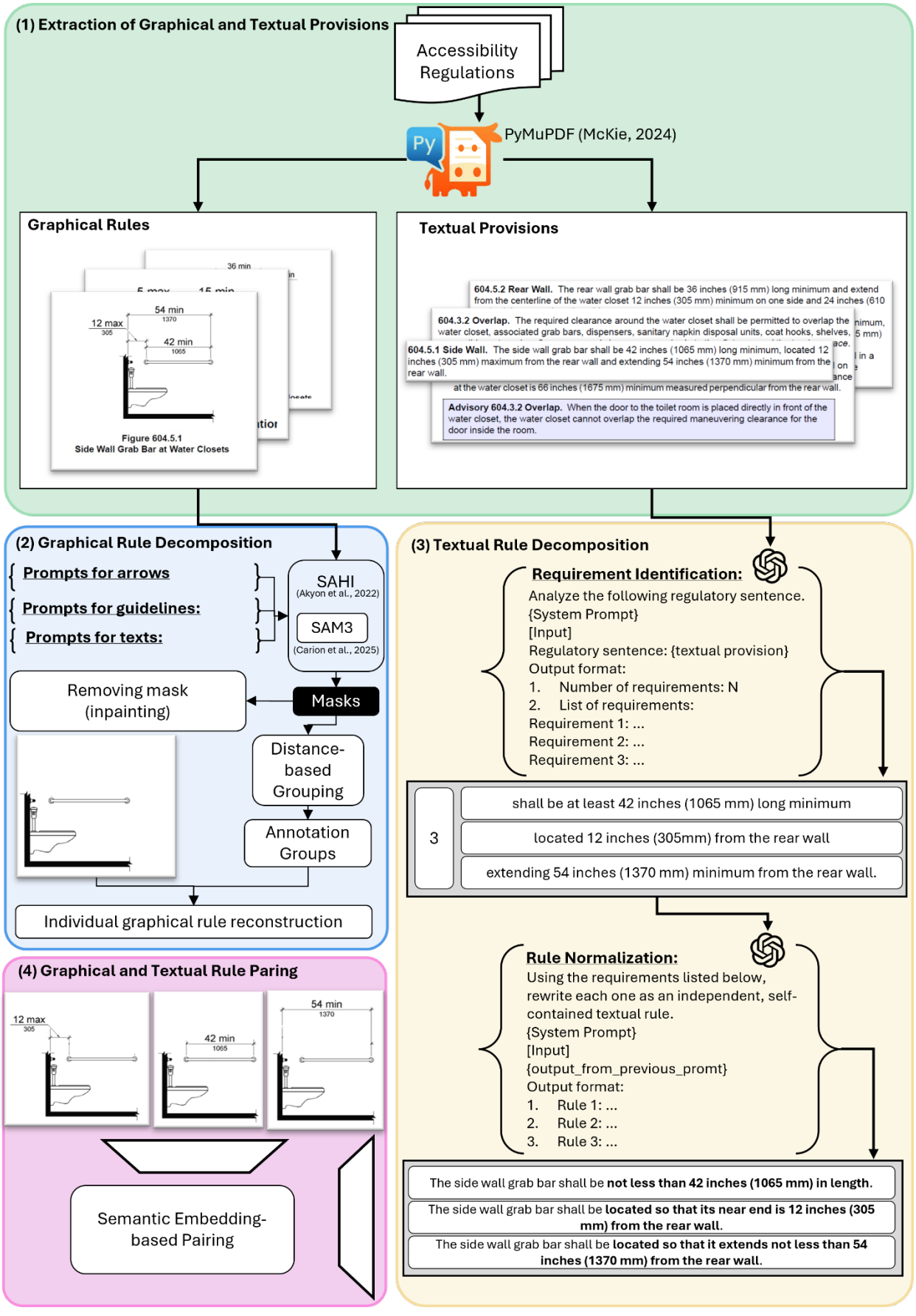


Figure 1: Overview of the proposed framework for graphical rule decomposition and pairing

and (4) all relevant measurement references and conditions are explicitly stated in a concise manner.

An instruction-only prompting strategy is adopted throughout both stages, without using any few-shot examples. The output of the requirement identification stage is provided as input to the rule normalization stage, enabling progressive refinement of the textual decomposition while preserving the original normative meaning.

Graphical and textual rule pairing

Finally, each reconstructed graphical rule is paired with its corresponding textual rule. To achieve this, semantic embeddings are computed using a pretrained CLIP model (ViT-B/32), which maps images and text into a shared embedding space.

Rule pairing is performed by maximizing cosine similarity between graphical and textual embeddings. The resulting graphical–textual rule pairs can be directly used for downstream interpretation by MLLMs.

Experimental design

The experimental evaluation focuses on assessing whether the proposed method correctly decomposes composite graphical rules into atomic, rule-level graphical rules that each correspond to a single regulatory requirement.

Based on the proposed method, this section describes the experimental design used to evaluate its effectiveness.

The evaluation is designed to assess two key aspects: (1) whether the decomposed diagrams represent a single regulatory requirement, and (2) whether they preserve the same regulatory meaning as the ground truth.

To this end, an MLLM-as-a-judge framework (D. Chen et al., 2024) is employed as the primary evaluation method. The framework is used to compare automatically decomposed graphical rules with manually decomposed ground-truth diagrams and to assess whether the predicted results preserve the intended regulatory meaning.

Ground truth generation

For each composite graphical rule, a set of ground-truth rule-level diagrams was manually generated using image editing software. Each ground-truth diagram isolates exactly one regulatory requirement by retaining only the annotation elements relevant to that requirement, together with the minimal graphical context necessary for interpretation.

In total, ground truth was constructed from 48 composite graphical rules, resulting in 134 individual graphical rules. The manual decomposition process was guided by the semantics of the corresponding textual provisions to ensure that each ground-truth diagram represents a single, independent regulatory rule without ambiguity or overlap.

These manually decomposed rule-level diagrams serve as the reference standard for evaluating the correctness of graphical rule decomposition.

Baseline definition

To contextualize the performance of the proposed method, a prompt-based annotation filtering baseline is employed. The baseline adopts a multi-step prompting strategy using a multimodal generative model to produce rule-specific diagrams by selectively suppressing irrelevant annotations in the original composite graphical rule.

First, the model is prompted to identify and enumerate the individual regulatory requirements encoded in the input composite diagram. These extracted requirements are organized into a structured list and serve as the basis for subsequent annotation filtering.

Next, for each identified regulatory requirement, the original composite diagram is provided again, and the model is instructed to remove all annotation elements that are unrelated to the target requirement, while preserving the original geometric drawing and the annotation elements corresponding to the selected rule. In this step, the model is explicitly guided to retain only the dimensional markers, symbols, and textual annotations necessary to convey the target regulatory requirement, and to suppress all others.

As a result, the baseline produces a set of rule-specific diagrams modified from the original image, each intended to express a single, self-contained regulatory requirement.

This baseline reflects recent prompt-driven approaches to diagram editing and controlled visual abstraction using generative multimodal models.

Evaluation metrics

The evaluation employs an MLLM-as-a-judge framework to assess rule-level semantic correctness.
For each decomposed graphical rule, the MLLM is first asked to determine whether the diagram represents exactly one regulatory requirement.
Next, the MLLM compares the automatically decomposed graphical rule with the manually decomposed ground-truth diagram and determines whether both convey the same regulatory requirement.

In addition to this semantic evaluation, we incorporate objective segmentation-level metrics to assess the quality of annotation isolation. Specifically, the mask-level intersection of union (IoU) is computed against manually annotated ground truth to quantitatively evaluate the accuracy of the extracted annotation elements.

Results and discussion

Experiment results

Table 1 summarizes the evaluation results obtained using the MLLM-as-a-judge framework.

Among the images generated by the proposed method, 84% were judged to contain exactly one regulatory requirement. In contrast, all images produced by the prompt-based baseline were classified as single-requirement diagrams. However, when comparing the regulatory semantics conveyed by each generated image with the manually decomposed ground truth, the proposed

Table 1: Overall experiment results from MLLM-as-a-Judge

MLLM-as-a-Judge	Proposed Method	Baseline
Has a single requirement	84 %	100 %
Same as ground truth	82 %	78 %

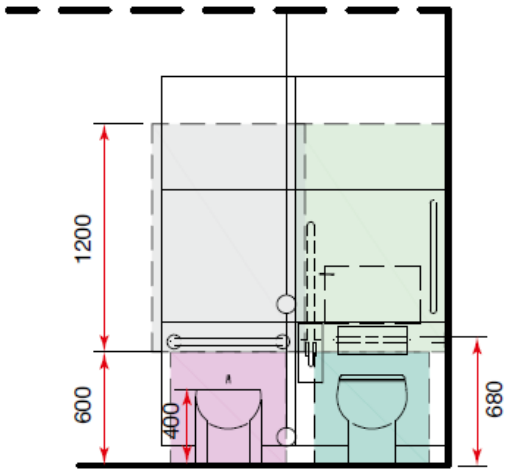
method achieved a higher agreement rate, showing a modest improvement (+4%) compared to the baseline.

In addition to semantic evaluation, we further assessed the quality of annotation isolation using segmentation-level metrics. The proposed method achieved a mask-level IoU of 88%, highlighting that segmentation quality remains a key bottleneck, and further improvements are required to achieve more reliable rule-level decomposition.

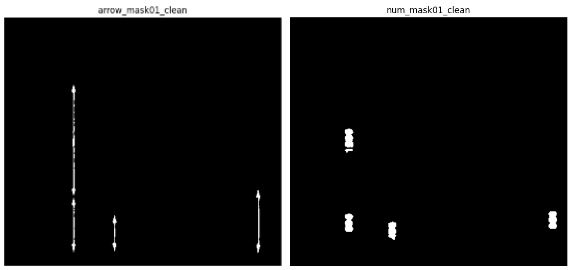
Discussions

To provide a clearer understanding of how the proposed method operates, we first present an intermediate example illustrating the decomposition process.

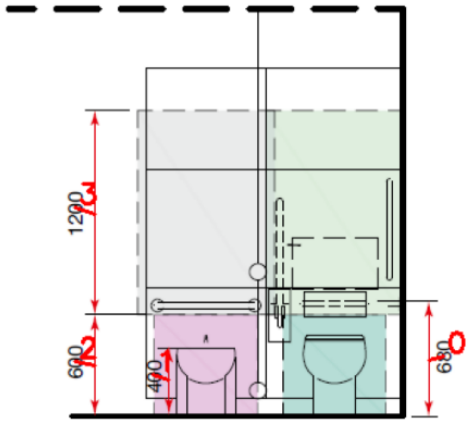
Figure 2 shows the intermediate results of the proposed pipeline, including annotation segmentation, clustering, and reconstruction.



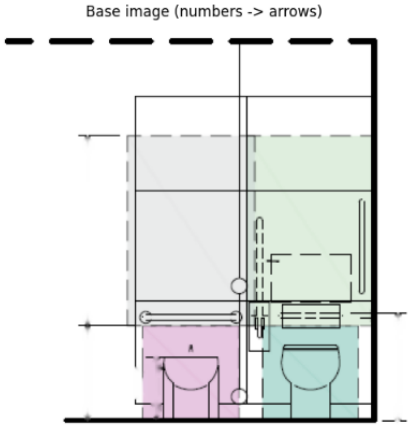
(a) Original composite graphical rule



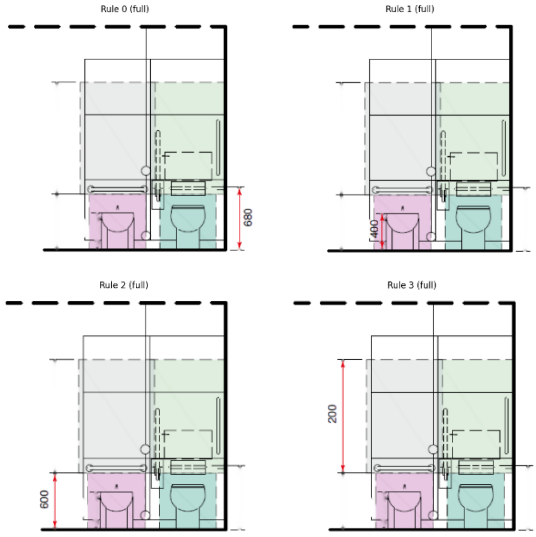
(b) Annotation segmentation results



(c) Grouping



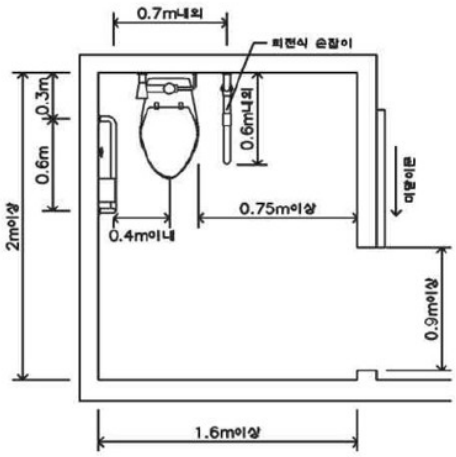
(d) Base image



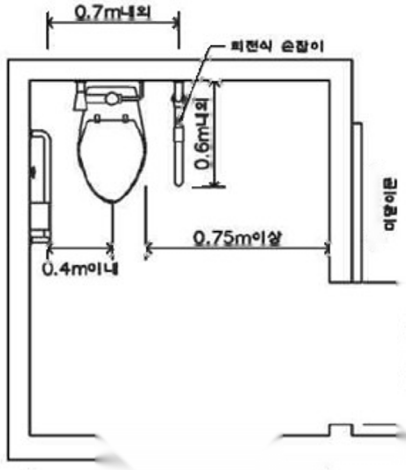
(e) Rule decomposition results

Figure 2: Intermediate results of the proposed pipeline for graphical rule decomposition. (a) Original composite graphical rule. (b) Annotation segmentation results generated by SAM3. (c) Grouping of annotation elements into rule-level clusters based on spatial proximity. (d) Base image obtained by removing annotation elements through inpainting. (e) Final rule-level graphical unit reconstructed by overlaying clustered annotations onto the base image.

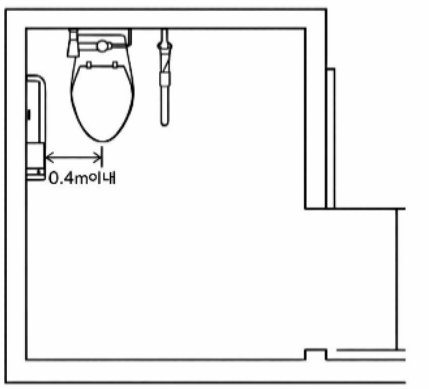
The experimental results indicate that the proposed method is less robust than the baseline in strictly decomposing composite diagrams into visually clean single-rule images. This limitation is particularly evident in challenging input cases, such as the example shown in Figure 3-(a), where failures occur during the annotation segmentation and grouping stages.



(a) A composite graphical rule for the decomposition



(b) A decomposition output from the proposed method



(c) A decomposition output from the baseline method

Figure 3: Example of a composite graphical rule (a) where the proposed method (b) fails to isolate single requirements, while the baseline succeeds in isolating single requirements and preserving the regulatory meaning.

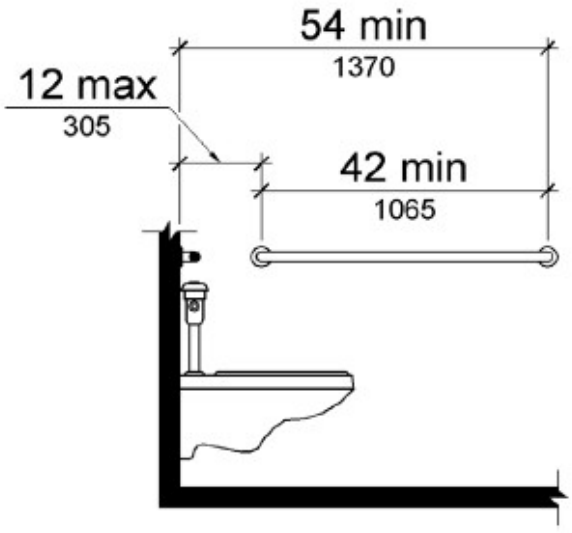
As shown in Figure 3, the baseline successfully isolates the regulation specifying that the spacing between the handrail and the toilet should be approximately 0.4 m, yielding a clear, self-contained graphical rule. In contrast, the diagram produced by the proposed method retains residual non-target graphical elements in the base drawing, indicating that unrelated annotations were not fully removed. Consequently, the resulting image does not form a fully isolated single graphical rule.

This failure mode is consistently observed in source images with low resolution or very small annotation elements, where reliable segmentation and grouping become particularly difficult.

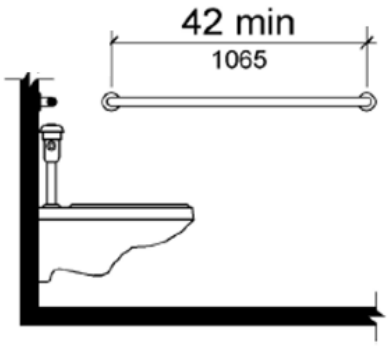
Despite this limitation, the proposed method demonstrates a complementary strength in preserving the regulatory semantics of the original graphical rules, as shown in Figure 4, where prompt-based baselines result in semantic distortion.

The diagram generated by the proposed method preserves the essential spatial relationships and numerical constraints required to convey the target regulation, despite minor visual imperfections introduced during the annotation isolation process. In contrast, the baseline produces a visually clean but semantically incorrect diagram that conveys a substantially different regulatory interpretation from the ground truth.

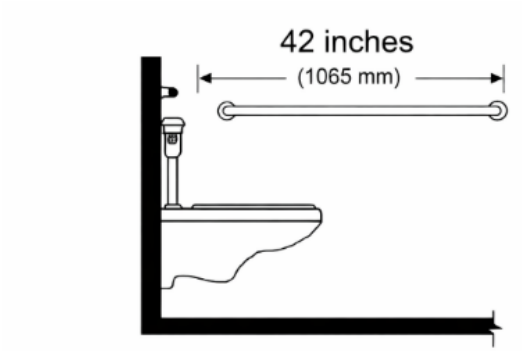
This tendency is also supported by the “same as ground truth” metric in Table 1, where the proposed method achieves 82%, slightly outperforming the baseline (78%). Although the numerical improvement is modest, it indicates that the proposed method more reliably preserves the intended regulatory meaning.



(a) A composite graphical rule for the decomposition



(b) A decomposition output from the proposed method



(c) A decomposition output from the baseline method

Figure 4: Example of a composite graphical rule (a) where the baseline (c) fails to preserve the regulatory meaning, while the proposed method (b) maintains semantic consistency with the ground truth.

Taken together, these results reveal a clear trade-off between the two approaches. The baseline is more effective at producing visually clean, strictly isolated single-rule diagrams, whereas the proposed method is more reliable at preserving regulatory semantics by maintaining the original geometric structure.

This behavior can be attributed to the reliance of the baseline on generative image synthesis, which reconstructs the diagram from scratch. While this enables strong visual simplification, it also increases susceptibility to hallucinations and unintended misinterpretation of regulatory constraints, leading to semantic drift.

In contrast, the proposed method operates by decomposing and reconstructing the original diagram, which can compromise visual cleanliness in some cases but helps prevent semantic distortion.

Conclusions

This paper presented a zero-shot framework for decomposing composite graphical rules into atomic graphical rules and pairing them with corresponding textual rules. By leveraging text-prompt-guided zero-shot segmentation with SAM3, context-aware inpainting for faithful graphical reconstruction, and instruction-only prompt chaining for textual rule decomposition, the proposed method enables rule-level understanding of graphical regulations without requiring labeled training data.

Experimental results suggest that the proposed method exhibits a trade-off compared to prompt-based image generation approaches. While it is less robust in enforcing visually clean single-rule isolation, it tends to better preserve regulatory intent and semantic consistency with the ground truth.

The evaluation combines semantic assessment using an MLLM-as-a-judge framework with segmentation-level metrics (e.g., IoU), but remains limited in several aspects, including reliance on a single MLLM-based evaluation, the absence of human validation and inter-rater reliability analysis, and a relatively small dataset.

Future work will incorporate human expert evaluation, in which domain experts define ground-truth interpretations of graphical rules and assess whether the decomposed outputs correctly match these references. Such an evaluation will enable more reliable validation of the proposed method.

Moreover, the current study focuses on dimensional technical drawings in which regulatory requirements are primarily conveyed through distance-based annotations, specifically within accessibility regulations for disabled restroom facilities. While this scope reflects a common form of graphical regulation, the experimental validation remains limited to a single facility type. In particular, the present study does not yet account for variations in drawing configurations. Therefore, extending the framework to a wider range of facility conditions and drawing variations, as well as to other types of graphical rules, remains for future work.

Acknowledgments

This work was supported by the Hans Fischer Senior Fellowship program at the Technical University of Munich – Institute of Advanced Studies (TUM-IAS) and the TUM Georg Nemetscheck Institute Artificial Intelligence for the Built World (GNI).

References

Amor, R. & Dimyadi, J. (2021) The promise of automated compliance checking. Developments in the Built Environment, 5, 100039.

Carion, N., Gustafson, L., Hu, Y.-T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., Afouras, T., Mavroudi, E., Xu, K., Wu, T.-H., Zhou, Y., Momeni, L., Hazra, R., Ding, S., Vaze, S., Porcher, F., Li, F., Li, S., Kamath, A., Cheng, H.K., Dollár, P., Ravi, N., Saenko, K., Zhang, P. & Feichtenhofer, C. (2025) SAM 3: Segment Anything with Concepts. arXiv preprint, arXiv:2511.16719.

Chen, D., Chen, R., Zhang, S., Liu, Y., Wang, Y., Zhou, H., Zhang, Q., Wan, Y., Zhou, P. & Sun, L. (2024) MLLM-as-a-Judge: Assessing Multimodal LLM-as-a-

Judge with Vision-Language Benchmark. arXiv preprint, arXiv:2402.04788.

Chen, N., Lin, X., Jiang, H. & An, Y. (2024) Automated Building Information Modeling Compliance Check through a Large Language Model Combined with Deep Learning and Ontology. Buildings, 14(7), 1983.

Dori, D. & Velkovitch, Y. (1998) Segmentation and Recognition of Dimensioning Text from Engineering Drawings. Computer Vision and Image Understanding, 69, p.pp.196–201

Goswami, K., Mathur, P., Rossi, R., Dernoncourt, F. (2025) PlotEdit: Natural Language-Driven Accessible Chart Editing in PDFs via Multimodal LLM Agents. arXiv preprint, arXiv:2501.11233.

Hettiarachchi, H., Dridi, A., Gaber, M.M., Parsafard, P., Bocaneala, N., Breitenfelder, K., Costa, G., Hedblom, M., Juganaru-Mathieu, M., Mecharnia, T., Park, S., Tan, H., Tawil, A.-R.H. & Vakaj, E. (2025) CODE-ACCORD: A Corpus of building regulatory data for rule generation towards automatic compliance checking. Scientific Data 12, 170

Hou, Y., Giledereli, B., Tu, Y. & Sachan, M. (2025) Do Vision-Language Models Really Understand Visual Language? arXiv preprint, arXiv:2410.00193.

Iversen, O. & Huang, L. (2026) Leveraging large language models for BIM-based automated compliance checking. Automation in Construction, 182, 106707.

Jeong, D., Kang, D., Park, J., Lee, H. & Paik, J. (2025) Structure-Preserving Zero-Shot Image Editing via Stage-Wise Latent Injection in Diffusion Models. arXiv preprint, arXiv:2504.15723.

Khan, M.T., Yong, Z., Chen, L., Tan, J.M., Feng, W. & Moon, S.K. (2025) Automated Parsing of Engineering Drawings for Structured Information Extraction Using a Fine-tuned Document Understanding Transformer. arXiv preprint, arXiv:2505.01530.

Kim, Y., Borrmann, A. & Lee, G. (2025) A Preliminary Study on Design Rule Derivation from Graphical Representations Using Multimodal Large Language Models. In: Proceedings of the 2025 European Conference on Computing in Construction, Porto, Portugal. European Council on Computing in Construction.

Li, Y., 2026. An Adaptive LLM-Based Agent Framework with Self-Extending Tool Library for Automated Building Code Compliance Checking. Master's thesis, Technical University of Munich, Munich, Germany.

Ma, N., Labib, R., Amor, R., Chong, A., Fan, C., Forth, K., Fu, X., Fuchs, S., Hong, T., Klimenkova, N., Koo, J., Li, S., McCullough, S.T., Park, J.Y., Shraga, R., Yoon, S., Zhang, L. & Zhang, Y. (2026) Ten questions concerning large language models (LLMs) for building applications. Building and Environment, 291, p.114260.

Madireddy, S., Gao, L., Din, Z.U., Kim, K., Senouci, A., Han, Z. & Zhang, Y. (2025) Large Language Model-Driven Code Compliance Checking in Building Information Modeling. Electronics 14, 2146.

Maupou, C., Yang, Y., Fodop, G., Qie, Y., Migliorini, C., Mehdi-Souzani, C. & Anwer, N. (2024) Automatic raster engineering drawing digitisation for legacy parts towards advanced manufacturing. Procedia CIRP 129, 234–239.

Ministry of Health and Welfare (MOHW) (2023) The enforcement rule of the act on promotion for persons with disabilities, the elderly, and pregnant women. Appendix 1: detailed standards for structure and materials of convenience facilities. National Law Information Center of Korea.

Picard, C., Edwards, K.M., Doris, A.C., Man, B., Giannone, G., Alam, M.F. & Ahmed, F. (2025) From concept to manufacturing: evaluating vision-language models for engineering design. Artificial Intelligence Review, 58, 288.

Lin S. & Ting C. (1997) A new approach for detection of dimensions set in mechanical drawings. Pattern Recognitition Letters, 18, p.pp.367–373.

U.S. Department of Justice (DOJ). 2010. ADA Standards for Accessible Design. U.S. Department of Justice. U.S. Department of Justice, Washington, DC

Wu, C., Yin, S., Qi, W., Wang, X., Tang, Z. & Duan, N. (2023) Visual ChatGPT: Talking, Drawing and Editing with Visual Foundation Models. arXiv preprint, arXiv:2303.04671.

Yang, D., Zhang, L., Yue, Z., Chen, L., Xu, Y., Wang, W. & Jin, Q. (2025) ChartM3: Benchmarking Chart Editing with Multimodal Instructions. https://doi.org/10.48550/arXiv.2507.21167

Ying, H. & Sacks, R. (2024) From Automatic to Autonomous: A Large Language Model- driven Approach for Generic Building Compliance Checking. arXiv preprint. In Proceedings of the CIB W78 Conference.

Zhang, W., Joseph, J., Yin, Y., Xie, L., Furuhata, T., Yamakawa, S., Shimada, K. & Kara, L.B (2023) Component segmentation of engineering drawings using Graph Convolutional Networks. Computers in Industry, 147, p.103885

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

SHARED EMBEDDING SPACES FOR AEC ONTOLOGIES: AN EMPIRICAL STUDY OF DISCOVERY AND ALIGNMENT

Tobias Heimig-Elschner[1,2], Jyrki Oraskari[2,3], and Jakob Beetz[2]
[1]Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR), Germany
[2]Design Computation, RWTH Aachen University, Germany
[3]Cluster of Excellence CARE, TU Dresden and RWTH Aachen University, Germany

Abstract

The adoption of ontologies in the AEC domain is increasing, yet ontology reuse remains constrained because identifying suitable ontologies and concepts is still largely manual. This paper investigates whether shared embedding spaces can capture semantic relatedness across AEC ontologies, thereby supporting ontology discovery and alignment. We test three hypotheses regarding expert-aligned similarity, the spatial proximity of related classes and the superiority of structure-aware over purely lexical models. We embedded several widely used AEC ontologies and evaluated ontology and class-level similarities. Results suggest that embeddings operationalize expert intuition and lower barriers to ontology reuse.

Introduction

A significant obstacle to ontology reuse is how many ontologies are designed: they often lack adequate documentation and provide no human-readable specifications, thereby imposing unnecessary friction on anyone attempting to adopt or extend them (Fernández-López et al., 2019). Many built-environment ontologies have been created recently, but users find it difficult to locate suitable ones, which limits their practical application (Akbarieh et al., 2025). Identifying suitable ontologies and concepts is still largely manual.

We assume that ontologies and their constituent entities can be represented in a vector space such that semantic relatedness systematically correlates with geometric proximity. We explore hypotheses about how the embeddings of lexical ontology notations relate to expert assessments and known alignments. Specifically, we suggest that:

- **H1 (Ontology-level similarity correlates with expert recommendations).** If ontology embeddings capture domain structure, then similarity in the embedding space should align with expert judgments of task or domain suitability.
- **H2 (Semantically corresponding classes are closer than unrelated classes).** Useful ontology embeddings must preserve class-relatedness.
- **H3 (Context-enriched lexical representations improve embedding quality).** Ontologies provide structural context, such as class connectivity and local relations.

We evaluate our hypotheses by embedding a curated set of widely used AEC ontologies and their classes using both text-based (labels + descriptions) and graph-based (RDF/OWL) strategies. These embeddings are employed for similarity-based ontology search and class-level alignment suggestions. Expert-guided evaluations and quantitative analyses of the embedding space reveal how closely the embeddings reflect expert judgment and semantic structure, and demonstrate that they can streamline ontology engineering workflows by lowering barriers to ontology reuse and integration in construction informatics.

Our main contributions include: (i) an LLM embedding–based, ontology-level semantic search framework; (ii) a class-level alignment method using similarity neighborhoods for an ontology base; (iii) an expert-aligned evaluation approach for comparing and assessing embedding spaces and (iv) the validation and development of a foundational embedding hypothesis for AEC ontologies.

Related Work

Linked Open Vocabularies (LOV) is a curated registry of RDF vocabularies that shows how they link, tracks usage in the Linked Data cloud, and offers search, SPARQL, and API access to find and reuse vocabularies. Despite strong reuse of LOV's core vocabularies, domain-specific vocabularies remain underlinked within and across domains (Maratsi et al., 2023). More recently, Akbarieh et al. (2025) presented BE-OLS, a built-environment-specific ontology lookup service that catalogs and organizes over 100 ontologies in the building environment domain with structured metadata and search/filter functions. It provides a custom, domain-aware evaluation framework (connectivity, accessibility, and documentation/reuse). Current limitations include an incomplete ontology collection, manually applied evaluation, and no LLM-based assistance for ontology search and analysis.

LLM embeddings (Tao et al., 2024) used in semantic searches encode text as high-dimensional vectors that capture semantic meaning. Models like BERT or all-MiniLM-L6-v2 (Reimers and Gurevych, 2024) convert text into vector representations, ensuring that semantically similar content is aligned closely and facilitating comparison based on meaning rather than exact words.

Wang et al. (2015) presented a method for short text categorization that combines semantic clustering with convolutional neural networks (CNN). Although newer tech-

niques such as transformers and large language models (LLMs) now outperform this approach, it remains effective for lightweight applications. This method employs a small, supervised CNN operating on semantic matrices, in contrast to LLMs that produce embeddings via unsupervised transformer pretraining with attention mechanisms, which offer richer context but require more computational resources.

A more recent method, introduced by Khan et al. (2020), uses sentence-embedding-based clustering to summarize discussion threads. It leverages Word2Vec[1] to create sentence embeddings and applies K-medoids clustering—a partitioning technique that selects actual data points as cluster centers—to group replies that are semantically similar. A ranking mechanism based on quality-related textual features then determines the most relevant responses for the summary. This fully automatic, domain-independent approach performs well on precision, recall, and F-measure. However, its reliance on Word2Vec and traditional clustering may limit its ability to capture complex semantic nuances compared with more advanced models, such as large language models (LLMs).

Teymurova et al. (2024) introduced OWL2Vec4OA, a version of OWL2Vec* tailored for ontology alignment by integrating confidence-weighted seed mappings into the random-walk process. This makes the learned embeddings more alignment-aware and improves matching performance in experiments. Its main limitations are dependence on seed-mapping quality, added computational complexity, and sensitivity to hyperparameters.

Methodology

Problem Setting and Representation Framework

To make the decisions made in representation and evaluation explicit, we introduce the following formal setting:

Let an ontology artifact be denoted by O. Let $\mathcal{D} = \{O_1, \ldots, O_N\}$ denote the ontology dataset, i.e. a finite collection of ontology artifacts processed jointly.

For each ontology O we assume a materialized RDF graph:

$$\mathcal{G}_O = (V, E) \tag{1}$$

where V denotes the set of RDF nodes and E the set of triples with predicates P. A triple is written as $(s, p, o) \in E$ with $s, o \in V$ and $p \in P$. Based on this, let C_O be the set of classes declared in O (e.g., resources typed as `owl:Class` or `rdfs:Class`) and P_O be the set of properties declared in O (e.g., `owl:ObjectProperty`, `owl:DatatypeProperty`, `rdf:Property`).

We define the set of representable entities [2] at both dataset and ontology level:

$$\mathcal{X}_O := \{O\} \cup C_O \cup P_O \tag{2}$$

$$C_{\mathcal{D}} := \bigcup_{O \in \mathcal{D}} C_O, \quad P_{\mathcal{D}} := \bigcup_{O \in \mathcal{D}} P_O, \quad \mathcal{X}_{\mathcal{D}} := \bigcup_{O \in \mathcal{D}} \mathcal{X}_O. \tag{3}$$

The formalism for representation and embeddings is defined as follows: Let $\mathcal{T}$ be the space of textual representations (strings), $f : \mathcal{T} \to \mathbb{R}^d$ be an embedding function and $\mathrm{t}_m : \mathcal{X} \to \mathcal{T}$ be a representation variant m. The embedding of an entity $x \in \mathcal{X}$ under variant m is:

$$\mathrm{v}_x^{(m)} := f(\mathrm{t}_m(x)) \tag{4}$$

All embeddings of interest for a fixed representation variant m live in a shared dataset-level embedding space

$$\mathcal{E} := \mathcal{E}_m = \{\mathrm{v}_x^{(m)} \mid x \in \mathcal{X}_{\mathcal{D}}\}. \tag{5}$$

Let $\delta(\cdot, \cdot)$ denote cosine distance on $\mathcal{E}$.

For any entity $x \in \mathcal{X}$, define three optional lexical fields: $\ell(x)$ – label or title string (identifier-like), $d(x)$– description or definition string (content-like) and $c(x)$– comment or remark string (additional notes). Extraction is performed via documented, stable predicate lists (implementation detail), but the formalism assumes only that $\ell(x), d(x), c(x)$ are string-valued or empty.

Use-case descriptions U are treated as external textual queries embedded via the same function f, but are not elements of $\mathcal{X}_O$.

Throughout the remainder of this section, $\mathbb{E}[\cdot]$ denotes the empirical expectation with respect to the sampling procedure specified in each context. Unless stated otherwise, sampling is performed uniformly over dataset-level entities $x \in \mathcal{X}_{\mathcal{D}}$ (or over constrained subsets such as relation-defined pairs).

Definition of a Well-Formed Embedding Space

Based on the foundational hypothesis that: *There exists an embedding space in which semantic relatedness between ontologies and ontology entities is reflected by geometric proximity*. We define a set of geometric (*G1–G3* and semantic properties *S1 & S2* that suggest that such an embedding space is well-formed. The following criteria define necessary geometric and semantic properties that an embedding space must satisfy to be considered suitable for ontology discovery and alignment tasks. G1–G3 are treated as necessary but not sufficient conditions for semantic validity.

G1 – Global Discriminability

In high-dimensional spaces, distances tend to concentrate, making nearest-neighbor relations unreliable; measuring distance contrast provides a global diagnostic of whether the embedding space preserves discriminative power (Beyer et al., 1999; Aggarwal et al., 2001). A well-formed embedding space must exhibit sufficient distance contrast, such that semantic distinctions are not lost due to high-dimensional distance concentration. This is operationalized via a distance contrast measure introduced by Beyer et al. (1999):

[1] `https://code.google.com/archive/p/word2vec/`

[2] When the context is clear, we write $\mathcal{X}$ for $\mathcal{X}_{\mathcal{D}}$ (dataset-level entities) and use $\mathcal{X}_O$ when referring to a specific ontology.

$$D = \{\delta(\mathrm{v}_i, \mathrm{v}_j) \mid i \neq j\}. \tag{6}$$

$$RC_q = \frac{\mathbb{E}[D] - Q_5(D)}{Q_5(D)} \tag{7}$$

where δ ranges over all pairwise distances in $\mathcal{E}$, with i, j indexing embeddings corresponding to entities $x_i, x_j \in \mathcal{X}_\mathcal{D}$.

G2 – Local Semantic Coherence

Relative neighborhood statistics remain meaningful in high-dimensional spaces because they rely on comparative distances rather than absolute magnitudes (Beyer et al., 1999; Radovanović et al., 2009). They serve as a robust replacement for classical clusterability tests (e.g. Hopkins (Hopkins and Skellam, 1954)), which degrade in high dimensions (Adolfsson et al., 2019). A well-formed embedding space must exhibit locally coherent neighborhoods, such that nearby vectors are significantly closer than randomly chosen vectors. To capture local coherence robustly, we adopt *relative k-nearest-neighbor distance ratio*, which compares expected distances to k nearest neighbors (k-NN) against those drawn from random pairs, in the spirit of neighborhood preserving metrics used in embedding evaluation (Boggust et al., 2022).

$$\bar{\delta}_k(x) := \frac{1}{k} \sum_{x' \in k\text{-NN}(x)} \delta(\mathrm{v}_x, \mathrm{v}_{x'}) \tag{8}$$

$$R_k = \frac{\mathbb{E}[\bar{\delta}_k(x)]}{\mathbb{E}[\delta(\mathrm{v}_x, \mathrm{v}_{x'})]} \quad \text{with} \quad R_k \ll 1 \tag{9}$$

where k-NN(x) denotes the set of the k nearest neighbors of x in $\mathcal{E}$, (x, x') a uniformly sampled random pair of entities, and expectations are taken over uniformly sampled entities $x \in \mathcal{X}_\mathcal{D}$. The parameter k controls the neighborhood size.

G3 – Non-Degenerated Dimensionality

Effective dimensionality captures whether embeddings meaningfully occupy a structured manifold rather than degenerating into noise or redundancy (Recanatesi et al., 2022; Giaffar et al., 2024) which can be assessed via the *participation ratio (effective dimensionality)* as proposed by Recanatesi et al. (2022):

$$D_{\text{eff}} = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2} \tag{10}$$

where λ_i are the eigenvalues of the covariance matrix of the embedding vectors in $\mathcal{E}$.
Following this, a well-formed embedding space satisfies:

$$1 \ll D_{\text{eff}} \ll d \tag{11}$$

While G1–G3 assess geometric well-formedness independently of domain semantics, S1 and S2 evaluate whether the embedding space aligns with ontological meaning and expert judgment.

S1 – Ontology-Aware Semantic Alignment

An intrinsic characteristic of ontologies is the explicit declaration of semantic relations between classes (intra- and inter-ontology), such as equivalence mappings, subsumption, and domain-specific predicates. A meaningful embedding space should preserve these relations, embedding semantically related classes closer together than unrelated ones. Intrinsic evaluation frameworks for ontology embeddings therefore emphasize explicit measurement of relational preservation (Alshargi et al., 2019).
Formally, let $\mathcal{R}$ denote the set of declared semantic relations between classes in the ontology dataset $\mathcal{D}$:

$$\mathcal{R} := \bigcup_{O \in \mathcal{D}} \mathcal{R}_O \cup \mathcal{R}_{\text{cross}}, \quad \mathcal{R} \subseteq C_\mathcal{D} \times C_\mathcal{D} \tag{12}$$

where $\mathcal{R}_O$ contains intra-ontology class pairs and $\mathcal{R}_{\text{cross}}$ inter-ontology class pairs. Class pairs $(i, j) \in C_\mathcal{D} \times C_\mathcal{D}$ are induced by predicate assertions (i, p, j) or (j, p, i). No logical inference or reasoning closure is applied. Let the binary semantic relatedness indicator $s_{ij} \in \{0, 1\}$ be

$$s_{ij} := \begin{cases} 1 & \text{if } (i, j) \in \mathcal{R}, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Define the sets of *related* and *unrelated* class pairs:

$$P_{\text{rel}} = \{(i, j) \mid s_{ij} = 1\}, \quad P_{\text{unrel}} = \{(i, j) \mid s_{ij} = 0\} \tag{14}$$

Pairs in P_{unrel} are uniformly sampled such that $|P_{\text{rel}}| = |P_{\text{unrel}}|$. Let $d_{ij} := \delta(\mathrm{v}_i, \mathrm{v}_j)$.
A well-formed embedding space satisfies

$$\mathbb{E}_{(i,j) \in P_{\text{rel}}}[d_{ij}] \; < \; \mathbb{E}_{(i,j) \in P_{\text{unrel}}}[d_{ij}]. \tag{15}$$

We further define the alignment ratio

$$R_{\text{align}} := \frac{\mathbb{E}_{(i,j) \in P_{\text{unrel}}}[d_{ij}]}{\mathbb{E}_{(i,j) \in P_{\text{rel}}}[d_{ij}]} \tag{16}$$

$R_{\text{align}} > 1$ indicates ontology-aware semantic alignment. Statistical significance is assessed via a stratified class-identity permutation test ($K = 1000$), where class identities are shuffled within ontology strata while distances are held fixed. The empirical one-sided p-value $P(\rho_{\text{null}} \geq \rho_{\text{obs}} \mid H_0)$ is computed from the resulting null distribution of Spearman $\rho(-d_{ij}, s_{ij})$.

S2 – External Semantic Consistency

While S1 evaluates semantic alignment intrinsic to the ontology structure, external semantic consistency assesses whether embedding-based similarity aligns with expert intuition in practical ontology selection scenarios. If an embedding space is well-formed, ontologies deemed relevant by domain experts for a given use case should appear among the most similar ontologies according to embedding-based rankings. Ontology-level similarity is computed using lexical and context-enriched lexical representations. Ontology embeddings are obtained directly from the corresponding textual representations.

Let U denote a use-case description and O the set of candidate ontologies. Let $E_U \subseteq O$ denote the set of ontologies identified as relevant by domain experts for use case U, allowing multiple ontologies to be considered equally relevant.

Let $r_{\text{embed}}(U) = (o_1, o_2, \dots)$ be the ontology ranking induced by embedding similarity between U and ontology representations, where $o_i \in O$. Define the set of top-k retrieved ontologies using precision-at-k and recall-at-k using the top-k set $R_k(U) := \text{TopK}(r_{\text{embed}}(U), k)$:

$$P_k(U) = \frac{1}{k} \sum_{o \in R_k(U)} 1[\, o \in E_U \,], \tag{17}$$

$$R_k(U) = \frac{1}{|E_U|} \sum_{o \in R_k(U)} 1[\, o \in E_U \,]. \tag{18}$$

Here, Precision@k measures the proportion of relevant ontologies among the top-k retrieved results, while Recall@k measures the proportion of all expert-relevant ontologies retrieved within the top-k. Since expert relevance sets may contain fewer than k ontologies, the maximum achievable precision is bounded by $|E_U|/k$, while recall remains unaffected.

To assess whether retrieval performance exceeds chance, observed $P_k(U)$ values are compared against a random baseline. Under random selection, the number of relevant ontologies among the top-k follows a hypergeometric distribution, with expected precision $\mathbb{E}[P_k(U)] = |E_U|/|O|$. A well-formed embedding space exhibits consistently high precision and recall values across use cases, indicating that expert-relevant ontologies are ranked among the most similar results, and achieves precision values that systematically exceed the corresponding random baseline, reflecting performance beyond chance.

Ontology Representations and Embeddings

In this paper, we define and evaluate two complementary lexical representations for ontologies O and their associated classes C_O. These representations rely solely on human-readable lexical information such as labels, descriptions, and comments, or enrich this information with bounded contextual signals derived from related entities. In addition, we include a purely structural RDF2Vec-based representation - as introduced by Ristoski and Paulheim (2016) - as an established baseline focusing on graph structure, used for comparative evaluation under criteria *G1–G3* and *S1*.

Lexical embedding v_{lex}

The *lexical representation* variant t_{lex} captures only the immediate human-readable lexical documentation attached to an entity, without incorporating any explicit structural context. For any $x \in \mathcal{X}$:

$$t_{\text{lex}}(x) := \text{concatFields}\big(\ell(x),\ d(x),\ c(x)\big) \tag{19}$$

Here, $\text{concatFields}(\cdot)$ denotes deterministic concatenation with a fixed ordering and separators.

The representation t_{lex} is deterministic, structure-independent, and fully interpretable. It serves as the lexical baseline for evaluating embedding quality and semantic alignment.

$$v_{\text{lex}}(x) = f(t_{\text{lex}}(x)) \text{ for all } x \in \mathcal{X} \tag{20}$$

Context-Enriched Lexical embedding v_{conlex}

The *context-enriched lexical representation* variant t_{conlex} extends t_{lex} by incorporating bounded and deterministic lexical context derived from related entities, while remaining text-based. For an ontology O, let $\bigoplus$ denote deterministic aggregation (concatenation) with a fixed ordering:

$$t_{\text{conlex}}(O) := \text{concatFields}\Big(t_{\text{lex}}(O),\ \bigoplus_{c \in C'_O} t_{\text{lex}}(c)\Big) \tag{21}$$

where $C'_O \subseteq C_O$ is a fixed, deterministic subset of representative classes. In this work, classes are ranked by first-order connectivity (number of RDF edges), with ties resolved by alphabetical ordering of IRIs. The top $\min(50, |C_O|)$ classes are selected.

For a class $c \in C_O$:

$$t_{\text{conlex}}(c) := \text{concatFields}\Big(t_{\text{lex}}(c),\ \bigoplus_{p \in P_c^{(n)}} t_{\text{lex}}(p)\Big) \tag{22}$$

where $P_c^{(n)} \subseteq P_O$ denotes properties connected to c within RDF graph depth n under a fixed neighborhood policy; depth $n = 1$ is used by default.

The representation t_{conlex} remains deterministic and human-interpretable, while introducing controlled local contextual signal in a reproducible manner.

$$v_{\text{conlex}}(x) = f(t_{\text{conlex}}(x)) \text{ for all } x \in \mathcal{X} \tag{23}$$

Structural Embedding v_{struct}

In addition to the lexical representation variants, we include a purely structural embedding baseline based on RDF2Vec. In contrast to v_{lex} and v_{conlex}, which follow the mapping $v_x^{(\bullet)} = f(t_\bullet(x))$, the structural variant directly induces embeddings from RDF graph structure. Formally, v_{struct} is defined only for class-level entities $c \in C_{\mathcal{D}}$, yielding embeddings $\mathrm{v}_c^{(\text{struct})} \in \mathbb{R}^d$. Ontology-level embeddings are derived via aggregation over class-level embeddings:

$$\mathrm{v}_O^{(\text{struct})} := \frac{1}{|C_O|} \sum_{c \in C_O} \mathrm{v}_c^{(\text{struct})}. \tag{24}$$

The resulting embeddings capture global structural patterns of the RDF graph, are stochastic but reproducible under fixed seeds, and are not directly human-interpretable.

Results

Implementation and Dataset

Implementation

The methodology was instantiated as a modular processing pipeline with separated responsibilities for identity management, semantic storage, and representation. Ontology identity and provenance are maintained via a JSON-based registry. Resolved ontology serializations are loaded into an Apache Jena Fuseki triple store[3], which serves as the authoritative semantic state and supports SPARQL-based querying. Normalized ontology artifacts and intermediate results are persisted as file-based materializations to ensure reproducibility. Vector representations derived from textual and graph-based encodings are stored separately in a Weaviate vector database, allowing symbolic semantic data and learned embeddings to be managed independently while remaining linked through stable ontology identifiers. Vector representations are derived from both textual and graph-based encodings. Textual embeddings are computed using the transformer-based model `snowflake-arctic-embed2` (1024 dimensions), selected as a representative open-weight model for semantic similarity tasks Yu et al. (2024). Structural embeddings are generated using RDF2Vec (100 dimensions) with random walks (depth $d = 4$, 100 walks per entity) and a Word2Vec backend, consistent with commonly used configurations in prior RDF2Vec studies. Traversal excludes high-degree hub predicates (e.g., `rdf:type`) to reduce structural noise Ristoski and Paulheim (2016). Full embedding configurations are provided in the appendix [4] .

Dataset

From the 144 ontologies listed in the Built Environment Ontology Lookup Service (BE-OLS) (Akbarieh et al., 2025), 101 ontologies provided a reachable reference or download URL. Of these, 100 could be successfully resolved and parsed. Two ontologies were fully contained as subgraphs of others already included and were therefore not treated as independent artifacts. The resulting dataset comprises 98 distinct ontologies.

To characterize the availability of lexical information, ontologies were classified as **meta** (no class definitions), **formal** (classes without lexical signals, i.e., only blank nodes or unlabeled IRIs), **structural** (fewer than 25% of classes provide at least one lexical signal (e.g., `rdfs:label`, `skos:prefLabel`, `rdfs:comment`, `dc:description`)), or **lexical** (at least 25% of classes provide lexical information). In addition, we assessed whether available lexical information follows common, human-readable documentation patterns suitable for downstream processing by large language models.

For inclusion in the vectorization experiments, ontologies were required to provide ontology-level lexical information, a resolvable serialization (i.e., not metadata-only), and a lexical profile classified as either structural or lexical. Applying these criteria yields a final subset of 84 ontologies, comprising 7,086 classes.

Evaluation of the Embedding Space

This section reports the results for three global well-formedness criteria (G1–G3) introduced above. All evaluations are reported for ontology-level and class-level embeddings and for all three representation variants (Lex, ConLex, Struct).

G1 – Global Discriminability

G1 evaluates global discriminability by measuring distance contrast across the embedding space. All representations satisfy this criterion at both the ontology and class levels, indicating that none of the embedding spaces suffers from distance concentration effects. Structural embeddings exhibit the highest contrast, reflecting the sparse and the role-oriented nature of graph-based representations, while lexical and contextualized lexical embeddings show moderate but stable contrast characteristic of dense language-model spaces. The consistency of G1 across granularities confirms that subsequent local semantic analyses are not confounded by degenerate global geometry. (see Table 1)

Table 1: G1 – Global discriminability (distance contrast).

Level	v_{lex}	v_{conlex}	v_{struct}
O	0.71 ✓	0.55 ✓	2.18 ✓
C_O	0.64 ✓	0.66 ✓	2.22 ✓

G2 – Local Semantic Correspondence ($\tau = 0.5$)

Table 2: G2 – Local semantic correspondence for $k = 10$.

Level	v_{lex}	v_{conlex}	v_{struct}
O	0.68 ×	0.73 ×	0.55 ×
C_O	0.42 ✓	0.43 ✓	0.06 ✓

G2 reveals a clear granularity effect (see Table 2). While none of the representations satisfy the criterion at the ontology level, all representations consistently meet G2 at the class level. This indicates that semantic correspondence is preserved locally between individual ontology entities, but not necessarily when embeddings aggregate heterogeneous concepts at the ontology scale. Rather than indicating a deficiency, this result confirms that class-level embeddings are the appropriate resolution for evaluating fine-grained semantic relatedness, whereas ontology-level embeddings primarily capture broader thematic or structural similarity.

G3 – Non-Degenerated Dimensionality

G3 evaluates whether embedding spaces exhibit non-degenerate effective dimensionality (see Table 3). All representations pass this criterion at both the ontology and class levels, with effective dimensionalities substantially

[3] `https://jena.apache.org/documentation/fuseki2/`

[4] The appendix is provided online at `https://git-ce.rwth-aachen.de/Tobias1/ec3-2026-shared-embedding-spaces-for-aec-ontologies/-/tree/v1.0-ec3-paper-2026`.

greater than one and far below the nominal embedding dimension. Lexical and contextualized lexical embeddings occupy a relatively low- dimensional subspace of the original 1024-dimensional space, reflecting semantic compression through aggregation, while structural embeddings exhibit even more compact but information-dense representations. The increase in effective dimensionality from ontology to the class level aligns with higher semantic granularity and complements the G2 findings.

Table 3: G3 – Effective dimensionality of embedding spaces.

Level	v_{lex}	v_{conlex}	v_{struct}
O	25.58 ✓	36.30 ✓	16.82 ✓
C_O	59.47 ✓	65.26 ✓	13.20 ✓

S1 – Ontology-Aware Semantic Alignment

All representations achieve statistically significant alignment ($R_{\text{align}} > 1$, $p < 10^{-3}$), with semantically related pairs consistently closer than unrelated ones. Lexical and context-enriched representations show moderate alignment ($R_{\text{align}} \approx 1.55$), while the structural embedding yields substantially stronger separation ($R_{\text{align}} = 2.94$), indicating that graph-based context reinforces semantic correspondence. The observed rank correlations ($\rho \approx 0.56$–0.60) confirm robust but non-trivial alignment between embedding distances and relatedness (see Table 4).

Table 4: S1 - Ontology-Aware Semantic Alignment Evaluation.

Metric	v_{lex}	v_{conlex}	v_{struct}
R_{align}	1.56	1.55	2.94
Spearman ρ	0.58	0.56	0.60
p_{perm}	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
$\mu_{\rho,\text{null}}$	0.08	0.06	0.05
Significance	✓	✓	✓

S2 - External Semantic Consistency

Across the five evaluated use cases, both lexical and contextualized lexical representations achieve precision@5 values above the random baseline, indicating that embedding-based retrieval captures some degree of expert-relevant signal. However, absolute performance remains low (P@5 ≈ 0.24–0.28), and no statistically significant results are observed across use cases. Lexical representations exhibit higher recall, while contextualized lexical representations achieve slightly higher precision (see Table 5), suggesting a trade-off between coverage and selectivity. Nevertheless, the overall performance indicates that ontology-level embeddings alone are insufficient to reliably reproduce expert ontology selection behavior under the tested conditions.

In contrast to S1, which evaluates intrinsic semantic alignment at the class level, S2 operates on highly aggregated ontology representations and sparse, non-ranked expert feedback, resulting in reduced discriminative power. Consequently, S2 does not provide confirmatory evidence for external semantic consistency in this setting.

Table 5: S2 – External semantic consistency (P@5 and R@5).

Metric	v_{lex}	v_{conlex}
Mean P@5	0.24	0.28
Mean R@5	0.30	0.15
Baseline $E[P@5]$	0.11	0.11
Baseline $E[R@5]$	0.07	0.07
$\Delta P@5$	0.13	0.17
$\Delta R@5$	0.23	0.08
Significant	0/5	0/5

Discussion

We investigated whether shared embedding spaces facilitate ontology discovery and alignment in the AEC domain (hypotheses H1–H3). All embedding variants met the geometric well-formedness criteria G1–G3, confirming stable, non-degenerate spaces. Semantic alignment at the class level (S1) was consistently observed across lexical, context-enriched lexical, and structural representations. In contrast, ontology-level semantic consistency (S2) yielded weak but above-random results, insufficient for reliable ontology selection. Thus, embedding-based similarity is valid primarily at the class-level abstraction, supporting the core hypothesis that semantic relatedness in ontologies is reflected by geometric proximity, while highlighting the abstraction levels where this relationship holds reliably.

With respect to H1, which hypothesizes that ontology-level embedding similarity correlates with expert recommendations, the results do not provide confirmatory evidence under the current evaluation setup. A primary limiting factor is the statistical underpowering of S2: unlike S1, which operates over thousands of class-level relation pairs, S2 relies on a small number of use cases and sparse, largely unranked expert relevance judgments. Combined with the strong aggregation inherent in ontology-level embeddings, this substantially reduces the sensitivity of retrieval-based evaluation measures such as precision@k and recall@k. Consequently, H1 cannot be rejected outright, but remains unconfirmed within the scope of this study and should be revisited using larger and more structured expert evaluations.

Beyond methodological considerations, the inconclusive results for H1 highlight a fundamental issue affecting ontology reuse. Ontology-level embeddings, derived solely from explicit information like labels, descriptions, and structure, fail to replicate expert selection behavior. This suggests that many ontologies lack sufficient information for reuse decisions, as experts also consider contextual factors like maturity, adoption, tooling, conventions, and experience—not encoded in standard documentation. This deficiency indicates broader communication issues within ontology ecosystems, echoing barriers caused by inadequate documentation. Additionally, S2 results imply that expert selection depends more on high-level lexical descriptions of scope and application than on detailed internal content, which is mainly reflected in class structure and relations.

The results strongly support H2, showing that semantic classes are embedded closer together when related, across all variants, and confirming that embedding spaces preserve fine-grained semantic relations within and across ontologies. This demonstrates their suitability for class alignment, correspondence discovery, and semantic exploration. The evaluation focuses on explicitly declared relations, isolating the effect of representation learning from reasoning variability. Including inferred relations could further improve semantic alignment by revealing implicit structures, highlighting the potential of knowledge graphs. The consistent results across different representations suggest that class-level semantic relatedness is a stable property of the embedding space, largely independent of the representation used.

The results support H3, which suggests that adding structural context improves embedding quality. Context-enriched lexical representations achieve results comparable to purely lexical ones in semantic alignment while remaining deterministic and interpretable, indicating that additional contextual aggregation does not consistently translate into improved alignment. Structural RDF2Vec embeddings have the highest alignment ratios, showing the importance of graph structure in maintaining ontological relations. Interestingly, large language model–based lexical embeddings already meet all geometric criteria (G1–G3) and achieve strong semantic alignment (S1), even though they weren't trained specifically for ontology embedding. This suggests that modern language models encode substantial latent domain knowledge, allowing them to recover meaningful semantic structures from sparse, diverse ontology data.

Taken together, these findings clarify the appropriate role of embedding-based methods in ontology engineering workflows. Embeddings should not be viewed as replacements for expert judgment at the ontology selection level (H1), but as powerful instruments for semantic exploration and alignment at the class level (H2), particularly when enriched with bounded structural context (H3). The context depth is intentionally limited to one hop ($n = 1$), focusing on directly related classes and properties while avoiding noise propagation and cyclic dependencies in graph traversal. Exploring deeper contextual representations may further enrich semantic signals, but introduces increased structural variance and is left for future work. By making explicit where semantic information is preserved and where it is absent, embedding-based representations both support ontology reuse and expose documentation gaps that warrant further attention.

Conclusion

This paper explored whether embedding-based representations can support ontology discovery and alignment in the AEC domain. Across all evaluated representations, the geometric well-formedness criteria (G1–G3) are satisfied, indicating that the resulting embedding spaces are stable and suitable for semantic analysis. At the class level, ontology-aware semantic alignment (S1) is consistently observed: semantically related classes are embedded closer together than unrelated ones, independent of representation choice. This confirms that embeddings preserve fine-grained semantic relationswithin and across ontologies and are well suited for class-level exploration and alignment tasks. These results further provide empirical support for the foundational hypothesis that semantic relatedness between ontology entities can, to a meaningful extent, be reflected by geometric proximity in a shared embedding space.

A key empirical finding is that LLM-based lexical embeddings perform surprisingly well in this setting, despite not being trained on ontologies or formal knowledge graphs. Even without explicit access to ontological structure, these embeddings satisfy all geometric criteria and exhibit statistically significant semantic alignment. While context-enriched lexical representations do not yield a significant improvement, incorporating structural information enhances alignment performance. Purely structural embeddings achieve the strongest alignment effects, albeit at the cost of interpretability.

At the ontology level, external semantic consistency with expert recommendations (S2) remains inconclusive. This result reflects both methodological limitations—strong aggregation and sparse expert feedback—and fundamental constraints on ontology reuse, where expert selection relies on contextual factors not explicitly encoded in ontology artifacts. Overall, the results support embeddings as a reliable semantic substrate for exploration and alignment at appropriate levels of abstraction, while clarifying their limitations for ontology-level decision-making.

Future Work

Future work should extend this study along three complementary directions. First, external semantic consistency (S2) should be evaluated using more mature and statistically robust designs, including a larger set of use cases, ranked or graded expert judgments, and repeated evaluations across expert groups. Such designs would enable stronger conclusions about ontology-level similarity and better isolate the factors influencing expert ontology selection. Second, the role of embeddings in concrete alignment tasks warrants systematic investigation. This includes evaluating their utility for candidate correspondence generation, semi-automated class alignment, and expert-in-the-loop alignment workflows, where embeddings serve as a semantic prior rather than a definitive decision mechanism. Third, embedding-based representations open the door to two-dimensional semantic mappings of the AEC ontology ecosystem. Future research should examine the stability, interpretability, and analytical value of such maps for ecosystem-level exploration, comparison, and gap analysis. Together, these directions would further clarify the scope and limits of embedding-based methods and support their principled integration into ontology engineering practice without overextending their claims.

Acknowledgments

Claims expressed in this article do not necessarily have to coincide with the positions of the BBR/BBSR. This work was partially funded by the DFG (German Research Foundation) under EXC 3115 - 533767731.

References

Adolfsson, A., Ackerman, M., and Brownstein, N. C. (2019). To cluster, or not to cluster: An analysis of clusterability methods. Pattern Recognition, 88:13–26.

Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001). On the Surprising Behavior of Distance Metrics in High Dimensional Space. In Goos, G., Hartmanis, J., Van Leeuwen, J., Van Den Bussche, J., and Vianu, V., editors, Database Theory — ICDT 2001, volume 1973, pages 420–434. Springer Berlin Heidelberg, Berlin, Heidelberg.

Akbarieh, A., Osadcha, I., Farghaly, K., Bosché, F., and Puust, R. (2025). Development of a built environment ontology lookup service (be-ols) with a new ontology evaluation. In Proceedings of the 2025 European Conference on Computing in Construction, volume 6 of Computing in Construction, Porto, Portugal. European Council on Computing in Construction.

Alshargi, F., Shekarpour, S., Soru, T., and Sheth, A. P. (2019). Metrics for evaluating quality of embeddings for ontological concepts. In AAAI Spring Symposium Combining Machine Learning with Knowledge Engineering.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When Is "Nearest Neighbor" Meaningful? In Goos, G., Hartmanis, J., Van Leeuwen, J., Beeri, C., and Buneman, P., editors, Database Theory — ICDT'99, volume 1540, pages 217–235. Springer Berlin Heidelberg, Berlin, Heidelberg.

Boggust, A., Carter, B., and Satyanarayan, A. (2022). Embedding Comparator: Visualizing Differences in Global Structure and Local Neighborhoods via Small Multiples. In 27th International Conference on Intelligent User Interfaces, pages 746–766, Helsinki Finland. ACM.

Fernández-López, M., Poveda-Villalón, M., Suárez-Figueroa, M. C., and Gómez-Pérez, A. (2019). Why are ontologies not reused across the same domain? Journal of Web Semantics, 57:100492.

Giaffar, H., Rullán Buxó, C., and Aoi, M. (2024). The effective number of shared dimensions between paired datasets. In Dasgupta, S., Mandt, S., and Li, Y., editors, Proceedings of the 27th International Conference on Artificial Intelligence and Statistics, volume 238 of Proceedings of Machine Learning Research, pages 4249–4257. PMLR.

Hopkins, Brian. and Skellam, J. G. (1954). A New Method for determining the Type of Distribution of Plant Individuals. Annals of Botany, 18(70):213–227.

Khan, A., Shah, Q., Uddin, M. I., Ullah, F., Alharbi, A., Alyami, H., and Gul, M. A. (2020). Sentence embedding based semantic clustering approach for discussion thread summarization. Complexity, 2020(1):4750871.

Maratsi, M. I., Alexopoulos, C., and Charalabidis, Y. (2023). A structured analysis of domain-specific linked open vocabularies (lov): Indicators for interoperability and reusability. In European, Mediterranean, and Middle Eastern Conference on Information Systems, pages 135–152. Springer.

Radovanović, M., Nanopoulos, A., and Ivanović, M. (2009). Nearest neighbors in high-dimensional data: The emergence and influence of hubs. In Proceedings of the 26th Annual International Conference on Machine Learning, pages 865–872, Montreal Quebec Canada. ACM.

Recanatesi, S., Bradde, S., Balasubramanian, V., Steinmetz, N. A., and Shea-Brown, E. (2022). A scale-dependent measure of system dimensionality. Patterns, 3(8):100555.

Reimers, N. and Gurevych, I. (2024). all-minilm-l6-v2: Sentence-transformers model. `https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2`. Accessed: 2025-11-26.

Ristoski, P. and Paulheim, H. (2016). RDF2Vec: RDF Graph Embeddings for Data Mining. In Groth, P., Simperl, E., Gray, A., Sabou, M., Krötzsch, M., Lecue, F., Flöck, F., and Gil, Y., editors, The Semantic Web – ISWC 2016, volume 9981, pages 498–514. Springer International Publishing, Cham.

Tao, C., Shen, T., Gao, S., Zhang, J., Li, Z., Hua, K., Hu, W., Tao, Z., and Ma, S. (2024). Llms are also effective embedding models: An in-depth overview. arXiv preprint arXiv:2412.12591. revised 2025.

Teymurova, S., Jiménez-Ruiz, E., Weyde, T., and Chen, J. (2024). Owl2vec4oa: Tailoring knowledge graph embeddings for ontology alignment. DOI: 10.48550/arXiv.2408.06310.

Wang, P., Xu, J., Xu, B., Liu, C., Zhang, H., Wang, F., and Hao, H. (2015). Semantic clustering and convolutional neural network for short text categorization. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 352–357.

Yu, P., Merrick, L., Nuti, G., and Campos, D. (2024). Arctic-embed 2.0: Multilingual retrieval without compromise.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A SEMI-AUTOMATED SCAN-TO-BIM APPROACH TO FIRE COMPARTMENTATION IN EXISTING BUILDINGS

Jorge Magalhães[1], Rafael Cabral[1], Jéssica Reis[2], Rui Gavina[2], José Aidos Rocha[3], Vaidotas Sarka[4], Tatjana Vilutiene[4], Ricardo Santos[1,5] and Diogo Ribeiro[1,5]

[1]iBuilt, Instituto Superior de Engenharia do Porto, Politécnico do Porto, Porto, Portugal

[2]VN2R-Engineering Innovation Consulting LDA, Porto, Portugal

[3]Exactusensu - Consultores Associados, Lda, Porto, Portugal

[4]Vilnius Gediminas Technical University, Faculty of Fundamental Sciences, Department of Engineering Graphics, Vilnius, Lithuania

[5]CONSTRUCT, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

Abstract

Fire safety compliance checks in existing buildings are often labor-intensive and prone to error, due to the lack of design data and the reliance on traditional inspection tools. This work proposes an alternative to conventional approaches through a semi-automated Scan-to-BIM methodology to create a BIM model directly from point cloud data. The methodology demonstrates robustness when applied to benchmark datasets. The results confirm the efficiency and accuracy of the methodology in developing early design and BIM models suited to support fire safety engineering and verification of compliance with fire safety regulations.

Introduction

According to the Portuguese National Emergency and Civil Protection Authority (ANEPC – Autoridade Nacionalk de Emergência e Proteção Civil) (Civil, 2024), 6,844 urban fires were confirmed, 603 of which occurred in unoccupied buildings of Portugal. Most of these buildings tend to be old and generally lack digital or even digitized geometrical information. These constraints create significant difficulties on the development of retrofitting projects, as building characteristics are typically obtained using traditional/manual methods that require excessive human intervention. One possible solution to overcome this challenge is the use of remote sensing methods, which enable the creation of point cloud data for generating *as is* geometric information (Oliveira et al., 2023, Ventura et al., 2026). For many building types, e.g. residential, office, or hotel, laser scanning is the standard, as it can accurately capture detailed reality with accurate dimensions in relatively straightforward field campaigns. The obtained raw data is the input for digital models' generation that can be further enriched to create BIM models and digital twins.

Fire safety is a critical aspect of building management, encompassing a comprehensive set of measures aimed at preventing and controlling fires Schönfelder et al. (2024). Among these measures, this work focuses on fire compartmentation requirements, in which a building is subdivided into compartments to isolated occupants and/or confine fire spread, thereby enhancing occupant safety. To address this issue, the present study proposes a semi-automated Scan-to-BIM methodology for generating BIM models promoting the implementation and assessment of fire compartmentation.

The Scan-to-BIM approach has become a prominent research topic in recent years. The scientific community has proposed several methods and alternatives to automate this process, though several approaches are still semi-automated and need manual intervention to identify structural elements (Zbirovský and Nežerka, 2025). The key step of this methodology is the contour extraction process. The point cloud contour extraction of a building seeks to obtain a clear description of the building's geometry, including aspects like shape, size, spatial positioning and internal layout. The existing building point cloud contour extraction methods mainly include image-based methods, geometric feature-based methods, and surface segmentation-based methods (Chen et al., 2025). Image-based methods consist of converting point cloud data into images for processing (Gao et al., 2024, Zbirovský and Nežerka, 2025). Geometric feature-based methods rely on extracting planes and line segments (Wang et al., 2022, Li et al., 2023), where the Random Sample Consensus (RANSAC) algorithm is the most used option for this end. Surface segmentation-based methods can be done by traditional point cloud segmentation strategies (edge detection, graph theory, clustering, hypervoxel) and Artificial Intelligence (AI) techniques (Chen et al., 2026, Mehranfar et al., 2024).

The transition toward full BIM adoption begins with the conversion of these models into the Industry Foundation Classes (IFC) standard, where building elements are explicitly classified into semantic entities and can be further enriched with structured parameters and properties. Zbirovský and Nežerka (2025) developed an open-source Scan-to-BIM tool, named Cloud2BIM, that automates the conversion of large-scale point clouds into BIM models in the IFC format. More recently, Chen et al. (2026) developed a framework that combines deep learning-based semantic segmentation with unsupervised clustering to generate models in IFC from point cloud data. This step is particularly relevant in the fire-safety domain, where Scan-to-BIM workflows are often limited to geometric reconstruction. Converting reconstructed models into IFC enables fire-safety engineers to

systematically augment *as is* models with domain-specific information, such as the fire resistance classes of construction elements and the fire reaction classes of linings the fire reaction classes of linings, supporting analysis, compliance verification, and decision-making processes.

Based on these findings and considering the diversity of building typologies, the automated generation of BIM models from existing buildings continues to be a prominent research topic and remains a challenging task. Moreover, point cloud data acquisition in buildings that are still in service often results in increased noise levels and incomplete geometric information, which can adversely affect the robustness of Scan-to-BIM frameworks. In the fire safety domain, the availability of BIM models for existing buildings can significantly enhance fire safety analyses, particularly in the assessment of compartmentation, where retrofit interventions may alter the original geometric layout.

The present work proposes a semi-automated Scan-to-BIM approach suitable for retrofit projects. The resulting digital models provide a semantic volumetric representation of the building structure at schematic design level, corresponding to a Level of Development (LOD) 200 (American Institute of Architects and buildingSMART International), and are intended to support decision-making processes related to fire compartmentation requirements. The conversion of point cloud data into BIM is evaluated using the IFC format, with a specific focus on the entities IFCSpace, IFCWall, and IFCSlab.

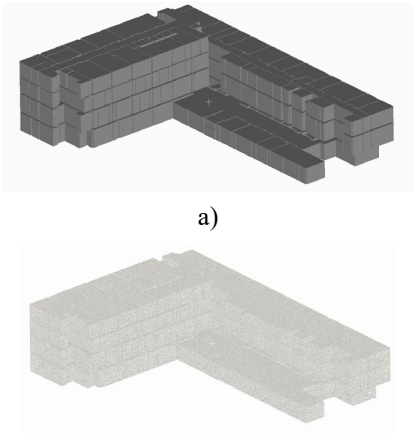
Proposed Methodology

The methodology is structured by four main steps, as shown on Figure 1. The step-by-step explanation will be illustrated by its implementation on a synthetic dataset composed of a part of Vilnius Gediminas Technical University (VILNIUS TECH). This data is characterized as an early design BIM model similar to indoor point clouds of empty buildings. The development was based on Python algorithms and applicable to multi-level Manhattan-world structures (Mehranfar et al., 2024).

Point cloud acquisition and registration

Point cloud acquisition and registration is characterized by transferring information from the real world into an available digital format. Terrestrial laser scanning (TLS) is typically the most suitable method for acquiring indoor point clouds. Normally, this procedure is performed in the proprietary software of the scanner brand used.

In the case of synthetic data, acquisition is performed through mesh disassociation, where a 3D model is converted into a point cloud based on a specified point density or number of points. Figure 2 presents the case study that supports the results of the application of the proposed framework.



a)

b)

Figure 2. VILNIUS TECH: a) Early design BIM model; b) point cloud.

Point cloud processing

Point cloud processing begins with a global reorientation of the dataset based on the clustering of estimated surface normal vectors. Surface normals were computed using a KD-tree–based hybrid neighborhood search, combining k-nearest neighbor (KNN) and radius-based criteria. Subsequently, the estimated absolute normal vectors were grouped using the k-means clustering algorithm.

The resulting clusters represent dominant surface orientations within the point cloud and are used to infer a consistent global orientation of the dataset. This procedure guarantees that the point cloud is centered and aligned with the orthogonal axis, which prevents aleatory

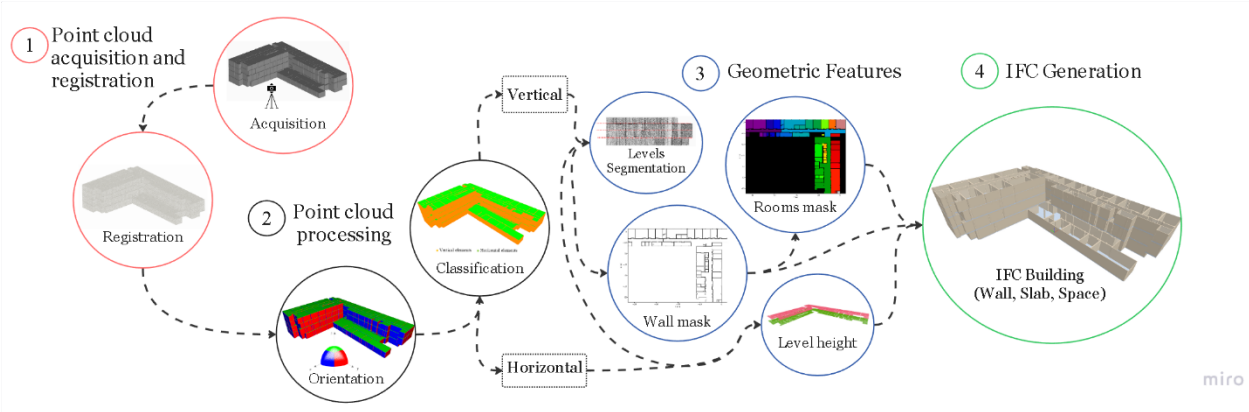


Figure 1. Flowchart of proposed methodology

position of the data in 3D space facilitating subsequent steps. The aligned point cloud is subjected to a second normal estimation stage to enable geometric classification into vertical and horizontal elements, as illustrated in Figure 3. After classification, a statistical outlier removal filter and a voxel downsampling is applied to clean each point cloud and eliminate spurious points.

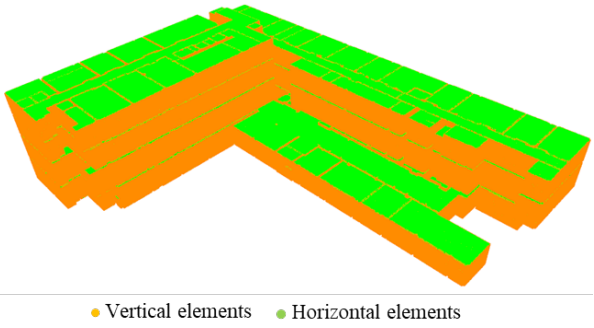


Figure 3. Point cloud classification.

Geometric Features

To automatically identify building storey levels, a height-based segmentation is performed using the vertical-point subset of the point cloud. Local minimum of a smoothed density function are used as cut elevations by two consecutive height intervals. After determining the set of level boundaries, each input point cloud class (e.g., vertical, horizontal) is segmented per level. On real point clouds, the quantity of noise can hinder the acquisition of entire contours of the existing rooms. On this work, the contour extraction for synthetic data is acquired from a z section of the vertical point cloud. Further ahead, on the Experimental analysis and discussion section, this extraction will be presented in more detail.

The point cloud contour extracted is given by:

$$\mathcal{P} = \{(x_k, y_k, z_k)\}_{k=1}^{N} \quad (1)$$

Where (x_k, y_k, z_k) represent the coordinate points of the point cloud contour then transformed into a 2D occupancy grid by discretizing the horizontal spatial domain into square cells of fixed resolution r. Each point projected onto the horizontal plane, and the grid cell c_{ij} is considered occupied (wall) if the number of projected points falling inside the cell (p_i) exceeds a predefined threshold τ:

$$c_{ij} = \begin{cases} 1, if\ p_{ij} \geq \tau \\ 0, otherwise \end{cases} \quad (2)$$

This procedure yields a binary wall mask that is robust to noise and suitable for image-based operations. These operations ensure that wall regions form coherent, topologically closed boundaries suitable for spatial reasoning. From the refined wall mask, a complementary free-space mask is obtained by logical inversion. To distinguish interior spaces from the exterior environment, a border-seeded flood fill is applied. This step guarantees that only enclosed regions bounded by walls are considered as candidate interior spaces. Therefore, a vectorization stage is applied to transform room masks into polygonal room boundaries (Figure 4), which can subsequently be extruded into 3D solids and IFCSpace.

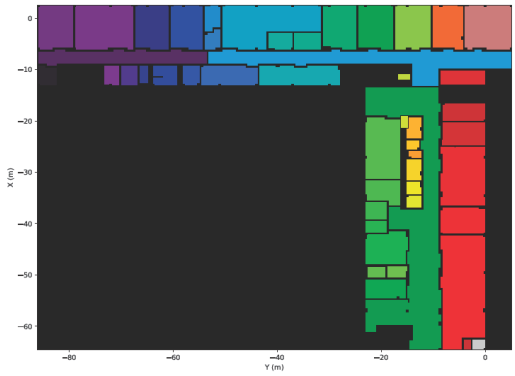


Figure 4. Room mask.

The binary regions are vectorized into metric polygons by enumerating boundary edges of occupied cells and colinear vertices are removed to yield simplified orthogonal loops. Ultimately, to mitigate corner gaps caused by discretization and later wall thickening, segments connected at junction endpoints are extended by half the nominal wall thickness, improving closure when generating IFCWall.

For level height, the RANSAC algorithm is applied on the horizontal point cloud to characterize the level. These geometric features are used to set the height of IFCSpace and to generate IFCSlabs.

IFCBuilding generation

The IFC generation was implemented using the ifcopenshell library. This stage instantiates a schema-compliant IFC4 spatial hierarchy (IfcProject–IfcSite–IfcBuilding–IfcBuildingStorey). For each storey, elements are positioned through hierarchical IfcLocalPlacement, ensuring consistent coordinate handling. The reconstructed 2D room boundaries are converted into IfcPolyline-based closed profiles and extruded along the +Z direction to form IfcSpace solids with a fixed height extracted from horizontal point cloud classification. Wall and slab elements are generated analogously by extruding closed profiles into IfcWall and IfcSlab solids; slabs are placed with a negative Z-offset so that their top surface coincides with the storey elevation, except on the first floor. As a result, a structured and information-rich IFC model is created, enabling BIM-based analysis, as can be observed in Figure 5.

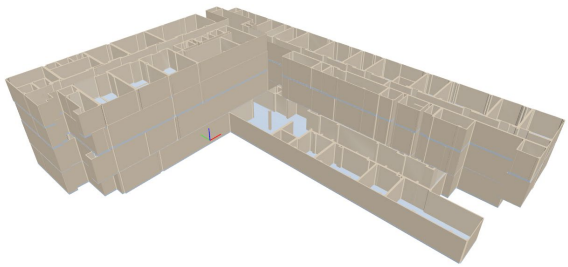


Figure 5. BIM model generated

Experimental analysis and discussion

This section presents the application of the proposed methodology to three public TLS datasets. The input point cloud, the resulting early design model and BIM model,

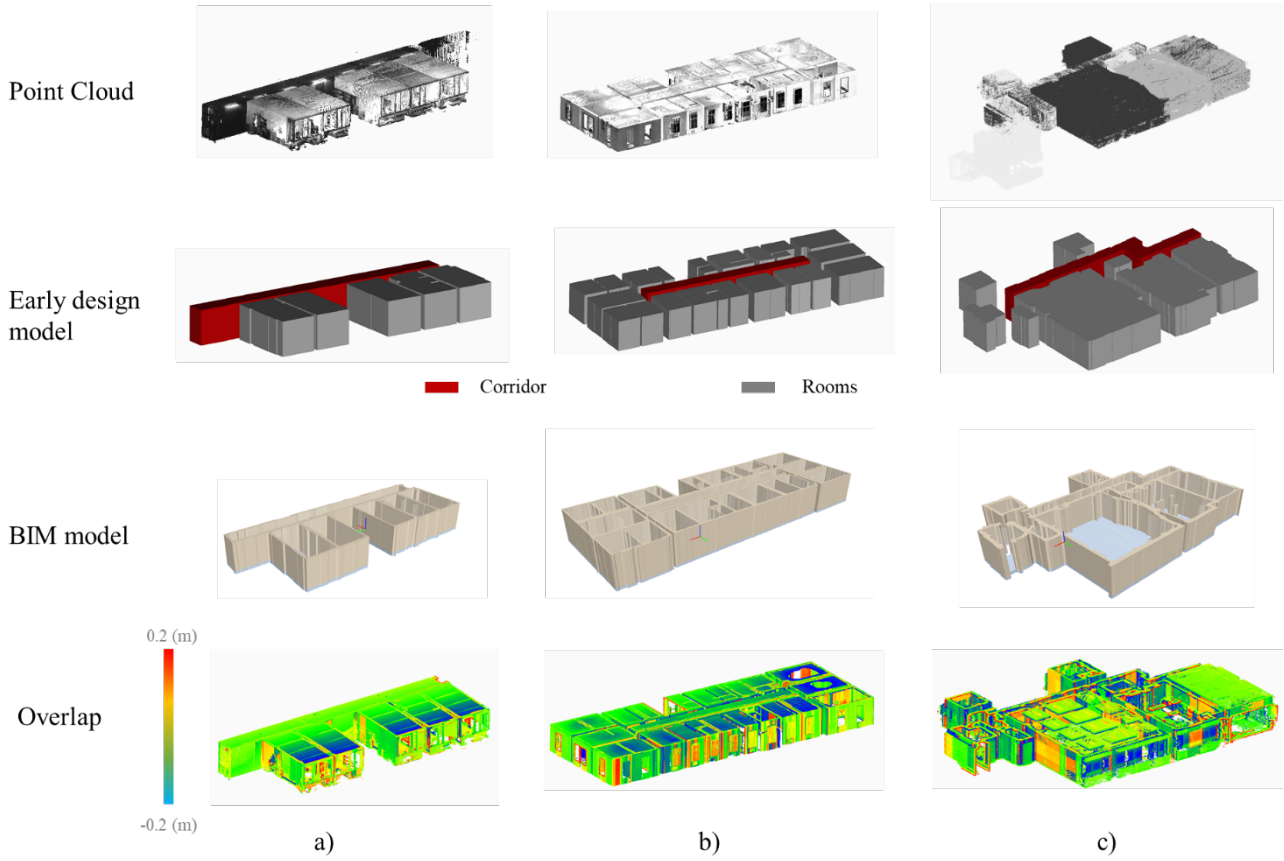


Figure 6. Experimental analysis of proposed methodology on real point clouds: a) University of Zurich (room3); b) Kladno railways station and c) University of Melbourne.

as well as the corresponding overlap results are shown in Figure 6.

For this type of data, the noise level of the point cloud has a significant impact on the contour extraction stage and consequently on the final result. To account for the diversity of building geometries, several trials were conducted to obtain contours that are both as clean and as complete as possible. In addition to vertical and horizontal surfaces, an inclined point cloud was extracted from the normal-based classification, comprising points whose unit normal vectors fall outside the predefined angular thresholds used to distinguish vertical and horizontal elements. This subset mainly represents transitional regions between horizontal and vertical surfaces and can reveal space-level separations.

Regarding the results obtained from the public datasets, the case studies of the University of Zurich (University of Zurich) (Figure 6a) and Kladno station (Urban, 2024) (Figure 6b) presents a good accuracy across most elements, achieving a mean distance of 0.059 m and 0.048 m, respectively. Nevertheless, variations in ceiling heights are visible. When the methodology was applied to a more complex dataset, such as the University of Melbourne case (Khoshelham et al., 2017) (Figure 6c), a reduction in performance was observed by a mean distance of 0.136 m. This dataset exhibits a higher noise level and includes a room that is not fully orthogonal, revealing limitations of the current approach. These results highlight the need for additional filtering of the extracted segments and for the generation of cleaner and smoother wall masks.

Contour extraction is therefore identified as the most critical stage of the entire process and remains the focus of ongoing research, with the objective of developing a method capable of handling a wider range of building typologies.

For the fire compartmentation evaluation, the early design model was manually modified to differentiate corridor spaces from rooms. This initial semantic distinction establishes the foundation for subsequent enrichment of the BIM model, particularly enabling the association of corridor-adjacent walls with fire-rated elements and supporting the definition of evacuation routes.

The proposed methodology ensures scalability and interoperability by transitioning from manual classification to a fully automated space classification framework, which builds on both property-based and relationship-based modeling strategies within the IFC schema.

At the property level, individual spaces (IfcSpace) may be enriched through dedicated property sets, such as Pset_SpaceCommon and Object Type, which explicitly encode whether a space is classified as a private room or a circulation area. This approach ensures compatibility with common BIM authoring and analysis tools while enabling straightforward querying and filtering operations. Complementarily, a higher-level semantic

grouping can be established through the use of IfcSpatialZone entities, allowing spaces to be aggregated into functional zones. This relationship-based structuring enhances the model's capacity to support fire safety analyses, including compartmentation logic, evacuation route modelling, and accessibility assessments, while preserving flexibility for multi-criteria classifications.

The transition towards automated classification is supported by a topology-driven strategy based on spatial adjacency. In this context, each space identified in the room mask (Figure 4) will be evaluated according to its connectivity with neighbouring spaces, derived from shared boundaries. Spaces exhibiting a number of adjacent connections above a predefined threshold will be interpreted as circulation elements (e.g., corridors or transitional zones), whereas spaces with limited connectivity are classified as enclosed functional units (e.g., rooms or offices). This graph-based interpretation of the spatial structure aligns with the inherent relational nature of the IFC schema and enables the derivation of higher-level semantic information from purely geometric and topological inputs. Furthermore, such an approach supports extensibility towards more advanced classification criteria, including geometric properties, accessibility patterns, or integration with external ontologies and classification systems.

Conclusions

The lack of architectural designs, detailed floor plans and 3D models for older buildings is a significant challenge in Portugal. To address this problem, this work presents a semi-automated scan-to-BIM methodology designed to support fire safety analysis in retrofitting projects.

The proposed methodology was first developed and described using a synthetic dataset and was subsequently validated on three different real-world point clouds from benchmark datasets. The results demonstrate the methodology's sensitivity when applied to noisy data, where the generated BIM models may contain irregularities. Nevertheless, validation across multiple datasets highlights the framework's primary advantage: its robustness and adaptability to diverse real-world conditions.

Conducted within the scope of the FireBIM project, this work provides a foundational contribution to fire safety analysis, specifically in assessing compartmentation within existing buildings. Furthermore, the resulting early-design model can be utilized to assist in planning fire safety evacuation routes, such as checking evacuation distances, width of exits and stairways, and determining the number of exits based on the capacity of the building's premises.

Future work will focus on developing more robust strategies for extracting accurate spatial contours from dense and noisy point clouds. This step is identified as critical to the entire workflow, as incomplete contours prevent the generation of successful BIM models. Moreover, an automated approach for BIM model enrichment will be incorporated, aimed at ensuring compliance with evacuation and compartmentation requirements. Additionally, the automated characterization of key building elements, such as stairs, doors, and openings, will be investigated to support connectivity modelling and fire safety analysis.

Acknowledgments

This work has been funded by the NORTE 2030 Programme through the European Regional Development Fund (ERDF), under project FireBIM, reference NORTE2030-FEDER-00358400. Also, this work was supported by: UID/04708/2025 and https://doi.org/10.54499/UID/04708/2025, of the CONSTRUCT - Instituto de I&D em Estruturas e Construções - funded by Fundação para a Ciência e a Tecnologia, I.P./ MECI, through the national funds.

References

American Institute of Architects and Buildingsmart International. Level of Development (LOD) Specification [Online]. Available: https://bimforum.org/resource/lod-level-of-development-lod-specification/ [Accessed 2026].

Chen, T., Wang, H., Li, D., Liu, J., Liu, P., Wu, Z., Xu, C. & Zhang, M. (2026). An automated framework for converting point cloud data to building information modeling with segmentation and refinement. Engineering Applications of Artificial Intelligence, 163, 112915, https://doi.org/10.1016/j.engappai.2025.112915

Chen, X., Mao, J., Zhao, B., Tao, W., Qin, M. & Wu, C. (2025). Building contour extraction from hypervoxel growth point cloud surface neighborhood azimuth geometric features. Journal of Building Engineering, 101, 111914, https://doi.org/10.1016/j.jobe.2025.111914

Civil, A. N. D. E. E. P. (2024). Anuário ANEPC [Online]. Available: https://prociv.gov.pt/pt/documentacao/anuario-de-seguranca-contra-incendio-em-edificios-2024/ [Accessed 2026].

Gao, X., Yang, R., Chen, X., Tan, J., Liu, Y., Wang, Z., Tan, J. & Liu, H. (2024). A New Framework for Generating Indoor 3D Digital Models from Point Clouds. 16, 3462,

Khoshelham, K., Díaz Vilariño, L., Peter, M., Kang, Z. & Acharya, D. (2017). THE ISPRS BENCHMARK ON INDOOR MODELLING [Online]. Copernicus Publications. Available: https://isprs-archives.copernicus.org/articles/XLII-2-W7/367/2017/ [Accessed XLII-2/W7].

Li, F., Shi, W., Tu, Y. & Zhang, H. (2023). Automated methods for indoor point cloud preprocessing: Coordinate frame reorientation and building exterior removal. Journal of Building Engineering, 76, 107270, https://doi.org/10.1016/j.jobe.2023.107270

Mehranfar, M., Braun, A. & Borrmann, A. (2024). From dense point clouds to semantic digital models: End-to-end AI-based automation procedure for Manhattan-world structures. Automation in Construction, 162, 105392, https://doi.org/10.1016/j.autcon.2024.105392

Oliveira, R., Ribeiro, D., Oliveira, P. & Gavina, R. (2023). Laser Scanner in Construction Quality Control of Steel Industrial Buildings. 28, 04023055, doi:10.1061/PPSCFX.SCENG-1389

Schönfelder, P., Aziz, A., Bosché, F. & König, M. (2024). Enriching BIM models with fire safety equipment using keypoint-based symbol detection in escape plans. Automation in Construction, 162, 105382, https://doi.org/10.1016/j.autcon.2024.105382

University of Zurich. VMML, ETH Zurich Office3 Point Cloud Dataset [Online]. Available: https://www.ifi.uzh.ch/en/vmml/research/datasets.html [Accessed 2026].

Urban, R. (2024). Cloud2BIM benchmark point cloud dataset [Online]. Available: https://zenodo.org/records/14221915 [Accessed 2026].

Ventura, J., Magalhães, J., Jorge, T., Oliveira, P., Santos, R., Cabral, R., Araújo, L., Moreira, R. F., Oliveira, R. & Ribeiro, D. (2026). Geometric Monitoring of Steel Structures Using Terrestrial Laser Scanning and Deep Learning. 26, 831,

Wang, Q., Zhu, Z., Chen, R., Xia, W. & Yan, C. (2022). Building Floorplan Reconstruction Based on Integer Linear Programming. 14, 4675,

Zbirovský, S. & Nežerka, V. (2025). Open-source automatic pipeline for efficient conversion of large-scale point clouds to IFC format. Automation in Construction, 177, 106303, https://doi.org/10.1016/j.autcon.2025.106303

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

3D SEGMENTATION AGENT INTEGRATING 3D GAUSSIAN SPLATTING AND VISION-LANGUAGE MODELS FOR AS-BUILT ANALYSIS

Boyu Wang[1], Jun Ma[1], and Borja García de Soto [2]
[1]The University of Hong Kong, Hong Kong, China
[2]New York University Abu Dhabi (NYUAD), Abu Dhabi, United Arab Emirates

Abstract

The transition from raw as-built data to semantic object representations is a bottleneck in construction digitalization, currently hampered by manual dependencies and the "black-box" nature of existing point cloud segmentation networks. This paper introduces an autonomous self-correcting 3D segmentation agent tailored for construction environments, utilizing 3D Gaussian Splatting (3DGS) as the underlying representation and Vision Language Models (VLMs) as the reasoning core. Unlike static segmentation methods, our agent adopts a dynamic approach by performing active scene roaming. Through a closed-loop feedback mechanism, the VLM analyzes visual data, detects objects, and rigorously assesses the plausibility of segmentation results. The agent iteratively corrects its own errors until a high-fidelity, logically consistent segmentation is achieved. This method demonstrates superior interpretability and accuracy, enabling downstream applications such as as-built verification, progress tracking, and quantity takeoff. By bridging generative rendering with semantic reasoning, our approach provides a robust framework for automated 3D scene understanding.

Introduction

Construction automation is progressively evolving through four stages: data acquisition, semantic understanding, human-robot interaction, and fully autonomous construction. While the latter stages remain nascent, the industry has made significant strides in the first stage. Data acquisition has matured from manual logging to high-precision 3D laser scanning and, more recently, robust mobile LiDAR-visual Simultaneous Localization and Mapping (SLAM) systems. These technologies accurately capture the 3D geometry of construction sites. However, the resulting data remains unstructured. To enable automated tasks like as-built modeling and quality assessment, this raw geometric data must be transformed into structured, semantic representations.

Semantic 3D segmentation is the critical bridge for this transformation, yet it faces two significant barriers in construction contexts: 1) Generalization gaps**:** Conventional models trained on structured datasets (e.g., offices) often fail in dynamic, cluttered construction environments due to domain shifts. 2) Limited robustness and interpretability: "Black-box" deep learning models lack physical reasoning capabilities, often yielding results that are geometrically or topologically implausible.

To address these system-level challenges, we propose a 3D segmentation agent integrating 3D Gaussian Splatting (3DGS) and vision-language models (VLMs) inspired by the cognitive workflow of human engineers. Unlike end-to-end models, human experts adopt an iterative strategy: they first perform a rough identification, evaluate the result's plausibility (e.g., checking for completeness and topological connectivity), and actively manipulate their view including zooming in, rotating, or translating to resolve ambiguities before finalizing the segmentation.

Mimicking this paradigm, our framework, illustrated in Figure 1, integrates a 3DGS Rendering Engine with two core modules. Unlike 'black-box' neural networks that rely on implicit latent spaces, our approach achieves high interpretability by decomposing the segmentation task into an explicit, human-readable cognitive workflow. The Analysis Module utilizes VLM-based detection and topological reasoning to assess segmentation quality ("Plausibility Analysis"). If the result is unsatisfying or ambiguous, the system triggers the Operation Module. This module executes active visual manipulation such as multi-scale zooming, 3D rotation, and cropping to acquire clearer details, thereby achieving a robust, self-correcting understanding of the 3D scene.

Literature Review

Deep Learning-Based 3D Segmentation

Point cloud segmentation has evolved from manual processing to automated deep learning. Fully supervised methods, employing MLP, convolutional, or attention-based architectures, have been extensively applied across AEC scenarios ranging from Mechanical, Electrical, and Plumbing (MEP) systems to bridges (Wang et al., 2025). However, their reliance on massive, labor-intensive annotations severely limits scalability.

To mitigate data costs, weakly supervised and few-shot paradigms were developed, utilizing sparse labels or prototype alignment. Yet, these approaches often compromise accuracy and remain constrained to specific target domains.

Ultimately, these traditional methods share a critical limitation: they are "closed-set" systems. They lack the open-vocabulary reasoning required to interpret the diverse, unstructured elements found in dynamic construction sites without task-specific retraining. This bottleneck necessitates a paradigm shift toward large-scale foundation models, which liberate segmentation from fixed label sets and enable generalized, zero-shot 3D understanding.

Vision-Language Models (VLMs)

VLMs have revolutionized perception, shifting from closed-set tasks to open-vocabulary learning. While early models like CLIP (Radford et al., 2021) focused on global alignment, recent advancements like Grounding DINO (Liu et al., 2024) and SAM 3 (Carion et al., 2025) integrate semantic reasoning with precise localization. To extend this to 3D, researchers employ 2D-to-3D projection, mapping either segmentation masks or feature embeddings from multi-view images onto point clouds (Wang et al., 2024). These methods have shown promise in AEC applications ranging from indoor scanning to infrastructure inspection.

However, a significant bottleneck persists: VLMs remain inherently 2D-centric. They struggle with complex 3D construction environment, where single-view occlusion and background clutter hinder accurate detection. This limitation suggests a synergistic opportunity: leveraging 3DGS to render simplified, optimal viewpoints, thereby enabling VLMs to progressively analyze the environment and derive accurate segmentation results.

3D Gaussian Splatting (3DGS)

3D scene reconstruction has transitioned from discrete point clouds, which suffer from sparsity and artifacts, to continuous Neural Radiance Fields (NeRF) (Mildenhall et al., 2021). While NeRF achieves photorealism, its implicit nature incurs high computational costs. 3DGS bridges this gap by representing scenes with explicit, anisotropic 3D Gaussians (Kerbl et al., 2023).

Technically, 3DGS optimizes Gaussian parameters (position, opacity, covariance) via differentiable rasterization, projecting 3D ellipsoids into 2D image space for real-time rendering. Its adaptive density control dynamically densifies or prunes primitives to capture fine details efficiently. Subsequent advancements like Scaffold-GS (Lu et al., 2024) have further enhanced geometric consistency and surface alignment.

Despite these strides in visual fidelity, a critical gap remains: existing research prioritizes visual reproduction over structural interpretation. There is a distinct lack of methodologies leveraging 3DGS to semantically decompose and interpret complex civil engineering components, leaving this frontier largely unexplored.

Methodology

To address the challenges of segmenting complex 3D structures in cluttered environments, we propose a novel 3D segmentation agent. As illustrated in Figure 1, our framework establishes a dynamic synergy between the semantic reasoning capabilities of VLMs and the flexible rendering power of 3DGS. Unlike traditional static pipelines that process a fixed set of views, our agent mimics human visual inspection, treating 3D segmentation as an iterative, active perception process.

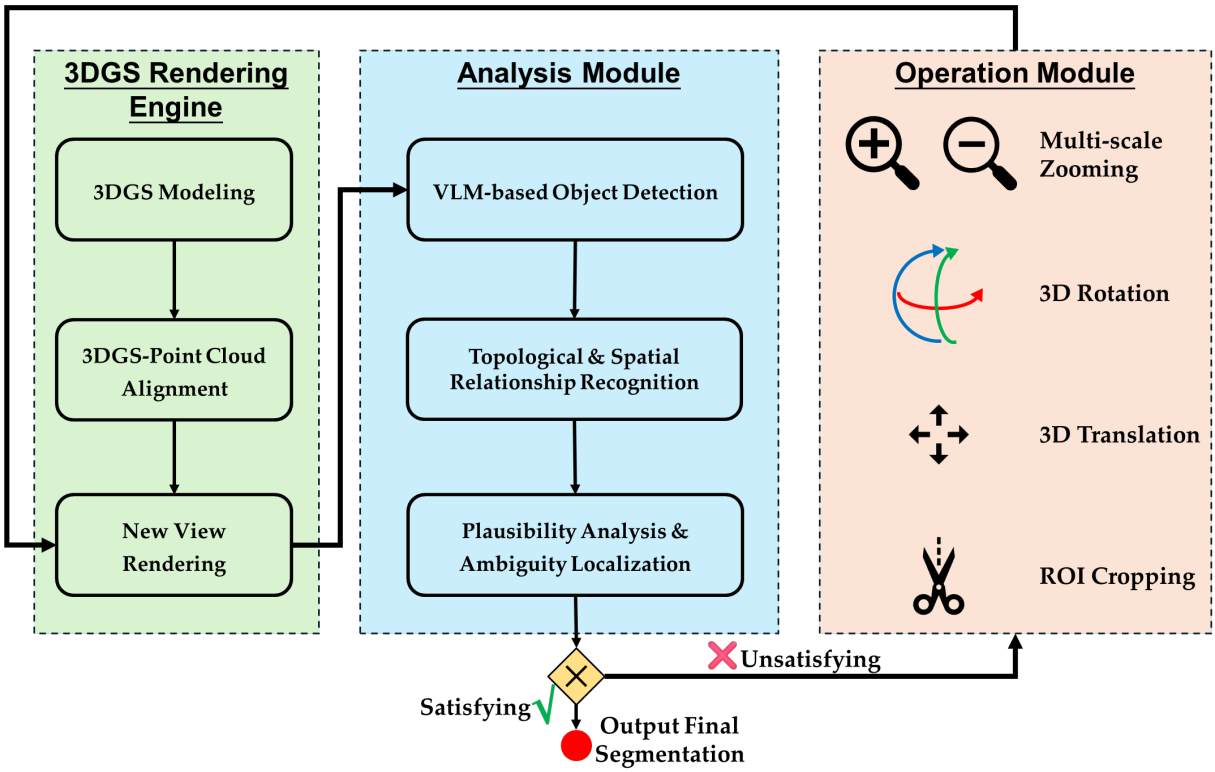


Figure 1: Overall framework of the proposed 3D segmentation agent system

The framework is structured around three interconnected components: the 3DGS Rendering Engine, the Analysis Module, and the Operation Module, forming a closed-loop active perception system:

- 3DGS Rendering Engine (Environment): This module constructs the scene via 3DGS Modeling and Point Cloud Alignment. It generates high-fidelity New View Renderings, providing the visual input necessary for semantic analysis.
- Analysis Module (Cognitive Core): Acting as the decision-maker, this module employs VLMs for object detection and topological relationship recognition. Crucially, it performs plausibility analysis to evaluate results. If the output is ambiguous, the system triggers the Operation Module.
- Operation Module (Active Control): To resolve identified ambiguities, this module executes targeted adjustments such as multi-scale zooming, 3D rotation, translation, or region of interest (ROI) cropping.

This cycle enables the agent to progressively refine its understanding of the scene until a high-confidence, topologically consistent segmentation is achieved. The following sections detail the implementation of these core components.

3DGS Rendering Engine

The 3DGS Rendering Engine acts as the bridge between the physical world and the agent's cognitive processes. Unlike traditional pipelines that rely solely on 2D images or sparse point clouds, our engine constructs a dual-representation environment that synchronizes photorealistic rendering with metric-accurate 3D geometry.

Data Acquisition and 3DGS Modeling: To capture complex as-built conditions, we utilize a handheld mobile scanning system integrating LiDAR and high-resolution RGB cameras. The system leverages Fast-LIVO2 (Zheng et al., 2024) for real-time SLAM, generating a dense point cloud with accurate global pose estimation. Simultaneously, we employ 3DGS to model the scene as a continuous radiance field. Initialized via sparse Structure-from-Motion (SfM) points, the scene is parameterized by volumetric 3D Gaussians, optimized iteratively to minimize photometric loss. This results in a representation that offers real-time, photorealistic rendering from arbitrary viewpoints, essential for the agent's active exploration.

Multi-modal Alignment and 2D-3D Projection: A critical feature of our engine is the rigorous alignment between the photorealistic 3DGS model and the metric-accurate SLAM point cloud. Since raw 3DGS reconstructions lack absolute scale, we register the two representations using a coarse-to-fine approach (anchor point estimation followed by Iterative Closest Point refinement). An example of 3DGS-Point Cloud alignment is shown in Figure 3.

This alignment establishes a deterministic mapping between the 2D rendered view and 3D space. By recording the extrinsic and intrinsic parameters of the 3DGS virtual camera for every rendered frame, we enable a precise pixel-to-point projection mechanism: Once the Analysis Module generates a 2D semantic mask, the engine could use the stored camera parameters to back-project mask pixels onto the aligned 3D point cloud. This mechanism allows the system to extract the exact spatial coordinates of semantic entities, effectively lifting 2D perception into 3D metric space for topological reasoning.

Analysis Module

The Analysis Module serves as the cognitive core of our agent, responsible for transforming raw visual data into a structured semantic understanding of the 3D environment. This process is divided into three progressive stages: VLM-based object detection, topological relationship recognition, plausibility analysis and ambiguous region identification.

VLM-based Object Detection and 3D Projection: To achieve robust segmentation without relying on pre-trained, domain-specific datasets, we adopt a "segment-then-recognize" strategy that integrates foundational 2D models with the reasoning power of Large VLMs.

The workflow begins with the 3DGS engine rendering an initial "initial look" image from a viewpoint as shown in the upper left corner of Figure 2. We utilize the Segment Anything Model (SAM) (Kirillov et al., 2023) to generate a set of high-quality, class-agnostic binary masks, where each mask corresponds to a distinct visual entity (e.g., a pipe segment, a wall patch) as shown in the upper right corner of Figure 2.

These binary masks, overlaid on the original RGB image, are fed into the VLM GPT-5.2. The VLM acts as a semantic parser, analyzing the visual context to assign specific semantic labels (e.g., "Water Pipe," "Concrete Wall," "Valve") to each mask ID as shown in the lower left corner of Figure 2. This step converts purely geometric segments into semantically meaningful objects.

To bridge the gap between 2D image space and 3D volumetric space, we perform a reverse projection. Using the known camera intrinsics and extrinsics from the 3DGS rendering process, each pixel within a 2D mask is back-projected to identify the corresponding 3D point clouds. This effectively "lifts" the 2D semantic labels onto the 3D point cloud, resulting in a semantically annotated 3D scene representation.

Topological and Spatial Relationship Recognition: Mere object detection is insufficient for complex engineering scenes; the agent must understand how objects interact. Once the 3D entities are labeled, the module extracts geometric primitives and infers spatial relationships.

The VLM analyzes these 2D visual features and 3D geometric features to construct a topological scene graph as shown in the lower right corner of Figure 2. This graph encodes connectivity and spatial adjacency, such as:

- Connectivity: Pipe 1 connects to Pipe 2 at Joint J1

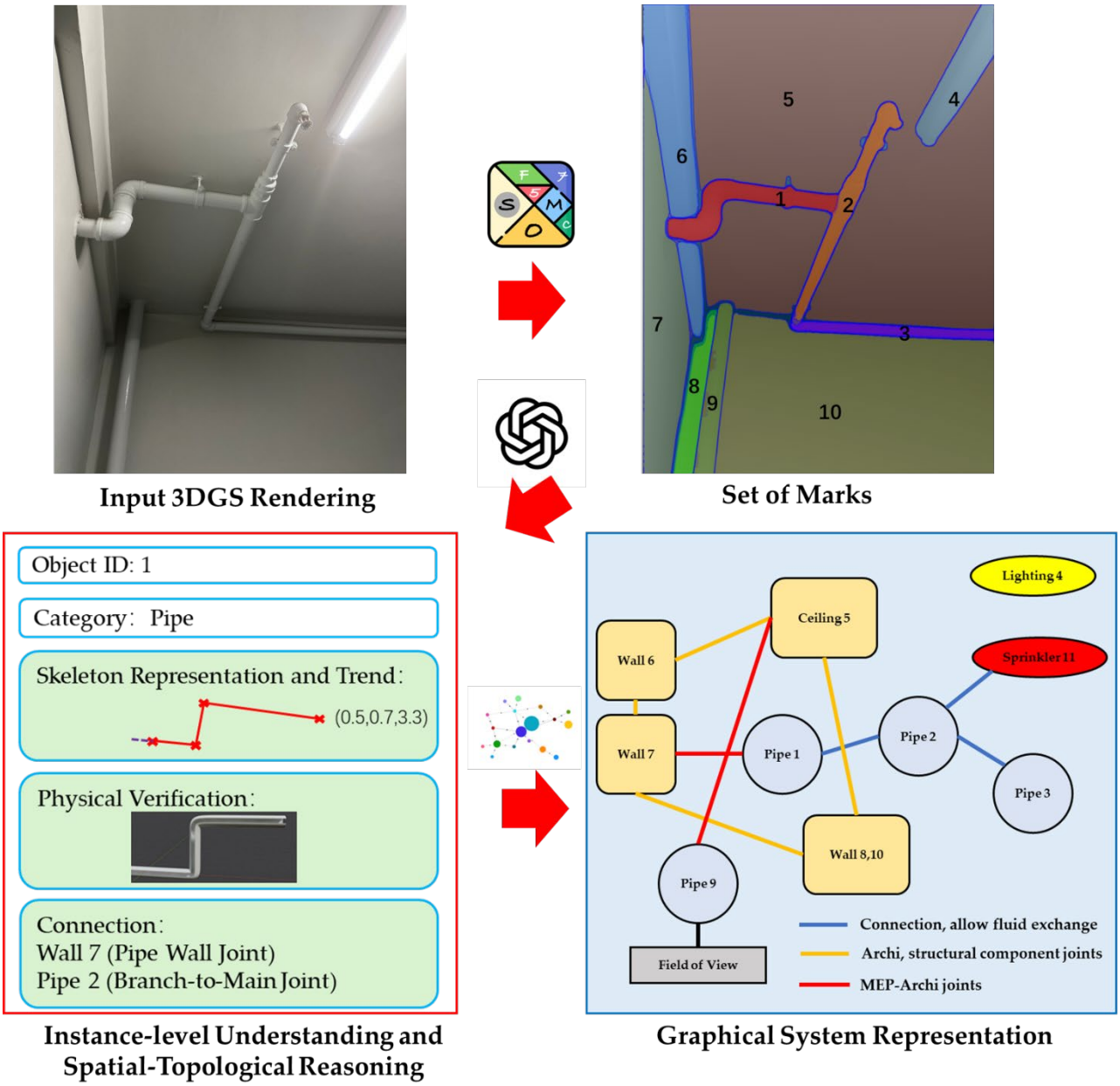


Figure 2: Analysis Module: VLM-based object detection and topological scene graph construction

- Attachment: Pipe 1 is mounted on Wall 7.
- Termination: Pipe 1 ends at coordinate (x, y, z)

Plausibility Analysis and Ambiguity Localization: With a structured scene graph established, the Analysis Module performs a "sanity check" based on physical priors. This is where the agent identifies ambiguities that require active investigation.

As illustrated in upper right corner of Figure 2, the initial segmentation successfully identifies the pipe network ("Pipe 1", "Pipe 2" and "Pipe 3") and the surrounding walls. Following this initial detection, the system evaluates the scene's topology using a hybrid verification mechanism that integrates geometric priors, structured prompting, and VLM-based visual judgment. Since the semantic labels and spatial coordinates (both 2D and 3D) of all detected components are known, the system first applies geometric priors (e.g., spatial proximity) to filter potential relationship candidates.

Applying this mechanism to the network, the topological analysis reveals a logical inconsistency regarding "Pipe 2". The system observes the skeleton of Pipe 2 and detects that it terminates abruptly in mid-air. To investigate this, a targeted logical verification chain is executed. For instance, during the Continuity Check, the system geometrically determines that "Pipe 1" is a viable candidate for connection. A targeted prompt then instructs the VLM to visually verify this specific connection within the 2D rendered image. Thus, while the verification logic is structurally constrained by geometric priors, the final determination relies on the VLM's visual reasoning. For the unattached end of Pipe 2, the VLM yields the following results:

- Continuity Check: Does the pipe connect to another segment? (Result: Negative)

- Boundary Check: Does the pipe extend outside the camera view? (Result: *Negative*)
- Structural Check: Does the pipe terminate into a wall surface? (Result: *Negative*)

According to established engineering priors, a water pipe cannot simply end in free space without a cap or a functional component. Consequently, the Analysis Module flags this region as having "Low Plausibility," hypothesizing that a small terminal device, likely a sprinkler head exists at this location but was missed due to the low resolution of the initial global view. This deduction seamlessly triggers the Operation Module, issuing a precise command: “Zoom in on the endpoint of Pipe 2 to verify the potential missing terminal component,” thereby initiating the active perception loop.

Crucially, this explicit, step-by-step verification chain provides a level of explainability absent in traditional end-to-end models. By structurally constraining the verification logic with geometric priors and outputting human-readable hypotheses (e.g., flagging a specific pipe termination as violating engineering norms), every decision made by the agent, from identifying an anomaly to actively moving the virtual camera, is fully traceable and transparent.

Operation Module

While the Analysis Module identifies what is missing or ambiguous, the Operation Module determines how to look closer. Acting as the agent's navigational controller, this module dynamically manipulates the virtual camera within the 3DGS environment to acquire supplementary views. The process is driven by a feedback loop that continues until the scene representation achieves both geometric high-fidelity and topological consistency. The Operation Module is triggered when the Analysis Module flags a region as "Ambiguous." To ensure rigorous execution, ambiguity is formally defined by two criteria: 1) Quantitative visual uncertainty, occurring when the SAM-predicted mask confidence score falls below 0.7 or the VLM semantic classification probability falls below 0.9; and 2) Qualitative topological inconsistency, occurring when a component fails the logical verification chain (e.g., a pipe exhibiting a "dangling edge" in the scene graph without connecting to a verified boundary or terminal). Upon detecting either condition, the Operation Module triggers a multi-scale zooming and rotation strategy.

- Focusing: For small or indistinct features (such as the potential sprinkler head identified in the last section), the module calculates a new camera pose that minimizes the distance to the target coordinate while maintaining the optical axis centered on the frame. This "zoom-in" operation increases the pixel density of the target, allowing the VLM to discern fine-grained details previously lost in the initial view.
- De-occlusion: If an object is suspected to be occluded, the module executes a 3D rotation around the target's centroid. By generating views from oblique angles, the system bypasses obstacles to reveal the hidden geometry.

Spatial Continuity and Boundary Extension: A common challenge in large-scale engineering scenes is that linear structures (like pipes or ducts) often extend beyond the current field of view (FoV).

- Truncation Detection: If a segmented mask touches the boundary of the rendered image frame, the system recognizes this as an incomplete observation.
- Trajectory Extension: The Operation Module computes the directional vector of the truncated object in 3D space. It then performs a 3D Translation (panning) along this vector, shifting the virtual camera to capture the subsequent segment of the structure. This ensures that long-span objects are tracked continuously across multiple frames rather than being fragmented into disjointed parts.

Active Perception Loop and Termination Criteria: The core mechanism of this module is an iterative Active Perception Loop. The system does not settle for a single pass; it continuously refines its understanding based on specific termination criteria. The loop cycles through Rendering Analysis Operation until the following three conditions are simultaneously met:

- Geometric Completeness: All segmented objects within the target block possess high-confidence mask scores (≥ 0.7), indicating clear boundaries and minimal visual noise.
- Topological Consistency: The scene graph is logically sound. There are strictly zero unexplained "dangling" edges; every segment must either connect to another component, terminate at a verified surface, or end with a recognized terminal device.
- Global Coverage: The union of all recognized entities fully occupies the spatial volume of the target block, ensuring the remaining uninspected void volume falls below a minimal threshold.

The loop iterates until all three conditions are simultaneously met, signifying convergence to a high-fidelity semantic model. To prevent infinite loops in cases of severe physical occlusion or missing scan data, the system enforces a maximum iteration limit (e.g., $N_{max} = 5$) per ambiguous target, after which the unresolved region is explicitly flagged for human review.

Experiment

To evaluate the effectiveness of the proposed pipeline, experiments were conducted on a real-world public toilet captured via mobile laser scanning. The evaluation focused on the object detection accuracy of architectural and functional components, assessed by comparing the predicted labels to manual annotations. The tested scene serves as a representative real-world case intended to validate the feasibility of the proposed pipeline, rather than to provide a comprehensive performance benchmark. Broader validation across multiple sites and building configurations is left for future work.

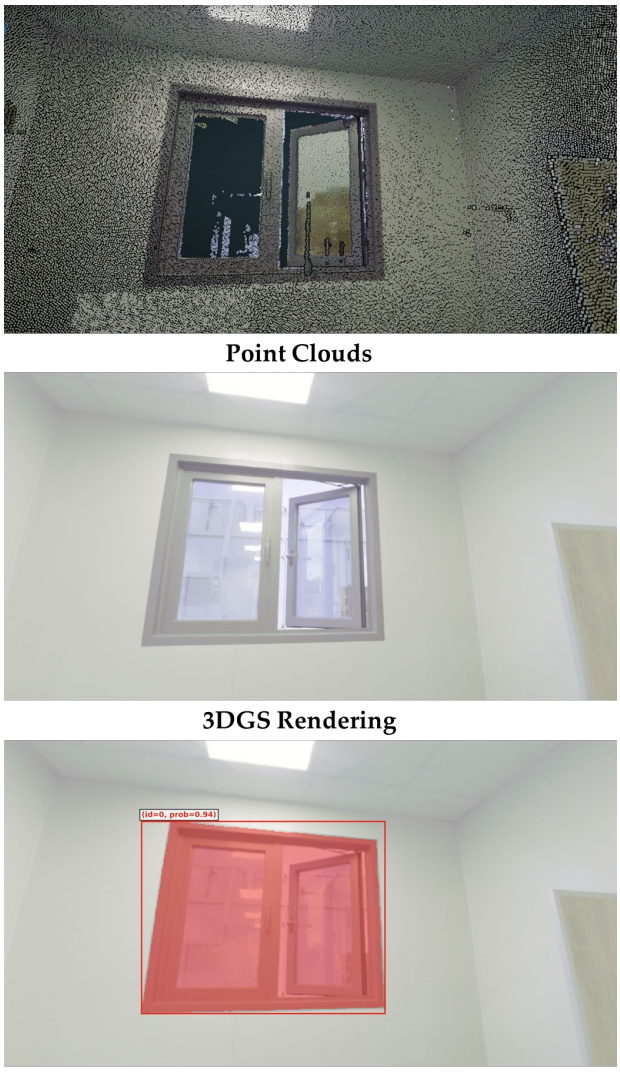


Figure 3: Alignment between 3DGS and point clouds from Rendering Engine and VLM-based object detection results in the experiment

Figure 3 shows a step-by-step breakdown of a specific detection instance (a window). The process relies on the precise alignment between the raw point cloud and the 3DGS model. The top row displays the raw point cloud view. It could be observed the laser scan data is sparse and noisy, lacking sufficient geometric definition for traditional shape-based detectors to reliably identify the object. The middle row shows the corresponding view rendered by the 3DGS. By leveraging the continuous volumetric representation, 3DGS effectively fills in the missing visual information, recovering the window frame structure and the reflective properties of the glass with photorealistic quality. The bottom row demonstrates the final detection result. The VLM performs inference on the rendered image. Because the visual features are clearly restored, the model successfully detects the target "window" with a high confidence score of 0.94, generating an accurate bounding box and segmentation mask. This example confirms that the proposed 3DGS rendering engine allows the system to overcome the sparsity inherent in point clouds, enabling accurate semantic recognition in complex scenarios.

Table 1: Results of public toilet component detection

Component Type	True Positive	False Positive	False Negative
Door	5	0	1
Window	4	0	0
Vent	5	0	0
Lighting	16	0	0
Urinal	2	0	0
Plate	11	0	0
Basin	6	0	0

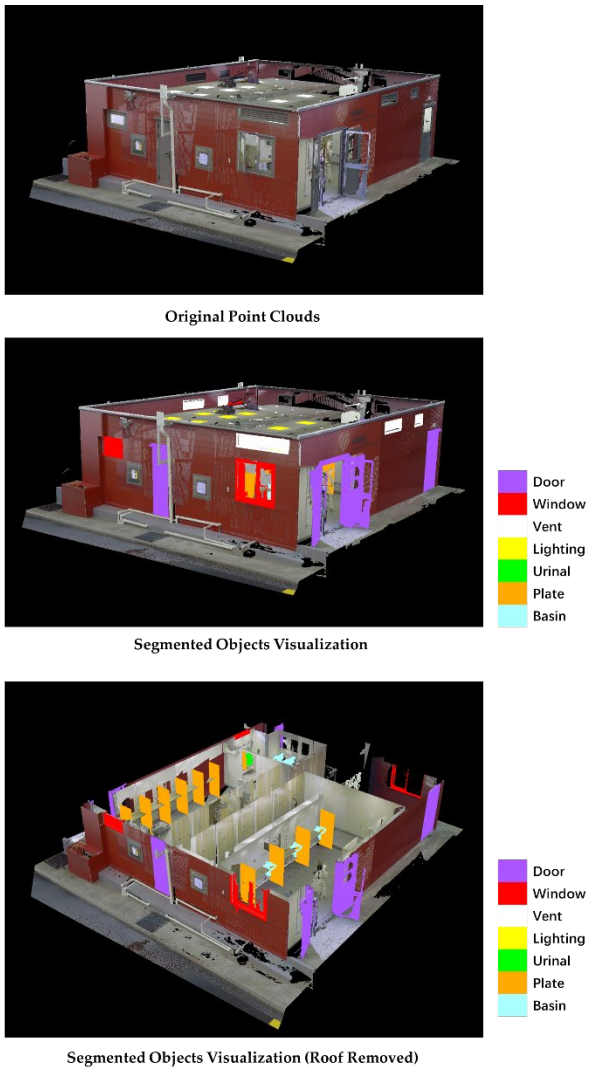


Figure 4: 3D segmentation results performed by the proposed agent

As shown in Table 1, A total of 50 components were annotated and used as ground truth. The results show that the proposed method achieves near-perfect precision

across almost all categories. Notably, repetitive functional elements such as Lights (16 instances), Partition Plates (11 instances), and Basins (6 instances) were detected with 100% precision and recall. The only detection error occurred in the Door category, where one instance was missing (5 out of 6 detected). This specific false negative was attributed to severe data incompleteness; the missing door was in a transition zone that was not fully covered by the scan path, resulting in a low-quality 3DGS for the VLM to form a valid semantic mask, even after the active operation module attempted to refine the view. However, the successful detection of most of the components confirms that the integration of 3DGS modeling and VLM-based segmentation enables high-quality detection with strong generalization across diverse indoor assets. The 3D segmentation results is shown in Figure 4.

Conclusions

This paper presents a novel 3D segmentation agent for as-built analysis by integrating 3DGS and VLMs. Through the synergy of a high-fidelity 3DGS rendering engine and an interactive operation module, the system dynamically adjusts observation perspectives via multi-scale zooming, 3D rotation, and translation. This active perception capability enables the analysis module to effectively perform VLM-based object detection, recognize complex topological and spatial relationships, and conduct plausibility analysis to localize and resolve ambiguities in the scene. Consequently, the system provides a highly interpretable and traceable decision-making process, successfully overcoming the 'black-box' limitations of conventional deep learning approaches.

Experimental results demonstrate the agent's robustness in interpreting complex as-built environments, achieving high accuracy in 3D segmentation and semantic reasoning. The approach successfully bridges the gap between geometric reconstruction and semantic understanding. However, some limitations remain, such as the automation level of the operation module is limited and the testing scenario is not complex enough. The efficiency of the proposed agent in large-scale scenes needs further validation. Furthermore, while our 3DGS-based re-rendering approach theoretically overcomes the limitations of standard 2D-to-3D projection methods, which rely on static raw images and struggle with physical occlusions, by enabling dynamic novel view synthesis, quantitative comparisons are currently absent. As part of our future work to expand this research into a comprehensive journal-length manuscript, we plan to conduct extensive quantitative ablation studies. These studies will systematically compare our active roaming framework against standard 2D-to-3D projection baselines to rigorously demonstrate the performance gains in segmentation accuracy and topological consistency. Overall, the proposed approach lays the foundation for intelligent as-built data analysis, providing a reliable and automated solution for as-built inspection and semantic enrichment.

Acknowledgments

This research was partially supported by the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES), funded by Tamkeen under the NYUAD Research Institute Award CG013.

References

Carion, N., Gustafson, L., Hu, Y. T., Debnath, S., Hu, R., Suris, D., ... & Feichtenhofer, C. (2025). Sam 3: Segment anything with concepts. arXiv preprint arXiv:2511.16719.

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. ACM Trans. Graph., 42(4), 139-1.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4015-4026).

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., ... & Zhang, L. (2024, September). Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In European conference on computer vision (pp. 38-55). Cham: Springer Nature Switzerland.

Lu, T., Yu, M., Xu, L., Xiangli, Y., Wang, L., Lin, D., & Dai, B. (2024). Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20654-20664).

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1), 99-106.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021, July). Learning transferable visual models from natural language supervision. In International conference on machine learning (pp. 8748-8763). PmLR.

Wang, B., Lin, F., Li, M., Liang, Z., Chen, Z., Wang, M., & Cheng, J. C. (2025). Informative As-Built Modeling as a Foundation for Digital Twins Based on Fine-Grained Object Recognition and Object-Aware Scan-vs-BIM for MEP Scenes. Advanced Engineering Informatics, 65, 103382.

Wang, B., Chen, Z., Li, M., Wang, Q., Yin, C., & Cheng, J. C. (2024). Omni-Scan2BIM: A ready-to-use Scan2BIM approach based on vision foundation models for MEP scenes. Automation in Construction, 162, 105384.

Zheng, C., Xu, W., Zou, Z., Hua, T., Yuan, C., He, D., ... & Zhang, F. (2024). Fast-livo2: Fast, direct lidar-inertial-visual odometry. IEEE Transactions on Robotics.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AGENTIC AI – THE TRANSFORMATION OF DECISION-MAKING PROCESSES THROUGH ARTIFICAL INTELLIGENCE

Charlotte-Maria Esch[1], Marijana Srećković[2], and Josef Windsperger[1]
[1]University of Vienna, Vienna, Austria
[2]Vienna University of Technology, Vienna, Austria

Abstract

Agentic AI describes a new generation of AI that automates tasks, pursues goals independently, makes decisions and proactively interacts with its environment. This study - based on a qualitative exploratory research design with expert interviews - aims at understanding how the use of such systems affects internal decision-making logic and organizational structures in companies. Results show that Agentic AI changes decision-making processes. Decisions are being made by systems and humans, with humans mainly taking on supervisory or control functions. We provide a theoretical contribution to existing decision-making and organizational theories with practical recommendations for dealing with such autonomous AI systems.

Introduction

Traditional decision-making processes are reaching their limits due to the availability of large amounts of data, the increasing speed at which market changes occur, and the need to remain competitive in the global marketplace. The dynamic development of AI has the potential to fundamentally change the way companies operate in different industries (IT-P, 2024). In Architecture, Engineering and Construction (AEC), AI adoption is still very limited as reported in the global RICS survey (2025).

AI systems are often used in content generation, data analysis, and forecasting. However, research shows that increasingly powerful autonomous systems, e.g. Agentic AI (Stryker, 2024) are finding its way into practice. This technology goes beyond traditional AI systems in that it is capable of solving complex, multi-stage problems that have no fixed rules or predefined solutions. Agentic AI thus represents an advanced form of AI that can assume a supportive role, as well as act independently. This offers companies the opportunity to optimize decision-making processes and increase efficiency. The Harvard Business Review describes how these systems are transforming human-machine collaboration through their capabilities, for example in virtual nursing or AI-supported travel planning (Purdy, 2024), managing smart homes (Sepkota et al., 2026) or assisting in Urban Planning (Finn & Downie, 2025). Companies like PricewaterhouseCoopers (2024) emphasize the transformative influence of multimodal frameworks in which specialized AI agents work together to optimize efficiency and decision-making processes on an unprecedented scale. At the same time, research by Berkeley SCET highlights the opportunities and risks of these systems, in particular their adaptability to changing conditions and their ability to achieve specific goals autonomously (Ghose, 2024). Simultaneously, key questions arise regarding the area of application and the comprehensive effects, particularly with regard to transparency, accountability, and the delegation of decision-making powers to AI within the internal organization (Ghose, 2024; Purdy, 2024; Stryker, 2024).

In this context, it is becoming increasingly important to examine the organizational and strategic implications of Agentic AI. It is necessary to clarify in which business areas this technology can be used effectively, how it should be introduced, and what capabilities are required for its implementation. This raises the following research question: *"How is Agentic AI integrated into organizational systems to optimize decision-making processes, and what impact does this integration have on the internal structure?"* Hence, this scientific paper examines the effects of Agentic AI on decision-making processes and organizational structures in companies.

First, we examine the status-quo on Agentic AI and decision-making processes. Second, we conduct a qualitative study to explore the experts´ perspectives on AI integration and its implementation effects in organizational settings. This is followed by an analysis and discussion of the interview results, as well as theoretical and practical implications. Finally, we conclude with limitations of the study and related starting points for possible further research in this area.

Theoretical foundation

Decision-making processes in organizations

Research on organizational decision-making has evolved from idealized rational models to more realistic approaches that account for human and structural limitations. The classical rational model, grounded in the *homo economicus* assumption, posits fully informed decision-makers who maximize utility (Suchanek, 2021).

In practice, however, constraints such as incomplete information, time pressure, and bounded cognitive capacity prevail. Simon (1955) addressed these realities with the concept of bounded rationality, according to

which decision-makers seek satisfactory rather than optimal solutions. Complementary models highlight the roles of intuition, incrementalism, and uncertainty (Lindblom, 1959; Dane & Pratt, 2007; Cohen et al., 1972; Kahneman & Tversky, 1979).

Decision-making also varies by planning horizon. Operational decisions address short-term, routine matters, tactical decisions focus on medium-term resource allocation, and strategic decisions shape long-term organizational direction. Complexity and uncertainty increase with the decision level, influencing information needs and responsibility distribution (Chopra & Meindl, 2007; Misni & Lee, 2017).

Organizational structure and culture further shape decision processes. Bottom-up approaches, common in agile settings, emphasize decentralized and participatory decision-making, while top-down hierarchies concentrate authority. Group-based decisions, such as those in interdisciplinary AEC project teams, benefit from diverse perspectives but risk groupthink (Turner & Pratkanis, 1998).

This variety raises a central question: how should decision-making rights (DR) be distributed — centrally or decentrally? In modern organizations, DR are increasingly differentiated. Specific rights apply to codifiable processes, whereas residual rights remain in complex, uncertain situations. Sreckovic and Windsperger (2023) demonstrate, using blockchain-based organizations, that the degree of centralization depends largely on the codifiability of knowledge. The more a process can be formalized, the greater the potential for automation. Although derived from blockchain, this framework provides a valuable analogy for analyzing the impact of Agentic AI on organizational decision-making.

Limits of classic decision-making models

Modern decision-making faces increasing complexity, rapid market changes, and vast data volumes, exposing the limitations of traditional models (Davenport & Harris, 2007). Decisions are frequently made under conditions of incomplete information, time pressure, and cognitive constraints, which foster biases such as confirmation and overconfidence, as well as emotional misjudgments. Structural barriers—including information silos, lack of real-time data, and limited scalability—further restrict human decision-making capacity, particularly in dynamic environments (Berthet, 2021).

These constraints explain the growing importance of technology-based support systems. While business intelligence (BI) enables data-driven insights and real-time decision support (SAP, n.d.-b), Agentic AI goes further by independently analyzing situations, making decisions, adapting to change, and acting proactively (Stryker, 2024).

AI Agents and Agentic AI

Agentic AI is a key technology for automating complex tasks and enabling autonomous decision-making in dynamic, unpredictable environments (Finn & Downie, 2025). Unlike traditional AI systems that act reactively,

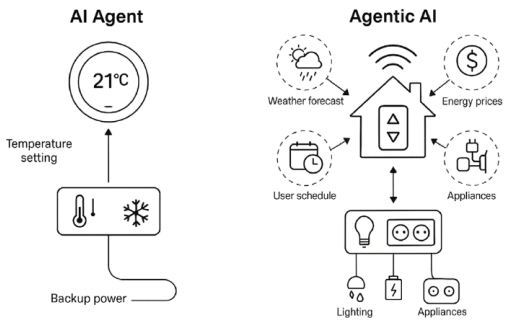


Figure 1: Comparative illustration of AI Agent vs. Agentic AI (Sapkota et al., 2026, p. 8)

agentic systems operate independently: they recognize goals, plan multi-step tasks, and orchestrate external tools or sub-agents to achieve outcomes (Stryker, 2024; Weiler. 2024). By combining the linguistic versatility of large language models with the structured precision of traditional programming, Agentic AI enables end-to-end automation of complex processes (Guan et al., 2024; Stryker, 2024; Finn & Downie, 2025). Rather than generating content on demand, these systems act proactively as goal-oriented collaborators, as illustrated in the smart home example (Figure 1).

The integration of Agentic AI into business processes shifts decision-making from passive suggestions to active, autonomous execution (Ismail, 2025; Fincken, 2024). While generative AI focuses on linguistic content, agentic systems introduce a new automation logic through strong reasoning, modular tool use, and goal pursuit, expanding their applicability across industries (Deslandes, 2025; Zhang, 2025). Key distinctions between AI paradigms are summarized in Table 1 (Sapkota et al., 2026).

Table 1: Key Differentiating Attributes of AI Agents, Agentic AI, and Generative Agents

Aspect	AI Agent	Agentic AI	Generative Agent
Primary Capability	Task execution	Autono-mous goal setting	Content generation
Planning Horizon	Single-step	Multi-step	N/A (content only)
Learning Mechanism	Rule-based or supervised	Reinforce-ment/meta-learning	Large-scale pretraining
Interaction Style	Reactive	Proactive	Creative
Evaluation Focus	Accuracy, latency	Engagement adaptability	Coherence, diversity

AI in AEC

The literature in the AEC domain shows that AI implementation is primarily focused on automated design generation, building performance optimization, system diagnostics enhancement, digital twin operations, and assistance in urban-scale planning, with the broader context of the built environment missing such as the building life-cycle (Lee et al., 2025; Khodabandelu &

Park, 2021; Piccialli et al., 2025). In their review of currently existing AI deployment limitations, Lee et al. (2025) mention: 1) predominant reliance on user-side prompt engineering (lack of a structured knowledge engineering framework); 2) missing comprehensive AI workflows and no knowledge integration with existing systems; 3) deficiencies in standardization and real-world validation (absence of in-depth real-world application resp. evaluations for benchmarking). Regarding collaborative integration and decision-making, Lee et al. (2025) argue that Agentic AI can enhance existing workflows and accelerate the transition toward Human-AI collaboration in the built environment.

The question remains here – how would these scalable integrations of AI work in the organizational setting and decision-making context? So far on the global level, the construction sector shows significant barriers to adopting and scaling AI use, including lack of skilled personnel; integration with existing systems – where companies strive to find flexible and transparent integration options from technology vendors, i.e. interoperable systems for AI integration with their current existing solutions, organizational structures and workflows; data quality and availability; and high implementation costs (RICS, 2025).

Methodology

Study design and sample

Agentic AI represents a novel and under-researched field, particularly regarding its organizational integration and implications. Given the absence of consolidated theoretical frameworks, a qualitative exploratory research design was adopted to generate new insights rather than test existing hypotheses (Corbin & Strauss, 2008).

Semi-structured expert interviews were selected as the data collection method, as they are well-suited for exploring complex, dynamic, and emerging topics and provide access to specialized expert knowledge (Patton, 2002; Saldaña, 2013). The interview guide was divided into five thematic blocks: (1) professional background of the interviewees and their points of contact with Agentic AI, (2) areas of application and potential of Agentic AI, (3) integration into internal company decision-making processes, (4) organizational and procedural implications, and (5) future prospects and integration recommendations. The guide was slightly adapted during the course of the survey in order to specifically address and further develop findings from previous interviews, such as the distinction between Agentic AI and Generative AI (Urquhart et al., 2010).

The interviewees were selected through criterion sampling, focusing exclusively on experts with direct experience in agent-based AI systems. The final sample comprised seven experts from the SAP corporate ecosystem, including SAP in-house specialists, implementation partners, and customer-side users already deploying Agentic AI productively. This diversity of roles — spanning technical, strategic, and innovation perspectives — enabled a multifaceted understanding of both technological and organizational aspects of Agentic AI integration. The focus on the SAP environment was deliberate, given the company's leading position in enterprise software and its substantial strategic investment in Agentic AI (Katanich, 2025; SAP, 2025). The ecosystem provides rich insights that are expected to be transferable to the AEC industry.

Data collection and analysis

The seven semi-structured interviews were conducted in German and English from June to July 2025, based on an interview guide that included open-ended questions.

The interview data was evaluated using the qualitative analysis software ATLAS.ti as part of a combined deductive-inductive process. The coding in ATLAS.ti yielded some insightful and interesting findings regarding the integration and impact of Agentic AI in decision-making processes. Table 2 shows a coding example.
Key topics were derived from the literature, systematically queried, and the corresponding responses were coded deductively. This was supplemented by an inductive approach, in which new categories and patterns, developed from the material, were identified and coded in summary form (Gioia et al., 2012, p. 10; Mayring & Fenzl, 2019, p. 641; Saldaña, 2013, p. 9).
Generated from ATLAS.ti, Table 3 provides an overview of all categories, including the corresponding number of codes, the assigned code group based on the structure of the interview guide, and the number of associated quotes. A total of 74 codes were applied to 325 quotes from seven

Table 2: Exemplary coding

Code group	Category	Code	Quote
Applications & Potential of Agentic AI	Current Applications	Variety of Agents	*"And we already have an agent in use, so to speak—this Classification Agent—which is currently helping in our call centers to structure customer inquiries. It essentially identifies the correct category for the inquiry so that the right agent can then work on the ticket. That was a big step forward. "* (Interviewee No. 2) *"It looks at your code, makes suggestions, writes code for you—whatever you want. Cursor is probably the agent I use the most. It basically executes code for you, writes code for you, tests the code for you, pushes it to GitHub into this external environment where you share the code with others, and also performs the operations you want there."* (Interviewee No. 7)
Effects of Integration	Structural Changes	Regulated Decision-Making & Decision Support	*"[Fully autonomous decision-making] will probably come eventually, but I don't see it happening right now; though I actually think it's quite good."* (No. 1) *"Yes and no—regulated decision-making, that is. It's not simply the system deciding a large part of our business plan, for example. It's about specific and limited decisions for specific environments."*(No. 7)

Table 3: Coding overview

Category	Nr. of Codes	Group / Document Group	Nr. of Quotations
Agentic AI Characteristics	3	Context & Definition	16
Prerequisites for Use	6	Agentic AI Integration in Decision-Making Processes	40
Action Recommendations	5	Outlook & Recommendations	57
Long-term Effects	5	Outlook & Recommendations	46
Codifiability of Processes	7	Agentic AI Integration in Decision-Making Processes	15
Potential Areas of Application	5	Areas of Application and Potentials of Agentic AI	17
Current Areas of Application	8	Areas of Application and Potentials of Agentic AI	54
Critical Areas of Application	5	Areas of Application and Potentials of Agentic AI	15
Structural Changes	7	Effects of Integration	63
Processual Changes	9	Effects of Integration	64
Risks	9	Agentic AI Integration in Decision-Making Processes	25
Opportunities	5	Agentic AI Integration in Decision-Making Processes	12

interviews. Some quotes were assigned to multiple codes, as they address various aspects simultaneously. The coding was carried out in several cycles according to Saldaña (2013). In the first cycle, in-vivo coding (recording of direct quotes) and structured coding (thematic categorization based on the interview guide) were combined to capture both the interviewees' verbatim statements and the key thematic areas. Deductive coding was applied to, among others, "Agentic AI characteristics: Autonomous," "Structural changes: Centralization," or "Codifiability of processes," as these were specifically addressed in the interview guide. Additionally, categories with codes were developed inductively from the empirical material—for example, "Prerequisites for deployment: Cost-effectiveness over hype," "Structural Changes: Agents as Team Members," or "Long-Term Effects: Changed Work Methods & New Competency Requirements." In the second cycle, thematically related codes were bundled, abstracted, and merged into higher-level codes using *pattern coding*. This thematic grouping reduced the complexity of the analysis without losing depth of content. To avoid excessive differentiation through individual codes and the resulting high complexity of the analysis given an exceptionally large number of codes, thematically related codes were consolidated. For example, statements that were originally coded separately, such as "Test locally, then scaled centrally," "Test user groups," "Iterative approach," and "Trial & Error," were grouped into a single subcategory: "Test phase & scaling".

Discussion of results

This study provides answers to the research question of *how Agentic AI is integrated into corporate systems to optimize decision-making processes and what effects this integration has on the internal structure*.

The interview results show that Agentic AI is currently being implemented primarily as intelligent assistance that supports, accelerates, and partially automates decision-making processes—but without completely delegating control. Integration is taking place gradually and selectively, depending on the technological and organizational maturity of the company.

A common thread runs through the interviews: there is agreement on many aspects – in particular on the supporting role of Agentic AI, the importance of technical prerequisites such as data quality and system transparency, and the need for human control in the sense of a human-in-the-loop principle. At the same time, differing perspectives and uncertainties arise on key issues such as the codifiability of processes: some respondents consider it an indispensable prerequisite for the use of agents, while others see modern LLM-based architectures as sufficiently flexible to automate even semi-structured processes. The integration of Agentic AI is already having structural effects. New roles have been identified, such as agent architects and strategic AI managers, who do not perform operational tasks but instead coordinate the technology in a targeted manner. At the same time, Agentic AI is transforming traditional decision-making processes and contributing to the reorganization of workflows, for example, by automating repetitive tasks. Simultaneously, there is an apparent discrepancy between the way it is presented in publications and its actual implementation in practice.

All interviewees recognize the potential of Agentic AI, particularly with regard to efficiency gains, process improvements, and new forms of decision support. The central challenge is how and under what conditions it can be sustainably integrated into existing structures.

Organizational prerequisites and design factors

The successful integration of Agentic AI into corporate decision-making processes is not solely a question of technological feasibility- it strongly depends on organizational and infrastructural prerequisites. Many of those are not only conditions for meaningful
use, but offer concrete recommendations for action for successful and sustainable implementation. They are therefore consolidated in Table 4 as key success factors. To enhance practical utility, we augment each factor with concrete implementation examples drawn from the expert interviews and measurable success metrics.
Current market studies also illustrate that the use of Agentic AI often fails due to a lack of system transparency, fragmented IT landscapes, and inadequate infrastructure. Belcic and Stryker (2025) and Sj (2025)

Table 4 : Key Success Factors with Metrics

Success Factor	Description	Implementation Example (from interviews / practice)	Suggested Success Metrics
Economic efficiency over hype & Strategic selection	Focus on business value rather than technology trends; prioritize use cases	One SAP partner expert described starting with a procurement classification agent that flags non-compliant invoices before full automation (reducing manual review by ~40% in pilot). Similar to Coupa's autonomous procurement intelligence	ROI of agentic initiatives; cost savings vs. implementation effort; prioritized use-case coverage rate
System transparency & process clarity	Clear system landscape SAP workflows first and integrable, data-supported workflows	Interviewees emphasized mapping existing SAP workflows first (e.g., end-to-end procure-to-pay) beforeintroducing agents. DHL-style real-time logistics agents require interoperable date pipelines	Process mapping completeness score; quality index; system integration error rate
Technological & organizational implementation capabilities	Digital infrastructure foundation plus internal resources	Experts recommended pilot projects in logistics/supply chain using SAP Joule as a digital assistant before scaling to multi-agent orchestration	Infrastructure readiness score; training hours per employee; internal AI skill coverage
Governance & agent logic	Defined responsibilities, control structures, and comprehensible agent objectives	SAP customer-side expert described a multi-level approval workflow; agents handel routine procurement decisions up to €50k threshold, with automatic escalation to human "AI managers" above that	Agent goal alignment score; transparency/explainability index; governance policy adherence rate
Human control & monitoring	Human-in-the-loop structures with continuous oversight	All interviewees stresses "supervisory" roles; one example involved a quality-control agent in manufacturing that requests human input on edge cases (e.g., unusual material defects)	Human override rate (<15% target for codifiable processes); escalation frequency trend
Responsibility & Supervision /Change management & test phase	Iterative introduction with pilot phases and stakeholder participation	Experts advocated starting with 3-6 months pilots (e.g., ticket classification or invoice agent) followed by feedback loops and role definition workshops	Adoption rate; employee trust/satisfaction score (via pulse surveys); pilot-to-scale success rate

emphasize the need for robust, scalable architectures, while Noureen (2025) points out that despite a projected potential of over USD 450 billion, many companies do not have the technological maturity and organizational capability to implement it. These findings are consistent with the interviews conducted for this study, in which statements regarding the success factors "system transparency & process clarity" and "technological & organizational feasibility" was mentioned repeatedly. In addition, Shavit et al. (2023) provide concrete governance practices for the responsible use of Agentic AI. They emphasize the importance of transparent agent logic, automated monitoring, and limiting scope for action, thereby underscoring the relevance of the empirically derived factors "governance & agent logic" and "human control & monitoring".

Table 5 shows examples of areas of application and associated agent solutions that have already been discussed theoretically in this paper and have now been empirically confirmed (Capgemini, 2025; Dudley & DelMastro, 2025; Finn & Downie; Weiler, 2024).

Organizational transformation and decision-making rights

Agentic AI is currently understood primarily as a support system that structures, accelerates, and supplements decision-making processes, but not as an autonomous decision-maker. The experts surveyed describe a hybrid control model in which humans increasingly act as "AI managers" (Nicoud, 2025; Raskovich et al., 2024).
Responsibility remains deliberately with humans, especially in sensitive or strategic contexts, in contrast to simple and predominantly operational decisions, which can be automated to a greater extent. This tension between automation and control is consistent with the literature on human-in-the-loop principles and the emphasis placed by Shavit et al. (2023).

Table 5: Exemplary areas of application

Administration	Granola Note-taker, follow-up emails after meetings
Logistics & Supply Chain	Quality control agent, invoicing agent
E-commerce	Shopping agent
Marketing, Sales & Service / CX Area	Ticket Classification Agent, User Story Refinement Agent
Travel Industry	Booking Agent
Support	Code Generation / Cursor Pro / GitHub, Web Application, Security Agent
Wide range of applications	Digital Assistant Joule, Test Automation Agent
AEC	Autonomous calculations, code compliance check; proactive BIM model validation, clash resolution suggestions; risk-monitoring etc.

To move from current supporting roles toward more advanced hybrid decision-making, organizations can follow this practical pathway (derived from interviewee forward-looking statements and aligned with emerging industry roadmaps):

1. ***Pilot Stage (Awareness & Selective Automation)***: Focus on narrow, high-volume, codifiable tasks (e.g., invoice classification, routine scheduling). Emphasize data quality, transparency, and human oversight. The goal: Prove value and build internal trust.
2. ***Scaled Integration Stage (Hybrid Orchestration)***: Expand to multi-step processes involving sub-agents

(e.g., procurement agent coordinating with logistics and compliance agents). Introduce new roles such as agent architects and AI governance leads. Governance frameworks become central.

3. ***Optimized Autonomy Stage (Strategic Human-AI Collaboration)****:* Agents handle end-to-end operational decisions with minimal intervention in well-defined domains, while humans focus on strategic oversight, exception handling, and innovation. Continuous learning loops and ethical guardrails are fully embedded. This pathway helps companies progress gradually while mitigating risks such as loss of control or competence erosion.

The interviews also show that Agentic AI is increasingly seen as a team member rather than just a technical tool. Agents take on preparatory tasks, while final responsibility remains with humans—a division of roles that is also reflected in theory. At the same time, there is tension between decentralized use (in specialist areas) and centralized control through governance and monitoring (Accelirate, 2025; Nicoud, 2025). Efficiency and scalability can be ensured through central standards and control.

The codifiability framework also finds confirmation here. While standardized, codifiable decisions are increasingly being made automatically, residual decision-making rights in unclear, dynamic situations remain with human actors. In practice, therefore, Agentic AI is not used as a substitute, but as a cognitively complementary system (Sreckovic & Windsperger, 2023).

At the same time, organizational shifts are becoming apparent: new roles and skills are emerging, such as prompt designers, agent architects, and agent orchestrators; traditional process responsibilities are being replaced by cross-functional coordination and monitoring tasks. Successful integration of Agentic AI is not only a technical challenge, but an organizational transformation process – with clear responsibilities and new roles (Ismail, 2025; Nicoud, 2025).

Areas of tension and open questions

Despite the recognizable progress and potential associated with the use of Agentic AI, key areas of tension and unresolved issues remain. These relate not only to technological and organizational dimensions, but also to ethical, legal, and social aspects. A key area of tension is the discrepancy between public expectations and corporate reality. While media reports and specialist publications often outline the increasing dominance of autonomous, proactive agents, practical reports and the present study show that companies are currently still struggling with fundamental challenges: a lack of process standards, limited data quality, fragmented IT landscapes, and cultural reservations are hindering implementation (Belcic & Stryker, 2025; Dudley & DelMastro, 2025; Noureen, 2025). Another key area of tension concerns the relationship between autonomy and control. The more powerful Agentic AI becomes, the more pressing the question of legal, ethical, and operational frameworks becomes. Without clear accountability, comprehensible agent goals, and ethical guidelines, there is a risk of losing control—not only technically, but also socially (Weiler, 2024; Zhang, 2025). In addition, questions of efficiency versus competence development are gaining in importance. As agents increasingly take on routine tasks, junior roles in particular are coming under pressure. This is preventing learning and development opportunities, hence poses risk to the long-term retention of competence in organizations.

Furthermore, authors such as Ghose (2024) and Shavit et al. (2023), in line with the experts in this study, call for robust governance approaches not only at the corporate level, but also by social actors. In this context, governance refers to the entirety of institutional regulations, processes, and control mechanisms for the responsible design, control, and monitoring of technology use. This includes, e.g. legal minimum standards for transparency, data protection, the control of algorithmic biases, and the traceability of agentic decisions.

Finally, security risks should not be underestimated. Lack of traceability, potential for abuse, inadequate training, or hallucinations can have legal, technical, commercial, and social consequences. These risks, which were addressed in the interviews, are also confirmed in other articles (Ghose, 2024; Ramel, 2024).

New AEC Implications

Although the empirical findings are rooted in the SAP enterprise ecosystem, they carry relevant implications for the AEC industry, where AI adoption remains limited and fragmented (RICS, 2025; Lee et al., 2025). Construction projects are characterized by high uncertainty, fragmented workflows, data silos across design, procurement, and on-site execution, and stringent requirements for safety, compliance, and lifecycle performance. Agentic AI can address these challenges by enabling proactive, multi-step automation that goes beyond current reactive tools. For instance, agentic systems could autonomously validate BIM models in real time, detect and resolve clashes, monitor project risks through integration with digital twins and sensor data, and optimize resource allocation across supply chains while escalating complex or high-stakes decisions to human supervisors.

Drawing on the success factors, AEC organizations could begin with targeted pilots in codifiable areas such as automated code compliance checks or progress reporting, then progress toward hybrid orchestration involving coordinated agents for design optioneering, predictive scheduling, and sustainability assessments. The human-in-the-loop governance structures and phased transition pathway proposed here are particularly valuable in construction's risk-averse environment, where full autonomy must be balanced with accountability and regulatory compliance. As noted by Lee et al. (2025),

Agentic AI has strong potential to accelerate the shift toward true Human-AI collaboration in the built environment, yet success will depend on overcoming barriers such as poor data interoperability, skills gaps, and fragmented IT landscapes—precisely the organizational prerequisites emphasized in our findings. Future AEC-

specific applications could therefore benefit from adapting the codifiability framework and governance logic to the sector's unique project-based and multi-stakeholder context.

Limitations and further research

The results are based on a limited number of expert interviews with seven participants from the SAP corporate environment. This provided focused insight into the current practice and perception of Agentic AI, but limits the representativeness and generalizability of the results. In addition, all interview partners come from the Austrian/Germany/Switzerland region, which results in geographical homogeneity and limits the application to other cultural or economic contexts. Furthermore, the study is a snapshot of a dynamically developing field of research. Agentic AI is still in the pilot stage in many organizations, which is why numerous assessments are based on hypothetical assumptions or early experience. Rapid technological development can mean that individual findings become obsolete in a short period of time. Conceptual limitations are also apparent at the theoretical level. The analysis is primarily based on decision-making and organizational theory as well as the concept of codifiability. Other theoretical perspectives, such as ethical or legal approaches, were not systematically included. While this deliberate focus enables a targeted analysis of organizational decision-making processes, it also opens up space for future interdisciplinary research.

Despite methodological rigor, a certain degree of subjective interpretation is unavoidable when evaluating qualitative data. The subjectivity of the statements and potential distortions pose further methodological challenges. Future studies should specifically address these limitations. Quantitative studies with larger and more diverse samples could help to validate the patterns identified in this study and test their transferability to other contexts. There is also a need for further research in the form of longitudinal studies that examine changes in decision-making structures, role profiles, and governance systems over a longer period of time.

Conclusion

In the long term, the use of agentic AI will not only bring technological innovations, but also profound changes for organizations. This is not just a matter of more efficient processes, but of structural change. Working methods are changing, new roles are emerging, and the requirements for skills are shifting significantly—especially in the coordination and responsible control of intelligent agents. At the same time, the tension between technological progress, regulatory frameworks, and sustainable action remains. While autonomous agents appear as key decision-makers in some future scenarios, the question remains as to how much responsibility such systems can—and should—actually assume. The increasing pressure to innovate and rapid developments in the field of AI require companies not only to invest in technology, but also to develop a new understanding of leadership, trust, and responsibility, as well es AI Capabilities. These areas of tension mark central topics for future research and the governance of Agentic AI design.

References

Accelirate. (2025) Understanding Agentic AI Architecture: Key Components and Their Impact on AI Efficiency.

Belcic, I., & Stryker, C. (2025) AI agents in 2025: Expectations vs. reality, IBM Think.

Berthet, V. (2021) The Impact of Cognitive Biases on Professionals' Decision-Making: A Review of Four Occupational Areas. Front Psychol, 12, p-pp. 802439.

Capgemini. (2025) Capgemini accelerates enterprise adoption of agentic AI for industries with NVIDIA. Capgemini.

Chopra, S., & Meindl, P. (2007) Supply Chain Management. Strategy, Planning & Operation. In: C. Boersch & R. Elschen (Eds.), Das Summa Summarum des Management: Die 25 wichtigsten Werke für Strategie, Führung und Veränderung p-pp. 265-275. Gabler.

Cohen, M. D., March, J. G. & Olsen, J. P. (1972) A Garbage Can Model of Organizational Choice. Administrative Science Quarterly, 17(1), p-pp. 1-25.

Corbin, J. & Strauss, A. (2014) Basics of qualitative research: Techniques and procedures for developing grounded theory. Sage publications

Dane, E. & Pratt, M. G. (2007) Exploring Intuition and Its Role in Managerial Decision Making. The Academy of Management Review, 32(1), p-pp. 33-54.

Davenport, T., & Harris, J. (2007) Competing on Analytics: The New Science of Winning.

Deslandes, N. (2025) 2025 Informed: the year of Agentic AI. TechInformed - a Leading Media and Marketing Solutions Platform for IT Decision Makers.

Dudley, B., & DelMastro, T. (2025) The Next Frontier: The Rise of Agentic AI.

Fincken, T. (2024) Blick in die Zukunft: Die Evolution von Agentic AI. Computerwoche.

Finn, T. & Downie, A. (2025) Agentic AI vs. generative AI. IBM Think.

Dudley, B., & DelMastro, T. (2025). The Next Frontier: The Rise of Agentic AI.

Fincken, T. (2024). Blick in die Zukunft: Die Evolution von Agentic AI. Computerwoche.

Finn, T. & Downie, A. (2025) Agentic AI vs. generative AI. IBM Think.

Ghose, S. (2024) The Next "Next Big Thing": Agentic AI's Opportunities and Risks. UC Berkeley.

Gioia, D. A., Corley, K. G. & Hamilton, A. L. (2013) Seeking Qualitative Rigor in Inductive Research:Notes

on the Gioia Methodology. Organizational Research Methods, 16(1), p-pp. 15-31.

Guan, Y., Wang, D., Chu, Z., Wang, S., Ni, F., Song, R. & Zhuang, C. (2024) Intelligent agents with llm-based process automation. Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining. pp. 5018-5027

Ismail, N. (2025) Agentic AI: Revolutionary and radical impact on the horizon. HCLTech.

IT-P. (2024) Wie Künstliche Intelligenz (KI) die Entscheidungsfindung verändert: Ein Einblick in verschiedenen Branchen.

Katanich, D. (2025) SAP has just become Europe's most valuable company, dethroning Novo Nordisk. Euronews.

Kahneman, D. & Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), p-pp. 263-291.

Khodabandelu, A. & Park, J. (2021) Agent-based modeling and simulation in construction. Automation in Construction, 131, p-pp. 103882.

Lee, J., Song, J., Koo, J., Choi, S., Hwang, J., Hasnain Saif, S. M., Li, Y., Li, J., Yoo, J., Lee, G., Seok, M. & Yoon, S. (2025) Agentic Built Environments: a review. Energy and Buildings, 346, p-pp. 116159.

Lindblom, C. E. (1959) The Science of "Muddling Through". Public Administration Review, 19(2), p-pp. 79-88.

Mayring, P. (1991). Qualitative Inhaltsanalyse. In: U. Flick, E. v. Kardoff, H. Keupp, L. v. Rosenstiel, & S. Wolff (Hrsg.), Handbuch qualitative Forschung : Grundlagen, Konzepte, Methoden und Anwendungen, pp.. 209-213. München: Beltz - Psychologie Verl. Union.

Misni, F., & Lee, L. S. (2017) A Review on Strategic, Tactical and Operational Decision Planning in Reverse Logistics of Green Supply Chain Network Design. J. of Computer and Communications, 5(8), p-pp. 83-104.

Nicoud, A. (2025) Why agentic AI needs data (and humans). IBM Think.

Noureen, R. (2025) AI Agents Could Unlock $450B by 2028, But Businesses Aren't Ready Yet. Petri.

Patton, M. Q. (2002) Two Decades of Developments in Qualitative Inquiry:A Personal, Experiential Perspective. Qualitative Social Work, 1(3), p-pp. 261-283.

Piccialli, F., Chiaro, D., Sarwar, S., Cerciello, D., Qi, P. & Mele, V. (2025) AgentAI: A comprehensive survey on autonomous agents in distributed AI for industry 4.0. Expert Systems with Applications, 291, p-pp. 128404.

PricewaterhouseCoopers. (2024) Agentic AI – the new frontier in GenAI: An executive playbook.

Purdy, M. (2024) What is Agentic AI, and how will it change work? Harvard Business Review.

Ramel, D. (2024) Gartner Names Agentic AI Top Tech Trend for 2025. THE Journal - Technology Horizons in Education.

Raskovich, K., Briggs, B., Bechtel, M., & Ravinutala, A. (2024) What's next for AI? Deloitte. Insights.

RICS. (2025) Artificial intelligence in construction report https://www.rics.org/news-insights/artificial-intelligence-in-construction-report

Saldaña, J. (2013) The Coding Manual for Qualitative Research. SAGE Publications.

SAP. (2025) How agentic AI is transforming IT: A CIO's guide. SAP.

Sapkota, R., Roumeliotis, K. I. & Karkee, M. (2026) AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges. Information Fusion, 126, p-pp. 103599.

Shavit, Y., O'Keefe, C., Eloundou, T., McMillan, P., Agarwal, S.,….Robinson, D. G. (2023). Practices for Governing Agentic AI Systems. OpenAI.

Simon, H. A. (1955) A Behavioral Model of Rational Choice. The Quarterly Journal of Economics, 69(1), p-pp. 99-118.

Sj, K. (2025). The need for robust AI infrastructure. techradar.

Srećković, M. & Windsperger, J. (2023) Blockchain, (De-)Centralization and Organization Design of Intra-and Interfirm Networks, SSRN

Stryker, C. (2024) Agentic AI: 4 reasons why it's the next big thing in AI research. IBM Think.

Suchanek, A. (2021) Definition: Was ist "Homo oeconomicus"? Gabler Wirtschaftslexikon.

Turner, M. E. & Pratkanis, A. R. (1998) Twenty-Five Years of Groupthink Theory and Research: Lessons from the Evaluation of a Theory. Organizational Behavior and Human Decision Processes, 73(2), p-pp. 105-115.

Urquhart, C., Lehmann, H. & Myers, M. D. (2010) Putting the 'theory' back into grounded theory: guidelines for grounded theory studies in information systems. Information systems journal, 20(4), p-pp. 357-381.

Weiler, C. (2024) Was ist Agentic AI? Blue Delta

Zhang J. (2025) Why Agentic AI is the next frontier of generative AU. Forbes

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

NEURO-SYMBOLIC REINFORCEMENT LEARNING FOR MULTI-CATEGORY CONSTRUCTION RELATIONSHIP WITH REASONING AND SEQUENCE PLANNING

Akarsth Kumar Singh and Shang Hsien Heish
National Taiwan University, Taipei, TAIWAN

Abstract

Construction planning depends on tacit expertise that is rarely captured, causing recurring knowledge loss and weak standardization across projects. This paper presents a neuro-symbolic reinforcement learning framework for multi-category construction relationship reasoning that converts execution logic into explicit, domain-grounded constraints. It integrates synthetic WBS generation, LLM-based constrained reasoning, and Group Relative Policy Optimization (GRPO) with symbolic rewards to generate explainable relationship graphs at inference time. Results demonstrate stable learning, over 75% structural validity and reasoning quality across categories, bounded policy divergence, and reduced truncation, supporting more consistent, reusable, and auditable automated construction planning.

Introduction

Construction planning in practice is rarely a standardized or transferable process. Instead, it relies heavily on planners' tacit knowledge, informal judgment, and accumulated site experience. This knowledge is seldom codified. When a project concludes, the sequencing logic and coordination strategies developed during execution leave with the planner, while organizations retain only static schedules that record outcomes rather than the reasoning behind them (Koo et al., 2007). As a result, planning knowledge is repeatedly lost, forcing each project to relearn previously acquired lessons and contributing to persistent inefficiencies in construction delivery (Wu et al., 2010; Amer et al., 2023).

This limitation is compounded by the fact that construction execution is governed by a wide range of construction execution constraints that extend far beyond what is explicitly represented in formal schedules. These constraints encompass interdependent structural, temporal, resource, cross-discipline, quality, and safety considerations that are typically managed implicitly through informal coordination rather than codified planning artifacts. When such coordination fails, the consequences include rework, schedule delays, cost overruns, and safety risks. The persistent gap between documented plans and the constraints that must be coordinated on site remains a primary source of inefficiency in construction projects (Pfitzner et al., 2024; Soman and Molina-Solana, 2022). Capturing and reasoning over these interacting constraints simultaneously imposes a cognitive burden that exceeds the limits of manual planning processes, particularly for large-scale projects.

In this context, artificial intelligence becomes a necessary enabler rather than a mere productivity enhancement. While other industries have adopted AI-driven planning and coordination, construction planning remains largely manual and experience-dependent (Rashid and Louis, 2019). Recent advances in large language models and reinforcement learning offer an opportunity to formalize and automate the reasoning processes that planners currently apply implicitly. However, for AI to deliver meaningful value in construction, it must move beyond isolated automation of scheduling logic and instead reason explicitly across interacting execution constraints in an explainable and transferable manner. Such relationship modeling is also foundational for downstream applications, including 4D BIM simulation, automated project controls, and delay analysis.

Existing research has addressed individual aspects of this problem. Reinforcement learning has been applied to optimize resource-constrained sequencing (Yao et al., 2024; Kedir et al., 2022), natural language processing models have inferred scheduling dependencies from historical data (Amer et al., 2023), BIM-driven methods have automated schedule generation from geometric models (Faghihi et al., 2014; Gan, 2022), and knowledge graphs have been used to structure construction process information (Pfitzner et al., 2024; Shan et al., 2025). Large language models have further shown promise in compliance checking and construction knowledge extraction (Yang and Zhang, 2024; Zhang et al., 2025). Despite these advances, existing approaches remain fragmented, typically addressing only a limited subset of construction relationships in isolation. None provides a unified framework for reasoning jointly across the full spectrum of execution constraints.

Based on this gap, three core challenges must be addressed to enable practical and transferable AI-driven construction planning.

Challenge 1: Multi-category reasoning. Construction relationships span multiple interacting categories, including scheduling logic, structural and physical constraints, resource requirements, temporal constraints, cross-discipline coordination, quality and inspection requirements, and safety and permitting rules. These

categories are tightly coupled, as physical and regulatory constraints directly shape feasible sequencing and temporal dependencies. Effective reasoning therefore requires these relationships to be considered jointly rather than independently.

Challenge 2: Symbolic grounding. Construction relationships are governed by domain invariants derived from physical laws, material behavior, regulatory codes, and established construction methodologies. Predictions that violate load-path logic, omit mandatory inspections, or ignore curing requirements are not merely inaccurate but potentially unsafe. Learning-based models must therefore be grounded in explicit symbolic representations that enforce these constraints.

Challenge 3: Explainability and standardization. For AI-generated plans to be trusted, adopted, and reused across projects, each predicted relationship must be accompanied by a clear justification traceable to domain knowledge. Explicit reasoning enables standardization and knowledge transfer, directly addressing the chronic loss of planning knowledge observed in current practice.

To address these challenges, this research proposes a neuro-symbolic reinforcement learning framework for multi-category construction relationship reasoning. The framework integrates symbolic domain knowledge with learned reasoning policies through this research pipeline. First, a synthetic Work Breakdown Structure generation process produces structurally valid construction inputs at scale. Second, multi-category relationships are inferred using an ensemble of large language models constrained by formalized domain knowledge, generating reasoning-annotated dependency datasets across multiple execution constraint categories. Third, a 7B-parameter small language model is fine-tuned using Group Relative Policy Optimization with a composite reward function that encodes relationship correctness, directed acyclic graph validity, temporal plausibility, constraint compliance, and category coverage. Finally, the trained model generates explainable, multi-category relationship graphs for unseen inputs at inference time. Symbolic knowledge serves a dual role by grounding reasoning during data generation and providing verifiable reward signals during policy optimization.

The contributions are threefold. First, this research introduces a unified neuro-symbolic formulation for construction relationship reasoning, converting tacit planning knowledge into explicit, verifiable constraints rather than relying on symbolic rules or learning models alone. Second, it develops an LLM-driven data generation pipeline that produces reasoning-annotated construction relationship datasets with explicit justifications. Third, it provides a reproducible performance reference for multi-category dependency prediction using reinforcement learning with symbolic rewards, where LLM reasoning serves as the pre-optimization benchmark and supports explainable, standardized knowledge transfer across projects and organizations.

Related Work

AI for Construction Planning and Scheduling

Research on automated construction planning has progressed through several methodological paradigms. Early work emphasized constraint-based and rule-driven approaches using predefined methods, sequencing templates, and resource packages. For example, Wu et al. (2010) applied simulation and pattern-matching to generate schedules under explicit constraints, while Koo et al. (2007) formalized sequencing rationale through a constraint ontology for representing activity logic. Subsequent BIM-driven approaches leveraged geometric and structural information. Faghihi et al. (2014) employed genetic algorithms to derive structurally feasible sequences, and Gan (2022) introduced a graph-based model integrating IFC semantics and spatial topology for modular design generation. Despite their effectiveness, these methods typically address isolated relationship categories, such as scheduling or structural constraints, and depend on hand-coded rules that limit adaptability across projects.

Recent studies have shifted toward data-driven learning of construction sequencing. Amer et al. (2023) used transformer models to infer dependency constraints and generate lookahead plans, while Rashid and Louis (2019) demonstrated activity pattern recognition using deep learning with synthetic data. However, these approaches lack explicit symbolic grounding, failing to enforce domain invariants such as load transfer, curing constraints, and inspection requirements, thereby limiting reliability in safety-critical planning contexts.

Reinforcement Learning in Construction

Reinforcement learning has emerged as a promising paradigm for optimizing construction scheduling decisions under uncertainty and resource constraints. Yao et al. (2024) proposed a deep reinforcement learning framework with valid action sampling and graph convolutional feature extraction, achieving reduced project durations compared to heuristic baselines. Kedir et al. (2022) combined reinforcement learning with agent-based modeling and graph embeddings to optimize activity sequencing under constrained resources. Soman and Molina-Solana (2022) applied Q-learning with linked-data-based constraint checking to automate lookahead schedule generation, illustrating the potential of reinforcement learning as a decision support tool for short-term planning. More recently, Yang et al. (2026) introduced the Construction Markov Decision Process framework, employing multi-agent proximal policy optimization to simulate crew-level decision-making across multiple construction stages with spatial dynamics.

Despite these advances, reinforcement learning applications in construction have largely focused on optimizing scheduling objectives such as project duration or resource utilization within narrowly scoped problem formulations. Symbolic domain knowledge is typically applied as hard constraints or pre-processing checks rather than being embedded directly into the reward

structure. As a result, these models neither reason across multiple categories of construction relationships nor provide explicit, explainable justifications for predicted dependencies, limiting their usefulness for knowledge transfer and practitioner adoption.

Knowledge Graphs, Ontologies, and Semantic Reasoning in Construction

Knowledge representation through ontologies and knowledge graphs has a long-standing role in construction informatics. Pauwels et al. (2017) enhanced the ifcOWL ontology with optimized geometric representations to improve semantic BIM data exchange. Zhang and El-Gohary (2017) integrated semantic NLP with logical reasoning for automated building code compliance and later extended this work using transformer-based models to align IFC entities and regulatory concepts within a knowledge-graph-enhanced classification framework Zhang and El-Gohary (2023). Ding et al. (2016) proposed an ontology-based framework for construction risk knowledge management, enabling semantic inference of risk interdependence in BIM environments. Collinge et al. (2022) developed a BIM-based safety ontology to support prevention-through-design by structuring tacit and explicit knowledge. More recently, Pfitzner et al. (2024) constructed knowledge graphs from continuous site image data, linking detected objects to spatiotemporal metrics for progress monitoring.

These studies demonstrate the value of structured knowledge representations for specific domains such as safety, compliance, risk management, and progress tracking. However, most knowledge graph–based systems remain static or query-driven and are not integrated with learning-based policy optimization. Moreover, they are typically domain-specific and do not provide a unified representation that jointly captures scheduling, structural, resource, temporal, cross-discipline, quality, and safety relationships required for automated construction planning.

Large Language Models and Neuro-Symbolic Approaches in AEC

Large language models have recently been introduced to address knowledge-intensive tasks in the AEC domain. Yang and Zhang (2024) demonstrated prompt-based automation of building code information transformation, achieving high precision with limited training data. Zhang et al. (2025) proposed a JSON-based agent schema for LLM-driven building energy analysis, enabling reusable agent workflows. Shan et al. (2025) applied LLMs with in-context learning to automate knowledge graph construction for building health resilience using multi-modal data. More broadly, Pan et al. (2023) highlighted the potential of integrating LLMs, which encode parametric knowledge, with knowledge graphs, which represent explicit symbolic knowledge, to overcome limitations inherent in either paradigm alone.

Despite these promising developments, current LLM applications in construction remain predominantly zero-shot or few-shot and rely heavily on prompt engineering rather than systematic learning. Generated outputs are rarely validated against domain knowledge, often lack consistency across relationship categories, and provide no guarantees of physical or regulatory correctness. Furthermore, no prior work combines LLM-based symbolic reasoning for dataset generation with reinforcement learning–based policy optimization for construction dependency prediction. The framework proposed in this study addresses this gap through a neuro-symbolic pipeline in which formalized domain knowledge grounds LLM reasoning during data generation and simultaneously serves as a verifiable constraint signal during reinforcement learning optimization The reasoning-annotated outputs generated by the LLM during data construction serve as the reference against which the RL-trained model is evaluated, implicitly establishing prompted LLM reasoning as the performance reference point that policy optimization is designed to exceed.

Problem Formulation

Given a Work Breakdown Structure (WBS) instance

$$W = \{h, p, l, O, T\}$$

where h denotes the construction head (discipline), p the work package, l the location (building, floor, or zone), O=$\{o1,\ldots,om\}$a set of structural or MEP objects, and T=$\{t1,\ldots,tn\}$ a set of construction tasks without explicit dependencies, the objective is to infer a multi-category construction relationship graph

$$G = \{(r_i, c_i, l_i, e_i)\}_{i=1}^{|G|}$$

Each relationship tuple consists of:

- ri=$(from,to,link_type,lag)$, a directed dependency between two tasks or objects with an associated precedence type (FS, SS, FF, SF) and lag duration;
- $ci \in C$, the relationship category (scheduling, structural, resource, temporal, cross-discipline, quality, or safety);
- li, a natural-language justification traceable to domain knowledge;
- $ei \in [0,1]$, a confidence score.

Figure 1 illustrates a sample multi-category relationship graph for a column construction WBS scope, where nodes represent tasks and directed edges encode relationship tuples across structural, quality, temporal, and scheduling categories with associated confidence scores.

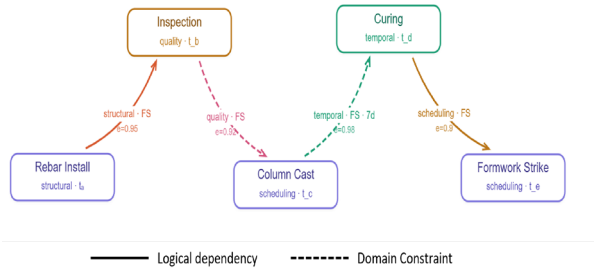


Figure 1: Example multi-category construction relationship graph G inferred from a column construction WBS instance

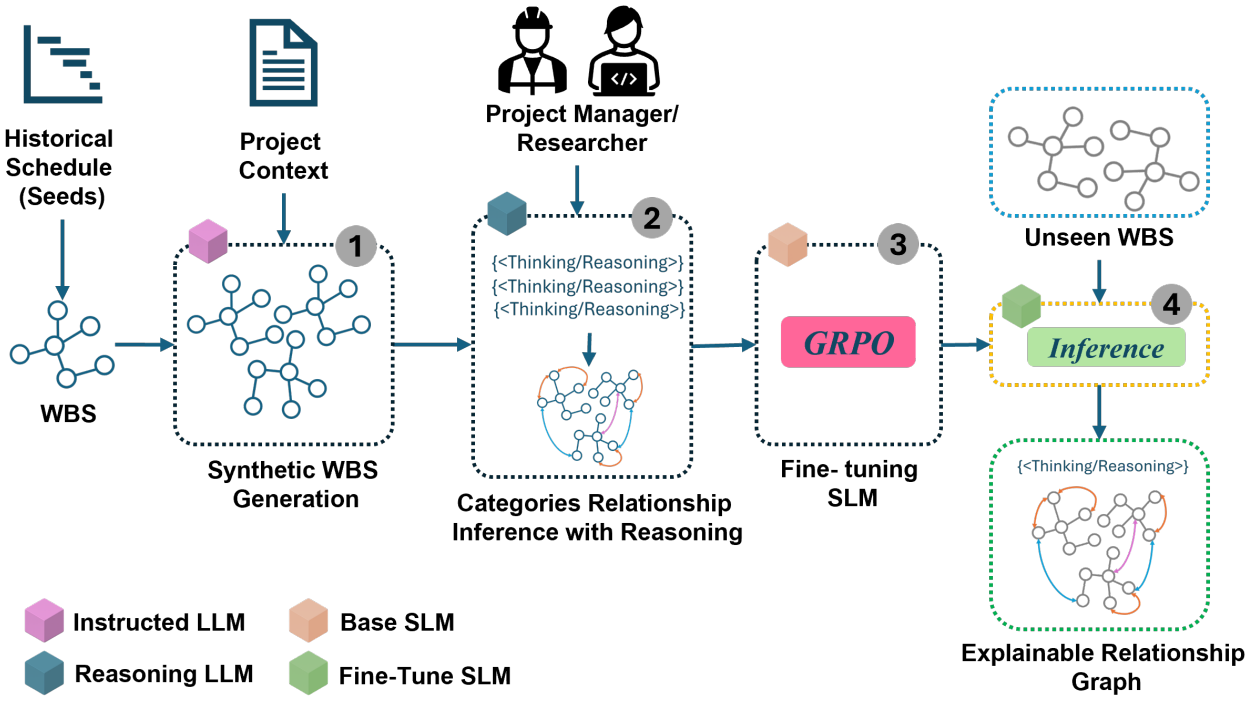


Figure 2: Overview of Framework

The predicted graph G must satisfy three validity conditions:

(1) graph validity, such that scheduling relationships form a directed acyclic graph;

(2) physical feasibility, ensuring compliance with gravity-driven load transfer and construction logic; and

(3) domain compliance, meaning all applicable construction constraints are respected.

Methodology

The proposed methodology integrates synthetic data generation, domain-constrained reasoning, and reinforcement learning to enable multi-category construction relationship inference. As illustrated in Figure 2, the framework begins with the generation of diverse Work Breakdown Structure representations reflecting realistic construction contexts without explicit dependency information. Construction relationships are then inferred category-by-category using large language model reasoning grounded in explicit domain constraints, generating directed edges between task nodes with accompanying explanatory traces. These outputs train a small language model through reinforcement learning, where symbolic constraints are embedded directly into the optimization objective as verifiable reward signals. During inference, the trained model generates explainable construction relationship graphs validated for scheduling acyclicity and physical compliance before being finalized across all execution categories.

Domain Knowledge Representation

Construction knowledge that is typically implicit in planners' experience is formalized as a set of verifiable domain constraints expressed as predicates over the relationship graph G — for example, $\text{supports}(o_i, o_j) \rightarrow \text{precedes}(\text{task}(o_i), \text{task}(o_j))$ for structural load transfer, or $\text{requires_inspection}(t_i) \rightarrow \text{precedes}(\text{inspection}(t_i), t_j)$ for safety compliance. These constraints encode principles derived from structural mechanics, material behavior, construction methodology, trade coordination practice, and regulatory requirements. Each constraint maps to one or more relationship categories and can be automatically evaluated against a predicted graph, enabling both dataset validation and reward computation without manual annotation.

Synthetic WBS Generation

Labeled multi-category construction data is scarce due to proprietary schedules and inconsistent documentation practices. To address this, synthetic WBS instances are generated using template-based augmentation grounded in real project schedules. Recurring WBS patterns are first extracted from reinforced-concrete building projects, capturing typical relationships among construction heads, work packages, locations, objects, and tasks. These patterns are then provided to an instruct-tuned LLM, which generates diverse yet domain-consistent WBS instances by varying contextual attributes such as floor levels, object types, and task labels. Each generated WBS is normalized into a unified schema containing tasks and objects but no explicit dependencies, ensuring that all relationships must be inferred by the model.

Multi-Category Relationship Inference via LLM Reasoning

Each synthetic WBS instance is processed by a reasoning-oriented large language model to infer construction relationships across scheduling, structural, resource, temporal, cross-discipline, quality, and safety categories. Inference is performed in a category-aware manner, reasoning over one category at a time while maintaining

global context consistency. The model receives structured WBS inputs with category-specific domain knowledge to generate structured relationship outputs while verifying assumptions. Category-specific schemas define representations such as precedence and lag (scheduling), load-transfer logic (structural), and clearance or inspection requirements (quality and safety). For each category, the model outputs relationships with justifications and confidence scores. Outputs are validated for logical and structural consistency and aggregated into a unified multi-category graph, retaining reasoning traces to support learning of reasoning quality alongside output correctness.

Reinforcement Learning with Symbolic Rewards

The small language model is fine-tuned using Group Relative Policy Optimization (GRPO) to learn construction relationship reasoning policies. For each WBS input, eight candidate relationship graphs (K=8) are sampled and ranked relative to one another using a symbolic, multi-component reward function (weights reported in the Experiments section), and policy updates reinforce higher-quality reasoning trajectories.

$$R(G) = \alpha_1 R_{correct} + \alpha_2 R_{format} + \alpha_3 R_{structure} + \alpha_4 R_{constraint} + \alpha_5 R_{reason}$$

$R_{correct}$ evaluates semantic agreement with reference relationship graphs generated during data construction.

R_{format} enforces strict adherence to the required output schema.

$R_{structure}$ validates category-specific structural properties, such as acyclic scheduling graphs and physically feasible dependencies.

$R_{constraint}$ verifies consistency with construction domain knowledge relevant to each relationship category.

R_{reason} rewards clear, internally consistent reasoning traces accompanying each prediction.

Each reward component is independently scaled to [0,1] prior to weighted aggregation, and GRPO advantages are computed as each candidate's reward minus the group mean, normalized by group standard deviation across the K=8 sampled graphs. This formulation allows multiple valid solutions while penalizing infeasible or poorly justified outputs. By embedding symbolic validation directly into the reward, the model internalizes domain-consistent reasoning rather than relying on post-hoc filtering. This design directly targets the known limitations of zero-shot LLM reasoning, inconsistency across relationship categories and lack of guaranteed domain compliance, as mentioned in the Related Work section.

Inference and Explainable Outputs

At inference time, a WBS instance is provided to the fine-tuned SLM, which generates a multi-category relationship graph with concise natural-language justifications for each dependency, reflecting the domain principles that motivate relationships and enabling transparency and interpretability. Since each WBS instance is processed independently, multiple instances can be inferred in parallel, enabling full project-scale schedule generation with inference time scaling linearly with the number of instances. Automated post-processing verifies scheduling acyclicity, structural feasibility, and domain compliance, flagging rather than discarding violations to preserve explainability. The final output can be directly consumed by scheduling software, BIM-based 4D simulation tools, or project control systems, and because each dependency includes a traceable justification, the generated plans are inherently transferable across planners and projects, directly addressing the knowledge loss problem identified in the Introduction.

Experiments

Dataset Construction

The experimental dataset was constructed using the proposed Stage 1 and Stage 2 pipelines, with a real-world residential construction schedule as the seed source. The project includes 28 Work Breakdown Structure (WBS) instances comprising 125 activities across substructure, superstructure, and finishing works. To enhance structural diversity and generalization, these WBS patterns were extended into synthetic variants representing multiple structural systems, including masonry, reinforced concrete, composite, precast, steel, and timber structures, specifically to mitigate the residential sequencing and coordination bias introduced by the single-project seed source. The final dataset contains 619 unique construction tasks spanning disciplines, locations, and object types. Using the Stage 2 reasoning pipeline, multi-category relationships were inferred for each task across seven execution categories, yielding 4,333 relationship annotations. The dataset was designed to maintain balanced representation across categories, preventing bias during training.

Reinforcement Learning Fine-Tuning

The policy model was initialized from Qwen3-4B-Base and fine-tuned using Group Relative Policy Optimization (GRPO). For each WBS input, eight candidate relationship graphs were sampled per training step and ranked using a symbolic, multi-component reward function. The reward combined output correctness (40%), format compliance (15%), output structure validity (25%), rule application consistency (10%), and reasoning quality (10%), ensuring balanced supervision across semantic accuracy, structural validity, and explanatory coherence. Training was conducted for five epochs with a batch size of 8 and a learning rate of 1×10^{-4}, using a cosine learning-rate scheduler with warm-up to improve training stability. All experiments were performed on a workstation equipped with an NVIDIA RTX 6000 Ada GPU (48 GB VRAM); the 4B-parameter model is additionally compatible with deployment on GPUs with 24 GB VRAM using standard quantization, broadening accessibility beyond high-end research hardware.

Results and Discussion

Training-Time Analysis (GRPO Optimization Dynamics)

Training results indicate that the proposed neuro-symbolic GRPO framework converges in a stable and

controlled manner. As shown in Figure 3, improvements in total reward are primarily driven by sustained gains in output correctness and rule-consistent reasoning, indicating that the policy progressively learns semantically valid construction relationships rather than merely satisfying structural or formatting constraints. Rule application improves steadily throughout training, confirming increasing internalization of symbolic domain knowledge, while format compliance and structural validity saturate early and function mainly as stabilizing constraints.

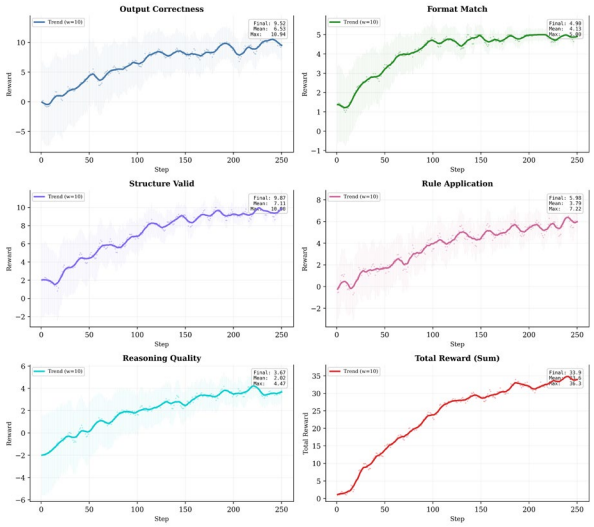


Figure 3: Evolution of GRPO reward components during training

Completion-length behavior further explains the improved optimization dynamics. As shown in Figure 4, increasing the maximum completion length reduces truncation, with the mean length stabilizing at approximately 1,370 tokens per WBS instance and an average inference time of 25–30 seconds per instance on the described hardware, indicating a predictable and bounded cost suitable for practical planning workflows. The truncation ratio decreases to 13.5%, preventing partial reasoning traces from corrupting reward computation and enabling more reliable evaluation of correctness and reasoning-related rewards.

Policy stability during GRPO optimization is supported by the KL divergence profile shown in Figure 5. After an initial adjustment phase, divergence remains bounded around a mean of approximately 0.22, with no upward drift over the training horizon, indicating controlled policy updates and regulated exploration. Overall, these results demonstrate reliable reward optimization,



Figure 4: Completion length and truncation behavior

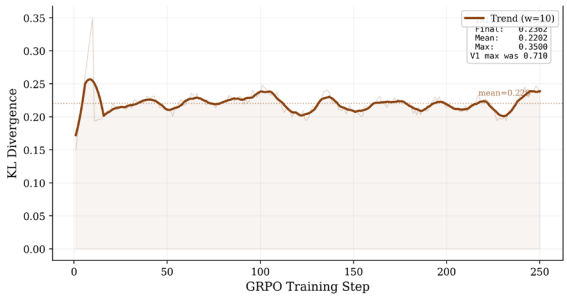


Figure 5: KL divergence during GRPO optimization

mitigation of truncation-induced noise, and stable learning dynamics, providing a robust foundation for inference-time evaluation on unseen construction planning instances.

Inference-Time Performance and Category-wise Analysis

Inference-time evaluation on 100 unseen WBS instances demonstrates that the trained model generalizes consistently across all reward components. As shown in Figure 6, the distributions of output correctness, structural validity, and format compliance are strongly concentrated near their upper bounds, indicating stable and reliable prediction behavior. Reasoning quality also exhibits a tight distribution around high scores, confirming that the model maintains coherent explanatory traces beyond training data. Rule application shows greater variance, reflecting the inherent difficulty of enforcing complex, context-dependent construction constraints during inference.

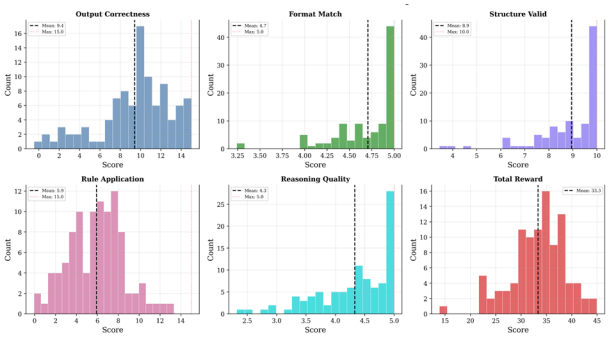


Figure 6: Inference-time reward distribution

A category-wise breakdown of inference rewards further highlights the multi-category robustness of the proposed framework. As illustrated in Figure 7, structural validity, format compliance, and reasoning quality consistently exceed the 75% performance threshold across all construction categories, including scheduling, structural, temporal, cross-discipline, quality, and safety. Output correctness varies more across categories, with lower scores observed for resource and cross-discipline reasoning, which require more implicit coordination knowledge. Nevertheless, safety and quality categories achieve the highest overall performance, reflecting the effectiveness of symbolic constraints in regulating critical execution logic. Overall, these results confirm that the

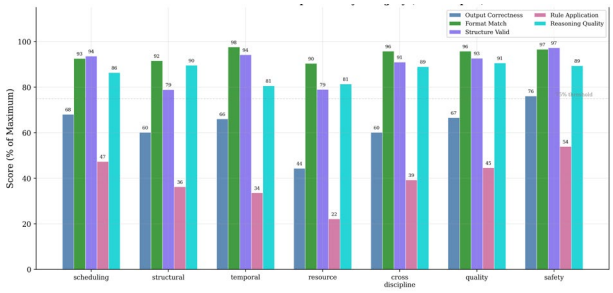


Figure 7: Category-wise inference performance

proposed neuro-symbolic GRPO framework produces explainable, structurally valid, and category-consistent construction relationship graphs at inference time, supporting its suitability for deployment in practical construction planning and downstream project control applications.

Inference and Interpretability

Figure 8 shows an example of inference output for a foundation excavation task. From a dependency-free WBS input, the model generates coherent multi-category relationships, including scheduling precedence, temporal constraints, resource requirements, cross-discipline coordination, and safety considerations. The inferred relationships are context-aware, category-consistent, and aligned with practical construction logic. Non-applicable categories are omitted, indicating controlled reasoning rather than template-based generation. This example highlights the model's ability to produce interpretable and deployment-ready construction relationship graphs at inference time.

Conclusions

This study presented a neuro-symbolic reinforcement learning framework for multi-category construction relationship reasoning, addressing key limitations in planning practice, namely the loss of tacit knowledge and the absence of standardized, explainable representations of execution constraints. Unlike prior approaches focused on isolated scheduling logic or post-hoc validation, the framework embeds symbolic domain knowledge into both data generation and policy optimization, enabling planning-time reasoning across interacting constraints.

Construction relationship inference is formulated as a structured graph prediction problem spanning scheduling, structural, temporal, resource, cross-discipline, quality, and safety categories. To mitigate limited labeled data, a synthetic Work Breakdown Structure generation pipeline is combined with reasoning-guided annotation using large language models constrained by domain knowledge. A small language model is subsequently fine-tuned using Group Relative Policy Optimization with a composite symbolic reward enforcing structural validity, temporal plausibility, constraint compliance, and reasoning quality, shifting learning from output memorization to transferable reasoning policies.

Experimental results indicate stable training dynamics, characterized by controlled policy updates, reduced truncation effects, and consistent reward convergence. Inference on unseen WBS inputs demonstrates robust generalization across all relationship categories, producing structurally valid and explainable dependency graphs. Performance is strong in structural and safety-critical relationships, while higher variability in resource and cross-discipline reasoning reflects the complexity of implicit coordination tasks. Each predicted relationship includes an explicit justification traceable to domain knowledge, supporting transparency and planner trust.

System Prompt

```
Analyze the construction task and identify scheduling relationships.

Output format: Array of scheduling_relationships with:
- from_task, to_task: Task names
- link_type: FS (Finish-to-Start), SS (Start-to-Start), FF (Finish-to-Finish), SF (Start-to-Finish)
- lag_days: Numeric lag in days
- reason: Justification citing relevant domain rules
- is_driving: Boolean indicating if this is on critical path
- total_float: Float days available

Output your analysis in this format:
<start_working_out>
[Your reasoning about the task, mentioning relevant domain rules]
<end_working_out>

<OUTPUT>
[JSON array of scheduling relationships]
</OUTPUT>
```

Inference Input (User Prompt)

```
Country: India, Project Type: Residential Building, Structure Type: RCC Frame, Work Package: Sub Structure (STRUCTURAL), Location: Foundation Area, Object: Foundation Footings, Task: Excavation Work

Analyze this task from a SCHEDULING perspective.
```

Model Reasoning (Intermediate)

```
<start_working_out>
Excavation work is a critical foundation activity and cannot begin until soil conditions and access safety are confirmed.
Per R-007 and R-010, temporary works such as shoring must be completed before excavation starts.
Soil assessment directly governs excavation feasibility and is therefore a driving predecessor.
Once excavation is underway, some verification activities such as subgrade testing can proceed in parallel in completed zones, forming a Start-to-Start relationship.
However, foundation formwork requires a short stabilization period after excavation completion, consistent with R-001 and R-009, introducing a Finish-to-Start dependency with lag.
Only preparatory and formwork-start dependencies are critical; intermediate testing activities retain float.
<end_working_out>
```

Final Output (Scheduling Relationships)

```
<OUTPUT>
[{ "from_task": "Soil Condition Assessment", "to_task": "Excavation Work","link_type": "FS","lag_days": 0, "reason": "Soil bearing capacity and type must be verified prior to excavation to ensure design compliance and safety.", "is_driving": true, "total_float": 0 },
 { "from_task": "Temporary Works Completion", "to_task": "Excavation Work", "link_type": "FS", "lag_days": 0, "reason": "Shoring and temporary stabilization systems must be in place before excavation to prevent collapse and allow safe access.", "is_driving": true, "total_float": 0 },
 { "from_task": "Excavation Work", "to_task": "Subgrade Compaction Testing", "link_type": "SS", "lag_days": 0, "reason": "Compaction and density testing can start in completed zones while excavation continues elsewhere.", "is_driving": false, "total_float": 2 },
 { "from_task": "Excavation Work", "to_task": "Foundation Formwork Start", "link_type": "FS", "lag_days": 1, "reason": "A minimum stabilization and preparation period is required after excavation before formwork erection begins.", "is_driving": true, "total_float": 0 }
]
</OUTPUT>
```

Figure 8: Inference Output Showing Explainable Multi-Category Construction Relationships

The study is based on a residential project dataset augmented with synthetic WBS generation, enabling scalable training but introducing bias toward residential construction logic. Although methodologically general, the learned policy may not transfer directly to other typologies, such as suspension bridges, tunnels, or petrochemical plants, where execution constraints differ significantly and the domain constraint layer must be adapted to project-specific physics, regulations, and construction methods. Future work will extend the constraint framework to additional project types, incorporate typology-specific constraints and uncertainty-aware reasoning, and validate on diverse real-world datasets to assess cross-domain transferability. Overall, this work demonstrates a practical pathway toward

explainable and transferable AI-driven construction planning that respects the domain-specific nature of construction knowledge while enabling reuse across projects and organizations.

Acknowledgments

The authors thank the National Science and Technology Council, Taiwan, for supporting this research through research grant Nos. NSTC 114-2119-M-002-034.

References

Amer, F., Jung, Y. & Golparvar-Fard, M. (2023) Construction schedule augmentation with implicit dependency constraints and automated generation of lookahead plan revisions. Automation in Construction, 152, 104896.

Collinge, W.H., Farghaly, K., Mosleh, M.H., Manu, P., Cheung, C.M. & Osorio-Sandoval, C.A. (2022) BIM-based construction safety risk library. Automation in Construction, 141, 104391.

Ding, L.Y., Zhong, B.T., Wu, S. & Luo, H.B. (2016) Construction risk knowledge management in BIM using ontology and semantic web technology. Safety Science, 87, pp.202–213.

Faghihi, V., Reinschmidt, K.F. & Kang, J.H. (2014) Construction scheduling using genetic algorithm based on building information model. Expert Systems with Applications, 41(16), pp.7565–7578.

Gan, V.J.L. (2022) BIM-based graph data model for automatic generative design of modular buildings. Automation in Construction, 134, 104062.

Kedir, N.S., Somi, S., Fayek, A.R. & Nguyen, P.H.D. (2022) Hybridization of reinforcement learning and agent-based modeling to optimize construction planning and scheduling. Automation in Construction, 142,104498.

Koo, B., Fischer, M. & Kunz, J. (2007) Formalization of construction sequencing rationale and classification mechanism to support rapid generation of sequencing alternatives. Journal of Computing in Civil Engineering, 21(6), pp.423–433.

Pan, J.Z., Razniewski, S., Kalo, J.-C., Singhania, S., Chen, J., Dietze, S., Jabeen, H., Omeliyanenko, J., Zhang, W., Lissandrini, M., Biswas, R., de Melo, G., Bonifati, A., Vakaj, E., Dragoni, M. & Graux, D. (2023) Large language models and knowledge graphs: Opportunities and challenges. arXiv preprint, arXiv:2308.06374.

Pauwels, P., Krijnen, T., Terkaj, W. & Beetz, J. (2017) Enhancing the ifcOWL ontology with an alternative representation for geometric data. Automation in Construction, 80, pp.77–94.

Pfitzner, F., Braun, A. & Borrmann, A. (2024) From data to knowledge: Construction process analysis through continuous image capturing, object detection, and knowledge graph creation. Automation in Construction, 164, 105451.

Rashid, K.M. & Louis, J. (2019) Times-series data augmentation and deep learning for construction equipment activity recognition. Advanced Engineering Informatics, 42, 100944.

Shan, T., Zhang, F., Chan, A.P.C., Zhu, S. & Li, K. (2025) Large language models-empowered automatic knowledge graph development based on multi-modal data for building health resilience. Advanced Engineering Informatics, 68, 103655.

Singh, A.K. & Hsieh, S.H. (2026) Multi-LLM-based augmentation and synthetic data generation of construction schedules and task descriptions with SLM-as-a-judge assessment. Advanced Engineering Informatics, 69, 103825.

Soman, R.K. & Molina-Solana, M. (2022) Automating look-ahead schedule generation for construction using linked-data based constraint checking and reinforcement learning. Automation in Construction, 134, 104069.

Wu, I.C., Borrmann, A., Beißert, U., König, M. & Rank, E. (2010) Bridge construction schedule generation with pattern-based construction methods and constraint-based simulation. Advanced Engineering Informatics, 24(4), pp.379–388.

Yang, B., Liu, B., Han, Y., Meng, X., Wang, Y., Yang, H. & Xiao, J. (2026) Multi-agent reinforcement learning-enhanced decision-making of crew agents in floor construction processes. Advanced Engineering Informatics, 69, 103996.

Yang, F. & Zhang, J. (2024) Prompt-based automation of building code information transformation for compliance checking. Automation in Construction, 168, 105817.

Yao, Y., Tam, V.W.Y., Wang, J., Le, K.N. & Butera, A. (2024) Automated construction scheduling using deep reinforcement learning with valid action sampling. Automation in Construction, 166, 105622.

Zhang, J. & El-Gohary, N.M. (2017) Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. Automation in Construction, 73, pp.45–57.

Zhang, L., Fu, X., Li, Y. & Chen, J. (2025) Large language model-based agent schema and library for automated building energy analysis and modeling. Automation in Construction, 176, 106244.

Zhang, R. & El-Gohary, N. (2023) Transformer-based approach for automated context-aware IFC-regulation semantic information alignment. Automation in Construction, 145, 104540.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

FROM BUILDING PERMIT REVIEW COMMENTS TO COMMENT-BASED PROCESS INDICATORS: AN ESTBERT-BASED CLASSIFICATION WORKFLOW

Pille-Riin Peet[1], Ergo Pikas[2], and Aime Ruus[1]
[1] Tartu College School of Engineering, Tallinn University of Technology, Puiestee 78, 51008, Tartu, Estonia
[2] Building Lifecycle Research Group, Department of Civil Engineering and Architecture, Tallinn University of Technology, Ehitajate tee 5, 19086, Tallinn, Estonia

Abstract

Current building permit evaluation methods use broad metrics like duration and interviews, lacking detail and objectivity for targeted improvements. This paper introduces a workflow to convert unstructured reviewer comments into process indicators using EstBERT, a taxonomy, and supervised multi-label classification. Comments are categorized by issue type and domain, then combined into indicators of rework reasons, inconsistency, and transparency. On a validation set of 365 comments, the classifier scores micro-F1 of 0.470 and macro-F1 of 0.395. Results indicate reviewer comments can reliably support baseline assessment and inform targeted interventions across applications, municipalities, and work types.

Introduction

The building permit process and related procedures are a critical regulatory step in the delivery of construction projects (Hagedorn et al., 2025). However, it is often described as lacking transparency and predictability, with interpretive subjectivity and inconsistencies stemming from the volume and fragmentation of regulations, which lengthen the process and increase construction project costs (Peet et al., 2025). The process is complex, involving numerous stakeholders and disciplinary domains, and its implementation varies across jurisdictions due to differences in administrative practices, values, legal frameworks, and institutional structures (Fauth et al., 2024).

Given this complexity, commonly used metrics, such as total processing time and activity counts (e.g., review rounds and number of remarks), are insufficient because they describe outcomes without providing insight into why rework and iterations occur. On the other hand, effectiveness is frequently assessed through interviews and stakeholder surveys that capture concepts such as transparency, robustness, and citizen-friendliness. These perception-based assessments can be influenced by role-dependent and reporting biases, limiting their suitability as objective baselines. Emerging quantitative approaches, such as stakeholder management analysis metrics (Motahar et al., 2025), provide complementary insights into coordination and communication structures. However, they do not offer scalable, content-level evidence on the issues that cause review iterations in individual cases. In other words, current methods do not adequately support the diagnosis of process problems in a way that is sufficiently scalable and objective to enable benchmarking and monitoring of improvements.

Building permit process activity logs include unstructured comments that capture what reviewers found missing, unclear, inconsistent, or non-compliant, thereby providing direct evidence of why iterations occur. This paper examines whether building permit reviewer comments can be classified into a structured two-layer taxonomy with sufficient reliability to support the development of comment-based process indicators. More specifically, the study evaluates whether an Estonian-language transformer model (EstBERT) (Tanvir et al., 2021) can support multi-label classification of reviewer comments by issue type and disciplinary domain in a way that is useful for baseline assessment and intervention monitoring. The aim is to complement qualitative analyses with scalable indicators that reveal content-level reasons for rework requests across applications, municipalities, and work types. Rather than treating classification as an end in itself, this study uses it as an intermediate step in a broader workflow for transforming unstructured reviewer comments into structured evidence for process diagnosis.

Thus, the main contributions of this study are as follows:

- a workflow for deriving comment-based process indicators from unstructured building permit activity logs;
- a two-layer taxonomy for structuring permit review comments by issue type and disciplinary domain;
- an EstBERT-based multi-label classification pipeline that enables scalable application of this taxonomy to permit review comments.

This paper is structured as follows. The Methodology section describes the data source and the main workflow. In Data Preprocessing, we describe how the data were cleaned, how multi-item remarks were split, and how anonymization was implemented. Comment Classification covers the taxonomy, labeling strategy, and the EstBERT-based multi-label classification method. Deriving Metrics defines comment-based indicators and evaluation measures. The Results section describes the main findings, and the Discussion section discusses the implications, limitations, and possible directions for future research.

Methodology

The building permit process activity logs are from the Estonian Building Registry (EBR). EBR is a database designed to store, provide, and disclose information on planned, under-construction, and existing buildings, as well as related procedures. It is a nationwide environment used for all formal communication between the applicant and the reviewers during the building permit process. For this study, five municipalities representing a range of municipal sizes and annual application volumes (see Table 1) were selected. The building permit activity logs query was submitted to the EBR analyst for all applications in the five selected municipalities that ended in 2025 and started after 01.06.2022 (this date was selected because, in May 2022, EBR underwent a major update affecting all available services and related data).

Table 1: Selected municipalities by population size and annual application volume

	Size of the Municipality (inhabitants as of 2025)	Building permit and building notification applications in a year
A	< 5000	< 500
B	5000 – 10 000	< 500
C	10 000 – 20 000	< 500
D	20 000 – 50 000	500 – 1000
E	> 50 000	> 1000

Types of work planned in the building:

1. Renovation of a building
2. Building construction
3. Building extension
4. Demolition of a building
5. Replacement of part of a building with an equivalent

Table 2: Count of applications by municipality and work type

	A	B	C	D	E
1	0	24	47	111	1029
2	28	105	18	436	632
3	2	13	5	62	176
4	1	2	3	30	93
5	0	0	3	14	52

The dataset was exported as a .csv file containing administrative fields and process names, with column names in Estonian. To improve readability, we use English labels for these fields throughout the paper, while retaining the original column names in an accompanying data dictionary for reproducibility. This paper includes only the subset of fields used for preprocessing, model training, and indicator derivation. Additional fields present in the exports are documented but not used.

The proposed workflow uses reviewer comments as a source of structured evidence on the causes of rework in building permit review. To enable this, comments are classified using a two-layer label system and then aggregated into municipality- and work-type-level comment-based process indicators. Because reviewer comments are unstandardized and vary substantially in style, length, and specificity, the workflow combines rule-based preprocessing with supervised classification. Comment texts were tokenized using the tartuNLP/EstBERT checkpoint, accessed via the Hugging Face model hub and the Transformers library. EstBERT generated (Tanvir et al., 2021) contextual word embeddings were then used to train a multi-label logistic regression classifier using scikit-learn (Pedregosa et al., 2012). All preprocessing, classification, and indicator computation scripts are available on GitHub (https://github.com/pillepidu/EBR-comment-classification.git).

Data preprocessing

Before analysis, the raw building permit activity logs (91,056 rows) were cleaned to ensure data quality and relevance. The dataset was first filtered to include only document types classified as building permit applications ('Ehitusloa taotlus') and building notifications ('Ehitusteatis'), as these were the focus of the analysis. Rows where the comment goal ('eesmark') was marked as internal purpose ('EESMARK_SISEMINE') were removed, as these comments were deemed irrelevant for external classification. Furthermore, any rows with entirely empty values across the comment ('kommentaar'), reason ('pohjus'), and comment goal fields were excluded, as these likely represented system entries without meaningful information. To facilitate chronological analysis and ensure uniform representation, date columns such as application start ('algus'), application end ('lopp'), and date of entry ('tehtud_kp') were converted into pandas `datetime` objects (dd/mm/yyyy), and missing dates are kept as null values. The personal identification code ('isikukood') and registration code ('reg_kood') columns were converted into binary indicators, showing whether each column has a value or not.

The raw comment ('kommentaar') fields contained embedded HTML syntax, which is extraneous for NLP tasks and could introduce noise. To address this, HTML syntax was systematically removed using the `BeautifulSoup` library. Regular expressions were used to identify additional image tags and replace various image tag formats, such as `<img src="data:image/...">` (base64-encoded images) and `<img src="//:0">` (other image references), with the standardized placeholder `[pilt]`. This approach preserved the semantic information that an image was present in the original comment while removing lengthy and irrelevant character sequences.

In the 34,250 rows of activity logs, some of the comments didn't have a value, and many comments were a list of multiple comments, either bulleted (e.g., using '·') or numbered (e.g., '1. Text', '1) Text'). To enable granular analysis and accurate classification, rows with empty comments were filtered out. Comments that were lists were split into individual comment entries. The splitting logic specifically targeted patterns indicative of valid list items: a number followed by a period or parenthesis, and then an uppercase letter (marking the start of a new sentence). This allowed differentiation from legal references, which often contain numbers but lack the specific pattern indicating the start of a sentence. Each identified list item was then treated as a separate row, with all other metadata fields duplicated from the original entry. This resulted in 19,585 unique rows. The classifier was trained and evaluated on these split comment items. Indicator calculations then aggregate predicted labels over all comment items associated with the same application ID (and, where relevant, review round), treating a label as present if it occurs at least once within the group.

To protect personal data and comply with privacy regulations, a three-step anonymization process was applied to the comment fields. In the first two steps, regular expressions were used to anonymize email addresses and phone numbers. Standard email address patterns (e.g., `example@domain.ee`) were identified and replaced with the placeholder `[e-mail]`. Phone numbers that comply with Estonian regulations (*Eesti Numeratsiooniplaan–Riigi Teataja*, 2018) were detected and replaced with `[telefoni number]`. To prevent misclassification of regulation references, numbers containing a slash or preceded by 'nr.', 'nr', or 'nr.' were excluded from this anonymization. The third step was the anonymization of persons' names using the EstNLTK `NerTagger` (Laur et al., 2020), which assigned the 'PER' (person) tag to identify personal names. Once identified, these names were replaced with a sequence of 'X' characters of equal length (e.g., 'Kalle Kask' became 'Xxxxx Xxxx'). This method ensured that personal identifiers were obscured while preserving the textual structure.

Comment classification

To capture the content and context of building permit review comments beyond general procedural status labels, we developed a two-layer label system in which the primary labels refer to the type of problem raised in the comment, and the second layer, the domain tag, describes the disciplinary context. The initial list of categories was developed based on an existing legal analysis of the requirements for construction projects in Estonia (Sorainen & TalTech, 2024). Because reviewer comments often contain multiple requests or refer to closely related problems, the label system was designed as a multi-label scheme in which a single comment may receive multiple primary labels and multiple domain tags. Some category boundaries are necessarily adjacent, particularly between comments about missing information, inconsistency, incorrect content, and requests for clarification. To reduce ambiguity, short decision rules were used during manual labeling. However, some overlap remained and is treated as an inherent characteristic of the comment data rather than as a purely technical modeling issue.

To train a multi-label classifier, we first extracted a subset of 2048 unique preprocessed comments for manual labeling. The subset of comments was randomly selected; to ensure an even distribution across municipalities and application types, we ensured that there were 200 unique comments for each municipality and application type. If there weren't enough of them, then all of them were chosen for the manual training set.

During the manual labeling phase, the label set was refined to better reflect recurring patterns in the reviewers' comments. The manually labeled subset was divided into a training set of 1682 labeled comments and a validation set of 365 labeled comments. The training set was used to train the initial classifier, which was applied to automatically label the remaining comments.

Primary Labels:

P1. Missing Documentation or Information
Required documents, parameters, information, or attachments are missing and must be provided.

P2. Incorrect or Error
There is an incorrect value, detail, or error in the submitted content.

P3. Inconsistent
Two submitted documents contradict one another (e.g., an explanatory document and a site plan).

P4. Non-Compliance with Regulations
Comment states non-compliance with law/regulation/standard (often with legal reference).

P5. Non-Compliance with Design Requirements or Detail plans or General Plans
Conflict with the detailed plan, general plans, or design conditions. Often explicitly stated.

P6. Unclear or Needs Clarification
Information exists but is ambiguous. The reviewer requests clarification or justification.

P7. Format or Presentation (e.g., missing or wrong signatures)

P8. Process or status (e.g., EBR form or procedural issues, formal Statements of approval or refusal without explanation)
Remarks regarding EBR forms or workflow issues rather than technical content.

P9. Requirements for stakeholder approval
Requires consent or approval from another authority or party (e.g., neighbors, utility service providers, Environmental agencies).

P10. No Issues
Explicit statement that there are no remarks.

Domain Labels:

D1. Fire Safety
D2. Heritage
D3. Site
D4. Utilities

D5. Administrative or Procedural
D6. Environmental
D7. Waste Management
D8. Energy efficiency

EstNLTK (Laur et al., 2020) was used to preprocess Estonian-language comment texts before automatic classification. Each comment was sentence-segmented and tokenized into word tokens. For tokenization, each comment text was converted into an EstNLTK `Text` object and tokenized using the `words` layer. For sentence segmentation, the same `Text` object was annotated with the `sentence` layer to provide sentence boundaries for downstream NLP processing. Each comment was then tokenized using EstBERT (Tanvir et al., 2021) based embeddings, a transformer-based language model trained on Estonian text. Contextual embeddings were generated using EstNLTK `BertTagger`, which assigns a BERT embedding to each token in the comment text after annotating the text with the `words` and `sentences` layers. Token-level embeddings were then aggregated into a single fixed-length vector per comment by calculating the arithmetic mean across all token embeddings (mean pooling). This comment-level tokenization supports downstream similarity-based analysis while allowing comments to be associated with multiple issue categories in subsequent steps, even when a single comment contains multiple distinct issues.

Deriving metrics

Following model training, the final classifier was applied to the complete preprocessed dataset to assign primary labels and domain tags to each reviewer comment. These predicted labels were then added to the activity logs and aggregated into comment-based process indicators for rework, inconsistency (e.g., review rounds and recurring issue types), and procedural transparency (e.g., the presence of explicit legal references). The indicators were calculated for the full dataset and also by municipality and intended construction work type. For multi-label outputs, each assigned label was counted once when calculating category frequencies. For application-level metrics (e.g., whether an application contains issues about missing information or documentation), a category is counted as present if it occurred in at least one comment associated with that application.

Model performance was evaluated using the validation set of 365 manually labeled comments (20% of the labeled dataset). Because the task is multi-label and the label distribution is imbalanced, the primary evaluation metrics are micro- and macro-averaged precision, recall, and F1 score, together with per-label precision, recall, and F1 for the primary labels. For completeness, we also report label-cell accuracy as a supplementary metric. For each label ℓ, we compute true positives (TPℓ), false positives (FPℓ), and false negatives (FNℓ) from binary predictions. Macro-averaged precision, recall, and F1 were calculated by averaging the per-label scores, giving equal weight to each label. Micro-averaged precision, recall, and F1 are calculated by first summing TPℓ, FPℓ, and FNℓ across all labels, and then computing precision, recall, and F1-score, with label weights proportional to their frequency. Additionally, we conducted a qualitative error analysis by reviewing misclassified samples to identify the causes of systemic failures (e.g., overlapping category boundaries, multi-problem comments not fully separated during preprocessing, and domain-specific phrasing).

In addition to the predicted labels, each comment was tagged for the presence of an explicit legal reference (has_legal_reference: True/False) using regular expression pattern matching for citation markers and keywords commonly used in Estonian, such as "§", "EhS", "määrus", "standard", "juhend", "seadus", and "nõuded". This flag was used to quantify the proportion of comments that provide explicit legal justification, both overall and by municipality.

Table 3 summarizes the comment-based process indicators used in the quantitative analysis.

Table 3: Overview of metrics derived from EBR activity logs

Theme	Question	Analysis and derived metrics
Inconsistencies and subjectivity	How many applications require multiple review rounds?	Calculates R=max('ringi_nr') per application. Reports the percentage of applications where R > 1 overall, by municipality, and work type.
	What are the most common reasons for revision requests?	Identifies comments, both primary and domain tags. Reports the most common labels overall, by municipality, and work type.
Procedural transparency and a generally worded legal framework	How many comments contain an explicit reference to any legal document, standard, or guide as justification for the comment?	Identifies comments where `has_legal_reference`=True. Reports the percentage of comments with legal references overall, by municipality, and by work type.
Inconsistency and procedural transparency	How often do requirements shift across review rounds within the same application?	Tracks both the primary and domain tags per application that first appear when the review round > 1. Reports the percentage of applications with late-appearing labels, the most common late-appearing labels overall, and how many times they occur.

Results

Classification performance

Table 4 summarizes overall performance using micro- and macro-averaged precision, recall, and F1, reported separately for the primary labels (P1–P10), the domain labels (D1–D8), and all labels combined. Label-cell accuracy is also reported as a supplementary metric across the validation set.

Overall, domain labels (D1–D8) are predicted more reliably than primary labels (P1–P10), with higher label-cell accuracy (0.893 vs 0.778) and macro-F1 (0.437 vs 0.362). The gap between micro and macro scores indicates uneven performance across labels, with labels with less training data lowering the macro-averaged results. Table 5 reports per-label performance for the primary labels. Performance varies substantially across labels, with the highest F1 observed for "Process or status" (P8, F1 = 0.662) and the lowest for "No issues" (P10, F1 = 0.143).

Table 4: Overall performance of the classifier

	Primary labels (P1-P10)	Domain labels (D1-D8)	All labels (primary + domain)
Micro Precision	0.332	0.395	0.351
Micro Recall	0.689	0.761	0.712
Micro F1	0.448	0.520	0.470
Macro Precision	0.274	0.349	0.307
Macro Recall	0.620	0.638	0.628
Macro F1	0.362	0.437	0.395
Accuracy	0.778	0.893	0.829

Table 5: Performance for primary labels (P1-P10). P1 - Missing Documentation or Information, P2 - Incorrect or Error, P3 – Inconsistent, P4 - Non-Compliance with Regulations, P5 - Non-Compliance with Design Requirements or Detail plans or General Plans, P6 - Unclear or Needs Clarification, P7 - Format or Presentation, P8 - Process or status, P9 - Requirements for stakeholder approval, P10 - No Issues

	Precision	Recall	F1
P1	0.437	0.874	0.582
P2	0.126	0.621	0.209
P3	0.358	0.763	0.487
P4	0.125	0.438	0.194
P5	0.136	0.75	0.231
P6	0.363	0.589	0.449
P7	0.194	0.619	0.295
P8	0.62	0.71	0.662
P9	0.247	0.686	0.364
P10	0.133	0.154	0.143

Additionally, we conducted a qualitative error analysis by manually reviewing misclassified validation samples (false positives and false negatives) to identify recurring causes of systematic failures. Two main problems were detected. First, overlapping category boundaries led to confusion between similar labels, especially when a comment could belong to more than one primary category. Second, the splitting preprocessing function failed to capture multiple requests or issues in a single text. The model tended to either assign only a subset of the relevant labels (false negatives) or add an additional label triggered by a specific phrase (false positives). This helps explain the uneven per-label performance and highlights that both label-definition clarity and data-preparation exceptions are key drivers of systematic misclassification.

Frequency and reasons for rework

Table 6 reports the share of applications that required multiple review rounds (R > 1) by municipality (marked A-E as described in Methodology) and intended work type (marked 1-5 as described in Methodology). Across the full dataset (N = 2,804 applications), 1,498 applications (53.4%) required more than one review round, with an average of 2.10 rounds per application.

Table 6: Percentage of applications that require multiple review rounds by municipality and work type (%)

	A	B	C	D	E
1	0.0	41.7	38.3	43.2	50.4
2	25.0	31.4	61.1	46.8	72.2
3	50.0	38.5	20.0	54.8	63.6
4	0.0	0.0	33.3	30.0	62.4
5	0.0	0.0	33.3	85.7	25.0

Table 7 summarizes the most common issue types driving revision requests, broken down by municipality and work type. Across all municipalities and work types, the classifier assigned 51,051 primary labels in total (each label counted once per comment). The three most frequent issue types were "Missing Documentation or Information" (P1, 9,573), "Unclear or Needs Clarification" (P6, 8,519), and "Incorrect or Error" (P2, 7,711). This indicates that rework mainly results from incomplete submission documentation, unclear information, and specific errors in the submitted documents.

Table 7: Most common reasons for revision requests by municipality and work type (Top 3, P - Primary label). P1 - Missing Documentation or Information, P2 - Incorrect or Error, P3 – Inconsistent, P4 - Non-Compliance with Regulations, P5 - Non-Compliance with Design Requirements or Detail plans or General Plans, P6 - Unclear or Needs Clarification, P7 - Format or Presentation, P8 - Process or status, P9 - Requirements for stakeholder approval, P10 - No Issues

	Top3	A	B	C	D	E
1	1	-	P1	P1	P1	P1
	2	-	P2	P6	P2	P2
	3	-	P6	P2	P8	P6
2	1	P6	P1	P1	P8	P1
	2	P8	P8	P2	P1	P6
	3	P1	P9	P6	P6	P2
3	1	P1	P2	P1	P1	P1
	2	P9	P1	P2	P6	P6
	3	P8	P4	P6	P2	P2
4	1	P2	P1	P8	P8	P6
	2	P6	P6	P1	P6	P1
	3	-	P8	P2	P1	P2
5	1	-	-	P1	P1	P8
	2	-	-	P3	P8	P1
	3	-	-	P2	P3	P3

Table 8 summarizes the most common issue domains driving revision requests by municipality and work type. Across all municipalities and work types, the classifier assigned 20,481 domain labels. The three most frequent were “Administrative or Procedural” (D5, 5,790), “Site” (D3, 5,094), and “Waste Management” (D7, 2,622). This shows that revision requests are influenced not only by technical compliance issues but also by procedural requirements and site-specific factors.

Table 8: Most common reasons for revision requests by municipality and work type (Top 3, D - Domain label). D1 - Fire Safety, D2 - Heritage, D3 - Site, D4 - Utilities, D5 - Administrative or Procedural, D6 - Environmental, D7 - Waste Management, D8 - Energy efficiency

	Top3	A	B	C	D	E
1	1	-	D1	D5	D5	D5
	2	-	D5	D2	D1	D3
	3	-	D7	D3	D3	D7
2	1	D5	D5	D5	D5	D3
	2	D3	D3	D3	D3	D5
	3	D1	D1	D7	D7	D7
3	1	D3	D1	D5	D5	D5
	2	D5	D5	D1	D1	D3
	3	-	D3	D7	D2	D7
4	1	-	D7	D5	D5	D5
	2	-	D5	D3	D3	D3
	3	-	-	D2	D7	D7
5	1	-	-	D5	D5	D5
	2	-	-	-	D3	D3
	3	-	-	-	D7	D1

Taken together, Tables 7 and 8 demonstrate that the primary drivers for revision requests are not limited to a single technical discipline. Instead, they combine issues related to submission completeness, information clarity and correctness, administrative procedures, and site-specific requirements. This supports the value of comment-based process indicators, as they offer a more nuanced understanding of the causes of rework than high-level process metrics alone.

Transparency of the rework requests

Table 9 reports the share of comments that explicitly reference a legal document, standard, or guide. Overall, 26.9% of comments contained explicit legal references. Variations in referencing practices across municipalities and work types indicate differences in how reviewers justify remarks and the transparency of the written feedback.

Table 9: Percentage of comments that contain an explicit reference to any legal document, standard, or guide as justification for the comment by municipality and work type (%)

	A	B	C	D	E
1	0.0	28.8	38.0	32.8	32.2
2	38.7	36.2	48.6	34.9	20.4
3	0.0	27.9	16.7	24.0	31.5
4	0.0	50.0	0.0	29.8	28.4
5	0.0	0.0	20.0	39.4	44.5

Table 10 reports the share of applications in which issue types or domain tags first appear after the first review round (“late-appearing labels”). Overall, 4.7% of applications had late-appearing labels, suggesting that most issues are identified early, but that a subset of cases experience shifting requirements across rounds. The five most frequent late-appearing labels were “Process or status” (P8, 436 times), “Inconsistent” (P3, 416), “Administrative or Procedural” (D5, 414), “Requirements for stakeholder approval” (P9, 356), and “Unclear or Needs Clarification” (P6, 350).

Table 10: Percentage of applications with late-appearing labels by municipality and work type (%)

	A	B	C	D	E
1	0.0	12.5	19.6	2.8	4.2
2	25.0	10.6	27.8	5.8	2.5
3	0.0	23.1	0.0	0.0	0.6
4	0.0	0.0	33.3	6.7	2.3
5	0.0	0.0	33.3	14.3	0.0

Discussion

This study shows that unstructured building permit review comments can be structured into categories and aggregated into comment-based process indicators that complement both qualitative assessments (e.g., perceived transparency) and high-level quantitative measures (e.g., total application duration). The results indicate substantial variation in review rounds (Table 6), the types of issues raised in revision requests (Tables 7–8), and the use of explicit legal references in comments (Table 9). Taken together, these indicators provide a more diagnostic view of the permitting processes by showing not only how much rework occurs, but also what kinds of issues drive it and how review practices differ across jurisdictions and work types.

The proposed indicators should be interpreted as signs of process characteristics rather than direct measures of performance quality or administrative performance. For example, a higher number of comments or legal references may reflect a thorough and transparent review rather than inefficiency. Conversely, fewer comments could indicate either efficient early-stage collaboration between the applicant and the reviewer or limited feedback from the reviewer. Additional review rounds may result from project complexity, applicant experience, or local administrative practices. For this reason, comparison across municipalities or work types should be interpreted in context and, where possible, alongside qualitative analysis.

The evaluation conducted in this study is based on activity logs from the Estonian Building Registry (EBR) and on an Estonian-language transformer model, EstBERT. The taxonomy was developed through analysis of Estonian law, and the label definitions reflect local regulations and administrative practices. Applying the workflow in other contexts would therefore require adaptations, including revising the label set, collecting new annotated data, and retraining or adjusting the language model for the target language and document style. Nevertheless, the core workflow - preprocessing comments, multi-label categorization, and aggregation into comment-based process indicators - remains reusable across contexts.

Manual labeling may introduce bias, particularly when performed by a single labeler. Certain categories, such as "Missing Documentation or Information," "Unclear or Needs clarification," "Incorrect or Error," and "Inconsistent," are semantically adjacent and may therefore be interpreted differently across labelers. Short definitions and decision rules were used to reduce ambiguity, but future research should involve multiple labelers and report inter-labeler agreement. This would strengthen the reliability of label assignments, clarify category boundaries, and improve the interpretability of both the classification results and the indicators derived from them.

The activity logs show variation in municipal reviewing practices. Some municipalities provide brief comments, whereas others use long lists or aggregate multiple review requests into a single entry. The splitting method used in this paper extracts list items but fails to separate multiple issues expressed within a single sentence, or may incorrectly split text that resembles a list. These preprocessing problems can affect both classification performance and resulting indicators. Future improvements should therefore focus on more robust sentence-level splitting, improved list detection, and targeted handling of complex comment structures identified through error analysis.

The classification approach used in this study relies on EstBERT-based embeddings with mean pooling and logistic regression, providing a transparent and locally executable baseline for privacy-sensitive administrative text. However, this approach may overlook fine-grained contextual cues within longer or more complex comments and may perform weakly on rare or semantically adjacent labels. This is reflected in the gap between the micro- and macro-averaged results, reinforcing the need for cautious interpretation of category-level outputs. A promising direction for future research is to extend the workflow toward a more explainable AI setting in which the classification-based approach presented in this paper is complemented by an LLM-based method applied in parallel. In such a setup, the structured classifier would support auditable category assignments, while the LLM-based component could help interpret more complex or ambiguous comments, for example, by directing low-confidence predictions to LLM-based classification, decomposing multi-issue comments before labeling, or generating post hoc justifications for human review. Potential improvements, therefore, include threshold tuning, alternative aggregation strategies, stronger handling of long comments, and comparative or parallel evaluation against more recent language-model-based approaches, where privacy and deployment constraints permit.

Conclusions

This study demonstrates that reviewer comments in building permit activity logs can be structured into comment-based process indicators that make recurring causes of rework and review practices visible at scale. In contrast to high-level process measures such as total duration or number of review rounds, these indicators provide more specific evidence about the content of revision requests and the issues that drive iterative review.

Their main value lies in supporting baseline assessment, cross-case comparison, and the identification of dominant issue patterns across municipalities and work types. The three-stage workflow, consisting of preprocessing, classification, and indicator aggregation, is transferable to other jurisdictions by replacing the target-language transformer and the locally derived taxonomy and training set, while the indicator definitions remain unchanged.

The proposed workflow can support future building permit process improvement efforts by helping identify dominant issue categories, such as missing documentation, unclear submissions, or procedural bottlenecks within each jurisdiction. This helps inform improvement efforts, such as clearer reviewer templates, better applicant guidance, or the addition of a pre-consultation phase. Future research should evaluate how such interventions affect specific comment-based indicators over time and should further strengthen the workflow through improved preprocessing, stronger annotation procedures, and comparative or parallel model evaluation.

Acknowledgments

This work has been supported by the Estonian Research Council grant (PSG963).

References

Eesti numeratsiooniplaan–Riigi Teataja. (2018). https://www.riigiteataja.ee/akt/110042024005

Fauth, J., Nørkjær Gade, P., Kaiser, S., Raj, K., Goul Pedersen, J., Olsson, P.-O., Nisbet, N., Mastrolembo Ventura, S., Hirvensalo, A., Granja, J., Urban, H., Rutešić, S., Verstraeten, R., Raitviir, C.-R., Kallinen, A.-R., Schranz, C., Stojanov, T., & Tekavec, J. (2024). Investigating building permit processes across Europe: Characteristics and patterns. Building Research & Information, 0(0), 1–18. https://doi.org/10.1080/09613218.2024.2400467

Hagedorn, P., Fauth, J., Zentgraf, S., Seiß, S., König, M., & Brilakis, I. (2025). OntoBPR: An ontology-based framework for performing building permit reviews using standardized information containers. Advanced Engineering Informatics, 66, 103369. https://doi.org/10.1016/j.aei.2025.103369

Laur, S., Orasmaa, S., Särg, D., & Tammo, P. (2020). EstNLTK 1.6: Remastered Estonian NLP Pipeline. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds), Proceedings of the Twelfth Language Resources and Evaluation Conference (pp. 7152–7160). European Language Resources Association. https://aclanthology.org/2020.lrec-1.884/

Motahar, H., Fauth, J., & Melzner, J. (2025). Evaluating Stakeholder Management Practices in the Building Permit Process. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 17(3), 04525035. https://doi.org/10.1061/JLADAH.LADR-1310

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., & Louppe, G. (2012). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12.

Peet, P.-R., Pikas, E., & Ruus, A. (2025). Lost in Regulations: Digitalization as the Key to Transparent and Efficient Building Permit Evaluation in Estonia. Computing in Construction, 6, 0–0. https://doi.org/10.35490/EC3.2025.432

Sorainen & TalTech. (2024). Ehitusprojektile esitatavate nõuete õigusanalüüs. https://kliimaministeerium.ee/sites/default/files/documents/2024-12/Ehitusprojektile%20esitatavate%20n%C3%B5uete%20%C3%B5igusanal%C3%BC%C3%BCs.pdf

Tanvir, H., Kittask, C., Eiche, S., & Sirts, K. (2021). EstBERT: A Pretrained Language-Specific BERT for Estonian. In S. Dobnik & L. Øvrelid (Eds), Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa) (pp. 11–19). Linköping University Electronic Press, Sweden. https://aclanthology.org/2021.nodalida-main.2/

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

STRUCTURED INNOVATION IN CONSTRUCTION VALUE ENGINEERING THROUGH LLM-ASSISTED TRIZ REASONING

Fuhao Zu[1] and Linghui Xie[2]
[1]Chongqing University, Shapingba District, Chongqing, China
[2]The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

Abstract

This study proposes an integrated framework that enhances idea generation in construction VE by leveraging LLMs to facilitate the application of TRIZ. The framework systematically incorporates project-specific contextual information into an LLM-driven TRIZ reasoning process. It enables context-aware abstraction of technical contradictions, retrieval of relevant inventive principles, and generation of solution ideas. By bridging TRIZ-based structured ideation with VE context, this approach supports more relevant and feasible creativity without requiring deep TRIZ expertise, demonstrating the feasibility of integrating LLM-assisted TRIZ reasoning into VE workflows.

Introduction

Value Engineering (VE) has been widely adopted in the construction industry as a structured method to maximize project functional value while minimizing cost and resource use (Chen et al., 2022). In current VE practice, the idea generation phase has been criticized for relying on unstructured brainstorming, limited reuse of innovation knowledge, and inconsistent solution quality, ultimately constraining the overall effectiveness of VE implementation (Park et al., 2017).

TRIZ (the Theory of Inventive Problem Solving) provides a systematic framework that introduces structured reasoning to enhance the creativity and consistency of generated alternatives during VE execution (Mao et al., 2009). Developed through the systematic analysis of over 200,000 invention patents, TRIZ aims to uncover repeatable patterns in resolving technical problems through innovation (Sheu et al., 2020). Its core tools include the Contradiction Matrix, the 40 Inventive Principles, the Separation Principles, and a series of Standard Solutions, all designed to guide engineers in reformulating specific problems into generalized structures that can be resolved through heuristic reasoning (Kalevi & Ellen R., 2007). This structured logic makes the TRIZ conceptually well aligned with the tasks during VE execution, particularly in developing alternative ideas. TRIZ has, therefore, been introduced into VE frameworks to support more systematic idea generation and reduce reliance on unguided brainstorming (Mao et al., 2009; Wu et al., 2020; Zhang et al., 2009). However, the integration of TRIZ into practical VE applications remains limited. One of the major challenges lies in the high level of abstraction embedded in TRIZ guidance tools, such as the contradiction matrix and inventive principles (Spreafico, 2022). Though conceptually powerful, these tools are typically presented in a condensed and domain-agnostic format, requiring both deep familiarity with TRIZ logic and advanced abstraction skills for effective use (Chen et al., 2024). These requirements are rarely met in typical VE teams. As a result, the cognitive effort required to map project-specific problems to usable TRIZ models has hindered its broader adoption in construction-related VE practice (Jiang et al., 2025).

Recent advancements in large language models (LLMs) offer promising avenues for addressing the cognitive and expertise-related barriers that limit the practical application of TRIZ. These models, exemplified by GPT-4 (OpenAI, 2023), DeepSeek (DeepSeek-AI, 2025), and similar architectures, exhibit strong capabilities in multi-step reasoning, analogical mapping, pattern recognition, and natural language generation. These competencies are well aligned with the analytical demands of TRIZ reasoning. In particular, LLMs can assist in abstracting technical contradictions, retrieving relevant inventive principles, and articulating generalized solution paths (Chen et al., 2024), thereby acting as intelligent assistants in the TRIZ reasoning process (Lee et al., 2024). Previous efforts have explored the integration of LLMs into TRIZ-based workflows. Relevant studies and prototype systems demonstrate that LLMs can automatically map problem statements to TRIZ contradiction models (Guarino et al., 2022), query inventive principles from pre-structured knowledge bases (Lee et al., 2024), and generate exploratory solution ideas (Chen et al., 2024; Jiang et al., 2025). These findings suggest that LLMs can reduce the reliance on expert-level TRIZ knowledge and make structured ideation more accessible.

However, while the feasibility of LLM-driven TRIZ reasoning has been preliminarily demonstrated, its application in the VE context remains underdeveloped. This is particularly significant given that prior attempts to integrate TRIZ into VE have struggled with practical adoption due to the abstract and domain-agnostic nature of TRIZ tools, which require considerable expertise to interpret and apply (Wu et al., 2020). With their capacity for abstraction, analogical reasoning, and guided knowledge retrieval, LLMs offer a viable solution to this

challenge by assisting non-expert users in applying TRIZ. Moreover, current work regarding LLM-assisted TRIZ remains largely confined to generic problem-solving scenarios, with limited consideration of domain-specific constraints (Lee et al., 2024). Existing approaches, such as AutoTRIZ and TRIZ-GPT, primarily focus on mapping problem descriptions to TRIZ models and generating solution ideas, but typically treat the input as standalone problem statements without systematically incorporating structured domain-specific information. Unlike general design settings, VE is inherently embedded in rich project-specific contexts (Gunduz et al., 2022). Before idea generation, VE involves extensive information gathering, including clearly defined objectives, requirements, constraints, and functional decompositions, among other things (Zu & Zhang, 2025). These are central to evaluating the feasibility and value of the proposed ideas. In contrast, this study explicitly integrates such VE-derived contextual information into the TRIZ reasoning process, enabling context-aware abstraction and idea generation aligned with project-specific constraints and value considerations.

Based on this, this study proposes an integrated framework that embeds project-specific contextual information, generated through the structured VE workshop, into LLM-driven TRIZ reasoning, to support the idea generation phase of VE practice. VE inherently clarifies project context through its early-stage activities, such as information gathering and function-cost analysis, providing a foundation for following idea generation (Zu & Zhang, 2025). By systematically incorporating this contextual information into the LLM-driven TRIZ reasoning process, the proposed framework aims to enhance the relevance and feasibility of the generated ideas while preserving the inventive rigor of TRIZ. In doing so, the framework bridges the gap between TRIZ-based systematic ideation and VE-oriented contextual understanding, offering a novel, LLM-powered approach tailored to the creativity phase of VE. This approach enables project teams to access structured innovation support without requiring in-depth TRIZ expertise, providing a context-aware pathway toward effective, value-driven VE practices in construction projects.

Methodology

This section presents the proposed methodology to integrate LLMs into the TRIZ reasoning in the context of construction VE. It aims to enhance the idea generation phase by combining TRIZ innovation strategies with the intelligent reasoning capabilities of LLMs.

Framework overview

As shown in Figure 1, the framework builds upon the traditional VE process, replacing the commonly used brainstorming approach with a structured, LLM-assisted TRIZ reasoning pipeline. In this way, it leverages the systematic innovation logic of TRIZ while addressing the expertise barrier that has historically limited its adoption in construction projects.

The process begins with the information and function analysis phases of VE, where project-specific information is collected, including project context (objectives, stakeholder priorities, and constraints, denoted as C) and functions with value improvement potential (denoted as F). This information serves as the foundation for defining the specific engineering challenge to be improved. Next, LLMs are engaged through prompt engineering techniques, leveraging few-shot examples and structured templates, to abstract the specific challenge into a formal TRIZ contradiction model. This abstraction translates domain-specific challenges into standardized TRIZ contradictions consistent with TRIZ theory. The contradiction model is then used to query the TRIZ matrix, producing a set of relevant inventive principles through structured retrieval. Finally, LLMs recontextualize these general TRIZ strategies into domain-specific, actionable alternatives. In this step, context-aware prompt engineering combined with a two-step process enables the LLM to fuse the retrieved TRIZ principles with detailed project context and functional descriptions. This ensures that the generated alternatives are inventive, feasible, targeted, and aligned with the project requirements, constraints, and stakeholder objectives. Through this, the framework not only addresses the expertise barriers in TRIZ application but also provides a scalable pathway for integrating AI-driven structured reasoning into VE practice.

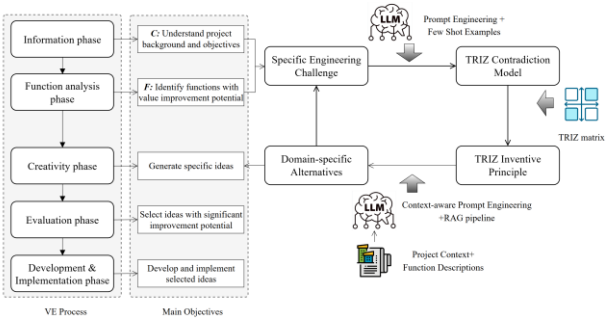


Figure 1: Framework of context-aware TRIZ reasoning supported by LLMs for VE

LLM-Assisted abstraction of TRIZ contradictions

The first critical step in the proposed framework is to abstract engineering challenges into TRIZ contradiction models, bridging contextualized VE information with systematic TRIZ problem-solving tools. The objective is to identify pairs of engineering parameters representing the desired improvement and the accompanying negative trade-off, which define the row and column of the TRIZ contradiction matrix. Traditionally, this abstraction relies heavily on expert judgment, requiring practitioners to interpret specific challenges through the 39 standardized TRIZ parameters. In the proposed framework, this task is performed by LLMs guided through prompt engineering and few-shot learning techniques.

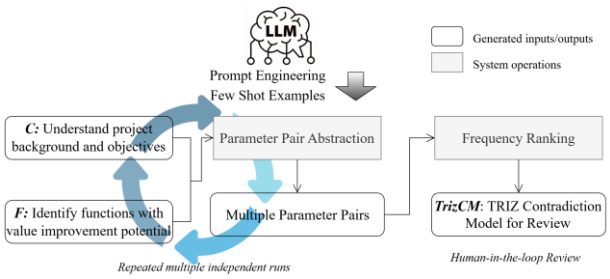


Figure 2: LLM-assisted abstraction and human-in-the-loop review of TRIZ contradictions

As depicted in Figure 2, to enhance LLM performance in TRIZ abstraction, the model is provided with carefully designed prompts and few-shot examples demonstrating how similar challenges have been mapped to TRIZ parameters. This enables the LLM to derive logical mappings from the input context, i.e., descriptions of project objectives and functional improvement targets, to appropriate technical contradictions. For example, the model may infer that enhancing structural durability at the cost of increased material use corresponds to a contradiction between "strength" and "production expense." Recognizing that complex design challenges often admit multiple perspectives, the framework incorporates multiple independent reasoning cycles to explore this diversity. In each cycle, the LLM generates distinct improving-worsening parameter pairs, offering alternative interpretations of the current challenge. The generated pairs are aggregated, and the most frequent are identified as the outputs of this phase. To ensure transparency and alignment with the project context, the framework incorporates Human-in-the-loop Review. As depicted in Figure 2, the ranked outputs are presented to VE teams for validation or adjustment based on domain knowledge, practical feasibility, and stakeholder priorities. In this step, users are expected to assess whether the identified contradiction pairs accurately capture the core design trade-offs and remain relevant to the specific project context. When discrepancies arise, they are resolved through discussion within the VE team, following standard workshop practices, and the selected pairs are refined accordingly.

This approach balances automation and flexibility. While the LLM handles the complexity of abstraction, human oversight preserves contextual appropriateness and critical judgment, with users evaluating the relevance of contradiction pairs, resolving disagreements through structured discussion, and refining selections to improve their contextual suitability for subsequent reasoning. The number of parameter pairs per cycle and the number of cycles are configurable, allowing adaptation to specific VE workflows or project complexity.

TRIZ principle mapping and recontextualization

The core objective of this stage is to transform the TRIZ contradiction model into domain-specific alternative ideas that align with VE objectives. This stage ensures that the systematic innovation strategies of TRIZ are effectively translated into practical design proposals that are inventive and also feasible within the specific context of construction projects. The process consists of two integrated components: (1) retrieval of inventive principles from the TRIZ matrix and (2) context-aware recontextualization of these principles through LLM-supported reasoning, as depicted in Figure 3.

In the retrieval phase, the framework employs a rule-based procedure grounded in TRIZ theory. Since the improving and worsening parameters of the TRIZ contradiction model have been identified, they map deterministically to inventive principles via the TRIZ matrix. This ensures that for any given contradiction, the inventive principles retrieved are traceable to historically validated solutions for similar technical conflicts, preserving methodological rigor. In the generation phase, the framework employs LLMs enhanced with context-aware prompt engineering to adapt the abstract inventive principles to the specific project environment. The model receives three types of input: (i) the retrieved inventive principles, (ii) the project objectives and constraints, and (iii) the functional improvement targets. These inputs are synthesized through carefully designed prompts that introduce the TRIZ strategies while embedding key elements of the project context, as well as functional descriptions, into the reasoning process. While the inventive principles retrieved for each contradiction pair are defined by the TRIZ matrix, the number and diversity of recontextualized ideas generated per principle can be configured according to project complexity or preferences. In this study, for demonstration purposes, five alternative ideas are generated for each inventive principle, providing project teams with a manageable yet diverse set of options for evaluation and selection. This ensures that the generation phase remains flexible and adaptable to different VE contexts.

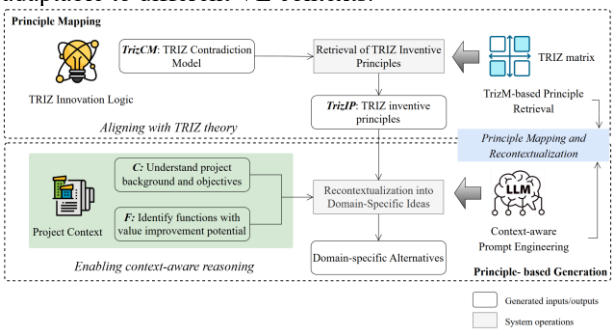


Figure 3: Two-step TRIZ reasoning and contextualized alternative generation

This framework integrates TRIZ-based retrieval with LLM-assisted recontextualization to systematically support the creativity phase of VE. By building directly on the outputs of the information and function analysis phases, it enables automated idea generation that aligns seamlessly with the VE workflow and reduces reliance on traditional brainstorming methods. At the same time, the framework ensures that the generated ideas are rooted in the structured logic of TRIZ and adapted to the specific goals, constraints, and functional requirements of the project, thereby enhancing their relevance, feasibility, and value for subsequent evaluation and decision-making. This offers a structured, scalable alternative to the

conventional creativity phase in the VE workshop, supporting more consistent, rigorous idea generation.

Demonstration

This part presents the development of the proposed prototype and demonstrates its functional workflow through a representative case study.

Prototype development

To demonstrate the feasibility of the proposed LLM-assisted TRIZ reasoning framework for construction VE, a prototype system is developed. The platform integrates LLMs, TRIZ theory, and human-in-the-loop validation to support automated, context-aware idea generation that aligns with the standard VE workflow.

As depicted in Figure 4, the system comprises four main components: (1) a user interface module implemented using Python Tkinter for structured input collection, human-in-the-loop review, and result visualization; (2) an LLM-assisted abstraction module for mapping project context and functional requirements to TRIZ parameter pairs; (3) a TRIZ knowledge base module that stores the 39 standard TRIZ parameters, the contradiction matrix, and inventive principles in structured JSON format; and (4) a context-aware idea generation module that implements the two-step process, combining principle retrieval with LLM-based recontextualization.

In the parameter pair abstraction phase, the system invokes ChatGPT-4o via Azure OpenAI to identify improving and worsening parameters from the user-provided project context and functional improvement targets. The prompt design for this phase follows a template-based structure. 39 TRIZ parameters and their definitions are embedded statically within the prompt, together with the designed few-shot examples relevant to the construction domain (e.g., balancing size v.s. performance, or durability v.s. maintainability). This structured design provides information for understanding and replicating the core prompting logic, although full prompt templates are not included due to space constraints. The project-specific inputs are inserted into this static template at runtime. This design ensures that the LLM reasons consistently within a defined framework while accounting for variations in project descriptions. To enable the LLM to generate multiple reasonable interpretations of the problem for frequency aggregation, the temperature is set to a moderate level (0.5-0.7). The system performs multiple independent abstraction cycles, aggregates the resulting parameter pairs, and ranks them by frequency for human validation, thereby enhancing robustness against individual model variances. In this study, five reasoning cycles were conducted, each generating three parameter pairs. The frequency of each contradiction pair was calculated based on its occurrence across the five cycles, and the two highest-ranked pairs were provided for subsequent analysis. When ties occurred, final selection was determined through human review based on project context.

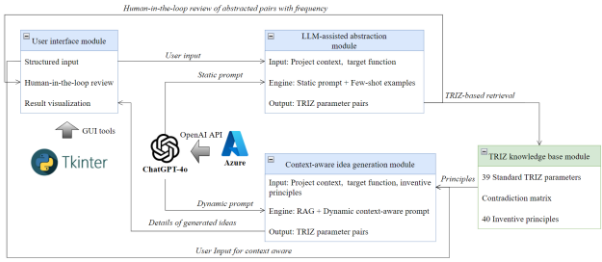


Figure 4: Architecture of the developed prototype

From a deployment perspective, multiple LLM calls introduce additional cost and latency, which are controlled in this prototype by limiting the number of reasoning cycles to 5 and outputs per cycle to 3. Each cycle completes within a few seconds, resulting in an overall response time suitable for interactive VE use. Potential failure modes include incorrect contradiction mapping and inconsistent or unsupported reasoning. To mitigate these risks, the system constrains the LLM using a fixed prompt structure embedding the 39 TRIZ parameters, incorporates domain-specific few-shot examples to guide mapping consistency, and applies frequency-based aggregation to filter unstable outputs across cycles. These measures complement the human-in-the-loop validation by reducing the likelihood of erroneous candidates being presented to users.

In the TRIZ strategy retrieval and idea generation phase, the recognized parameter pairs are used to query the stored contradiction matrix (in JSON format) to retrieve the associated inventive principles. The LLM is then engaged to generate domain-specific alternative ideas based on the retrieved principles and project context. The prompts for this phase are dynamically constructed in real time, designed to synthesize the inventive principles, project objectives, functional targets, and key constraints. The LLM is explicitly instructed in these prompts to reason how each inventive principle can be adapted to the specific project conditions and to articulate the rationale behind each generated idea, enabling context-aware reasoning that tailors idea generation to the unique conditions of the project. To encourage creative and diverse idea generation, the temperature is increased to a high level (0.8-1.0) at this stage. This configuration supports the generation of innovative alternatives that remain traceable to TRIZ theory and adapted to the specific realities of the VE project.

This architecture demonstrates the technical feasibility of embedding LLM reasoning and TRIZ logic into the VE process to systematically enhance the creativity phase, providing a system foundation for functional demonstration and evaluation in subsequent sections.

Functional demonstration

To verify the functionality and practical feasibility of the developed prototype, a demonstration was conducted using a representative case study aligned with typical VE applications in construction projects. The case focuses on enhancing passenger throughput in an urban transportation hub during peak periods, while minimizing energy consumption and space utilization.

Abstraction of TRIZ contradictions

The first step of the prototype corresponds to the information and function analysis phases of VE. As shown in Figure 5, the user provides structured inputs describing the project context and functional improvement target. These inputs are derived from typical VE artifacts, where the project context summarizes key constraints, stakeholder requirements, and system characteristics, while the functional improvement target reflects the prioritized functions identified through function analysis and function-cost evaluation. The project context specifies the development of a large-scale urban transportation hub designed to accommodate high passenger volumes while ensuring safety, comfort, and compliance with environmental and cost constraints. The functional improvement target focuses on increasing passenger throughput during peak periods with minimal energy consumption and space usage.

Once the input is submitted, the system engages the LLM-assisted abstraction module to identify potential contradiction pairs. ChatGPT-4o (via Azure OpenAI) is invoked multiple times, with each reasoning cycle leveraging a static prompt that embeds the 39 TRIZ parameters, their definitions, and domain-relevant few-shot examples. The project-specific input is dynamically inserted into this template at runtime. To promote diverse interpretations of the problem, the temperature is set to a moderate value of 0.5 in the demonstrated case. The system executes five independent abstraction cycles, with each cycle generating three candidate contradiction pairs. The resulting pairs from all cycles are aggregated, and their frequency distribution is computed.

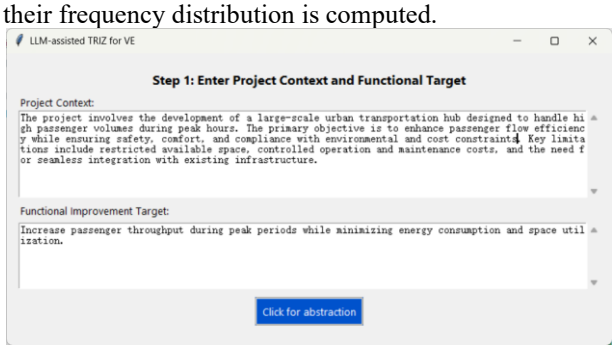


Figure 5: Interface for specifying project context and functional improvement target

As shown in Figure 6, the system visualizes the top-ranked pairs in a bar chart along with textual explanations of their origins. At this point, a human-in-the-loop review is conducted. The user examines the generated contradictions and their contextual explanations, and selects two pairs for further analysis. This interactive step ensures alignment of the automated abstraction results with domain knowledge and project-specific priorities, enhancing the transparency and adaptability of the process.

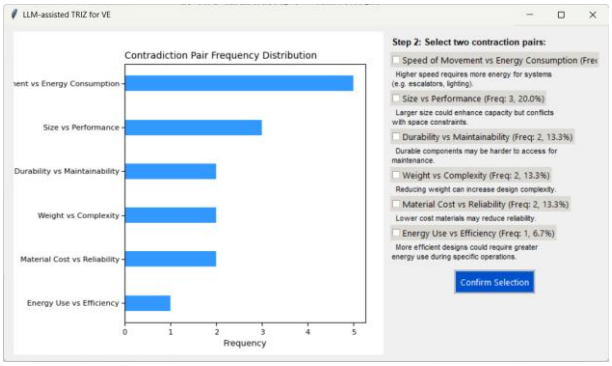


Figure 6: Interface for human-in-the-loop review.

Retrieval of TRIZ principles

After the user confirms their selection, i.e., Speed of Movement vs Energy Consumption and Size vs Performance, the system queries the embedded TRIZ contradiction matrix to retrieve the associated inventive principles. The context-aware idea generation module then dynamically constructs prompts combining the retrieved inventive principles, project objectives, functional targets, and key constraints. The LLM is explicitly instructed within these prompts to reason how each inventive principle can be adapted to the specific project conditions and to articulate the rationale for each proposed idea. The temperature is raised to a higher range, 1.0 in this case, to foster creativity and encourage diverse solution pathways.

Context-aware idea generation

As illustrated in Figure 7, the system generates multiple domain-specific alternative ideas for each inventive principle. Each idea is accompanied by a clear rationale demonstrating how the principle is contextualized within the specific VE problem. For example, under the Preliminary Action principle, the system suggests deploying smart signage that activates before peak periods to proactively guide passenger flow, thus reducing reactive energy demand. Under Mechanics Substitution, proposals such as AR-based wayfinding and virtual queuing are generated, offering innovative alternatives that minimize reliance on physical infrastructure. These generated ideas, together with their reasoning logic, form structured and traceable outputs that can seamlessly feed into the subsequent VE evaluation phase, where design alternatives are assessed for value contribution, feasibility, and alignment with stakeholder objectives. The human-in-the-loop review at this stage further supports the collaborative refinement and selection of the most promising alternatives.

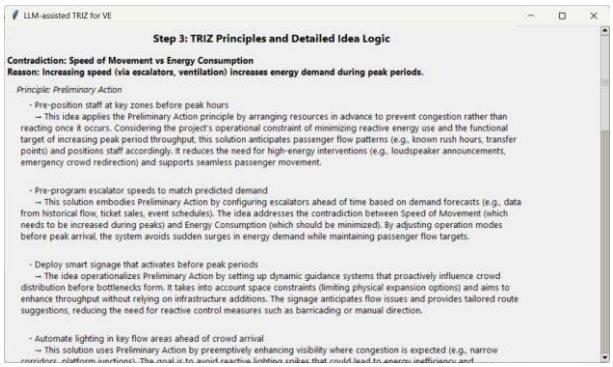


Figure 7: Interface for context-aware TRIZ idea generation with principle-specific rationales

Conclusions

This study presents a framework for integrating LLM-assisted TRIZ reasoning into construction VE. The results demonstrate its feasibility in supporting structured, context-aware, and traceable idea generation during the creativity phase. Its main contributions can be summarized in three dimensions. From a methodological perspective, the proposed framework demonstrates how LLMs can function as structured reasoning assistants, guiding the mapping of project-specific challenges to standardized TRIZ contradictions and supporting context-aware generation of alternative ideas. This advances the theoretical foundation for enhancing the creativity phase in VE by combining AI reasoning with established innovation models. From a technical standpoint, the prototype demonstrates the feasibility of the proposed framework. It supports TRIZ-inspired reasoning in identifying contradictions and generating alternative ideas. These ideas span both technological solutions and operational considerations, such as space optimization and energy efficiency. The inclusion of clear, traceable reasoning chains linking the identification of contradictions, the retrieval of principles, and the rationale for solutions supports subsequent evaluation and decision-making in VE phases. Finally, in terms of practical relevance, this study highlights how LLM-assisted innovation reasoning can be embedded within real-world VE workflows. The tool can serve as a supportive aid in group discussions, providing structured suggestions and reasoning cues that participants can iteratively review and refine. The integration of human-in-the-loop ensures that automation is balanced with expert oversight, making the system adaptable to collaborative and time-constrained workshop settings across varied project contexts.

Despite these contributions, there remain limitations. Specifically, the prototype is developed as a standalone tool and has not yet been integrated into a comprehensive VE management platform. It focuses on demonstrating methodological feasibility and does not include a systematic evaluation of the quality of generated ideas. Future work will incorporate user studies and expert-based comparisons with conventional brainstorming and expert-led TRIZ processes. In addition, the mapping of project-specific challenges to TRIZ parameters relies on LLM-based reasoning, which may introduce inconsistencies across contexts and therefore requires human validation; improving the reliability of this process remains an important direction for future work.

References

Chen, L., Song, Y., Ding, S., Sun, L., Childs, P., & Zuo, H. (2024). TRIZ-GPT: An LLM-augmented method for problem-solving, arXiv:2408.05897. https://arxiv.org/abs/2408.05897

Chen, W. T., Merrett, H. C., Liu, S.-S., Fauzia, N., & Liem, F. N. (2022). A Decade of Value Engineering in Construction Projects. Advances in Civil Engineering, 2022, 1–13. https://doi.org/10.1155/2022/2324277

DeepSeek-AI. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning, arXiv:2501.12948. https://arxiv.org/abs/2501.12948

Guarino, G., Samet, A., & Cavallucci, D. (2022). PaTRIZ: A framework for mining TRIZ contradictions in patents. Expert Systems with Applications, 207, 117942. https://doi.org/10.1016/j.eswa.2022.117942

Gunduz, M., Aly, A. A., & El Mekkawy, T. (2022). Value Engineering Factors with an Impact on Design Management Performance of Construction Projects. Journal of Management in Engineering, 38(3). https://doi.org/10.1061/(asce)me.1943-5479.0001026

Jiang, S., Li, W., Qian, Y., Zhang, Y., & Luo, J. (2025). AutoTRIZ: Automating engineering innovation with TRIZ and large language models. Advanced Engineering Informatics, 65(PC), 103312. https://doi.org/10.1016/j.aei.2025.103312

Kalevi, R., & Ellen R., D. (2007). Simplified TRIZ:New Problem Solving Applications for Engineers and Manufacturing Professionals, Second Edition. In Auerbach Publications (Second Edition). https://doi.org/10.1201/b22111

Lee, C. K. M., Liang, J., Yung, K. L., & Keung, K. L. (2024). Generating TRIZ-inspired guidelines for eco-design using Generative Artificial Intelligence. Advanced Engineering Informatics, 62(PC), 102846. https://doi.org/10.1016/j.aei.2024.102846

Mao, X., Zhang, X., & AbouRizk, S. M. (2009). Enhancing Value Engineering Process by Incorporating Inventive Problem-Solving Techniques. Journal of Construction Engineering and Management, 135(5), 416–424. https://doi.org/10.1061/(asce)co.1943-7862.0000001

OpenAI. (2023). GPT-4 Technical Report, arXiv:2303.08774, 2023. https://doi.org/10.48550/arXiv.2303.08774

Park C.S., Kim H.J., Park H.T., Goh J.H., & Pedro A. (2017). BIM-based Idea Bank for Managing Value Engineering Ideas. International Journal of Project

Management, 35, 699–713. https://doi.org/10.1016/j.ijproman.2016.09.015

Sheu, D. D., Chiu, M. C., & Cayard, D. (2020). The 7 pillars of TRIZ philosophies. Computers and Industrial Engineering, 146(101), 106572. https://doi.org/10.1016/j.cie.2020.106572

Spreafico, C. (2022). Can TRIZ (Theory of Inventive Problem Solving) strategies improve material substitution in eco-design? Sustainable Production and Consumption, 30, 889–915. https://doi.org/10.1016/j.spc.2022.01.010

Wu, Y., Zhou, F., & Kong, J. (2020). Innovative design approach for product design based on TRIZ, AD, fuzzy and Grey relational analysis. Computers and Industrial Engineering, 140(August 2019), 106276. https://doi.org/10.1016/j.cie.2020.106276

Zhang, X., Mao, X., & AbouRizk, S. M. (2009). Developing a knowledge management system for improved value engineering practices in the construction industry. Automation in Construction, 18(6), 777–789. https://doi.org/10.1016/j.autcon.2009.03.004

Zu, F., & Zhang, X. (2025a). Integrating a DfMA Guideline Matrix to Facilitate Value Engineering Workshops for Construction Projects. Journal of Construction Engineering and Management, 151(7), 04025083. https://doi.org/10.1061/JCEMD4.COENG-15777

Zu, F., & Zhang, X. (2025b). Streamline the information phase of the value engineering workshop through a web-based platform integrating BIM and GIS. Engineering, Construction and Architectural Management. https://doi.org/10.1108/ECAM-06-2024-0715

Industrialization, Automation & Robotics

Sub-Tracks:

Lean & Industrialized Construction Methods

Integrated Workflows (BIM / IoT / AI)

Supply Chain Analytics & Optimization

Prefabrication & Modular Construction

3D Printing & Additive Manufacturing

On-Site Robotics & Autonomous Systems

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

FORECASTING MACHINE WORKING TIMES IN OFF-SITE CONSTRUCTION: A MACHINE LEARNING BENCHMARK FOR SUPPORTING PRODUCTION PLANNING

Andrea Sbaragli[1], Xiang Xie[2], John Hedley [2] and Mohamad Kassem[2]
[1]School of Architecture, Technology and Engineering, University of Brighton, Brighton, BN2 4AT, UK
[2]Department of Engineering, Newcastle University, Newcastle, NE1 7RU, UK

Abstract

Off-site construction relocates building processes to controlled manufacturing environments, enhancing productivity but increasing the risk of cost overruns due to the non-repetitive nature of module fabrication. This work presents a machine learning benchmark for a cyber-physical system that digitises manufacturing operations by capturing machine working times alongside design and production features. The benchmark spans tree-based and support vector regressors, with the best-performing model embedded in an industrial dashboard for data-driven scheduling. Validated in a UK steel frame manufacturer across eight months of production data, Random Forest achieved 10.84% median absolute percentage error on 420 frames, enhancing production planning and scheduling.

Introduction

Off-site manufacturing has increasingly been adopted within construction by drawing on industrial production and assembly systems methodologies. Approaches such as prefabrication and modular construction involve manufacturing building components in controlled factory environments before transporting them to the site for installation (Fan et al., 2024). Although this paradigm shift improves project planning, overall productivity, and delivery time, several threats may disrupt production cycles (Mehdipoor and Iordanova, 2021). Construction projects present a wide range of variability, arising from geographical conditions, budgetary constraints, intended building functions, and client personalisation. These drivers may lead to design–production misalignment and prevent manufacturing standardisation. Delays and inaccuracies in translating Building Information Models (BIM) into manufacturing processes can generate bottlenecks (e.g., reworks) that negatively impact profit margins and expose projects to penalties arising from missed deadlines (Jowett et al., 2024). These environments require an expanded digital footprint to enable continuous monitoring of industrial Key Performance Indicators (KPIs) and to support their control within desired efficiency levels (Saif et al., 2024). The fourth industrial revolution vastly digitised and automated manufacturing environments, exploiting several enabling technologies. While robotics and automation have progressively replaced human involvement in repetitive and hazardous tasks, workers continue to play a pivotal role in manufacturing value creation, performing operations where automation remains either technically unfeasible or economically unviable. Yet monitoring the human factor poses particularly significant challenges. Unlike machine processes, human behaviour is inherently variable, shaped by cognitive states, fatigue, skill levels, and environmental conditions, making it difficult to model and standardise. Compounding this, human-related data are often noisy, context-dependent, and subject to rapid fluctuations, further hindering reliable and continuous assessment (Sbaragli and Nardello, 2025). Several cyber-physical systems (CPS) have been developed to support decision-making in industrial processes by leveraging multiple architectural layers. Acquisition layers enable the automated collection of data from shop-floor resources, while cyber layers process these data streams and apply machine learning algorithms to compute Key Performance Indicators (KPIs) (Thiede et al., 2016; Alsakka et al., 2024). Before model deployment, various algorithmic structures and configurations are systematically benchmarked to identify the most performing solution for the specific use case (Sbaragli and Nardello, 2025; Albanese et al., 2022). These industrial metrics are subsequently integrated into dashboards that automate key tasks for manufacturing stakeholders and enhance the operational visibility of industrial processes (Thiede et al., 2016). Among the various industrial applications, CPS have been extensively investigated to support production planning and scheduling (Brunelli et al., 2019). In this context, product-related industrial and design features are used as inputs to machine learning models to predict manufacturing processing times (Çakıt and Dağdeviren, 2023). Despite providing a robust foundation for improving production planning and scheduling operations, these data-driven applications suffer from two main limitations. First, these methodologies offer proof-of-concept solutions, where acquisition layer designs are not discussed, affecting successful deployments in production environments (Çakıt and Dağdeviren, 2023; Masoumi and Bond, 2025). Second, the evaluation of cyber layers and related machine learning models is often constrained to a limited set of indicators, resulting in incomplete evaluations and discussion of their weaknesses (Masoumi and Bond, 2025; Çakıt and Dağdeviren, 2023). Based on these limitations, this work proposes a CPS to support production planning in human-

centric off-site construction environments, with the core contribution being the systematic benchmarking of different machine learning regression models to accurately forecast semi-automated machine working times. The remainder of this work is organised as follows. A literature review of studies on forecasting resources' working time in production environments is presented. This is followed by the presentation of our developed CPS, its evaluation, and the discussion of findings. The last Section ends this work by outlining the conclusions and further research opportunities.

Literature Review

Production planning involves the systematic forecasting, coordination, and scheduling of activities, resources, and materials required to execute manufacturing processes efficiently (Göppert et al., 2024). Traditionally, operations research models, such as multi-objective programming and metaheuristic algorithms, have been extensively applied to support these decisions by iteratively identifying suboptimal industrial configurations (Demartini et al., 2021). For instance, Auer et al. (2021) proposes a heuristic to minimise production throughput time under constraints such as resource capacities and set-up times. However, these approaches often rely on static parameters (e.g., standard cycle times), which limit their effectiveness in dynamic manufacturing environments. In this context, machine learning-powered CPS approaches are increasingly adopted to enhance the robustness and responsiveness of scheduling decisions. Recent advances demonstrate the effectiveness of supervised, unsupervised, and reinforcement learning techniques in forecasting key operational variables, such as machine utilisation and processing times (Göppert et al., 2024). By enabling data-driven and adaptive decision-making, machine learning methods effectively complement the limitations of traditional optimisation approaches. For instance, Çakıt and Dağdeviren (2023) designs a CPS benchmarking machine and deep learning models to predict cycle times of bus manufacturing operations. The trained dataset consists of several features, including the number of sub-assemblies, welding operations, and metal forming processes. The k-nearest neighbor emerges as the most performing classifier with a Root Mean Squared Error (RMSE) equal to 1.48, granting superior visibility into manufacturing operations before scheduling jobs into the shop floor. Despite these promising results, this methodology has some notable limitations. The acquisition layer does not discuss how the dataset features are automatically gathered and thus fed into computational models. Furthermore, some dataset features, such as environmental conditions (products' complexity and ergonomics patterns), are highly subjective as they depend on evaluators' perception as well as workers' experience and physiology. The algorithms' validation does not discuss percentage errors between actual and predicted working times, and falls short in validating the methodology with a production sequence. This analysis is pivotal to discussing how potential under- or overestimation may affect production environments and the related stakeholders' activity. Similarly, Masoumi and Bond (2025) validates an approach to optimizing and estimating the lead time of wooden furniture. The dataset maps chair products as a function of design (e.g., leg shapes) and industrial (e.g., machine configurations) features. Neural networks and tree-based regressors validate the CPS performances, showing a Mean Absolute Percentage Error (MAPE) equal to 1.63%. Despite the excellent performance in forecasting production times, this approach suffers from two main limitations. First, the authors do not discuss the design of the acquisition layer, limiting the deployment into manufacturing environments. Second, the investigation targets fully automated environments, neglecting the pivotal role that workers play in production environments. In particular, the intrinsic entropy of the human factor during task execution may challenge the learning capabilities and thus the classification and regression performances of cyber layers (Sbaragli et al., 2025). Besides traditional manufacturing industries, production planning digital systems have been leveraged in off-site construction environments to better schedule building phases and collaborate with project stakeholders (e.g., site installation). Aghajamali et al. (2022) proposes a CPS to estimate fitting and welding stations' working times in a Canadian construction specialist business. Dataset features relate to each resource involved in the productive value-creation. While fitting station targets the amount of bolted and cut parts, the welding process relies on the product's linear length. These datastreams are fed into a linear regressor and a fully connected neural network to estimate industrial resources cycle times. Despite the neural network's superior performance, the discussion solely evaluates the Mean Squared Error (MSE), falling short in outlining percentage errors between actual and predicted working times. Moreover, authors do not investigate the aggregate errors in a test production sequence in order to identify how potential overestimation may affect the sequential operations of other stakeholders. Data-driven approaches have also been developed to predict construction wall frames production cycle times (Alsakka et al., 2024). In particular, an acquisition layer digitizes the shop floor by leveraging MES and BIM datastreams and Internet of Things sensors such as cameras. This information is fed into neural networks and simulation models to predict working times in real-time. This CPS is validated with a run sequence of 25 assembly frames, showing an underestimation of 22 minutes compared to the actual cycle time. This approach massively outperforms a baseline model that relies on frames' linear meters. However, the sensors' cost, along with their potential occlusions, may limit the applicability of this human-centric solution in other off-site use cases. In summary, the reviewed CPS studies provide valuable methodologies and actionable metrics to support production planning. However, their applicability to off-site construction environments can be further enhanced from two key perspectives.

- **Acquisition layers design and ownership cost:** (Çakıt and Dağdeviren, 2023; Masoumi and Bond, 2025) do not address the design of acquisition layers for automated production data collection. Such functionality is essential for deploying production-ready CPS, as it enables the continuous integration of shop-floor data into algorithms to derive actionable metrics. Furthermore, acquisition systems should be cost-effective to facilitate their adoption in real-world operating environments (Alsakka et al., 2024).
- **Performance benchmarking & evaluation:** the benchmark of machine learning regressors is often based on a limited set of metrics, failing to highlight strengths and limitations in predicting work times. Also, these data-driven approaches are required to investigate how models' predictions affect manufacturing run sequence, potentially affecting other businesses' operations (e.g., site installation) (Masoumi and Bond, 2025; Çakıt and Dağdeviren, 2023).

Building on this discussion, this paper proposes a machine learning-powered CPS to address the identified limitations. The system comprises three interconnected components. First, an acquisition layer consisting of two cloud-based Extract, Transform, Load (ETL) pipelines automates the collection of real-time production data and product design workflows through a cost-effective approach. Second, a cyber layer pre-processes the collected data streams via feature extraction techniques and systematically benchmarks five machine learning regression models to forecast machine working times, including operators' manual material handling activities. Third, a web-based dashboard streamlines production scheduling by automatically dispatching selected frames to stand-alone machines and notifying manufacturing decision-makers via email with predicted working times, while providing key metrics to monitor operations and support integration with other business functions. The CPS is validated in a job shop of a UK-based off-site construction company that manufactures 16 steel frame types. The dataset contains more than 2000 frames manufactured during the first eight months of 2025. Among the benchmarked regressors, the Random Forest achieves the best performance, with a Median Absolute Percentage Error of 10.84% and a percentage residual error of 0.02% on a production order of 420 frames.

Cyber physical system

The digital system is developed to support decision-makers in monitoring and planning fast-paced human-centric off-site production environments. To meet these complementary goals, the CPS consists of four conceptual entities (Fig.1). The first layer targets any human-centric construction environment in which workers execute production cycles by interacting with a diverse set of resources, including machines and assembly stations. An acquisition layer digitizes process execution in these fast-paced, non-standardized settings, replacing reliance on plant managers' experience and lean methodologies. While products' design features and machine working times are acquired using cloud-based ETL pipelines, the Scanning App is an in-house built MES that records product and worker-dependent manufacturing events. These data streams, after specific pre-processing computations, are fed into the third system's entity, the cyber-physical layer. In this regard, machine learning regressors and time-driven algorithms forecast products' working times and derive industrial KPIs. These metrics flow in a dashboard, divided into four different areas. While the first three monitors manufacturing outputs and efficiencies of resources leveraging data coming from the MES (see Fig.1), the planning area presents an interactive user interface to support manufacturing stakeholders in scheduling production run sequences on the shop floor. This preliminary work investigates the feasibility of supporting production planning through the integration of machine learning algorithms within a cyber-physical system. The core contribution lies in benchmarking different regression models for predicting machine working times and in presenting the functionalities of the dashboard's planning area to support run sequence scheduling. The remaining CPS components, such as the MES, are beyond the scope of this paper and are therefore not discussed in further detail.

Acquisition Layer

The acquisition layer runs two parallel ETL pipelines to gather design data streams and machine working time from the shop floor (see Fig. 2). Starting with the first, the structural team publishes the modeled frames on Asite, where any stakeholder may track the product workflow. Any product is uniquely identified by an assembly reference and two supplementary files. While the PDF shows the assembly drawing useful to assist the shop floor task executions, the spreadsheet lists the coil parameters, stud length, and their machine-specific operations (e.g., dimples, notches, swages, etc.) along with a 3D frame position. Each frame also carries a design status, such as On Hold or Approved IFC, along with a revision identifier. A Databricks recursive pipeline (see green lines in Fig.2) continuously accesses all frames regardless of their design status, storing stud identifiers and required machine operations into a cloud database table. As a result, when a frame transitions from On Hold to Approved IFC, the corresponding database records are updated rather than newly created. Therefore, each frame and revision generates multiple entries, one per manufactured stud, with the assembly reference and stud identifier forming the composite primary key. The pipeline runs three times per day to ensure that any revision or status change is promptly reflected in the database. Only the latest revision of frames with an Approved IFC design status can be scheduled for production. Once the production is scheduled and running, the stand-alone machine processes frame-specific spreadsheets and records studs working times in TXT logs. The machine is network-connected and exposed to a static IP address that facilitates the data retrieval of the second pipeline (see blue

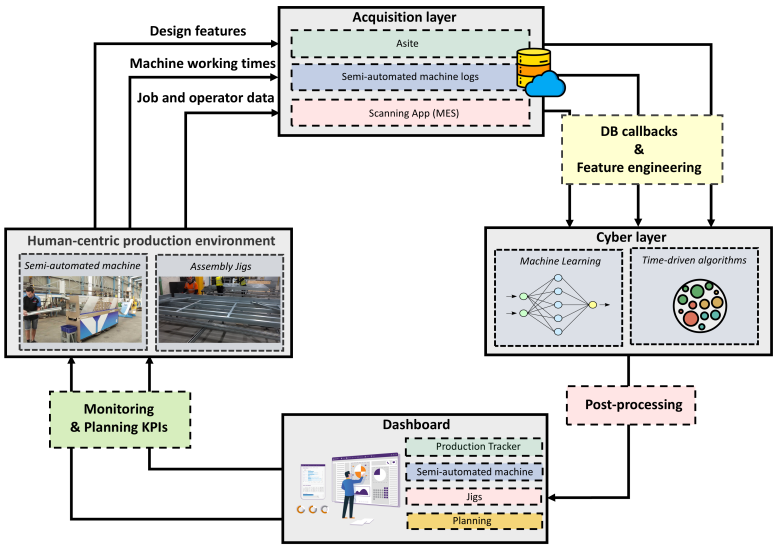


Figure 1: Cyber-physical system to monitor and plan human-centric offsite construction environments

lines in Fig.2). Production data are stored in a separate table but share the same composite primary key of the design flow. The machine pipeline time resolution is equal to 5 minutes, enabling future close-to-real-time shop floor operations monitoring.

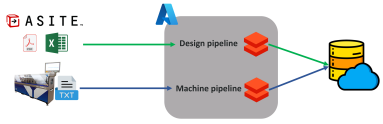


Figure 2: Acquisition Layer data flow

Cyber Layer - Machine Learning

Acquired datastreams are retrieved using database callbacks and subsequently joined on the shared composite primary key. Multiple machine learning models, commonly applied in industrial applications, are then trained on the aggregated dataset to forecast steel frames production times based on design features (Sbaragli et al., 2025). The dataset was partitioned into training and testing subsets using an 80%-20% split, and model performance was evaluated using 5-fold cross-validation. Models' hyperparameters are sub-optimized through the grid search algorithm. The following bullet point lists the evaluated models:

- **Baseline:** this model assumes a direct and linear relationship between frames' length and machine work time, as investigated by Alsakka et al. (2024). The training defines a coefficient as $minutes/millimeter$ that is used to infer process times in the test set.
- **Linear Regressor (LR):** a parametric model that assumes a linear mapping between predictors and the target variable. As LR requires no hyperparameter tuning, it represents a computationally efficient baseline against which to compare more flexible models (Aghajamali et al., 2022).
- **Support Random Vector (SVR):** this kernel-based method fits a regression function within an error tolerance. Its predictive behavior is governed by several parameters. While the regularization parameter C controls the trade-off between model complexity and tolerance to errors, the kernel coefficient γ determines the influence of individual training samples. Finally, the insensitive loss margin ϵ defines the tolerance zone around predictions where no penalty is applied (Çakıt and Dağdeviren, 2023).
- **Random Forest (RF):** an ensemble approach that develops multiple decision trees on bootstrapped samples and aggregates their predictions to reduce variance. Performance depends on parameters such as the number of trees, node-splitting thresholds, and feature selection strategy (Sbaragli et al., 2025).
- **Light Gradient Boosting Machine (Light GMB):** this boosting-based ensemble regressor builds trees sequentially, where each iteration corrects the residuals of the previous one. Key parameters include the number of boosting rounds, learning rate, boosting type (e.g., GBDT or DART), and the learning objective set for regression (Masoumi and Bond, 2025).

After having trained and tested these classifiers, the most performing model is used to support manufacturing engineers in production planning and scheduling. While the comparison benchmark of models is discussed in the Results Section, the following subsection outlines the benefit of leveraging this data-driven CPS in off-site manufacturing environments.

Dashboard - Planning Area

Building on the discussed cyber-layer capabilities for predicting machine work times, this final subsection discusses the dashboard's planning area. The frontend presents an interactive table that reflects the design workflow captured by the ETL pipeline. As shown in Fig. 3, the table maps assembly references (e.g., ARef) against key attributes including project, coil configuration, frame length

(e.g., Len), and the number of screws to be fitted in the jig area. Since an assembly reference may go through multiple design revisions, the table displays only the latest revision for each, ensuring that engineers always work with the most up-to-date information. Each assembly entry also displays its current design status and scheduled production day. Manufacturing engineers can schedule a frame to the shop floor only when two conditions are met: i) the structural manager has set the design status to Approved IFC in Asite, and ii) the frame has not yet been assigned to a production date (i.e., the field is null in Fig.3). To support efficient navigation of large datasets, a filter button allows users to narrow the table view by attributes such as project, design status, and coil configuration. Once the desired eligible items have been selected for scheduling, the user clicks the Schedule Selected button, which triggers two back-end operations. First, the machine files for the selected frames are transferred via the File Transfer Protocol to a designated folder on the stand-alone machine, which the shop floor operator can then open and automatically load. Second, a spreadsheet summarising the expected production times and linear coil consumption for each configuration is automatically emailed to the relevant production stakeholders, such as manufacturing engineers and the production director. By automating both file transfer and the manual development of run sequences, the dashboard significantly reduces the non-value-added administrative burden on engineers, ultimately driving meaningful productivity enhancements across the production workflow.

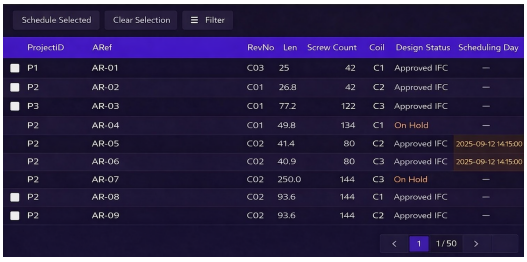


Figure 3: Interactive web-based application to support production scheduling

This novel data-driven methodology provides the business a key opportunity to strengthen its operational value creation. Indeed, the manufacturing team has greater operations visibility that enables better coordination and collaboration with other business areas:

- **Purchasing & Accounting:** the known a priori coil configurations consumption facilitates a just-in-time approach that minimizes stock occupancy and avoids cost overruns in low-rotating or project-specific material. At the same time, project cash flows, such as labor and material costs, are seamlessly integrated.
- **Site:** forecasted due dates of run sheets enable a data-driven definition of frames shipping windows, avoiding project delays and related penalties for missed deadlines.

Results

This section validates the CPS by feeding structural frames' design features into machine learning models to forecast semi-automated resource working times, as captured by the ETL pipeline. Notably, these times encompass material handling activities, which may challenge regressors' learning capabilities. Before discussing the performance of the benchmarked regressors, the following subsection discusses 1) the off-site use case manufacturing operations and 2) the obtained dataset by collecting 8 months of steel frame structural production.

Off-site use case: structural steel frame production

The discussed CPS is tested and validated in a UK-based off-site construction specialist. The monitored human-centric job shop manufactures several steel frame products involving different resources. Fig.4 outlines the off-site manufacturing operations, starting from the steel coil, which is distinguished by different configurations of depth and width (e.g., C1, C2, etc.). Once structural frames are modeled and set to Approved IFC, they are released into the shop floor by using the discussed interactive web-based dashboard. The manufacturing process is then carried out by a semi-automated machine, in which a skilled operator loads the machine file before execution and moves sequentially processed studs into work-in-progress (WIP) storage units. A jolly worker may join the process to facilitate the material handling of studs up to 6.4 meters long. Based on the final product's complexity, a structural steel frame may consist of up to 74 studs. Each processed stud is distinguished by a set of design features (e.g., dimples, swages, notches, length, weight, etc.) that affect the machine's working cycle times. However, ETL-captured working time logs, whose resolution and structure are dictated by the machine output rather than experimental design, also include material handling events. This may challenge machine learning regressors due to the intrinsic process variability (e.g., differences in stud size) and variability in human commitment and experience (Sbaragli and Nardello, 2025). The construction job shop follows a one-piece flow approach in normal operating conditions, and thus, WIP units are moved next to the available or less saturated assembly jig. Four jigs assemble steel frame products in parallel, leveraging up to 4 skilled operators each. Based on the designed assembly drawings, manual operations require stud fittings and screwing operations, as well as potential reworks to address some modelling and/or machine errors. Completed frames are classified into 16 products and thus follow separate production flows (see Fig.4). While Product IDs 1 and 2 are sent to the construction site for their installation, the remaining ones represent WIP modules for other in-plant job shops.
Based on this discussion on manufacturing operations, the following subsection details the acquired datasets from the semi-automated machine.

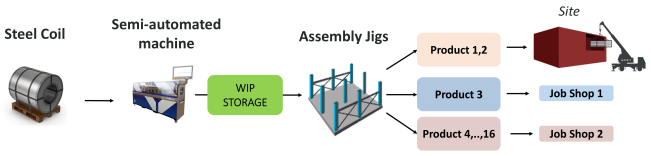


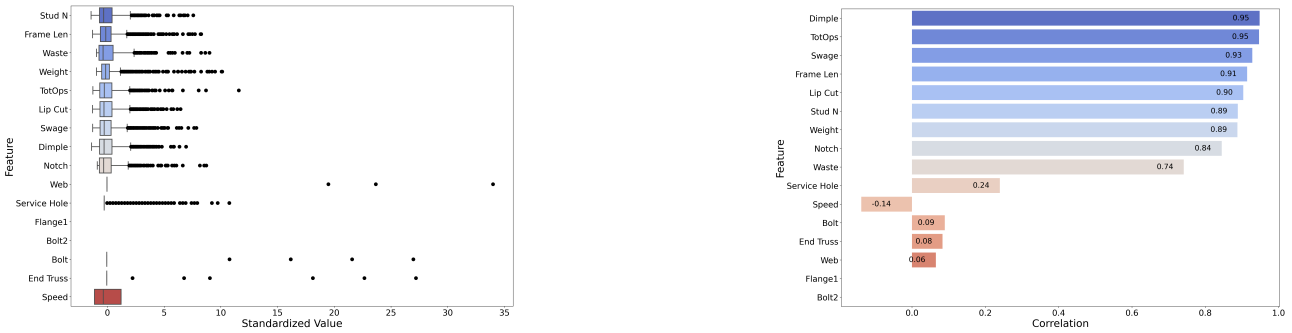
Figure 4: Human-centric production environment

Dataset

To validate the proposed methodology in forecasting machine processing times, 8 production months in 2025 (e.g., from January to August) are retrieved from the resource-specific database involving 5 construction projects across the United Kingdom. The stud-based dataset comprises 26500 samples, with individual work times ranging from 2 to 71.8 seconds. The average process time and standard deviation are equal to 20.4 and 14.2 seconds, respectively. This low-level dataset is grouped by assembly reference to achieve a frame-based representation, where one ARef may group 6 to 74 studs based on the final product complexity and application. This aggregation is implemented for two main aspects. First, the studs' working times are highly exposed to the intrinsic uncertainty of the human factor during task executions. The dedicated operator often pauses the machine to perform audit checks and move manufactured items into WIP storage units. Despite being cyclic, these events have high degrees of entropy, affecting the studs' working times and limiting the learning capabilities of any machine learning model. Second, frames might be assembled only if all studs are processed and stored in a specific location (see Fig. 4). The new compressed dataset includes 2099 steel frames having working times ranging from 0.2 to 26.4 minutes. The average process time and standard deviation are equal to 4.3 and 2.9 minutes, respectively. Once the machine dataset is compressed, a separate callback downloads and post-processes data from the design database table. Similarly to the machine dataset, frame features are obtained by summing or counting the studs descriptors while grouping on the ARef field. These frames' features include the total amount of required manufacturing operations (e.g., Lip Cut, Notch, Web, Service Hole, etc.), the aggregate dimensions (e.g., length and weight), and the count of grouped studs. Finally, the two datasets are joined together on ARef, the common and unique column. Fig. 5 (a) outlines the entire set of frame features, where waste and speed represent steel scrap and machine velocity in manufacturing a given product. It is worth noting that Stud N and TotOps represent the total amount of studs and machine operations on every frame, respectively. The remaining structural features are not detailed further due to the company non non-disclosure agreement. The job shop's production output is very customized to meet the customer and construction requirements of different healthcare projects. Fig.5 (a) streamlines features relative variability and distribution patterns through z-scores. The 16 different variables have standardized values ranging from 0 to approximately 35 and might be divided into different sub-groups. Structural information (e.g., Lip Cut, Swage, Dimple, Notch, etc.), dimensions (e.g., Weight and Frame Len), and Waste present compact distributions with upper quartiles below the standardized unit 5. However, many outlier reiterates the uniqueness of structural modules within projects. Less-exploited design specifications, such as Web, Bolt, and Service Hole, present very narrow distribution close to zero but with several distinct outliers at higher values. Flange1 and Bolt2 are not represented in this dataset. Finally, the Speed has no whiskers, underlying machine velocities equal to 20%, 50%, and 70% of the maximum. This set of features is ranked to reduce computational load and simplify the learning process for machine learning models. The data distributions are non-Gaussian, may lack linear relationships with machine work time, and include many outliers. Therefore, the Spearman correlation is applied to evaluate the strength and direction of monotonic associations. Fig. 5 (b) ranks the pair-wise correlation between each feature and the machine work time. It is worth noting that the Speed correlation captures a weak inverse relationship with the processing time. Feature selection employed a correlation threshold of 0.7 to minimize dimensionality for machine learning models. Therefore, the subsequent performance benchmark discussion excludes all descriptors with correlations below this cutoff, from Service Hole to Bolt2.

Working times prediction: algorithms' performance benchmarking

The structural dataset has been validated using the five different regressors discussed in the cyber layer subsection. The training process involves five-fold cross-validation and grid-search hyperparameter optimisation. Regressors are benchmarked using Training Time (TrTime), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2), which provide standard error and goodness-of-fit measures, alongside Median Absolute Error (MdAE) and Median Absolute Percentage Error (MdAPE) as robust, scale-independent alternatives less sensitive to outliers. Tab.1 benchmarks the performance of the algorithms to forecast the semi-automated machine work times by having as input the features discussed in the previous section. The baseline approach that leverages frame length delivers the weakest performance. The high error levels (e.g., RMSE = 1.19, MAE = 0.78, MAPE = 19.09%) and a relatively low coefficient of determination (e.g., R^2 = 0.84) reflect the model's inability to capture the underlying data patterns and relationships. LR substantially improves these metrics, achieving the highest R^2 and competitive error values. RMSE and MAE are equal to 0.72 and 0.51, respectively. However, LR's percentage-based metrics (e.g., MAPE = 14.51%, MdAPE = 11.36%) remain higher compared to tree-based methods, suggesting reduced robustness when predicted working times are



(a) Z-score features distribution

(b) Spearman Correlation

Figure 5: Dataset features distribution and ranking

Table 1: Performance comparison of regression models.

Regressor	TrTime [sec]	RMSE	MAE	MAPE	MdAE	MdAPE	R^2
Baseline	NA	1.19	0.78	19.09%	0.50	17.81%	0.84
LR	NA	0.72	0.51	14.51%	0.37	11.36%	0.94
SVR	8.2	0.88	0.55	17.26%	0.39	12.39%	0.92
RF	**44.0**	**0.85**	**0.52**	**13.79%**	**0.33**	**10.84%**	**0.92**
Light GBM	14.8	0.85	0.55	16.02%	0.37	11.92%	0.92



(a) Absolute Residuals

(b) Relative Residuals

Figure 6: Aggregate model performances on the test set

assessed relative to actual values. SVR performs moderately well, even though the RMSE of 0.88 and MAPE equal to 17.26% indicate less stability compared to LR and RF models. Light GBM achieves comparable performance to SVR (e.g., RMSE = 0.85 and MAE = 0.55), but MAPE and MdAPE equal to 16.02% and 11.92% show limitations in effective generalization. Within this benchmark, the RF regressor demonstrates the most balanced and robust performance metrics. Despite a training time equal to 44 seconds and slightly higher exposure to outliers compared to LR (e.g., RMSE = 0.85), RF yields the lowest MAPE (13.79%) and MdAPE (10.84%), indicating that its predictions are not only accurate but also less sensitive to variations in scale. The performance benchmark is further expanded by investigating how models' performances affect the run sequence of the test set (e.g., 420 frames). In this regard, Fig. 6 depicts the aggregate absolute and relative residuals after having summed the actual machine working times and the ones predicted by each algorithm. RF is the only model to achieve an aggregate absolute residual close to zero (0.45 minutes), whereas LR, SVR, and LightGBM display absolute and relative deviations ranging from 6.05 to 8.20 minutes and from 0.34% to 0.46%, respectively. The baseline remains the worst regressors, accounting for an underestimation of 179.23 minutes and 9.94%.

Operational implications

These results confirm that the RF regressor provides the most reliable and operationally robust performance among the tested models (Tab. 1 and Fig. 6). Crucially, the dashboard workflow automates production scheduling by generating optimised run sequences directly from the web application, eliminating the need for engineers to interact with multiple company software systems (e.g., Asite), manually download files, or transfer machine files to the stand-alone resource. Preliminary investigations on run sequences ranging from 20 to 30 structural frames demonstrate a 75% reduction in planning time. At the same time, machine learning-based predictions enhances visibility into manufacturing operations, allowing stakeholders to better anticipate workloads and coordinate downstream processes. By integrating this capability across departments, the system supports tighter alignment across

key functions. For planning, it enables more realistic deadline scheduling against production constraints. For purchasing, it facilitates just-in-time material deliveries and reduces stock occupancy. For finance, it provides advance visibility into material consumption, supporting improved cost forecasting and budget control.

Conclusions & Further Research

Off-site construction environments tend to suffer from cost overruns and constantly delayed deadlines due to poor operations management and a lack of standardization across projects. Therefore, construction factories are required to leverage Industry 4.0 enabling technologies to support production planning and achieve a more reliable operations visibility. This work proposes a machine-learning-powered CPS that supports manufacturing stakeholders in forecasting machine cycle times and automating run sequences on the shop floor. While two ETL pipelines acquire design features and machine work times, the cyber layer benchmarks 5 different algorithms, establishing the RF as the most performing model to forecast frame production times. This tree-based regressor is embedded in an interactive web-based dashboard, enabling manufacturing stakeholders to forecast production times before scheduling run sequences into the shop floor. This CPS has been validated in a UK-based steel frame job shop production using 8 months of data in 2025. Further research should validate this methodology in the jigs area and use metaheuristic algorithms to optimise production scheduling given a set of constraints.

Acknowledgments

This work has been supported by Innovate UK under the Knowledge Partnership Program with grant number 10099000.

References

Aghajamali, K., Suliman, A., Alattas, A., and Lei, Z. (2022). Data-driven cycle time prediction of fitting and welding stations in steel fabrication. Modular and offsite construction (MOC) summit proceedings, pages 122–129.

Albanese, A., Nardello, M., Fiacco, G., and Brunelli, D. (2022). Tiny machine learning for high accuracy product quality inspection. IEEE Sensors Journal, 23(2):1575–1583.

Alsakka, F., Yu, H., El-Chami, I., Hamzeh, F., and Al-Hussein, M. (2024). Digital twin for production estimation, scheduling and real-time monitoring in off-site construction. Computers & Industrial Engineering, 191:110173.

Auer, P., Dósa, G., Dulai, T., Fügenschuh, A., Näser, P., Ortner, R., and Werner-Stark, Á. (2021). A new heuristic and an exact approach for a production planning problem. Central European journal of operations research, 29(3):1079–1113.

Brunelli, L., Masiero, C., Tosato, D., Beghi, A., and Susto, G. A. (2019). Deep learning-based production forecasting in manufacturing: a packaging equipment case study. Procedia Manufacturing, 38:248–255.

Çakıt, E. and Dağdeviren, M. (2023). Comparative analysis of machine learning algorithms for predicting standard time in a manufacturing environment. AI EDAM, 37:e2.

Demartini, M., Tonelli, F., Pacella, M., Papadia, G., et al. (2021). A review of production planning models: Emerging features and limitations compared to practical implementation. Procedia CIRP, 104:588–593.

Fan, J., Chen, L., and Chen, K. (2024). Digitalizing industrialized construction projects: Status quo and future development. Applied Sciences, 14(13):5456.

Göppert, A., Kaven, L., Baum, J., Melnychuk, O., and Schmitt, R. (2024). Machine learning for online scheduling in manufacturing: A systematic literature review. Procedia CIRP, 130:154–160.

Jowett, B., Edwards, D. J., and Kassem, M. (2024). Field bim and mobile bim technologies: a requirements taxonomy and its interactions with construction management functions. Construction Innovation, 24(1):134–163.

Masoumi, A. and Bond, B. H. (2025). Machine learning-based prediction of processing time in furniture manufacturing to estimate lead time and pricing. European Journal of Wood and Wood Products, 83(1):35.

Mehdipoor, A. and Iordanova, I. (2021). Systematic literature review on the combination of digital fabrication, bim and off-site manufacturing in construction—a research road map. In Canadian Society of Civil Engineering Annual Conference, pages 283–296. Springer.

Saif, W., RazaviAlavi, S., and Kassem, M. (2024). Construction digital twin: a taxonomy and analysis of the application-technology-data triad. Automation in Construction, 167:105715.

Sbaragli, A., Ghafoorpoor, P. Y., Thiede, S., and Pilati, F. (2025). A cyber-physical architecture to monitor human-centric reconfigurable manufacturing systems. Journal of intelligent manufacturing, pages 1–23.

Sbaragli, A. and Nardello, M. (2025). From sensing to segmentation: Transformer-based worker activity recognition for industrial assembly. IEEE Access, 13:207475–207487.

Thiede, S., Juraschek, M., and Herrmann, C. (2016). Implementing cyber-physical production systems in learning factories. Procedia Cirp, 54:7–12.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

EMERGING BUSINESS-MODEL ARCHETYPES OF A SMART MOBILE FACTORY FOR CONSTRUCTION 4.0

Alexander N. Walzer[1], Irfan Čustović[2], Konrad Graser[1] and Daniel M. Hall[2]
[1] Zurich University of Applied Sciences, Winterthur, Switzerland
[2] Department of Management in the Built Environment,
Delft University of Technology, The Netherlands

Abstract

Smart Mobile Factories (SMFs) bring prefabrication to the construction site, enabling just-in-time delivery, shorter haulage distances and lower carbon emissions. Although integrated digital twins and modular factory designs have demonstrated the technical feasibility of SMFs in construction, their value-creation mechanisms for operators and clients remain unclear. Adapting the Business-Model Canvas to SMFs, we identify four emerging archetypes. Two case vignettes illustrate the transition from a low-tech, asset-centric to a high-tech, data-centric service model. Finally, we propose a mapping that relates organizational digital maturity to suitable SMF business-model archetypes, which allows to indicate a possible path toward profitable and sustainable adoption.

Introduction

Construction activities account for roughly 6% of European gross domestic product, a percentage that has steadily declined in the past two decades (UNECE, 2025), trailing manufacturing in terms of productivity, schedule reliability, and environmental performance (Hamza et al., 2019). Off-site prefabrication has long been held as a strategy for increasing productivity, but the approach is still suffering from the spatial and temporal disconnect between off-site prefabrication plants and on-site assembly zones. Transporting large components over long distances inflates fuel consumption and forces projects to maintain excess on-site inventory, a phenomenon often called the just-in-time delivery paradox (Johansson et al., 2024). Smart Mobile Factories (e.g., Figure 1) have emerged as technological response to these limitations, reintroducing production flexibility into the construction context through mobility (Matt, Rauch & Dallasega, 2015).

Moreover, recent research suggests that mobile factories can introduce a new level of efficiency and flexibility into construction projects by relocating production closer to the point of use (Wagner et al., 2020; Nahmad Vazquez, Garivani & Dackiw, 2024). Beyond just differentiated in logistics, SMFs are typically conceived as highly digitalized production systems, embedding sensors, data integration, and automated control within a controlled, yet relocatable environment (Ibid.).

This setting facilitates process monitoring and data-driven coordination that is difficult to achieve in conventional, exposed, and constantly changing construction sites. The combination of mobility and digital control therefore distinguishes SMFs operationally and in terms of their data generation and decision-making potential.



Figure 1: Smart Mobile Factory for Infrastructure (SMF4INFRA, 2025)

Despite increasing conceptual maturity and pilot applications, the economic rationale for SMFs remains weakly articulated. Existing research has largely focused on technical configurations and productivity gains, with limited attention to value capture, revenue generation, and risk allocation (Matt, Rauch & Dallasega, 2015). At the same time, insights from servitization and Industry 4.0 research highlight the potential of digital technologies to enable service- and performance-based business models, but these are primarily derived from stationary manufacturing contexts and insufficiently address the mobility, project-based organization, and site-specific constraints of construction (Grubic, 2014; Frank et al., 2019). This gap is rather critical, as digitalization requires coordinated business model alignment across multiple actors (Kohtamäki et al., 2019), and the high upfront investment associated with SMFs amplifies financial risk perceptions when revenue logic is unclear (European Commission, 2025; Walzer et al., 2024b). Consequently, the relationship between organizational digital maturity and viable, scalable business models of SMFs remains poorly understood and is the center of this study.

State of the Art

Smart Mobile Factories

A mobile factory is defined as a distributed production facility composed of reconfigurable and mobile production systems that can adapt capacity and functionality to site conditions (Dallasega et al., 2024).

By physically relocating prefabrication activities to the construction site, these factories challenge the traditional separation between manufacturing and assembly. Producing closer to the site reduces transportation distances and delivers clear environmental and economic advantages (Ma et al., 2025). This proximity also supports more responsive production planning, reducing storage requirements and buffer inventories that remain a persistent source of inefficiency in construction logistics (Matt, Rauch & Dallasega, 2015). Moreover, the relocatable nature of mobile factories allows prefabrication benefits to be extended to remote or otherwise difficult project locations. From a sustainability perspective, SMFs contribute to lower transportation-related emissions, improved resource efficiency, and reduced waste, while enabling the use of energy-efficient production technologies (Matt, Rauch & Dallasega, 2015).Smartness in SMFs is achieved through the integration of advanced digital technologies that connect factory operations with site-level management systems (Čustović et al., 2023b). These systems provide a digital representation of processes, materials, and equipment, enabling more transparent and coordinated production control. Over time, accumulated production data supports performance benchmarking and improves the prediction of system behavior under varying conditions (Thelen et al., 2022). Recent empirical studies show that IoT sensor networks enhance quality monitoring in mobile factory settings (Tripathi et al., 2025). In addition, digital twin (DT) approaches allow production plans to be adjusted dynamically in response to uncertainty (Čustović et al., 2025) and support informed decisions regarding factory relocation, considering schedule, cost, and environmental impacts (Cao et al., 2024; Ma et al., 2025). These developments position SMFs as data-intensive systems at the intersection of Construction 4.0 and service-oriented manufacturing (Jäkel et al., 2024; Karmakar & Delhi, 2021; Sawhney et al., 2020).

Business Model Aspects of Smart Mobile Factories

Despite technological progress, the business model implications of SMFs remain weakly developed. The majority of SMF-related studies frame these systems primarily as technical solutions, emphasizing automation, system design, and efficiency gains (Matt, Rauch & Dallasega, 2015). As a result, commercially critical issues such as pricing mechanisms, asset and data ownership, lifecycle cost allocation, and inter-organizational governance are rarely examined in depth. This narrow focus limits understanding of how technical performance translates into economic value across project networks. By contrast, the servitization and Industry 4.0 literature studies the role of digital capabilities in enabling outcome-based services and collaborative value creation (Grubic, 2014; Frank et al., 2019; Čustović et al., 2023a). Yet, these models typically assume stable production environments and long-term customer relationships, assumptions that do not align with the temporary, mobile, and project-based nature of construction. When applied uncritically to SMFs, such assumptions obscure challenges related to demand volatility, site-specific customization, and short-term contractual arrangements.

This disconnect becomes particularly problematic given the high capital intensity and demand uncertainty associated with industrialized construction solutions (Čustović et al., 2023b). Without explicit mechanisms for value capture and risk sharing, these characteristics reinforce financial risk perceptions and slow adoption. Although recent construction management reviews recognize that construction business models are evolving (Li et al., 2025), they offer limited guidance on how digital maturity, such as the integration of BIM, IoT, and DTs (Revolti et al., 2025), can be monetized in SMFs. Consequently, the state of the art lacks an integrated perspective that links digital maturity, mobility, and performance outcomes to sustainable business models for SMFs.

Therefore, in this paper we investigate the research question: "*Which business-model configurations are currently emerging for Smart Mobile Factories for Construction 4.0, and how can practitioners select the most appropriate configuration given their digital maturity?*"

Methodology

We pursued a design science approach, yet the nascent state of the field prevents iterative artifact testing (Hevner & Chatterjee, 2004; Peffers et al., 2007; Kuechler & Vaishnavi, 2008). Thus, we started from the canonical Business-Model Canvas (BMC) by Osterwalder & Pigneur (2010) and enriched it with three SMF-specific dimensions (smartness, mobility, site-specificity). The resulting SMF-specific BMC (Table 1) served as the a-priori coding framework for all subsequent document analysis (Hsieh & Shannon, 2005). Seven sources (partly peer-reviewed and a project report) were selected using a purposive sampling approach and structured reading because they (i) discuss SMFs or related mobile-prefabrication concepts, and (ii) contain at least one explicit reference to business-model elements. The corpus comprises: Dallasega et al., 2024; SMF4INFRA, 2025; Martínez et al. , 2013; Park et al., 2019; Frank et al., 2019; Grubic, 2014; Tripathi et al., 2025. Using the SMF-specific BMC as a coding sheet, each document was read line-by-line, and every passage that corresponded to one of the nine BMC blocks (value proposition, customer segments, channels, revenue streams, key resources, key activities, key partnerships, cost structure, customer relationships) was extracted. Extracted excerpts were entered into a shared Microsoft Excel workbook (one column per BMC block, one row per source). The spreadsheet was then synchronised on a collaborative MIRO board so that all co-authors could visualise the coding and discuss emerging ambiguities in real time. Two pilot projects, the SMF4INFRA hyperloop prototype for linear infrastructure (Figure 1) (SMF4INFRA, 2025) and the Flexible Field Factory for housing construction (Figure 2) (Martínez et al., 2013), are further used as case vignettes.

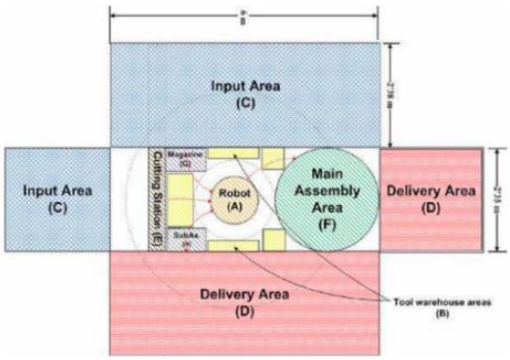


Figure 2: Flexible Field Factory (Martínez et al., 2013)

Result Analysis

Several SMF elements for each of the nine blocks of the BMC are reported in Table 1. The coded BMC allows to reveal the SMF's value-creation logic: the value proposition centres on just-in-time on-site prefabrication that cuts haulage distance and CO_2 emissions; channels and key resources (DT platform, IoT sensor network, mobile plant) describe how the service is delivered, while revenue streams (fixed-term lease, per-component assembly fees, transaction fees, DaaS subscriptions) reveal how value is captured. From these elements, two overarching insights emerge.

First, the plant's mobility reduces on-site inventory and creates a distinct cost item (relocation logistics) that can be monetised with low-to-mid digital-maturity business models (see Table 2). Second, the continuous sensor feed within the DT can enable a high-margin Data-as-a-Service (DaaS) revenue stream, but only when the SMF reaches a very high digital-maturity level. These patterns form the basis for the four business-model archetypes presented in the next section, each grouping a distinct combination of BMC blocks along the servitisation continuum.

Table 1: SMF-specific Business-Model Canvas (BMC)

BMC block	SMF-specific element derived from the literature
Value proposition	Just-in-time on-site prefabrication that eliminates long haulage, cuts CO_2, and compresses the schedule, improved quality and quality control
Customer Segments	Large-scale infrastructure owners (rail, tunnel, hyperloop); modular-housing developers; public-private partnership (PPP) consortia
Channels	Direct turnkey contracts; digital marketplace for on-demand modules (online portal); BIM-integrated procurement API
Revenue streams	Fixed-term lease fee for the mobile plant; per-component assembly charge; KPI-linked performance bonus/penalty (schedule, CO_2); transaction fee on marketplace sales; subscription for real-time production data (DaaS)
Key resources	Mobile modular plant (cranes, CNC, material mixers); BIM/DT platform (cloud server, GIS); IoT sensor network (temperature, vibration, GPS); skilled on-site crew (fabricators, data analysts)
Key activities	Relocation logistics (move plant, set up utilities); production of components; on-site assembly of prefabricated components; continuous data acquisition & analytics (DT updates); life-cycle-assessment feedback loop (track CO_2, adjust processes)
Key partnerships	Equipment OEMs (provide modular plant modules); local utilities (power, water, waste); logistics providers (fuel, auxiliary material transport); municipal authorities (permits, road closures)
Cost structure	Capital purchase/depreciation of the mobile plant; energy consumption (diesel generators, electricity); IT/cloud services (BIM licenses, DT hosting); training & certification of the mobile crew; maintenance & spare-part inventory
Customer Relation-ships	Long-term service contracts (maintenance, upgrades); outcome-based SLAs with bonus/penalty clauses; platform support desk for marketplace users; data-portal access for DaaS subscribers

Emerging Business-Model Archetypes of SMFs

Based on the BMC elements presented in Table 1, we derived four distinct business-model (BM) archetypes. The four business-model archetypes presented in Tables 2-5 are not directly observed in the assessed literature as such but constitute an original typology constructed by the authors through pattern induction across the coded BMC elements. This is understood as being rather projective.

Each archetype was interpreted as a coherent revenue logic and labelled according to the servitisation continuum (product-centric → service-centric → data-centric). In a second step, we linked each cluster to the digital-maturity level required to realise the associated activities (low to very-high). Tables 2-5 summarizes the four archetypes, their defining BMC pattern, the primary adoption drivers, the construction sectors where they are most common, and the digital-maturity prerequisite.

Product-Plus (low digital maturity): In this archetype the SMF is treated as a capital asset that is bought or leased, and revenue comes from a per-module assembly fee. Early pilots (Martínez et al., 2013) report exactly this cost-plus arrangement, and servitisation theory locates it at the product-centric end of the continuum (Grubic, 2014; Frank et al., 2019). Minimal sensor integration means the owner only needs a basic BIM model; risk is shifted to the contractor through the fixed-lease payment. Product-Plus therefore provides a baseline from which higher-maturity models can be differentiated.

Table 2: SMF-specific Business-Model: Product-Plus

Core BMC pattern(s)	Revenue: fixed lease plus per-component assembly fee; Key Activities: on-site assembly; Cost: CAPEX amortisation
Adoption driver(s)	Need for predictable cash-flows and CAPEX spreading; high repeatability of components makes a fixed lease attractive; owners prefer to outsource assembly while retaining design control
Dominant sectors	Low-rise modular and residential housing, industrial plant construction
Required maturity	Low (basic BIM, no continuous data)

Outcome-Based (medium digital maturity): This archetype ties revenue to the achievement of predefined KPIs such as schedule compliance or CO_2 reduction. Performance-based contracts are documented in construction research (Li et al., 2024; Frank et al., 2019), and understood as used in the SMF4INFRA hyperloop case (Cao et al., 2024; Ma et al., 2025). Implementing this model requires a modest digital stack, such as a BIM-linked DT for schedule forecasting and a limited set of IoT sensors that feed KPI data to the client's dashboard. Outcome-Based models adds a performance-monitoring layer to Product-Plus, preparing the SMF for more data-intensive configurations.

Table 3: SMF-specific Business-Model: Outcome-Based

Core BMC pattern(s)	Value: schedule + CO_2 guarantees. Revenue: performance-linked bonus / penalty; Customer Relationships: SLAs
Adoption driver(s)	Contracts tie funding to schedule and carbon-emission KPIs; risk-sharing incentives push both parties toward performance-based remuneration
Dominant sectors	Linear infrastructure (rail, tunnel), infrastructure renewal (bridges, dams)
Required maturity	Medium (real-time KPI dashboard, basic sensors)

Platform-Marketplace (high digital maturity): This archetype turns the SMF into a digital hub that coordinates third-party suppliers through an online marketplace. Platform-based servitisation is well established (Grubic, 2014; Kohtamäki et al., 2019) and BIM-linked API that would enable external vendors to submit designs and receive automated compatibility checks (Cao et al., 2024; Thelen et al., 2022). This configuration demands a full-scale BIM model, open-API interfaces, and a cloud-hosted portal; revenue is generated via transaction fees and optional analytics subscriptions. Network effects increase platform value as more suppliers join. Platform-Marketplace occupies the service-centric pole of the servitisation continuum, extending Outcome-Based by adding active ecosystem orchestration.

Table 4: SMF-specific Business-Model: Platform-Marketplace

Core BMC pattern(s)	Channels: online portal; Revenue: transaction fee; Partnerships: OEMs + utilities; Key Resources: DT platform
Adoption driver(s)	Coordination of many third-party suppliers along a moving front requires a shared digital ecosystem; economies of scale and network effects make a marketplace financially viable
Dominant sectors	Complex infrastructure projects with many suppliers; emerging urban-retrofit projects
Required maturity	High (full BIM-linked twin, open APIs)

Data-as-a-Service (very high digital maturity): this archetype monetises the continuous IoT stream from the SMF, offering analytics-as-a-subscription. IoT-driven construction services can show high margins for subscription-based data models and digital-twin literature confirms that a cloud-hosted twin can act as a data broker (Thelen et al., 2022; Cao et al., 2024). DaaS requires end-to-end IoT streaming, a cloud analytics stack, and secure APIs that expose anonymised sensor feeds to subscribed parties (e.g., OEMs, insurers). Revenue here can stem from subscription fees, pay-per-analysis charges, and licensing of aggregated datasets, making the model largely volume-independent. DaaS sits at the far end of the servitisation ladder, representing the ultimate digital transformation of SMFs and delivering a high-margin, recurring-revenue stream.

Table 5: SMF-specific Business-Model: Data-as-a-Service

Core BMC pattern(s)	Key Resources: IoT sensor network; Revenue: subscription for data/analytics; Value: real-time insights for OEMs
Adoption driver(s)	Monetisation of sensor data reduces rework, warranty costs, and downtime; high-value OEMs seek fleet-wide performance analytics; space constraints make just-in-time logistics dependent on real-time data
Dominant industry sectors	Modular housing, high-rise construction, and critical industrial-plant upgrades that use predictive-maintenance services (power stations)
Required maturity	Very High (continuous IoT streaming, cloud analytics)

As of writing this study in 2026, it is understood that both Platform-Marketplace and DaaS are considered emerging and aspirational configurations based on technical feasibility rather than current mature market standards.

Mapping of Case Vignettes to BM Archetypes

The SMF4INFRA hyperloop pilot is a mobile prefabrication plant that manufactures pre-cast concrete tunnel segments for a Swiss hyperloop tunnel (SMF4INFRA, 2025).

The Flexible Field Factory (Martínez et al., 2013) is a truck-mounted production plant that manufactures prefabricated elements for low-rise residential housing in Spain. The two illustrative projects are examined by populating the SMF-specific Business-Model Canvas with the information extracted from their respective reports and checking the consistency of each case with the block patterns that define the four archetypes (Figure 3). The SMF4INFRA hyperloop pilot aligns mostly with the Platform-Marketplace archetype in terms of ambition. In this project, a DT-enabled marketplace can coordinate a large number of component suppliers through an ecosystem, and a transaction fee can be levied on each component sale. Because the plant also streams continuous IoT data for performance monitoring, a Data-as-a-Service element can be added on top of the primary platform model. In contrast, the Flexible Field Factory (Martínez et al., 2013) exemplifies mostly the Product-Plus archetype. It is assumed the plant is leased on a fixed-term basis and generates revenue through per-module assembly fees. No continuous sensor data are reported for this case, so the model remains at a low-to-medium digital-maturity level, and no secondary archetype is evident. Taken together, these two vignettes provide partial validation of the four-archetype typology.

The hyperloop pilot demonstrates how a high-digital-maturity SMF naturally adopts a platform-centric business model (with an optional DaaS add-on), whereas the Flexible Field Factory shows that even a low-to-medium-maturity SMF can achieve notable productivity gains by employing a Product-Plus configuration. This mapping exercise is only an early suggestion that the digital-maturity level of an SMF is a decisive factor in determining which business-model archetype best fits a given project and organization.

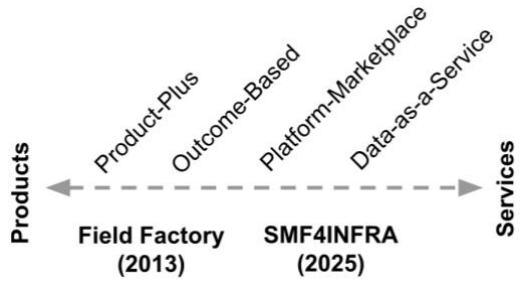


Figure 3: The case vignettes and archetypes mapped on the servitisation continuum

Proposed Business-Model Archetype Map

To help practising stakeholders to identify a suitable business-model archetype for their SMF plans, we propose a simple to use business-model archetype map (Figure 4). This map positions the envisioned SMF along three practical dimensions: the intraorganizational digital maturity, the interorganizational maturity and the servitisation continuum. Taken together, this leads to a three-step screening logic for early-stage feasibility assessment: stakeholders first populate the SMF-specific BMC shown in Table 1 with the project's actual resources, activities, and partners; then assess intra- and interorganizational digital maturity (low, medium, high, very high) based on BIM depth, sensor coverage, and API openness; and finally identify the matching archetype(s) described in Tables 2-5 (Product-Plus, Outcome-Based, Platform-Marketplace, and/or Data-as-a-Service).

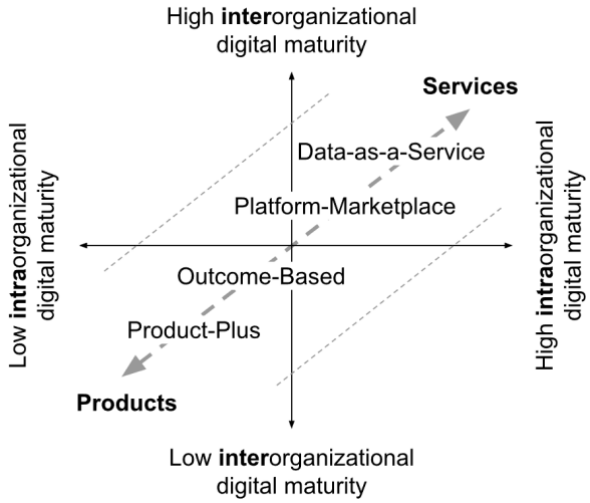


Figure 4: BM archetypes mapped along digital maturity levels

Discussion

This study makes three original contributions to the emerging field of research on SMFs for Construction 4.0. First, we create the first typology of SMF business models. Second, we illustrate and partially test these typologies with two independent case vignettes. Third, we synthesize these findings in a Business-Model Archetype Map (Figure 4), which helps stakeholders, such as contractors, investors, and public agencies, to align their digital capabilities (BIM, IoT, DT) with an appropriate revenue structure, and capitalize on the benefits documented in SMF life-cycle studies. This map aligns digital maturity in SMF contexts with the servitization continuum.

The four archetypes uncovered in the results correspond in principle to the servitisation continuum described by Grubic (2014) and Frank et al. (2019). The case mapping shows that a high-maturity DT (SMF4INFRA) enables a Platform-Marketplace revenue logic, whereas a low-maturity plant (Flexible Field Factory) relies on a Product-Plus structure. This pattern mirrors the servitisation continuum: firms progress from product-plus offerings to data-driven services as their digital capabilities mature. Furthermore, the sectoral distribution in Table 2 shows distinct servitisation drivers. In linear infrastructure, the dominant driver is performance-based risk allocation: governments and financiers attach penalties / bonuses to cost and time schedules, and CO_2 KPIs, which align naturally with the Outcome-Based archetype (Frank et al., 2019). The concurrent use of a Platform-Marketplace arises because a moving front requires real-time coordination of multiple suppliers, a rather classic ecosystem-centric servitisation pattern (Grubic, 2014). In modular housing, the primary driver is cost predictability; developers may prefer a fixed lease

that spreads the capital burden of the mobile plant (most closely related to Product-Plus).

But as data-collection technologies become cheaper, housing firms may add a Data-as-a-Service layer to extract value from sensor streams, reflecting the "data-driven servitisation" stage identified by Frank et al. (2019). Further, industrial plant construction can combine both drivers: high CAPEX pushes firms toward leasing, while simultaneously stringent certification deadlines create a need for outcome-based guarantees. Finally, the nascent adoption of data-centric models in urban retrofit projects is understood as primarily driven by space scarcity and the necessity for just-in-time logistics, which arguably only a real-time data service can reliably support.

Limitations and Future Research

The evidence base for this study is rather small, purposively selected, and heavily Euro-centric. Consequently, (i) emerging or niche SMF revenue models may be missing, (ii) sustainability conclusions rely on a limited set of LCA studies, (iii) no primary data (surveys, interviews, or field observations) were collected, (iv) cost-estimation and unit-economic analyses are absent, and (v) rapid advances in IoT, blockchain, AI and performance-contracting could generate business-model configurations beyond the four archetypes presented. To address these gaps, future research will certainly have to:

(1) expand the literature pool by adding recent scientific papers, industry white papers, technical reports and region-specific case studies (e.g., North-America, Asia-Pacific) to capture additional revenue models and test the universality of the presented typology;

(2) collect mixed-methods stakeholder data (surveys and semi-structured interviews) to validate the archetypes, map preferences (Wu et al., 2024), perceived barriers (Tan et al., 2025), and willingness-to-pay for services such as Data-as-a-Service;

(3) apply appropriate cost-estimation / unit-economic analysis methods (see Marcinkeviciūtė et al., 2025; Walzer et al., 2024a; Walzer et al., 2026) to quantify CAPEX, OPEX and ROI for each archetype;

(4) track firms longitudinally to document the evolution from Product-Plus to Data-as-a-Service and identify the critical capability-building turning points;

(5) conduct comparative, longitudinal case studies (for example high-rise urban vs. linear-infrastructure projects) to evaluate net-present value, CO_2 emissions, schedule adherence and risk exposure; and, lastly,

(6) refine this initial three-step screening logic into a calibrated multi-criteria decision-analysis (MCDA) tool that integrates financial, environmental and risk criteria from a standpoint of epistemological and methodological plurality (Walzer, 2025) that also considers social aspects, such as human-technology interactions (Johansen et al., 2024; Tan et al., 2026).

Conclusions and Implications

This study may be a first to analyse the business model aspects of SMFs across varying degrees of digital maturity. This was achieved by adapting the Business-Model Canvas to the construction context.

The analysis revealed a four-archetype continuum (Product-Plus, Outcome-Based, Platform-Marketplace, Data-as-a-Service) that aligns with the servitisation ladder, and is ordered by increasing digital-maturity requirements. The three-step screening logic operationalises this insight and two real-world pilots (SMF4INFRA hyperloop and the Flexible Field Factory) demonstrate that SMF projects can illustrate the identified archetypes fairly well, yet require further validation.

The typology itself yields concrete implications: For practitioners, the four archetypes make explicit which digital investments (depth of BIM models, IoT sensor layers, openness of APIs) are prerequisites for moving from low-value (Product-Plus) to high-value (Platform-Marketplace, Data-as-a-Service) revenue models, enabling staged technology roll-outs and, respectively, risk-adjusted financing. Policymakers can steer SMF deployments toward the most sustainable structures by embedding CO_2-performance KPIs in outcome-based contracts and utilising, for example, tax-credit incentives. For scholars, the work extends servitisation theory to a mobile, construction-site context and establishes a baseline for longitudinal research on business-model evolution in Construction 4.0 environments. Moreover, this work provides a taxonomy that can be used as a coding scheme in future empirical studies of digital-enabled construction ventures.

In conclusion, SMFs for Construction hold the promise of reconciling the speed and precision of factory production with the flexibility demanded by on-site construction. By clarifying how value can be captured, this paper helps advance SMFs from a purely technical curiosity into a potentially commercially viable, sustainability-enhancing construction paradigm. The archetypes and the screening logic presented herein lay the groundwork for both immediate industry adoption and a future agenda of scholarly inquiry.

Acknowledgments

We are thankful for the constructive and inspiring discussions within the SMF4INFRA Consortium, in which this work has been developed. A special gratitude is directed towards Prof. Guillaume Habert of ETH Zurich, who enabled the collaboration of the co-authors. This study has received funding from the Swiss National Science Foundation under grant agreement No. 204852 and the Autonomous Province of Bolzano/Bozen – South Tyrol. We further thank the Institute of Virtual Design and Construction (VDC) at FHNW for adopting this work within their didactic framework, allowing it to inform the coursework of the students in a recent semester project.

References

Cao, J., Čustović, I., Soman, R., & Hall, D. (2024). Reinforcement learning for smart mobile factory operation in linear infrastructure projects. In Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 41, pp. 738-744). IAARC Publications. https://doi.org/10.22260/ISARC2024/0096

Čustović, I., Cao, J., Hall, D. M., & Wamelink, H. (2025). Enhancing situation awareness of construction managers using human-centered digital twins. Advanced Engineering Informatics, 67, 103557. https://doi.org/10.1016/j.aei.2025.103557

Čustović, I., Cao, J., & Hall, D. M. (2023a). Cloud manufacturing for industrialized construction: Opportunities and challenges for a new manufacturing model. Journal of Infrastructure Intelligence and Resilience, 2(1), 100027. https://doi.org/10.1016/j.iintel.2023.100027

Čustović, I., Soman, R. K., Pauwels, P., & Hall, D. M. (2023b). Towards integrated digital twins for construction and manufacturing. In Proceedings of the 2023 European Conference on Computing in Construction and the 40th International CIB W78 (Vol. 4). European Council on Computing in Construction. https://doi.org/10.35490/EC3.2023.256

Dallasega, P., Kaushal, I., Revolti, A., & Miori, N. (2024). Life-cycle analysis for the concept design of a smart mobile factory (SMF) for infrastructure construction projects. In M. Thürer, R. Riedel, G. von Cieminski, & D. Romero (Eds.), Advances in Production Management Systems (pp. xx-xx). Springer. https://doi.org/10.1007/978-3-031-71637-9_3

European Commission: Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs. (2025). Market potential of offsite construction for housing supply : report. Publications Office of the European Union. https://data.europa.eu/doi/10.2873/3569150

Frank, A. G., Mendes, G. H. S., Ayala, N. F., & Ghezzi, A. (2019). Servitization and Industry 4.0 convergence in the digital transformation of product firms: A business-model innovation perspective. Technological Forecasting & Social Change, 141, 341-351. https://doi.org/10.1016/j.techfore.2019.01.014

Grubic, T. (2014). Servitization and remote monitoring technology: A literature review and research agenda. Journal of Manufacturing Technology Management, 25(1), 100-124. https://doi.org/10.1108/JMTM-05-2012-0056

Hamza, M., Shahid, S., Bin Hainin, M. R., & Nashwan, M. S. (2019). Construction labour productivity: review of factors identified. International Journal of Construction Management, 22(3), 413–425. https://doi.org/10.1080/15623599.2019.1627503

Hevner, A. R., & Chatterjee, S. (2004). Design science research in information systems. In Design research in information systems (pp. 1-26). Springer. https://doi.org/10.1007/978-1-4419-5653-8

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. Qualitative Health Research, 15(9), 1277-1288. https://doi.org/10.1177/1049732305276687

Jäkel, J.–I., Fischerkeller, F., Oberhoff, T., & Klemt-Albert, K. (2024). Development of a maturity model for the digital transformation of companies in the context of Construction Industry 4.0. ITcon, 29, 778-809. https://doi.org/10.36680/j.itcon.2024.034

Johansson, P. E., Bruch, J., Chirumalla, K., Osterman, C., & Stålberg, L. (2024). Integrating advanced digital technologies in existing lean-based production systems: Analysis of paradoxes, imbalances and management strategies. International Journal of Operations & Production Management, 44(6), 1158-1191. https://doi.org/10.1108/IJOPM-05-2023-0434

Johansen, S.S., A. Burden, E. Schneiders, and A.N. Walzer (2024). “Designing for People in Human-Robot Collaboration”. In: Interaction Design and Architecture(s) Journal - IxD&A 61. Preface, pp. 5–10. doi: 10.55612/s-5002-061-001psi.

Karmakar, A., & Delhi, V. S. K. (2021). Construction 4.0: What we know and where we are headed? ITcon, 26, 526-545. https://doi.org/10.36680/j.itcon.2021.028

Kohtamäki, M., Parida, V., Oghazi, P., & Gebauer, H. (2019). Digital servitization business models in ecosystems: A theory of the firm. Journal of Business Research, 104, 380-392. https://doi.org/10.1016/j.jbusres.2019.06.027

Kuechler, W., & Vaishnavi, K. (2008). Design science research methods and patterns: Innovating information and communication technology. ACM SIGMIS Database, 39(1), 28-45.

Li, Y., Maxwell, D. W., & Moehler, R. (2025). Evolving business models in the construction industry: A comprehensive review and analysis. Production Planning & Control, 37(2), 109–127. https://doi.org/10.1080/09537287.2025.2463370

Ma, J., Cao, J., Benedetti, L., Revolti, A., & Zea Escamilla, E. (2025). Environmental and economic optimisation of relocation strategies for mobile prefabrication factories in infrastructure projects. Developments in the Built Environment, 24, 100744. https://doi.org/10.1016/j.dibe.2025.100744

Marcinkeviciūtė, D., Schildknecht, L., Huber, M., Pancera, M., Gschwind, J., Badertscher, M., & Crevillen, J. (2025). Simplification and enrichment of IFC models for cost estimation. In M. Hajdu, M. Skibniewski, & Z. Turk (Eds.), Proceedings of the 14th Creative Construction Conference (CCC 2025),

(pp. 154-163). International Association for Automation and Robotics in Construction (IAARC). https://doi.org/10.22260/CCC2025/0021

Martínez, S., Jardón, A., Víctores, J. G., & Balaguer, C. (2013). Flexible field factory for construction industry. Assembly Automation, 33(2), 175-183. https://doi.org/10.1108/01445151311306708

Matt, D. T., Rauch, E., & Dallasega, P. (2015). Trends towards distributed manufacturing systems and modern forms for their design. Procedia CIRP, 33, 185-190. https://doi.org/10.1016/j.procir.2015.06.034

Nahmad Vazquez, A., Garivani, S., & Dackiw, J.-N. (2024).Decentralized, data-informed, robotic-based digitaltimber micro-factories. Construction Robotics, 8(2), 24. https://doi.org/10.1007/s41693-024-00132-y

Osterwalder, A., & Pigneur, Y. (2010). Business Model Generation. Wiley.

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45-77.

Park, K. T., Nam, Y. W., Lee, H. S., Im, S. J., Noh, S. D., Son, J. Y., & Kim, H. (2019). Design and implementation of a digital-twin application for a connected micro smart factory. International Journal of Computer Integrated Manufacturing, 32(6), 596-614. https://doi.org/10.1080/0951192X.2019.1599439

Revolti, A., Benedetti, L., Kaushal, I., & Dallasega, P. (2025). Empirical validation of BIM-to-digital twin transformation for smart mobile factories: A case study. In Proceedings of the 2025 European Conference on Computing in Construction (Vol. 6). European Council on Computing in Construction. https://doi.org/10.35490/EC3.2025.238

Sawhney, A., Riley, M., & Irizarry, J. (Eds.). (2020). Construction 4.0: An Innovation Platform for the Built Environment (1st ed.). Routledge. https://doi.org/10.1201/9780429398100

SMF4INFRA. (2025). Smart Mobile Factory for Infrastructure Projects (SMF4INFRA). Retrieved from https://www.smf4infra.net/overview

Tan, T., Graser, K., Walzer, A. N., Ng, M. S., & Hall, D. M. (2025). Exploring the implementation of digital fabrication in construction: Combining system approaches and a revelatory case study from Switzerland. Journal of Construction Engineering and Management, 151(12), 05025016. https://doi.org/10.1061/JCEMD4.COENG-16680

Tan, T., Liu, X., Walzer, A.N. et al. Human–robot interaction design with augmented reality in construction: A systematic review of frameworks, applications, and future directions. ENG. Manag. (2026). https://doi.org/10.1007/s42524-026-5039-0

Thelen, A., Zhang, X., Fink, O., Lu, Y., Ghosh, S., Youn, B. D., Todd, M. D., Mahadevan, S., Hu, C., & Hu, Z. (2022). A comprehensive review of digital twin — part 1: Modeling and twinning enabling technologies. Structural and Multidisciplinary Optimization, 65(12), 354. https://doi.org/10.1007/s00158-022-03425-4

Tripathi, A., Kaushal, I., Revolti, A., Benedetti, L., & Dallasega, P. (2025). Digital twin-driven quality control in smart mobile factories: A framework in precast manufacturing. Procedia CIRP, 134, 801-806. https://doi.org/10.1016/j.procir.2025.02.194

UNECE. (2025). Value of the construction industry as a share of the gross domestic product (GDP) in selected European countries from 2005 to 2024 [Graph]. In Statista. Retrieved January 12, 2026, from https://www.statista.com/statistics/1600838/gdp-share-of-the-construction-industry-in-major-european-countries

Wagner, H. J., Alvarez, M., Kyjanek, O., Bhiri, Z., & Menges, A. (2020). Flexible and transportable robotic timber construction platform – TIM. Automation in Construction, 120, 103400. https://doi.org/10.1016/j.autcon.2020.103400

Walzer, A. N. (2025). Assessing multi-level uncertainties in construction robotics entrepreneurship by means of methodological pluralism (Doctoral dissertation). ETH Zurich. https://doi.org/10.3929/ethz-b-000705101

Walzer, A. N., Kozlova, M., Mohite, A., Yeomans, J. S., & Hall D. M. (2026). How to scale a 3D concrete printing facility? A stochastic decision-support framework for production investment. Journal of Information Technology in Construction (ITcon), 31, 180-200. https://doi.org/10.36680/j.itcon.2026.008

Walzer, A. N., Kozlova, M., & Yeomans, J. S. (2024a). Printing homes: Unit cost estimation for additive manufacturing in construction. In M. Kozlova & J. S. Yeomans (Eds.), Sensitivity Analysis for Business, Technology, and Policymaking Made Easy with Simulation Decomposition (pp. 116-141). Routledge. https://doi.org/10.4324/9781003453789-8

Walzer, A. N., Tan, T., Graser, K., & Hall, D. M. (2024b). Bug or feature? Institutional misalignments between construction technology and venture capital. Construction Management and Economics, 43(2), 130-152. https://doi.org/10.1080/01446193.2024.2401818

Wu, S., Walzer, A. N., Kahlert, A., Dillenburger, B., & Hall, D. M. (2024). Understanding stakeholders' intention to use construction robots: A fuzzy-set qualitative comparative analysis. Construction Robotics, 8(5), Article 5. https://doi.org/10.1007/s41693-024-00115-z

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

EXPERIMENTING WITH DIGITAL TWINS: A LEGO®-BASED MODULAR TESTBED FOR SYSTEMATIC EVALUATION

Irfan Čustović [1], Emilio Murillo Sierra [2], Ranjith K. Soman [2], Daniel M. Hall [1]
[1] Department of Management in the Built Environment, Delft University of Technology, The Netherlands
[2] Department Materials, Mechanics, Management & Design, Delft University of Technology, The Netherlands

Abstract

Digital twins (DTs) promise improved decision-making under project uncertainty, but real-world implementations are costly and hard to compare. As DT research moves toward tighter feedback loops and greater autonomy, affordable and controllable environments for evaluating such DTs are lacking. This paper presents a modular DT testbed that combines LEGO®-based physical components with Raspberry Pi computers to support controlled variation of DT scenarios, feedback loops and automation levels. Two studies conducted in the testbed illustrate how DTs can be systematically evaluated across computational processes and physical execution. The testbed facilitates cost-effective and repeatable future research on autonomous DTs for construction management.

Introduction

DTs for construction management are receiving significant attention because project execution remains inefficient (Kookalani *et al.*, 2026). The basic promise is straightforward: link on-site conditions with a digital representation that supports situation awareness and more informed control. DT research has moved from isolated twins for a single building in operation to integrated systems intended to support complex project-level decisions (Whyte *et al.*, 2025). This shift matters because project disruptions are not isolated events; they emerge from complex interactions across trades, zones and time.

Current DT research follows two intertwined trajectories: speeding up feedback loops and increasing autonomy. Faster, more frequent data integration is motivated by the need for timely responses to disruptions and the desire for preventive action (Boje *et al.*, 2020). Accordingly, DT applications span timeframes from weekly coordination to more continuous monitoring under unstable conditions (Chacón *et al.*, 2024; Soman *et al.*, 2025). At the same time, DTs are envisioned to evolve from supporting data analysis in routine situations toward autonomous execution of non-routine actions, often realized through mobile on-site robotics (de Soto and Skibniewski, 2020; Agrawal *et al.*, 2023). However, accelerating feedback loops and automation levels do more than improve responsiveness; they reshape how work and decisions are distributed between humans and digital systems (Parasuraman, Sheridan and Wickens, 2000).

The value of DT implementations depends on alignment between automation level, organizational context, and intended use case, rather than on technical sophistication alone (Čustović *et al.*, 2025). Without such alignment and early user involvement, DTs risk underperforming when exposed to real project variability.

This situation creates a methodological bottleneck: transitioning DT concepts into operational prototypes and systematically studying their effects remain difficult. The inherent variability and the complex system-of-systems that characterize construction projects make it difficult to isolate the effects of specific DT features (Zhu *et al.*, 2024). It becomes unclear whether improvements are driven by the DT itself, by users' responses to it, or by coincidental project dynamics. While real-world DT studies provide valuable insights, they largely focus on monitoring and low-autonomy use cases and often depend on large, well-funded consortia (Brell-Çokcan and Schmitt, 2024; Soman *et al.*, 2025). As DT research shifts toward tighter feedback loops and greater autonomy, the field lacks affordable, controllable and repeatable environments that allow DT design choices to be tested and compared under realistic but manageable conditions.

To address this, we introduce a modular DT testbed that links digital intelligence with physical sensing and actuation in a miniaturized construction environment. The testbed builds on a LEGO®-based haptic simulation approach combined with Raspberry Pi single-board computers (SBCs), enabling controlled experimentation with bi-directional DTs and management strategies without access to full-scale construction sites. To support reuse and further development, the source code and digital models are available as supplementary data to this article.

Point of Departure

Digital Twins for Construction Management

DTs are virtual representations of physical objects, processes, or systems that are linked to the physical domain through bi-directional communication and are instantiated to perform specific operational or analytical functions (Čustović *et al.*, 2023; AlBalkhy *et al.*, 2024). The literature characterizes DTs as information systems that integrate heterogeneous data in real time, provide advanced analysis, predictive models for early disruption detection and automated recommendations (Sacks *et al.*, 2020; Saif, Doukari and Kassem, 2026).

Implementations and empirical testing are essential to demonstrate the functionality of a digital artifact beyond conceptual design (Tommelein, 2020). In construction, the dynamic nature of sites and the one-off character of projects make such testing significantly more difficult than in industries with stable and repeatable processes, such as manufacturing. Consequently, most field-based DT studies emphasize data acquisition and status visibility rather than active control. Virtual-to-physical interaction is generally realized by human managers, with DTs providing structured, timely input that informs planning adjustments and operational responses within existing supervisory workflows (olde Scholtenhuis and Soman, 2026). Even when DTs are designed to operate with limited autonomy, practical challenges persist.

Challenges with Digital Twin Implementations

Establishing fast feedback loops and high levels of DT autonomy in construction is difficult due to changing site conditions and organizational constraints (Saif, RazaviAlavi and Kassem, 2024). Chacón *et al.* (2024) show that technical feasibility alone is insufficient: although data collection and system deployment were successful, the insights generated were not adopted in site management procedures. Similarly, Zheng et al. (2025) report that data ownership concerns and commercial policies limit the application of DT systems in complex project settings involving multiple subcontractors. Soman *et al.* (2025) develop a comprehensive DT that integrates multiple production data streams to support weekly project meetings. Yet, despite the significant resources of a consortium setting, the study does not assess how the DT influenced managerial decisions or project performance, nor does it execute any physical work. Even when the DTs are deployed onsite, controlling the many interdependent variables to ensure research validity is very difficult.

In addition to methodological and organizational challenges, financial constraints strongly shape DT research. Deep process integration, real-time monitoring, and autonomous action typically require substantial investments (AlBalkhy *et al.*, 2024). For instance, the *Internet of Construction* project enabled controlled investigation by setting up a dedicated construction site with a simple load-bearing structure, but this was only possible through substantial financial and organizational resources of a large consortium (Brell-Çokcan and Schmitt, 2024). Higher levels of DT autonomy depend on cyber-physical systems capable of executing decisions, i.e., construction robotics. Although these have been widely studied, the significant capital investment required for robotic hardware has led to slow adoption (Walzer, 2025). Together, these examples illustrate the substantial practical constraints that limit testing of operational DT systems intended to support construction management.

Laboratory Environments for Construction Management Digitalization

Experiments are a valid method for evaluating the effectiveness of information system designs because they allow researchers to test defined design choices under controlled conditions (Hevner, March and Park, 2008). Laboratory testbeds provide a practical setting for such experiments. They enable researchers to specify the system design and boundary conditions, allowing selected parameters to be systematically controlled and varied across trials. Construction research has long relied on lab-like setups to address the scale, messiness, and risk of real projects, where controlled trials are hard to stage (Herbsman, 1986). In construction research, serious games have become a popular experimental tool, especially for studying lean construction practices (Bhatnagar *et al.*, 2023; Martinez *et al.*, 2023).

For DTs, however, the experimental requirements differ. DTs depend on a tight coupling between physical assets and virtual representations. This duality means that computer-only simulations or offline board games often cannot test key DT functions such as sensing, feedback, and control across physical and virtual layers. Laboratory testbeds for DT research, therefore, need to include both layers while still allowing researchers to control conditions and observe outcomes.

Only a few studies have built such DT-oriented laboratory setups, and they often address narrow technical goals. Jiang *et al.* (2022) use a 6-axis robot arm to simulate assembly processes of prefabricated modules. However, such robotic testbeds can be expensive, safety-critical, and difficult to reproduce when documentation is limited. Hasan *et al.* (2022) present a DT of a laboratory tower crane with a strong focus on physical realism and crane control. Yet, the setup does not connect DT capabilities to construction management decisions or project scenarios. Li *et al.* (2025) develop a DT testbed for status monitoring using computer vision. However, it focuses on monitoring and does not include actuators or user interfaces for instructing workers or coordinating actions. A further weakness across these efforts is their monolithic, purpose-built nature, which restricts replication and makes reconfiguration to other DT use cases difficult, especially when source code and reusable modules are not provided.

Taken together, the literature provides limited experimental means to test how DTs, when combined with robotics as autonomous hardware, affect construction management outcomes. Existing work tends to study DTs without meaningful physical actuation, or robots as isolated machines without management-related use cases. We address this gap by proposing a low-cost, scalable laboratory environment that integrates DT features from both the physical and virtual worlds to support controlled experiments in DT-enabled construction management.

Design Requirements

The development of the DT testbed is based on translating insights and challenges from prior DT studies into concrete system features. A primary requirement concerns representational breadth: the environment must support multiple construction use cases that can benefit from DTs, including project monitoring and control, dynamic replanning under uncertainty and workflow

analysis for performance evaluation. A second requirement addresses the virtual-to-physical twinning, requiring the inclusion of actuation for at least one mechanized site process, such as material lifting, to allow digital decisions to be evaluated against physical outcomes. A third requirement concerns decision authority, mandating support for both human-in-the-loop supervision and autonomous action execution in routine situations with explicit logging of decision rationales. In addition to these use-case-driven requirements, the testbed must realize system-level DT features including real-time communication, reliable data logging and querying, and adaptability across construction scenarios and scripting languages. We address these requirements through a combination of LEGO®-based haptic simulation with Raspberry Pi computers.

LEGO®-based mock-ups demonstrate effectiveness in modeling construction management tasks and in evaluating the impact of digital tools such as Building Information Modeling (BIM) or virtual reality (VR) (Sacks, Esquenazi and Goldin, 2007). Extending LEGO® bricks with programmable LEGO® motors enables actuation and creates closed-loop cyber-physical systems without extensive hardware development (Gawthrop and McGookin, 2004). The LEGO® Spike platform offers modular, preconfigured components that support rapid prototyping and system expansion.

Raspberry Pi SBCs provide a cost-effective yet performative computational backbone and are widely adopted as prototyping tools in scientific research across varying degrees of complexity (Mathe *et al.*, 2024). Their Linux-based ecosystem and compatibility with additional electronics hardware via GPIO interfaces and the Build HAT support integrated implementation of sensing, actuation, and computational functions.

Digital Twin Testbed

The DT testbed consists of (i) a LEGO®-based gantry robot and (ii) interactive workstations for human-DT interaction. Figure 1 presents the integrated setup.

Gantry Robot

The gantry robot serves as a programmable transport system for LEGO® elements, autonomously managing material flows, logging operational data and exchanging information with external services.

The mechanical structure consists of a crane track supported by four pillars on each side, carrying a crane bridge that moves longitudinally on four wheels. Mounted centrally on the bridge is a Raspberry Pi 5 SBC, fixed in position to ensure stable computation during operation. A LEGO® Build HAT is attached to the Raspberry Pi, enabling direct control of four motors that together define the robot's four degrees of freedom (DoF). Horizontal motion is realized through a trolley unit that travels across the crane bridge and carries the gripping mechanism. This gripper is guided by vertical rods and operates through a circular opening-and-closing motion. Except for two components, all mechanical elements are made from original LEGO® parts. A custom-designed container houses the Raspberry Pi computer. In addition, the gripper claws are 3D printed and shape-optimized to improve gripping reliability.

Actuation is achieved using LEGO® SPIKE motors, with each motor controlling a single DoF independently (see Figure 2). Calibration happens by moving each motor toward a fixed mechanical endpoint and detecting stalling. This sensor-free approach simplifies system design but implicitly assumes consistent mechanical behavior, which becomes a critical assumption as operational complexity increases.

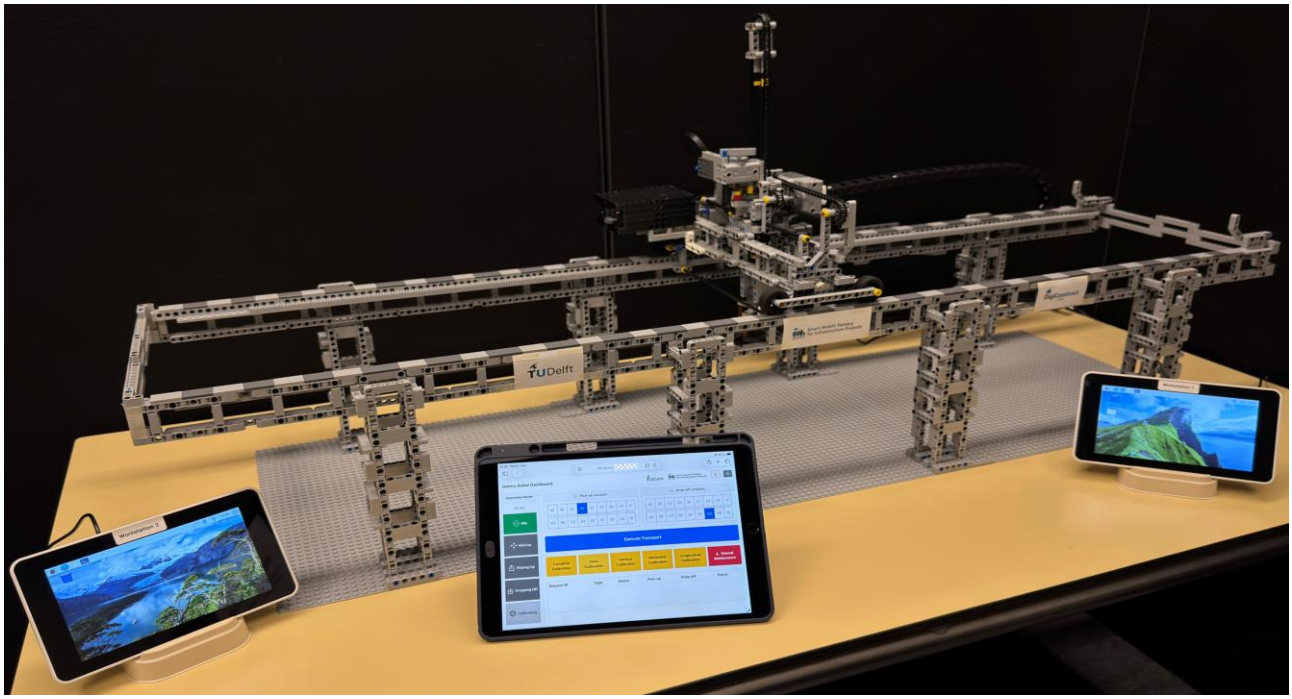


Figure 1: Overview of the modular digital twin testbed, consisting of the gantry robot and two interactive workstations. The tablet (optional) displays the gantry robot's web-based UI served via a FastAPI endpoint that is accessible from devices on the WLAN.

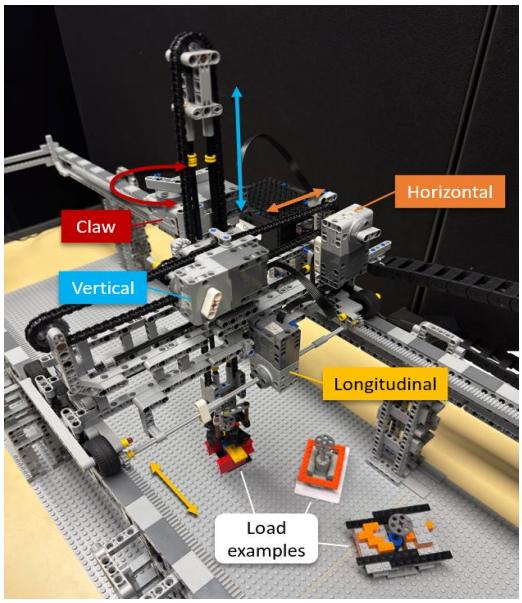


Figure 2: Gantry Bridge with 4 motors and 3 different examples of loads that can be transported.

The digital layer runs on a Linux distribution on the Raspberry Pi and is structured using Docker containers to enable adaptability. This digital system separates responsibilities into service, motor control, and storage containers. The service container functions as the system integrator, linking physical actuation, data storage, and order processing. It runs a FastAPI instance that exposes REST endpoints for querying system state and submitting transport commands, while a dedicated service script manages the order queue and database communication. Motor execution is delegated to a dedicated container that interfaces directly with the hardware through the BuildHAT Python library. Data storage is handled by a third container that supports interchangeable database systems for long-term storage of movements, transport requests and operation modes. Both GraphDB and MongoDB containers were tested in this setup.

User interaction is provided through a web-based interface served directly by the FastAPI framework using Jinja templates. The interface retrieves the gantry robot's operation mode via REST calls at configurable polling intervals and allows users to trigger calibration and transport commands by sending JSON payloads. This approach enables wireless control through common devices such as smartphones and tablets (Figure 1). While polling is sufficient for the demonstrated use cases, FastAPI also supports WebSockets.

Interactive Workstations

The interactive workstations represent the user-facing side of the DT testbed and operationalize direct human interaction within the distributed DT architecture. A workstation operates as an independent edge node, allowing for local data processing and storage. This configuration enables shifting interaction logic away from the actuated system core and toward decentralized components that remain tightly coupled via network communication.

Each workstation consists of a Raspberry Pi computer paired with a 7″ touchscreen. The touchscreen serves as the sole interaction modality, unifying user input and system feedback without reliance on external devices. By omitting actuators, the workstation maintains a clear functional boundary between interaction and physical execution within the DT architecture. At the same time, hardware extensibility remains available through Raspberry Pi's GPIO interface. For example, an RFID reader was connected in the testbed to explore contactless identification as a means of reducing repetitive user input and lowering the likelihood of interaction errors.

The digital architecture follows the same principles as the gantry robot by running Docker container services on a Linux platform (see Table 1). A minimal container setup is sufficient, consisting of a FastAPI-based service container that manages the touchscreen interface and communication logic, and a database container for persistent storage. The interface is generated using HTML Jinja templates and displayed directly on the touchscreen, reinforcing consistency across the gantry robot and the interactive workstations. Data exchange between the robot and workstations is implemented via REST-based communication, with WebSocket support available but not required for the demonstrated scenarios.

Digital Twin Experiments Performed with the Testbed

The following subsections describe how the testbed was used in other studies, illustrating its versatility and applicability for evaluating DT-focused research. Specific details on the experiment variables and evaluation criteria will be provided in separate publications. At the time of writing, they have yet to be published.

Agents in Construction (AiC) Ontology Verification

Study 1 used the testbed to functionally verify the *Agents in Construction (AiC)* ontology as an operational, machine-readable representation for real-time site coordination (Čustović, Cao and Soman, 2025). In this study, the AiC ontology served as the shared model for coordination-relevant facts. Accordingly, the evaluation examined whether information expressed using AiC concepts and relations could coordinate a construction process digitally across repeated runs.

Verification was conducted in the testbed because DT experiments on active construction sites often depend on enabling conditions that are difficult to ensure simultaneously, including integrated sensing and actuation, continuous capture of worker interactions, and uninterrupted wireless connectivity. The modular testbed provided these conditions in a controllable environment.

The coordination scenario was instantiated as a LEGO®-based simplified infrastructure assembly (see Figure 3). The elements of the infrastructure were decomposed into batches of LEGO® bricks within the gantry's carrying capacity and the workspace was segmented into multiple work zones plus a storage zone. Within this setup, the gantry robot operated as a transport agent moving

Table 1: Overview of Digital Twin Features of the Gantry Robot and the Interactive Workstations.

DT Feature	Gantry Robot	Interactive Workstation
User Interface / Visualization	Web-based UI (HTML / CSS / JavaScript)	Web-based UI (HTML / CSS / JavaScript); 7-inch mounted touchscreen UI
Computing Unit	Raspberry Pi 5, 8 GB RAM, BuildHAT	Raspberry Pi 5, 8 GB RAM
Application / Service Layer	Linux distribution; containerized with Docker	
Network Interface	REST API (FastAPI endpoints)	
Data Management	Time-series DB (e.g., MongoDB); graph DB (e.g., GraphDB); relational DB (e.g., PostgreSQL)	
Sensing	Internal motor state and telemetry from LEGO® SPIKE Prime / Powered Up Motors	RFID reader via GPIO interface
Actuation	LEGO® Spike motor actuation; 4 degrees of freedom (4-DoF)	-

materials between storage and site locations. The interactive workstations served as DTs for the human agents, displaying and enabling updates to information during execution. Four participants played different roles: two construction workers were building the LEGO® structure and used interactive workstations to request material delivery from the crane; a manager responsible for task assignment; and a packer assisting the gantry robot by positioning the pieces in the pickup area. In the unlikely event that the gantry robot fails to grip or release a load, this fourth operator serves as a human backup (DT has routine autonomy).

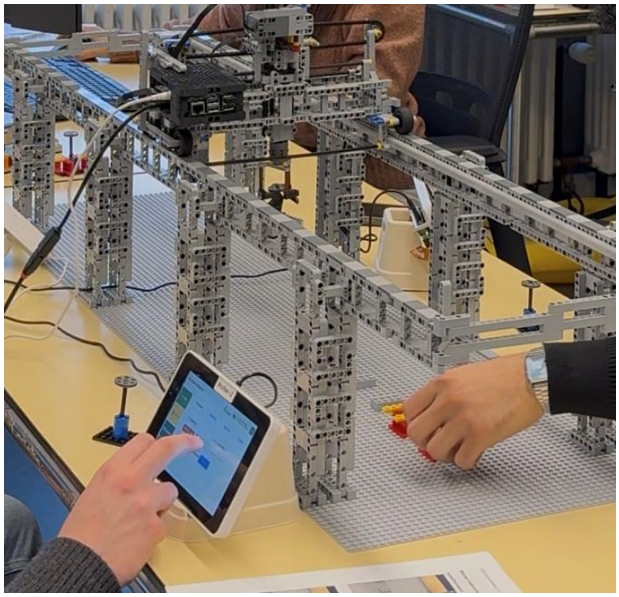


Figure 3: Scene from the AiC ontology verification experiment conducted using the DT testbed.

To perform the evaluation and monitor agents and agent-to-agent communication, the AiC ontology was implemented as the data model, and GraphDB was deployed on all Raspberry Pi computers. Participants viewed and updated operation modes, process lists, material and equipment usage, and locations as tasks advanced, while the manager exchanged process assignments with the workstations via REST calls; these interactions exercised coordination loops repeatedly and were captured for subsequent analysis.

The modular testbed enabled real-time investigation of DT-based site coordination that relied on the AiC ontology. It was possible to repeat the same coordination scenario and vary relevant variables (worker availability or the task execution sequence) while logging time-stamped coordination messages.

Serious Game with an Autonomous Site Crane

Study 2 examined how humans work when selected construction decisions and actions are taken autonomously by a DT. The goal was to simulate a site scenario in which workers collaborate and must coordinate around a shared, autonomous robotic resource.

Running comparable studies on live construction sites is difficult. Conditions on site vary in ways that are hard to control; the number of robots that can operate autonomously in changing environments is limited, and cost and safety constraints are high. Nevertheless, the development of such autonomous systems needs systematic evidence on how they shape coordination and work practices. We therefore used the DT testbed as a controllable, observable environment that supports repeatable autonomous execution together with structured human-DT interaction, allowing us to capture effects that would be hard to isolate on site.

These needs were operationalized in the DT testbed by using the gantry robot as an autonomous site resource. The robot followed scripts corresponding to predefined autonomy levels; in addition to the routine autonomous transport examined in Study 1, Study 2 incorporated shared-resource allocation decisions, which are common in construction (see Figure 4): the gantry robot DT decided which participant would be served next, requiring participants to plan and coordinate around a single robotic bottleneck. The testbed also enabled systematic variation in decision autonomy by altering the information available to the robot's decision scripts. In the routine autonomy condition, the robot relied on limited state data; in the non-routine autonomy condition, the gantry robot

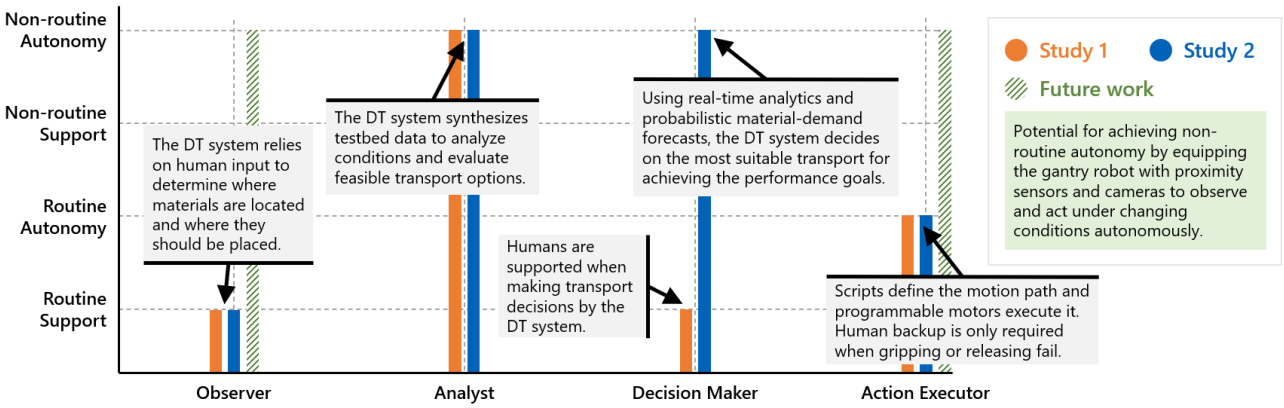


Figure 4: Different DT autonomy levels that were examined in the testbed mapped to the framework by Agrawal et al. (2023).

DT used real-time progress-data updates and probabilistic material-demand estimates. This design isolated the effects of autonomy and information availability while holding the physical system, task structure, and interaction channels constant.

Discussion

The laboratory testbed advances construction automation research by addressing a persistent gap at the decision-making and actuation layer of DTs. While many DT studies articulate data acquisition, integration and visualization in detail, their realization lacks fast feedback loops and high degrees of DT autonomy (olde Scholtenhuis and Soman, 2026). The programmable gantry robot introduces a tangible site actor that autonomously transports materials in routine situations while exposing operation conditions and commands through defined interfaces. The gantry robot supports systematic experimentation across different levels of autonomous analysis, decision-making, and action execution (see Figure 4). Combined with the interactive workstations, the modular DT testbed allows for controlled investigation of how human actors interact with and respond to automated execution. This aligns with the categorization of DTs that highlights human-system interaction and autonomy as defining dimensions in construction contexts (Agrawal *et al.*, 2023). Rather than assuming idealized coordination between human and digital actors, the testbed enables examination of frictions, delays, and behavioral adaptations that emerge under real-time conditions with immediate feedback.

At a broader level, the testbed supports scenario-based experimentation in which construction management conditions are deliberately constructed and compared (Li *et al.*, 2025). Real-time representation of site-like activities, combined with controllable execution, enables repeatable experiments that are rarely feasible in real construction projects. This capacity supports theory-driven inquiry into DT features, rather than one-off demonstrations of technological feasibility.

The contribution is further strengthened through the full-stack implementation of DT building blocks (AlBalkhy *et al.*, 2024; Saif, Doukari, and Kassem, 2026), spanning sensing, communication, storage, analytics, visualization, and action implementation, using embedded computing and containerized services. The reliance on low-cost hardware and open-source or free software addresses financial barriers that constrain DT experimentation and replication across institutions (Zhu *et al.*, 2024). The distributed interactive workstations extend this contribution by enabling a structured investigation of how information is distributed, interpreted, and acted upon by human participants, highlighting organizational and cognitive aspects that are often underrepresented in DT research (Čustović *et al.*, 2025). The modular hardware and software architecture positions the testbed as an evolving research platform that can be reconfigured or expanded with additional sensors and actuators. Docker containers can be replaced or redeployed across alternative hardware with minimal architectural change.

The testbed is intentionally simplified and therefore differs in scale and context from real construction sites, so the transferability of results must be explicitly assessed in follow-up studies. While the testbed provides the technological backbone for serious games to be conducted within it, the transferability of the resulting findings to real-world contexts depends on how well reality is modeled through game elements and mechanics (Freese and Bekebrede, 2025). Besides, the gantry robot is a material-movement actor rather than a building or assembly robot. It also inherits fundamental workspace constraints, limited DoFs, inability to reach around obstacles, and strong collision-avoidance requirements, which restrict the representation of complex access conditions or overhangs. Data quality further depends on human input at interactive workstations, which creates the risk of missing or inconsistent updates; audio hardware for voice control or automated sensing (e.g., computer vision) could reduce this dependency. The gantry robot cannot sense whether it has properly gripped a load, so it only supports routine autonomy for actions that require human backup. Precision and timing fidelity are limited by LEGO® motor accuracy, sensor-free calibration drift, and latency from containerization and REST-based communication, making the setup better suited for coordination and decision-logic studies than for high-precision or hard real-time control validation.

Conclusions

This paper introduced a scalable testbed for experimenting with DT capabilities in construction management scenarios. The testbed combines a LEGO®-based haptic simulation with programmable motors, Raspberry Pi 5 SBCs and touchscreen workstations, enabling a digital-physical feedback loops and human-DT interaction. The testbed can be reconfigured to represent different DT scenarios by changing the physical layout, the digital functions and the number of stations involved.

The testbed offers value for academia and industry. For researchers, it provides a replicable, configurable environment for testing a wide range of DT hypotheses under controlled conditions. Its design deliberately relies on consumer-ready hardware and software, reducing costs and setup effort and supporting adoption by other research groups. For the industry, it provides a concrete demonstration environment to communicate DT concepts to practitioners and to trial DT features against company-specific use cases before large financial commitments.

As an initial working prototype, the testbed provides a solid basis for extensions that improve realism and broaden the experimental scope. Key priorities are enhanced sensing for reliable state capture (e.g., precise position tracking with proximity sensors and RFID-based element detection), camera vision (fixed or gantry-mounted), more robust messaging, such as MQTT, to support scalable device communication, and wearable hardware that goes beyond an interactive workstation. Further avenues include integrating additional programmable controllers or components from the LEGO® SMART Play™ system (The LEGO Group, 2026) to represent other task types. Finally, future work can extend the setup with local AI models running on the gantry robot or the interactive workstations and develop educational programs around the testbed.

Supplementary Data

The source code for the gantry robot alongside the digital models of the LEGO® structure and the custom 3D-printed elements can be found in this repository:

https://github.com/DigiConstructLab-TU-Delft/Digital-Twin-Testbed/

Acknowledgments

This work is carried out within the project "Smart Mobile Factory for Infrastructure Projects (SMF4INFRA)" and supported by the Swiss National Science Foundation [grant no. 204852]. The authors would like to thank Ahmed Shalabi, C. Koray Bingöl and Lukas Böhm for their participation during the first demonstration of the testbed in September 2025. LEGO is a trademark of the LEGO Group of companies, which does not sponsor, authorize or endorse this research. Raspberry Pi is a trademark of Raspberry Pi Ltd.

References

Agrawal, A., Thiel, R., Jain, P., Singh, V. and Fischer, M. (2023) 'Digital Twin: Where do humans fit in?', *Automation in Construction*, 148, p. 104749. doi: 10.1016/j.autcon.2023.104749.

AlBalkhy, W., Karmaoui, D., Ducoulombier, L., Lafhaj, Z. and Linner, T. (2024) 'Digital twins in the built environment: Definition, applications, and challenges', *Automation in Construction*, 162, p. 105368. doi: 10.1016/j.autcon.2024.105368.

Bhatnagar, S., Jacob, G., Devkar, G., Rybkowski, Z.K., Arefazar, Y. and Obulam, R. (2023) 'A systematic review of lean simulation games in the construction industry', *Architectural Engineering and Design Management*, 19(6), pp. 701–719. doi: 10.1080/17452007.2022.2155604.

Boje, C., Guerriero, A., Kubicki, S. and Rezgui, Y. (2020) 'Towards a semantic Construction Digital Twin: Directions for future research', *Automation in Construction*, 114, p. 103179. doi: 10.1016/j.autcon.2020.103179.

Brell-Çokcan, S. and Schmitt, R. (eds) (2024) *IoC - Internet of Construction: Informationsnetzwerke zur unternehmensübergreifenden Kollaboration in den Fertigungsketten des Bauwesens*. 1st ed. Wiesbaden: Springer Fachmedien Wiesbaden GmbH.

Chacón, R., Posada, H., Ramonell, C., Jungmann, M., Hartmann, T., Khan, R. and Tomar, R. (2024) 'Digital twinning of building construction processes. Case study: A reinforced concrete cast-in structure', *Journal of Building Engineering*, 84, p. 108522. doi: 10.1016/j.jobe.2024.108522.

Čustović, I., Cao, J. and Soman, R.K. (2025) 'Agents in Construction (AiC) Ontology'. Available at: https://w3id.org/aic#.

Čustović, I., Cao, J., Wamelink, H. and Hall, D.M. (2025) 'Enhancing Situation Awareness of Construction Managers using Human-centered Digital Twins', *Advanced Engineering Informatics*, 67, p. 16. doi: 10.1016/j.aei.2025.103557.

Čustović, I., Soman, R.K., Pauwels, P. and Hall, D.M. (2023) 'Towards integrated digital twins for construction and manufacturing', *EC3 Conference 2023*. Crete, Greece. doi: 10.35490/EC3.2023.256.

Freese, M. and Bekebrede, G. (2025) 'Game research by design in project management and beyond', *Project Leadership and Society*, 6, p. 100174. doi: 10.1016/j.plas.2024.100174.

Gawthrop, P. and McGookin, E. (2004) 'A LEGO-based control experiment', *IEEE Control Systems*, 24(5), pp. 43–56. doi: 10.1109/MCS.2004.1337857.

Hasan, S.M., Lee, K., Moon, D., Kwon, S., Jinwoo, S. and Lee, S. (2022) 'Augmented reality and digital twin system for interaction with construction machinery',

Journal of Asian Architecture and Building Engineering, 21(2), pp. 564–574. doi: 10.1080/13467581.2020.1869557.

Herbsman, Z. (1986) 'Project Management Training Using Microcomputers', *Journal of Management in Engineering*, 2(3), pp. 165–176. doi: 10.1061/(ASCE)9742-597X(1986)2:3(165).

Hevner, A.R., March, S.T. and Park, J. (2008) 'Design science in information systems research', *MIS Quarterly*, 28(1), pp. 75–105. doi: 10.2307/25148625.

Jiang, Y., Li, M., Guo, D., Wu, W., Zhong, R.Y. and Huang, G.Q. (2022) 'Digital twin-enabled smart modular integrated construction system for on-site assembly', *Computers in Industry*, 136, p. 103594. doi: 10.1016/j.compind.2021.103594.

Kookalani, S., Green, S., Luo, P., Alavi, H., Parn, E., Sun, Z. and Brilakis, I. (2026) 'Mapping digital twin applications in infrastructure and the built environment across research types, methods, sectors, phases, and scales', *Automation in Construction*, 182, p. 106778. doi: 10.1016/j.autcon.2026.106778.

Li, J., Yeung, T., Ying, H., Grodsky, S., Schaumann, D. and Sacks, R. (2025) 'A laboratory setup for exploring digital twin construction planning and control', in *Proceedings of the 2025 european conference of computing in construction (2025 EC3)*. Porto, Portugal.

Martinez, J.G., Yeung, T., Sacks, R., Shahaf, Y. and Sharoni, L.-O. (2023) 'Situational Awareness in Construction Using a Serious Game', *Journal of Construction Engineering and Management*, 149(3), p. 04022183. doi: 10.1061/JCEMD4.COENG-12521.

Mathe, S.E., Kondaveeti, H.K., Vappangi, S., Vanambathina, S.D. and Kumaravelu, N.K. (2024) 'A comprehensive review on applications of Raspberry Pi', *Computer Science Review*, 52, p. 100636. doi: 10.1016/j.cosrev.2024.100636.

Parasuraman, R., Sheridan, T.B. and Wickens, C.D. (2000) 'A model for types and levels of human interaction with automation', *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3), pp. 286–297. doi: 10.1109/3468.844354.

Sacks, R., Brilakis, I., Pikas, E., Xie, H.S. and Girolami, M. (2020) 'Construction with digital twin information systems', *Data-Centric Engineering*, 1, p. e14. doi: 10.1017/dce.2020.16.

Sacks, R., Esquenazi, A. and Goldin, M. (2007) 'LEAPCON: Simulation of Lean Construction of High-Rise Apartment Buildings', *Journal of Construction Engineering and Management*, 133(7), pp. 529–539. doi: 10.1061/(ASCE)0733-9364(2007)133:7(529).

Saif, W., Doukari, O. and Kassem, M. (2026) 'Stakeholder-centric whole-lifecycle framework for guiding the development and implementation of construction digital twins', *Automation in Construction*, 183, p. 106773. doi: 10.1016/j.autcon.2026.106773.

Saif, W., RazaviAlavi, S. and Kassem, M. (2024) 'Construction digital twin: a taxonomy and analysis of the application-technology-data triad', *Automation in Construction*, 167, p. 105715. doi: 10.1016/j.autcon.2024.105715.

olde Scholtenhuis, L. and Soman, R.K. (2026) 'Trapped in technocentric thinking? Revisiting Digital Twins through a pragmatic framework', in *Proceedings of the 41st annual ARCOM conference*. Dundee, UK.

Soman, R.K., Farghaly, K., Mills, G. and Whyte, J. (2025) 'Digital twin construction with a focus on human twin interfaces', *Automation in Construction*, 170, p. 105924. doi: 10.1016/j.autcon.2024.105924.

de Soto, B.G. and Skibniewski, M.J. (2020) 'Future of robotics and automation in construction', in *Construction 4.0*. Routledge, pp. 289–306.

The LEGO Group (2026) 'LEGO® SMART Play™ System | Official LEGO® Shop US'. Available at: https://www.lego.com/en-us/smart-play (Accessed: 25 January 2026).

Tommelein, I.D. (2020) 'Design science research in construction management: multi-disciplinary collaboration on the SightPlan system', *Construction Management and Economics*, 38(4), pp. 340–354. doi: 10.1080/01446193.2020.1718723.

Walzer, A.N. (2025) *Assessing Multi-Level Uncertainties in Construction Robotics Entrepreneurship by Means of Methodological Pluralism*. ETH Zurich. doi: 10.3929/ETHZ-B-000705101.

Whyte, J., Soman, R.K., Sacks, R., Mohammadi, N., Naderpajouh, N., Hong, W.-T. and Lee, G. (2025) 'Using Digital Twins for Managing Change in Complex Projects', in A. GhaffarianHoseini, F. Rahimian, and M. Babu Purushothaman (eds). *International Conference on Smart and Sustainable Built Environment (SASBE 2024)* (Lecture Notes in Civil Engineering), pp. 1575–1582. doi: 10.1007/978-981-96-4051-5_150.

Zheng, Y., Al Barazi, A., Seppänen, O., Abou-Ibrahim, H. and Görsch, C. (2025) 'Semantic digital twin framework for monitoring construction workflows', *Automation in Construction*, 176, p. 106301. doi: 10.1016/j.autcon.2025.106301.

Zhu, H., Hwang, B.-G., Tan, Y.Z. and Wei, F. (2024) 'Building on Digital Twin: Overcoming Barriers and Unlocking Success in the Construction Industry', *Journal of Construction Engineering and Management*, 150(10), p. 04024142. doi: 10.1061/JCEMD4.COENG-13991.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

ENRICHING VERSION INCREMENTS WITH ENGINEERING INTENT FOR ENHANCED BIM MODEL CHECKING

Sebastian Esser [1,2], Jiabin Wu [1,2], Suhyung Jang [1,2], Lena Fahrner [3], Robert Amor [4], and André Borrmann [1,2]
[1] Chair of Computing in Civil and Building Engineering, Technical University of Munich, Germany
[2] TUM Georg Nemetschek Institute, Technical University of Munich, Germany
[3] Landmarken AG, Aachen, Germany
[4] University of Auckland, School of Computer Science, Auckland, New Zealand

Abstract

BIM-based model checking helps improve the quality of design data throughout the planning phase of a built asset. However, rule-based checking can be time-consuming for large, frequently updated BIM models. Increment descriptors for BIM models are enriched with engineering intent, thereby unveiling additional information about the changes made. The proposed method enhances model-checking efficiency by executing only the rules affected by changes, rather than reprocessing the entire model. A case study demonstrates that selective rule triggering reduces redundant checks by re-executing only the rules affected by a given change, automatically identifying which open issues a model update resolves, thereby improving traceability between design edits and compliance outcomes.

Introduction

In recent years, substantial progress has been achieved in coordinating interdisciplinary models in the Architecture, Engineering, and Construction (AEC) industry, largely through the adoption of Building Information Modeling (BIM) approaches into collaborative design workflows. Design data is primarily stored in discipline-specific BIM models, which users typically access as distinct monolithic files. An especially important application of model-based workflows is the validation of models against a predefined rule set, a process often referred to as BIM-based model checking (Eastman et al., 2009) These rules are derived from client requirements or codes and standards that the multidisciplinary design must comply with. Any identified problems are recorded as issues and returned to the designer, who is responsible for adjusting the design models so that they ultimately conform to the stated rules.

Model checking is typically performed at predefined milestones and on a complete model without reusing results from previous runs. The intermediate steps designers take to resolve issues are neither documented nor communicated. Consequently, current checking systems do not support incremental validation on updates, increasing checking time and resulting in unnecessary rework. So far, no mechanisms are in place to automatically track whether the changes have resolved the previously identified issues. As a result, engineers or reviewers must manually review the model-checking results to determine whether new issues need to be documented and assigned, and which existing issues were resolved in the latest design iteration.

Contribution

This paper introduces a new method that applies version control concepts to transmit only the changes made to a BIM model, rather than sending entire model files. These version increments are subsequently used to provide the rule system with more precise information about which modified model elements need to be checked. To this end, the graph transformations that constitute the version increments are supplemented with higher-level annotations that express the engineering operations performed and identify the affected parts of the BIM model. Consequently, only those rules influenced by recent model changes need to be re-executed when a model increment is shared. Furthermore, associating each version increment with its enriched descriptors and the rule outcomes generates a valuable information resource for the subsequent issue management, as it reveals which issues may need a status update and directs the engineer to areas that warrant particular attention.

Related Works

BIM-based Model Checking and Issue Processing

The fundamental concepts of BIM-based model checking were introduced by Eastman et al. (2009) and subsequently adopted and further developed by other scholars. In summary, a model-checking routine typically involves four main stages: translating the rules into a machine-readable language, preparing the building model data, carrying out the checking process, and finally presenting the results.
Lee et al. (2026) provides an overview of recent advances in code-compliance checking, highlighting three key research areas: interpreting regulatory text into computable formats, model checking workflows with BIM and IFC data, and integrated workflows linking interpretation, checking, and reporting. To overcome manual translations, Fuchs et al. (2024) employed Large Language Models (LLMs) to convert regulatory documents into actionable, machine-executable representations. The execution logic of the checking routines must also be presented clearly and transparently. Preidel and Borrmann (2016) proposed a visual programming framework that accelerates the creation of rules and provides more flexibility than conventional model-checking tools. As an extension to these previous works, Iversen and Huang (2026) have investigated whether LLMs can directly be used for rule

interpretation, information extraction from BIM models, rule execution, and reporting bypassing manual preparation of distinct elements in a model-checking workflow. Even though these advancements are promising, most software products on the market provide predefined rule templates that can be adapted to specific scenarios to be checked.

Despite these advances in rule interpretation and representation, existing automated compliance checking workflows still determine the scope of checking through model-level filtering. This mechanism presupposes that the entire BIM model is available and consistently structured, but more importantly, it prevents the system from identifying which elements are affected by design changes. As a result, the current approaches cannot support incremental compliance checking and instead require full re-evaluation after every update.

Design Intent and Rationale

Every decision made during the design and planning phase of a built asset entails reflections on *how* and *why* it is constructed. Although these modes of thinking are at the heart of the design and planning disciplines in the AEC sector, they are rarely explicitly documented in design models. Wyke and Munch Lindhard (2025) report that it would be highly advantageous if the models also contained the underlying engineering intent that caused a change, as well as the reasoning behind the modifications performed. Developing effective methods to capture the rationale of a design has been a longstanding goal in engineering design (Siddharth and Sarkar, 2018). Without an explicit representation of design intent, design changes can easily introduce inconsistencies in building planning. For example, changes in architectural design can cause sprinkler layouts to become misaligned, making MEP revisions necessary to reestablish the intended coordination with the updated architecture (Sacks et al., 2022).

Currently, issue-related communication for BIM models is primarily facilitated via the BIM Collaboration Format (BCF). Because the BCF provides only very limited metadata about identified issues, the actions taken to resolve them can currently be recorded only as written comments attached to the issue. Such textual notes can convey intent and rationale only to a very limited extent, which is insufficient to support further automation of downstream processes. Consequently, the engineering intent behind a model change is generally implicit and is expected to correspond to the issues and defects revealed in the preceding model check. To support the design decision-making process, Wu et al. (2025) introduce a Design Healing framework that searches for code-compliant alternatives to resolve identified building code violations while remaining close to the originally intended solution. This framework focuses on localized changes, such as wall movements, and avoids modifications that would alter the existing building design topology. However, design changes in practice often alter the original building topology, for example, by creating or deleting building elements, and there is still limited support for systematically analyzing the underlying engineering intent behind such changes. Hu et al. (2021) capture the dependency between clashes and generate an optimal component change list to minimize the impact on the project. This method addresses the challenge posed by highly interdependent building systems, where clashes can propagate across related components.

Change propagation and version control of BIM models and issue representations

To establish a connection between implemented changes and unresolved issues, a robust method is needed to record and reveal the modifications made to the design models. Precisely monitoring changes across multiple model versions is labor-intensive and has to be performed again for every new version. This situation was primarily driven by the lack of version control techniques for BIM models that track changes based on the underlying object structures in the past.

As a result, changes between different model states often require substantial manual investigation before potential impacts on other disciplines can be evaluated. Whereas some methods have focused solely on algorithms for performing diff operations directly on the underlying data structures of BIM models (Shi et al., 2018; Liu et al., 2026), Esser et al. (2022) proposed dedicated version control mechanisms grounded in graph-based representations of BIM models, enabling more fine-grained, efficient management of model changes. In essence, BIM models are transformed into Labeled Property Graphs. Changes are then classified as either topological modifications covering inserted or removed nodes and edges, or semantic modifications, capturing attribute-level edits to structurally unchanged nodes. Any change information is ultimately assembled into a graph transformation rule p_{ij}, which is exchanged as an incremental change description. Figure 1 illustrates the approach schematically.

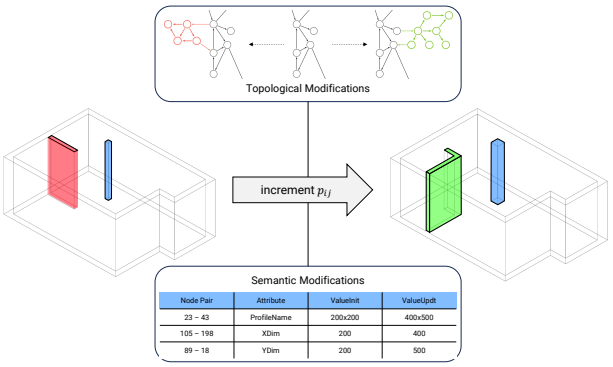


Figure 1: Incremental Version Control of BIM models.

Further efforts have focused on embedding version-control concepts directly into the issue representations. Oraskari et al. (2022) explored different ways to capture the history and evolution of issues within bcfOWL, which encodes issue data as a triple-based semantic graph in accordance

with the Resource Description Framework (RDF) principles.

There has also been research conducted to analyze the as-happened BIM authoring process using BIM logs. As data sources for understanding incremental authoring events, BIM logs provide information about how models are revised over time. Although BIM logs have initially been developed for monitoring and maintenance of authoring systems, several studies have enhanced their metadata to better capture modeling activities (Jang et al., 2023). Yarmohammadi and Castro-Lacouture (2018) introduced a bounding-box-based building element logs to quantify modeling performance. Gao et al. (2021) proposed a graph-structured event log that links modeling commands to the geometry they affect, enabling the analysis of trial-and-error patterns during authoring. Although these efforts commonly focus on improving the representation of element-level modeling events and revision trajectories, such logs do not capture the inter-element relationships that are essential for understanding model compliance. Moreover, they primarily record the state of BIM elements affected by modeling actions, whereas compliance checking requires tracking change propagation, which is significant in model checking.

Identified research gap

As discussed, model checking not only requires a carefully prepared set of rules but also well-prepared BIM models to be subjected to the checking process. However, current rule-checking workflows typically re-execute the entire rule set for each new design model file, regardless of what has changed in the interim, reflecting a lack of mechanisms to identify incremental checking targets. Moreover, existing approaches do not provide explicit representations to capture how changes in one model element propagate to related elements and their associated rules, making it difficult to reason about the true impact of a modification. Finally, there is no structured way to record compliance issues, the changes made in response, and how those issues are resolved over time. Incorporating version control principles into BIM-based workflows could address these gaps by systematically tracking edits, modeling their propagation across related elements and rules, and maintaining structured records that support more transparent, targeted, and repeatable compliance checking. Therefore, this paper aims to bridge these gaps by linking model updates to the subset of rules required to evaluate the performed changes, connecting checking results with subsequent issue information for coherent tracking across model updates, compliance assessments, and project communication items.

Proposed Methodology

Our approach aims to improve the preparation of the building model and the rule execution process by determining which rules to apply based on the exact changes made to the design model. When new components are first introduced to the design model, they are evaluated against the complete set of rules relevant to that modification scope. Any subsequent changes are then evaluated only against the subset of rules affected by those updates. In particular, by applying version control to the design models as they evolve, we gain direct access to detailed information about the changes made. The version increment that records these modifications describes the alterations at a strictly data-representation level following concepts of set and graph theory. Consequently, we propose additional enrichment and processing of this information to enable its effective use in the subsequent workflow.

Processing Pipeline

The envisioned processing pipeline is illustrated in Figure 2 and considers a designer who incrementally enhances their BIM model containing information about a specific discipline or trade, and a model checking workflow is used to validate whether the model complies with a given set of rules, indicated in the right part of Figure 2. The modification details captured in the version increment are enriched to determine which rules must be executed, enabling a more efficient assessment of the model's compliance. To accomplish this, four processing steps need to be incorporated into the standard model checking workflow:

Rule setup: Each rule performs checks on specific elements contained in the BIM model. In this way, we can understand which elements are checked by which rules and what additional context they require.

Increment Interpreter: In this stage of processing, we apply pattern matching to the version increments to derive engineering-level insights from the underlying graph transformation.

Assignment Module: This module uses the version increment and the identified modification scope to determine and trigger which rules must be re-validated in light of the design changes made.

Reasoning Module: In this step, the outcome of the performed check, along with the version increment, is used to decide whether a new issue must be created or whether satisfying the rule should instead trigger an update to an issue's status or its closure.

Rule setup

Figure 3 schematically shows the configuration of the rules. Each rule is defined by its application target, the necessary model context, a modification scope to which the rule is sensitive, and the specific checking procedure. Together, these elements form the digital representation of a particular aspect derived from a code or standard that the design is required to comply to.

The *application target* defines the primary focus of a rule, specifying which elements are explicitly examined. Targets can be selected based on element types and/or spatial or functional containment, or their combinations (e.g., all windows from the second floor upward). In addition to the application target, a rule can specify *supplementary model context*. This context provides additional informa-

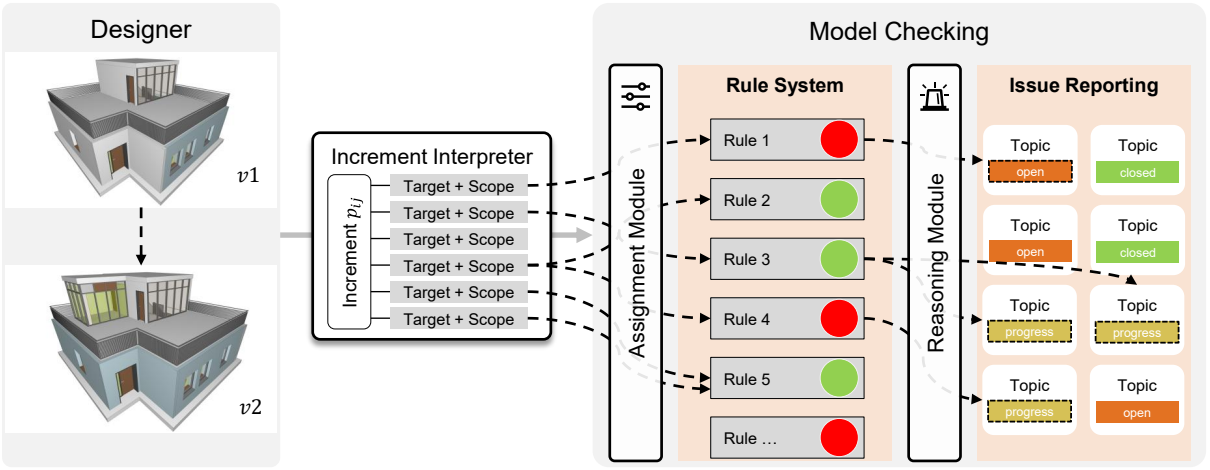


Figure 2: Overall processing pipeline: a model author creates two versions of a BIM model, from which version control principles produce an increment representing the performed changes. The increment is enriched before being fed into the model checking system.

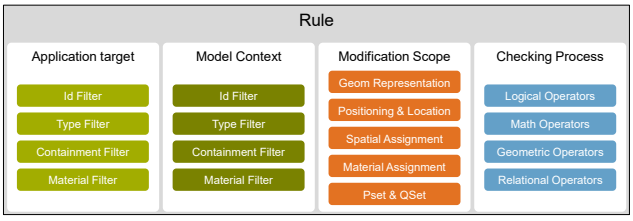


Figure 3: Rule Setup with the Application Target, Model Context, Modification Scope, and Checking Process.

tion that the checking process may require for execution. It can be retrieved from the model at any time, independently of whether the corresponding elements were modified in the latest design iteration.

For example, consider a rule that checks whether doors intersect with furnishings when opened. The application target of this rule would be the insertion of doors or modifications to their geometry and position. However, the furnishing elements are part of the model context and are fetched from the BIM model when the checking process needs them for geometric overlap calculation. Because rules can depend both on their application target and on their model context, rule execution must be defined in a commutative way: the rule must be re-executed not only when the application target changes, but also when elements forming the model context are modified — even if the application target itself would not otherwise be triggered.

The *modification scope* serves as a trigger interface that determines which types of model modifications the rule must be assessed for, both in terms of the application target and the model context. It consists of various categories of changes that can occur within a BIM model and is used to determine, for both the specific elements addressed by the rule (application target) and the surrounding conditions in the BIM environment (model context), under which circumstances the rule should be executed. The operators shown in the checking process section in Figure 3 correspond to those introduced by Preidel and Borrmann (2016), but can be expanded further to address specific issues.

Increment Interpreter

The version increment that captures the applied model modifications is formalized as a graph transformation rule and adheres to well-established notations from fundamental graph and set theory. In the method presented by Esser et al. (2022), a Double-Push-Out formalism is used to define how the underlying object network forming the BIM model must be altered by deleting and inserting subgraphs to transform the model from its initial state to its updated state. These subgraphs are described in the Left-Hand-Side (LHS), the Interface graph I, and the Right-Hand-Side (RHS), reflecting the required transformation steps.

The modification operations captured in a version increment describe changes at the raw data level, interacting only with nodes and edges in the graph. While model updates at any level of granularity can be reliably expressed in this way, additional interpretation is required to derive higher-level information from these version increments that can guide our decisions about which rules need to be re-evaluated. Consequently, the increment interpreter is introduced as a preprocessing step before the data enters the model checking system (as illustrated in Figure 2). Figure 4 illustrates the interpretation process.

In this processing step, we examine the contents of the version increment and perform pattern matching on the graphs that constitute the underlying transformation rule, describing the necessary actions to update the BIM model (i.e., its graph representation) from version 1 to version 2. More formally, this means that we look for an occurrence of an operation pattern O_i within the graph patterns that form the Left-Hand Side (LHS) and the Right-Hand Side (RHS) of the transformation rule. The operation patterns O are assumed to be predefined and are supplied as a library. For each pattern, the library provides additional logic that extracts information about the target of the modification (e.g., the globally unique identifier of the object

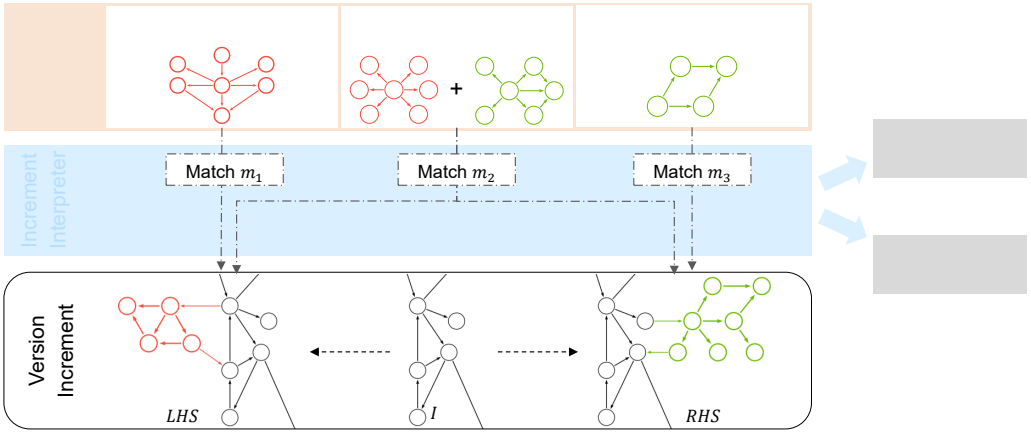


Figure 4: Increment interpreter: The version increment obtained from comparing two versions of a BIM model is shown at the bottom. The Interpreter's goal is to identify occurrences of the graph patterns defined in the operation library to reveal the scope and its target.

being modified) and the modification scope. Both aspects form the output of the increment interpreter, which will then be utilized in the assignment module.

Naturally, semantic changes also provide an informational foundation, meaning that the graph paths uniquely identifying nodes affected by semantic changes can likewise be used to infer higher-level statements about the target and the extent of the modification. Common examples of semantic changes include modifications to the Cartesian coordinates that define an element's location, as well as alterations to its name or description attributes, which are attached to nodes representing instances of subclasses derived from IfcRoot.

Assignment Module

The assignment module is responsible for linking the modification scopes and application targets, as determined by the Increment Interpreter, to the rules that must be executed in response to the detected model changes. In essence, it decides which rules are affected by which changes and orchestrates their re-execution. When existing components in the model are modified, the corresponding rules are selectively re-triggered, restricted to the identified modification scope, so that only the impacted portions of the model are reconsidered. In contrast, when elements are removed, any rules that reference them within the model context must be re-evaluated. If deleted elements were the original targets of specific rules and had already caused problems, those rules may be closed once the elements are removed. Similarly, when new elements are introduced, the assignment module identifies all rules that are applicable to these elements and ensures that they are executed accordingly. This targeted rule assignment enables incremental, efficient, and consistent evaluation of model changes without recomputing the entire rule set.

Reasoning Module

After the rules have been evaluated, their outcome is essentially a basic pass-or-fail result, typically accompanied by information about the elements that were checked and failure information about missing properties or violated parts of rules. These results are then further processed to support the system operator in managing issues.

In this step, the Reasoning Module preserves an explicit link between the performed model changes, the rules that were (re)evaluated as a consequence, and the issues that arise from those rule activations. It maintains the connection from the originating rule, through its evaluation outcome, to the corresponding problem descriptions and the resulting change requests addressed to the model authors. The issues themselves are recorded using the BIM Collaboration Format (BCF), which stores key metadata for each identified problem, such as its name, description, and the GUIDs of the affected elements. On top of this, the issues are enriched with identifiers that allow us to trace back precisely which rule triggered the documented problem or topic, and which model revision it relates to.

Because we can track how modifications to the model affect each rule's output, we can also determine which open issues are likely to be resolved by the considered model update. Although much of this analysis can be automated using the maintained links between model changes, rule activations, and issues, the final decision on each issue's status is ultimately left to the model reviewer.

Rationale of the Processing Pipeline

The proposed approach integrates into existing model-based workflows as a preprocessing layer, without requiring fundamental changes to the underlying rule system or issue management infrastructure, while the rule setup module enables practitioners to define checking targets at a finer granularity than conventional model-level filtering. The increment interpreter replaces the implicit assumption that reviewers can manually identify the scope of a modification, and the assignment module translates this information directly into rule-execution decisions, ensuring that only the affected subset of rules is re-triggered, thereby reducing the computational overhead associated with frequent model updates. Beyond efficiency gains, the reasoning module closes the feedback loop by connect-

ing rule outcomes to issues, introducing a level of process transparency that current workflows lack and enabling reviewers to trace why a particular check was triggered and which design decision prompted it. Over time, the accumulated change-issue pairs constitute a structured record of a project's compliance history, which could inform both design decisions and quality assurance processes in future projects.

Case Study

The following case study demonstrates the key ideas of the proposed method using a small-scale building model. The designer arranged a main hall as the central meeting space, with several auxiliary service rooms located at the rear of the building. To verify the design compliance, two specific rules out of a larger rule set are considered:

- Exit-Positioning rule: When two exits are required from an exit access area, they must be separated by at least one-half of the longest diagonal of the area, measured in a straight line [1].
- FireRatingPSet rule: Each door and window must be assigned a property set that includes its fire rating information.

Each rule provides the system with the parameters that it requires to run. For the Exit-Positioning rule, doors constitute the primary application target, and the rule requires the adjacent space geometry as model context in order to derive the minimum clearance distance to be maintained between individual doors. The rule is therefore sensitive to any modification in the geometric representation of the doors or their placement within the spatial configuration. In comparison, the FireRatingPSet rule applies specifically to doors and windows and focuses on fire-resistance data. Consequently, this rule reacts to modifications in the corresponding pSet attributes.

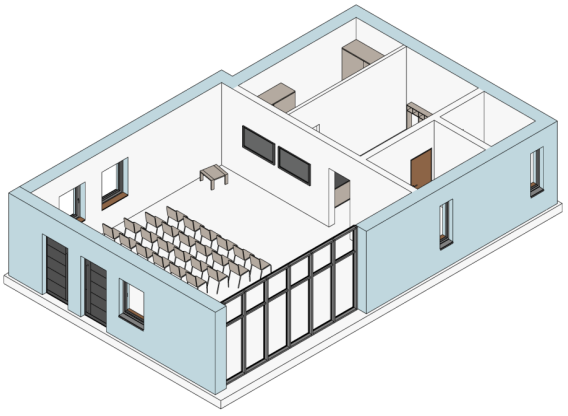


Figure 5: Case Study Model.

[1] Summarized from IBC Rule 1007.1.1, available `https://codes.iccsafe.org/content/IBC2018P6/chapter-10-means-of-egress#IBC2018P6_Ch10_Sec1007.1.1` (last access 2026-01-26)

Figure 5 illustrates the model in its initial version state, which is checked against any rule, as any information contained in the design model was newly inserted. In the initial design, two primary exit doors are positioned next to each other to provide access to the main hall. Given that the diagonal of this room is approximately 11.30 m, the spacing between these doors is clearly insufficient. Consequently, an issue is generated directly from the detected violation, and the association between the rule and the resulting issue is recorded. All doors and windows in the initial design contained adequate property set data; therefore, the FireRatingPSet rule confirmed their compliance. After returning the detected issues to the designer, they revised the model. Figure 6 illustrates the design modifications performed. One door has been repositioned on another wall, and a new window has been inserted.

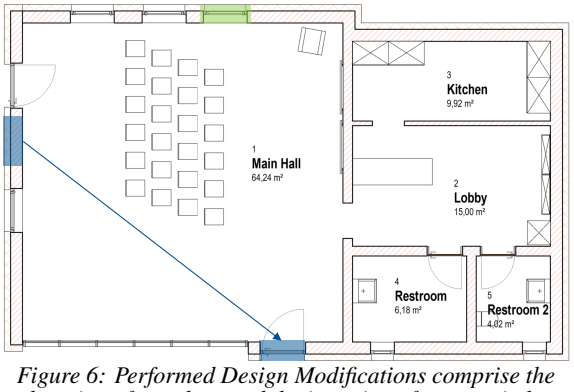


Figure 6: Performed Design Modifications comprise the relocation of one door and the insertion of a new window.

The changes are recorded as a version increment, comprising 287 nodes and 282 relationships in the removed pushout pattern, and 613 nodes and 637 relationships to be inserted via the inserting pushout pattern. Furthermore, 1817 node attributes have been edited. These nodes have not been altered in their topological position within the model graph but underwent semantic changes.

Figure 7 shows the operation patterns O that are available in the operation library. The increment interpreter applies these patterns to recognize that an entrance door has been modified and then reports the changes to the object's position and its nesting assignment, since the door is now placed on a different wall. As the version increment does not alter the fire resistance parameters of the relocated door, it is concluded that the door continues to comply with the fire resistance rule. Consequently, only the Exit-Positioning rule is re-assessed. As this rule reports no violations, the issue raised based on the original design model can be marked as resolved by the model modification and may be closed by the model reviewer, given the positive evaluation result. However, the newly inserted window is checked against the FireRatingPSet rule to ensure it meets the required element criteria. If compliance is confirmed, no issue needs to be created; if not, a new issue is generated that retains information about the originating rule and the non-compliant element.

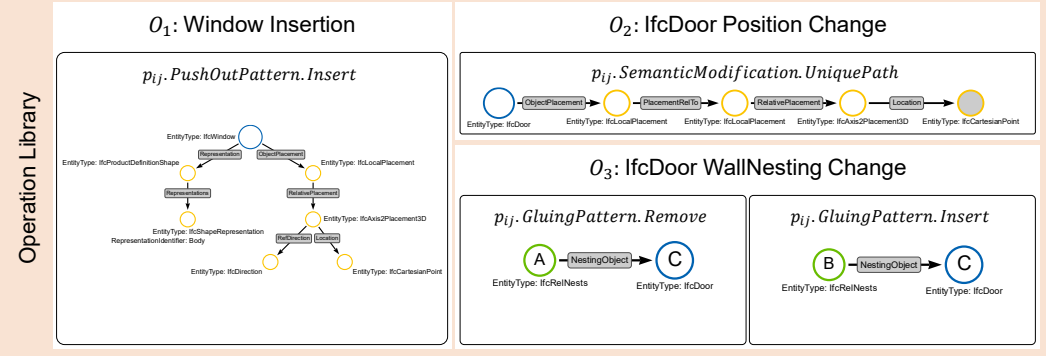


Figure 7: Case Study: Operation library used to identify the performed design modifications.

Limitations

Despite the potential benefits of the proposed approach, several limitations must be acknowledged.

First, the proposed concept relies heavily on an operation library that maps concrete design modifications to corresponding rule triggers. A core limitation is the assumption that this operation library is complete and accurate. In reality, constructing such a comprehensive operation library is challenging as certain modeling operations may not be fully captured or formalized. Furthermore, new modeling patterns and tools can introduce operations that are not yet known to the increment interpreter. If the operation library is incomplete or inaccurately specified, relevant triggers may not be fired, resulting in missing relevant checks. Consequently, the overall reliability of the approach is directly dependent on the continuous maintenance, validation, and extension of this library.

The current approach focuses on design changes that explicitly trigger rule checks. However, there can be intricate rules, which can be violated implicitly, even when no formal change trigger has been activated. Modifications in one part of the model may indirectly affect constraints or assumptions embedded elsewhere in the rule system, leading to undetected non-compliance. For example, design compliance with evacuation time depends on many interrelated factors such as space connectivity, door and corridor widths, and fire compartmentation. A local change, such as moving a door or partition, may not trigger an explicit evacuation check, yet it can still invalidate the relevant computed egress path or partially undermine its underlying design intent. While running the complete rule system on a periodic basis (e.g., weekly) partially mitigates this risk, it does not eliminate the possibility of temporary non-compliant states between full runs. In practice, this means that some violations may only be discovered with delay, and selective checks cannot yet guaranty full coverage of all rule interactions.

Another limitation of this study can be found in the limitation to single-discipline problems. Many industrial model-checking scenarios, however, involve multidisciplinary design models, where several domains are simultaneously represented. Interdisciplinary dependencies among building elements are common, and a change in one discipline may require corresponding updates in others to maintain conformance with relevant constraints and compliance requirements, for example, moving a wall may require relocating pipes and connected equipment in the MEP model. The current reasoning approach primarily targets change analysis within a single domain and must be extended to serve for interdisciplinary aspects, accordingly. In summary, while the method improves responsiveness and focuses rule-checking effort around actual changes, its effectiveness is constrained by the completeness of the operation library, the handling of implicit rule interactions, and the ability to reason over multidisciplinary models.

Summary and Outlook

In this paper, we introduced a novel approach to improving model-checking workflows by leveraging incremental version control as a fine-grained source of information for intelligent rule selection and subsequent reasoning about issue documentation. The approach decomposes the overall problem into smaller processing steps. First, version control principles are applied to track the evolving BIM model on the author's system. Performed changes are encoded as version increments, which are essentially graph transformations leveraging the underlying object network in the BIM model from its initial to an updated state. Then, we use these version increments to guide informed rule selection and focused model checking. Instead of running the full rule set after every modification, our method constructs a dependency mapping between model-edit operations and the rules that may be affected by those edits. Guided by this mapping, the model checking is invoked selectively, concentrating only on rules whose input parameters have been modified. By linking the outcomes of checks to the resulting issues, the approach enables both retrospective and predictive reasoning. Detected issues are associated with the specific changes that introduced or resolved them, thereby building a corpus of annotated change–issue pairs.

By exposing the relationship between detected issues and the underlying model changes, our method offers deeper insights into model evolution and the root causes of problems. Future work should extend the current illustra-

tive case study through experiments on real-world projects to examine more explicitly the reduction in rule executions enabled by enhanced traceability in model checking workflows. Moreover, collecting and learning from this data may enable future systems to propose solutions to issues with similar characteristics, paving the way for more proactive, intelligent support in model-based design environments.

Use of AI tools

While preparing this work, the authors utilized Large Language Models inside Overleaf Writefull AI to refine the language and identify spelling errors. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

Acknowledgments

The research presented has been financially supported by the Federal Ministry of Education and Research (Bundesministeriums für Bildung und Forschung) in the frame of the project BauPuls360 under grant number 02K23A096. Furthermore, the work was supported by the funds of TUM Georg Nemetschek Institute. The responsibility for the content of this publication lies with the authors.

References

Eastman, C., Lee, J.-m., Jeong, Y.-s., and Lee, J.-k. (2009). Automatic rule-based checking of building designs. Automation in Construction, 18(8):1011–1033. doi: 10.1016/j.autcon.2009.07.002.

Esser, S., Vilgertshofer, S., and Borrmann, A. (2022). Graph-based version control for asynchronous BIM collaboration. Advanced Engineering Informatics, 53:101664. doi: 10.1016/j.aei.2022.101664.

Fuchs, S., Witbrock, M., Dimyadi, J., and Amor, R. (2024). Using Large Language Models for the Interpretation of Building Regulations. Journal of Engineering, Project, and Production Management, 14(4). doi: 10.32738/JEPPM-2024-0035.

Gao, W., Wu, C., Huang, W., Lin, B., and Su, X. (2021). A data structure for studying 3D modeling design behavior based on event logs. Automation in Construction, 132:103967. doi: 10.1016/j.autcon.2021.103967.

Hu, Y., Castro-Lacouture, D., Eastman, C. M., and Navathe, S. B. (2021). Component Change List Prediction for BIM-Based Clash Resolution from a Graph Perspective. Journal of Construction Engineering and Management, 147. doi: 10.1061/(asce)co.1943-7862.0002092.

Iversen, O. and Huang, L. (2026). Leveraging large language models for BIM-based automated compliance checking. Automation in Construction, 182:106707. doi: 10.1016/j.autcon.2025.106707.

Jang, S., Lee, G., Shin, S., and Roh, H. (2023). Lexicon-based content analysis of BIM logs for diverse BIM log mining use cases. Advanced Engineering Informatics, 57:102079. doi: 10.1016/j.aei.2023.102079.

Lee, J. H., Ji, S. Y., and Ostwald, M. J. (2026). Automated compliance checking across the building lifecycle: Systematic and semantic review integrating PRISMA and deep search. Automation in Construction, 185:106859. doi: 10.1016/j.autcon.2026.106859.

Liu, H., Ma, J., Gao, G., and Gu, M. (2026). Decoupling IFC models for reliable modification of shared references. Automation in Construction, 181:106670. doi: 10.1016/j.autcon.2025.106670.

Oraskari, J., Schulz, O., and Beetz, J. (2022). Towards describing version history of bcf data in the semantic web. In LDAC 2022: 10th Linked Data in Architecture and Construction Workshop, page 87.

Preidel, C. and Borrmann, A. (2016). Towards Code Compliance Checking on the Basis of a Visual Programming Language. Journal of Information Technology in Construction, 21:402–421.

Sacks, R., Wang, Z., Ouyang, B., Utkucu, D., and Chen, S. (2022). Toward artificially intelligent cloud-based building information modelling for collaborative multidisciplinary design. Advanced Engineering Informatics, 53:101711. doi: 10.1016/j.aei.2022.101711.

Shi, X., Liu, Y. S., Gao, G., Gu, M., and Li, H. (2018). IFCdiff: A content-based automatic comparison approach for IFC files. Automation in Construction, 86(June 2016):53–68. doi: 10.1016/j.autcon.2017.10.013.

Siddharth, L. and Sarkar, P. (2018). A Multiple-Domain Matrix Support to Capture Rationale for Engineering Design Changes. Journal of Computing and Information Science in Engineering, 18(2). doi: 10.1115/1.4039850.

Wu, J., Nousias, S., and Borrmann, A. (2025). Design Healing framework for automated code compliance. Automation in Construction, 171:106004. doi: 10.1016/j.autcon.2025.106004.

Wyke, S. and Munch Lindhard, S. (2025). Understanding Design Rationale and Intent through Natural Language Processing analysis: A search for Consensus. Journal of Information Technology in Construction, 30:631–649. doi: 10.36680/j.itcon.2025.026.

Yarmohammadi, S. and Castro-Lacouture, D. (2018). Automated performance measurement for 3D building modeling decisions. Automation in Construction, 93:91–111. doi: 10.1016/j.autcon.2018.05.011.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A COMPUTATIONAL DEA-BASED DECISION SUPPORT FRAMEWORK FOR GREEN SUPPLY CHAIN MANAGEMENT IN CONSTRUCTION

Erawati Ravindra Kadam, and Ann Francis
Department of Civil and Environmental Engineering, Indian Institute of Technology Delhi, India

Abstract

In construction, greening supply chains requires balancing economic performance with environmental impact, making systematic evaluation and comparison difficult. Current approaches rely on single indicators or subjective assessments, limiting their ability to identify relatively efficient configurations. This paper proposes a Data Envelopment Analysis (DEA)-based framework to computationally evaluate green supply chain management (GSCM) scenarios. Economic and environmental parameters are treated as inputs and outputs in the DEA model. An application to 81 scenarios identifies supply chain configurations that achieve environmental improvements without disproportionate economic penalties, addressing a key challenge in implementing GSCM in the construction sector.

Introduction

The construction industry plays an important role in the world's economic growth and infrastructure development. Apart from construction activities, a substantial amount of construction-related environmental and economic impact is embedded in upstream material production, transportation, and logistics decisions (Wibowo, Handayani and Farida, 2022). As a result, attention has increasingly shifted toward construction supply chains as a key leverage point for improving overall project performance. In response, green supply chain management (GSCM) has emerged as an approach that integrates conventional supply chain decisions with environmental considerations, while maintaining economic and operational performance (Farida, Handayani and Wibowo, 2019). In construction, GSCM practices include reducing material wastage, improving energy efficiency, lowering carbon emissions, and optimising transportation and logistics operations (Castellano *et al.*, 2023). However, these strategies rarely improve all performance dimensions simultaneously.

A primary challenge in implementing GSCM in construction lies in managing trade-offs among multiple, often conflicting parameters, including cost, emissions, waste generation, time, and profitability (Wibowo, Handayani and Farida, 2022). Therefore, evaluating alternative options in GSCM configurations is a complex multi-criteria decision-making problem. Existing practices focus on a single dimension or a descriptive comparison, such as cutting emissions or reducing costs and time. Because of focusing on one dimension, these approaches often fail to capture the multidimensional nature of supply chain performance and typically require subjective weighting of criteria.

These constraints are particularly evident in project-based construction contexts where total material demand is fixed, and variations among alternative supply chain configurations arise primarily from differences in resource utilisation efficiency rather than scale. Under such conditions, conventional evaluation methods struggle to provide objective benchmarking across green supply chain scenarios. There remains a lack of structured, scenario-based analytical frameworks tailored to project-level construction supply chains, capable of simultaneously capturing economic, environmental, and operational performance while enabling objective efficiency benchmarking under constant material demand conditions. Addressing this gap requires an efficiency-based evaluation approach that can handle multiple inputs and outputs concurrently, avoid subjective weighting of performance criteria, and benchmark alternatives relative to best-performing configurations.

Data Envelopment Analysis (DEA) offers a comprehensive solution to this issue. DEA is a non-parametric method for evaluating efficiency that measures the relative performance of similar entities by examining their capacity to convert various inputs into preferred outputs, without the necessity for predetermined weights (Charnes, Cooper and Rhodes, 1978; Banker, Charnes and Cooper, 1984). DEA has shown potential applications in supply chain studies as well as environmental and sustainability research to evaluate performance independently (Zhou, Ang and Poh, 2008). These attributes make DEA particularly suitable for assessing GSCM in construction projects that involve competing economic and environmental objectives.

Therefore, this study proposes a structured, scenario-based DEA framework for assessing GSCM efficiency in project-level construction supply chains under constant demand conditions. Alternative supply chain configurations are systematically developed using literature-based economic and environmental parameters and evaluated as decision-making units within a variable returns to scale (VRS) DEA model. By focusing on project-oriented supply chain contexts where material

quantities are design-driven, the study provides a targeted analytical approach for comparing green supply chain configurations and identifying improvement pathways. The findings contribute methodological support for efficiency-based decision-making in construction supply chains and offer practical insights for managers and policymakers evaluating sustainable supply chain interventions.

Research background

Conventionally, construction supply chains have been designed primarily to minimise costs, improve efficiency, ensure on-time deliveries, and maintain sufficient inventory to prevent construction delays. However, these priorities often conflict with sustainability goals, leading to overproduction, excessive transport, and material wastage (Cataldo *et al.*, 2022) .

The rising urgency of climate change and resource scarcity has placed sustainability at the forefront of supply chain management. Governments, investors, and clients now demand construction projects that reduce carbon emissions, minimise waste, and adopt circular practices(Chen, Feng and Garcia de Soto, 2022). For cement, reducing emissions is a global priority, as its production contributes significantly to overall greenhouse gas levels (GCCA India-TERI, 2025) . But project stakeholders cannot ignore cost pressures, productivity requirements, and delivery deadlines. This dual pressure creates a fundamental tension over whether to focus on environmental outcomes or efficiency-focused measures (Golpîra, 2021).

This tension is especially evident in decisions around production and inventory (Castellano *et al.*, 2023). For example, producing and storing more material than immediately needed helps prevent material shortages on site, but increases emissions, costs, and risks of wastage (Chen, Feng and Garcia de Soto, 2022). In construction, the problem is more constrained to the project level. Total material demand is often fixed by design, and performance differences among alternatives arise from differences in per-unit costs, emissions intensity, and operational losses, such as material wastage. Transportation choices further complicate the trade-off, as frequent deliveries may improve responsiveness but generate higher emissions (Suhaib, Amir Babak Rasmi and Türkay, 2023).

In response, GSCM enhances conventional supply chain management by including environmental impacts into procurement, production, transportation, and waste management decisions. In the construction industry, GSCM methods typically encompass the adoption of low-carbon resources, energy-efficient technology, optimal logistics planning, and minimisation of material waste (Farida, Handayani and Wibowo, 2019; Wibowo, Handayani and Farida, 2022).

A variety of techniques have been utilised to assess green supply chain effectiveness. Life Cycle Assessment (LCA) is extensively employed to measure environmental consequences throughout the supply chain and identifies emission hotspots in construction materials and processes (Backes and Traverso, 2022). However, LCA predominantly emphasises environmental factors but does not explicitly include economic efficiency or the trade-offs among various performance metrics.

Multi-Criteria Decision-Making (MCDM) methodologies, like the Analytic Hierarchy Process (AHP), TOPSIS, have been widely utilised to assess green supply chain options by consolidating various criteria into a singular score(Paul *et al.*, 2021). Although these systems provide flexibility, they depend on predetermined or gathered weights to reflect decision-making preferences. Numerous assessments indicate that MCDM outcomes are significantly influenced by weighting methods, potentially introducing subjectivity and constraining transparency, especially when objective benchmarking is necessary (Paul *et al.*, 2021)

In project-based construction environments, where possibilities are evaluated under set demands and standardised functional units, dependence on subjective weighting is a considerable constraint. Thus, there is a necessity for performance evaluation methodologies that may objectively assess many options based on empirical performance data without necessitating prior determination of preference weights.

DEA gives a non-parametric framework for assessing the relative efficiency of analogous decision-making units utilising various inputs and outputs without predetermined weights (Charnes, Cooper and Rhodes, 1978; Banker, Charnes and Cooper, 1984). The DEA formulates an efficiency frontier based on observable data, designating efficient units as benchmarks for their inefficient counterparts, while also measuring prospective improvement objectives.

DEA has been extensively utilised in energy, environmental, and sustainability research to evaluate eco-efficiency by including costs, energy consumption, emissions, and resource losses as inputs, while financial performance indicators serve as outputs (Zhou, Ang and Poh, 2008). In supply chain research, DEA has been employed to assess companies, logistics networks, and supply chain designs, providing insights about efficiency rankings and opportunities for enhancement (Zhou, Ang and Poh, 2008).

In contrast to LCA and MCDM methodologies, DEA is especially appropriate for objective benchmarking scenarios that need the simultaneous assessment of numerous quantitative metrics. Its capacity to eliminate subjective weighting and to offer reference sets and proportional improvement metrics renders it very appropriate for assessing the green supply chain for the construction industry, contexts defined by fixed demand and various efficiency determinants. Despite existing contributions, a distinct vacuum persists in the systematic benchmarking of various GSCM scenarios within construction project-based frameworks. Current research predominantly applies descriptive comparisons, utilises MCDM methods with subjective weighting, or focuses on firm-level efficiency

assessments instead of scenario-based evaluations under constant project demand.

This paper proposes a DEA-based framework for assessing the efficiency of green supply chains to solve this gap. The approach utilises an input-oriented VRS-DEA model to benchmark literature-informed scenarios, facilitating objective comparisons of economic and environmental performance, identification of efficient configurations, and quantification of prospective improvements. This enhances both the theoretical and practical aspects of decision-making in GSCM for the construction industry.

Methodology

This study proposes a structured, DEA-based multi-stage framework to address the need for systematic, objective benchmarking methods for assessing GSCM scenarios in the context of project demand. The methodology aims to facilitate transparent scenario construction, minimise dependence on subjective performance weighting, and promote quantitative efficiency benchmarking across various supply chain configurations.

The proposed methodological framework shown in Figure 1 combines literature-informed parameter identification, scenario development, DEA model formulation, computational implementation, improvement potential analysis, and decision support interpretation to facilitate a comprehensive and replicable assessment of GSCM efficiency.

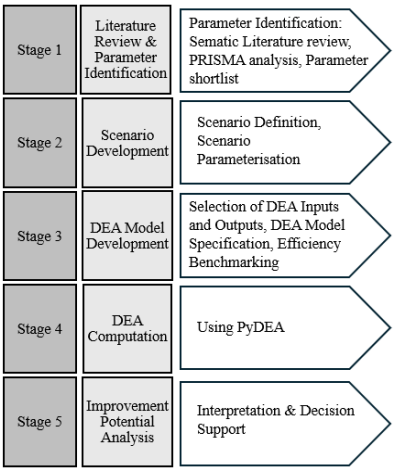


Figure 1: DEA-Based Methodological Framework

Stage 1: Literature Review & Parameter Identification

A systematic literature review using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was performed to identify performance parameters associated with the efficiency of GSCM(Page *et al.*, 2021). Relevant studies were sourced from prominent academic databases such as Google Scholar, Web of Science, and Scopus. Based on their applicability to construction supply chains and their utilisation of quantitative performance indicators. From the collected literature, economic, environmental, and operational parameters frequently used to assess green supply chain performance were extracted and categorised. Table 1 represents the shortlisted Economic, Environmental, and Operational parameters with their sources. The selection of parameters was guided by their relevance to GSCM objectives, the availability of quantitative benchmarks, and their appropriateness for efficiency-based evaluation. This approach ensures that scenario development and subsequent DEA analysis are anchored in established literature rather than arbitrary assumptions.

Table 1: Shortlisted Parameters

Category	Parameter	Description	Source
Economic	Production cost (₹/t)	Costs associated with material manufacturing and processing	1,2,3
	Transport cost (₹/t)	Cost of conveying materials from the origin to the site	1,2,3,4
	Carbon cost/Tax (₹/t)	Costs related to carbon pricing or emissions Taxes	7,8
	Net profit (₹/t)	The financial worth produced per unit of material provided	1,2,3
Environmental	CO_2 emissions (production) (tCO_2/t)	Emissions produced during the manufacturing of materials	4,5,6
	CO_2 emissions (transport) (tCO_2/t)	Emissions produced during the transportation of materials	4,5,6
Operational	Material wastage (%)	Percentage of material wasted due to handling and inefficiencies	9,10

1.(Khaiyum, Sarker and Kabir, 2023), 2.(Castellano *et al.*, 2023), 3.(Bhagwat and Sharma, 2007), 4.(Francis and Thomas, 2020a), 5.(GCCA India-TERI, 2025), 6.(Okay, Sencer and Taskin, 2024),7. (Pan *et al.*, 2024),8.(Santos, 2022),9.(Gulghane and Khandve, 2015), 10.(Golzarpoor and González, 2013)

Stage 2: Scenario Development

This study characterises a scenario as a unique construction supply chain configuration distinguished by varying operational and environmental management approaches, with each scenario considered as a Decision Making Unit (DMU) within the DEA framework. Scenarios are designed deterministically to represent realistic and comparative supply chain options. A case study construction project is utilised to ensure practical relevance, with the corresponding cement demand serving as the analytical foundation. All scenarios are constructed with regard to this case study.

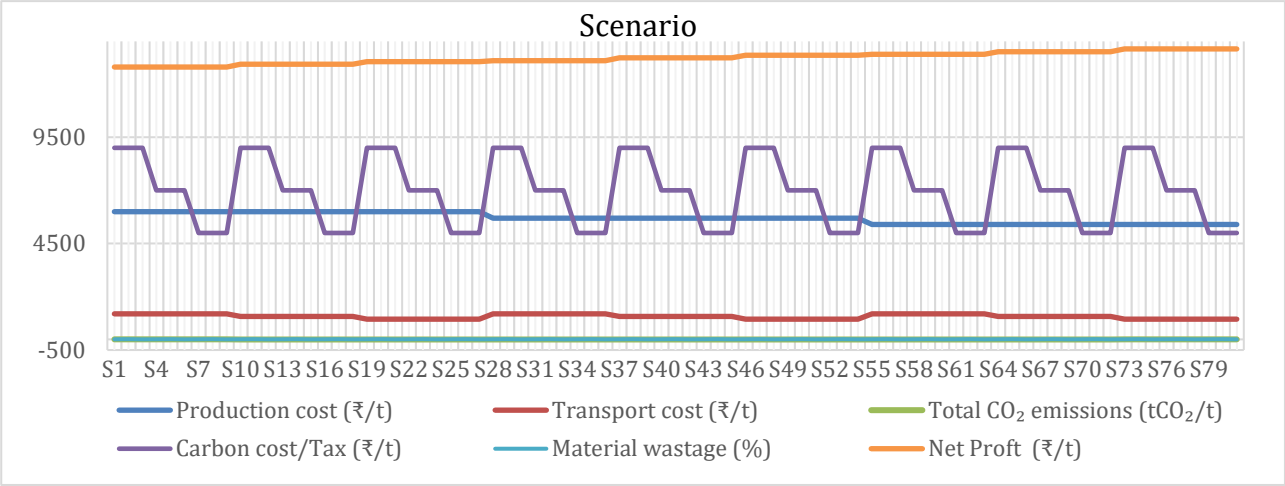


Figure 2: GSCM Scenario Design for DEA-Based Evaluation

A constant project demand assumption is used across all scenarios, representing standard construction conditions in which material quantities are dictated by design and scheduling limitations. This guarantees that the observed discrepancies in efficiency stem from differences in resource utilisation rather than scale effects. A total of 81 scenario archetypes are developed to assess performance across various management strategies, reflecting systematic differences in cost structure, emission intensity, material waste, logistics efficiency, and carbon cost internalisation. Figure 2 shows a graphical representation of all developed scenarios.

Table 2: Parameter Range Summary Table

Parameter	Max	Mid	Min
Production cost (₹/t)	6000	5700	5400
Transport cost (₹/t)	1200	1075	950
Carbon cost/Tax (₹/t)	9000	7000	5000
CO_2 emissions production (tCO_2/t)	0.9	0.65	0.54
CO_2 emissions transportation (tCO_2/t)	2.02	1.9	0.65
Material wastage (%)	5	3.5	2

Table 2 provides a parameter-range summary used in scenario development. The parameter values were selected from industry evidence and published studies. Wastage percentages were defined using industry evidence. Gulghane and Khandve (2015) report that cement wastage in Indian construction typically ranges from 2–3% in well-managed sites, while poorly managed projects experience higher levels. Other studies note wastage levels of 2.6–4% on active projects, which justifies a maximum assumption of 5% and a reduction to 2% for scenarios (Golzarpoor and González, 2013). Cost assumptions were anchored in UltraTech Cement's financials for 2025, which report an average realisation of ₹5,052 per tonne and logistics costs of ₹1,167 per tonne (UltraTech Cement Limited, 2025). Scenarios apply 5–10% premiums in line with early adoption of low-carbon cement, which global benchmarks suggest may be 50–45% more expensive. Finally, CO_2 emission parameters were aligned with sectoral roadmaps.

Francis and Thomas (2020b)shows that Indian housing can achieve up to 40% reductions through material substitution and policy measures, while the Decarbonization Roadmap for the Indian Cement Sector: Net-Zero CO_2 by 2070 (GCCA India-TERI, 2025), outlines feasible reductions of 35–50% by mid-century. Accordingly, the parameter value applies a 50% reduction for both production and site consumption emissions, with incremental reductions at Low and Medium levels.

Carbon cost numbers were articulated in Indian Rupees per tonne of CO_2 to ensure compatibility with project cost data. International carbon tax rates and emissions trading prices indicate that carbon pricing levels often vary from below USD 1/tCO_2 to over USD 100/tCO_2(Santos, 2022; Pan *et al.*, 2024; United Nations Environment Programme and Global Alliance for Buildings and Construction, 2025). Utilising a realistic currency rate, this equates to around ₹100–₹10,000 per tCO_2. Consequently, low, medium, and high carbon cost levels were established within this spectrum to represent voluntary carbon markets, average compliance pricing, and stringent policy scenarios, respectively. While India does not presently enforce a comprehensive carbon tax, this study incorporates carbon costs to account for evolving international carbon pricing frameworks and prospective regulatory conditions, facilitating a forward-looking assessment of the efficiency of GSCM.

Stage 3: DEA Model Development

The selection of DEA inputs and outputs is based on GSCM concepts and the attributes of project-based construction systems. Inputs signify unfavourable economic, environmental, and operational constraints that need to be reduced, and outputs denote favourable economic results.

Each construction supply chain scenario is regarded as a DMU defined by a vector of inputs and outputs. The DEA input set comprises production cost (₹/t), transport cost (₹/t), carbon cost/tax (₹/t), total CO_2 emissions (tCO_2/t),

and material wastage (%). Net profit (₹/t) is designated as the primary output variable as it represents an aggregate performance indicator reflecting the combined effects of production cost, transport cost, carbon cost, and material wastage, which are explicitly modelled as inputs. The quantity of material is omitted as an output variable because project demand remains constant across all scenarios and therefore does not influence efficiency differentiation. All variables are articulated on a per-tonne basis to guarantee comparability and eliminate scale-induced error. Table 3 represents inputs and outputs used in DEA.

Table 3: DEA Inputs and Outputs

Variable type	Parameter	Unit	DEA role
Input	Production cost	₹/t	Minimized
Input	Transport cost	₹/t	Minimized
Input	Carbon cost	₹/t	Minimized
Input	Total CO_2 emissions	tCO_2/t	Minimized
Input	Material wastage	%	Minimized
Output	Net profit	₹/t	Maximized

An input-oriented DEA model is used to assess the relative efficiency of green supply chain scenarios. The aim is to reduce economic and environmental burdens while sustaining the same level of revenue generation. This viewpoint aligns with construction project conditions, where output quantities are limited by fixed demand and efficiency enhancements are primarily realised through reductions in costs, emissions, and waste.

VRS are assumed using the Banker–Charnes–Cooper (BCC) DEA model, which allows efficiency to vary with scale and isolates pure technical efficiency from scale effects (Charnes, Cooper and Rhodes, 1978; Banker, Charnes and Cooper, 1984).

DEA Multiplier Formulation,

For a given scenario, the DEA efficiency score is obtained by solving the following linear programming problem:

$$Max\ E_o = \sum_{r=1}^{s} u_r y_{ro}$$

Subject to,

$$\sum_{r=1}^{s} v_i x_{io} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} \leq \sum_{i=1}^{m} v_i x_{ij}, \quad \forall_j = 1,2, \dots \dots \dots, n$$

$$u_r \geq 0, \quad v_{i \geq 0}$$

The objective function maximises the weighted sum of outputs for scenario o, normalised by the weighted sum of its inputs. The second constraint ensures that the identical combination of weights does not yield an efficiency score exceeding 1 in any other scenario, thereby establishing the efficiency frontier.

For computational implementation, the equivalent input-oriented VRS envelopment model is used:

$$Min\ \theta_o$$

Subject to,

$$\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_o x_{io}, \forall_i = 1,2, \dots \dots, m$$

$$\sum_{j=1}^{n} \lambda_j x y_{ij} \leq y_{ro}, \forall_r = 1,2, \dots \dots, s$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0$$

This study employs DEA formulations based on the basic input-oriented multiplier and envelopment models initially introduced by Charnes et al., (1978) and then modified to VRS by Banker et al., (1984). Table 4 shows a description of the notation.

Table 4: Notation and Variable Description

Symbol	Description	Unit	Corresponding parameter
n	Number of scenarios (DMUs)	–	81 scenarios
m	Number of inputs	–	5 inputs
s	Number of outputs	–	1 output
x_{1j}	Production cost of scenario (j)	₹/t	Production cost
x_{2j}	Transport cost of scenario (j)	₹/t	Transport cost
x_{3j}	Carbon cost of scenario (j)	₹/t	Carbon cost
x_{4j}	Total CO_2 emissions of scenario (j)	tCO_2/t	Total CO_2 emissions
x_{5j}	Material wastage of scenario (j)	%	Material wastage
y_{1j}	Net profit of scenario (j)	₹/t	Net profit
u_r	Weight assigned to output (r)	–	DEA output weight
v_i	Weight assigned to input (i)	–	DEA input weight
λ_j	Intensity variable	–	Peer contribution
θ_o	Efficiency score of scenario (o)	–	DEA efficiency

Stage 4: DEA Computation Using PyDEA

The DEA model outlined in the preceding part is executed using PyDEA, a Python-based application for data envelopment analysis (PyDEA Developers, 2023). PyDEA calculates efficiency scores by resolving a linear programming problem for each scenario and determines efficient benchmarks and reference sets. These outputs serve as the foundation for later efficiency benchmarking

and analysis of improvement possibilities, which will be addressed in the upcoming section.

Validation and Robustness Checks

This study emphasises the credibility and robustness of the proposed DEA-based framework over prediction accuracy. Conceptual validity is achieved by basing scenario development and parameter selection on recognised literature in GSCM and construction supply chains. All scenarios represent realistic and comparable configurations under constant project demand.

A sensitivity study was performed to evaluate the robustness of the DEA results by adjusting key input parameters within ±10% of their baseline values. The efficiency rankings and benchmark scenario identifications were predominantly consistent, suggesting that the DEA results are not affected by slight data fluctuations. These evaluations increase confidence in the reliability of the proposed framework and its relevance for decision assistance in GSCM planning for construction projects.

Stage 5: Improvement Potential Analysis

The next phase of the methodology focuses on synthesising DEA outputs to facilitate decision-making in GSCM planning for the construction industry. Efficient scenarios are recognized as benchmark configurations exemplifying best practices under constant project demand, whereas inefficient scenarios are assessed in comparison to their reference counterparts and their potential for improvement. The next part discusses detailed efficiency scores, benchmarking correlations, and opportunities for improvement.

Results and Discussion

The proposed DEA-based framework assessed 81 scenario-based DMUs representing different green supply chain configurations in the construction industry. Figure 3 represents the efficiency score across all scenarios with the considered parameters. All scenario configurations were developed within constrained and literature-informed parameter ranges to ensure realistic techno-economic and environmental relationships across production and transport stages. Production costs, transportation costs, transportation and production carbon emissions, carbon cost levels, and material wastage levels for the economic and environmental inputs are integrated into each scenario. Performance is evaluated using an economic outcome indicator under uniform functional assumptions. With scenarios S1–S27 obtaining an efficiency score of roughly 0.727, scenarios S28–S54 achieving an intermediate efficiency of roughly 0.88, and scenarios S55–S81 forming the efficient frontier with an efficiency score equal to 1, the resulting DEA efficiency scores indicate a structured pattern.

Scenarios S1–S27 exhibit high-intensity production configurations, featuring a production cost of ₹8,800 per tonne and production-related emissions of 0.90 tCO_2 per tonne, reflecting typical high-emission production regimes where higher emissions are associated with less efficient and more resource-intensive processes. Despite the reduction of wastage from 5% to 2% and decreasing the carbon tax from ₹9,000 to ₹5,000 per tCO_2e, these scenarios continue to exhibit relative inefficiency in terms of DEA. An efficiency score of 0.727 indicates that, on average, these scenarios necessitate a 27% decrease in their aggregate economic and environmental inputs to align with the performance of frontier scenarios. From a GSCM perspective, this outcome suggests that downstream green practices, such as waste minimisation and logistics optimisation, enhance absolute performance by reducing material losses and carbon emissions; however, their efficacy is limited when upstream production technologies are significantly carbon- and resource-intensive.

Scenarios S28–S54 indicate medium-intensity, characterised by a production cost of ₹7,200 per tonne and decreased production emissions of 0.65 tCO_2 per tonne. These scenarios attain an efficiency score of roughly 0.88, indicating the synergistic impact of moderate emission intensity and enhanced cost structure. This also reflects a realistic transition regime where improvements in emission intensity are accompanied by more efficient cost structures rather than conflicting trade-offs. Within this cohort, fluctuations in wastage and transportation emissions result in quantifiable enhancements in net profit and carbon cost liability. For instance, a 3 % point reduction in waste effectively reduces material losses and associated carbon costs, thereby improving economic performance even at elevated carbon tax rates. The results closely align with GSCM principles, indicating that operational greening and emission-control strategies significantly enhance supply chain performance when industrial systems are not overly carbon-intensive. Nevertheless, despite these advancements, none of the scenarios under this category attain the efficiency frontier, showing that even incremental greening cannot entirely offset the lack of low emissions during production.

Scenarios S55–S81 constitute the DEA efficient frontier, characterized by low-intensity production configurations, with a production cost of ₹6,400 per tonne and production emissions of 0.54 tCO_2 per tonne, alongside relatively reduced transport emissions. These configurations represent technologically feasible low-emission production pathways rather than idealised or unconstrained combinations of low cost and low emissions. Across all examined policy and operational situations, these configurations currently encompass various aspects of optimal green supply chain practices. Under this system, slight alterations in wastage or carbon tax do not lead to shifts in dominance, as the scenarios function near the least viable combination of economic and environmental costs within the specified scenario space. In the context of GSCM, these scenarios exemplify advanced, cohesive green supply chains where technological decisions, operational rigour, and regulatory readiness together enhance eco-efficiency.

The analysis shows that waste reduction and carbon taxes affect supply chain performance through production and logistics frameworks rather than as standalone efficiency

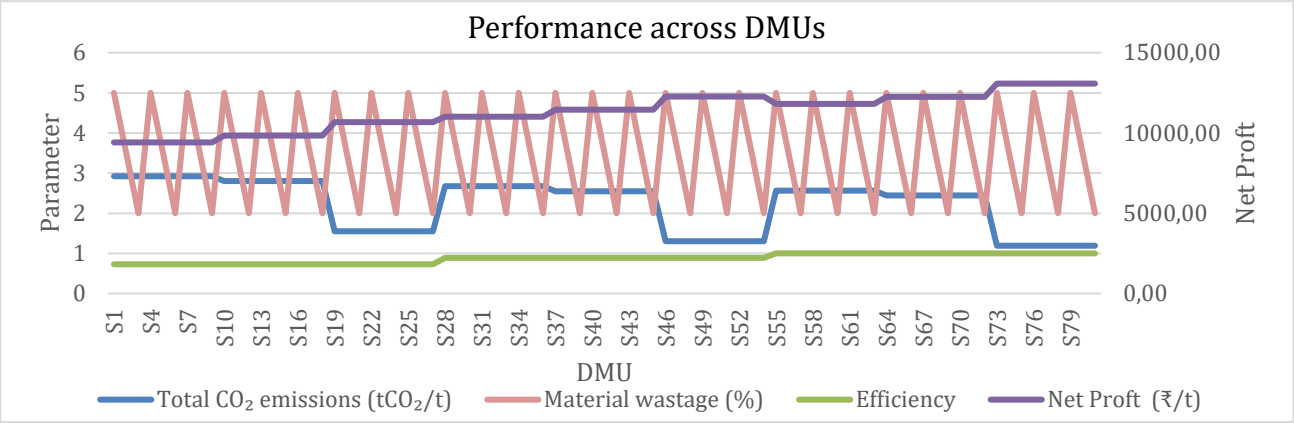


Figure 3: Integrated Environmental, Operational, and Economic Performance across DMUs

drivers. Under the fixed-revenue assumption, wastage and emission reductions improve absolute profitability and regulatory resilience but may not always improve relative efficiency scores compared to fundamentally cleaner setups. This supports GSCM theory, which distinguishes between incremental efficiency improvements within a technological regime and frontier-shifting transformations through low-carbon production and sourcing.

The results support the GSCM principle that technological, operational, and regulatory choices must align to sustain environmental advances. Performance and resilience improve with incremental greening, while structural investments in low-emission manufacturing and logistics redefine the efficiency frontier. This multi-scenario, multi-input–output framework ranks scenarios and reveals structural efficiency drivers, allowing decision-makers to compare supply chain configurations, identify non-dominated pathways, and understand how technology, operations, and policy changes may affect performance in a complex, sustainability-oriented decision environment.

Conclusions

This study proposes a structured, scenario-based DEA framework designed for project-level green construction supply chain decision-making under constant demand conditions. The primary contribution lies not merely in the application of DEA, but in the systematic development and benchmarking of alternative supply chain configurations using multi-dimensional economic, environmental, operational, and regulatory parameters. By organising scenario-based supply chain strategies as decision-making units within a consistent functional boundary, the framework enables objective efficiency comparison without requiring subjective weighting of sustainability indicators. This project-oriented perspective provides a targeted analytical approach for evaluating green supply chain strategies in construction contexts where material demand is design-driven and efficiency improvements must be assessed across feasible operational and technological alternatives.

The framework is useful for decision-makers in complex green supply chains where cost efficiency, carbon emissions, material waste, and regulatory demands are interdependent. Managers and policymakers can examine how production technology, transport emissions, wastage, and carbon taxation affect efficiency and net profit using multi-scenario analysis. Instead of offering a single solution, the framework shows dominance relationships and improvement potentials, helping stakeholders distinguish between operational and fundamental technology inefficiencies. This helps decide whether greening should focus on incremental operational improvements or deeper transitions to low-emission industrial technology.

The results are relative efficiency outcomes under fixed modelling assumptions and scenario-based inputs; therefore, they enable decision-making rather than prescriptive optimization for real-world implementation. While the use of a single output simplifies the analysis, it enables clearer interpretation of relative efficiency, with scope for extending the framework to multi-output DEA formulations in future work. In practice, demand variability due to design changes or schedule pressure can alter resource utilisation, cost structures, and emission intensity, potentially affecting relative efficiency rankings. Future research can address this limitation by extending the model using stochastic or dynamic DEA approaches that incorporate probabilistic demand variation, time-dependent resource allocation, and uncertainty in economic and environmental parameters.

References

Backes, J. G. and Traverso, M. (2022) 'Life Cycle Sustainability Assessment in the construction sector – actual application and future outlook', *E3S Web of Conferences*, 349, p. 04007. doi: 10.1051/e3sconf/202234904007.

Banker, R. D., Charnes, A. and Cooper, W. W. (1984) 'Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis', *Management Science*, 30(9), pp. 1078–1092. doi: 10.1287/mnsc.30.9.1078.

Bhagwat, R. and Sharma, M. K. (2007) 'Performance measurement of supply chain management: A balanced scorecard approach', *Computers and Industrial Engineering*, 53(1), pp. 43–62. doi: 10.1016/j.cie.2007.04.001.

Castellano, D. *et al.* (2023) 'Optimizing production-inventory replenishment and lead time decisions under a fill rate constraint in a two-echelon sustainable supply chain with quality issues', *International Journal of Systems Science: Operations and Logistics*, 10(1). doi: 10.1080/23302674.2023.2173540.

Cataldo, I. *et al.* (2022) 'SUSTAINABLE SUPPLY CHAIN MANAGEMENT IN CONSTRUCTION: AN EXPLORATORY REVIEW FOR FUTURE RESEARCH', *JOURNAL OF CIVIL ENGINEERING AND MANAGEMENT*, 28(7), pp. 536–553. doi: 10.3846/jcem.2022.17202.

Charnes, A., Cooper, W. W. and Rhodes, E. (1978) 'Measuring the efficiency of decision making units', *European Journal of Operational Research*, 2(6), pp. 429–444. doi: 10.1016/0377-2217(78)90138-8.

Chen, Q., Feng, H. and Garcia de Soto, B. (2022) 'Revamping construction supply chain processes with circular economy strategies: A systematic literature review', *Journal of Cleaner Production*, 335, p. 130240. doi: 10.1016/j.jclepro.2021.130240.

Farida, N., Handayani, N. U. and Wibowo, M. A. (2019) 'Developing Indicators of Green Construction of Green Supply Chain Management in Construction Industry: A Literature Review', in *IOP Conference Series: Materials Science and Engineering*. Institute of Physics Publishing. doi: 10.1088/1757-899X/598/1/012021.

Francis, A. and Thomas, A. (2020a) 'Decoupling Housing Growth from Environmental Impact: A Pathway to Sustainability through a System Dynamics Simulation Approach', in *Construction Research Congress 2020*. Reston, VA: American Society of Civil Engineers, pp. 268–277. doi: 10.1061/9780784482858.030.

Francis, A. and Thomas, A. (2020b) 'Exploring the relationship between lean construction and environmental sustainability: A review of existing literature to decipher broader dimensions', *Journal of Cleaner Production*, 252, p. 119913. doi: 10.1016/j.jclepro.2019.119913.

GCCA India-TERI (2025) *Decarbonization Roadmap for the Indian Cement Sector Net-Zero CO 2 by 2070*.

Golpîra, H. and T. E. B. (2021) 'Optimization for Construction Supply Chain Management', *Application of Mathematics and Optimization in Construction Project Management*, pp. 165–183. Available at: https://link.springer.com/chapter/10.1007/978-3-030-81123-5_7 (Accessed: 21 August 2024).

Golzarpoor, H. and González, V. (2013) 'A green-lean simulation model for assessing environmental and production waste in construction', in *21st Annual Conference of the International Group for Lean Construction 2013, IGLC 2013*, pp. 845–854.

Gulghane, A. A. and Khandve, P. V (2015) *Management for Construction Materials and Control of Construction Waste in Construction Industry: A Review*, *Journal of Engineering Research and Applications www.ijera.com*. Available at: www.ijera.com.

Khaiyum, M. Z., Sarker, S. and Kabir, G. (2023) 'Evaluation of Carbon Emission Factors in the Cement Industry: An Emerging Economy Context', *Sustainability*, 15(21), p. 15407. doi: 10.3390/su152115407.

Okay, N. C., Sencer, A. and Taskin, N. (2024) 'Quantitative indicators for environmental and social sustainability performance assessment of the supply chain', *Environment, Development and Sustainability*. doi: 10.1007/s10668-024-05210-3.

Page, M. J. *et al.* (2021) 'The PRISMA 2020 statement: an updated guideline for reporting systematic reviews', *BMJ*, p. n71. doi: 10.1136/bmj.n71.

Pan, J. *et al.* (2024) 'To tax or to trade? A global review of carbon emissions reduction strategies', *Energy Strategy Reviews*, 55, p. 101508. doi: 10.1016/j.esr.2024.101508.

Paul, A. *et al.* (2021) 'Sustainable Supply Chain Management and Multi-Criteria Decision-Making Methods: A Systematic Review', *Sustainability*, 13(13), p. 7104. doi: 10.3390/su13137104.

PyDEA Developers (2023) *PyDEA: A Python package for data envelopment analysis*. Available at: https://github.com/araith/pyDEA (Accessed: 28 January 2026).

Santos, G. (2022) 'Climate change policy and carbon pricing', *Energy Policy*, 168, p. 112985. doi: 10.1016/j.enpol.2022.112985.

Suhaib, Amir Babak Rasmi, S. and Türkay, M. (2023) 'Sustainability analysis of cement supply chains considering economic, environmental and social effects', *Cleaner Logistics and Supply Chain*, 8, p. 100112. doi: 10.1016/j.clscn.2023.100112.

UltraTech Cement Limited (2025) *Leading with Scale, Growing in Strength and Winning with Purpose UltraTech Cement Limited*.

United Nations Environment Programme and Global Alliance for Buildings and Construction (2025) *Global Status Report for Buildings and Construction 2024/2025*. United Nations Environment Programme. doi: 10.59117/20.500.11822/47214.

Wibowo, M. A., Handayani, N. U. and Farida, N. (2022) 'Developing Indicators of Green Operation and Maintenance of Green Supply Chain Management in Construction Industry', in, pp. 1193–1206. doi: 10.1007/978-981-16-7924-7_77.

Zhou, P., Ang, B. W. and Poh, K. L. (2008) 'A survey of data envelopment analysis in energy and environmental studies', *European Journal of Operational Research*, 189(1), pp. 1–18. doi: 10.1016/j.ejor.2007.04.042.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DEFINING DEMONSTRATOR PROJECTS FOR DIGITAL FABRICATION (DFAB) ADOPTION IN ARCHITECTURE, ENGINEERING AND CONSTRUCTION (AEC)

Jennifer Li[1], Konrad Graser[2], Alexander N. Walzer[3], and Daniel M. Hall[1]
[1]Technical University of Delft, Delft, the Netherlands
[2]Zurich University of Applied Sciences, Zurich, Switzerland
[3]University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland

Abstract

Digital Fabrication (DFAB) proposes to increase productivity and digitalisation in Architecture, Engineering and Construction (AEC). Demonstrator projects are used to showcase novel DFAB innovations to disseminate knowledge, learning, and attract future investment to support their transfer from research to commercialisation. Currently, no unified definition of what constitutes a DFAB demonstrator in AEC exists, thus, this paper aims to address that. This offers distinction to the phenomenon reflecting its real-world context, subsequently effecting how humans think and act on it. A definition is proposed based on analysis from literature and interviews to inform future studies on DFAB demonstrator projects in AEC.

Introduction

Architecture, Engineering and Construction (AEC) is the world's largest industry by share of gross domestic product. However, it is also ranked as one of the lowest in terms of productivity and digitalisation (Ribeirinho *et al.*, 2020). The complexity of AEC sets it apart from other industries (e.g., manufacturing, automotive) due to the individuality of each built asset's design and construction needs based on unique (site) conditions and stakeholder requirements (Garcia De Soto *et al.*, 2018). One systemic innovation proposed as a solution to increase productivity and digitalisation in AEC is Digital Fabrication (DFAB). DFAB is pushing the boundaries of creativity, design and construction of the built environment and is seen as a building block of realising the digital transformation of AEC into Construction 4.0 (Graser, 2023). Examples of DFAB in practice include the MX3D Printed Bridge, a robot printed steel structure once located in Amsterdam's Red Light District in the Netherlands (Figure 1); and DFAB House, the world's most comprehensive demonstration of DFAB in the AEC domain (Figure 2).

DFAB is "*a fabrication or building process driven by the seamless conversion of design and engineering data into fabrication code for digitally controlled tools*" (Gershenfeld, 2012, cited in Kaseman and Graser, 2020, p. 189). A relatively new field of innovation, DFAB was initiated to AEC in Japan in the 1980s with the deployment of robotics systems to automate and standardise the construction of buildings (Kaseman and Graser, 2020). It is described as a system-level innovation



Figure 1: MX3D Printed Bridge, Amsterdam, The Netherlands (Arup, 2025)

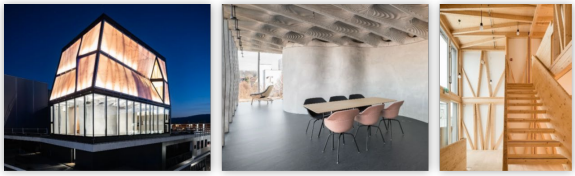


Figure 2: DFAB House, Dübendorf, Switzerland (Graser, 2023)

that combines digital design with advanced manufacturing technologies to automate construction activities and enhance the quality of built assets (Ng, Graser and Hall, 2023). Today, it is seen as a potential solution to solve some of the AEC industry's most pressing issues, such as the need to build more sustainably (Khadim *et al.*, 2023), increase productivity in light of a stagnant workforce (Mischke *et al.*, 2024), and increase digitalisation (Alsharef *et al.*, 2024) to increase efficiency. Digital technology adoption has been found to have positive effects on economic, social and environmental performance (Li *et al.*, 2022).

Despite the positive outlook for DFAB, is has yet to achieve mainstream adoption across the AEC industry due to a "*large gap between research feasibility and industry adoption*" (Graser, Kahlert and Hall, 2021, p. 853). This gap is often related to socio-technical factors such as business models and processes, integration into existing practices, product architecture and organisational structures (Graser, Kahlert and Hall, 2021). Ng et al. (2022) offer several reasons for this including the need for a new level of systemic thinking to seamlessly integrate the design and construction processes that requires "*more detailed knowledge, participation, and computation than*

traditional craft production", and the lack of understanding of DFAB as a "*standardized process or an integrated digital system*" (Ng *et al.*, 2022, p. 1).

Demonstrator projects

Demonstrator projects are one way the prevailing challenges of integrating DFAB into AEC praxis are being addressed. They constitute a phase in technology development that provides a vehicle to showcase technologies, methods and processes beyond the research boundary to attract interest and further investment with the goal of transferring innovations to industry and commercialisation. Moultrie (2015) describes demonstrator projects as *boundary objects* acting as a mediator between different social actors. They typically embody some level of risk (e.g., technical, production, commercial, operational), and require substantial investment (Aaron and McGill, 2012). They are receiving increasing attention as evident in literature where many demonstrator projects are published (e.g., in the FABRICATE series whose focus is on the integration of digital design and manufacturing processes (FABRICATE, 2025); and international exhibitions such as the International Architecture Exhibition, established in 1980 as part of La Biennale di Venezia, that showcases some of the world's cutting edge architectural solutions to address the most pressing challenges seen in society and the built environment (La Biennale di Venezia, 2024)). Despite their growing application to demonstrate the potential of new technologies, methods and processes, demonstrator projects are often discussed as failures as vehicles to facilitate transfer of innovations from research to industry (Sjöö and Frishammar, 2019), and there is limited literature on the objective, types and designs of demonstrator projects (Bobbe *et al.*, 2023). In addition, their focus is often on technical feasibility but omits the social aspects that are required for the success of a new innovation (Graser, 2023; Preston, Sujan and Habli, 2025).

Despite these challenges, application of demonstrator projects in the development of new technologies including DFAB is increasing. They are an effective method of exploration, their implementation increases acceptance (if not yet adoption) of DFAB in AEC, and they are instrumental in the establishment of a new praxis for DFAB (Graser, 2023). Given this growing attention, this paper aims to provide a comprehensive definition for demonstrator projects in the domain of DFAB in AEC. It does so by evaluating current understanding of demonstrator projects generally through literature, and engages with researchers and practitioners of DFAB in AEC specifically. Providing a useful definition is an important step in enhancing understanding of demonstrator projects' philosophical role for DFAB in the AEC community. Definitions provide "*distinctions that mirror common language use and real-world phenomena while paying special attention to context and implications*" (Mitcham and Schatzberg, 2009, p. 27). The act of "*sorting out, classifying, and defining […] remain fundamental to human acting and thinking*" (Mitcham and Schatzberg, 2009, p. 56). The research question for this paper is therefore: *How do we define demonstrator projects for DFAB in AEC to enhance understanding of their purpose in the transfer of DFAB innovations from research to commercialisation?*

The remainder of this paper is structured as follows: the next section presents the methodology adopted for this paper, then definitions for demonstrator projects are presented. Following analysis, a comprehensive definition for DFAB demonstrator projects in AEC is proposed and its implications discussed. Finally, the paper is concluded.

Methodology

As there is little literature available on the topic of demonstrators and even less on demonstrators in DFAB for the AEC industry, this research adopts an interpretivist epistemology to uncover new theory around the practice of demonstrators. The study adopts a qualitative research enquiry that engages with researchers and practitioners with experience of demonstrators in DFAB. The steps taken to conduct this study are shown in Figure 3.



Figure 3: Methodology

First, a traditional literature review was conducted to identify any existing definitions of demonstrator projects in DFAB and beyond. The criteria for selecting papers included: demonstrator projects as the main subject of the paper, inclusion of a definition of demonstrator projects, and related to the production of something physical (i.e., this study is not concerned with software demonstrations). Then, semi-structured interviews were conducted with participants of the DFAB House demonstrator to understand what they believed their notion of the purpose and effects of a demonstrator to be. The interviews were conducted as part of a wider longitudinal study of DFAB House and to establish a methodology for DFAB demonstrator projects in AEC. However, the scope of this paper is concerned only with defining demonstrators and, therefore, only one question related to this subject is included here: *How do you define a demonstrator project in the context of DFAB in AEC?* DFAB House was chosen as the case study as the most comprehensive representation of the use of DFAB in AEC to date (Graser *et al.*, 2020). DFAB House combines six different DFAB objects as a demonstration of some of the leading advances in DFAB showcasing its feasibility and potential to the global community.

The sample included eight academics and one client (specific roles can be seen in Table 1), selected for their work on DFAB House during its design, development and/or construction, and because some were previously interviewed for a study on DFAB House shortly after its completion as they form part of the longitudinal study. The interviews were recorded and transcribed in Microsoft Teams and the interviewees were asked to approve the transcript post processing. Relevant

Table 1: Definitions of demonstrator projects in literature

#	Reference, domain	Definitions
R1	Aaron and McGill (2012), defence/public services	"short term projects (or project activities) that help demonstrate the level of maturity of technologies in support of capability management decisions" (UK Ministry of Defence cited in Aaron and McGill, 2012, p. 2).
R2	Bobbe et al. (2023), Internet of Things (IoT) and robotics	"(a) an embodiment or representation, (b) of a hypothesis about the future and (c) that is communicated to a specific audience" (p. 4); "an artefact which is created during the scientific process to support science itself, but also the dissemination and communication of it" (Moultrie, 2015, cited in Bobbe *et al.*, 2023, p. 2).
R3	Graser (2023), AEC	"tools of digital fabrication research and development: Demonstrator projects implemented on the construction scale span the boundaries of research and implementation practice, usually focusing in-depth on a single building system or fabrication technology" (p. 7).
R4	Husain, Wehrmeyer and Reeby (2023), demonstrator projects	"the process by which new thinking is tried and tested at scale and proactively communicated to target audiences, in order to improve processes, products, and services" (p. 32).
R5	Karlström and Sandén (2004), fuel cells and hydrogen supply systems for public transport vehicles	"a finite initiative to test hypotheses about a technology" (p. 287); "a finite initiative to test a technology according to the project objectives. […] The project starts with a conceptual design and by deciding project objectives. It ends when the technology is implemented and the results are evaluated and communicated" (p. 287).
R6	Moultrie (2015), design	"technological prototypes which are close to market" (p. 2).
R7	NASA (2023), space exploration	TRL 6: "System/sub-system model or prototype demonstration in an operational environment"; TRL 7: "System prototype demonstration in an operational environment".
R8	Sjöö and Frishammar (2019), sustainability transitions	"'technological infrastructures' often intended for multiple uses by a number of actors (rather than being geared towards one firm)" (p. 331); "they 'bridge' basic and small-scale knowledge generation in laboratories with large-scale industrial application and commercialization" (p. 331).
R9	Stelvaga and Fortin (2022), aerospace	"a project devoted to the development of a functional piece of technology with the goal of showing its capabilities" (p. 275).
R10	Sviridova and Verlinden (2023), product development and industrial design	"at least a partially physical one-off design object that facilitates the interests of designer, client, user, and technology, and communicate them to the audience in an interactive way" (p. 950).

responses to the question asked above were extracted and put into a table to allow for coding and content analysis to be performed along with analysis of the definitions found in literature. This consisted of finding recurring terms and phrases across both sets of definitions as well as identifying any terms and phrases that are applicable. The terms and phrases were then examined to establish whether they were indeed relevant for the domain of DFAB in AEC and the results were formulated into a comprehensive definition.

Defining demonstrator projects for DFAB in AEC

Finding a unified definition for DFAB demonstrator projects is challenging, in part because many projects have the intention of 'demonstrating', e.g., a new method or process (Husain, Wehrmeyer and Reeby, 2023), but they are not necessarily a demonstrator project per se. For example, a prototype project can be used to demonstrate something but the scale of development cannot yet be considered a demonstrator project. In current literature, terms are often interchangeable, such as in Wang, Hunhevicz and Hall (2025) who refer to both demonstrator and prototype, without clearly defining either; the Rijksdienst voor Ondernemend Nederland (the Netherlands Enterprise Agency) (RVO, 2025) indicates use of the Technology Readiness Levels (TRLs) system and refers to demonstration of a prototype (TRL 7), which mixes terms; and often demonstrators are not defined by papers that write about them, which could mean a) authors assume the concept is understood by the reader, or b) the authors are unsure themselves of how to define them.

The literature review conducted for this study found only one paper with a definition in the context of AEC. To supplement this, definitions were found in papers from other domains bringing the total to 10. The definitions along with the domain to which they belong can be seen in Table 1.

To ensure the study's ability to position demonstrators specifically in DFAB in AEC, interviews were conducted with people who had been involved in the DFAB House project. The participants responses are shown in Table 2.

Analysis and discussion

Across these two sets of definitions, there are several recurring themes that appear to be relevant to formulating a comprehensive definition for DFAB demonstrator projects in the AEC domain. Relevance was determined

Table 2: Direct quotes from interview participants on the definition of demonstrator projects for DFAB in AEC

Interviewee	Definitions
P1: project manager	"the use of a digital or digital fabrication technology, or method, on the full scale, meaning the scale of buildings, and I think it needs to be in an environment that resembles the process of construction"; "in the technical sense, a demonstrator is usually something you do once the technology is already at a certain [TRL] level and the demonstrator is basically your final proof".
P2: researcher	"the demonstrator is one way to prove the feasibility of a certain approach, maybe a technological methodology or a complicated workflow"; "first and foremost, geometric scalability".
P3: researcher	"an attempt to show that this is not just an academic exercise, and that there is the possibility to do something with a meaningful scaling"; "it has to leave the lab, but it can be produced in the lab".
P4: client	"takes technology, systems, materials, concepts from a laboratory level to a near market level".
P5: PI	"a real world installation involving the outcomes so that the finished research is demonstrated"; "something that goes beyond the function of a demonstrator; it is actually something that demonstrates by using the research outcomes"; "it would ideally be something that is then also in use"; "the confrontation of research with the real world scenario"; "a demonstrator should be out of the lab".
P6: PI	"an attempt to benefit from the challenges of transferring research to application very early on in research by actually making the full scale application of digital fabrication".
P7: PI	"it needs to be enough to leave the lab"; "there is an expectation to the outcome that is beyond only demonstrating literally a new technology, so that you add the challenge of giving it a function, either a more infrastructural and architectural and engineering kind of function perhaps"; "the threshold of a demonstrator project: did you have to face the real conditions? So even though the client is set up, even the budget might be not typical, but at least all the other aspects like building codes and stuff like that"; "demonstration is challenging your concept to a real scale".
P8: member of NCCR DFAB (role hidden to maintain anonymity)	"It's a demonstration of the innovation that you've created, but it's also a demonstration that you or the surrounding team around you know how to make that thing and make it ready for the real world"; "a demonstrator is a proof-of-concept, whereas a prototype is a development of an idea. A prototype is a working document, and a demonstrator is the document itself, it's the finalised version of the document"; "in DFAB, a demonstrator has industry partners working on it. That's a major difference between a prototype and a demonstrator, that you engage people who are going to learn from this demonstration already in the making of the demonstrator".
P9: member of NCCR DFAB (role hidden to maintain anonymity)	"a demonstrator project is whether it's with or without industry, but usually it's with industry, is that you upscale from the lab scale to the real scale, to the one-to-one construction site scale. And demonstrator project means actually to demonstrate that all the processes work out at the larger scale"; "to demonstrate that all the processes can work in reality given the building codes and other constraints".

by frequency of occurrence (set as a minimum of two) as well as recognition of importance from the authors based on their knowledge and experience of the topic. The results of the coding process are shown in Table 3 ordered in terms of frequency.

While many of the definitions from literature are specific to the subject of the individual publication, the material aspects of them can be applicable to other fields, i.e., AEC. The interviewees' definitions are more specific to the AEC context, but could equally be applicable outside of that if reference to, for example, buildings and the construction site are removed. The themes are described below in order of frequency as they appear in Table 3, not necessarily in order of importance.

Scale: many of the definitions explicitly refer to the scale of the demonstrator project mentioning full- or one-to-one scale. This scaling is what provides the technology, method or process the experiential learning of transitioning outside of the laboratory context, such as the requirements to move from research to production at scale for mainstream adoption and commercialisation.

Environmental context: demonstrators take place in an environment that exposes the potential innovation to real-world risks and challenges that cannot be experienced in a laboratory including but not limited to permitting, regulations, and site conditions.

Maturity and readiness: demonstrators provide an opportunity to demonstrate readiness of a technology, method or process, as well as an opportunity for learning during the project. The project forces the concept to be implemented in a real-world context that it has not yet seen in the laboratory. Thus, a concept may increase its maturity and readiness through the demonstrator process.

Inter- or trans-disciplinarity: as a boundary object, demonstrators bring different disciplinary actors together to co-create. At some point in the project, all parties contribute to the production of an artefact and through doing so contribute to the knowledge created and learning

Table 4: Coding of literature and interview definitions of demonstrators

Code	Occurrence	Frequency
Scale	R3 R4 R8 P1 P2 P3 P5 P6 P7 P9	10
Environmental context	R3 R7 R8 P1 P3 P5 P6 P7 P9	9
Maturity and readiness	R1 R6 P1 P2 P4 P8	6
Inter- or transdisciplinary	R3 R8 R10 P8 P9	5
Communication and dissemination	R2 R4 R5 R9 R10	5
Functionality	R8 R9 P5 P7	4
Novelty	R4 R10 P7	3
Social feasibility	R4 P7 P9	3
Temporality	R1 R5	2
(Hypothesis) testing	R2 R5	2

experienced. The level of disciplinarity often occurs at the higher end of the disciplinarity scale i.e., interdisciplinary or even transdisciplinary projects. This inevitable collision of social groups often occurs at the boundary between research and industry.

Communication and dissemination: demonstrators have the purpose of communicating and disseminating knowledge, learning and level of maturity of the technology, method or process beyond the project team to interested industry practitioners and potential future investors.

Functionality: demonstrators should produce something that has some level of functionality or use beyond the end of the demonstrator project (i.e., the construction phase of the demonstrator) that proves its utility in the real world over a (defined) period of time.

Novelty: a common theme across the definitions is novelty, that what is being demonstrated is offering something new to complement or replace existing praxis, or establish new praxis.

Social feasibility: the focus of demonstrators tends to be on the technical feasibility and often omits the social feasibility. However, as a boundary object, demonstrators force project participants to communicate outside of their day-to-day environment by considering the organisational context, business model(s) and process(es) of implementing a new technology, method or process along with any changes required to current operations.

Temporality: demonstrator projects are short term and finite endeavours.

(Hypothesis) testing: testing is central to demonstrator projects to test hypotheses about research and its ability to meet the project's objectives across many of the themes discussed above, such as scale, readiness, context and feasibility.

Based on the analysis here and the recurring themes identified across the definitions from literature and the interviews, the following definition for demonstrator projects in DFAB in the AEC domain is proposed:

> *Digital fabrication (DFAB) demonstrator projects in Architecture, Engineering and Construction (AEC) are finite endeavours of co-creation to materialise the maturity and readiness of a novel technology, method or process. This is done through production and testing of a functional artefact at full scale in a realistic, non-laboratory-based, construction-specific environment.*
>
> *A demonstrator project spans the boundaries of research and industry to generate knowledge of technical and social feasibility for active communication and dissemination to parties beyond the project team. The goals are to inform industry of the potentials of DFAB, and attract investment for future development to achieve transfer to commercialisation.*

To answer the research question: *How do we define demonstrator projects for DFAB in AEC to enhance understanding of their purpose in the transfer of DFAB innovations from research to commercialisation?*, this paper has collected definitions from literature across different sectors (due to a lack of definitions from within the AEC domain) and from AEC-specific researchers and practitioners to establish what the key aspects of DFAB demonstrator projects in the AEC industry are. This has ensured the definition proposed here encompasses the most important aspects of demonstrator projects to provide a comprehensive and useful definition to enhance understanding within research and outward to industry practitioners. In line with Mitcham and Schatzberg (2009), such a definition is important in the discourse on demonstrator projects for DFAB in AEC because definitions offer a distinction to a phenomenon that reflects its real-world context, that can then effect how humans think and act on it. While acceptance of a unified definition remains the longer-term goal for DFAB demonstrator projects in AEC, we propose here a definition designed to be immediately useful across disciplinary contexts and argue that such practical utility is a necessary condition for any definition to eventually achieve wider adoption.

The definition can solidify the purpose of conducting demonstrators for researchers and practitioners by understanding how they fit into the wider process of technology development. What comes before and after a demonstrator project is as important as the demonstration itself to ensure its success. This definition both enhances understanding of DFAB demonstrator projects and encourages a more holistic approach to the development of DFAB in AEC. This is particularly important if we approach it from a *needs* perspective. In a "push" scenario where academia is developing a new DFAB technology, it can support them in their efforts to seek industry buy-in. In an industry-led "pull" scenario where industry is proposing a technology to solve an industry issue, the definition can support them in seeking validation in collaboration with research.

Finally, highlighting that a demonstrator project is a process of co-creation encourages researchers and industry to focus collaboratively on this transitionary phase of technology transfer and the mechanisms required for it to be successful.

Conclusions

DFAB is a growing systemic innovation that could contribute to solving some of the AECs most pressing issues such as needing to build more sustainably and increase productivity and digitalisation. However, DFAB requires new systemic thinking to address some of its barriers to adoption, compounded by lack of a unified understanding of DFAB. One phase in the trajectory of technology development is demonstration to support the transfer of DFAB innovations from research to commercialisation. Demonstrator projects are still relatively new in the diffusion of innovations and as a result there is little academic discourse on the subject. This paper is part of a wider study to look at how demonstrators take place, how they create learning, how they play an important part of the research process in DFAB, and how they can be used to increase the transfer of innovations from research into industry and commercialisation. The aim of this paper was to define the role of demonstrators for DFAB in the AEC industry that will form the basis for the wider study, that is to develop a methodology for DFAB demonstrator projects in AEC.

The paper identified several definitions of demonstrator projects in literature, only one of which came from the AEC domain. These were supplemented by interviews with researchers and a client for the DFAB House project. These definitions were analysed and a comprehensive definition for DFAB demonstrator projects in AEC was proposed. This definition can provide a distinction to the phenomenon of DFAB demonstrators that can effect how humans think about and act on it. It is hoped that this definition will provide some consistency for DFAB in the AEC industry to support its understanding as a viable stage in technology development across research and industry and form the basis for a unified definition in the future.

Limitation and future research

While the wider study is a work-in-progress, this paper is limited by the small sample size and skew of interview participants toward the side of research, as opposed to providing a balanced view with the inclusion of more industry practitioners. This will be addressed in the next steps of the study by conducting further interviews with academics and industry practitioners with demonstrator experience, to evaluate the success of demonstrator projects using DFAB House as a case study, and to formulate a methodology for conducting DFAB demonstrator projects in AEC that can a) support students and researchers in their endeavours to innovate in DFAB, and b) support practitioners with the transfer of those innovations to industry and commercialisation.

Acknowledgments

This project was funded by the Swiss National Centre for Competence in Research (NCCR) Digital Fabrication.

References

Aaron, A. and McGill, I. (2012) 'A comparative review of Australian technology demonstration and commercialisation initiatives in sustainable energy'. Available at: https://www.ceem.unsw.edu.au/sites/default/files/documents/A%20comparative%20review%20of%20Australian%20technolog%20demonstration%20and%20commercialisation%20initiatives%20in%20sustainable%20energy%20final.pdf (Accessed: 13 August 2024).

Alsharef, A. *et al.* (2024) 'Biggest Challenges Facing the Construction Industry', pp. 652–660. Available at: https://doi.org/10.1061/9780784485286.065.

Arup (2025) *MX3D: Amsterdam's 3D printed steel bridge*. Available at: https://www.arup.com/projects/mx3d-bridge/ (Accessed: 3 February 2026).

Bobbe, T. *et al.* (2023) 'Design for communication: how do demonstrators demonstrate technology?', *Design Science*, 9, p. e3. Available at: https://doi.org/10.1017/dsj.2023.1.

DFAB House (2026) *DFAB HOUSE – Building with robots and 3D printers*. Available at: https://dfabhouse.ch/ (Accessed: 6 February 2026).

FABRICATE (2025) 'fabricate.org | Design & Making'. Available at: https://www.fabricate.org/ (Accessed: 14 November 2025).

Garcia De Soto, B. *et al.* (2018) 'Rethinking the roles in the AEC industry to accommodate digital fabrication', in *Creative Construction Conference 2018 - Proceedings*. *Creative Construction Conference 2018*, Budapest University of Technology and Economics, pp. 82–89. Available at: https://doi.org/10.3311/CCC2018-012.

Graser, K. *et al.* (2020) 'Dfab House: A Comprehensive Demonstrator of Digital Fabrication in Architecture', in *Fabricate 2020*. UCL Press (Making Resilient Architecture), pp. 130–139. Available at: https://doi.org/10.2307/j.ctv13xpsvw.21.

Graser, K. (2023) *From Invention to Innovation: Socio-technical Processes in the Adoption of Digital Fabrication to Architecture, Engineering and Construction Praxis*. Doctoral Thesis. ETH Zurich. Available at: https://doi.org/10.3929/ethz-b-000615698.

Graser, K., Kahlert, A. and Hall, D.M. (2021) 'DFAB HOUSE: implications of a building-scale demonstrator for adoption of digital fabrication in AEC', *Construction Management and Economics*, 39(10), pp. 853–873. Available at: https://doi.org/10.1080/01446193.2021.1988667.

Husain, S.A., Wehrmeyer, W. and Reeby, C. (2023) 'Critical success factors facilitating the propagation of demonstrator projects', *Socialis Series in Social Science*, 4, pp. 31–44. Available at: https://doi.org/10.20319/socv4.3144.

Karlström, M. and Sandén, B.A. (2004) 'Selecting and assessing demonstration projects for technology assessment: The cases of fuel cells and hydrogen systems in Sweden', *Innovation*, 6(2), pp. 286–293. Available at: https://doi.org/10.5172/impp.2004.6.2.286.

Kaseman, K. and Graser, K. (2020) 'Digital fabrication in the construction sector', in *Construction 4.0*. Routledge.

Khadim, N. *et al.* (2023) 'Whole building circularity indicator: A circular economy assessment framework for promoting circularity and sustainability in buildings and construction', *Building and Environment*, 241, p. 110498. Available at: https://doi.org/10.1016/j.buildenv.2023.110498.

La Biennale di Venezia (2024) *Biennale Architettura 2025 | Homepage 2025*, *La Biennale di Venezia*. Available at: https://www.labiennale.org/en/architecture/2025 (Accessed: 2 February 2026).

Li, Y. *et al.* (2022) 'Effects of Digital Technology Adoption on Sustainability Performance in Construction Projects: The Mediating Role of Stakeholder Collaboration', *Journal of Management in Engineering*, 38(3), p. 04022016. Available at: https://doi.org/10.1061/(ASCE)ME.1943-5479.0001040.

Mischke, J. *et al.* (2024) *Delivering on construction productivity is no longer optional*. McKinsey & Company. Available at: https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional (Accessed: 3 December 2025).

Mitcham, C. and Schatzberg, E. (2009) 'Defining Technology and the Engineering Sciences', in *Philosophy of Technology and Engineering Sciences*. Elsevier, pp. 27–63. Available at: https://doi.org/10.1016/B978-0-444-51667-1.50006-9.

Moultrie, J. (2015) 'Understanding and classifying the role of design demonstrators in scientific exploration', *Technovation*, 43–44, pp. 1–16. Available at: https://doi.org/10.1016/j.technovation.2015.05.002.

NASA (2023) *Technology Readiness Levels Definitions*. Available at: https://www.nasa.gov/wp-content/uploads/2017/12/458490main_trl_definitions.pdf (Accessed: 8 October 2025).

Ng, M.S. *et al.* (2022) 'Designing for Digital Fabrication: An Empirical Study of Industry Needs, Perceived Benefits, and Strategies for Adoption', *Journal of Management in Engineering*, 38(5), p. 04022052. Available at: https://doi.org/10.1061/(ASCE)ME.1943-5479.0001072.

Ng, M.S., Graser, K. and Hall, D.M. (2023) 'Digital fabrication, BIM and early contractor involvement in design in construction projects: a comparative case study', *Architectural Engineering and Design Management*, 19(1), pp. 39–55. Available at: https://doi.org/10.1080/17452007.2021.1956417.

Preston, K.E., Sujan, M.A. and Habli, I. (2025) 'Assurance of AI and autonomous technology in complex environments: Safety Critical Systems Symposium', in *Proceedings, Safety Critical Systems Symposium SSS'25*. Safety Critical Systems Club. Available at: https://scsc.uk/e1099prog (Accessed: 6 February 2026).

Ribeirinho, M.J. *et al.* (2020) *The next normal in construction: How disruption is reshaping the wrold's largest ecosystem*. McKinsey. Available at: https://www.mckinsey.com/~/media/McKinsey/Industries/Capital%20Projects%20and%20Infrastructure/Our%20Insights/The%20next%20normal%20in%20construction/The-next-normal-in-construction.pdf.

RVO (2025) *Technology Readiness Levels (TRL)*, *RVO.nl*. Available at: https://www.rvo.nl/onderwerpen/trl (Accessed: 12 August 2025).

Sjöö, K. and Frishammar, J. (2019) 'Demonstration projects in sustainable technology: The road to fulfillment of project goals', *Journal of Cleaner Production*, 228, pp. 331–340. Available at: https://doi.org/10.1016/j.jclepro.2019.04.302.

Stelvaga, A. and Fortin, C. (2022) 'Development of a Methodology for Technology Demonstration Projects Evaluation', *Proceedings of the Design Society*, 2, pp. 273–282. Available at: https://doi.org/10.1017/pds.2022.29.

Sviridova, A. and Verlinden, J.C. (2023) 'Materialization of the future: The demarcation line between prototypes and demonstrators', in *From Abstractness to Concreteness – experiential knowledge and the role of prototypes in design research. International Conference 2023 of the Design Research Society Special Interest Group on Experiential Knowledge (EKSIG)*, Milan, Italy: Politecnico di Milano , 2023, pp. 942–957.

Wang, H., Hunhevicz, J.J. and Hall, D.M. (2025) 'From automation to agency: Prototype for self-owning intelligent buildings enabled by blockchain', *Automation in Construction*, 177, p. 106309. Available at: https://doi.org/10.1016/j.autcon.2025.106309.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A TWO-STAGE MEP INSTALLATION CONSTRUCTABILITY OPTIMIZATION SYSTEM INTEGRATING DSM-BASED MODULE DIVISION AND 4D SIMULATION

Ang Li[1], Zhengyi Chen[2], and Jack C.P. Cheng[1]
[1] The Hong Kong University of Science and Technology, Hong Kong SAR, China
[2] Southeast University, Nanjing, China

Abstract

While modularization techniques can reduce installation complexity and improve the efficiency of MEP projects, existing approaches predominantly focus on static constraints and overlook the dynamic environmental interactions in assembly process. Addressing these challenges, this paper developed an MEP installation optimization framework integrating DSM-based module division and a simulation-based feedback mechanism. The framework extracts IFC data for dynamic modeling, clusters elements into modules, and employs a feedback loop to optimize installation sequences. Validated on a real-world project, the system successfully generates collision-free kinematic paths and significantly reduces installation costs, demonstrating enhanced efficiency and constructability for MEP systems.

Introduction

Mechanical, Electrical, and Plumbing (MEP) systems are essential in building systems, providing essential services to maintain a comfortable environment (Wang et al., 2025). However, the installation of these systems is challenging due to their inherent complexity, characterized by high component density and intricate spatial layouts (Lee, 2014). Consequently, the installation constructability of MEP systems, defined as the ease and efficiency of installation, has become a critical determinant of overall project constructability (Wang et al., 2016). The successful installation of MEP systems within pre-existing building structures relies heavily on a robust installation plan (Wu et al., 2022), specifically the precise sequencing and path planning of components. Suboptimal installation sequences often lead to on-site congestion and spatial conflicts (Yan et al., 2024), severely compromising construction efficiency. Therefore, developing an automated system to optimize the installation plan is imperative for enhancing MEP constructability.

To mitigate the inefficiencies and installation difficulty in traditional on-site MEP installation, off-site modular construction and multi-trade integrated MEP (MiMEP) has been widely adopted in MEP project. This approach divides the MEP system into different modules, then pre-fabricates and pre-installs the module to reduce the on-site installation (Zhao et al., 2026), thereby enhancing the installation constructability. For MiMEP, the installation plan includes determining the module division and determining the installation sequence and path for each module. However, the granularity of module division (i.e., module dimension and quantity) critically dictates the feasibility of the delivery path (Zhao et al., 2024). Determining the optimal module division remains a significant challenge, as it must balance the benefits of prefabrication against the rigorous spatiotemporal constraints imposed by the constructed environment.

Existing matrix-based methodologies have demonstrated significant efficacy in MEP modularization. For example, Samarasinghe et al. (2019) utilized the Dependency Structure Matrix (DSM) to quantify the interaction strength between MEP components, which effectively clusters MEP elements into cohesive modules. However, current module division methods predominantly prioritize static physical constraints, such as stiffness, structural integrity, and weight. Consequently, these methods often fail to adequately account for the dynamic environmental interactions inherent to the installation process, specifically overlooking critical factors like transportation trajectories and evolving spatial availability.

Furthermore, current research usually formulates the MEP installation sequences schedule as a mathematical optimization problem to determine optimal sequence. For instance, Yin et al. (2024) modeled the sequencing task using mixed integer linear programming (MILP), aiming to minimize potential spatial conflicts. However, such approaches often overlook the reciprocal influence between the installation sequence and the dynamic site environment, which fails to consider how an installed module alters the feasible path planning for subsequent components. Moreover, these methods typically lack a quantitative framework for evaluating installation costs and efficiency. Therefore, integrating 4D simulation into the optimization framework is necessary, as it provides a dynamic testbed that can rigorously capture the evolving spatiotemporal environment, ensuring that the continuous changes in site constraints are fully accounted for during sequence planning.

For the implementation of 4D simulation, the Robot Operating System (ROS) has emerged as a predominant platform. It offers a robust ecosystem featuring advanced physics engines and modular libraries that enable high-fidelity modeling of kinematic interactions and collision dynamics (Kim et al., 2021). While recent studies have

successfully demonstrated the conversion of building information modeling (BIM) data into ROS environments for general construction simulations (Liang et al., 2025), research specifically targeting the complex domain of MEP systems remains limited. Furthermore, given that MEP optimization inherently involves evolving spatiotemporal constraints, there is a distinct lack of a unified framework that effectively integrates 4D simulation with the dynamic constraint management required for the MEP installation process.

Addressing this contextual backdrop, this paper proposes a two-stage MEP installation optimization framework that integrates DSM-based module division and 4D simulation. The system aims to generate optimal installation sequences and collision-free kinematic paths, thereby maximizing constructability within the dynamic on-site environment.

Methodology

To enhance installation constructability through precise sequence planning, this study proposes a comprehensive MEP installation optimization framework driven by 4D simulation, presented in Figure 1. The methodology comprises three integral steps: 1) Dynamic environmental modeling, this step extracts constructability-related data from IFC files to instantiate a physics-based 4D simulation environment within the ROS; 2) MEP module division, a DSM-based approach is employed to decompose complex MEP systems into discrete, manageable modules, thereby facilitating off-site prefabrication and on-site assembly. 3) Installation sequence optimization, the optimal installation sequence and kinematic paths are determined via a two-stage optimization algorithm with feedback mechanisms.

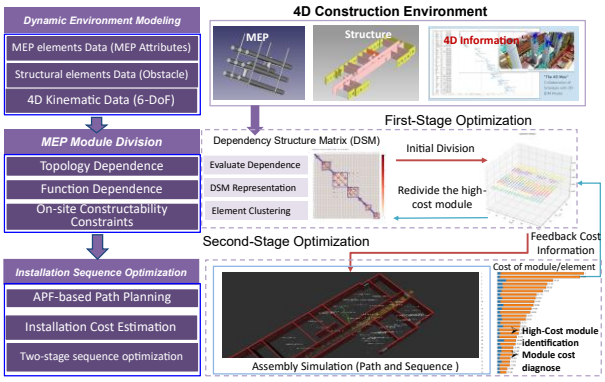


Figure 1: MEP Installation Optimization Framework

Dynamic environment modeling for MEP installation

To address the complexities of the dynamic construction environment, this study establishes a high-fidelity kinematic model (Robot Operating System, ROS) of the dynamic environment, which can precisely simulate the installation trajectories of MEP components to calculate installation cost and analyze sequence conflicts, thereby supporting subsequent optimization of installation paths and sequences. The establishment of dynamic environment involves two steps: (1) Extraction of Constructability-related Information from IFC data, and (2) Establishment of kinematic simulation environment within ROS.

Extraction of Constructability-related information from IFC data

To facilitate the precise extraction of constructability-related data for MEP installation, the IfcOpenShell library is employed to parse the BIM data in Industry Foundation Classes (IFC) files. Table 1 summarizes the specific instances of the MEP and structural elements extracted to support the optimization process. The study focuses on four fundamental categories of MEP elements: segments, fittings, controllers, and terminals. These categories encompass the primary components of MEP systems and provide a sufficient basis for analyzing and enhancing installation constructability. Additionally, structural elements, specifically entities under IfcBuildingElement, are extracted and treated as static obstacles (i.e., spatial constraints) during the installation process.

Table 1: Constructability-related IFC Information

Element	Entity	MEP Instance
MEP Element	IfcFlowSegment	IfcPipeSegment IfcDuctSegment IfcCableSegment
	IfcFlowFitting	IfcPipeFitting IfcDuctFitting IfcCableFitting
	IfcFlowController	IfcValve IfcFlowMeter IfcSwitchingDevice
	IfcFlowTerminal	IfcAirTerminal
Structural Elements	IfcBuildingElement	IfcWall IfcColum IfcBeam IfcSlab

To support dynamic modeling, the constructability-related information for each MEP element (the target of the optimization) is categorized into three distinct domains:

1) Geometric Information: The geometric configuration and spatial positioning of MEP elements are prerequisites for developing the assembly kinematics model used in installation simulations. The shape representation of each element is derived from IfcShapeRepresentation and stored as a triangulated boundary representation (B-Rep). Furthermore, the position of each element is retrieved from IfcObjectPlacement and transformed into a unified global coordinate system by traversing the relevant chain of local coordinate systems.

2) Production Information: This domain encompasses critical fabrication and installation attributes, including material type (e.g., galvanized steel, PVC), assembly method (e.g., flange, weld), and element weight. These attributes significantly influence the complexity of MEP

installation. For instance, modules requiring heterogeneous assembly methods necessitate frequent switching between different tooling or processes on-site, thereby increasing installation difficulty. Regarding data extraction, material types are queried from the IfcMaterial of the associated entities, while assembly methods are derived from specific property sets (e.g., Pset_DistributionPortTypeDuct) attached to IfcDistributionPort. Additionally, element weights are retrieved directly from IfcQuantityWeight.

3) Functional Information: This category defines the functional classification of an MEP element and its functional-related elements. The functional classification of the element identifies the specific purpose of an element (e.g., pipe system or duct system) and is used for clustering elements based on their utility. This information is extracted from the PredefinedType attribute (e.g., IfcDuctSegmentTypeEnum) of the respective IFC entities. Furthermore, the functional-related elements refer to the connected elements to ensure the systemic functionality, which can be derived from IfcRelConnectsElements or IfcRelFlowControlElements of the associated entities.

Establishment of Simulation Environment for Path and Sequence Planning

To develop the simulation test platform, the extracted IFC information is instantiated within the ROS environment (Figure 2), which defines the boundary specifications (obstacles) and environmental constraints necessary for MEP elements installation. The ROS environment provides a physics-based computational framework that enables spatiotemporal optimization of construction sequences and installation path planning under dynamic on-site environmental constraints.

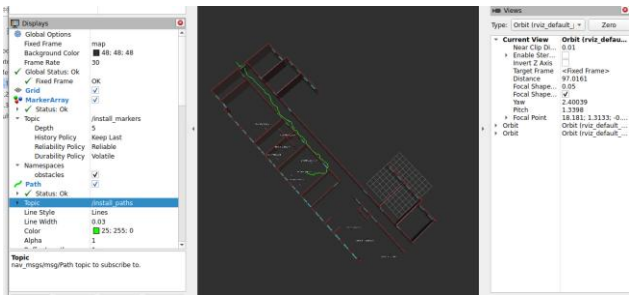


Figure 2: Interface of ROS

The configuration of ROS for 4D installation optimization includes 3 parts as follows:

1) Static Environmental Constraints: Structural elements (e.g., columns, beams, and slabs) completed prior to the MEP phase are instantiated as static rigid bodies. These elements serve as hard constraints, explicitly delineating the occupied volumetric space that must be avoided during the operation.

2) Traversable Space Definition: The system identifies architectural openings, such as corridors, doors, and windows, to define the navigable volume. This volume maps the available free space designated for the logistical transport of MEP components.

3) Configuration Space Configuration: By integrating the static structural data with the traversable volume, the environment establishes a collision-free configuration space. This representation serves as the foundational input for subsequent path planning algorithms, ensuring that any generated trajectory strictly adheres to the physical realities of the partially built structure.

MEP Model Division considering constructability constraints

To mitigate the inefficiencies inherent in traditional on-site MEP installation, off-site modular construction has been widely adopted. This approach decomposes complex, large-scale MEP networks into discrete, manageable modules. These modules are pre-assembled in a controlled off-site environment, thereby reducing on-site operations to relatively simple inter-modular connections. However, determining the optimal division (i.e., quantity and dimensions) of these modules presents a significant challenge, particularly given the spatial constraints imposed by the on-site constraints. The module division strategy critically impacts both assembly cost and installation complexity; therefore, the system must be divided to minimize complex on-site assembly operations.

To obtain the optimum module division, this study develops a dependence structure matrix (DSM)-based framework for module division and facility the module installation. The DSM-based method uses a matrix represent the dependence (i.e., interaction strength) of each MEP elements and cluster the high-dependent elements into a cohesive module. The framework comprises two primary stages: (1) Evaluation of MEP element dependence in DSM; (2) Module division considering on-site constraints.

Evaluation of MEP element dependence in DSM

The interdependence of the MEP elements is depicted in matrix form using DSM, which identifies the clusters of the elements have stronger inner interactions by mapping the interaction between elements in the system. DSM typically has a square symmetric matrix, where the diagonal axis of the matrix represents the elements and off-diagonal points represent the dependent value of one element with another.

The dependence values are estimated based on Eq. (1), which includes two components i.e., the topology dependence and function dependence. The detailed indexes for dependence estimation is shown in Table 2.

$$d = d_{Topology} + d_{Function} \quad (1)$$

Module division considering on-site constraints

The module division phase employs a Constrained Agglomerative Clustering algorithm to group MEP elements into different modules. This process is governed by a greedy optimization strategy as the following stages:

1) Prioritization via Greedy Strategy: To maximize intra-module cohesion, the algorithm prioritizes element pairs with the strongest coupling. All elements (D_{ij}) in the DSM are sorted in descending order based on their dependent value. This ensures that elements with the highest

Table 2: MEP elements dependence evaluation

Dependence type	Index	Description
Topology Dependence	Connectivity Constraint	If two elements are physically connected (i.e., sharing a common flow port or joint), the dependency $d_{Topology}$ score is incremented by 1.
	Adjacent Elements	To account for potential spatial clustering, the $d_{Topology}$ score is incremented by 1 if the Euclidean distance between two elements is less than a predefined threshold of 0.5 meters.
	Loop Integrity	If elements constitute a closed-loop sub-system or a local hydraulic cycle (e.g., bypass piping loops, valve manifolds), their $d_{Topology}$ score is incremented by 1 to preserve the integrity of these functional groups during modularization.
Function Dependence	Homogeneity	The $d_{Function}$ score is increased by 1 if the pair of elements belongs to the same IFC entity class or functional category, ensuring consistency in component handling.
	Process Compatibility	The $d_{Function}$ score is incremented by 1 if the elements share identical connection mechanisms or assembly interfaces (e.g., flange-to-flange connections)
	Operational Coupling	The $d_{Function}$ score is incremented by 1 for element pairs exhibiting operational interdependency

topological and functional dependencies are evaluated for merging with higher priority.

2) Cluster Aggregation Process: Initially, each MEP element is treated as a discrete singleton cluster. The algorithm iterates through the sorted dependence value. For any two distinct clusters with high dependence value, a tentative merge operation is initiated to form a potential candidate module.

3) Dynamic Constraint Verification: Before finalizing the union of two clusters, the candidate module must pass a rigorous feasibility check to ensure compliance with on-site limitations (e.g., logistics and installation), which is shown in Table 3. The Union operation is executed if and only if the candidate module satisfies all constraints. If a constraint violation is detected, the merge is rejected, and the clusters remain independent to prevent the formation of infeasible super-modules.

Installation Sequence Optimization with feedback mechanism

While off-site modularization significantly mitigates on-site complexity, overall construction efficiency and cost remains critically sensitive to the assembly sequence. A suboptimal sequence often leads to spatial conflicts, where previously installed components obstruct the feasible installation trajectories of subsequent modules. To address this, this study implements a two-stage optimization strategy, it utilizes the Artificial Potential Field (APF) method for path planning, integrated with a feedback mechanism that triggers module redivision when viable paths cannot be found.

Table3: The on-site constraints for module division

Constraint	Description
Dimensional Constraints	The geometric dimensions of each module are validated against the maximum transportable envelope (e.g., $3m \times 6m \times 4m$) determined by project logistics.
Payload Constraints	The cumulative mass of the module must strictly adhere to the rated lifting capacity of the on-site hoisting equipment.
Gravity Stability	The module's center of gravity is constrained to fall strictly within its support footprint to guarantee static stability during hoisting and placement.
Clustering Coefficien	A module should not contain two dense clusters connected solely by a slender or fragile element, which is extremely prone to breakage during transportation.

Formulation of installation optimization

The objective of the optimization is to obtain installation sequence $S = [m_1, m_2, \dots, m_N]$ with lowest total installation cost, where m_i is the module installed in the i-th step. The objective function can be formulated as Eq. (2). The E_{i-1} is the installation environment for i-th step which includes the static structural elements (e.g., walls, columns) and the installed MEP elements in previous i-1 steps. The installation path should satisfy the collision free constraint as Eq. (3) represented, specifically, for each module m_i, a continuous path must exist that remains strictly collision-free with respect to both static environmental obstacles and the cumulative volume of all previously installed modules $(m_1, \dots, m_{i-1})$.

$$Obj: \text{Min } Cost = \sum_{i=1}^{N} Cost(m_i|E_{i-1}) \quad (2)$$

$$Constraint: \text{Path}_{m_1} \cap E_{i-1} = \emptyset \quad (3)$$

To calculate the comprehensive installation cost and improve the installation efficiency, the installation cost for m_i consist of the weighted cost of translation and rotation manipulation ($Cost_M$), the cost penalty for operational difficulty caused by proximity ($Cost_P$), the cost for trajectory smoothness ($Cost_S$), which are shown in Eq. (4).

$$Cost_{m_i} = Cost_M + Cost_P + Cost_S \quad (4)$$

$Cost_M$ is calculated based on the length of the installation path, since rotational manipulation typically incurs significantly higher costs and control efforts compared to translational displacement, a higher weight coefficient is assigned to the rotational cost, which is shown in Eq. (5).

$$\text{Cost}_M = \text{w}_1 T + \text{w}_2 RL \quad (5)$$

Where w_1 and w_2 are the weight for translational cost and rotational cost, respectively, $\text{w}_1 < \text{w}_2$, T is the total amount of translational manipulation, R is the total amount of rotational manipulation (rad), L is the length of the module.

The $Cost_P$ quantifies the operational difficulty imposed by spatially constrained environments (e.g., narrow pipe shafts). Limited space increased the control precision of manipulation and reduced the maneuvering speeds.

$$Cost_P = \int_{p_0}^{p_1} \max\big(0, D_S - \text{D}(p,\text{O})\big)\, dp \quad (6)$$

Where p_0 and p_1 are the start point and end point of the path, respectively, D_S is the safe distance between MEP and obstacles, $\text{D}(p,\text{O})$ is the actual distance between MEP and obstacles.

The $Cost_S$ evaluates the geometric continuity of the trajectory. Even in obstacle-free zones, trajectories exhibiting high tortuosity (e.g., frequent sharp turns or erratic zigzag maneuvers) are operationally inefficient due to the requirement for frequent acceleration and deceleration.

$$Cost_S = \sum_{q=1}^{P} \left|\Delta\theta_q\right| L \quad (7)$$

where $\Delta\theta_q$ represents the turning angle (i.e., the change in heading) between adjacent path segments at waypoint q.

The simulation-based heuristic search algorithm employed to determine the optimal installation sequence in ROS platform. Specifically, the proposed approach adopts a Greedy Strategy with Look-ahead to iteratively select the next module for installation. The procedure for each installation step is detailed below:

1) Candidate Generation: To reduce the search space, a subset of potential candidates is identified from the remaining uninstalled module set, denoted as R. A heuristic filter selects the top- K modules ($CandidateSet \subset R$) based on geometric volume, prioritizing larger modules to resolve major spatial constraints early in the sequence.

2) Feasibility Simulation: For each candidate module $c \in CandidateSet$, a physics-based simulation is executed to verify installation feasibility. The system invokes the path planner (using the Artificial Potential Field method in this study) to generate a collision-free trajectory from the supply point to the installation target, considering the current environmental state (static structures plus previously installed modules). Candidates for which a valid path cannot be generated are deemed infeasible in the current iteration.

3) Cost Evaluation and Selection: The installation cost, $Cost(c)$, is computed for each feasible candidate using the multi-objective metric defined previously (incorporating manipulation effort, safety clearance, and path smoothness). The module yielding the minimum cost is selected as the optimal choicem_i for the current step.

4) Dynamic State Update: Upon selection, the geometry of module m_i is instantiated into the simulation environment and registered as a dynamic obstacle. Consequently, the path planning for all subsequent steps $(i+1)$ must strictly avoid the volume occupied by m_i, ensuring the physical validity of the evolving assembly sequence.

Artificial potential field-based module installation path planning

In this study, the APF algorithm is employed as the core trajectory generator, which provides computationally efficient, real-time motion planning with simple mathematical formulation (Liu et al., 2020). The fundamental objective is to compute a collision-free path for each module from its supply point to the installation target. The supply point refers to the storage location of prefabricated MEP components on the construction site. The installation target is the geometric location of the installing MEP element exported from design model.

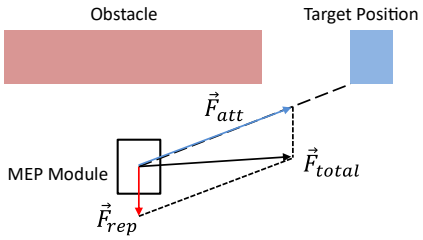


Figure 3: Artificial Potential Field Method

The algorithm conceptually abstracts the MEP module as a point mass moving within a composite virtual force field, driven by the interaction of two distinct potential functions:

1) Attractive Force ($\vec{F}_{att}$) Generated by the target installation position, this force exerts a gravitational-like pull on the module, guiding it towards the destination. The magnitude of the attractive force is typically proportional to the Euclidean distance between the module and the target, ensuring global convergence.

$$\vec{F}_{att} = k_{att} \cdot \left(\vec{P}_{goal} - \vec{P}_{current}\right) \quad (7)$$

Where k_{att} is the coefficient for attractive force.

2) Repulsive Force ($\vec{F}_{rep}$) Generated by obstacles to enforce collision avoidance. The obstacle set comprises static constraints (e.g., structural walls, columns) and dynamic constraints (i.e., the set of previously installed modules). Analogous to magnetic repulsion, these obstacles exert a pushing force that increases exponentially as the module approaches the obstacle boundary (ρ_0), thereby preventing physical intersection.

$$\vec{F}_{rep} = \eta\left(\frac{1}{\rho} - \frac{1}{\rho_0}\right)\frac{1}{\rho^2} \quad (7)$$

Where η is the repulsive gain coefficient.

The motion of the module is governed by the superposition of the attractive and repulsive forces. At each iterative step, the system calculates the resultant force vector acting on the module. The module's position is updated incrementally along the direction of this resultant force (following the negative gradient of the potential field) until it successfully converges at the target coordinates.

$$\vec{F}_{total} = \vec{F}_{att} + \vec{F}_{rep} \quad (7)$$

Feedback high-cost information for module redivision

Suboptimal module partitioning often leads to computational instability during path optimization, manifesting as deadlocks (local minima) or oscillatory trajectories. This typically necessitates excessive manual tuning of stochastic perturbation parameters to achieve convergence. To address this, a closed-loop feedback mechanism is introduced. When a module exhibits an anomalous trajectory, the aggregate navigation cost is decomposed to evaluate the contribution of individual constituent elements within the DSM. An element is identified as a high-cost contributor if it violates specific kinematic or spatial criteria:

Spatial Constriction: The element maintains persistent proximity to obstacles, indicating a clearance bottleneck.

Kinematic Amplification: The element is located at the distal end of the rotation radius, resulting in high linear velocity and an excessive swept volume.

Based on this diagnosis, these high-cost elements are imposed as separation constraints within the DSM to trigger module re-partitioning, followed by a re-optimization of the installation sequence.

Results

Case Description

To rigorously validate the feasibility and robustness of the proposed MEP installation optimization framework, a real-world construction project in Figure 4 was selected as the empirical testbed. This case study represents a large-scale, high-density engineering scenario, comprising 17283 distinct MEP elements and 762 structural entities (e.g., beams, columns, and shear walls). Characterized by its intricate spatial layout and complex topological entanglements, this project presents a significant computational challenge, making it an ideal benchmark to verify the effectiveness of the proposed method in handling high-dimensional combinatorial optimization problems, which need to consider both the module division and installation path.

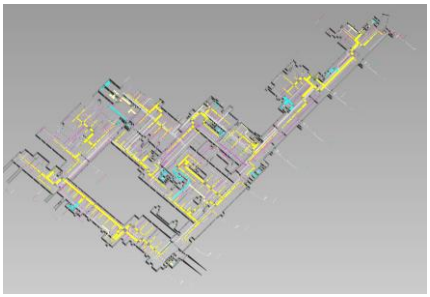


Figure 4: MGY underground parking garage MEP model

Module Division Results

By applying the proposed DSM-based framework which integrates topological dependencies with physical constraints, the 17283 MEP elements in the MEP design were successfully divided into 2564 distinct modules.

To provide a granular visualization of the partitioning efficacy, a representative sub-region comprising 100 contiguous elements was selected for detailed analysis. As illustrated in Figure 5, these elements were algorithmically grouped into 17 modules, where distinct colors denote separate modular assignments. The visualization confirms that physically connected and spatially adjacent components are coherently grouped, ensuring geometric continuity.

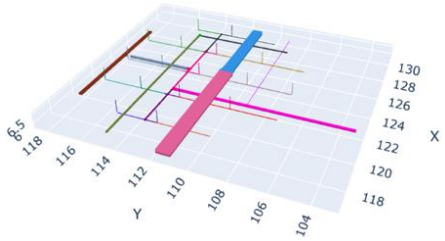


Figure 5: Module division example (100 elements)

The corresponding clustered Dependency Structure Matrix (DSM) is presented in Figure 6. The blue bounding boxes along the main diagonal delineate the boundaries of the identified modules. The heatmap reveals a distinct block-diagonal structure: high-dependency interactions (indicated by darker red hues) are concentrated within the modules, signifying strong intra-module cohesion. Conversely, the sparse distributions outside the diagonal blocks indicate minimal inter-module coupling, validating the algorithm's effectiveness in minimizing complex on-site connections.

Optimization Results

Following the initial module partitioning, the proposed two-stage optimization algorithm was employed to determine the globally optimal installation sequence. Figure 7 illustrates the generated optimal trajectory for a representative module. Crucially, the simulation environment is dynamic; modules scheduled earlier in the sequence are instantiated as static obstacles, thereby

enforcing strict spatiotemporal constraints on the current path planning task.

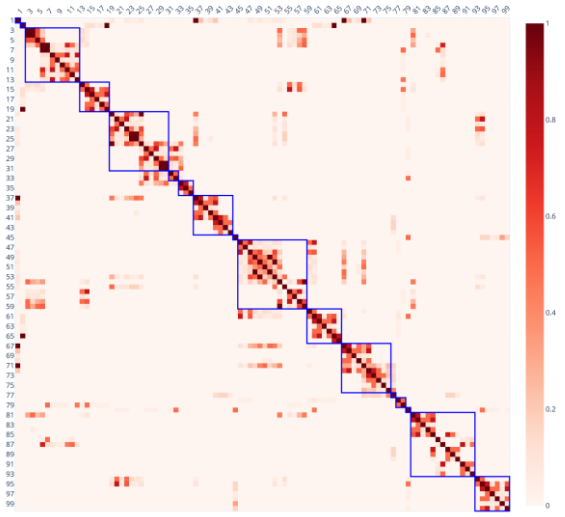


Figure 6: DSM heatmap and element dependence (100 elements)

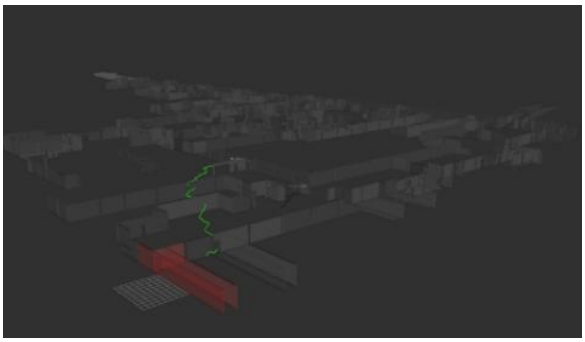


Figure 7: Module installation path diagram

Figure 8 presents the optimized installation sequence and corresponding installation costs for the initial 20 MEP modules. To ensure the installation robustness, the kinematic anomalies of the paths are identified through oscillatory motions detection (i.e., repetitive back-and-forth movements), which is highlighted in red in Figure 8.

These identified anomalies necessitate further optimization via the feedback loop. As illustrated in Figure 9, such kinematic instabilities (e.g., zig-zag trajectories) are often attributed to suboptimal module division, specifically inappropriate module dimensions. The feedback mechanism is utilized to identify high-risk elements that contribute disproportionately to the installation cost.

As detailed in Table 5, these high-risk elements are characterized by high-risk scores (e.g., 0.97), resulting from either persistent proximity to obstacles (high Prox value) or excessive swept volumes due to large rotation radius (Radius close to 1.0). These metrics serve as the basis for imposing separation constraints in the DSM for re-partitioning.

Install Order	Module ID	Module Name	Total Cost
1	MOD_2966	Module MOD_2966	95.99
2	MOD_1283	Module MOD_1283	65.40
3	MOD_2637	Module MOD_2637	156.40
4	MOD_3460	Module MOD_3460	167.79
5	MOD_2649	Module MOD_2649	142.23
6	MOD_527	Module MOD_527	204.97
7	MOD_2627	Module MOD_2627	151.92
8	MOD_2643	Module MOD_2643	219.23
9	MOD_5526	Module MOD_5526	197.75
10	MOD_642	Module MOD_642	243.78
11	MOD_10827	Module MOD_10827	226.76
12	MOD_5830	Module MOD_5830	242.43
13	MOD_1458	Module MOD_1458	274.44
14	MOD_4354	Module MOD_4354	258.99
15	MOD_1276	Module MOD_1276	428.86
16	MOD_4659	Module MOD_4659	247.51
17	MOD_9048	Module MOD_9048	329.98
18	MOD_10447	Module MOD_10447	664.98
19	MOD_10437	Module MOD_10437	726.70
20	MOD_10355	Module MOD_10355	973.94

Figure 8: Module installation sequence

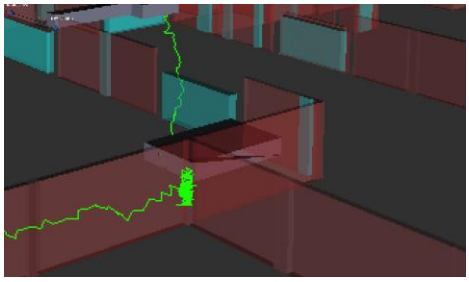


Figure 9: Kinematic anomalies of installation path (Oscillation example)

Table 5: High-Risk Elements (Top 5)

Risk	Swept	Prox	Type	ID
0.97	0.94	1	IfcDuctFitting	8254523
0.74	0.96	0.53	IfcDuctSegment	8254517
0.71	1.00	0.42	IfcDuctFitting	8254538
0.69	0.96	0.42	IfcDuctSegment	8254535
0.66	0.52	0.80	IfcDuctSegment	8254506

Finally, the overall efficacy of the framework is validated through comparative cost analysis.

Table 6: Total installation cost of different method

Method	Optimal Installation cost
Two-stage Optimize	602006
No redivision feedback	1014465
No module division	2752630

As summarized in the comparison Table 6, the proposed two-stage optimization achieves a significant installation cost reduction compared with the baseline methods, which delivers a 78 % lower installation cost than the traditional direct component installation strategy (i.e., without module division) indicating the advantage of modular installation. The two-stage method also shows a 41% cost reduce over no redivision feedback method,

confirming that the integration of modularization with feedback-driven optimization yields the most efficient construction process.

Conclusions

To address the critical challenge of modeling dynamic on-site environments and optimizing installation sequences under complex on-site installation constraints, this paper proposes a two-stage MEP installation optimization framework that integrates DSM-based module division and 4D simulation. The system aims to generate optimal installation sequences and collision-free kinematic paths, thereby maximizing constructability within the dynamic on-site environment. Our proposed MEP installation optimization system has three major contributions: 1) IFC-based 4D Simulation Environment, this study establishes 4D simulation environment within ROS, rigorously transforming IFC building information into dynamic spatiotemporal constraints to support virtual construction planning. 2) DSM-based MEP Module Division, the DSM-based framework is developed to streamline off-site fabrication, which employs constrained agglomerative clustering to decompose complex MEP systems into transportable modules based on functional and topological dependence considering on-site constraints. 3) MEP Installation Sequence Optimization with Feedback Mechanism, installation efficiency is maximized via a two-stage sequence optimization algorithm incorporating a closed-loop feedback mechanism, which refines inappropriate module division to resolve kinematic conflicts and path planning deadlocks.

The proposed methodology was validated on a real-world MEP project characterized by high component density and complex layouts. The results indicate that our proposed framework can successfully divide the MEP elements in different modules based on DSM and optimize the installation path and sequence of the MEP module based on 4D simulation. As Table 6 shows, this two-stage MEP installation optimization can significantly reduce the total installation cost, thereby enhancing the installation constructability and efficiency.

With our proposed MEP installation optimization system, users can obtain the robust and collision-free installation plan for MEP project in an efficient and automated way. Additionally, our system leverages the 4D simulation to simulate the changing on-site environment, which can integrate the dynamic installation constraints into the optimization framework. Future research will focus on extending the framework's applicability to more domain-specific and highly complex scenarios, such as intricate HVAC plant room assemblies.

Acknowledgments

The authors would like to thank the Innovation and Technology Fund, HKSAR (No. ITS/394/23FP) for partially supporting this research. The authors declare no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Kim, S., Peavy, M., Huang, P. C., & Kim, K. (2021). Development of BIM-integrated construction robot task planning and simulation system. Automation in Construction, 127, 103720.

Lee, G. (2014). Parallel vs. sequential cascading MEP coordination strategies: A pharmaceutical building case study. Automation in Construction, 43, 170-179

Liang, Z., Yang, L., Liang, Z., Chan, J. C. F., Zhang, Z., Wang, M., & Cheng, J. C. (2025). Optimized UAV view planning for high-quality 3D reconstruction of buildings using a modified sparrow search algorithm. Advanced Engineering Informatics, 65, 103344.

Liu, J., Xu, C., Wu, Z., & Chen, Y. F. (2020). Intelligent rebar layout in RC building frames using artificial potential field. Automation in Construction, 114, 103172.

Samarasinghe, T., Gunawardena, T., Mendis, P., Sofi, M., & Aye, L. (2019). Dependency Structure Matrix and Hierarchical Clustering based algorithm for optimum module identification in MEP systems. Automation in Construction, 104, 153-178.

Wang, B., Lin, F., Li, M., Liang, Z., Chen, Z., Wang, M., & Cheng, J. C. (2025). Informative As-Built Modeling as a Foundation for Digital Twins Based on Fine-Grained Object Recognition and Object-Aware Scan-vs-BIM for MEP Scenes. Advanced Engineering Informatics, 65, 103382.

Wang, J., Wang, X., Shou, W., Chong, H. Y., & Guo, J. (2016). Building information modeling-based integration of MEP layout designs and constructability. Automation in construction, 61, 134-146.

Wu, Q., Chen, L., Shi, P., Wang, W., & Xu, S. (2022). Identifying impact factors of MEP installation productivity: An empirical study. Buildings, 12(5), 565.

Yan, G., Yang, Y., Zhang, H., Li, Z., Chen, S., Zhao, X., ... & Liu, L. (2024). A review of life cycle construction process and cutting-edge technology in prefabricated MEP installation engineering. Buildings, 14(3), 630.

Yin, J., Luo, S., Rui, J., Yao, J., Zhang, F., Fu, B., & Kassem, M. (2024). BIM-based detection and optimization of spatial-temporal clashes in underground pipeline construction. Automation in Construction, 166, 105616.

Zhao, X., Li, J., Yang, Q., Fan, X., Liu, S., Cao, L., ... & Zhang, Z. (2026). Corridor MEP pipeline integration and hanger-layout optimization using BIM-based genetic partitioning and performance analysis. Journal of Intelligent Construction, 9180114.

Zhao, X., Yang, Q., Yan, G., Fan, X., Yang, Y., Zhang, H., & Chen, S. (2024). Length optimization of MEP pipeline integrated modular based on genetic algorithm. Buildings, 14(12), 3826.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

EMERGING OPPORTUNITIES FOR INTEGRATING DIGITAL PRODUCT PASSPORT INTO LEAN CONSTRUCTION WORKFLOWS: A CONCEPTUAL PERSPECTIVE

Yanh Ribeiro[1], Pedro Mêda1[1,2], and Jorge Teixeira[1]
[1]ICS, CONSTRUCT, Faculty of Engineering, University of Porto, Portugal
[2] CONSTRUCT, Faculty of Engineering, University of Porto, Portugal

Abstract

Digital Product Passports (DPPs) are emerging as mandatory, standardised sources of product information under EU policy. Yet their implications for construction-phase Lean workflows remain underexplored. Current practices still suffer from late discovery of product-related constraints, missing documentation, and weak traceability, undermining task readiness and plan reliability. This study combines a scoping search with conceptual mapping of DPP data categories to digital Last Planner System (LPS) decision points. Results include a DPP-to-LPS integration framework and construction-oriented guidelines for make-ready, commitments, and receiving/QA. The contribution is a practical research agenda for operationalising DPPs in Lean execution and enabling data continuity (e.g., DBLs-related).

Introduction

Lean Construction aims to improve production reliability and workflow stability on construction sites by reducing variability and preventing tasks from starting without the necessary prerequisites. In practice, however, construction-phase planning and control often experience challenges related to product information availability, such as missing or outdated specifications, unclear compliance evidence, limited traceability across changes, and inconsistent access to installation and handling instructions. These shortcomings often surface late, during make-ready or even during execution, contributing to disruptions, rework, and "making-do" situations where crews proceed despite incomplete inputs (Amaral et al., 2023).

In parallel, European policy has moved towards mandating Digital Product Passports (DPPs) as structured, product-level information containers intended to support transparency, traceability, and circularity across value chains. Within the construction domain, recent work has discussed DPP information needs and adoption barriers, emphasising issues of stakeholder roles, data governance, and interoperability with existing information practices (Mêda et al., 2024; Mêda et al., 2023). Research on material and product passports has also highlighted that data availability is uneven and that explicit data templates are needed to operationalise passport concepts in real settings (Çetin et al., 2023). This aligns with findings from construction information systems, where the classification and standardization of item parameters within work breakdown structures are fundamental to enable integration and interoperability (Ribeiro et al., 2023b). Yet, despite the growing DPP discourse, relatively little has been published on how DPP data could be operationalised to support construction-phase production planning and control mechanisms, particularly Lean workflows on site.

While product-related information is typically generated during design and planning phases, it may become fragmented, outdated, or misaligned with site and product reality during construction execution. This disconnect undermines the effective use of information at the moment when it is most needed for decision-making. DPPs are an instrument to bridge this gap by making product information more available and accessible, while maintaining it structured, standardised, and traceable across project phases and stakeholders (Preamble 32, European Parliament, 2024).

In this context, the relevance of DPPs lies in their potential to improve process reliability and stability in construction workflows by enhancing the consistency and traceability of product information during execution. By supporting more reliable and usable information flows, DPPs can help reduce errors associated with information gaps and strengthen feedback loops within production control.

From this perspective, the DPP can be positioned as a practical enabler of Lean execution, not by "making construction lean" on its own, but by improving the timeliness, structure, and verifiability of product information at decision points where Lean depends on reliable inputs. Key DPP attributes, such as unique identifiers, technical specifications, and traceability events, can support earlier constraint detection, more robust make-ready screening, and more reliable decisions. These capabilities align with the requirements of Last Planner System (LPS) implementations, which increasingly act as centralised production control environments for coordinating commitments, constraints, and feedback loops (McHugh et al., 2025), and with emerging evidence that structured product data integration can improve site logistics and flow reliability (Berroir et al., 2023).

This paper makes three main contributions to the body of knowledge of computing in construction. First, it

proposes a structured mapping between Digital Product Passport (DPP) data categories, construction-phase decision points in planning and control, and expected Lean mechanisms such as improved task readiness and reduced making-do, grounded in digital LPS workflows. Second, it introduces a conceptual integration framework that positions DPPs as an operational information layer supporting make-ready, commitment, and receiving decisions during execution. Third, it derives a set of actionable, system-oriented guidelines that translate DPP requirements into traceable and usable data needs for site-level decision-making. To support these contributions, the study adopts a targeted scoping search to confirm the scarcity and fragmentation of DPP–Lean Construction research and concludes by discussing implementation considerations (e.g., identifiers, access control, interoperability) and outlining a research agenda for empirical validation in project settings.

Method

Research design

This study adopts a conceptual and scoping-based research design (Arksey and O'Malley, 2005; Jaakkola, 2020). The research combines a targeted scoping search to assess the maturity and thematic focus of literature connecting Digital Product Passports (DPPs) with construction workflows and Lean production control, with a conceptual synthesis that develops an integration framework and operational guidelines for DPP-enabled digital Last Planner System (LPS) workflows.

The scoping component aims to identify the extent to which existing research addresses the operational use of product-level digital information within construction-phase planning and control. The conceptual component builds on this evidence to propose a structured mapping between DPP data categories, construction-phase decision points, and Lean execution mechanisms.

Targeted scoping search

A targeted and iterative scoping search was conducted to explore research at the intersection of Digital Product Passports (DPPs), construction workflows, and Lean production control. The purpose of this search was not to provide an exhaustive systematic review, but to identify the existence, maturity, and thematic characteristics of research connecting these domains.

Searches were conducted in the Scopus and Web of Science databases due to their extensive coverage of engineering, construction management, and digitalisation research. The search strategy followed an iterative logic, beginning with the focal research intersection and progressively expanding the terminology when limited or no results were retrieved.

The initial search string combined the central concepts of this study: "digital product passport" AND "lean construction". This search returned no results in either database, suggesting a potential research gap at the explicit intersection between DPPs and Lean Construction.

To verify whether conceptually related research existed under broader terminology, the search scope was expanded by relaxing the Lean Construction concept. The following query was applied: "digital product passport" AND "lean". This search returned two records in each database. Screening revealed that these studies addressed lean manufacturing or used the term "lean" in alternative contexts rather than Lean Construction workflows and were therefore treated as domain-drift results.

A third search aimed to map DPP research within the construction domain more broadly: "digital product passport" AND "construction". This search produced a broader set of records, enabling thematic screening and categorisation of existing research related to DPP implementation in construction.

Considering that Digital Product Passports are frequently operationalised through related constructs such as material passports and structured product data templates, additional searches were conducted using these proxy concepts. This decision was supported by prior research indicating that material passports and data template structures represent foundational building blocks for DPP implementation (Honic et al., 2024). Therefore, the following search strings were applied: "material passport" AND "lean"; "data template" AND "lean".

Across all searches, records were screened to assess their relevance to: (a) construction-phase planning and control workflows; and (b) Lean Construction mechanisms such as constraint management, make-ready processes, and Last Planner System routines. Studies addressing Lean concepts outside the construction domain were considered false positives and excluded from thematic synthesis.

Scoping results and evidence of scarcity

The targeted search "digital product passport" AND "lean construction" returned no results in either database. The expanded search "digital product passport" AND "lean" returned two results in Scopus and two in Web of Science; however, screening confirmed that these studies addressed lean manufacturing or used the term "lean" outside the Lean Construction domain.

The search "digital product passport" AND "construction" produced 32 results in Scopus and nine in Web of Science. After duplicate removal across databases, 32 unique publications were retained for thematic screening.

The retrieved publications were categorised inductively based on their dominant research contribution. Categories were defined to reflect the functional role of the research in enabling, structuring, governing, or operationalising DPP implementation within the construction sector, and are presented in Table 1.

The screening indicates that current research on DPP implementation in construction is predominantly focused on circularity, lifecycle management, interoperability, and enabling digital infrastructures. No studies were identified that explicitly address the operational use of DPP data within construction-phase Lean production

control workflows. This observed gap motivates the conceptual framework and operational guidelines proposed in this study.

Table 1: Thematic distribution of DPP+Construction publications

Category	Count	Example references
Enabling digital and data infrastructures	6	Adu-Duodu et al., 2025; Byers et al., 2025
Data structuring, templates and interoperability	8	Aragón et al., 2025; Vangelova et al., 2025
Stakeholders, governance and adoption conditions	4	Mêda et al., 2024; Mêda et al., 2023; Ruismäki et al., 2025
Circularity, lifecycle and end-of-life applications	9	Çetin et al., 2023; Pedroso et al., 2025
Construction-phase and production control	0	-
Peripheral or indirect mentions of DPP	5	Cieślewicz et al., 2025; Mengyao Li, 2025

Complementary searches

To ensure that relevant contributions were not overlooked, broader sensitivity searches were conducted using combinations of synonyms and adjacent concepts related to product passports, data templates, circular economy, production planning, and construction workflows.

The searches combined DPP-related concepts, Lean production synonyms, product data approaches, circular economy–related concepts, and construction-related terms (e.g., digital product passport, material passport, product passport, data template, product data, circular economy, circularity, EPD, lean construction, last planner, production control, workflow, construction, site, building). Results retrieved from Scopus and Web of Science consistently showed that the identified studies either focused on adjacent domains or reflected domain drift, such as workflow digitalisation, lifecycle sustainability, data modelling, or Lean Manufacturing. Most records either overlapped with previously identified studies or addressed related research areas without explicitly linking Digital Product Passports to construction-phase Lean production control workflows.

Overall, the complementary searches did not reveal additional directly relevant studies beyond those identified in the primary search. Consequently, these results did not alter the thematic synthesis and were not included in further analysis. This approach is consistent with scoping-based research aimed at assessing the extent and boundaries of emerging research domains rather than exhaustively synthesising all available literature.

Conceptual synthesis and outputs

Given the limited availability of empirical evidence directly linking Digital Product Passports (DPPs) to Lean Construction execution, the study proceeds with a structured conceptual synthesis to operationalise the potential interaction between product-level digital information and construction-phase production control. The synthesis followed an iterative, multi-step process designed to ensure traceability between theoretical constructs and operational outcomes.

First, DPP data categories relevant to construction activities were derived through documentary analysis of emerging standardisation efforts and construction-focused DPP literature. Key sources included ISO 22057 data template frameworks (ISO, 2022) and research addressing data availability and structuring requirements for material and product passports in construction (Çetin et al., 2023; Vangelova et al., 2025). This step aimed to consolidate recurring product information attributes that are technically feasible and relevant to construction execution.

Second, operational decision points within Lean Construction production control were identified through review of canonical Last Planner System literature and recent developments in digital LPS implementations (Ballard and Howell, 2003; McHugh et al., 2025). This step focused on identifying planning and control moments where reliable product information is required to support workflow stability, including lookahead/make-ready planning, weekly work planning commitments, receiving and quality verification, and exception management.

Third, DPP data categories and LPS decision points were systematically mapped to identify potential interaction mechanisms between product information and Lean execution processes. The mapping aimed to establish how specific DPP attributes could support Lean objectives such as improved constraint removal, increased task readiness, and reduction of making-do and rework.

Fourth, operational system guidelines were derived from the conceptual mappings with emphasis on practical applicability within construction workflows and digital LPS environments. The guidelines were formulated to translate conceptual relationships into actionable information management practices for site-level planning and control.

The synthesis process resulted in two main research outputs: (1) a conceptual DPP-to-LPS integration framework linking product information categories to production control decision points, and (2) a set of construction-oriented system guidelines to support operational implementation.

DPP-to-LPS integration framework

This section presents the conceptual DPP-to-LPS integration framework, linking structured product-level information to construction-phase decision points within digital Last Planner System workflows. It focuses on how DPP data can support key Lean execution processes, including lookahead planning, weekly commitments,

receiving/QA, and exception management, by improving the reliability and availability of product information at decision-making moments.

Lean Construction conceptualises production not as a sequence of isolated tasks, but as an integrated process oriented towards flow reliability and waste reduction, enabling more predictable planning and control (Garcés et al., 2025). Lean theoretical foundations further emphasise the coordination of value flows over mere task conversion, reflecting the conceptual arguments originally introduced by Koskela and consolidated in subsequent literature (Porras Díaz et al., 2014). These foundations provide the theoretical grounding for the proposed framework and guidelines.

Although Lean Construction may not necessarily suffer from a lack of information, its effectiveness is closely linked to the availability of reliable and timely information at key decision points. The Last Planner System, as a production control mechanism, is designed to improve planning reliability and manage workflow variability through collaborative commitment-making and continuous learning, relying on the availability of sound prerequisites for task execution (Ballard and Howell, 2003). In this context, research on lookahead planning shows that unresolved constraints, often related to missing or unclear product information, can affect the ability to anticipate and remove obstacles before execution (Hamzeh et al., 2015). At the operational level, such information gaps may result in making-do situations, contributing to variability and rework (Amaral et al., 2023).

Framework overview

The framework is organised around three elements: (i) DPP data categories relevant to construction execution, (ii) LPS decision points during construction-phase planning and control, and (iii) Lean execution mechanisms expected to benefit from improved product information reliability.

DPP data categories for construction execution

Based on the construction-focused DPP literature, product/passport data templates, and interoperability discussions, we define the following DPP data categories as most relevant to the construction phase:

- Identification and versioning
- Technical specifications and performance
- Compliance and documentation
- Traceability and chain-of-custody events
- Installation, handling, and safety information

These categories reflect recurring needs reported for construction products, including harmonisation challenges across passport terminologies and the requirement for machine-interpretable and interoperable product data (Aragón et al., 2025; Çetin et al., 2023; Vangelova et al., 2025). They are also consistent with stakeholder expectations and barriers such as balancing transparency with intellectual property concerns (Ruismäki et al., 2025).

Construction-phase decision points in digital LPS workflows

Digital LPS systems provide a natural integration point for DPP data because they coordinate commitments, constraint management, and feedback loops across planning horizons. In particular, DPP information can support decision-making at:

- **Lookahead planning and make-ready screening:** verifying prerequisite availability and identifying product-related constraints early
- **Weekly Work Planning commitments:** supporting reliable "can" decisions by providing verified product requirements and documentation
- **Receiving and quality checks:** confirming that delivered products match specified versions/batches and that required documentation is available
- **Exception management:** handling substitutions, non-conformities, and late changes with controlled traceability and version awareness.

Digital LPS implementations emphasise centralised information management and situational awareness for coordination, making them suitable entry points for structured product data (McHugh et al., 2025).

Mapping DPP data to Lean execution mechanisms

Table 2 summarises the mapping between DPP data categories, LPS decision points, and the primary operational purpose supporting Lean execution. For example, reliable identifiers and versioning can reduce planning failures caused by incorrect substitutions; traceability events can strengthen logistics reliability and receiving controls; and compliance documentation can prevent work starting without required approvals, thereby reducing making-do situations. Prior Lean research has linked missing prerequisites and product-related issues to waste and rework, reinforcing the importance of timely and reliable product information in production control (Amaral et al., 2023). Related work on product data integration for lean site logistics further supports the relevance of structured product information to improve flow reliability (Berroir et al., 2023).

Table 2: Mapping DPP data categories to LPS decision points and expected Lean effects

DPP category	Lookahead	WWP	Receiving/QA	Exceptions
ID & versioning	Right item/version	Reliable "can"	Delivery match	Substitution control
Technical specs	Prerequisites/compatibility	Task readiness	Properties confirmed	Equivalence check
Compliance & docs	Approvals available	Commit with evidence	Acceptance gate	Ncr re-approval
Traceability	Logistics status	Availability certainty	Batch/serial verify	Rca & targeted fix
Install/handling/safety	Handling constraints	Safe method prereqs	Prevent damage/ncr	Corrective guidance

System guidelines for DPP-enabled digital LPS workflows

This section translates the proposed framework into practical guidelines for construction teams and for digital LPS implementations that aim to incorporate DPP information during the construction phase. The guidelines are written from a production planning and control perspective, focusing on how DPP information should be used to support make-ready, commitments, logistics reliability, and quality assurance.

These guidelines describe how project teams can use DPP information within LPS routines (lookahead/make-ready, weekly commitments, receiving/QA, and learning) to increase task readiness and reduce avoidable variability. When a digital LPS platform is used, these guidelines translate into practical requirements for what the platform should make visible, record, and enforce during planning and control.

Guideline 1 — Define a "minimum DPP dataset" per product group for construction execution

For each relevant product group (e.g., façade elements, MEP equipment, firestopping products), define a minimum set of DPP data categories that must be available before work can be made ready. This minimum should prioritise identification/version, technical requirements affecting installation, and mandatory compliance documentation, avoiding information overload while ensuring decision usability (Cetin et al., 2023; Ruismäki et al., 2025).

Guideline 2 — Treat DPP-based checks as explicit make-ready constraints

In lookahead planning, create a dedicated constraint type for "product information readiness" and link it to DPP evidence (e.g., verified documentation available, correct version identified). A task should only move to "can" when these DPP-related constraints are removed.

Guideline 3 — Use DPP identifiers to control substitutions and protect plan reliability

When changes happen (due to availability, lead times, or design late decisions), record the product replacement using its DPP identifier/version and require an explicit check of: (i) equivalence/compatibility with the intended design intent, and (ii) compliance implications. This makes change a managed, informed and compliance-checked decision rather than an informal field workaround.

Guideline 4 — Link receiving to plan: confirm "right product, right version, right evidence"

At receiving, confirm that delivered items match the planned product identity (and, when relevant, batch/serial) and that required evidence is available (e.g., declarations/certificates, handling instructions). If not, open a constraint immediately and block release to installation until resolved. This creates a direct bridge between logistics and LPS make-ready.

Guideline 5 — Make installation prerequisites explicit at the moment of commitment

When a task is proposed for the Weekly Work Plan, make the key product-specific prerequisites visible and confirm them explicitly (e.g., substrate readiness, environmental limits, curing times, required tools, safety prerequisites). The aim is not to add bureaucracy, but to reduce tacit assumptions that frequently trigger making-do.

Guideline 6 — Classify failures so learning becomes actionable

When tasks are not completed as planned, classify reasons for variance using categories that isolate "product information" problems (e.g., missing documentation, unclear requirements, wrong version/substitution, traceability uncertainty). Over time, this helps quantify how much unreliability is driven by product information and where DPP use would prevent recurrence.

Guideline 7 — Establish responsibility for DPP information completeness in the supply chain

Assign clear responsibilities for ensuring that required DPP fields are available and validated before delivery and installation (e.g., supplier/manufacturer provides documentation; contractor verifies readiness; trade confirms prerequisites). This is essential to avoid shifting the burden to site crews at the last minute.

Guideline 8 — Use DPP to support work packaging and logistics planning

For work packages and kitting/logistics plans, use DPP identifiers to define package contents (correct items, versions, and quantities) and to confirm completeness at handoff points. This guideline supports more reliable material flow to the workface and reduces making-do triggered by missing components (Amaral et al., 2023; Berroir et al., 2023).

Guideline 9 — Roll out incrementally: start where constraints are most frequent and costly

Do not attempt full DPP integration across all products at once. Start with product groups that are (i) high-cost, (ii) high-risk for rework or non-compliance, or (iii) recurrent sources of constraints in lookahead planning. An incremental strategy is consistent with observed implementation priorities in construction contexts (Mêda et al., 2024).

Guideline 10 — Maintain an auditable decision trail for DPP-based readiness

At key points (task made ready, weekly commitment, receiving acceptance, installation completion), record which product identity/version and which evidence supported the decision. This should be lightweight (not a full dossier), but sufficient to answer later: what was installed, where, when, and under which requirements/evidence? This guideline supports accountability, reduces dispute ambiguity, and enables structured downstream information continuity.

Link to Digital Building Logbooks (DBL)

The evidence trail in Guideline 10 also creates a straightforward pathway to future Digital Building Logbook (DBL) use. A DBL can only function reliably if product-level information is connected to where products were installed and which version/batch and documentation applied. If LPS routines already require teams to confirm readiness and record product identities at natural control points, extending this practice to populate DBL-relevant records becomes a small, non-disruptive additional step rather than a separate documentation effort. In this sense, making DPP use operational in construction-phase control is a prerequisite for credible DBL data continuity across the asset lifecycle.

Discussion and research agenda

The proposed framework and guidelines position the DPP as an operational information layer that can reduce construction-phase variability driven by product uncertainty. The scoping results suggest that most DPP-in-construction research has prioritised circularity and end-of-life use cases, interoperability with BIM/EPDs, and enabling infrastructures, while giving limited attention to how DPP data can support production planning and control on site. By focusing on digital LPS routines (lookahead/make-ready, weekly commitments, receiving/QA, and learning cycles), this paper highlights a practical pathway to translate product information into earlier constraint removal and more reliable commitments.

A key implication is that the value of DPPs in construction depends less on the mere existence of product data and more on whether such data is effectively used at critical decision points within planning and control workflows. Existing approaches to managing product information, including conventional documentation and BIM-based systems, are often limited to the design phase and not consistently maintained during execution, which limits their effectiveness in supporting real-time decision-making (Çetin et al., 2023; Ribeiro et al., 2023b).

If DPP information remains disconnected from planning and control processes, it risks becoming a compliance-oriented repository that is consulted too late to influence workflow reliability. In contrast, integrating minimum product information checks into make-ready processes and linking receiving/QA activities to planning commitments can shift the identification of issues upstream. This contributes to reducing making-do, rework, and informal substitutions, reinforcing flow reliability.

These observations are consistent with prior findings that structured product data integration can support lean site logistics and improve production stability (Berroir et al., 2023), as well as with research highlighting the importance of actionable data templates and stakeholder-aligned information requirements for operationalising passport-based approaches (Çetin et al., 2023; Ruismäki et al., 2025).

The guidelines also clarify governance and adoption challenges. Construction stakeholders highlight intellectual property concerns, unclear ownership of data, and the need to tailor DPP content by product group (Ruismäki et al., 2025). From a site perspective, overly complex or excessive product information can create friction and be ignored under time pressure, particularly when it exceeds what is operationally necessary for task readiness and decision-making. Insights from Lean Information Management (LIM) suggest that the success of such systems hinges on ensuring information flows are reliable, timely, and waste-free. Case studies of construction information systems designed as Lean enablers show that benefits are realized when the system reduces the eight wastes of information (e.g., waiting, overprocessing, defects), thereby increasing the reliability and usability of data for decision-makers (Ribeiro et al., 2023a). The “minimum dataset” approach proposed here is intended to balance usability with compliance and reliability by prioritising a core set of essential attributes that support effective and traceable decisions without overburdening site actors (Elkhidir et al., 2025).

Finally, the evidence trail created at LPS decision points can support lifecycle continuity beyond construction. If product identities and versions are recorded where they are installed and under which evidence they were accepted, the transition to building-level information structures such as Digital Building Logbooks becomes substantially less disruptive. In this view, operational DPP use during construction is not only a Lean enabler but also a prerequisite for credible downstream information management.

Limitations and practical barriers to implementation

It is important to acknowledge the limitations of this conceptual study, and the practical barriers to implementing the proposed framework. First, the framework assumes a level of digital maturity (e.g., use of digital LPS, connectivity on site) and supply chain readiness (e.g., DPP data provision by manufacturers) that is not yet universal. Second, the guidelines introduce new responsibilities and checks; their adoption may face cultural resistance or be perceived as adding bureaucratic overhead, especially if not seamlessly integrated into existing workflows. Third, issues of data ownership, intellectual property, and the cost of implementing the required technological infrastructure are significant real-world hurdles highlighted in stakeholder studies (Ruismäki et al., 2025; Mêda et al., 2024). These barriers underscore the need for the incremental, pilot-based rollout suggested in Guideline 9, and highlight that technological and procedural solutions must be accompanied by aligned contractual and incentive structures.

Research agenda

Future work should focus on empirical validation of the proposed framework through case-based and longitudinal studies in live construction projects. Priority directions include:

- **Pilot implementations in digital LPS environments:** implement a limited set of DPP-based readiness checks for selected product groups and assess effects on planning reliability and constraint removal practices.
- **Measurement using Lean-compatible indicators:** track changes in task readiness, Percent Plan Complete (PPC), and "reasons for variance" categories related to materials/product information, as well as non-conformance and rework events attributable to product documentation or substitutions.
- **Product group tailoring and minimum dataset definition:** test how the "minimum dataset" varies across trades and products, and identify thresholds where information becomes either insufficient for risk control or excessive for site usability.
- **Data governance and responsibility allocation:** study contractual and organisational arrangements that ensure product information is available when needed without shifting the burden to site crews.
- **Continuity to DBL and asset information:** investigate how construction-phase evidence trails can be structured to support handover and operations, including links from installed locations to DPP references.

Conclusions

This paper explored how Digital Product Passports (DPPs) can support Lean Construction execution by strengthening the availability, structure, and verifiability of product information at construction-phase decision points. A targeted scoping search indicates that DPP-in-construction research is currently concentrated on circularity, interoperability, and enabling infrastructures, with limited explicit attention to on-site production planning and control. In response, we proposed a DPP-to-LPS integration framework that maps DPP data categories to digital Last Planner System workflows and expected Lean mechanisms such as improved constraint removal and task readiness.

The paper also provided construction-oriented guidelines for operationalising DPP information in lookahead/make-ready, weekly commitments, receiving/QA, and learning cycles. Beyond near-term Lean benefits, the proposed "evidence trail" approach offers a practical pathway to support future lifecycle continuity, including the population of building-level records such as Digital Building Logbooks. Future research should validate the framework through project pilots and assess impacts using planning reliability, variance reasons, and rework/non-conformance indicators.

As DPP become the norm, a new paradigm emerges in which the challenge shifts from data availability to its effective use. As a philosophy grounded in continuous improvement, Lean Construction cannot disregard this valuable source of structured information, which can support increasingly reliable decision-making in construction execution.

Acknowledgments

This work was financially supported by: UID/04708 of the CONSTRUCT - Instituto de I&D em Estruturas e Construções - funded by Fundação para a Ciência e a Tecnologia, I.P./ MCTES through the national funds and DPP4SBE project: Digital Product Passports for Sustainable Building Environment (COMPETE2030-FEDER-01456700, n. 17784) funded by PORTUGAL 2030 program, cofinanced by European funds.

References

Adu-Duodu, K., Wilson, S., Li, Y., Oladimeji, A., Huraysi, T., Barati, M., Perera, C., Solaiman, E., Rana, O., Ranjan, R. and Shah, T. (2025) A Circular Construction Product Ontology for End-of-Life Decision-Making [Online]. In: Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing, March 31, 2025. Catania International Airport Catania Italy: ACM, pp. 1943–1952. Available from: <https://doi.org/10.1145/3672608.3707870>.

Amaral, T. G. D., Maciel, C. C. M., Filho, R. R. D. G., Pessoni, R. C. D. S., Paula, G. S. D. and Silva, S. V. D. (2023) Results of the Causes and Impacts of Making-Do Wastes in Production in Fortaleza, Ceará, Brazil [Online]. Lille, France, pp. 1395–1406. Available from: <https://doi.org/10.24928/2023/0250>.

Aragón, A., Nieto, O., Rønning, A., Schulze, E., Alberti, M. G. and Pavón, R. M. (2025) Gaps in the Machine-Interpretability of ISO 22057 EPDs: Identification and Proposals for a Revised International Standard. Developments in the Built Environment, 24 December, p. 100790. Available from: <https://doi.org/10.1016/j.dibe.2025.100790>.

Arksey, H. and O'Malley, L. (2005) Scoping Studies: Towards a Methodological Framework. International Journal of Social Research Methodology, 8 (1) February, pp. 19–32. Available from: <https://doi.org/10.1080/1364557032000119616>.

Ballard, G. and Howell, G. (2003) Lean Project Management. Building Research & Information, 31 (2) January, pp. 119–133. Available from: <https://doi.org/10.1080/09613210301997>.

Berroir, F., Pyszkowski, M., Maatar, O. and Mack, N. (2023) Construction Supply Chain Product Data Integration for Lean and Green Site Logistics [Online]. Lille, France, pp. 1662–1673. Available from: <https://doi.org/10.24928/2023/0154>.

Byers, B. S., Emmenegger, P., Hunhevicz, J. J., Schumm, D., Heisel, F., Hall, D. M. and De Wolf, C. (2025) Decentralized Phygital Identifier Systems for Digital Passports in Circular Construction: A Design Science Evaluation. Construction Innovation, December, pp. 1–35. Available from: <https://doi.org/10.1108/CI-02-2025-0074>.

Çetin, S., Raghu, D., Honic, M., Straub, A. and Gruis, V. (2023) Data Requirements and Availabilities for

Material Passports: A Digitally Enabled Framework for Improving the Circularity of Existing Buildings. Sustainable Production and Consumption, 40 September, pp. 422–437. Available from: <https://doi.org/10.1016/j.spc.2023.07.011>.

Elkhidir, E., Rotimi, J. O. B., Patel, T., Moshood, T. D. and Wilkinson, S. (2025) Toward Standardised Construction Pipeline Data: Conceptual Minimum Dataset Framework. Buildings, 15 (15) August, p. 2797. Available from: <https://doi.org/10.3390/buildings15152797>.

European Parliament and Council of the European Union (2024) Regulation (EU) 2024/1781 of the European Parliament and of the Council of 13 June 2024 Establishing a Framework for the Setting of Ecodesign Requirements for Sustainable Products, Amending Directive (EU) 2020/1828 and Regulation (EU) 2023/1542 and Repealing Directive 2009/125/EC.

Garcés, G., Forcael, E., Osorio, C., Castañeda, K. and Sánchez, O. (2025) Systematic Review of Lean Construction: An Approach to Sustainability and Efficiency in Construction Management. Journal of Infrastructure Preservation and Resilience, 6 (1) January, p. 6. Available from: <https://doi.org/10.1186/s43065-025-00119-1>.

Hamzeh, F. R., Saab, I., Tommelein, I. D. and Ballard, G. (2015) Understanding the Role of "Tasks Anticipated" in Lookahead Planning through Simulation. Automation in Construction, 49 January, pp. 18–26. Available from: <https://doi.org/10.1016/j.autcon.2014.09.005>.

Honic, M., Magalhães, P. M. and Van Den Bosch, P. (2024) From Data Templates to Material Passports and Digital Product Passports [Online]. In: De Wolf, C., Çetin, S. and Bocken, N. M. P. ed., A Circular Built Environment in the Digital Age. Cham: Springer International Publishing, pp. 79–94. Available from: <https://doi.org/10.1007/978-3-031-39675-5_5>.

ISO, I. S. O. (2022) ISO 22057:2022 Sustainability in Buildings and Civil Engineering Works — Data Templates for the Use of Environmental Product Declarations (EPDs) for Construction Products in Building Information Modelling (BIM).

Jaakkola, E. (2020) Designing Conceptual Articles: Four Approaches. AMS Review, 10 (1–2) June, pp. 18–26. Available from: <https://doi.org/10.1007/s13162-020-00161-0>.

McHugh, K., Egan, P. and Dave, B. (2025) Lean Thinking for a Digital Last Planner System [Online]. Osaka and Kyoto, Japan, pp. 517–528. Available from: <https://doi.org/10.24928/2025/0259>.

Mêda, P., Calvetti, D., Hjelseth, E. and Sousa, H. (2024) Digital Product Passports in Construction – Barriers and Opportunities at People, Process and Technology Dimensions [Online]. Available from: <https://doi.org/10.35490/EC3.2024.233>.

Mêda, P., Munir, M., Calvetti, D. and Sousa, H. (2023) Information Needs in Digital Products Passport – Discussing Data Framework and Reasonability [Online]. Available from: <https://doi.org/10.35490/EC3.2023.244>.

Mengyao Li (2025) Psychological Differences in Luxury Consumption between China and the United States and Strategies for Building a Sustainable Consumption Market. Environment and Social Psychology [Online], 10 (11) November. Available from: <https://doi.org/10.59429/esp.v10i11.4252>.

Pedroso, P. F., Silvestre, J. D. and Costa, A. A. (2025) Archetypes of Deconstruction: A Framework Introducing LCA to the End-of-Life. IOP Conference Series: Earth and Environmental Science, 1568 (1) December, p. 012023. Available from: <https://doi.org/10.1088/1755-1315/1568/1/012023>.

Porras Díaz, H., Sánchez Rivera, O. G. and Galvis Guerra, J. A. (2014) Filosofía Lean Construction Para La Gestión de Proyectos de Construcción: Una Revisión Actual. Avances Investigación en Ingeniería, 11 (1) December, pp. 32–53. Available from: <https://doi.org/10.18041/1794-4953/avances.1.298>.

Ribeiro, Y., Teixeira, J., Mêda, P., Moreira, J., Sousa, R. and Sousa, H. (2023a) A Construction Information System as a Lean Information Management Enabler – Case Study [Online]. Lille, France, pp. 104–115. Available from: <https://doi.org/10.24928/2023/0166>.

Ribeiro, Y., Teixeira, J., Mêda, P., Moreira, J., Sousa, R. and Sousa, H. (2023b) Information Systems for Construction 4.0: Classification of Contents for Integration and Interoperability – Case Study [Online]. Available from: <https://doi.org/10.35490/EC3.2023.289>.

Ruismäki, W., Zu Castell-Rüdenhausen, M., Pohjalainen, E. and Wahlström, M. (2025) Stakeholder Perspectives on Digital Product Passports for Construction Products. Cleaner Logistics and Supply Chain, 17 December, p. 100275. Available from: <https://doi.org/10.1016/j.clscn.2025.100275>.

Vangelova, S., Forth, K. and De Wolf, C. (2025) Towards a Compatible Digital Product Passport Data Template in Construction. IOP Conference Series: Earth and Environmental Science, 1554 (1) November, p. 012053. Available from: <https://doi.org/10.1088/1755-1315/1554/1/012053>.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATIC MOUNTAIN TUNNEL INVERT EXCAVATION WITH DIGITAL TWIN MANAGEMENT IN NON-GNSS ENVIRONMENTS

Nobuyoshi Yabuki[1], Hitoshi Ishida[2], Sadatoshi Oomori[2], Yoichi Moriya[2], and Shinji Fujita[2]
[1]Tokyo City University, Tokyo, Japan
[2]Penta-Ocean Construction Co., Ltd., Tokyo, Japan

Abstract

This research proposes a real-time tracking method using Light Detection and Ranging (LiDAR) to capture reference points as a positioning technique for non-Global Navigation Satellite System (GNSS) environments, such as mountain tunnel construction sites. Using this to acquire backhoe coordinates during tunnel construction enabled the automation of rough excavation for the invert excavation. Furthermore, by providing an overview visualization of the processes before and after automated construction, the surrounding conditions, and related tasks on a digital twin, it demonstrated safe automated construction management from remote locations. This research contributes to the automation and unmanned operation in mountain tunnels.

Introduction

In open-cut earthwork, high-precision positioning of construction machinery using the Global Navigation Satellite System (GNSS) enables semi-automated machine operation. In contrast, underground spaces such as mountain tunnels are non-GNSS environments where GNSS signals cannot be received, making high-precision position estimation difficult and hindering automation. For mounting tunneling, either the conventional timbering method or the New Austrian Tunneling Method (NATM) is employed. Among mountain tunnel construction methods, invert construction involves excavating the tunnel floor into an inverted arch shape (concave bottom) and reinforcing it with concrete. This aims to withstand pressure from below in weak ground, enhance overall structural strength, and facilitate drainage. Excavation work in invert construction requires both precise excavation shapes and the ability to move autonomously during operations, making position estimation critical and automated construction particularly difficult. The finished form of the invert is managed to ensure its thickness meets or exceeds the design value, typically limiting over-excavation to within a few centimeters. Excessive over-excavation increases both the volume of soil disposal and the amount of concrete required. Therefore, from the perspectives of work efficiency and environmental consideration, it must be avoided.

Furthermore, in mountain tunnel construction, multiple types of works typically alternate while working in confined spaces. This necessitates meticulous coordination by tracking the positions and operational status of construction machinery, materials, and workers. Implementing autonomous operation requires not only high-precision position estimation for the targeted machinery but also real-time, accurate awareness of its spatial relationships and operational status relative to other trades. Traditionally, site staff and workers visually observed conditions and communicated status via radio or similar means. Recently, however, digital twins have been employed as a method to visualize real-time conditions at construction sites during operations. Digital twins offer the advantage of enabling remote personnel to intuitively grasp the entire site, making it easier to understand the relationship between autonomous operation, surrounding tasks, and preceding/subsequent work.

On the other hand, mountain tunnel construction has historically been plagued by fatal rockfall accidents. Additionally, significant safety risks exist, such as water ingress and the generation of combustible gases. There are also substantial health risks, including dust generation from excavation, muck hauling, concrete spraying, and rock bolt installation. Consequently, reducing or eliminating personnel inside the tunnel is highly desirable. Therefore, it is also desirable to enable remote management for personnel overseeing operations within the tunnel. On the ground of the background, this study targets mountain tunnel construction in non-GNSS environments. It aims to propose and demonstrate the usefulness of automated invert excavation operations and a management method for automated operations using digital twins for remote monitoring.

Thus, the research questions of this paper are as follows; 1) How can we automate the invert excavation work of mountain tunnels in non-GNSS environments? 2) How can digital twins be utilized for the automated and unmanned invert excavation operation by safe remote monitoring?

Previous research

Automation in mountain tunnel construction

Mountain tunnel construction typically involves multiple types of work processes in succession within confined spaces. While automation research has advanced for NATM excavation tasks such as drilling and blasting, shotcrete application, steel support installation, and rock bolting, no automation efforts have been observed for

invert excavation. This is likely due to the challenges involved in invert excavation: the backhoe must acquire precise coordinates, move to the excavation position based on those coordinates, and control the excavation depth. Acquiring accurate coordinates while moving within a mountain tunnel is difficult. Furthermore, there are no initiatives for automatically moving to construction sites or automatically switching between work types. This is likely because all work types require auxiliary manual labor on-site either before, after, or during automated construction.

Yuhong et al. (2022) and Zijian et al. (2023) proposed using digital twins to visualize safety risk management in tunnel construction. Mountain tunnels face numerous daily risks, such as face collapses, water inflows, and combustible gas emissions, with conditions constantly changing. By consolidating site data—including ground conditions and worker locations—into a digital twin, it becomes possible to manage overall site operations and predict hazards. As mentioned earlier, digital twins can be utilized as a management method to visualize the flow of work types and worker assignments, enabling appropriate automated construction even in environments where human labor is mixed. Li et al. (2024) clustered many papers on digital twins for intelligent tunnel construction into various areas, including sensor networks, Internet of Things (IoT), computer vision-based twin data acquisition, communications, natural language processing (NLP), automatic control-based connection, and geometric, semantic, and analytical integrated twin modeling. However, no studies have been found that have conducted automated invert excavation of mountain tunnels and employed digital twins for unmanned safe construction and monitoring in non-GNSS environments.

Self-localization in non-GNSS environments

Nakao, Wakayama et al. (2023) conducted position estimation experiments for transport trucks in outdoor yards using LiDAR Simultaneous Localization and Mapping (SLAM). They verified accuracy at varying speeds and reported that while errors increased at higher speeds of 10, 20, and 30 km/h, position estimation was generally achievable with an error of 30–40 cm. They also reported that for driving within non-GNSS environments like tunnels, position estimation was possible with an error of approximately 30–40 cm, 20 and 30 km/h, and the error generally remained within 30–40 cm. They concluded that this method is practical for operation within tunnels, which are non-GNSS environments. Furthermore, Ishida et al. (2022) demonstrated that during automated construction targeting face excavation work (using breakers) in a mountain tunnel, LiDAR SLAM enabled self-position estimation with an error of approximately 30–40 cm and allowed for automated movement while maintaining awareness of its own position. However, these methods struggle to automate excavation tasks requiring quality control with sub-centimeter accuracy, such as invert excavation. The cause of this error is thought to be that the road surface of mountain tunnels under construction is not flat. The laser beams from the LiDAR used in SLAM encounter the tunnel walls or various equipment installed on the walls due to minor vibrations caused by the vehicle's movement, causing distance fluctuations. Depending on whether the beams hit the tunnel wall or this equipment, differences of several tens of centimeters occur in the measured distance. Resolving this issue would require a significant overhaul of the current construction equipment, which is not practical.

Yimin et al. (2022) used AI to detect objects and acquire targets, enabling real-time coordinate alignment for flying UAVs. Similarly, if coordinates for construction machinery inside mountain tunnels could be acquired from known points, more accurate self-position estimation would be possible, facilitating smooth automated construction.

Proposed method

The method proposed in this study consists of two approaches: a technique for high-precision automated invert excavation in mountain tunnels under non-GNSS conditions, and a technique for managing automated operations by remotely monitoring site conditions via a digital twin.

Moving to the worksite and improving localization accuracy

The constructed autonomous machine is shown in Figure 1. To achieve automated invert excavation, precise automatic movement to the excavation site is essential. Therefore, a 2D LiDAR was installed at the front of the vehicle to acquire surrounding point clouds in real time. Regarding the automation of backhoe control, an unmanned construction backhoe (a remote-controlled machine for disaster response) equipped with a hydraulic control system was modified. Four inclinometers for machine guidance (mounted on the body, boom, arm, and bucket link) and a slewing angle sensor measuring the slewing angle between the traveling section (lower part of the backhoe) and the working section (upper part of the backhoe) were installed. This enables the computer to determine the backhoe's posture, and a slewing angle sensor measuring the angle between the undercarriage (lower part of the backhoe) and the upper structure (upper part of the backhoe). This enables computer acquisition of the backhoe's posture. Furthermore, by controlling the hydraulic controller via this computer, a system was created that performs hydraulic control while acquiring travel and posture data. This system can calculate the cutting edge coordinates with the same precision as the guidance system for manned operation. Therefore, this section focuses on the positioning of the backhoe body itself, which is expected to cause particularly large errors in automated construction. It describes methods for acquiring the surrounding environment to enable accurate automated movement and techniques for moving precisely to the excavation position.

Previous studies have shown that SLAM using 2D LiDAR can achieve positional accuracy within 30 to 40 cm when precise positioning is not required, such as for movement

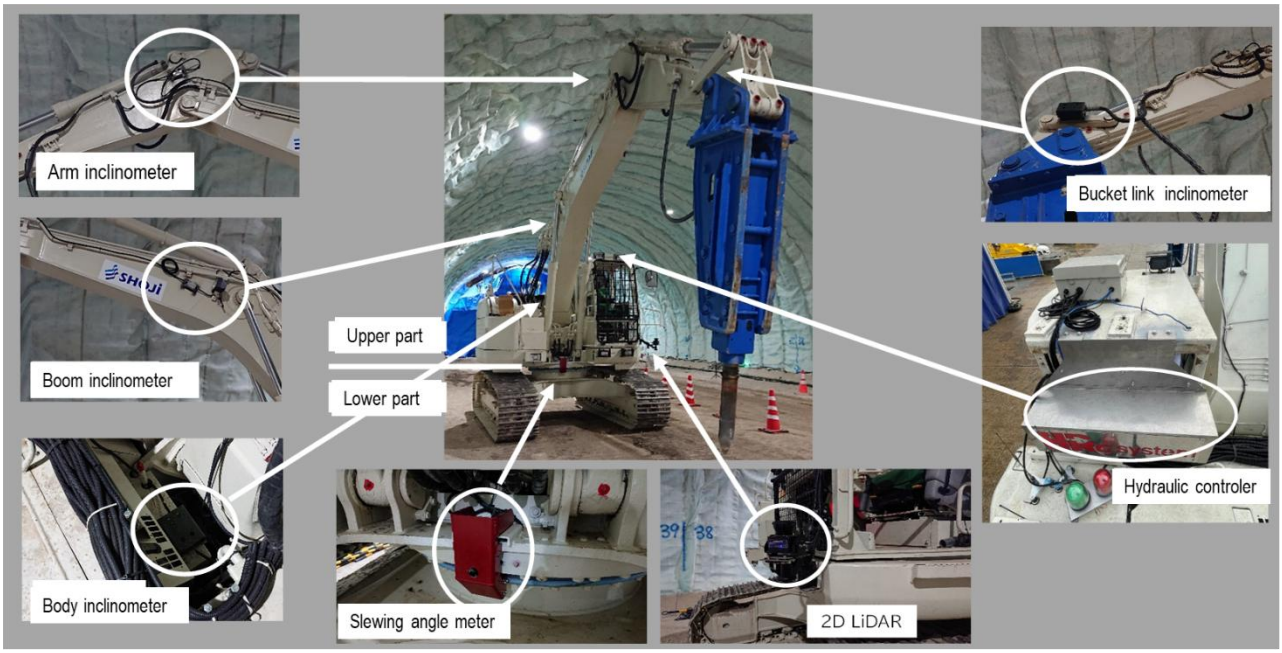


Figure 1: Developed autonomous backhoe for tunnel invert excavation

within a tunnel bore. However, the breaker tip requires accuracy within 5 cm, and the construction precision of the machine guidance system used in this study is also within 5 cm. Therefore, positional accuracy of at least within 5 cm was also deemed necessary. Therefore, as shown in Figure 2, two cylindrical reference points (barrier drums) were installed at the construction site to improve position estimation accuracy. Barrier drums were chosen as reference points because they are numerous on-site, have sufficient diameter and weight for high stability, and their large diameter allows acquisition of a large number of LiDAR measurement points. A typical and used barrier drum's dimensions are D580 mm x H820 mm, and 200 liters.

The procedure is as follows:

1) Pre-install the barrier drums near the construction site, acquire the center coordinates of the circle using a total station (TS), and register them in the system.
2) Drive the backhoe while scanning its front using the 2D LiDAR mounted on the backhoe body. Detect objects. When three or more points are acquired, perform Circle Fitting (Figure 3) using the XY coordinates to determine the shape. If the shape is a cylinder of a specified diameter, it is considered a calibration point.
3) When two calibration points are captured, the center coordinates from Circle Fitting are set to the coordinates acquired by the TS.
4) The coordinates of the two points are used to estimate the autonomous vehicle's own position and the orientation of the working equipment.

Detecting two calibration points enables calculation of the vehicle body's orientation. This method achieves higher accuracy than SLAM. The 2D LiDAR used in this research is VFT-35LX-F0 and detects calibration points in real time. The ranging accuracy of the used LiDAR is ±5 cm within 15 m and ±10 cm beyond 15 m when measuring white paper (a highly reflective target).

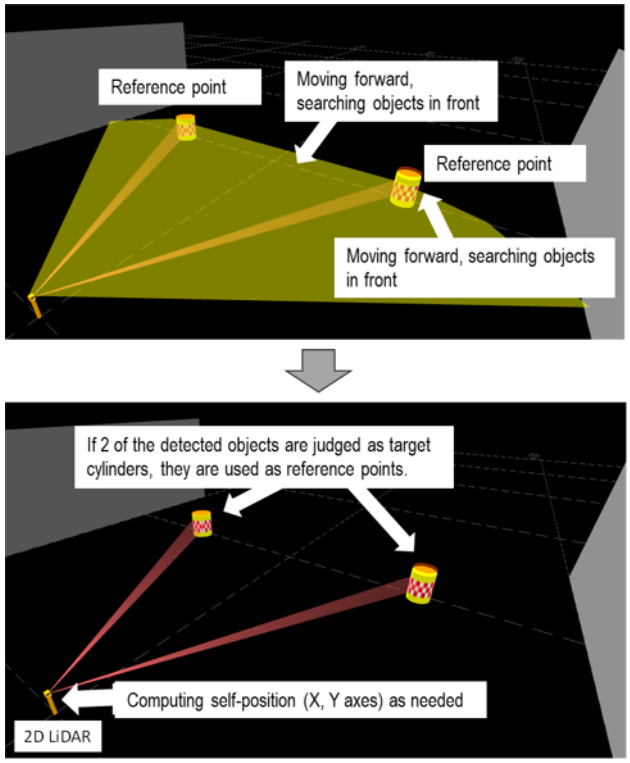


Figure 2: Measuring self-position using reference points

Sufficient accuracy may not be achievable at long distances. In contrast, this study assumes automated construction per block (primarily 10.5 m in length, based on the standard 10.5 m length of a typical tunnel lining form (traveller) as an example of a construction unit for mountain tunnel projects. When the autonomous machine enters within 10 m of a reference point, the system automatically switches from position estimation via SLAM to position estimation using the reference points,

thereby improving accuracy. These functions enable automated invert excavation in non-GNSS environments, demonstrating the proposed method's effectiveness.

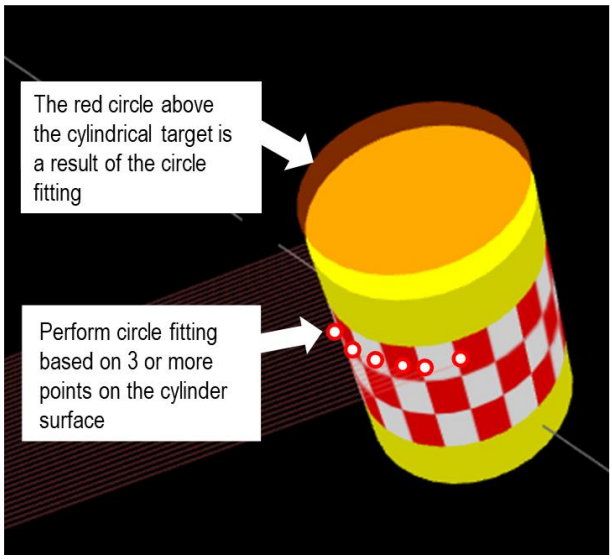


Figure 3: Detection of reference points using point cloud data observed by the 2D LiDAR

Digital twin-based remote site monitoring and automated construction management/control

Figures 4 and 5 show the digital twin used in this research. Figure 4 displays the WebGL view, showing the overall status viewed via a web browser, while Figure 5 shows immersion in the digital twin via VR. Figure 6 shows the system architecture of the digital twin developed in this study. This research proposes using the digital twin for unified management of automated construction and manual operations, anticipating scenarios where automated operation and manual work coexist on-site. This digital twin reflects the site conditions in real time, enabling intuitive grasp of the entire site situation from remote locations. The digital twin displays not only the design geometry but also as-built measurement results and IoT sensor measurement data. For IoT sensors, real-time acquisition of dust concentration, illuminance, and the signal strength of BLE (Bluetooth Low Energy) beacons worn by workers was performed to manage the work environment and worker locations. For the invert excavation section targeted for automation and its surrounding sections, managing the presence of manual labor is critical for safety assurance. Therefore, worker positions acquired via BLE beacons are reflected in the digital twin, and autonomous machines are restricted from entering sections where manual work is in progress.

After invert excavation completes in a section, manual formwork and rebar assembly enter the same section, while the invert excavation moves to an adjacent section. While automated construction and manual work alternate and proceed adjacent to each other, the digital twin manages and controls the automated construction. This allows supervisors to visually confirm the situation while also preventing automated construction from occurring in manual work sections through system control. The digital twin pre-registers the designed shape for invert excavation. The current terrain is captured using a 3D laser scanner immediately before invert excavation begins. This data is imported into the digital twin, which then automatically generates planning data based on both the current terrain and the designed shape (Figure 7). These functions enabled remote monitoring of the site conditions, control of automated operations while separating manned work from automated tasks, and demonstrated the effectiveness of this approach.

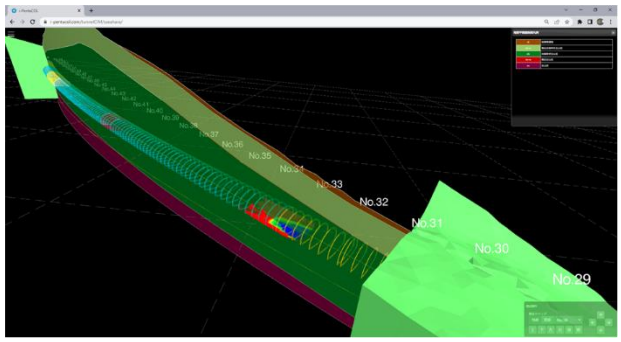


Figure 4: Image of the whole construction site in the developed digital twin (WebGL)

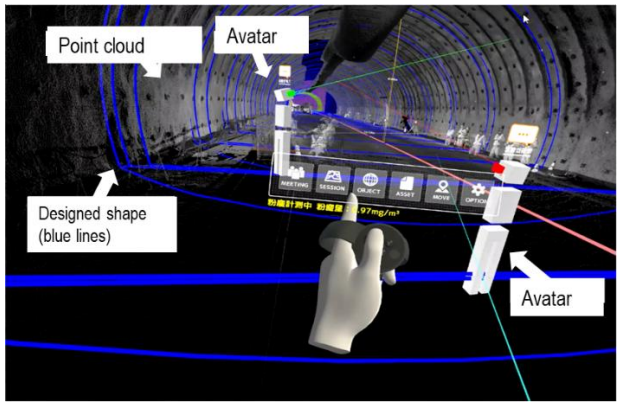


Figure 5: VR view of the tunnel construction site on the developed digital twin

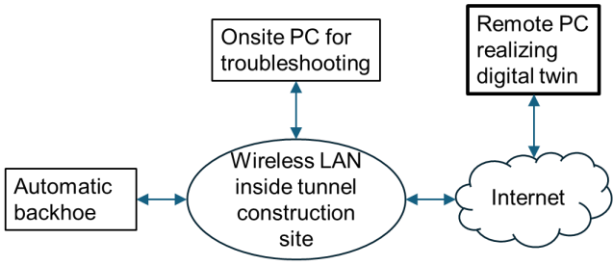


Figure 6: System architecture of the developed digital twin

Demonstration at a tunnel construction site

The proposed method in this study was demonstrated at the National Route 57 Sasahara Tunnel construction project (Location: from Sasahara-cho to Amatsu-cho, Uto City, Kumamoto Prefecture, Japan) commissioned by the Kyushu Regional Development Bureau. This is a mountain tunnel construction project with a width of approximately 14 m, a height of approximately 8 m, and a total length of 679 m.

The autonomous machine moves from the standby location to the construction site using SLAM-based position estimation, automatically avoiding temporary obstacles along the way. After detecting cylindrical reference points, it switches to position estimation based on these points. Subsequently, it automatically moves the cutting edge to the pre-set excavation location and performs excavation using a breaker. The accuracy verification method involved stopping the breaker at its deepest point during excavation and measuring the deviation of the guided cutting edge. The excavation guidance status display is a schematic diagram showing the spatial relationship between captured reference points, the automated machine, and the mesh diagram of the construction target area. The number of mesh points increases as the construction target area expands.

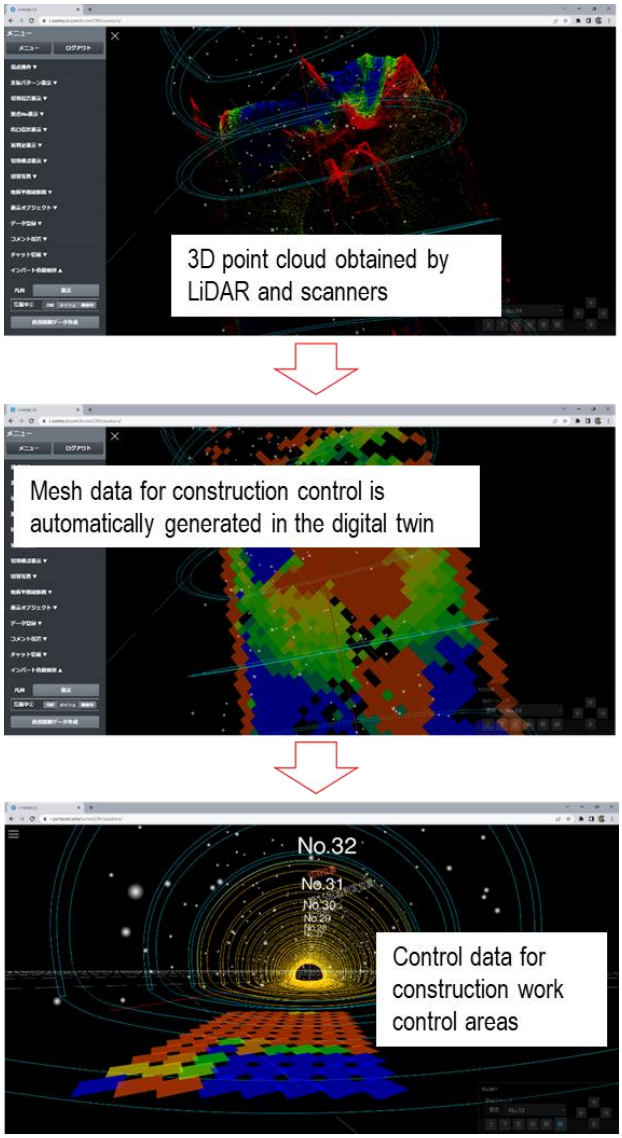


Figure 7: Automated generation of autonomous construction plans in the developed digital twin

The vehicle's front-rear direction was defined as the X-axis, left-right direction as the Y-axis, and depth direction as the Z-axis. For depth measurement, a marker was installed on the upper part of the breaker, and depth was calculated by measuring the height of this marker. The accuracy verification results are shown in Table 1. When an operator is in the cabin performing manual excavation, the positional accuracy of the breaker tip is approximately 5 cm. In this automated operation demonstration, the average error exceeded 5 cm, indicating overall lower accuracy compared to manual operation and unsuitability for finishing excavation. However, for rough excavation targeting 50 cm above the finishing excavation level, an error margin of up to 50 cm is acceptable. Therefore, high precision is not required, and it is sufficiently applicable. Regarding measurement point 1 in Table 1, the depth-direction error is larger compared to other measurement points. For mountain tunnel applications, the breaker unit is heavier than conventional units, and it has been confirmed that this causes greater vibration at points farther from the vehicle body. This is considered to be the cause of the error. Overall, vehicle body vibration is significant when the arm is extended, and the cutting edge often wobbles considerably before reaching the intended position. Therefore, for the end construction, it was judged that contact with the wall surface was possible, and construction was performed using onboard operation.

Table 1: Accuracy verification result of excavation control

Measuring point No.	Average error for each axis		
	x (cm)	y (cm)	z (cm)
1	8.3	9.2	22.3
2	6.0	8.7	12.3
3	5.7	7.0	16.0
4	7.0	5.7	9.3
5	10.3	8.3	12.3
6	7.0	6.0	8.3
7	10.0	8.3	7.7
8	8.3	7.7	5.7

During invert excavation, three workers typically measure the as-built condition on-site as needed to guide the excavation to the specified depth. This not only requires frequent manning of the construction site but also necessitates three workers standing by nearby during excavation. In contrast, this demonstration method implemented automated as-built management using 3D LiDAR (Figure 8). Consequently, invert excavation proceeded unmanned, and measurements could be performed with only one surveyor stationed away from the construction site. This demonstrates that merely automating the construction machinery itself is insufficient to achieve an unmanned construction site. To realize unmanned operation, it is necessary to automate or remote-control associated tasks such as as-built verification.

The preceding process for the invert excavation work in this study was rock bolt installation. However, sufficient time had elapsed since the preceding process, so no particular work coordination was required. Furthermore, since rock bolt installation and invert work are not

performed consecutively, it is considered that no similar work coordination would be required outside this demonstration project. Figure 9 shows an example of the area of a single invert excavation and the surrounding conditions. Since invert excavation reduces ground load, excavation is often performed alternately on each side. After excavation, the schedule is typically adjusted to perform concrete placement and backfilling as early as possible. To use the excavated soil for backfilling during invert excavation, temporary stockpiles are often placed nearby for construction efficiency. Consequently, the actual site is divided into even smaller sections, with manual labor occurring. Manual labor is performing formwork and rebar assembly for the post-invert excavation process. The left side served as a passageway for transporting materials. In the next phase, invert excavation moves to the opposite side. Manual concrete placement for the invert occurs at the tunnel rear, and the completed excavation side becomes the concrete transport route. Managing these chronologically changing conditions around automated construction is also necessary. This study utilized radio signals from BLE beacons worn by workers to manage the location of manual operations, monitor formwork/rebar assembly status and material transport status, and automatically halt automated construction when workers or vehicles approached the automated machinery. However, frequent material transport causes delays in automated construction, necessitating measures to relocate transport routes away from the construction zone.

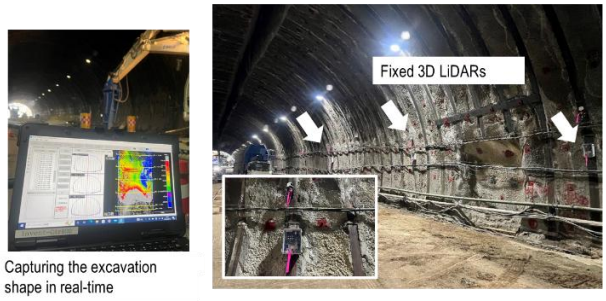


Figure 8: Fixed 3D LiDARS were used for automated as-built management

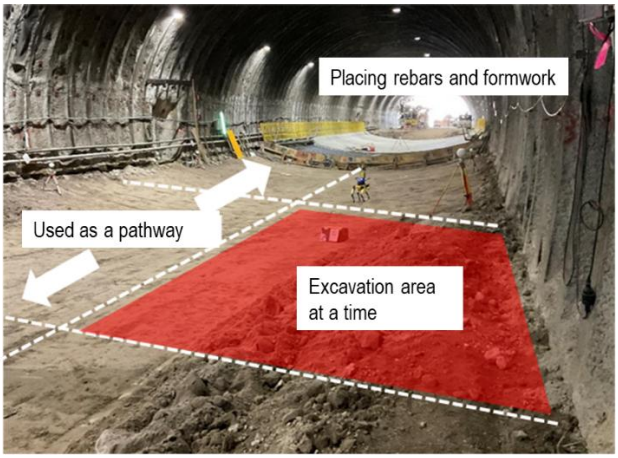


Figure 9: An invert excavation area at a time

Regarding the demonstration content and results, multiple veteran site supervisors with extensive mountain tunnel construction experience evaluated the system. The evaluation was done both in writing and interviewing in terms of precision, safety, productivity, and miscellaneous of automation of invert excavation and the digital twin. Initially, these supervisors considered automated construction in mountain tunnels impractical due to the tunnel's confined space and the complex interplay of numerous work types requiring frequent fine-tuning between tasks. However, through this research demonstration, they gained confidence in the feasibility of automated construction in mountain tunnels. This feasibility stems not only from the automation of construction machinery but also from establishing a system where digital twins enable personnel to oversee the entire site while allowing human management intervention. This allows progress to be made by combining automated construction with conventional methods, even in response to minute changes in conditions.

Conclusions

This research demonstrated two key elements for realizing automated construction in mountain tunnels: 1) automation of invert excavation in non-GNSS environments, and 2) management of automated construction using a digital twin. Regarding 1) automation of invert excavation in non-GNSS environments, while its lower construction accuracy compared to operator-controlled excavation limited its application to rough excavation, it was confirmed that even for rough excavation alone, eliminating the need for heavy equipment operators during simple repetitive tasks offers significant labor-saving potential. According to site personnel, automating even just rough excavation allows operators to spend about half a day on other tasks. 2) Regarding automated construction management and control using digital twins, beacons enabled clear identification of both manual worker locations and automated construction positions. Furthermore, integrating human vision and intuitive judgment through the digital twin made it possible to proceed with a mix of automated construction and conventional manual work.

Furthermore, even with automated construction machinery, frequent on-site tasks such as measuring work progress and guiding during automated construction occur. It was found that advancing automation in these areas also improves the efficiency of automated construction. Moving forward, visualizing movements during handoffs between different work types using the digital twin and defining the operating conditions of automated machines according to each set of conditions will enable smooth coordination with other work types.

Regarding position estimation methods, the approach used in this study requires the installation and removal of calibration points, leading to additional on-site work. At the construction site, only 2 barrier drums were used for referencing. The effects of the number of barrier drums and their locations are our future work. An alternative

approach could involve assigning coordinates to existing feature points, such as formwork, rock bolts, or immovable temporary equipment, and performing position estimation by automatically detecting these points.

While this research managed invert excavation using a digital twin, applying this approach to other trades is feasible. By acquiring detailed real-time site information via 3D data and IoT data, human intervention can be reduced. Future research should explore integrating digital twin-based safety risk management, management of manned/unmanned zones, and automated construction control.

Mountain tunnel construction involves multiple trades operating within a narrow, extended space, necessitating detailed understanding and meticulous coordination of site conditions. While research on automated construction for mountain tunnels has advanced across many trades, an integrated coordination mechanism is now required. As automation and remote operation advance, we aim to proceed with introducing automated and remote construction while analyzing how many workers will remain in the tunnel for each task and the workflow for progressing each task

Acknowledgment

Field verification of this methodology was conducted with funding from the Ministry of Land, Infrastructure, Transport and Tourism's "Project on the Introduction and Utilization of Innovative Technologies to Dramatically Improve Construction Site Productivity."

References

Ishida, H., Yabuki, N., Oomori, S., Moriya, Y. & Fujita, S. (2023) Development of an automated driving method for rock scaling in mountain tunnels in a non-GNSS environment. Journal of Japan Society of Civil Engineers, F3 (Civil Engineering Informatics), 79 (22).

Li, T., Li, X., Rui, Y., Ling, J., Zhao, S., & Zhu, H. (2024) Digital twin for intelligent tunnel construction. Automation in Construction, 158, 105210.

Nakao, Y. & Wakayama, M. (2023) Autonomous driving experiment of 2-ton truck using SLAM technology. JSCE 78th Annual Conference, VI-795.

Yimin. Z, Zhixiong, Y. & Zhuang, M. (2022) UAV based indoor localization and object detection, Frontiers in Neurorobotics, 16.

Yuhong, Z., Naiqiang, W. & Zhansheng, L. (2022) An established theory of digital twin model for tunnel construction safety assessment. Applied Sciences, 12 (23).

Zijian, Y., Ying, Y., Chengping, Z., Zhiming, Z., Wei, L., Xuejie, W., Lei, W. & Libin, W. (2023) A digital twin approach for tunnel construction safety early warning and management. Computers in Industry, 144.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

PROACTIVE SAFETY MANAGEMENT IN EARTHWORK OPERATIONS USING IOT-DRIVEN REAL-TIME RISK ANALYSIS

A N Ledin Ullah[1], Tulika Choudhury Puja[1], Jhonattan G. Martinez[1,3] , Lorenzo Loyola[1], Istiaque Hossain[2] and Freddy Perez[4]
[1] Aarhus University, Aarhus, Denmark
[2] Samsung R & D Institute, Dhaka, Bangladesh
[3]Fundacion De Educacion Superior San Jose, Colombia
[4]The Cooperative University of Colombia

Abstract

Earthwork operations are among the most hazardous construction activities due to frequent interactions between heavy machinery and workers in dynamic, constrained environments. Traditional safety practices remain largely reactive and insufficient for real-time accident prevention. This paper presents an IoT-driven real-time risk analysis framework for proactive safety management in earthwork operations. The system integrates machine-mounted telematics, spatial analytics, and dynamic heat-map visualization to assess collision risks at machine and zone levels. A three-week deployment on a construction site identified spatiotemporal risk hotspots and recurring near-miss patterns using meter-level positioning accuracy, demonstrating a shift from reactive control to proactive, data-driven safety management.

Introduction

In European construction, particularly in Denmark, earthwork operations pose a significant hazard due to frequent collisions involving heavy machinery. Confined workspaces amplify these inherent risks, leading to a high rate of severe accidents. Poor on-site conditions and human factors can lead to equipment-worker collisions, jeopardizing safety and operational efficiency (Son and Kim, 2021; Talmaki and Kamat, 2022). Machinery collisions are a significant concern on all construction sites. The close-quarters operation of heavy equipment, combined with poor visibility and miscommunication among operators, frequently leads to serious accidents. The shared workspace between workers and machinery increases the risk of injuries, making it essential to thoroughly investigate mitigation strategies (Kim et al., 2023). Existing methods, which rely on manual supervision and post-incident analysis, are reactive and insufficient for preventing collisions. These approaches often cause costly delays and make safety hazards worse (Soltanmohammadlou et al., 2019; Zhang et al., 2017) Therefore, the motivation to drive this research is to make a transition from reactive to proactive safety management systems that can effectively predict and prevent incidents on construction sites through innovative technology.

Despite advancements, construction safety still relies heavily on manual oversight and post-incident reviews. The use of flagmen and basic rear cameras exemplifies a reactive safety approach, underscoring the need for a shift toward real-time, technology-driven risk management. . There is a notable gap in using real-time data to prevent equipment collisions during rapidly changing earthwork operations (Waqar et al., 2023). Incorporating technologies such as the Internet of Things (IoT), Geographic Information Systems (GIS), and machine learning is crucial for advancing construction safety beyond traditional methods. However, current practices often fail to fully utilize these tools to build a cohesive, real-time monitoring system that adapts to the dynamic nature of construction sites. Existing literature lacks studies that show how to effectively integrate technology into safety systems that both identify risks and provide real-time insights to stakeholders (Talmaki and Kamat, 2022). The absence of real-time, tech-enabled risk assessment frameworks remains a significant barrier to safer construction sites (Häikiö et al., 2020). Dynamic risk assessments that combine real-time data with visualization tools, such as heat maps, enable swift hazard identification, supporting quick, informed decisions (Talmaki and Kamat, 2022). This research presents a novel, integrated IoT framework for construction safety that explicitly addresses the current gap by linking machine telematics, spatial analytics, and dynamic heat-mapping visualization to enable proactive safety management during earthwork operations. This research emphasizes equipment-based risk monitoring and does not address interactions between workers and equipment, which continue to be a major cause of fatal accidents. As a result, the proposed framework should be viewed as a partial safety measure rather than a complete safety system. The proposed system provides continuous, real-time spatial risk awareness, enabling immediate on-site decision-making, unlike traditional post hoc or isolated sensing approaches.

Literature review

State of practice in safety management for earthwork operations

Across the EU, 20% of accidents and deaths occur on construction sites (Hong and Teizer, 2024). Construction remains one of the most hazardous industries in the UK, accounting for 27% of workplace deaths in 2020/21, with 18% caused by moving vehicles (Walker et al., 2021). Earthwork operations, involving heavy machinery like

excavators and bulldozers, pose significant safety risks, particularly collisions, which research shows are a leading cause of injuries (Shen et al., 2016; Teizer, 2015). Sites with multiple heavy equipment units operating simultaneously can create confusion regarding navigation and operational priorities. The presence of numerous personnel can further complicate coordination and increase the likelihood of accidental collisions (Kukoyi and Adebowale, 2021). Trenching and excavation activities are particularly hazardous due to the risk of collapse. OSHA reports that trenching cave-ins are among the top causes of construction-related fatalities, often resulting in severe injuries or deaths (Asadzadeh et al., 2020). The primary causal factors include inadequate protective systems and geological instabilities that are insufficiently assessed before excavation. Furthermore, Poor planning and unclear equipment paths often cause construction accidents. Research has linked inadequate design and maintenance negligence to collisions, highlighting the need for careful planning and regular inspections to ensure safe operation of machinery (Zhang et al., 2017).

Traditional safety management on construction sites during earthwork operations with heavy equipment has primarily relied on manual inspections, visual checks, and physical barriers to reduce the risk of collisions. For Instance, communication methods, such as walkie-talkies or visual signals, are not always effective in ensuring timely warnings, especially in sites with high noise levels or low visibility (Jo et al., 2019). Consequently, the risk of collisions among heavy equipment persists, resulting in serious accidents and injuries. Other methods, such as physical barriers and cones to delineate safe work zones, were prominent but not foolproof; they did not account for all variables that could lead to collisions between equipment and personnel (Kayhani et al., 2018).

Enhancements in communication, such as alarms and visual alerts, were limited and did not provide comprehensive real-time proximity awareness (Jo et al., 2019). Safety performance was largely reactive, focusing on accident responses rather than proactive hazard assessments and preventative measures (Leonardo et al., 2020).

Digital Technologies in Safety Management of Earthwork Operations

Innovative technologies like IoT are transforming construction safety through real-time data collection. Using wearable sensors and environmental monitoring, IoT enables continuous tracking of workers, machines, and site conditions. Similarly, Asadzadeh et al. (2020) discuss how sensor-based safety management integrated with IoT technologies improves responsiveness to dynamic working conditions (Asadzadeh et al., 2020). This shift towards automated, proactive monitoring systems has emerged as a necessity given the inadequacies of passive safety measures traditionally employed in construction settings (Häikiö et al., 2020; Kim et al., 2023). Another key innovation, heat maps, offers visual data representations that enhance decision-making for worker safety and operational efficiency.

Wearable sensors now play a vital role in construction safety by providing real-time monitoring to identify hazards before accidents happen (Zamani et al., 2020). Strategies for sensor placement, such as those by (Talmaki and Kamat, 2022), improve machinery safety. Positioning technologies like Global Positioning System (GPS), Radio Frequency Identification (RFID), Light Detection and Ranging (LiDAR), and computer vision enable real-time tracking of equipment and workflows(Kim and Xiong, 2025). GPS supports spatial awareness on large sites, reducing near-miss incidents (Cheng and Teizer, 2013), while LiDAR offers 3D mapping to identify high-risk zones (Rao et al., 2022). RFID enhances equipment tracking and accountability (Poudel and Assaad, 2025). Together, these technologies create integrated safety systems that improve visibility and hazard detection on construction sites. Golovina et al. developed heat maps that visually represent high-risk areas where worker-equipment interactions are most likely, supporting predictive safety planning (Golovina et al., 2016). These visuals help safety managers identify hazards and develop mitigation strategies by prioritizing areas needing immediate attention. Heat maps are also widely used outside construction, such as in traffic safety management, to analyze accident hotspots and implement preventive measures (Chung et al., 2023). They highlight high-traffic equipment zones, enabling proactive interventions before incidents occur. Managers can modify equipment routes or schedules based on this data to reduce interactions and risks.

A recent perspective by Bęś et al. (2025) highlights the integration of IoT sensors with digital twins, GIS/BIM, and real-time visualization for construction safety. It emphasizes sensor networks, RTLS, and connected systems for dynamic site monitoring, enabling data-driven dashboards and risk visualization (Bęś et al., 2025). Modern safety management combines automation, real-time visualization, and dynamic risk assessment, but fully integrated solutions remain limited. Advances in IoT, digital twins, and AI now allow automated systems to track hazards, assess risks, and support decisions through live dashboards and heatmap-style risk views. An ideal system would unify hazard detection, risk prioritization, and KPI monitoring into a single, scalable platform for real-time safety management (Khajavi et al., 2023)

Research gaps & objectives

The integration of IoT-based monitoring systems with heat map generation holds significant promise for improving safety in earthwork operations, yet several gaps limit its effectiveness. One identified gap is that current systems lack real-time data analytics capable of processing high-volume sensor data quickly enough to provide immediate hazard detection and dynamic heat map updates (Liu et al., 2021; Niu et al., 2019). Moreover, most tools rely on historical data rather than proactive hazard prediction models, thereby limiting their

ability to anticipate collision risks. Machine learning approaches are needed to analyze equipment movement and generate predictive, forward-looking heat maps (Moe et al., 2023). Real-time integration of IoT-based construction safety systems remains constrained by data heterogeneity, latency, and weak governance, which limit the seamless integration of sensor streams (e.g., wearables, RTLS, environmental data) with dynamic heatmap or risk visualization outputs. Even where real-time dashboards exist, they often rely on static thresholds or simplified models, reducing the accuracy and responsiveness of risk prioritization in fast-changing site conditions. Additionally, many solutions lack robust governance frameworks that link sensor data to safety actions, decision logs, and audit trails across the project lifecycle, raising concerns about data reliability and regulatory compliance. While digital twin approaches offer promise for auditable and traceable safety management, their large-scale deployment e,g. With consistent rule versioning and intervention tracking, it remains underdeveloped. Furthermore, existing studies are often limited to specific use cases, with insufficient evidence on how these integrated systems perform across design, procurement, and on-site execution, hindering broader industry adoption (Arya et al., 2025; Catalán et al., 2025). Addressing these gaps calls for research that advances real-time analytics, develops predictive modeling capabilities, and incorporates user-centered design principles to enable practical adoption and enhance the effectiveness of digital safety management in construction. This study contributes by delivering a practical, real-time IoT-integrated safety framework that unifies data, visualization, and actionable risk monitoring, bridging the gap between fragmented tools and scalable, proactive construction safety management

Methodology

This section outlines the core theme of developing a real-time risk management system, including the design framework, data collection methods, and technology integration strategy used to enable proactive safety monitoring in earthwork operations. This study will follow a Design Science Research (DSR) methodology structured into four sequential steps:

1. *Identification of research gaps and system requirements* through a systematic literature review and expert consultation.
2. *Evaluation Metrics for Safety Performance* to evaluate system performance, a set of safety KPIs, including proximity violations, duration of unsafe interactions, near-miss detection, and severity classification based on a risk matrix, was defined.
3. *System design and technical development* of IoT-based heat map solutions integrating real-time analytics and predictive modeling.

Step #1: system requirements:

This stage ensured technical robustness and practical relevance by grounding the system design in two approaches: (1) a systematic literature review (on IoT safety, positioning, and heat maps), and (2) semi-structured interviews with three experienced safety managers to translate theoretical advances into operational needs.

The project requirements spanned three areas: Functional (system capabilities, including continuous tracking and heat map generation), Technical (performance, with a focus on low-latency, real-time processing of diverse data streams), and User-Oriented (practicality for safety managers, requiring an intuitive interface and clear, actionable visuals).

Step #2: evaluation metrics for safety performance

To strengthen the evaluation of the proposed system, a set of quantitative safety performance indicators (KPIs) was defined based on proximity analysis, temporal patterns, and risk severity classification. These indicators provide a structured framework to assess unsafe interactions and interpret safety conditions beyond simple tracking performance. The primary KPIs include: (1) proximity violations per operational hour, representing the frequency of safety threshold breaches between machines or between machines and static hazards; (2) duration of unsafe interactions, capturing the exposure time during which equipment operates within unsafe proximity ranges; and (3) severity classification of safety violations, derived from a dynamic 5 × 5 risk assessment matrix combining likelihood and consequence levels. In this study, near-miss events are defined as interactions occurring within critical distance thresholds (e.g., <5 m), thereby enabling the identification of high-risk situations before actual incidents occur.

Table 1: Assessment of metrics

KPI	Description	Unit
Proximity violations per hour	Number of safety threshold breaches between entities	events/hour
Duration of unsafe interactions	Time spent within unsafe proximity ranges	Seconds
Severity classification	Risk level based on likelihood × consequence matrix	Categorical (1-5)
Near-miss events	Interactions below the critical distance threshold (<5 m)	events

Step #3: system design and implementation:

Employing a Design Science Research (DSR) methodology, this study developed a real-time safety monitoring framework. The system uses IoT sensors on heavy machinery to collect positional and operational data. This data is managed, stored, processed, and analyzed efficiently by a hybrid InfluxDB (time-series database)–AWS architecture, which facilitates dynamic heat map and risk matrix visualization. The system comprises three layers: sensor, database, and visualization (Figure 1).

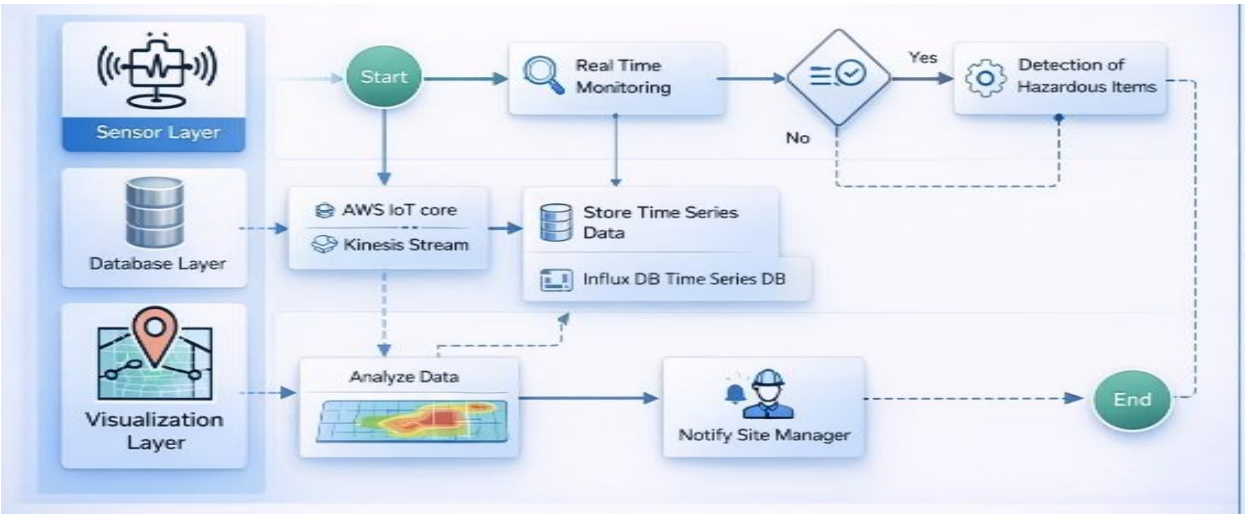


Figure 1:System overview

Sensor layer: IoT devices with GNSS-enabled telematics continuously stream high-frequency positional and kinematic data (location, movement) from heavy machinery. This data enables precise, real-time mapping of equipment trajectories, providing the technical foundation for dynamic heat maps and proactive hazard assessments. These sensors provide latitude–longitude coordinates at approximately 1 Hz, with a typical horizontal accuracy of 0.5-1.5 m under open-sky conditions. Such accuracy is consistent with prior construction safety studies that employ GNSS-based equipment tracking and is sufficient to detect spatial overlaps and proximity risks during earthwork operations.

By incorporating engine RPM and load, the system reliably classifies machine activity, as these variables strongly correlate with operational status (Talmaki & Kamat, 2022). The system distinguishes among idle, down, and active working time, enabling granular monitoring and a more realistic representation of operational risk, both of which are essential for safety management.

Database layer: The database layer uses a hybrid InfluxDB–AWS architecture for scalable, real-time safety. InfluxDB efficiently handles high-frequency data ingestion (via REST APIs) and time-based queries for hazard detection. AWS services (IoT Core, Kinesis, S3) ensure scalability, secure connection management, real-time ingestion, and long-term storage. This setup delivers reliable, low-latency insights crucial for proactive safety visualization management. Machine-installed telematics securely stream GNSS, motion, and engine data via APIs.

Visualization layer: The visualization layer delivers a dynamic spatial and temporal view of machine movements across a specific geofenced earthwork site. This area is defined by GNSS coordinates in the WGS84 reference frame: (55.719581, 9.128638) → (55.719998, 9.126529) → (55.722117, 9.127383) → (55.721882, 9.129044) → (55.722466, 9.129268) → (55.722419, 9.129801) → (55.719581, 9.128638), forming a closed polygon that marks the operational boundary of the jobsite. Machine telemetry is processed as continuous time-series data and refreshed at fixed intervals to maintain current situational awareness. Operational states are classified into time windows based on proximity events and hazard interactions, then mapped to a heatmap layer using color-coded risk indicators: blue (Idle), yellow (Down), and red (Working). Static hazard zones are defined through structured user input, with each hazard characterized by its spatial location, influence radius, and severity classification. These hazards are modeled as persistent spatial entities that serve as fixed reference points against which the movements of live machines are continuously evaluated for proximity violations.

Machine-level safety indicators are derived from live telemetry streams updated at a fixed refresh rate. For each machine, a continuous safety evaluation is performed, capturing the amount of unsafe work it performs at any given time. Unsafe work is identified when the real-time distance between machines, or between a machine and a static hazard zone, falls below predefined safety thresholds.

Zone Safety Status is derived by continuously evaluating unsafe activity within each spatial zone. For a given zone, safety status indicates the level of safety at any time, based on the machines operating within its boundaries and their ongoing proximity violations. A zone is considered increasingly unsafe when multiple machines, or the same machine repeatedly, perform unsafe work—defined by distance-based threshold violations with other machines or static hazard zones—while operating in that zone. The zone safety status is updated in real time and visualized as a zone-level indicator, enabling supervisors to rapidly identify areas where unsafe work is accumulating due to congestion, overlapping tasks, or hazardous spatial configurations. Both machine-level and zone-level safety states are translated into a standardized 5 × 5 risk assessment matrix to support consistent interpretation and prioritization of safety conditions. In this mapping, likelihood is derived from the extent and persistence of unsafe work, as reflected in the machine or Zone Safety

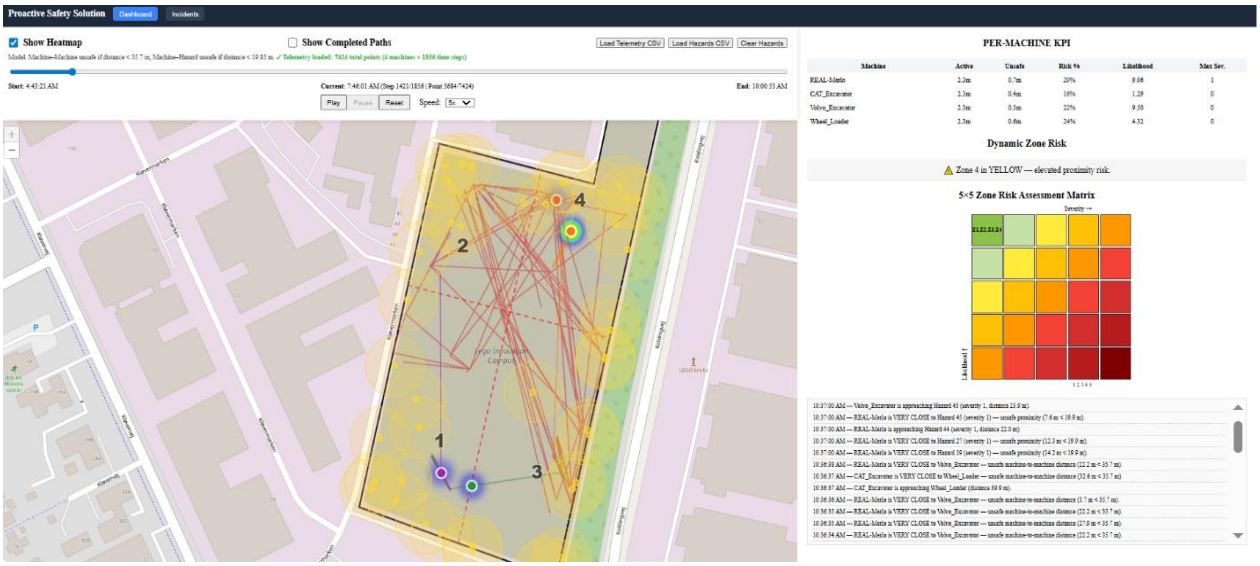


Figure 2: Visualization layer

KPI, capturing how frequently and for how long distance-based safety thresholds are violated. Severity is determined by the classification of the interacting hazards or machines involved in those violations, reflecting the potential consequences of the unsafe condition. By combining these two dimensions, each machine and zone is assigned to a corresponding risk cell that represents its current safety state. The matrix is continuously updated as KPIs evolve and rendered in the visualization interface to provide a clear, comparable view of escalating or diminishing risk across the site. Additionally, live commentary is provided to enhance the system's decision-making robustness.

The visualization layer is implemented as a fully client-side JavaScript dashboard using Vue 3 for reactive state management and Leaflet for heat plugin, for geospatial rendering. Vue components maintain live state for machine telemetry, static hazard definitions, safety KPIs, and risk states, while Leaflet renders the geofence polygon, 2 × 2 zone partitions, machine markers, trajectories, static hazard zones, and a continuously updated risk heat map. The heat map visualizes the spatial accumulation of unsafe activity by projecting machine- and zone-level safety KPIs onto the map surface, thereby highlighting areas of elevated risk. Telemetry and hazard data are ingested via browser-based CSV parsing and replayed at a fixed refresh rate using a JavaScript timing loop. Distance-based proximity evaluation and safety status updates are computed in JavaScript and propagated through Vue's reactivity system, triggering synchronized updates of the heat map, map layers, machine and zone KPIs, and the 5 × 5 risk assessment matrix. This fully browser-based architecture enables continuous, low-latency visualization of evolving safety conditions without server-side processing.

Step #3: System implementation:

The system was implemented and evaluated on a live construction site in Billund, Denmark, covering approximately 19,000 m² and involving intensive earthwork activities, including excavation, grading, and site preparation. For this pilot study, four machines were instrumented and monitored simultaneously: a telescopic handler (13013-MerloP40.17), two excavators (Volvo and CAT), and a wheel loader (Figure 2). These machines are the primary sources of dynamic interaction and collision risk on earthwork sites, owing to overlapping workspaces and frequent material-handling operations. The system was deployed over a three-week monitoring period, providing a realistic operational context for continuous telemetry acquisition and for evaluating real-time hazard detection, zone-level risk accumulation, and heat-map-based visualization within a space-constrained, dynamically evolving site. In addition to mobile equipment, predefined static hazardous items (Table 1) were incorporated prior to system execution and remained fixed throughout the monitoring period, serving as continuous reference points for distance-based safety evaluation.

Table 2: Static hazardous items in jobsite with severity

Group Name	Total Items	Severity Level (Out of 5)
Access & Road	15	3
Perimeter & Boundary	11	2
Storage & Stockyard	9	4
Cranes, Panels & Heavy Machinery	8	5
Facilities	2	1

The system captured live location, movement, and engine performance data, enabling accurate machine tracking. This data was integrated with a digital jobsite map (including obstacles/boundaries) to create dynamic heat maps. These maps moved beyond simple movement tracking to identify actual collision risks, significantly enhancing safety awareness (see Figure 3, which highlights high-risk zones).

Results and discussion

During the three-week monitoring period, the machines continuously streamed high-frequency GNSS data (location, movement, status). This positioning information was integrated with the site's digital boundaries. The data's consistency with planned routes and physical constraints confirmed its reliability for proximity-based safety analysis, even without ultra-high precision. This study focuses on equipment-centric risk monitoring and does not capture worker–equipment interactions, which remain a primary source of fatal incidents; therefore, the proposed framework should be interpreted as a partial safety solution rather than a comprehensive systemThe system successfully tracked equipment movement patterns across 504 operational hours, generating comprehensive telemetry records that captured real-time positioning and operational status. Analysis of this data revealed distinct patterns of spatial clustering and temporal risk concentration. Machine activity was visible in the heatmap, including down periods (equipment powered off during breaks and end-of-shift), idle states (engine running but stationary during excavation or loading work), and active movement between work locations. This operational state classification, derived by combining GNSS positioning with RPM and load-sensor data, proved essential for accurate risk assessment.

Dynamic heatmaps clearly visualized equipment activity and identified spatiotemporal safety hotspots near excavation fronts, loading zones, access routes, heavy machinery areas, and excavation perimeters. These high intensity clusters showed elevated collision risk caused by dense machine movement and repeated machine to hazard proximity violations, with extreme cases recording minimum distances below 1 m. The analysis also revealed that risks increased during peak operational periods when multiple machines worked in confined spaces, allowing the framework to distinguish hazardous zones from safer staging areas and provide a live, transient safety assessment beyond static safety methods

Additionally, the dynamic heat map visualization provides situational awareness superior to fixed zones or manual checks. By continuously displaying shifting risk patterns and congestion clusters, it enables safety managers to visually identify emerging hazards and facilitate timely, targeted interventions (e.g., rerouting). This shifts safety management from post hoc analysis to real-time, proactive spatial intelligence. The system's capacity to update visualizations in near-real-time (refresh intervals under 30 seconds) allowed supervisors to observe evolving safety conditions and respond before critical thresholds were breached.

The proposed system is designed to support safety managers or designated personnel in monitoring and managing safety risks in earthwork operations in real time. When the dashboard detects unsafe proximity between machines or hazards, risks are immediately visualized through heat maps, KPI dashboards, risk matrices, and live commentary, enabling rapid identification and communication to operators and nearby workers via tools such as walkie-talkies, radios, or mobile devices, or through public announcement systems when necessary. If no corrective action is observed, the system supports escalation by continuously monitoring the condition, triggering repeated alerts, and enabling stronger interventions such as issuing urgent warnings, halting machinery, or clearing workers from risk zones. In critical situations involving high-severity hazards, operations may be paused entirely until safe conditions are restored. By combining real-time detection, communication, and escalation mechanisms, the system facilitates a structured response process that enhances situational awareness, supports timely intervention, and enables continuous improvement of safety practices through the analysis of recurring risk patterns

This study presents the methodology for an integrated IoT safety monitoring system based on commercial telematics and cloud infrastructure. Our findings indicate that meter-level positional consistency is adequate for proactive safety decision-making and identifying hazardous zones, effectively overcoming the limitation of lacking quantitative accuracy validation. The system's reliance on industry-standard GPS accuracy (typically 2–5-meter precision) proved sufficient for safety zone monitoring, confirming that sub-meter precision is not a prerequisite for effective proximity-based risk assessment in open earthwork environments. Future work is necessary to integrate predictive analytics and model complex multi-equipment interactions for broader applicability.

The analysis revealed that unsafe interactions were not uniformly distributed but concentrated in specific spatial zones and time periods. Temporal risk analysis revealed high-risk operational periods, while spatial clustering identified recurring hotspots near excavation fronts and in high-traffic areas. Additionally, machine-specific risk profiles indicated that certain equipment types exhibited higher frequencies of proximity violations, highlighting opportunities for targeted safety interventions. While no historical baseline or incident logs were available for direct comparison, the system detected multiple proximity violations and near-miss patterns over 504 operational hours, demonstrating its ability to uncover latent safety risks that are typically not observable with traditional monitoring approaches. These results support the system's role in enabling proactive safety management by providing real-time, data-driven insights into hazardous interactions. Future work will focus on validating these indicators through longitudinal studies and integrating incident records and near-miss reporting systems to establish quantitative links between detected risk patterns and actual safety outcomes.

Limitations and conclusions

The proof-of-concept system was tested for only three weeks at a single site, limiting generalizability. Machine telematics lacked survey-based validation, so it supports relative risk identification rather than precise localization. GPS positioning also introduces accuracy uncertainty, especially near obstructions, though this was minimal at

the study site. The absence of baseline safety records limits the ability to quantify absolute reductions in incidents, and future work will address this through before–and–after comparative analysis.

Finally, the study was limited to two machine types and omitted critical factors like worker-on-foot interaction, multi-machine dynamics, and environmental/behavioral context. The absence of worker tracking represents a significant limitation, as pedestrian-equipment interactions constitute a major source of construction site accidents. Additionally, the system did not account for visibility conditions, operator experience levels, or communication patterns among crews, all of which influence real-world safety outcomes.

Despite these constraints, the system demonstrates strong technical feasibility and practical value by integrating IoT telemetry with dynamic heat maps for proactive safety management. Continuous risk visualization reveals high-risk spatiotemporal patterns, providing actionable insights beyond traditional static plans. Its reliance on operational context confirms the effectiveness of real-time spatial analytics, while hazard detection and visual risk zones address key safety gaps.

From an implementation perspective, the system uses commercially available telematics and cloud infrastructure, making it accessible without major customization. Its compatibility with existing fleet systems, along with standard data formats and open-source tools, supports easy adoption, scalability, and site-specific customization

The research demonstrates that IoT-enabled proximity monitoring can transition construction safety from reactive incident response to predictive hazard prevention. By providing continuous spatial awareness and enabling early intervention before accidents occur, the system exemplifies how digital construction technologies can fundamentally enhance worker safety. The framework's capacity to identify recurring risk patterns also supports organizational learning, allowing firms to refine operational procedures and training programs based on empirical evidence of actual site behavior rather than theoretical hazard assessments.

To advance this work, future efforts must quantitatively validate positioning accuracy and assess system robustness across various sites. Research should incorporate new data sources, including worker tracking and multi-equipment modeling. Expanding the framework to include wearable sensors for workers on foot would enable comprehensive monitoring of all site personnel, thereby addressing the current limitation of equipment-only tracking. Integration of computer vision systems could complement GPS-based positioning, particularly in confined spaces or in areas with degraded GPS signal. Crucially, the integration of machine learning for predictive risk forecasting will be key to transitioning construction safety management from reactive to a predictive practice. Advanced analytics could leverage historical proximity patterns to forecast collision probabilities from current operational configurations, enabling preemptive interventions when risk thresholds are projected to be exceeded.

Long-term deployment across multiple sites can establish industry benchmarks and validate risk patterns. Integrating BIM and construction schedules can contextualize proximity data, enabling a scalable, data-driven approach to improving construction safety and preventing accidents.

References

Arya, V., Colace, F., Lombardi, M., Lorusso, A., Santaniello, D., Valentino, C., 2025. DTSafe-O. Int. J. Semantic Web Inf. Syst. 21, 1–36. https://doi.org/10.4018/ijswis.395999

Asadzadeh, A., Arashpour, M., Li, H., Ngo, T., Bab-Hadiashar, A., Rashidi, A., 2020. Sensor-based safety management. Autom. Constr. 113, 103128. https://doi.org/10.1016/j.autcon.2020.103128

Bęś, P., Strzałkowski, P., Górniak-Zimroz, J., Szóstak, M., Janiszewski, M., 2025. Innovative Technologies to Improve Occupational Safety in Mining and Construction Industries—Part I. Sensors 25, 5201. https://doi.org/10.3390/s25165201

Catalán, J., Santos-Olmo, A., Rosado, D.G., Sánchez, L., Fernández-Medina, E., 2025. Cybersecurity Risk Analysis and Management in IoT Environments: The MARISMA-IoT Pattern. https://doi.org/10.21203/rs.3.rs-7491380/v1

Cheng, T., Teizer, J., 2013. Real-time resource location data collection and visualization technology for construction safety and activity monitoring applications. Autom. Constr. 34, 3–15. https://doi.org/10.1016/j.autcon.2012.10.017

Chung, W.W.S., Tariq, S., Mohandes, S.R., Zayed, T., 2023. IoT-based application for construction site safety monitoring. Int. J. Constr. Manag. 23, 58–74. https://doi.org/10.1080/15623599.2020.1847405

Golovina, O., Teizer, J., Pradhananga, N., 2016. Heat map generation for predictive safety planning: Preventing struck-by and near miss interactions between workers-on-foot and construction equipment. Autom. Constr. 71, 99–115. https://doi.org/10.1016/j.autcon.2016.03.008

Häikiö, J., Kallio, J., Mäkelä, S.-M., Keränen, J., 2020. IoT-based safety monitoring from the perspective of construction site workers. Int. J. Occup. Environ. Saf. 4, 1–14. https://doi.org/10.24840/2184-0954_004.001_0001

Hong, K., Teizer, J., 2024. Automated Construction Site Safety Monitoring Using Preidentified Static and Dynamic Hazard Zones. Presented at the 41st International Symposium on Automation and Robotics in Construction, Lille, France. https://doi.org/10.22260/ISARC2024/0051

Jo, B.-W., Lee, Y.-S., Khan, R.M.A., Kim, J.-H., Kim, D.-K., 2019. Robust Construction Safety System (RCSS) for collision accidents prevention on construction sites. Sens. Switz. 19. https://doi.org/10.3390/s19040932

Kayhani, N., Taghaddos, H., Behzadipour, S., 2018. Construction Equipment Collision-Free Path Planning Using Robotic Approach. https://doi.org/10.22260/isarc2018/0169

Khajavi, S.H., Tetik, M., Liu, Z., Korhonen, P., Holmström, J., 2023. Digital Twin for Safety and Security: Perspectives on Building Lifecycle. Ieee Access 11, 52339–52356. https://doi.org/10.1109/access.2023.3278267

Kim, D., Xiong, S., 2025. Enhancing Worker Safety: Real-Time Automated Detection of Personal Protective Equipment to Prevent Falls from Heights at Construction Sites Using Improved YOLOv8 and Edge Devices. J. Constr. Eng. Manag. 151. https://doi.org/10.1061/JCEMD4.COENG-14985

Kim, S., Hong, S.H., Kim, H., Lee, M., Hwang, S., 2023. Small object detection (SOD) system for comprehensive construction site safety monitoring. Autom. Constr. 156. https://doi.org/10.1016/j.autcon.2023.105103

Kukoyi, P.O., Adebowale, O.J., 2021. Impediments to Construction Safety Improvement. J. Eng. Proj. Prod. Manag. https://doi.org/10.2478/jeppm-2021-0020

Leonardo, M., Berardo, N., Carbonari, A., Luigi, R., Giuda, G.M.D., 2020. Development of a Digital Twin Model for Real-Time Assessment of Collision Hazards 14–19. https://doi.org/10.3311/ccc2020-003

Liu, C., Su, X.-L., Li, C., 2021. Edge Computing for Data Anomaly Detection of Multi-Sensors in Underground Mining. Electronics 10, 302. https://doi.org/10.3390/electronics10030302

Moe, S.J.S., Kim, B.W., Khan, A.N., Xu, R., Tuan, N.A., Kim, K., Kim, D.H., 2023. Collaborative Worker Safety Prediction Mechanism Using Federated Learning Assisted Edge Intelligence in Outdoor Construction Environment. Ieee Access 11, 109010–109026. https://doi.org/10.1109/access.2023.3320716

Niu, Y., Lu, W., Xue, F., Liu, D., Chen, K., Fang, D., Anumba, C., 2019. Towards the "third wave": An SCO-enabled occupational health and safety management system for construction. Saf. Sci. 111, 213–223. https://doi.org/10.1016/j.ssci.2018.07.013

Poudel, O., Assaad, R.H., 2025. A Real-Time Intelligent Acoustic IoT-Enabled Embedded Construction Site Monitoring and Alert System: Integrating Deep Learning–Based Machine-Listening Algorithms, Edge Computing, and Cloud Computing. J. Constr. Eng. Manag. 151, 04025075. https://doi.org/10.1061/JCEMD4.COENG-15938

Rao, A.S., Radanovic, M., Liu, Y., Hu, S., Fang, Y., Khoshelham, K., Palaniswami, M., Ngo, T., 2022. Real-time monitoring of construction sites: Sensors, methods, and applications. Autom. Constr. 136, 104099. https://doi.org/10.1016/j.autcon.2021.104099

Shen, X., Marks, E., Pradhananga, N., Cheng, T., 2016. Hazardous Proximity Zone Design for Heavy Construction Excavation Equipment. J. Constr. Eng. Manag. 142, 05016001. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001108

Soltanmohammadlou, N., Sadeghi, S., Hon, C.K.H., Mokhtarpour-Khanghah, F., 2019. Real-time locating systems and safety in construction sites: A literature review. Saf. Sci. 117, 229–242. https://doi.org/10.1016/j.ssci.2019.04.025

Son, H., Kim, C., 2021. Integrated worker detection and tracking for the safe operation of construction machinery. Autom. Constr. 126, 103670. https://doi.org/10.1016/j.autcon.2021.103670

Talmaki, S.A., Kamat, V.R., 2022. Sensor Acquisition and Allocation for Real-Time Monitoring of Articulated Construction Equipment in Digital Twins. Sensors 22, 7635. https://doi.org/10.3390/s22197635

Teizer, J., 2015. Magnetic Field Proximity Detection and Alert Technology for Safe Heavy Construction Equipment Operation. Presented at the 32nd International Symposium on Automation and Robotics in Construction, Oulu, Finland. https://doi.org/10.22260/ISARC2015/0062

Walker, R., Smith, S., Bosché, F., 2021. Enabling operational autonomy in earth-moving with real-time 3D environment modelling. Presented at the 38th International Symposium on Automation and Robotics in Construction, Dubai, UAE. https://doi.org/10.22260/ISARC2021/0022

Zamani, V., Banihashemi, S.Y., Abbasi, A., 2020. How can communication networks among excavator crew members in construction projects affect the relationship between safety climate and safety outcomes? Saf. Sci. 128. https://doi.org/10.1016/j.ssci.2020.104737

Zhang, M., Cao, T., Zhao, X., 2017. Applying Sensor-Based Technology to Improve Construction Safety Management. Sensors 17, 1841. https://doi.org/10.3390/s17081841

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

PERCEIVED BENEFITS OF 5D BIM IN CONSTRUCTION COST MANAGEMENT: A EUROPE-WIDE SURVEY

Enes Gönen and Semra Çomu
Boğaziçi University, Istanbul, Türkiye

Abstract

Examining contextual factors shaping the success of 5D BIM implementation in Europe, results reveal that while 91.5% utilize the technology for QTO and deep adoption remains limited, satisfaction increases significantly when a critical value threshold is exceeded, moving from single-use tasks to integrated workflows. Contractors are less satisfied by 5D BIM due to commercial risk, while medium-sized organizations and moderately complex projects exhibit a transition trap, resulting in lower satisfaction. Conversely, high national BIM maturity and Integrated Project Delivery amplify success. 5D BIM provides optimal value only when implemented as a comprehensive system, not merely as a standalone tool.

Introduction

The construction industry, despite its massive contribution to the global economy, has long been recognized as one of the least digitized sectors, characterized by stagnant productivity and a slow pace of technological adoption (Agarwal et al., 2016; Naji et al., 2024). However, the sector is currently undergoing a significant shift towards digital transformation to overcome these historic inefficiencies and modernize project lifecycles (Naji et al., 2024). Building Information Modeling (BIM), specifically its fifth dimension (5D BIM), offers a paradigm shift by integrating cost-related data with 3D models to improve efficiency, accuracy, and decision-making. The uses (features or functionalities) of 5D BIM include automated quantity takeoffs, cost estimation linked to model, cost planning and forecasting, change management, value engineering, cash flow analysis, cost control, progress payment validation, lifecycle cost analysis and claims management, delivering substantial advancements over traditional methods (Aibinu and Venkatesh, 2014; Charef et al., 2018; Kehily and Underwood, 2017; Lu et al., 2016; Stanley and Thurnell, 2014; Vigneault et al., 2020).

Despite these capabilities, the industry's perceptions of 5D BIM are divided. Enthusiasts praise it as a revolutionary tool that simplifies cost management, increases accuracy, and promotes collaboration, while skeptics argue that traditional tools are sufficient for similar tasks without the need for a BIM model, claiming that 5D BIM is overhyped. This divide highlights the need for robust empirical evidence to validate the benefits of 5D BIM and clarify its value proposition.

Existing research on 5D BIM has largely focused on technical case studies addressing the practical benefits of 5D BIM or developing frameworks and tools to support its adoption, with little research on which contextual factors shape practical outcomes. A critical gap persists in the literature: empirical studies investigating how specific factors influence the benefits of 5D BIM are scarce.

The motivation for this study stems from the lack of comprehensive research examining how project characteristics, user profiles, and features used affect the perceived success of 5D BIM in various construction projects. By addressing this gap, the study aims to provide actionable insights to optimize 5D BIM implementation. Through a survey of European construction professionals, it aims to provide evidence-based recommendations for developing 5D BIM adoption strategies.

Methodology

Data collection

A questionnaire survey was designed to collect data on respondent information, project characteristics, 5D BIM implementation, and perceived benefits of 5D BIM in construction projects. The survey targeted construction professionals with experience in projects where BIM was utilized in cost management processes across the European regions as defined by the Publications Office of the European Union (2026), excluding Russia and the Caucasus countries. This geographical scope was selected due to the widespread adoption of 5D BIM within the region and the researcher's ability to leverage an established professional network, which facilitated access to relevant and high-quality contacts.

The survey was structured into four sections. The first section, "Respondent Information", gathered demographic and professional data. Respondents were asked to specify their age, gender, total years of experience in the construction industry, BIM and 5D BIM, their current role/title and stakeholder category (e.g., client/owner, contractor, designer, consultant) with respect to both their current position and the reference project.

The second section, "Project Characteristics", collected specific details regarding the reference project. Respondents were asked to identify the project type (commercial, residential, metro, etc.), budget, duration, location, delivery method (design-bid-build, design-build, IPD, etc.), contract type (lump-sum, unit price, etc.), project's com-

plexity, BIM maturity level of the project country, and the BIM expertise level of the project team.

The third section, "5D BIM Implementation", examined the practical application and workflow of BIM in cost management processes. Respondents were asked to describe their workflow processes through an open-ended question, and to select the 5D BIM uses employed in the reference project and the software tools utilized for implementation.

The final section, "Benefits of 5D BIM in Cost Management", assessed the extent to which respondents agreed that 12 different expected benefits were observed in their reference projects using a 5-point Likert scale (ranging from 1-Strongly Disagree to 5-Strongly Agree). Respondents were asked to select "Not Relevant" if a specific outcome was not a targeted project goal. Additionally, an open-ended question was included to invite participants to share specific examples or anecdotes, adding qualitative depth to the analysis.

The survey was hosted on Google Forms, with settings configured to collect email addresses to ensure non-anonymous responses, allowing verification of respondents' identities and relevance.

A targeted survey was conducted with construction professionals experienced in 5D BIM within Europe. A multi-channel sampling approach, primarily utilizing targeted LinkedIn searches (e.g., "5D BIM cost") and professional referrals, built an initial pool of 236 potential participants. Candidates were screened against strict eligibility criteria requiring the practical use of at least one 5D BIM functionality (detailed in Figure 1) on a construction project in Europe. Excluding unreachable or ineligible contacts yielded 163 qualified invitees. Eligible professionals were invited via multiple communication channels, with information provided on the study scope and confidentiality. In total, 59 valid responses were collected between April and May 2025, corresponding to a response rate of 36.2% among qualified invitees. While this network-driven sampling ensured relevance, it inherently introduces a potential selection bias regarding professionals inactive on such platforms.

Data analysis

Survey data were initially screened and preprocessed using Microsoft Excel. Categorical responses were reviewed, consolidated where necessary, and infrequent categories were grouped under a general "Other" label. Project location data were aggregated into regional groupings based on EuroVoc classifications. For analysis, multiple-selection items were converted into binary variables, and two summary indicators were derived to capture the extent of 5D BIM adoption: the number of different 5D BIM functionalities used and the number of distinct software tools reported by each respondent.

Following preprocessing, the dataset was transferred to the Python programming environment for statistical analysis and visualization, utilizing the Pandas library for data manipulation and Matplotlib for generating figures.

Firstly, the extent of utilization was analyzed, where the frequency of use for each 5D BIM feature was calculated. The features were ranked based on the number of respondents utilizing them, and the results were visualized using a horizontal bar chart with a "progress-bar" style to clearly display the adoption rates relative to the total sample size.

Secondly, the perceived benefits were analyzed, where the industry's perception of 5D BIM benefits was assessed based on the 12 Likert-scale items. A diverging stacked bar chart was generated to illustrate the distribution of sentiments, separating positive (agree/strongly agree) and negative (disagree/strongly disagree) responses centered around the neutral option. Responses marked as "Not relevant" were excluded from the percentage calculations to ensure the analysis accurately reflected the views of those with direct experience.

Finally, to evaluate how different contextual factors impact the perceived success of 5D BIM, a composite metric termed "Overall Perceived Benefits" (OPB) was defined. This score was calculated for each respondent by averaging the numerical values of the 12 benefit items (scale 1–5), excluding "Not relevant" responses from the calculation. The distribution of OPB scores across various groups was analyzed using variable-width box plots, which display the median, quartiles, and range of the data. In these plots, the width of each box is proportional to the sample size of the respective subgroup, allowing for a simultaneous assessment of distribution and frequency. Given the total sample size of 59 and the resulting smaller subgroup populations, inferential statistical testing would lack the adequate statistical power to reliably detect effect sizes. Therefore, this descriptive visual approach was chosen over calculating precise confidence intervals. For visualization purposes, nominal categorical variables were sorted by median OPB score in descending order, while ordinal variables were sorted by their natural descending order. Even in cases where no significant relationship was observed, variables were presented to utilize the box widths representing frequency of the group as a descriptive tool, thereby providing essential context regarding the profile of the respondents and the characteristics of their referenced projects.

Open-ended responses regarding specific project experiences, challenges, and benefits were subjected to thematic analysis to identify recurring narratives. Relevant qualitative evidence was interwoven directly into the discussion of quantitative results, enabling recurring themes to serve as a corroborative context that offers explanatory depth to the statistical observations.

Results and discussion

A total of 59 valid responses were collected from professionals who utilized BIM for cost management in their reference projects located in 19 European countries. Figure 1 illustrates the adoption rates of various 5D BIM functionalities. The data reveals a striking disparity between the most common use case and all other functionalities. An overwhelming majority of respondents (91.5%) utilize

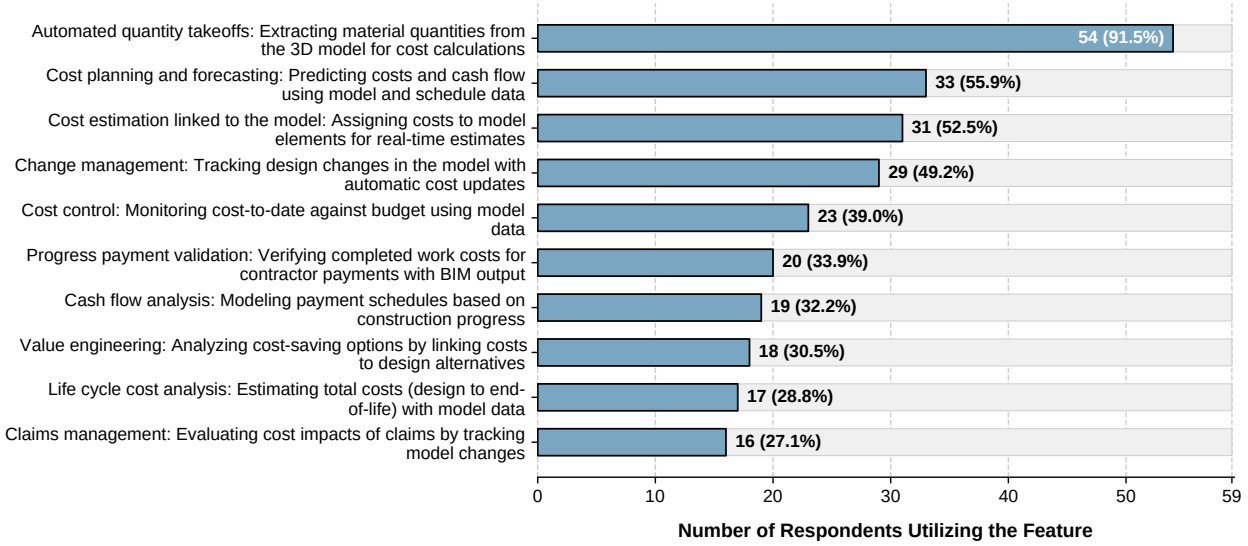


Figure 1: Utilization frequencies of 5D BIM features in reference projects (n=59)

5D BIM for ”Automated quantity takeoffs” making it the foundational, and for many, the only application of the technology.

In sharp contrast, the second most cited use drops significantly to 55.9%, indicating that many professionals equate 5D BIM solely with geometry-based quantity extraction rather than integrating cost data directly into the model environment or leveraging it for further financial analysis. This limited scope aligns with established literature; Mitchell (2012) observed that 5D is frequently reduced to the ”simple one-liner of Quantity Take Off” while Moses et al. (2020) confirmed that industry efforts have predominantly focused on using technologies for quantity extraction rather than fully integrated cost management. This exposes a prevalent definitional ambiguity: although Charef et al. (2018) identify a theoretical consensus defining 5D BIM as the ”cost” dimension, practical application reveals a stark divergence in what the cost dimension actually entails.

This quantitative observation is corroborated and critiqued by the qualitative findings from the open-ended survey questions. Several respondents argued that automated quantity extraction is merely a prerequisite but not the definition of 5D BIM. One respondent explicitly stated that ”true 5D BIM begins only after going beyond the quantity takeoff” implying that those who stop at quantity takeoff cannot claim to be fully utilizing 5D BIM. This dichotomy suggests that the industry is currently operating with a ”two-tier” definition of 5D BIM: a broad, entry-level tier focused on quantity takeoff (QTO), and a narrower, mature tier focused on comprehensive cost management.

Beyond automated quantity takeoffs, adoption rates cluster at moderate to low levels. Approximately half of the respondents reported using 5D BIM for ”Cost planning and forecasting” (55.9%), ”Cost estimation linked to the model” (52.5%), and ”Change management” (49.2%), indicating partial integration beyond geometry-based counting. All remaining functions including ”Cost control” (39.0%),

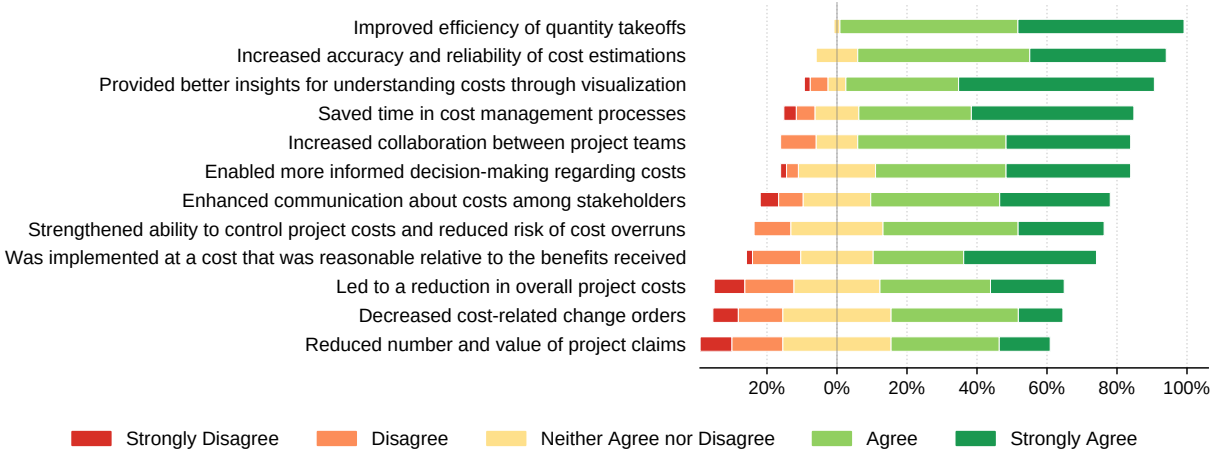


Figure 2: Perceived benefits of 5D BIM implementation (sorted by level of agreement)

”Progress payment validation” (33.9%), ”Cash flow analysis” (32.2%), and ”Value engineering” (30.5%) and ”Lifecycle cost analysis” (28.8%) and ”Claims management” (27.1%) were reported by fewer than 40% of respondents, reflecting that these specialized downstream tasks are currently the least common implementations of 5D BIM.

Turning from the breadth of adoption to the perceived impact, Figure 2 illustrates a predominantly positive reception of 5D BIM, based on the survey responses. Across all twelve queried benefits, the proportion of respondents expressing agreement (combining ”Agree” and ”Strongly Agree”) consistently outweighs those expressing disagreement. This consensus suggests that, despite the variations in adoption maturity discussed earlier, the technology delivers tangible value to those who implement it.

The most significant gains are perceived in the foundational tasks of cost management. ”Improved efficiency of quantity takeoffs” emerged as the top-rated benefit, followed closely by ”Increased accuracy and reliability of cost estimations.” This statistical finding is strongly corroborated by the qualitative feedback; respondents repeatedly emphasized that linking the Bill of Quantities (BOQ) directly to the 3D model creates an ”objective and consistent dataset” thereby accelerating the process and significantly reducing the risk of human error associated with manual calculations. This aligns with recent findings by Mohd Saupi and Ismail (2024), who reported that improved cost estimation remains a primary driver for adoption, with professionals emphasizing 5D BIM’s capacity to produce more precise and reliable financial projections compared to traditional methods.

Complementing these foundational gains, the literature emphasizes that 5D BIM extends its value into the dynamic phases of project delivery by enhancing decision-making and collaboration. Eldin et al. (2024) highlight that the technology facilitates real-time visualization and analysis of cost data, which allows for continuous financial forecasting and proactive risk mitigation rather than mere reactive corrections. Furthermore, Pishdad and Onungwa (2024) argue that integrating cost data into the model fosters a more transparent collaborative environment, effectively shifting cost management from isolated, convoluted spreadsheets to a shared, visual 3D format where assumptions can be easily challenged and verified by the entire project team.

To evaluate the impact of contextual factors on 5D BIM success, Overall Perceived Benefit (OPB) scores were analyzed using variable-width box plots. These figures simultaneously illustrate the statistical distribution of scores along the horizontal axis and the relative sample size of each subgroup through the vertical width of the boxes.

Figure 3 displays the distribution of OPB scores categorized by experience levels and Figure 4 by respondent age. The average age of the sample is approximately 35 years old, with 12 years of experience in construction, 8 years in BIM, and 6 years in 5D BIM, calculated based on the weighted mean of range midpoints. Both figures show stable medians and overlapping ranges across all groups,

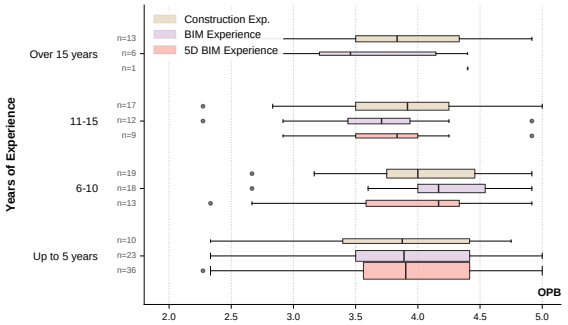


Figure 3: Distribution of OPB by experience in construction, BIM and 5D BIM

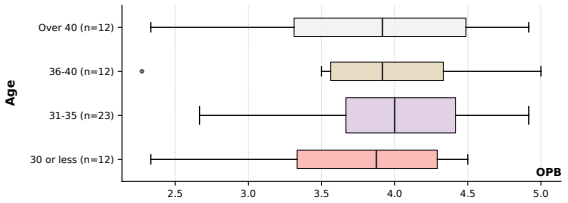


Figure 4: Distribution of OPB by respondent age

indicating that the positive perception of 5D BIM benefits remains consistent regardless of the length of professional experience and age.

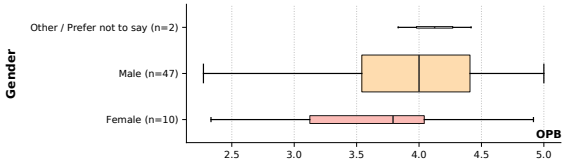


Figure 5: Distribution of OPB by gender

Figure 5 presents OPB scores by gender, reflecting a sample distribution with a larger male cohort (n=47) compared to the female group (n=10). While the male subgroup exhibits a slightly higher median, the notable difference in group sizes suggests that this observation should be interpreted within the context of the sample’s demographic composition, rather than as a definitive indicator of gender-based divergence in 5D BIM perception.

Figure 6 characterizes the respondents’ professional profiles at the time of the survey. The sample is notably dominated by ”BIM & Digital Construction” group (n=23), reflecting that 5D BIM is still largely a specialized discipline driven by domain experts rather than general project management staff. While smaller groups like ”Project Managers & Site Engineers” (n=10) and ”Executive Leadership & Founders” (n=7) suggest higher satisfaction, the BIM specialist group exhibits a wider range of OPB scores.

Figure 7 reveals a ”benefit asymmetry” across stakeholder roles. Although Contractors constitute the largest group in the reference project (n=22), they report a lower median OPB score compared to other groups. Qualitative findings further suggest that this lower perception among Contractors may be linked to commercial sensitivities rather than technical implementation alone. The high level of cost transparency required by 5D BIM may conflict with com-

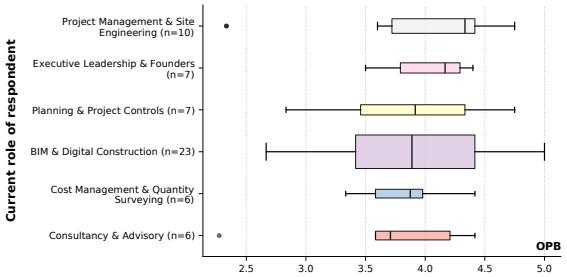


Figure 6: Distribution of OPB by current role of the respondent

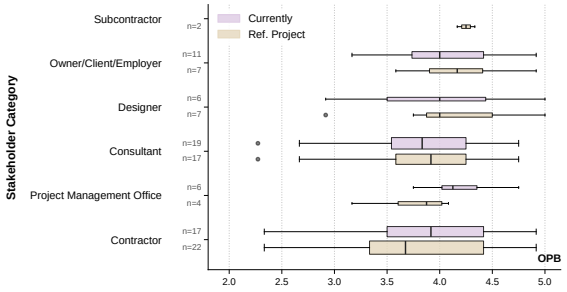


Figure 7: Distribution of OPB by current stakeholder category of the respondent and the stakeholder category in the reference project

petitive contracting environments where profit margins are guarded, and the requirement to share granular cost data within a shared model can be viewed as a commercial risk rather than a benefit, potentially tempering the enthusiasm for 5D adoption among contractor organizations. This stance is substantiated by Guo et al. (2019), who demonstrated that BIM-enabled transparency can directly impact contractors' interests by exposing proprietary resource allocations and cost structures that previously served as hidden sources of profit.

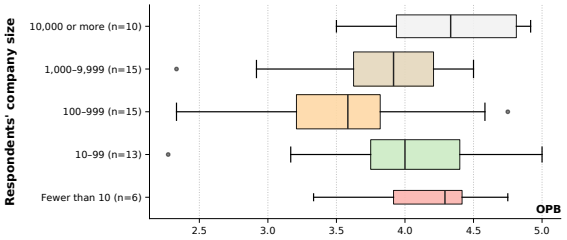


Figure 8: Distribution of OPB by the number of employees of the company that the respondent was/is working in the reference project

Figure 8 and Figure 9 illustrate a striking non-monotonic, U-shaped relationship across both company size and project complexity, where perceived benefits peak at the extremes and dip significantly in the intermediate categories. In terms of company size, the high satisfaction among small firms suggests organizational agility, while the success of large corporations (over 10,000 employees) likely stems from the resource abundance required to enforce standardization and specialized training, aligning with a resource-based view of firm capabilities.

Conversely, mid-sized firms (100–999 employees) appear caught in a "transition trap" facing the interoperability challenges of scaling without the institutional capacity of larger

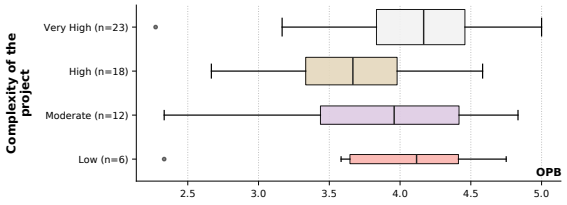


Figure 9: Distribution of OPB by project complexity

entities. This aligns with Bosch-Sijtsema et al. (2017), who argue that medium-sized contractors often face a specific disadvantage; they are compelled to compete with larger firms on complex projects yet lack the comparable in-house R&D resources and financial buffer to overcome the high investment costs and knowledge barriers of BIM implementation. A parallel U-shaped pattern in project complexity reinforces this observation, suggesting that while 5D BIM offers quick wins for simple projects and becomes indispensable for high-complexity ones, intermediate projects occupy a "zone of friction" where the implementation effort is felt most acutely relative to the rewards.

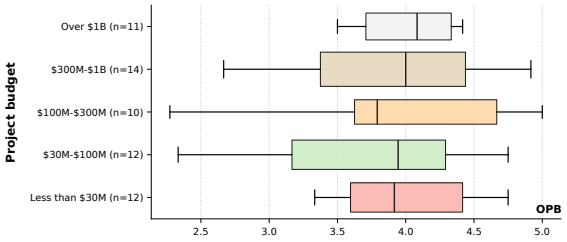


Figure 10: Distribution of OPB by project budget

Figure 10 displays the relationship between project budget and OPB scores. Within the studied sample, no significant trend was observed; the median OPB scores appeared consistent across all financial tiers.

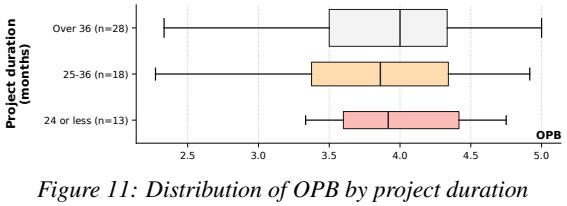


Figure 11: Distribution of OPB by project duration

Figure 11 presents the analysis regarding project duration. Due to data skewness, where nearly half the respondents (n=28) concentrate in the "Over 36 months" category, no meaningful relationship was observed.

Figure 12 demonstrates a clear positive correlation between the national BIM maturity level and the perceived benefits of 5D BIM. Unlike the variable project characteristics discussed earlier, the national environment shows a linear progression: as the maturity level rises from "Emerging or Low" (n=14) to "High" (n=8), the median OPB score steadily increases. This empirical finding supports the macro-micro relationship established by Prabhakaran et al. (2021), who argue that a mature national ecosystem characterized by established mandates and standardized protocols significantly amplifies organizational implementation suc-

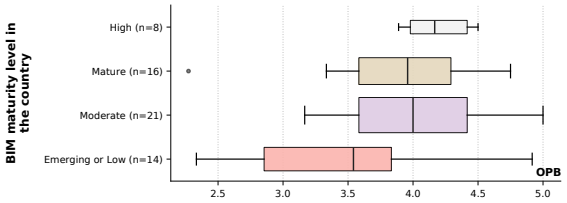


Figure 12: Distribution of OPB by BIM maturity level in the project country

cess by reducing external friction, reflecting the impact of institutional logics on technology adoption.

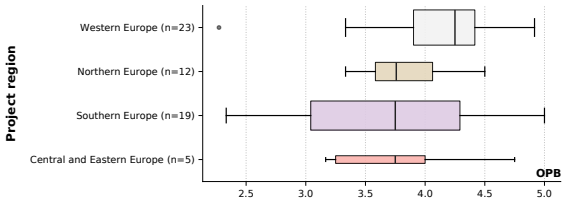


Figure 13: Distribution of OPB by project region

Figure 13 corroborates this finding through a geographical lens, highlighting Western Europe as the region with the highest perceived benefits. This group, which constitutes the largest segment (n=23) and includes respondents from the United Kingdom (14), Germany (5), Luxembourg (2), the Netherlands (1), and Austria (1), exhibits the highest median score. This superior performance can be attributed to the mature BIM adoption landscape in these countries, particularly the UK and Germany, where rigorous standards and widespread engineering familiarity create a conducive environment for 5D integration. In contrast, the remaining regions report comparatively lower median scores, likely reflecting varying degrees of market maturity. These include Southern Europe (n=19), dominated by respondents from Turkiye (16) alongside Portugal (2) and Italy (1); Northern Europe (n=12), comprising Lithuania (4), Denmark (2), Finland (2), Norway (2), Latvia (1), and Sweden (1); and Central and Eastern Europe (n=5), consisting of respondents from Hungary (1), Kosovo (1), Romania (1), Serbia (1), and Slovenia (1).

Qualitative insights from these regions suggest that these lower outcomes are driven by systemic barriers rather than individual lack of skill. Respondents specifically pointed to the absence of established national cost databases and a specialized BIM education as primary obstacles. This observation is empirically supported by the case of Türkiye, the dominant constituent of the Southern Europe group, where the lack of official government mandates and standardized cost data has been identified as a critical barrier, keeping adoption at a ”nascent” stage compared to the mature, policy-driven ecosystem of the UK (Sari and Pekeriçli, 2020; Omotayo et al., 2025). This validates the correlation between a supportive national infrastructure and the successful realization of 5D benefits.

Figure 14 examines the influence of the project team's BIM expertise on OPB scores. A visual inspection does not indicate a distinct, linear trend across all categories

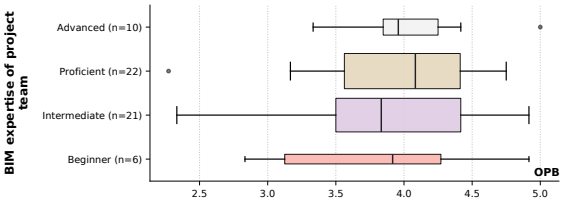


Figure 14: Distribution of OPB by BIM expertise of the project team

(from ”Beginner” to ”Advanced”). However, a closer look at the two largest subgroups, ”Intermediate” (n=21) and ”Proficient” (n=22), reveals a noticeable positive shift; the ”Proficient” group exhibits a higher median score compared to the ”Intermediate” cohort. This observation suggests that while the overall relationship remains complex across the spectrum, the transition from an intermediate to a proficient level of collective expertise serves as a discernible factor in realizing higher perceived benefits.

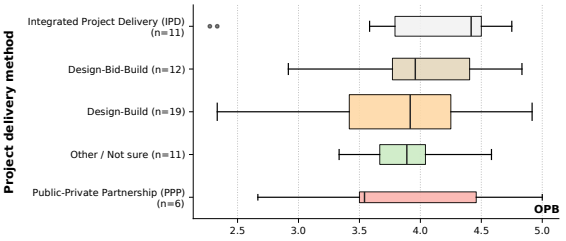


Figure 15: Distribution of OPB by project delivery method

Figure 15 displays the relationship between project delivery methods and OPB scores. The Integrated Project Delivery (IPD) group (n=11) demonstrates the highest median score with a relatively compact distribution. This aligns with the expectation that the collaborative structure of IPD creates a favorable environment for 5D BIM implementation compared to traditional methods such as Design-Bid-Build, Design-Build, Public-Private Partnership (PPP). This pattern is consistent with the broader literature, which identifies IPD as the optimal delivery method for BIM; its multi-party risk-sharing structure and reward mechanisms align commercial interests, helping to mitigate the financial transparency risks feared by contractors, thereby fostering early contractor involvement and seamless data integration, whereas traditional methods often fragment the very information flows that 5D BIM relies upon (Piroozfar et al., 2019; Karasu et al., 2022).

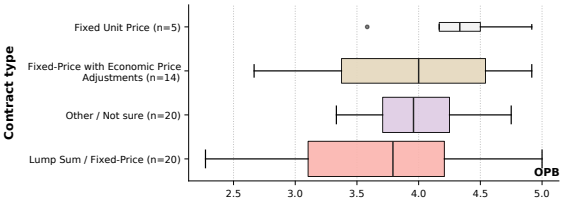


Figure 16: Distribution of OPB by contract type

Figure 16 examines OPB scores across contract types, revealing that despite minor variations between categories such as 'Fixed Unit Price' and 'Lump Sum / Fixed-Price,' the overall distribution patterns remain largely overlapping

with similar medians; consequently, no distinct relationship could be established between the specific pricing mechanism and the perceived benefits of 5D BIM.

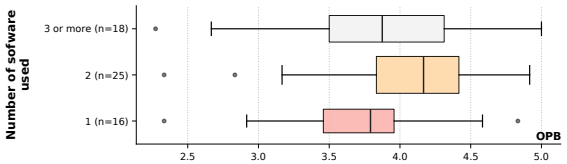


Figure 17: Distribution of OPB by the number of software programs each respondent mentioned using for 5D BIM

Figure 17 correlates the diversity of the software toolkit with OPB scores, revealing that the highest and most consistent satisfaction is found in the ”2 software” category (n=25). This suggests an optimal workflow often pairs a primary design tool, most notably Autodesk Revit (mentioned by 40 respondents), with a specialized coordination or estimation platform. In contrast, the ”1 software” group (n=16) reports a notably lower median score, likely reflecting the limitations of performing 5D tasks solely within a design authoring tool, while the ”3 or more” group (n=18) shows diminished returns, potentially due to complexity. Following Revit, the most frequently cited tools were Navisworks (29), RIB CostX (11), Synchro (10), Bexel Manager (8), Tekla (4), RIB iTWO (4), and Solibri (3), with 24 other programs mentioned by one or two respondents.

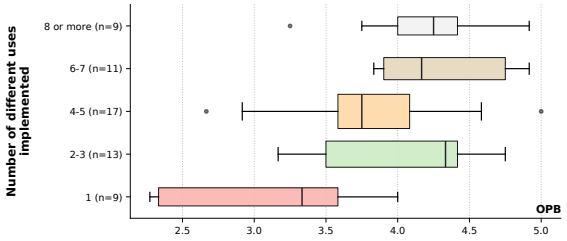


Figure 18: Distribution of OPB by the number of 5D BIM uses implemented

Figure 18 stands as the empirical validation of the ”True 5D BIM” concept, offering perhaps the most critical insight of the entire study. It establishes a robust positive correlation between the depth of implementation and perceived success. The group implementing only ”1 use” (n=9) reports the lowest median satisfaction, reinforcing the earlier observation from Figure 1 that limiting 5D BIM to a single function (predominantly quantity take-off) fails to unlock the technology’s potential.

However, the data reveals a dramatic ”value threshold” immediately after this initial stage. The jump to the ”2-3 uses” category (n=13) results in a sharp increase in the median OPB score, surpassing even the intermediate groups. This indicates that moving beyond isolation, by integrating just one or two additional functions (e.g., coupling QTO with Cost Estimation or Progress Monitoring), creates a synergistic effect that instantly amplifies value. This aligns with the findings of Disney et al. (2022) and Sun et al. (2024), who demonstrate that the interdependence of multiple BIM uses creates a ”systemic” benefit where the whole is greater than the sum of its parts, leading to significantly higher user satisfaction and project performance compared to fragmented tools. The trend continues to favor holistic application, with the ”8 or more” group (n=9) maintaining high satisfaction. This confirms that 5D BIM is not merely a tool for a specific task but a systemic methodology; the more comprehensively its diverse functions are utilized, the greater the return on investment.

Conclusions

This paper examined the contextual factors influencing the perceived success of 5D BIM in the European construction sector. By analyzing responses from 59 professionals across 19 countries, the research bridges the gap between theoretical potential and practical reality. The findings reveal a critical dichotomy: while most practitioners equate 5D BIM with simple quantity take-offs (QTO), the data confirms that this limited approach yields the lowest satisfaction. Instead, a distinct ”value threshold” exists where integrating cost estimation or progress monitoring via a dual-platform workflow significantly amplifies success, validating 5D BIM as a systemic, integrated methodology rather than a singular task.

Contextual analysis highlights that success follows a nonlinear, U-shaped pattern; mid-sized firms and intermediate projects often face a ”transition trap” lacking both the agility of small firms and the institutional resources of large corporations. Furthermore, a ”benefit asymmetry” emerges among stakeholders; contractors report lower satisfaction levels, likely due to the commercial risks associated with 5D BIM’s high transparency, which can conflict with competitive contracting norms by exposing proprietary cost structures. Conversely, the results confirm that a supportive ecosystem, characterized by high national BIM maturity and collaborative delivery methods like Integrated Project Delivery (IPD), is a fundamental catalyst for realizing the full potential of 5D BIM.

These results should be considered in light of several limitations. Most notably, the sample size, constrained by the specialized nature of the domain, reduces the statistical power for subgroup analyses, as small frequencies within certain subcategories preclude the calculation of robust effect sizes and result in wide confidence intervals. Furthermore, the analysis relies on perceived benefits as a surrogate for objective performance, a measure that indicates technological acceptance but does not necessarily equate to quantifiable financial profitability or hard project metrics. Additionally, the geographic distribution of the sample introduces potential geographic bias. Specifically, the heavy dominance of Turkish respondents in the Southern Europe category, while providing valuable context regarding emerging market barriers, may limit the generalizability of regional comparisons and mask country-specific differences, particularly regarding applicability to non-European regulatory contexts. Future studies should address these limitations by applying weighted benefit criteria and complementing perception-based surveys with quantitative case

studies that incorporate actual cost variance data and schedule performance metrics. This approach would validate the perception-behavior link and strengthen causal inference, while also expanding the assessment of return on investment (ROI) to encompass the operational stage of the facility's life cycle.

References

Agarwal, R., Chandrasekaran, S., and Sridhar, M. (2016). Imagining construction's digital future. McKinsey & Company. [Online] Available from: https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future [Accessed: 22 January 2026].

Aibinu, A. and Venkatesh, S. (2014). Status of BIM adoption and the BIM experience of cost consultants in australia. Journal of Professional Issues in Engineering Education and Practice, 140(3).

Bosch-Sijtsema, P., Isaksson, A., Lennartsson, M., and Linderoth, H. C. J. (2017). Barriers and facilitators for BIM use among Swedish medium-sized contractors - "we wait until someone tells us to use it". Visualization in Engineering, 5:1–12.

Charef, R., Alaka, H., and Emmitt, S. (2018). Beyond the third dimension of BIM: A systematic review of literature and assessment of professional views. Journal of Building Engineering, 19:242–257.

Disney, O., Roupé, M., Johansson, M., and Leto, A. (2022). Embracing BIM in its totality: a total BIM case study. Smart and Sustainable Built Environment, 13(3):512–531.

Eldin, A. M. S., Abdelalim, A. M., and Tantawy, M. (2024). Enhancing cost management in construction: The role of 5D building information modeling (BIM). Engineering Research Journal, 183(3):C164–C188.

Guo, H., Yu, R., and Fang, Y. (2019). Analysis of negative impacts of BIM-enabled information transparency on contractors' interests. Automation in Construction, 103:87–98.

Karasu, T., Aaltonen, K., and Haapasalo, H. (2022). The interplay of IPD and BIM: a systematic literature review. Construction Innovation, 22(3).

Kehily, D. and Underwood, J. (2017). Embedding life cycle costing in 5D BIM. Journal of Information Technology in Construction (ITcon), 22:145–167.

Lu, Q., Won, J., and Cheng, J. C. P. (2016). A financial decision making framework for construction projects based on 5D building information modeling (BIM). International Journal of Project Management, 34(1):3–21.

Mitchell, D. (2012). 5D BIM: Creating cost certainty and better buildings. In RICS COBRA 2012, Las Vegas, Nevada USA. RICS.

Mohd Saupi, N. D. H. and Ismail, N. A. A. (2024). The implementation of 5D building information modelling (BIM) in construction management: Benefits and challenges faced by construction professionals. Built Environment Journal, 21(SI):94–104.

Moses, T., Heesom, D., and Oloke, D. (2020). Implementing 5D BIM on construction projects: Contractor perspectives from the UK construction sector. Journal of Engineering, Design and Technology, 18(6):1867–1888.

Naji, K. K., Gunduz, M., Alhenzab, F. H., Al-Hababi, H., and Al-Qahtani, A. H. (2024). A systematic review of the digital transformation of the building construction industry. IEEE Access, 12:31461–31487.

Omotayo, T., Tanyer, A. M., Deng, J., Kaima, A., Akponeware, A., Pekericli, M. K., Shikder, S., and Ogunnusi, M. (2025). Advancing AI-powered BIM for circularity in construction in the UK and Turkiye: State-of-the-art review and capability maturity modelling. Buildings, 15(8):1224.

Piroozfar, P., Farr, E. R. P., Zadeh, A. M., Inacio, S., Kilgallon, S., and Jin, R. (2019). Facilitating building information modelling (BIM) using integrated project delivery (IPD): A UK perspective. Journal of Building Engineering, 26:100907.

Pishdad, P. and Onungwa, I. O. (2024). Analysis of 5D BIM for cost estimation, cost control and payments. Journal of Information Technology in Construction (ITcon), 29:525–548.

Prabhakaran, A., Mahamadu, A.-M., Mahdjoubi, L., Andrić, J. F., Manu, P., and Mzyece, D. (2021). An investigation into macro BIM maturity and its impacts: a comparison of Qatar and the United Kingdom. Architectural Engineering and Design Management, 17:496–515.

Publications Office of the European Union (2026). EuroVoc: Geography, Europe (concept 7206). [Online] Available from: https://eur-lex.europa.eu/browse/eurovoc.html?params=72%2C7206 [Accessed: 22 January 2026].

Sari, R. and Pekeriçli, M. K. (2020). An investigation of comparison and evaluation of official BIM documents released in the USA, UK and Turkey. Journal of Construction Engineering, Management & Innovation, 3(2):67–84.

Stanley, R. and Thurnell, D. (2014). The benefits of, and barriers to, implementation of 5D BIM for quantity surveying in New Zealand. Australasian Journal of Construction Economics and Building, 14(1):105–117.

Sun, H., Khoo, T. J., Esa, M., Mahdiyar, A., and Li, J. (2024). Critical factors driving construction project performance in integrated 5D building information modeling. Buildings, 14(9):2807.

Vigneault, M. A., Boton, C., Chong, H. Y., and Cooper-Cooke, B. (2020). An innovative framework of 5D BIM solutions for construction cost management: A systematic review. Archives of Computational Methods in Engineering, 27:1013–1030.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DATA-DRIVEN DECISION SUPPORT FOR CONSTRUCTION METHOD SELECTION IN LINEAR INFRASTRUCTURE PROJECTS: A SYSTEMATIC REVIEW

Maik Görgner[1,2], Simon Hoeng[1], Marc Schmailzl[1], Thomas Linner[1], Thomas Wolff[1], Mathias Obergrießer[1]
[1]Ostbayerische Technische Hochschule Regensburg, Seybothstraße 2, 93053 Regensburg
[2]Firmengruppe Max Bögl, Max-Bögl-Straße 1, 92369 Sengenthal

Abstract

The selection of construction methods is a critical factor in linear infrastructure projects, yet current practice remains largely experience-driven. As is evident from the extant literature, multi-criteria decision-making, rule-based and algorithmic approaches are applied to support method selection. The present systematic literature review analyses these approaches with respect to input parameters, output structures, and validation strategies. The findings indicate a predominance of project-level ranking models and limited spatial resolution. Key deficiencies identified include inadequate scalability, deficient integration of costs and restricted applicability to corridor-scale decision-making. The review identifies these gaps as central challenges for future data-driven decision-support systems.

Introduction

The accelerated expansion of energy transmission infrastructure is a key prerequisite for achieving climate and energy policy goals in many countries. In Germany, too, the expansion of transmission grids poses a particularly complex planning challenge. This complexity along underground transmission routes arises from heterogeneous environmental conditions, dense existing infrastructure and highly variable geotechnical conditions along the routes (Ahmels & Bruns, 2016; TenneT TSO GmbH, 2025). In particular, soil conditions, groundwater conditions, intersection types, access restrictions or environmental restrictions can change significantly over short distances. Consequently, the construction methods selected in response to these conditions along the corridor significantly influence costs and time, environmental impacts and technical risks. Inadequate choices can increase costs, prolong construction periods and intensify environmental-related constraints (Neuling, 2013). In this context, the selection of suitable construction methods, such as open-cut excavation or horizontal directional drilling (HDD), plays a decisive role in project execution. Linear infrastructure projects are typically realised along extended corridors with spatially varying boundary conditions, where construction-related constraints may change over short distances. As a result, construction method selection is not a single project-level decision but a sequence of locally constrained choices made along the corridor. Figure 1 illustrates such a construction context by showing an open-cut excavation along a linear infrastructure corridor. Furthermore it highlights the spatial setting in which construction method selection translates into locally constrained execution decisions along the route.



Figure 1: Open-cut construction method along a linear infrastructure corridor (source: Max Bögl Group)

Despite their relevance, the selection of construction methods for linear infrastructure projects is still predominantly based on expert judgement, experience-based heuristics, and manually prepared feasibility studies. While such approaches benefit from domain expertise, they are difficult to standardise, reproduce, or scale across long corridors with heterogeneous boundary conditions. In contrast, a data-driven construction method selection approach could have the potential to support spatially resolved, segment-oriented decision-making. Construction methods would be assessed locally along the corridor, enabling more transparent and reproducible decision-making.

Against this backdrop, digital planning tools are increasingly being explored to operationalise such spatially resolved decision-making in infrastructure projects. Therefore research and practice have made significant progress in the digitalisation of infrastructure planning. Geographic information systems (GIS) have become a core technology for the integration and analysis of spatial data in corridor and route planning, enabling the systematic consideration of environmental constraints, topography and existing infrastructure (Schito & Grassi, 2015).

GIS-based approaches are widely used for route optimisation, spatial conflict analysis, and corridor evaluation, often in combination with resistance-based models or multi-criteria decision analysis (MCDA) (Malczewski, 2006; Steuer & Donaubauer, 2012). Within the context of digital construction, these developments form part of a broader convergence of geospatial modelling and simulation-based planning and increasingly shape research in construction informatics (Linner, 2024). They primarily address the question of *where* infrastructure should be located, while higher levels of automated or algorithmic decision support for construction-related decisions remain limited in practice (Pan M. et al., 2018).

However, once a route has been determined due to approval decisions, land use restrictions or agreements with stakeholders, the focus of planning shifts from route selection to construction. At this stage, the key question is *how* the infrastructure should be built along heterogeneous corridor segments. While tunnel construction and microtunnelling research has extensively addressed method and machine selection under geological uncertainties, the systematic transfer of such decision-making concepts to corridor-based linear infrastructure remains limited (Pan M. et al., 2020; Borrmann, 2024). Existing GIS-based planning tools and digital planning environments predominantly support data management, visualisation, and documentation of route alignments, while providing limited support for the automated derivation of construction methods or execution variants (Pan M. et al., 2020; Borrmann, 2024). Similarly, BIM and simulation-based approaches increasingly enable detailed modelling of construction processes, but they often lack integration with large-scale geodata sets or the segment-by-segment decision logic required for long linear projects (Borrmann A. et al., 2015; Borrmann, 2024; Arashpour, M. et al., 2025). Consequently, planning and execution data remain fragmented, and decisions about construction methods are rarely supported by transparent, reproducible, and data-driven models that work along entire corridors.

Recent research highlights the growing potential of integrated, model-based decision support systems that link geospatial data, geotechnical information and execution-related parameters within a unified analytical framework (König, M. et al., 2021; Pache, Scharun, & Hochmuth, 2024). At the same time, several authors emphasise the ongoing challenges associated with data heterogeneity, expert dependency and limited interoperability between planning and execution models (Schito & Grassi, 2015; Arashpour, M. et al., 2025). These limitations are particularly evident in the selection of construction methods, which remains one of the least standardised and most experience-driven steps in linear infrastructure planning. Against this backdrop, a systematic understanding of existing data- and model-based approaches to the selection of construction methods is needed.

On the one hand, individual studies propose decision support models based on multi-criteria evaluation, rule-based logic or hybrid algorithmic approaches. On the other hand, there is still a lack of a consolidated overview of these methods, their input specifications and output structures, and their validation strategies.

This study addresses this need by conducting a systematic literature review and deriving a structured taxonomy of decision-support approaches for construction method selection in linear infrastructure projects.

Building on the identified gaps, ongoing research aims to develop GIS-based, segment-oriented decision-support workflows that translate corridor-scale constraints into transparent method assignments and link them to cost and schedule evaluation.

Research Objective and Questions

The objective of this systematic literature review is to identify data- and model-based approaches for selecting construction methods in linear infrastructure projects. These approaches will then be classified and summarised. The review focuses on methodological contributions that support the selection of construction methods under heterogeneous segment conditions. Beyond identifying families of methods, the overview aims to systematically consolidate the types of input parameters used for decision-making. Moreover, the different forms in which decision results are provided in the existing approaches are consolidated. In the final step, validation strategies and frequently reported limitations are to be evaluated.

To achieve this goal, three specific research questions were formulated.

RQ1: Which data- and model-based approaches for the selection of construction methods in linear infrastructure projects are described in the scientific literature?

RQ2: Which input parameters (e.g. geotechnical, geometric, ecological or economic factors) and which output structures (e.g. rankings, suitability assessments, cost or time indicators or method assignments) are used in these approaches?

RQ3: How are the proposed decision models validated and what methodological limitations are discussed in literature?

Methodology – Systematic Review Design

In order to address the research objective and the three research questions defined in Chapter 2, a systematic literature review was conducted. The methodological design enables a structured and transparent analysis of data- and model-based approaches to construction method selection in linear infrastructure projects. It further ensures the reproducibility of both the synthesis and the underlying selection process. The review follows established guidelines for systematic reviews in civil engineering research and is aligned with the PRISMA framework (Page, M. et al., 2021).

Review Protocol and Analytical Framework

The methodological workflow of the review is divided into five consecutive phases: definition, search, selection, analysis and synthesis. Figure 2 provides an overview of the analytical workflow used in this study.

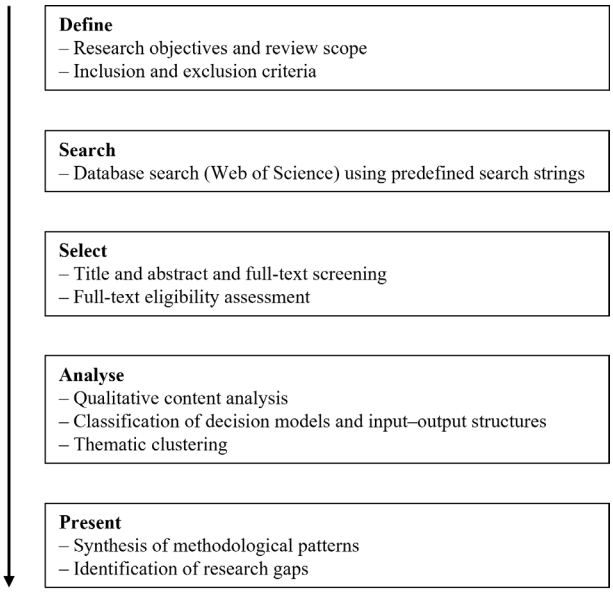


Figure 2: Methodological workflow of the systematic literature review

Literature Search Strategy

The literature search followed a transparent and reproducible procedure in accordance with the PRISMA framework and comprised literature identification, scope restriction, full-text eligibility assessment, and qualitative synthesis. The search was performed using the Web of Science Core Collection and limited to peer-reviewed journal articles and conference proceedings published in English between 2016 and January 2026.

Web of Science was selected as the primary database due to its structured indexing and broad coverage of engineering and construction-related outlets, supporting a consistent and reproducible review process.

Grey literature, such as technical reports or non-peer-reviewed project documentation, was not systematically included.

A predefined search strategy was applied on 12 January 2026 using three thematic keyword groups addressing (i) construction method selection, (ii) decision-support approaches, and (iii) linear infrastructure contexts. Keywords were combined using Boolean operators (OR within groups; AND between groups).

The search query was conducted using a topic search (TS), covering title, abstract, and author keyword fields. All search terms are summarised in Table 1. On top, subject-category filters were defined to restrict the results to relevant research fields. The selected categories included Engineering Civil, Construction & Building Technology, Materials Science Multidisciplinary, Physics Applied, Environmental Sciences, Computer Science Information Systems, Engineering Multidisciplinary and Geosciences Multidisciplinary. These categories were deliberately defined beyond the core field of civil engineering to ensure a broader coverage of related research domains and to capture potentially relevant methodological approaches from adjacent disciplines. Additional records were assigned to further categories but were not explicitly considered as primary filters.

After applying these filters, 177 publications remained from the initial 284 records for further screening.

Table1: Keyword groups and associated search terms used for the literature search

Keyword group	*Associated keywords*
Construction method selection	*construction method; construction technique; method selection; installation method; trenchless; horizontal directional drilling; HDD; tunnell; TBM selection*
Decision-support approaches	*decision support; select; multi-criteria; multicriteria; MCDA; MCDM; AHP; TOPSIS; ELECTRE; PROMETHEE; fuzzy; rule-based*
Linear infrastrucure context	*linear infrastructure; corridor; route; alignment; pipeline; cable; utility; underground infrastructure; transmission line*

This was followed by a two-stage screening process. First, titles and abstracts were reviewed for thematic relevance. Records that did not relate to civil engineering or construction or that did not focus on construction methods or processes were excluded at this stage. The full-text articles were then reviewed for suitability in terms of their methodological relevance. Studies were excluded in the full-text phase if they did not contain an explicit selection or comparison of construction methods, did not have a formal decision-making or evaluation model, or were purely descriptive case studies without transferable methodological contributions. The entire screening and selection process is summarised in the PRISMA flow diagram shown in Figure 3.

Data Extraction and Synthesis Approach

A structured data extraction scheme was used for the final selection of the included studies. Each publication was analysed with regard to (i) the type of decision support approach, (ii) the input parameters considered, (iii) the structure of the decision results, and (iv) the validation strategy applied. The extracted information was summarised using qualitative content analysis and thematic clustering. This process enabled the classification of decision support methods into methodological families, the identification of frequently used parameter categories, and the evaluation of validation practices and methodological limitations.

Results – Classification and Synthesis of Approaches

This section synthesises the results of the systematic review in accordance with the three research questions. The included studies were analysed with respect to their underlying decision-support approaches (RQ1), considered input parameters and output structures (RQ2), and applied validation strategies and methodological limitations (RQ3).
Table 2 provides an overview of the studies included in the qualitative synthesis.

Types of Decision-Support Approaches (RQ1)

The reviewed literature reveals three dominant families of decision-support approaches for construction method selection in linear infrastructure projects: (i) multi-criteria decision-making (MCDA) methods, (ii) rule-based and knowledge-based models, and (iii) hybrid or algorithmic decision approaches (Aschilean & Giurca, 2018; Matthews J. et al., 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Parka, A. et al., 2020; Alinizzi, M. et al., 2020; Lin, S.S. et al., 2023; Leu, S.S. et al., 2025).

In terms of digital decision support, the identified method families can be interpreted as distinct digital decision-support paradigms, ranging from expert-driven rule-based knowledge models to algorithmic and hybrid approaches with increasing degrees of automation (Pan M. et al., 2018; Pan M. et al., 2020).

MCDA-based approaches constitute the largest group of identified methods. These studies typically employ established techniques such as the Analytic Hierarchy Process (AHP), the Analytic Network Process (ANP) or fuzzy extensions to rank alternative construction methods based on weighted evaluation criteria (Aschilean & Giurca, 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Alinizzi, M. et al., 2020; Lin, S.S. et al., 2023). MCDA approaches are predominantly applied at the project level and aim to support early-stage decision-making under multiple, often conflicting criteria.
While these methods provide transparent and structured rankings, they generally rely on predefined weights and are sensitive to expert judgement (Aschilean & Giurca, 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Alinizzi, M. et al., 2020; Lin, S.S. et al., 2023).

Rule-based or expert-driven approaches formalise expert knowledge in the form of deterministic rules or decision trees (Matthews J. et al., 2018; Parka, A. et al., 2020). These models are commonly applied in more detailed planning or design stages and focus on technical feasibility and constraint satisfaction (Marinos V. et al., 2018; Parka, A. et al., 2020). Although such approaches enable consistent decision logic, their transferability is often limited by context-specific rule sets (Marinos V. et al., 2018; Parka, A. et al., 2020).

Hybrid and algorithmic approaches combine rule-based feasibility checks with optimisation or simulation techniques (Matthews J. et al., 2018; Leu, S.S. et al., 2025). These models are typically applied at a segment based decision level and explicitly account for spatial variability along linear infrastructure corridors (Matthews J. et al., 2018; Leu, S.S. et al., 2025).

Compared to purely MCDA-based methods, hybrid approaches allow the assignment of different construction methods to individual segments and thus offer a higher degree of operational realism (Matthews J. et al., 2018; Leu, S.S. et al., 2025).

Input Parameters and Output Structures (RQ2)

Across the reviewed studies, input parameters can be grouped into four main categories: geological and geotechnical conditions, geometric characteristics, economic factors, and environmental or social aspects (Aschilean & Giurca, 2018; Matthews J. et al., 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Alinizzi, M. et al., 2020; Ma, P. et al., 2023; Lin, S.S. et al., 2023; Leu, S.S. et al., 2025).

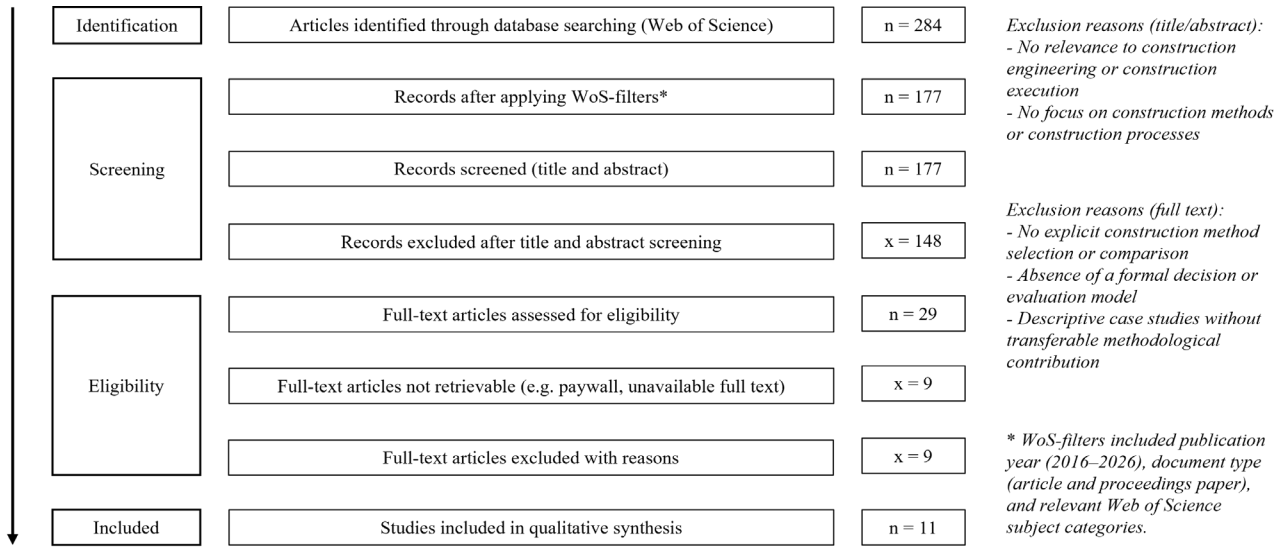


Figure 3: PRISMA flow diagram illustrating the literature search, screening, and study selection process

Geological and geotechnical parameters, such as soil type, groundwater conditions, and geomechanical properties, are considered in the majority of technically oriented models (Marinos V. et al., 2018; Matthews J. et al., 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Ma, P. et al., 2023; Leu, S.S. et al., 2025).

Geometric inputs typically include pipeline diameter, installation depth, alignment length, and crossing conditions (Aschilean & Giurca, 2018; Matthews J. et al., 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Ma, P. et al., 2023; Lin, S.S. et al., 2023).

Economic parameters, including construction cost and duration, are widely used across all methodological families (Aschilean & Giurca, 2018; Matthews J. et al., 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Parka, A. et al., 2020; Alinizzi, M. et al., 2020; Lin, S.S. et al., 2023; Leu, S.S. et al., 2025). Environmental and social criteria, such as emissions, traffic disruption, or societal impacts, are incorporated less consistently and are mainly addressed in sustainability-oriented MCDA studies (Zwierzchowska & Kuliczkowska, 2019; Alinizzi, M. et al., 2020; Rokstad, M. et al., 2025).

Regarding output structures, most approaches generate ranked lists or suitability scores for alternative construction methods (Aschilean & Giurca, 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Alinizzi, M. et al., 2020; Lin, S.S. et al., 2023). Some studies provide a single recommended method, while others report relative performance indicators or Pareto-optimal solutions (Marinos V. et al., 2018; Parka, A. et al., 2020; Leu, S.S. et al., 2025). Only a limited number of models explicitly assign construction methods at a segment level along a predefined alignment (Matthews J. et al., 2018; Leu, S.S. et al., 2025).

The findings highlight a clear dominance of ranking-based outputs and a lack of models that produce directly implementable, spatially explicit construction method assignments (Aschilean & Giurca, 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Alinizzi, M. et al., 2020; Lin, S.S. et al., 2023; Leu, S.S. et al., 2025).

Validation Strategies and Methodological Limitations (RQ3)

Validation strategies across the reviewed literature predominantly rely on illustrative case studies and scenario-based analyses (Aschilean & Giurca, 2018; Marinos V. et al., 2018; Matthews J. et al., 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Parka, A. et al., 2020; Alinizzi, M. et al., 2020; Ma, P. et al., 2023; Lin, S.S. et al., 2023; Leu, S.S. et al., 2025). Many studies demonstrate their proposed methods using single real-world projects or hypothetical examples (Aschilean & Giurca, 2018; Matthews J. et al., 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Parka, A. et al., 2020; Alinizzi, M. et al., 2020; Lin, S.S. et al., 2023; Leu, S.S. et al., 2025).

Simulation-based validation is applied in algorithmic and risk-oriented approaches, whereas sensitivity analyses are commonly used in MCDA-based models to assess robustness against weighting assumptions (Aschilean & Giurca, 2018; Matthews J. et al., 2018; Kwast-Kotlarek & Hełdak, 2019; Leu, S.S. et al., 2025).

Despite these validation efforts, several methodological limitations are recurrent. A strong dependence on expert input remains a key challenge, particularly in weighting criteria and defining rule sets (Aschilean & Giurca, 2018; Zwierzchowska & Kuliczkowska, 2019; Kwast-Kotlarek & Hełdak, 2019; Parka, A. et al., 2020; Lin, S.S. et al., 2023). Data availability and quality constraints limit the applicability of many models beyond their original case contexts (Marinos V. et al., 2018; Matthews J. et al., 2018; Ma, P. et al., 2023; Leu, S.S. et al., 2025).

Furthermore, most approaches lack integration with scheduling, cost control, or digital planning systems, which restricts their practical adoption in real-world project environments (Aschilean & Giurca, 2018; Matthews J. et al., 2018; Alinizzi, M. et al., 2020; Lin, S.S. et al., 2023; Leu, S.S. et al., 2025).

Table 2: Overview of studies included in the qualitative synthesis

ID	Author(s)	Year	Title (short)	Infrastrucure context
01	Aschilean et al.	2018	Sewer pipe rehabilitation (ANP)	Water pipelines
02	Marinos et al.	2018	TBM selection for microtunnelling	Microtunnelling
04	Matthews et al.	2018	Multi-segment trenchless method selection	Pipelines
05	Zwierzchowska et al.	2019	Fuzzy trenchless technology selection	Pipelines
06	Kwast-Kotlarek et al.	2019	HDD vs. open-cut method selection (AHP)	Gas pipelines
07	Parka et al.	2020	Rule-based pressure pipe lining selection	Water pipelines
09	Alinizzi et al.	2020	Fuzzy-QFD pipeline construction assessment	Large-diameter pipelines
11	Ma et al.	2023	Empirical analysis of pipe jacking practice	Pipe jacking
12	Lin et al.	2023	Optimal construction method evaluation	Underground infrastructure
15	Rokstad et al.	2025	Adoption of trenchless technologies	Water infrastructure
18	Leu et al	2025	Risk-based microtunneling machine selection	Microtunneling

Discussion

Interpretation of Results and Methodological Trends

The results of the systematic review show that research on data- and model-based construction method selection for linear infrastructure projects is methodologically diverse but limited in operational depth. Although numerous decision-support approaches have been proposed, most outputs remain project-level rankings or suitability assessments.

They primarily support the decision-making process in the early stages and rarely reflect the spatial variability of real-world corridors. Multi-criteria decision-making approaches dominate the literature, offering transparent evaluation structures and enabling the integration of technical, economic, and environmental criteria. Their outcomes strongly depend on expert-defined weights and project-specific assumptions.

Rule- and knowledge-based models provide more detailed feasibility assessments but often suffer from limited transferability due to context-specific rule sets. Against this background, a small number of core contributions represent a methodological advancement by addressing construction method selection at a segment or corridor level. Matthews J. et al. (2018) demonstrate how rule-based feasibility constraints combined with optimisation logic enable differentiated method assignment along pipeline corridors. Leu, S.S. et al. (2025) extend this perspective by incorporating risk and robustness considerations through Pareto-optimal solution spaces. In contrast, Alinizzi, M. et al. (2020) and Lin, S.S. et al. (2023) provide structured multi-criteria assessment frameworks that largely remain within ranking-based decision paradigms.

Identified Research Gaps

In spite of these advances, several persistent research gaps emerge from the synthesis.

First, construction method selection is rarely conceptualised as a spatially distributed decision problem along predefined alignments. Most approaches abstract from corridor geometry and do not provide explicit segment-level method assignments.

Second, a gap exists between the availability of detailed geospatial and geotechnical data and their systematic integration into decision-support models. Although such data are frequently cited as inputs, they are seldom combined within modular, rule-based frameworks that enable automated and consistent method selection along entire corridors.

Third, the operational integration of decision models remains limited. Most studies provide rankings or qualitative recommendations without translating them into quantitative indicators for cost, duration, or resource demand, thereby constraining their applicability within scheduling, cost control, or digital planning systems.

When interpreted through an AI and digital construction lens, the reviewed approaches distinguish between knowledge-driven decision-support models, such as rule-based systems, and more data-driven or algorithmic approaches. Whereas knowledge-based models offer transparency and traceability, they exhibit limited scalability and adaptability to changing boundary conditions. Automated and hybrid approaches remain rare and are often constrained by data availability, interoperability challenges, and limited uncertainty handling (Pan M. et al., 2020; Linner, 2024). This highlights construction method selection as a promising yet underexplored application field for AI-enabled decision-support within integrated digital construction workflows.

Although this review focuses on peer-reviewed journal articles and conference proceedings to ensure methodological transparency and reproducibility, practice-oriented knowledge documented in grey literature is not systematically reflected. Technical guidelines, industry reports, and project documentation often contain detailed procedural knowledge on construction method selection.

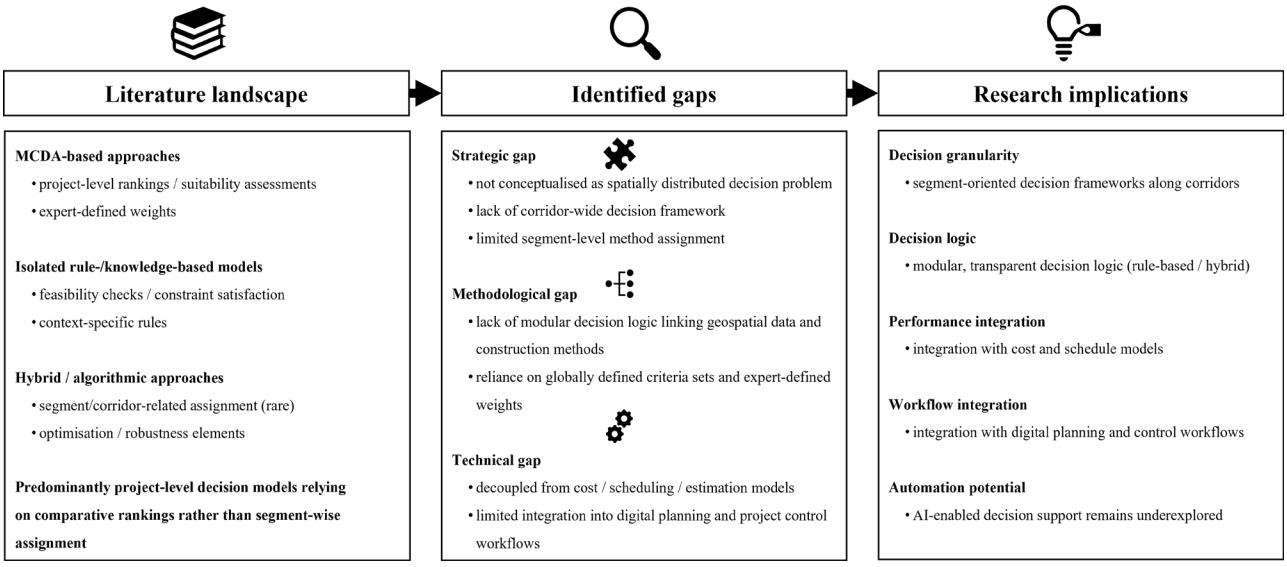


Figure 4: Conceptual synthesis framework derived from the systematic review

However, heterogeneous quality, limited traceability, and a lack of formal validation complicate systematic synthesis. From a scientific perspective, grey literature could be incorporated in complementary research designs, such as structured expert elicitation or qualitative case studies. While such approaches may support the transfer of practice-driven knowledge into reproducible decision-support models, they fall outside the scope of this systematic literature review.

Furthermore, another methodological limitation of this review lies in the restriction of the literature search to a single bibliographic database. While this supports consistency and reproducibility, relevant studies indexed exclusively in other databases (e.g., Scopus or Engineering Village) may not be fully captured. Future reviews could extend the search to additional databases and, where appropriate, complement peer-reviewed sources with structured evidence from practice-oriented outlets.

Methodological Implications for Decision-Support Approaches

Figure 4 summarises the review synthesis by linking the identified families of decision-support approaches, recurring limitations, and resulting methodological implications. Based on the findings of this systematic review, decision-support approaches for corridor-based projects should

(1) treat construction method selection as a segment-oriented decision problem with explicit spatial granularity rules,
(2) operationalise reusable and transparent decision modules linking ground and crossing conditions to feasible methods, and
(3) connect method assignment to cost, schedule, and resource evaluation through consistent data interfaces.

Building on these implications, the author's ongoing research focuses on integrating spatial segmentation, modular decision logic, and quantitative evaluation within a GIS-enabled framework. Accordingly, validation needs to extend beyond single-case demonstrations and include sensitivity analyses and validation against empirical project data to assess model behaviour under varying boundary conditions.
Complementing this, Figure 5 positions the systematic review within an end-to-end decision-support workflow.

Conclusions and Outlook

By categorising existing decision-support approaches and analysing their input parameters, output structures, and validation strategies, the review establishes a transparent reference framework for comparative analysis and future research. The synthesis identifies a central structural limitation in current research: construction method selection is predominantly addressed through project-level or ranking-oriented models and remains weakly integrated with geospatial information and quantitative cost or scheduling perspectives. As a result, decision-support concepts are often fragmented and offer limited applicability for corridor-scale, segment-oriented planning.

Overall, the findings underline that construction method selection represents a critical yet underdeveloped component of digital infrastructure planning. Addressing the identified integration deficit is essential for improving transparency and reproducibility in the planning and execution of linear infrastructure projects. In this context, the review provides a structured foundation for the author's research, which develops a modular hierarchy for construction method selection (e.g., open-cut and trenchless families and module types derived from soil and terrain conditions) and translates corridor constraints into reproducible module-based outputs (e.g., €/m and production rates). Such a framework aims to enable scalable, auditable, and quantitatively informed decisions along the entire alignment and to facilitate integration with downstream cost and schedule evaluation in digital planning workflows.

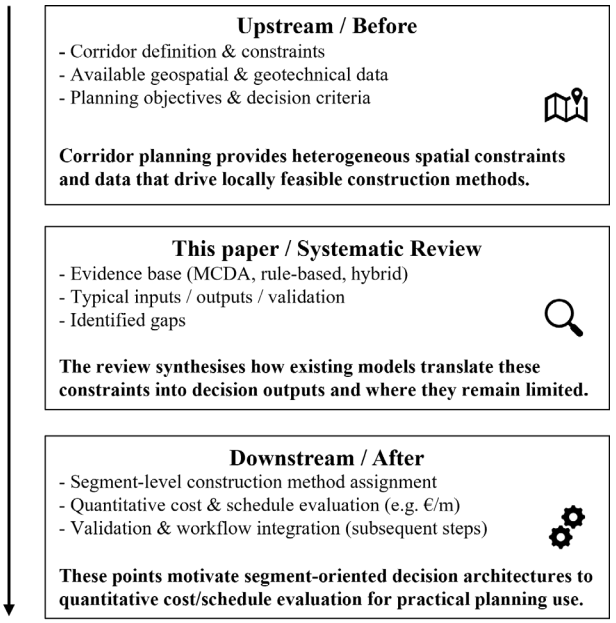


Figure 5: Systematic review positioned in an end-to-end corridor decision-support workflow.

References

Ahmels, P., & Bruns, E. (2016). Auswirkungen verschiedener Erdkabelsysteme auf Natur und Landschaft („EKNA" FKZ 3514 82 1600). Berlin: Deutsche Umwelthilfe.

Alinizzi, M. et al. (2020). Sustainability Assessment of Construction Technologies for Large Pipelines on Urban Highways: Scenario Analysis using Fuzzy QFD. sustainability, 12, 1-20. doi:10.3390/su12072648

Arashpour, M. et al. (Februar 2025). Enhancing infrastructure planning and design through BIM-GIS integration. Structure and Infrastructure Engineering, S. 1-19.

Aschilean, I., & Giurca, I. (15. April 2018). Choosing a Water Distribution Pipe Rehabilitation Solution Using

the Analytical Network Process Method. Water, 10, 1-23. doi:10.3390/w10040484

Borrmann A. et al. (2015). Multi-Scale Geometric-Semantic Modeling of Shield Tunnels for GIS and BIM Applications. Computer-Aided Civil and Infrastructure Engineering, 30, 263-281. doi:10.1111/mice.12090

Borrmann, A. (2024). Building Information Modeling für den Infrastrukturbau. In D. e.V., & R. T. e.V., Leitfaden Geodäsie und BIM - Version 4.0 (S. 42-54). Bühl; München.

GIS Consult GmbH. (2025). Produkte und Lösungen - nisStrom Trasse. (GIS Consult GmbH - Gesellschaft für angewandte geographische Informationssysteme) Abgerufen am 20. Oktober 2025 von https://www.gis-consult.de/produkte-und-loesungen/fachschale-nisstromtrasse/

König, M. et al. (Juni 2021). Prozessorientierte numerische Vortriebsmodellierung – Trassenplanung für den innerstädtischen Tunnelbau. tunnel - Fachzeitschrift für Tunnelbau und Infrastruktur, S. 48-53.

Kwast-Kotlarek, U., & Hełdak, M. (2019). Evaluation of the Construction and Investment Process of a High-Pressure Gas Pipeline with Use of the Trenchless Method and Open Excavation Method. Analytic Hierarchy Process (AHP). sustainability, 11, 1-18. doi:10.3390/su11082438

Leu, S.S. et al. (2025). Risk, robustness and microtunneling machine selection under geological uncertainty of clayed sand layer in Taiwan. Tunnelling and Underground Space Technology incorporating Trenchless Technology Research, 165, 1-9. doi:10.1016/j.tust.2025.106876

Lin, S.S. et al. (2023). Optimal construction method evaluation for underground infrastructure construction. Automation in Construction, 152, 1-17. doi:10.1016/j.autcon.2023.104921

Linner, T. (2024). Editorial: Convergence of digital construction, novel physical construction technologies and sustainable construction. Proceedings of the Institution of Civil Engineers –Engineering Sustainability, 177, 185-186. doi:10.1680/jensu.2024.177.4.185

Ma, P. et al. (2023). Transition of the pipe jacking technology in Japan and investigation of its application status. Tunnelling and Underground Space Technology incorporating Trenchless Technology Research, 139, 1-14. doi:10.1016/j.tust.2023.105212

Malczewski, J. (20. Februar 2006). GIS-based multicriteria decision analysis: a survey . International Journal of Geographical Information Science, S. 703-726.

Marinos V. et al. (2018). Selection of TBM and geotechnical assessment of a microtunnel in a difficult geological environment: a case of a natural gas pipeline beneath an active landslide (Albania). Bulletin of Engineering Geology and the Environment, 78, 1795-1813. doi:10.1007/s10064-018-1241-6

Matthews J. et al. (2018). Multi-segment trenchless technology method selection algorithm for buried pipelines. Tunnelling and Underground Space Technology, 73, 295-301. doi:10.1016/j.tust.2018.01.001

Neuling, E. (2013). Stromfluss unter der Erde - Einsatz von Erdkabeln beim Übertragungsnetzausbau. Berlin: Naturschutzbund Deutschland (NABU) e.V.

Pache, S., Scharun, S., & Hochmuth, M. (2024). BIM-GIS-Integration im Rahmen der Planung von Infrastrukturprojekten. In D. e.V., & R. T. e.V., Leitfaden Geodäsie und BIM - Version 4.0 (S. 237-240). Bühl; München.

Page, M. et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 372, 1-9. doi:10.1136/bmj.n71

Pan M. et al. (2018). A framework of indicators for assessing construction automation and robotics in the sustainability context. Journal of Cleaner Production, 182, 82-95. doi:10.1016/j.jclepro.2018.02.053

Pan M. et al. (2020). Influencing factors of the future utilisation of construction robots for buildings: A Hong Kong perspective. Journal of Building Engineering, 30, 1-13. doi:10.1016/j.jobe.2020.101220

Parka, A. et al. (2020). Selection of pressure linings used for trenchless renovation of water pipelines. Tunnelling and Underground Space Technology, 98, 1-11. doi:10.1016/j.tust.2019.103218

Rokstad, M. et al. (2025). Pardon my trench: reflections on the uptake of trenchless technologies in the Norwegian water sector. Water Science & Technology, 91, 83-92. doi:10.2166/wst.2024.397

Schito, J., & Grassi, S. (2015). Berechnung optimaler HSL-Pfade und Korridore für Hochspannungsleitung. In R. T. e.V., TUM, & R. T. e. V. (Hrsg.), Leitfaden 3D-GIS und Energie (S. 80-82). München.

Steuer, H., & Donaubauer, A. (2012). Geo-Web-Services im Rahmen der kollaborativen Trassenplanung. In Geoinformationssysteme - Beiträge zum 17. Münchener Fortbildungsseminar (S. 159-167). Berlin; Offenbach: Herbert Wichmann Verlag.

TenneT TSO GmbH. (2025). Tennet - Projekt SuedLink. (TenneT TSO GmbH) Abgerufen am 12. Juni 2025 von https://www.tennet.eu/de/projekte/suedlink

Zwierzchowska, A., & Kuliczkowska, E. (2019). The selection of the optimum trenchless pipe laying technology with the use of fuzzy logic. Tunnelling and Underground Space Technology, 84, 487-494. doi:10.1016/j.tust.2018.11.030

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

ONTOLOGY-BASED COST MANAGEMENT IN PUBLIC INFRASTRUCTURE PROJECTS: A CASE STUDY ON TRANSFERABILITY AND APPLICABILITY

Anna Nast[1], Jacopo Cassandro[2], Claudio Mirarchi[2], and Jürgen Melzner[1]
[1]Chair of Construction Engineering and Management, Faculty of Civil and Environment Engineering, Bauhaus-Universität Weimar (BUW), Weimar, Germany
[2]Department of Architecture, Built Environment, and Construction Engineering (DABC), Polytechnic of Milan, Milan, Italy

Abstract

Cost management in infrastructure projects is complex due to evolving cost information, multiple estimation methods, and national regulations. Ontology-based approaches combined with Building Information Modeling (BIM) can improve cost consistency and decision-making; however, their transferability and applicability across domains and countries remain insufficiently investigated. This paper examines how an existing ontology-based cost management approach can be transferred to German railway infrastructure projects. A case study explores the reuse of cost ontologies across construction domains and under a different national regulatory framework. The results suggest that ontology-based cost structures support cost development and semantic consistency in standardized infrastructure projects.

Introduction

Cost management is one of the most critical and challenging tasks in infrastructure project management. Total project costs are often unknown until after the conceptual design phase and, in some cases, even after construction completion (Moses, 2020; Ashworth & Perera, 2018). The use of different estimation methods, cost indicators, and classification systems throughout the project lifecycle frequently leads to inconsistencies, limited transparency, and error-prone decision-making (Cassandro et al., 2025).

Digital methods such as Building Information Modeling (BIM) have significantly improved the availability and integration of project data. In parallel, ontology-based approaches have been proposed to formalize cost knowledge, enable semantic interoperability, and support structured cost analyses. Previous studies indicate that ontologies enhance consistency and traceability in cost management by providing a shared semantic framework across disciplines and phases (Signorini et al., 2025).

Nevertheless, current ontology-based cost management approaches exhibit two main limitations. First, they often focus on highly detailed and static cost calculations without explicitly considering the evolution of cost information across different project phases. Second, most approaches are developed and validated within a single domain or national context, raising concerns regarding their transferability, applicability, scalability, and generalizability. This paper specifically addresses the second limitation by examining the application of an existing ontology-based cost management approach in the context of German railway infrastructure projects.

In this study, *transferability* refers to the reuse of the ontology-based cost model in a different construction domain, such as infrastructure instead of buildings, without fundamental changes to its core ontological classes and their relationships (Korstjens & Moser, 2018). *Applicability* refers to the feasibility of implementing the approach in a different regulatory and organizational context, including national cost management rules, such as the German instead of the Italian context. *Scalability* denotes the capability of the approach to support different project sizes and work types, for example small-scale rather than large-scale projects or demolition rather than new construction. *Generalizability* addresses the extent to which findings can be applied beyond a single case study, for example through the development of a reference project based on a larger project sample.

Following the recommendations of Farghaly et al. (2023), this paper emphasizes the reuse of existing ontologies rather than the development of new, isolated models. Specifically, it builds on the ontology-based cost management approach proposed by Cassandro et al. (2023), which has so far been validated primarily in building projects within the Italian regulatory framework, while its conceptual structure is designed to support multiple domains and national standards.

The objective of this paper is to explore the transferability and applicability of the ontology-based approach to infrastructure projects and to a distinct national context, namely the German railway sector. German railway infrastructure projects are characterized by specific cost management rules, standards, regulations, and organizational features that differ substantially from those applied in prior case studies in the Italian construction industry, making this context a suitable testbed for assessing cross-national applicability and methodological robustness.

While cost management procedures, aggregation methods, and regulatory requirements vary across national contexts and application domains, the underlying *cost items* – defined as price components and their

technical descriptions – remain conceptually comparable. In infrastructure projects, however, unit prices are often unstable and not directly reusable, as they strongly depend on site-specific conditions, construction sequences, access constraints, and track closures.

Therefore, the objective of ontology-based cost management is not to standardize prices, but to semantically structure cost item descriptions in a way that enables consistent association, comparison, and updating of cost values across projects. This study is based on the assumption that variations across contexts primarily affect how cost data are structured, interpreted, and managed, rather than the cost data themselves. By semantically structuring cost items and their descriptions through an ontology-based approach, cost information becomes machine-interpretable, verifiable, and consistently comparable across projects and national frameworks.

By transferring the existing ontology-based cost structure to the German railway context, this paper demonstrates its scalability, semantic depth, and ability to support cost data management despite differences in national regulations and industry practices. The case study of standardized German railway infrastructure projects therefore offers an exploratory validation of ontology reuse in a cross-domain and cross-national setting.

Background

Cost management

Cost management encompasses the processes of planning, estimating, budgeting, and controlling costs to ensure that projects remain financially viable and achieve their intended objectives (Project Management Institute, 2021). These activities enable practitioners to forecast expenditures, establish cost baselines, evaluate cost performance, and implement corrective actions when deviations occur.

In the construction sector, cost management spans the entire project lifecycle and addresses direct and indirect costs, including labor, materials, equipment, subcontractors, contingencies, overheads, and life-cycle expenditures (Ashworth & Perera, 2018; Smith, 2016). Effective cost management relies on structured mechanisms such as cost breakdown structures, detailed budgets, real-time cost tracking systems, and change control procedures (AACE International, 2022). Despite their maturity, these practices often remain dependent on manual processes, fragmented data sources, and heterogeneous classification systems.

The increasing adoption of the BIM technology has enabled more integrated 5D cost workflows. However, commercial 5D BIM tools typically rely on proprietary cost structures and require extensive manual mapping between model elements and cost items. This limits transparency, interoperability, and the systematic verification of consistency between geometric, quantitative, and cost domains (Cassandro et al., 2023). As a result, cost management processes rarely achieve the level of robustness required to meet public procurement, standardization, and auditability requirements, particularly in small-scale public works, where limitations in time, expertise, and budget hinder the adoption of complex digital ecosystems (Nast & Koch, 2023).

Ontology-based approaches

Ontologies, defined as “an explicit specification of a conceptualization” (Neches et al., 1991), have gained traction in the Architecture, Engineering, and Construction (AEC) sector as a means to address interoperability challenges and integrate heterogeneous datasets (Pauwels & Terkaj, 2016). They determine domain-specific concepts and relationships (Farghaly et al., 2023) and enable machine-interpretable knowledge representations using languages such as OWL (Web Ontology Language) and RDF (Resource Description Framework), supporting semantic interoperability, rule-based reasoning, and knowledge reuse. Recent systematic reviews confirm the growing adoption of ontology-based applications across AEC domains to address interoperability issues, although practical applications remain fragmented and constrained by data heterogeneity and scalability limitations (Farghaly et al., 2023; Palihakkara et al., 2025; Speiser et al., 2025).

In the cost domain, early research focused on representing work items, construction methods, and work conditions through ontologies to enable automated cost estimation from BIM models. Lee et al. (2014) proposed a BIM- and ontology-based cost-estimation approach that uses ontological reasoning to infer appropriate work items from BIM-derived work conditions, incorporating contextual information such as room usage and material attributes. Subsequent studies expanded the formalization of specifications and cost rules (Ma et al., 2016) or represented cost classification systems, such as the New Rules of Measurement (NRM), using ontological engineering approaches (Abanda et al., 2017). Similarly, Liu et al. (2016) developed a construction-oriented product ontology for quantity take-off (QTO) in light-frame construction, extending BIM schemas with domain-specific terms and enabling automated reasoning for QTO extraction. Zhang et al. (2022) used an ontological approach to conceptualize knowledge for budgeting real estate developments.

More recent research focuses on integrating cost ontologies with IFC-based models and Linked Building Data to enable semantic links between geometric entities, cost items, and resource information (Hagedorn et al., 2025). These approaches show that ontologies can support automated matching between BIM elements and cost structures, improve the transparency of cost logic, and reduce manual effort in traditional 5D BIM workflows. Further studies demonstrate how ontology-based models support the assessment of IFC class reliability for cost estimation and enable standardized digital cost representations for public procurement (Cassandro et al., 2023).

Infrastructure projects

Transportation infrastructure is typically grouped into asset classes such as roads and highways, railways, bridges, and related asset types (Bradley et al., 2016). However, the research literature offers no widely accepted, sector-wide thresholds for classifying project size. Accordingly, project size is usually defined case-by-case by the contracting authority or the implementing organization. Building on Nast & Koch (2023), this study operationalizes small-scale infrastructure projects primarily based on the physical scale of the asset. Examples include single-sided railway platforms, bus stops, rest areas, and service stations.

In the European infrastructure context, small-scale construction measures are particularly relevant. For example, in Germany, approximately 80–90 % of railway stations are small-scale facilities with side platforms (around 9,300 assets) (DB InfraGO AG, 2025). Additionally, railway infrastructure is selected as the focus of this study, as it represents a highly standardized yet cost-critical domain with substantial cost overruns compared to other infrastructure sectors, such as road construction (Flyvbjerg et al., 2003).

Standardization in AEC

Like other manufacturing industries, the AEC sector relies on standards, guidelines, and codes of practice to ensure safety and reduce errors. Standardization is widely recognized as a prerequisite for BIM implementation (Moses et al., 2020), for effective 5D BIM-based cost planning and management, and for the development of domain-specific ontologies (Signorini et al., 2025). In this context, standardization provides the base for semantic interoperability, and BIM aligns closely with the componentization and integration of building elements.

Cost management challenges in public projects

Public-sector projects must comply with standardized measurement rules, regional price lists, and strict auditability requirements. However, cost information is often encoded in natural language or non-interoperable software formats, hindering automated verification and cross-project reuse (OECD, 2017). Fragmentation is further exacerbated by the coexistence of diverse standards, such as NRM, UniClass, MasterFormat, and national or regional price books. Although classification systems are recognized as essential in both international and national contexts (Pupeikis et al., 2022), current digital workflows for estimation, budgeting, and cost control frequently lack semantic consistency, particularly in public projects where transparency, standardization, and reproducibility are mandatory.

This study aims to extend the use of existing ontology-based cost management approaches to the domain of small-scale public infrastructure projects. In doing so, it addresses the specific challenges of cost management in infrastructure projects while also helping to close a gap in the practical application of current ontology-based approaches.

Methodology

This study adopts a qualitative research design to investigate the transferability and applicability of existing ontology-based approaches for cost management. In line with Yin (2018), document analysis and expert interviews were used as primary sources of evidence. A representative small-scale railway project was first defined on the basis of publicly available documents from DB InfraGO AG, Business Unit Passenger Railway Stations, expert interviews, and anonymized project data. This reference case was then used to evaluate the ontology-based cost management approach proposed by Cassandro et al. (2023) during the procurement phase. Selected items from the Bill of Quantities (BoQ) were used in the experiment, and the approach was assessed with respect to transferability, applicability, scalability, and generalizability. Finally, the results were discussed and directions for future research were derived (see Fig. 1).

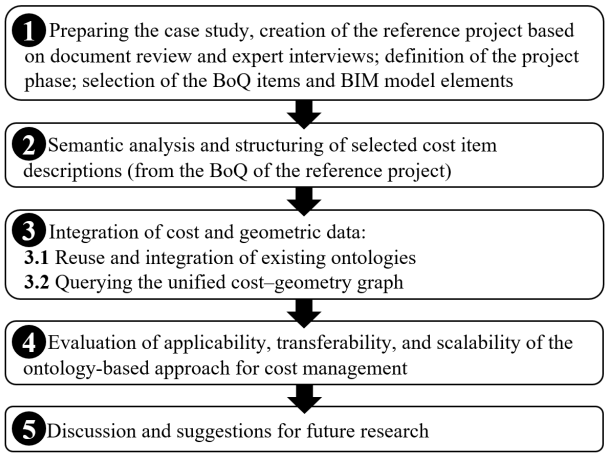


Figure 1: Overview of the research process

The adopted ontology-based approach explicitly separates the semantic definition of cost items from their associated economic values. While cost item descriptions remain stable and reusable across projects, cost values are modeled as context-dependent attributes that vary according to project-specific conditions, construction phases, and estimation scenarios. This distinction is particularly relevant for railway infrastructure projects, where cost estimation is largely driven by construction sequencing, logistics constraints, and site access conditions rather than standardized unit prices.

Case study

To address one of the key limitations of existing ontology-based approaches for cost-related activities – namely the insufficient validation and extension of existing ontologies – a case study was conducted in the German AEC context. The railway station sector was selected because it exhibits a high degree of standardization (DB InfraGO AG, 2025), remains underrepresented in research despite its practical relevance (Bawono et al., 2020), and combines both building and infrastructure construction characteristics.

The German railway infrastructure company DB InfraGO AG has established comprehensive construction standards for major facilities within the railway station domain (DB InfraGO AG, 2025). These standards govern planning, engineering, and construction activities at passenger stations and include standard drawings and details, sample service specifications (BoQ), structural design calculations, technical specifications, planning manuals, and approved building components.

A reference project was developed as a case study representing the typical characteristics of small-scale railway stations. The core scope comprises the construction of a new side platform, elevated and extended to ensure accessibility and improve infrastructure performance. In Germany, railway stations are rarely developed on greenfield sites. Therefore, the reference project involves the complete dismantling and reconstruction of the existing platform at the same location. All structural components, including platform edges, foundations, surface covering, and underground cabling, will be removed and replaced by new installations (see Fig. 2).



Figure 2: A typical railway station project

A key characteristic of the reference project is the strong influence of track closures, logistics constraints, and site conditions on cost calculation, particularly with respect to the selection and quantity of resources such as equipment, labor, and materials. Compared with typical building projects, construction scheduling must be considered much more explicitly when estimating costs for railway stations during the tender phase. Consequently, standardized unit prices are generally of limited applicability. In addition, railway infrastructure projects are subject to a large number of nationally specific guidelines, regulations, and approval procedures.

However, the main challenge is not limited to the coexistence of different national standards and classification systems. Cost information is also fragmented across heterogeneous document types and is often encoded in natural language or stored in non-interoperable formats. As a result, ontology-based representation and querying require concepts, quantities, and rules to be semantically disambiguated and formalized before they can be consistently linked across domains. Moreover, construction cost items are typically defined at a relatively high level of abstraction and do not adequately support resource-based cost analysis, which further complicates detailed semantic modeling.

Results

Based on the methodological steps illustrated in Figure 1, the results are organized to reflect the sequential application of the ontology-based approach to the selected cost items. Each result corresponds to a distinct phase of the research process, ranging from the semantic analysis of cost item descriptions to the integration of cost and geometric data and the evaluation of transferability and applicability.

Due to the limited scope of this paper, the procurement phase, based on the definitive estimate, was selected as the starting point for transferring the ontology-based approach to the infrastructure context. To ensure a focused analysis, the approach was applied to a selected subset of the BoQ items. Specifically, five representative cost items were examined, referring to three key construction elements that are fundamental to most railway station projects and include both construction and demolition activities: the precast foundation element, the precast platform edge, and the precast corner element (see Fig. 3). Despite the limited number of cost items analyzed, the selected elements represent standardized components of railway station projects and therefore provide a suitable basis for assessing the applicability, transferability, and scalability of the approach.

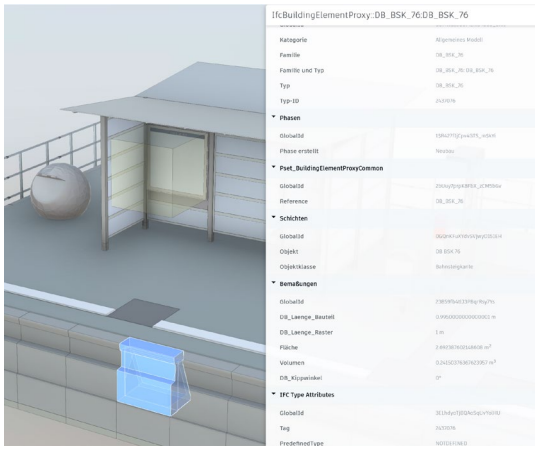


Figure 3: Selection of construction elements

Semantic structuring of cost items

In the first step, the case study focuses on the semantic analysis of railway cost items as defined in the project documentation. As illustrated in Figure 4, individual cost item descriptions typically combine heterogeneous information related to construction work, material resources, activities, and technical specifications within a single unstructured textual entry.

By adopting the ontology-based approach proposed in Cassandro et al. (2025) and Signorini et al. (2025), these cost items can be decomposed into semantically well-defined components, including:

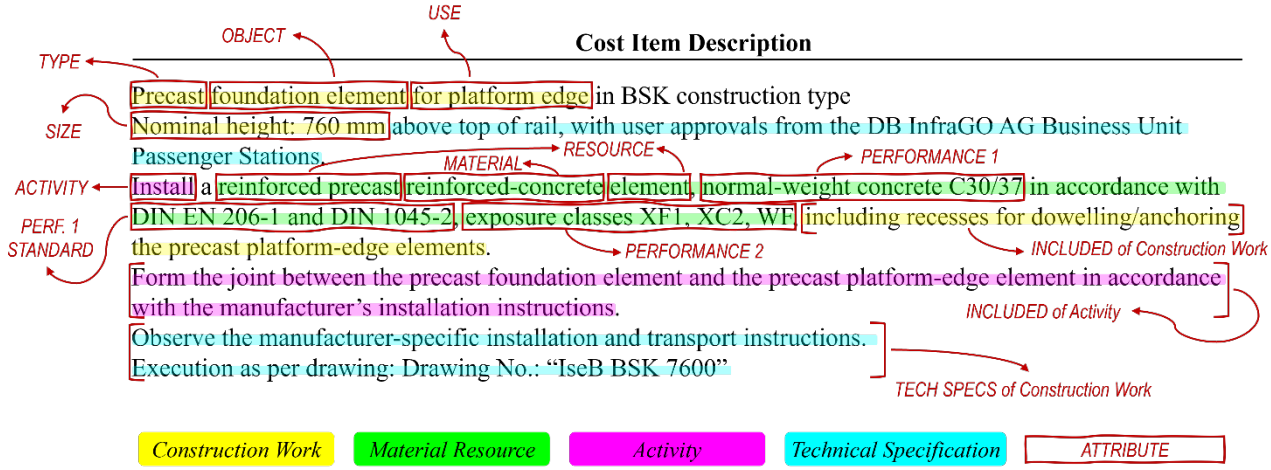


Figure 4: Semantic structuring of cost item 'Precast foundation element for platform edge'

- Construction Work Object (e.g., precast foundation element)
- Activity (e.g., installation)
- Material Resource (e.g., reinforced precast element)
- Technical and performance specifications (e.g., compliance with manufacturer-specific instructions).

This semantic structuring reveals several recurring issues in traditional cost descriptions, such as semantic overlap, implicit assumptions, and ambiguities between objects, processes, and requirements. By making these distinctions explicit and structuring the information into well-defined attributes (e.g., object, type, material, use, and performance), the ontology provides a clearer and more coherent definition of cost items while preserving their original meaning and enabling machine interpretability.

Reuse and integration of existing ontologies for cost – geometry linkage

In a second step, the study demonstrates how existing ontologies can be reused and integrated to structure cost items and establish semantic links to geometric data. Following ontologies were selected:

- *ifcOWL* is employed to generate an RDF representation of the IFC model, enabling access to geometric objects and their properties.
- *Cost Item Ontology (CI)* is used to formalize cost items and their internal structure based on the previously semantic structuring.
- *Construction Resource Ontology (CR)* is used to establish explicit relationships between geometric objects and cost items.

These ontologies were selected because they enable explicit semantic structuring of cost items and their direct linkage to geometric data. Many alternative ontologies focus on process automation or isolated knowledge representation and lack the semantic granularity required to jointly model and evaluate relationships between cost information and geometry (see Chapter 'Background'). In the cost domain, this explicit linkage is essential, as quantities, construction logic, and resource requirements are directly derived from geometric definitions.

Through the integration of these ontologies, a unified knowledge graph is generated in which each cost item is explicitly linked to one or more corresponding geometric objects (see Fig. 5). This linkage enables automated QTO extraction directly from IFC-derived data and their consistent association with cost items within a shared semantic framework.

As a result, quantities, cost item definitions, and unit prices are no longer managed as separate artifacts but are integrated into a single machine-interpretable model. This enables the generation of a computable and extensible BoQ, in which cost values can be updated, revised, or compared across scenarios, provided that such variations are driven by external cost factors that do not affect the semantic definition of cost items, their resource composition, or measurement rules. Typical examples include variations in labor, equipment, and material costs due to geographic location, market conditions, or alternative construction and procurement scenarios, rather than changes in construction logic or resource structure.

The absence of standardized unit prices in German railway projects does not affect the validity of the proposed approach. Project-specific unit prices can be assigned, updated, and modified independently of cost item definitions and object relationships, as long as such modifications reflect contextual cost variability rather than changes in quantity take-off logic or resource breakdown, thereby demonstrating the scalability, adaptability, and reuse potential of the ontology-based cost structure.

Querying the unified cost – geometry graph

Finally, the resulting knowledge graph is queried using SPARQL to demonstrate its practical applicability. The SPARQL queries are not intended for statistical analysis but are employed as a verification mechanism to demonstrate the internal consistency and operational applicability of the ontology-based model.

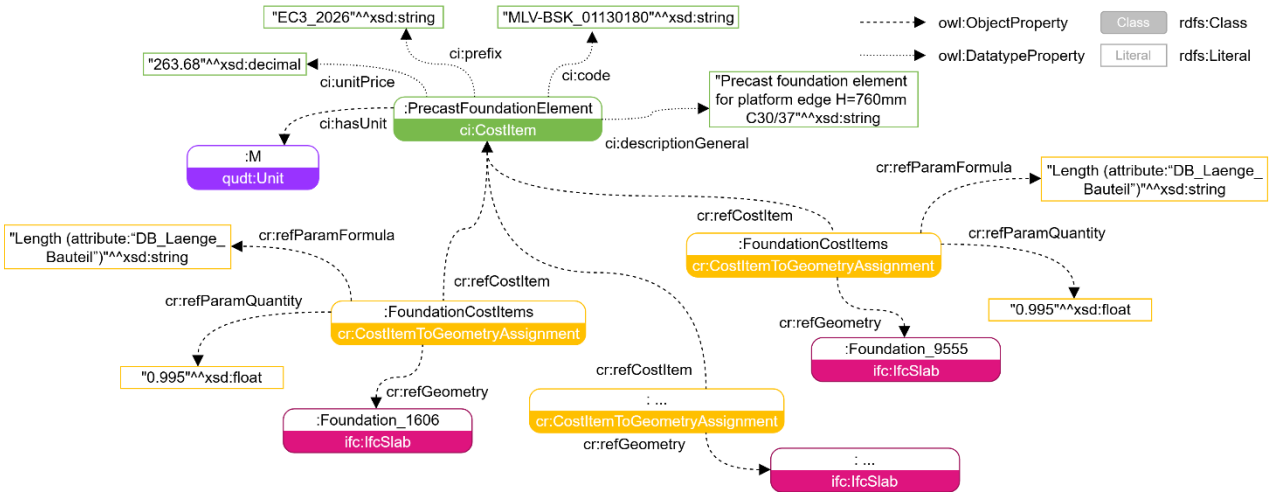


Figure 5: Unified knowledge graph for precast foundation element

The query shown in Listing 1 interrogates the integrated knowledge graph by explicitly crossing the semantic relationships between construction objects (`?obj`), cost assignments (`?assign`), and cost items (`?costItem`). It retrieves all construction objects (`?obj`) participating in a `cr:CostItemToGeometryAssignment`, independent of their specific IFC class. For each object, the query extracts its globally unique identifier (`?guid`) through the `ifc:globalId_IfcRoot` relationship and the associated `IfcGloballyUniqueId` literal (`express:hasString`), ensuring object-level traceability across geometric, cost, and quantity domains.

```
PREFIX rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ifc:
    <http://standards.buildingsmart.org/IFC/DEV/IFC2x3/TC1/OWL#>
PREFIX express: <https://w3id.org/express#>
PREFIX ci:      <https://w3id.org/ci#>
PREFIX cr:      <https://w3id.org/cr#>
PREFIX xsd:     <http://www.w3.org/2001/XMLSchema#>
SELECT DISTINCT ?obj ?guid ?qto ?code ?unitPrice
    ?measureCost ?formula
WHERE {
  ?obj rdf:type ?objType .
  ?obj ifc:globalId_IfcRoot ?guidNode .
  ?guidNode express:hasString ?guid .
  ?assign rdf:type cr:CostItemToGeometryAssignment .
  ?assign cr:refGeometry ?obj .
  ?assign cr:refCostItem ?costItem .
  ?assign cr:refParamQuantity ?qto .
  ?assign cr:refParamFormula ?formula .
  ?costItem ci:code ?code .
  ?costItem ci:unitPrice ?unitPrice .
  ?costItem ci:hasUnit ?measureCost .
  }
ORDER BY ?obj
```

Listing 1: SPARQL query

Cost information is accessed via the `cr:CostItemToGeometryAssignment` node, which acts as a semantic mediator between geometry and cost data. Through this assignment, the query retrieves:

- the referenced cost item (`cr:refCostItem`),
- the quantity parameter used for cost calculation (`cr:refParamQuantity`),
- and the applied cost formula (`cr:refParamFormula`).

From the linked cost item, the query extracts:

- cost code (`ci:code`), enabling classification and aggregation of cost items,
- the unit price (`ci:unitPrice`),
- and unit of measurement (`ci:hasUnit`).

The resulting dataset consolidates each cost – geometry association in a single row, explicitly linking a construction object to its cost item, quantity reference, unit price, unit of measurement, and calculation formula. By ordering the results by construction object, the query facilitates systematic object-level inspection and validation.

Unlike traditional document-centric cost management approaches (e.g., Excel files and IFC data) with rigid data schemas, ontologies and knowledge graphs enable an explicit, machine-interpretable representation of the semantic relationships between cost items, construction objects, and quantities. The query results demonstrate that cost and geometric information – traditionally distributed across separate documents, spreadsheets, and software systems – can be consistently retrieved from a single integrated model.

Evaluation of the ontology-based approach

The reuse of the selected ontology-based approach to cost data management would not be feasible if cost items could not be described in a sufficiently structured manner or if their semantic characteristics could not be mapped to the core ontological classes and properties, such as object, material, function, and form. In the present study, however, the cost item descriptions reveal a high degree of semantic consistency, while the alignment with the generic ontology classes is preserved, demonstrating the applicability and robustness of the approach.

Discussion

This case study shows that the ontology-based cost structure remains applicable and robust in German railway projects, even in the absence of standardized unit prices. Project-specific cost values can be assigned, updated, and compared independently of the underlying cost item definitions, which confirms the validity of separating semantic meaning from monetary values. Although pricing mechanisms differ between building construction and infrastructure projects, the study shows

a fundamental commonality in cost calculation: the semantic structure, granularity, and progressive development of cost information are consistent across domains, irrespective of national or organization-specific cost management practices.

The prior standardization of railway station projects proved to be a crucial prerequisite for transforming cost information into machine-interpretable structures. This finding further underlines the role of standardization as a fundamental enabler of digitalization and process automation in the AEC sector, particularly in the context of transparent and traceable cost management (Signorini et al., 2025). Establishing a reliable benchmark therefore requires the initial definition of a standardized project typology, followed by the incremental integration of cost information reflecting project-specific conditions.

The mapping between project data and ontology classes is performed manually. While this increases the workload and therefore constitutes a current limitation of the ontology-based approach, the mapping process could be automated using AI-based methods. However, such automation would require appropriate training data to ensure reliable mapping quality. Since sufficiently annotated datasets are not yet available, automated mapping was beyond the scope of the present study.

This paper focuses on the ontology-based approaches using BIM data, which typically represents the detailed stage of an infrastructure project. Early project phases are not covered, as they rely on other geometric data sources, and should be considered in future studies.

Conclusion

While Farghaly et al. (2023) emphasize the importance of testing, reusing, and transferring existing ontologies to new contexts, this paper extends the application of ontology-based cost management approaches to the domain of small-scale public infrastructure projects. In doing so, it explores the scalability, semantic depth, and robustness of the ontology-based cost structure under varying project characteristics, national regulations, and organization-specific cost management practices.

This study builds on previous research by Cassandro et al. (2025) and Signorini et al. (2025) on ontology-based approaches for structuring cost information and improving cost management, particularly with regard to transparency and decision-making. More specifically, it investigates the reuse of the Cost Item Ontology and the Construction Resource Ontology in the context of a German railway infrastructure, including both the representation of cost definitions and the queryability of the encoded data. In this way, the case study addresses one of the key limitations of current ontology-based approaches, namely their restricted applicability due to the lack of validation across different domains and national contexts.

Transferability was demonstrated by applying the ontology-based cost model to both building construction and infrastructure projects without requiring fundamental changes to its core ontological classes and relationships between them. *Applicability* was ensured by adapting the approach to a different national and organizational context, such as the German instead of the Italian context. *Scalability* was supported through the selection of small-scale projects relevant for industry adoption, as well as by including different construction activities such as prefabricated construction and demolition works. *Generalizability* was achieved through the use of a reference project derived from approximately 9,300 railway station platforms.

Across the case study, cost information for the reference project was represented in a structured, consistent, and transparent manner. The results confirm that cost values can be modified, compared, and reassigned without affecting the semantic integrity of cost items, directly supporting the methodological assumption of separating meaning from monetary value.

A limitation of this study is that only a subset of items from the reference BoQ was examined. Accordingly, the findings should be regarded as preliminary evidence of structural similarities between the Italian and German cost management contexts under investigation, as well as of the potential transferability of the ontology-based approach, rather than as confirmation of broad cross-context validity. Future studies should therefore expand the analysis to cover all construction trades.

Another limitation of this paper is its focus on detailed cost estimation during the procurement phase, which was necessary to assess transferability based on fully specified cost descriptions. Future research should extend the ontology-based approach to early cost estimation and demonstrate the dynamic growth and evolution of cost information across the project lifecycle.

Acknowledgement

The authors would like to thank the Lombardy region and DB InfraGO AG, Business Unit Passenger Railway Stations, for their support, knowledge sharing, provision of data and feedback.

Online Resources

The sources and documentation are available via https://github.com/Cassa97/2026_EC3_Infrastructure-Cost-Ontology.

References

Abanda, F.H., Kamsu-Foguem, B. and Tah, J.H.M. (2017) BIM – New rules of measurement ontology for construction cost estimation, Engineering Science and Technology, an International Journal, 20(2), pp. 443–459.

AACE International (2022) Recommended practices: https://web.aacei.org/resources/recommended-practices (Accessed: 8 February 2026).

Ashworth, A. and Perera, S. (2018) Cost studies of buildings. Abingdon: Routledge.

Bawono, A.A., von Schumann, C.M. and Lechner, B. (2020) Study of building information modelling

implementation on railway infrastructure, in Proceedings of the 18th ICCCBE 2020. Lecture Notes in Civil Engineering, 98. Cham: Springer, pp. 372–382.

Bradley, A., Li, H., Lark, R. and Dunn, S. (2016) BIM for infrastructure: an overall review and constructor perspective, Automation in Construction, 71, pp. 139–152.

Cassandro, J., Donatiello, M.G., Mirarchi, C., Zanchetta, C. and Pavan, A. (2023) Reliability of IFC classes in ontology definition and cost estimation of public procurement, in Proceedings of the 2023 EC³.

Cassandro, J., Hagedorn, P., Sigalov, K., Gatto, C., Mirarchi, C., Pavan, A. & König, M. (2025) Semantic Integration of Cost, Time and Geometry for a Resource-centred Perspective of Managing Construction Projects, in Proceedings of the 2025 EC³ and the 42nd CIB W78. Porto, Portugal, 14–17 July.

DB InfraGO AG – Passenger Railway Stations (2025) Information platform: https://infoplattform-personenbahnhoefe.deutschebahn.com/pbhf (Accessed: 8 February 2026).

Farghaly, K., Soman, R.K. and Zhou, S.A. (2023) The evolution of ontology in AEC: a two-decade synthesis, application domains, and future directions, Journal of Industrial Information Integration, 36, 100519.

Flyvbjerg, B., Skamris Holm, M.K. and Buhl, S.L. (2003) How common and how large are cost overruns in transport infrastructure projects? Transport Reviews, 23(1), pp. 71–88.

Hagedorn, P., Cassandro, J., Sigalov, K., Mirarchi, C., Pavan, A. and König, M. (2025) A Shared Construction Resource Ontology for Semantically Aligning Cost and Time Domains in Construction Projects, in Proceedings of the 13th LDAC 2025, Porto, Portugal, 9–11 July. CEUR Workshop Proceedings, 4113, pp. 10–24.

Korstjens, I. and Moser, A. (2018) Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing, European Journal of General Practice, 24(1), pp. 120–124.

Lee, S.-K., Kim, K.-R. and Yu, J.-H. (2014) BIM and ontology-based approach for building cost estimation, Automation in Construction, 41, pp. 96–105.

Liu, H., Lu, M. and Al-Hussein, M. (2016) Ontology-based semantic approach for construction-oriented quantity take-off from BIM models in the light-frame building industry, Advanced Engineering Informatics, 30(2), pp. 190–207.

Ma, Z., Liu, Z. and Wei, Z. (2016) Formalized representation of specifications for construction cost estimation by using ontology, Computer-Aided Civil and Infrastructure Engineering, 31(1), pp. 4–17.

Moses, T.G., Heesom, D. and Oloke, D.A. (2020) Implementing 5D BIM on construction projects: contractor perspectives from the UK construction sector, Journal of Engineering, Design and Technology, 18(6), pp. 1867–1888.

Nast, A. and Koch, C. (2023) BIM adoption in small-scale infrastructure projects – investigation on the German railway sector, in Proceedings of the 14th ECPPM, Trondheim, Norway, 14–16 September 2022. Abingdon: CRC Press/Balkema, pp. 695–702.

Neches, R., Fikes, R., Finin, T., Gruber, T.R., Patil, R.S., Senator, T.E. and Swartout, W.R. (1991) Enabling technology for knowledge sharing, AI Magazine, 12(3), pp. 36–56.

OECD (2017) Public procurement for innovation: Good practices and strategies. OECD Public Governance Reviews. Paris: OECD Publishing.

Palihakkara, A.D., Osorio-Sandoval, C.A. and Tizani, W. (2025) A review of the application of ontologies and semantic web for building information modelling and digital twins-based construction management, Journal of Information Technology in Construction (ITcon), 30, pp. 1208–1243.

Pauwels, P. and Terkaj, W. (2016) EXPRESS to OWL for construction industry: towards a recommendable and usable ifcOWL ontology, Automation in Construction, 63, pp. 100–133.

Project Management Institute (2021) The Standard for Project Management and a Guide to the Project Management Body of Knowledge (PMBOK® Guide). 7th edn. Newtown Square.

Pupeikis, D., Navickas, A.A., Klumbytė, E. and Šeduikytė, L. (2022) Comparative study of construction information classification systems: CCI versus Uniclass 2015, Buildings, 12(5), 656.

Signorini, M., Gatto, C., Cassandro, J., Pavan, A. and Lupica Spagnolo, S. (2025) Cost efficiency in buildings: an ontological perspective for sustainable life cycle management, Sustainability, 17(13), 5685.

Smith, J., Jaggar, D.M. and Love, P. (2016) Building cost planning for the design team. 3rd edn. Abingdon: Routledge.

Speiser, K., Seiß, S., Boukamp, F., Melzner, J. and Teizer, J. (2025) From fragmented data to unified construction safety knowledge: A process-based ontology framework for safer work, Automation in Construction, 176, 106293.

Yin, R.K. (2018) Case study research and applications: design and methods. 6th edn. Thousand Oaks, CA: SAGE Publications.

Zhang, Z., Liu, H., Xu, J., Shu, Y., Liu, H. and Xiao, J. (2022) Ontology-based integrated cost management system for real estate development, in ICCREM 2022. Reston, VA: American Society of Civil Engineers (ASCE), pp. 838–854.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

GRAPH-BASED GENERATIVE SCHEDULING IN CONSTRUCTION PROJECT MANAGEMENT

Jianpeng Cao[1], Hisham Said[2]
[1]The University of Hong Kong, Hong Kong S.A.R., China
[2]Santa Clara University, Santa Clara, CA, United States

Abstract

Factory-based modular construction resembles manufacturing systems, but production scheduling remains challenging due to design variability, product heterogeneity, and design–production coupling. Existing construction scheduling methods rely on predefined task structures and adapt poorly to design changes. This paper presents a graph-based generative scheduling framework that derives production plans directly from modular designs. A design graph is transformed through graph rewrite rules to generate tasks, resources, precedence constraints, and task durations. The resulting job shop scheduling problem is solved using constraint programming. Case studies demonstrate automated schedule generation, reveal resource bottlenecks, and support early-stage production capacity planning.

Introduction

The increasing adoption of modular and prefabricated construction has intensified the need for advanced production planning and scheduling methods that can effectively manage complex task precedence, resource constraints, and design variability (Said and Kandimalla 2018). Factory-based construction environments share many characteristics with manufacturing systems, yet construction-specific uncertainties, heterogeneous products, and tight coupling between design and production continue to limit the effectiveness of traditional scheduling approaches (Ballard and Howell 1998; Gibb and Isack 2003). In particular, translating design information into executable production schedules remains a largely manual and error-prone process.

Job shop scheduling models have been widely applied to prefabrication and industrialized construction to optimize makespan, resource utilization, and workflow reliability. Specifically, job shop scheduling has been applied to the fabrication of pipe spools (ElMenshawy et al. 2025; Karim et al. 2024; Liu et al. 2017; Safarzadeh et al. 2018), precast concrete elements (Anvari et al. 2016; Xie et al. 2023), and sheet metal assemblies (Said and Kandimalla 2018). However, most existing approaches rely on predefined task structures and static process models, which restrict adaptability to design changes and limit scalability across different module configurations. Recent research has highlighted the potential of graph-based representations to model production systems more flexibly, capturing both structural and operational relationships in a unified framework.

This paper proposes a graph-based generative scheduling methodology that directly links modular design representations with production planning through graph rewrite rules. By transforming a parametric module design graph into a task and resource graph, the approach enables automated formulation of a job shop scheduling problem with design-dependent task durations and constraints. While the underlying optimization follows standard job shop scheduling (JSP) formulations, the innovation lies in the automated, graph-based generation of the constraint space, a process that traditionally relies on manual, error-prone expert input. In this context, the constraint space encompasses the set of logical dependencies (such as precedence, resource availability, and spatial conflicts) that bound the feasible solutions for a project. By serving as a generative engine, the proposed graph transformation process maps architectural topology directly onto these constraints. The framework is demonstrated through a modular production case study and solved using constraint programming to generate optimized schedules.

Literature review

Production planning and scheduling for prefabricated and modular construction has been extensively studied using job shop and flexible job shop formulations. Early and foundational work adapted manufacturing-oriented models to construction fabrication contexts, including steel components, bridge girders, precast elements, and pipe spools, often focusing on minimizing makespan and balancing resource utilization (Hasan et al. 2019; Hu et al. 2014; Liu et al. 2017). Genetic algorithms, memetic algorithms, and heuristic methods have been commonly employed to address the computational complexity of these problems (Meng et al. 2018; Moghadam et al. 2017; Xie et al. 2023).

More recent studies have expanded these models to incorporate multi-objective optimization, distributed production, and just-in-time constraints, reflecting the increasing complexity of industrialized construction systems (Anvari et al. 2016; Flamini and Nicosia 2024; Wang et al. 2024). Data-driven and learning-based approaches have also emerged, leveraging historical production data to forecast labor hours and improve scheduling performance (ElMenshawy et al. 2025; Karim et al. 2024; Said and Kandimalla 2018).

Graph-based representations have been recognized as a powerful abstraction for scheduling problems, particularly through disjunctive graph formulations that explicitly encode precedence and resource conflicts (Błażewicz et al. 2000; Otala et al. 2021). Besides, these scheduling advancements, an increasing number of

studies have explored graph-based modeling for building design in offsite and modular construction, particularly due to its effectiveness in representing the physical topology and spatial relationships of prefabricated modules (Cao et al. 2024; Gan 2022). While these approaches offer strong theoretical foundations for both design and production independently, their application in construction has largely focused on modeling production processes rather than integrating upstream design information. Consequently, the linkage between modular design attributes and schedule generation remains weak.

The present study addresses this gap by employing graph rewriting to generatively derive task, resource, and constraint structures from a design graph, enabling automated, design-consistent job shop scheduling for modular construction.

Methodology

Graph Representation

The design of a prefabricated building module design is modeled as an undirected graph in which nodes represent rooms and walls. Room nodes indicate spaces with geometric attributes (width, length, area), while wall nodes are attributed by one of three types (internal, external, or partition wall) and are annotated with the wall's dimensional properties (height and length) and the presence of doors or windows. The edges between room and wall nodes capture boundary relationships, indicating which walls define the perimeter of each room. On the other hand, the edges between room nodes indicate the adjacency and accessibility between two rooms without the existence of a partition wall. This graph-based representation captures efficiently the room–wall topology of a module design that can be informative for production planning.

Rule Definition

This framework relies on a rule repository that bridges architectural design and production planning. To ensure the scalability of the generative approach, the framework categorizes rules into a multi-level hierarchy.

Element Level: At this granular level, rules specify the tasks and associated resources for individual components (e.g., interior walls, MEP fixtures). This ensures that every physical part of the parametric design triggers a corresponding manufacturing requirement.

Room Level: These rules capture the aggregated activities required for specific functional spaces, such as dry rooms versus wet rooms. This level introduces broader finishing tasks, including ceiling installation, flooring, and specialized commissioning.

Module Level: At the highest level of the hierarchy, the rules introduce an abstract "module node." This node associates the entire unit with holistic tasks such as structural assembly, final quality inspection, and logistics.

By organizing rules in this manner, the framework effectively deconstructs the fabricator's production expertise into a reusable Rule Library. While the initial formulation of these rules requires input from the fabricator crew to ensure the "constraint space" accurately reflects the reality of the shop floor, the resulting templates can be applied to new designs with minimal intervention. This reduces the user's burden from "creating logic" to "selecting appropriate rule sets," thereby enhancing the system's usability for non-expert project managers.

Design-to-Production Transformation using Graph Rewriting Rules

Graph rewriting, also known as graph transformation, is a rule-based approach for modifying graph structures by adding, removing, or reconfiguring nodes and edges according to predefined rules. In engineering design, it provides a structured yet flexible mechanism for automating complex transformations, encoding domain knowledge, and systematically exploring design alternatives (Kolbeck et al. 2022). Recent advances have demonstrated its effectiveness in modular and parametric design contexts, where rule-based transformations enable automated generation of design variants and support early-stage decision-making (Aksamija et al. 2010; Voss et al. 2023). Graph grammars extend this concept by offering a formal framework for design synthesis, allowing valid configurations to be generated through well-defined transformation rules while ensuring consistency and correctness through metamodels or type graphs (Campbell 2009; Schmidt and Rudolph 2016).

For the proposed design-production framework, graph rewriting rules are used to link process knowledge, including task, resources and constraints, to design graph. As such, there are two production-related nodes that are added to the design graph by the graph rewriting rules: 1) task nodes represent a discrete production activity with uninterrupted duration; and 2) resource nodes represent the production crews or equipment used to accomplish one task or more. Task assignment to each module design element (wall and room nodes) is created by adding edges "constructs" from task nodes to the design elements. Resource assignment is created by adding edges "uses" from task nodes to resource nodes representing required production capabilities. Afterwards, constraint assignment is created by adding edges "precedes" between task nodes to represent the logical sequencing between the fabrication tasks.

Graph Rule Assignment

Rule assignment is performed by iteratively matching the left-hand side (LHS) patterns of the graph rewrite rules against the module design graph. When a valid matching is found, the corresponding rule is applied by augmenting the graph with the right-hand side (RHS) components. The RHS introduces task nodes, resource dependency links, and task–task precedence relations, while preserving the original design entities and their topological relationships.

After the rule assignment, task durations are computed by associating the unit of the design entity with a predefined unit production rate stored in the task node. For example, the duration of the task "Light-gauge wall framing (studs

& door/window frames, exterior)" is calculated by multiplying the wall surface area extracted from the design graph by the corresponding production rate (hours per square meter of wall). This approach ensures the task durations are design-dependent and adapt to the design changes automatically, ensuring the consistency between design and production planning.

Scheduling Formulation

After the task assignment, the production planning is formulated as a job shop scheduling problem, where each module to be produced is modeled as a job. The input is the graph generated from the graph rewrite process, which defines the set of tasks, their precedence relationships, required resources, and design-dependent durations. Besides, two additional inputs, including the number of modules to be produced and the number of stations (resources), are defined according to order demand and factory production constraints. In traditional job shop models, stations are fixed in location. However, this proposed framework models them as moveable resources (crews) as most prefabricated volumetric modules are stationary and the fabrication crews move between them to execute the tasks. The scheduling objective is to minimize the overall makespan of the module production process.

Implementation

The proposed methodology is implemented as a Python-based computational pipeline that integrates graph modeling, task graph generation, and constraint-based scheduling. To demonstrate the methodology, a sample module design is used as a case study (Figure 1), which was obtained from a large building fabricator in Asia. In total, the analyzed module includes three exterior walls (EW), four interior walls (IW), two interior wet walls (WW), one dry room (R), and two wet rooms (WR). The differentiation between EW, IW, and WW wall elements is needed due to their different compositions and layers (partition material, insulation, existence of mechanical and electrical M&E systems, and finish material. A similar justification exists for the differentiation between dry and wet rooms (R and WR).

Wall Data

Wall	Dimension	Doors	Windows
EW1	3 x 2.12	-	1
EW2	3 x 2.51	-	1
EW3	3 x 4.63	-	2
IW1	3 x 6.88	1	-
IW2	3 x 6.88	-	-
IW3	3 x 4.10	-	-
IW4	3 x 4.10	1	-
WW1	3 x 2.12	-	-
WW2	3 x 4.10	-	-

Room Data

Room	Area
R1	4.63 x 6.88
WR1	2.12 x 4.10
WR2	2.51 x 4.10

EW = Exterior Wall
IW = Interior Wall
WW = Wet Wall
R = Room
WR = Wet Room

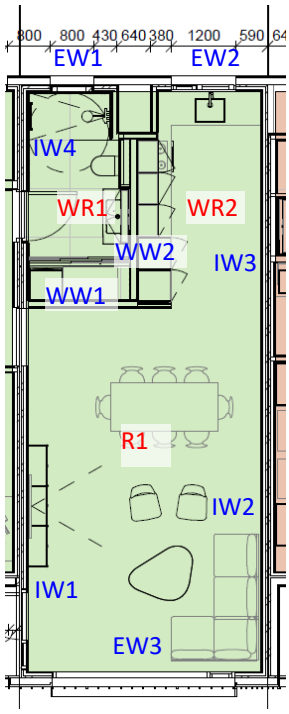


Figure 1: A sample module design

In addition to the prefabricated module design, it was critical to formulate a product-process recipe for each element in the module to facilitate the generative schedule process. Table 1 lists the product-process recipes for each element and aspect of a module, including production tasks, station/resource, unit type, duration, and precedence constraints. These recipes reflect the fabrication of the overall structural system of the module (structural steel post-beam with a concrete floor) and the fabrication tasks and resources needed to realize the different wall and room types. The recipes were formulated by the researcher by examining the production plan that were shared by the building fabricator. It should be noted that the fabricator shared a generic schedule for all their standard modules, which was useful to strategically outline the production process and resources but this generic schedule did not enable reliable scheduling and control of the production process. This observation confirms the need for this study and the proposed framework for generative scheduling of prefabricated building modules.

Table 1. Module production process

Element	ID	Production Task	Station / Resource	Unit	Duration (hrs)	Preceding Tasks
Strcutural	1	Construct the module structural steel frame & concrete for	Structural	Module	24	-
	2	Construct the module concrete floor	Structural	Module	8	1
	3	Strip the concrete formwork	Structural	Module	4	2, lag 32 hours
Exterior Wall	4	Light-gauge wall framing (studs & door/window frames, ex	Framing	m^2, wall	0.53	2, lag 16 hours
	5	Wall electrical installation	M&E	m^2, wall	0.27	21
	6	Insulation installation	Insulation	m^2, wall	0.20	5
	7	Interior gypsum board installation	Framing	m^2, wall	0.27	6
	8	Exterior water proofing	Insulation	m^2, wall	0.27	4
	9	Window & door installation and testing	Doors & Windows	Count, D&W	1.00	7,8
	10	Exterior galvanized panel installation and sealing	Framing	m^2, wall	0.53	9
Interior Regular Wall	11	Light-gauge wall framing (studs & door/window frames)	Framing	m^2, wall	2.00	21
	12	Wall electrical installation	M&E	m^2, wall	0.75	11
	13	Interior gypsum board installation	Framing	m^2, wall	0.50	12
	14	Door installation	Doors & Windows	Count, D	1.00	13
Interior Wet Wall	15	Light-gauge wall framing (studs & door/window frames)	Framing	m^2, wall	1.50	29
	16	Wall electrical installation	M&E	m^2, wall	0.75	15
	17	Wall plumbing installation	M&E	m^2, wall	1.25	15
	18	Interior gypsum board installation	Framing	m^2, wall	0.50	16, 17
	19	Wall waterproofing	Insulation	m^2, wall	0.20	18
	20	Door installation	Doors & Windows	Count, D	1.00	19
Dry Room	21	ceiling, Install mechanical and electrical	M&E	m^2, room	0.50	4
	22	ceiling, Light-gauge framing	Framing	m^2, room	2.00	21
	23	ceiling, Fireproofing	Insulation	m^2, room	0.25	22
	24	ceiling, install gypsum board	Framing	m^2, room	0.75	23
	25	Interior wall & ceiling mudding & taping	Painting	m^2, room+w	0.27	13, 24
	26	Vinyl flooring installation	Flooring	m^2, room	0.13	25
	27	Interior wall & ceiling painting	Painting	m^2, room+w	0.08	14, 26
	28	Power commissioning	M&E	room	2.00	12, 21
Wet Room	29	ceiling, Install mechanical and electrical	M&E	m^2, room	0.60	4
	30	Floor drain installation	M&E	room	3.00	29
	31	Floor waterproofing	Insulation	m^2, room	0.89	19, 30
	32	Tiling for floor and walls	Flooring	m^2, room+w	0.13	20, 31
	33	Bathroom flood testing	Insulation	room	4.00	32
	34	Cabinet installation	Carpentry	room	4.00	33
	35	Water fixture installation & commissioning	M&E	room	4.00	34
	36	Power commissioning	M&E	room	2.00	16, 29
Closeout & Logistics	37	Inspection	Closeout	Module	8.00	10, 28, 34, 36
	38	Module sealing and shippig	Closeout	Module	8.00	37

The implementation leverages *ReGraph* and *NetworkX* as the core graph modeling and transformation engines. *NetworkX* ("NetworkX" 2026) is used to construct and manage the underlying design and task graphs, providing a flexible data structure for representing nodes, edges, and associated attributes such as geometry, resources, and production parameters. *ReGraph* ("ReGraph" 2026) is employed to define and execute graph rewrite rules, enabling rule-based pattern matching and transformation of the design graph into a task and resource graph. Through left-hand-side pattern matching and right-hand-side graph augmentation, *ReGraph* supports the systematic addition of tasks, precedence relations, and resource dependencies while preserving the original design topology. This combination allows the implementation to separate graph data management from transformation logic, ensuring that design-dependent task generation is modular, extensible, and computationally efficient within the Python-based pipeline.

The task graph is then converted into a job shop scheduling problem and solved using Google OR-Tools CP-SAT ("OR-tools" 2026). The objective of the experimental evaluation is to analyze how the generative framework supports production planning through three key performance metrics: makespan, busy time, station saturation, and queue time. For the initial analysis, the production of a single module is examined under the assumption of deterministic task durations and fixed station capacities. Production station capacities are assumed based on typical factory conditions, including three framing stations, two M&E stations, two insulation stations, one doors-and-windows station, two painting stations, two flooring stations, one carpentry station, two structural workstations, and one closeout station.

Results

This section presents the outputs generated by the proposed framework. Figure 2 shows the graph representation of the sample module design. Rooms and walls are modeled as nodes, with edges capturing their topological relationships. Wall nodes store geometric properties such as dimensions and the number of doors and windows, and room nodes store the width, length and area. These geometric attributes are used to derive the duration calculation for each task.

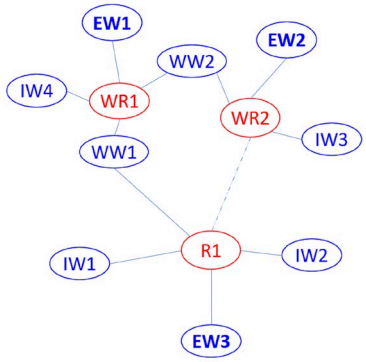


Figure 2: Graph representation of sample module design

Built upon the product-process recipes (Table 1), graph rewrite rules are applied to transform the design graph into a task graph, as illustrated in Figure 3. The resulting task graph represents design entities (blue circular nodes), production tasks (green squares) and resources/stations (orange diamonds). Solid edges represent task–entity and task–resource associations, while dashed edges indicate precedence constraints between tasks.

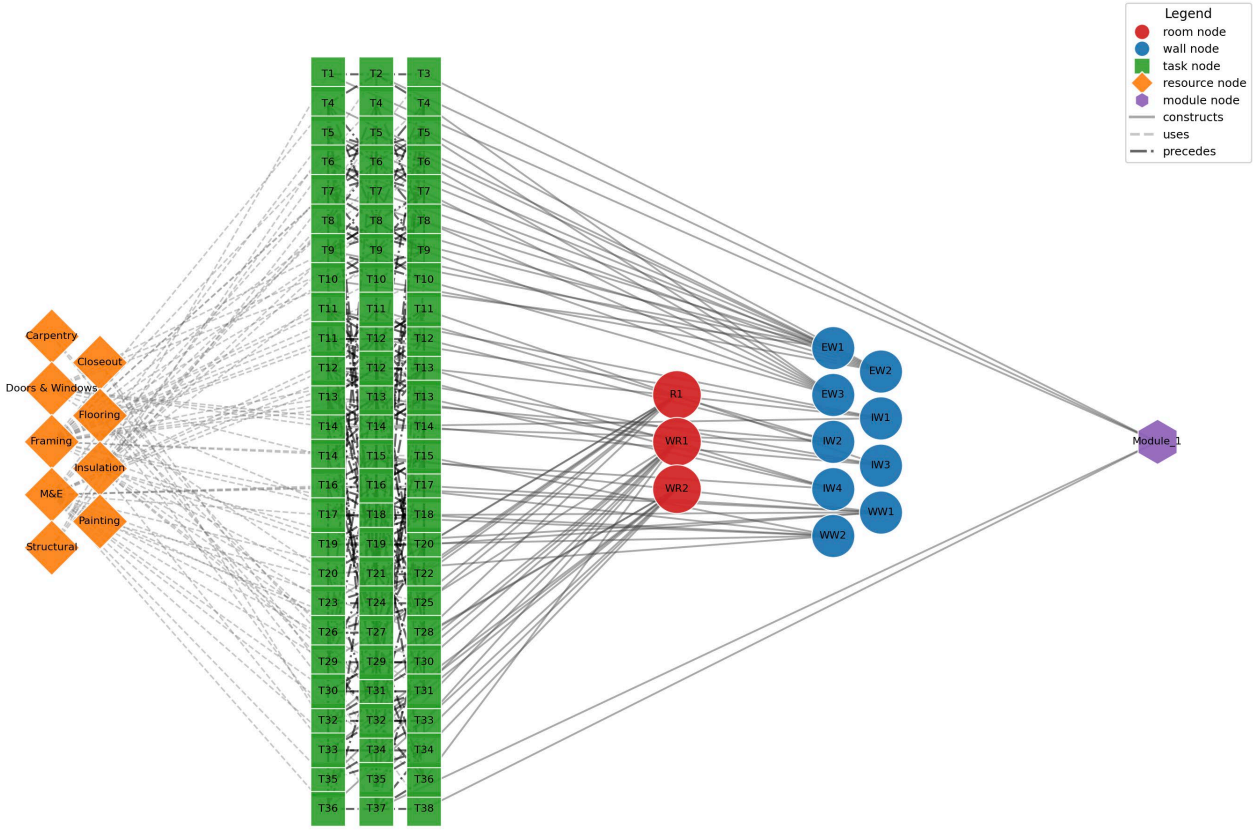


Figure 3: Task graph representation

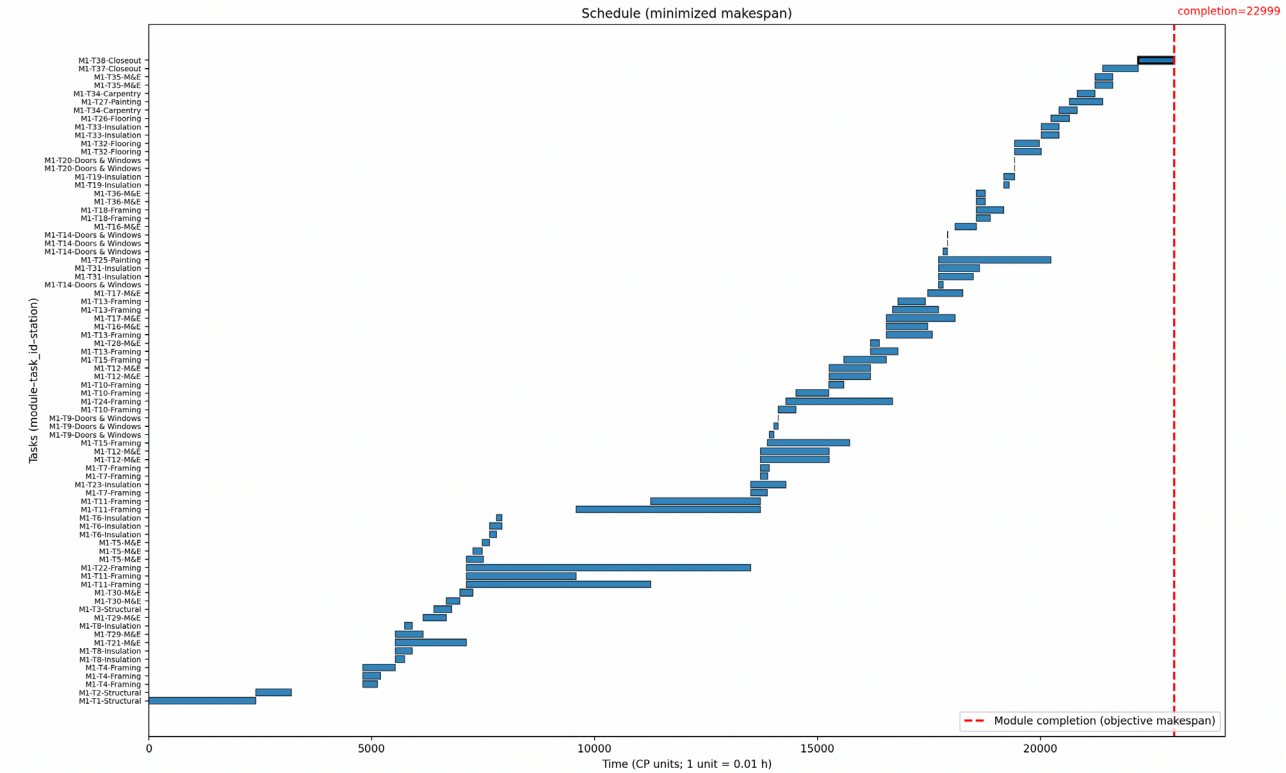


Figure 4: Optimized job shop schedule for modular production

The optimized schedule is presented as a Gantt chart (Figure 4) obtained by solving the job shop scheduling problem derived from the task graph. The vertical axis lists individual tasks, while each bar represents the corresponding task start and finish times. The dashed vertical line indicates the optimized makespan (230 hours), which corresponds to the module completion time.

Discussion

The results demonstrate the ability of the proposed graph-based generative scheduling framework to automatically translate a parametric modular design into a feasible and optimized production schedule. Beyond the single-module demonstration, additional experiments were conducted to derive important insights for the system scalability, resource bottlenecks, and the interaction between design-driven task generation and factory production constraints.

In contrast, professionals are enabled by the proposed generative framework to be more efficient by automating the mapping of architectural topology to scheduling constraints, the system eliminates manual "linking" errors, ensuring that the generated schedule is logically consistent and adheres strictly to design dependencies.

Expanding the analysis beyond the scheduling of a single module revealed the scalability limitation and the diminishing marginal efficiency of the modeled production system. Table 2 shows the resulting total makespan and the average module makespan for fabricating different numbers of modules. The multi-module experiments reveal a clear non-linear relationship between production volume and total makespan. While increasing the number of modules leads to economies of repetition in early stages, the average completion time per module eventually increases as shared resources become saturated. This diminishing marginal efficiency is a well-known phenomenon in prefabrication environments, where parallelism is constrained by finite station capacities and reentrant workflows. The observed reduction in average module time from three to approximately twelve modules indicates improved utilization of parallel stations, whereas the subsequent increase reflects congestion effects and resource contention. Importantly, these dynamics emerge naturally from the generated task and resource graph without manual adjustment of precedence logic or capacity rules. This highlights a key advantage of the proposed approach over traditional scheduling models that rely on static, predefined workflows. Because task durations and precedence relationships are derived directly from the design graph and rule assignments, the scheduling model remains consistent as production volume changes, allowing emergent system behavior to be analyzed rather than imposed.

Table 2: The resulting total makespan and average module makespan under different number of modules

Number of Modules	Total Makespan (hr)	Average Module Makespan (hr)
1	230.00	230.00
3	413.65	137.88
5	634.23	126.85
10	1,180.79	118.08
11	1,293.16	117.56
12	1,402.08	116.84
13	1,650.35	126.95
14	1,802.5	128.75
15	1,960.00	130.67
20	2,655.12	132.76

The station-level performance analysis further illustrates the explanatory power of the graph-based formulation. The framework was utilized to analyze the resource utilization and bottlenecks under the scenario of fabricating 20 modules. Three metrics were tracked for each station: 1) busy time: measures total processing workload executed on station; 2) saturation percentage: represents the fraction of time during which the station operates at full capacity, indicating how frequently the station constrains task execution; and 3) queue time: measures the cumulative waiting time of tasks assigned to a station before processing, capturing upstream congestion effects. Tables 3 lists the values of these metrics and the capacities for three sample stations: framing, mechanical and electrical, and insulation. Framing stations exhibited the highest busy time and saturation percentage, confirming their role as the primary bottleneck in the examined production configuration. This finding is consistent with practical observations in modular factories, where framing activities often dominate early production stages and constrain downstream flow. In contrast, insulation and finishing-related stations showed relatively low saturation, suggesting underutilized capacity and potential opportunities for rebalancing or task reallocation. Queue time metrics provided additional insight into upstream congestion effects that are difficult to capture using aggregate performance indicators alone. High queue times at framing and M&E stations indicate that delays propagate through the system not merely due to task durations, but because of structural dependencies encoded in the task graph. From a managerial perspective, this suggests that capacity expansion decisions should prioritize stations with both high saturation and high queue accumulation, rather than focusing solely on busy time.

Table 3: The station performance analysis results

Station	Capacity	Busy Time	Saturation %	Queue time
Framing	3	6531.2	81.3	255570.4
M&E	2	2830.6	53.0	46339.9
Insulation	2	993.0	17.6	6247.52

Conclusion

This study illustrates how graph rewriting can serve as an effective bridge between design representation and production scheduling. By separating design topology, process knowledge, and scheduling constraints into distinct but interoperable graph layers, the approach maintains both flexibility and interpretability. The use of constraint programming further ensures that the generated schedules are globally optimized with respect to the defined objective, rather than relying on heuristic sequencing alone. More broadly, the results suggest that graph-based generative scheduling can function not only as a planning tool, but also as an analytical instrument for understanding the systemic consequences of design and capacity decisions in industrialized construction.

A key contribution of this study lies in demonstrating how tight coupling between design and production can be operationalized rather than abstracted away. Because task generation and duration calculation are driven by geometric attributes extracted from the design graph, even minor design changes, such as wall lengths, room configurations, or the number of openings—would propagate automatically into the production schedule. This capability contrasts with most existing job shop scheduling approaches in construction, which assume fixed task sets and rely on manual updates when designs evolve. The experimental results suggest that such coupling is particularly valuable in early-stage design evaluation and production planning. Designers and production planners can use the framework to explore "what-if" scenarios, assessing how design decisions influence makespan, station saturation, and bottleneck severity before committing to detailed fabrication plans. In this sense, the framework supports a shift from reactive scheduling toward proactive, design-informed production optimization.

Nevertheless, the current implementation also has limitations. The experiments assume deterministic task durations and station capacities, whereas real-world factory environments are subject to stochastic disruptions, learning effects, and workforce variability. In addition, the case study focuses on a single module type, and while the framework is generalizable, heterogeneous module families and mixed-model production were not explicitly examined. These factors may amplify congestion effects and introduce additional scheduling complexity.

The findings from the extended experiments point to several promising avenues for future research. First, incorporating stochastic task durations and probabilistic disruptions would allow the framework to support robustness and resilience analysis. Second, extending the graph rewrite rules to handle multiple module typologies within a single production line would enable mixed-model scheduling studies. Third, coupling the scheduling outputs with cost, labor, or energy metrics could support multi-objective optimization and decision-making at the factory planning level. Fourth, the framework successfully transforms a parametric module design into a feasible schedule without manual intervention. Future work can leverage this capability to identify design features that have the greatest impact on production performance.

References

Aksamija, A., K. Yue, H. Kim, F. Grobler, and R. Krishnamurti. 2010. "Integration of knowledge-based and generative systems for building characterization and prediction." Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 24 (1): 3–16. https://doi.org/10.1017/S0890060409990138.

Anvari, B., P. Angeloudis, and W. Y. Ochieng. 2016. "A multi-objective GA-based optimisation for holistic Manufacturing, transportation and Assembly of precast construction." Automation in Construction, 71: 226–241. https://doi.org/10.1016/j.autcon.2016.08.007.

Ballard, G., and G. Howell. 1998. "What Kind of Production Is Construction?"

Błażewicz, J., E. Pesch, and M. Sterna. 2000. "The disjunctive graph machine representation of the job shop scheduling problem." European Journal of Operational Research, 127 (2): 317–331. https://doi.org/10.1016/S0377-2217(99)00486-5.

Campbell, M. 2009. "A Graph Grammar Methodology for Generative Systems." 25.

Cao, J., H. Said, A. Savov, and D. Hall. 2024. "Graph-Based Evolutionary Search for Optimal Hybrid Modularization of Building Construction Projects." Journal of Construction Engineering and Management, 150 (8): 04024098. American Society of Civil Engineers. https://doi.org/10.1061/JCEMD4.COENG-14687.

ElMenshawy, M., L. Wu, B. Gue, and S. AbouRizk. 2025. "Automating Pipe Spool Fabrication Shop Scheduling for Modularized Industrial Construction Projects Using Reinforcement Learning." Journal of Computing in Civil Engineering, 39 (3): 04025013. American Society of Civil Engineers. https://doi.org/10.1061/JCCEE5.CPENG-6042.

Flamini, M., and G. Nicosia. 2024. "A generalized disjunctive graph model for a complex production problem." Procedia Computer Science, International Conference on Industry Sciences and Computer Science Innovation, 237: 289–296. https://doi.org/10.1016/j.procs.2024.05.107.

Gan, V. J. L. 2022. "BIM-based graph data model for automatic generative design of modular buildings."

Automation in Construction, 134: 104062. https://doi.org/10.1016/j.autcon.2021.104062.

Gibb, A., and F. Isack. 2003. “Re-engineering through pre-assembly: client expectations and drivers.” Building Research & Information, 31 (2): 146–160. Routledge. https://doi.org/10.1080/09613210302000.

Hasan, M., M. Lu, and K. Bird. 2019. Planning and scheduling bridge girders fabrication through shop-floor operations simulation.

Hu, X., M. Lu, and S. AbouRizk. 2014. “BIM-based data mining approach to estimating job man-hour requirements in structural steel fabrication.” In: Proc. Winter Simul. Conf. 2014, 3399–3410.

Karim, M. R., B. Gue, L. Wu, and Y. Mohamed. 2024. “Job Labour-Hours Forecasting for a Pipe Spool Fabrication Shop Using Data Mining.” In: Proc. Can. Soc. Civ. Eng. Annu. Conf. 2023 Vol. 5, edited by S. Desjardins, G. J. Poitras, and M. Nik-Bakht, 437–449. Cham: Springer Nature Switzerland.

Kolbeck, L., S. Vilgertshofer, J. Abualdenien, and A. Borrmann. 2022. “Graph Rewriting Techniques in Engineering Design.” Frontiers in Built Environment, 7: 815153. https://doi.org/10.3389/fbuil.2021.815153.

Liu, J., M. Soleimanifar, and M. Lu. 2017. “Resource-loaded piping spool fabrication scheduling: material-supply-driven optimization.” Visualization in Engineering, 5 (1): 5. https://doi.org/10.1186/s40327-017-0044-3.

Meng, R., Y. Rao, Y. Zheng, and D. Qi. 2018. “Modelling and solving algorithm for two-stage scheduling of construction component manufacturing with machining and welding process.” International Journal of Production Research, 56 (19): 6378–6390. https://doi.org/10.1080/00207543.2017.1349949.

Moghadam, A., K. Y. Wong, and H. Piroozfard. 2017. “Solving a hybrid jobshop scheduling problem with space constraints and reentrant processesby using an improved hybrid genetic algorithm: A case study.” International Journal of Industrial Engineering : Theory Applications and Practice, 24: 483–504.

“NnetworkX.” 2026. Python. NetworkX.

“OR-tools.” 2026. C++. Google.

Otala, J., A. Minard, G. Madraki, S. Mousavian, J. Otala, A. Minard, G. Madraki, and S. Mousavian. 2021. “Graph-Based Modeling in Shop Scheduling Problems: Review and Extensions.” Applied Sciences, 11 (11). publisher. https://doi.org/10.3390/app11114741.

“ReGraph.” 2026. Python. Kappa-Dev.

Safarzadeh, S., S. Shadrokh, and A. Salehian. 2018. “A heuristic scheduling method for the pipe-spool fabrication process.” Journal of Ambient Intelligence and Humanized Computing, 9 (6): 1901–1918. https://doi.org/10.1007/s12652-018-0737-z.

Said, H. M., and P. Kandimalla. 2018. “Performance Measurement of Building Sheet-Metal Ductwork Prefabrication under Batch Production Settings.” Journal of Construction Engineering and Management, 144 (2): 04017107. American Society of Civil Engineers. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001423.

Schmidt, J., and S. Rudolph. 2016. “Graph-Based Design Languages: A Lingua Franca for Product Design Including Abstract Geometry.” IEEE Computer Graphics and Applications, 36 (5): 88–93. https://doi.org/10.1109/MCG.2016.89.

Voss, C., F. Petzold, and S. Rudolph. 2023. “Graph transformation in engineering design: an overview of the last decade.” AI EDAM, 37: e5. Cambridge University Press. https://doi.org/10.1017/S089006042200018X.

Wang, S., X. Li, L. Gao, and J. Li. 2024. “A multi-disjunctive-graph model-based memetic algorithm for the distributed job shop scheduling problem.” Advanced Engineering Informatics, 60: 102401. https://doi.org/10.1016/j.aei.2024.102401.

Xie, Y., H. Wang, G. Liu, and H. Lu. 2023. “Just-in-Time Precast Production Scheduling Using Dominance Rule-Based Genetic Algorithm.” IEEE Transactions on Neural Networks and Learning Systems, 34 (9): 5283–5297. https://doi.org/10.1109/TNNLS.2022.3217318.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

INTEGRATED MANY-OBJECTIVE OPTIMIZATION FOR DECISION-MAKING AND DYNAMIC PROJECT CONTROL

Aman Shukla, and Sahil Garg
Indian Institute of Technology Delhi, New Delhi, India

Abstract

Construction projects involve conflicting objectives beyond time and cost, including quality, safety risk, resource stability, and environmental impact. Many-objective optimization (MaOO) helps address this complexity in planning, but it provides limited support for decision-making during execution, and selecting one solution from Pareto-sets often depends on human judgment. To overcome these gaps, this paper proposes an integrated planning and control framework combining Opposition-Based NSGA-III (OBNSGA-III) for six-objective optimization, an Entropy-based-VIKOR model for compromise solution selection, and a Dynamic-MaOO re-optimization mechanism for execution control. A case study based on 25-activities is used to demonstrate effective decision support from planning to execution.

Introduction

Construction projects operate as short-term complex production systems that involve multiple stakeholders with different and competing objectives (Ward et al. 1991). Construction industry is one of the rapidly growing industries in developing nations that supports socio-economic development by contributing infrastructure growth and employment generation. (Fu & Zhang, 2016). Construction projects involve large number of individual but interlinked activities. (Singh and Ernst, 2011). Traditional construction projects focus only on optimization of Time and cost. But Construction projects involve large number of objectives such as Quality, safety risk, resource moment, and environmental impact due to construction. These are important objectives for stakeholders and have direct impact on the project.

Traditional multi-objective algorithms struggle to maintain solution diversity and convergence to optimize more than three objectives. Thus, simultaneously optimization of these objectives requires many objective optimizations model (MaOO). NSGA-III and its variants address this issue through reference-point-based selection mechanisms, enabling efficient search in high-dimensional objective spaces.

MaOO model gives pareto optimal solution sets. Decision makers often use subjective weighted-sum methods to select one solution from the Pareto solution sets for the project. This includes human factor and limiting reproducibility. Current construction MaOO research, optimize the project objectives during planning phase assuming that selected solution will remain valid during execution phase.

The proposed framework integrates Opposition-Based NSGA-III (OBNSGA-III) model, an Entropy-based-VIKOR model, and Dynamic-MaOO re-optimization mechanism. This integrated framework contributes as a new methodology in selection of single solution set from sets pareto-optimal solution and plays a key role in project control during execution phase as it moves beyond one-time optimization and supports planning as a continuous process and enables both initial decision-making and execution-stage adaptation.

Literature Review

Early research in construction projects optimization was largely focused on the optimization of time–cost and finding the relationship between project duration and direct cost. Abbasnia et al. (2008) applied fuzzy logic model to find uncertainty in time–cost relationships, while Zareei and Hassan-Pour (2015) introduced a resource-constrained formulation to improve the schedule feasibility. These studies showed that deterministic planning approaches are insufficient for realistic construction decision making. As research progressed, it is found that considering only two objectives are not sufficient, particularly in projects where schedule acceleration adversely affected other performance dimensions.

To address this, quality was introduced as a third optimization objective. El-Rayes and Kandil (2005) demonstrated that minimization of time and cost alone can significantly degrade the construction quality and used an integrated time–cost–quality optimization model. The emergence of evolutionary algorithms further enabled the treatment of such multi-objective problems. Deb et al. (2002) proposed a NSGA-II algorithm. This model was widely adopted in construction scheduling as it provides elitist selection, and it has ability to generate diverse Pareto-optimal solutions. Subsequent studies applied genetic algorithms and particle swarm optimization to incorporate additional constraints and objectives, including safety risk and resource utilization (Afshar, A., and Dolabi, 2014; Elbeltagi, E., 2016).

As construction performance evaluation expanded, researchers have studied the effects of optimization of

safety and environmental impact with time and cost. Afshar, A., and Dolabi (2014) showed that time–cost-optimized schedules often lead to higher safety risks and came with a time-cost-safety risk optimization model, while Ozcan-Deniz et al. (2012) demonstrated a significant trade-off between productivity and environmental impact. Further, multi-mode, resource-constrained formulations highlighted the conflict between stability of execution and traditional performance measures of project (Ghoddousi, P., et al., 2013; Kim, K., et al., 2015). At this stage, most models address three objectives, typically combining time, cost and either safety or environmental criteria.

However, as the number of objectives increased beyond three, the conventional dominance-based algorithms such as NSGA-II, the effectiveness and diversity of selected solutions decreased. To overcome this, Deb and Jain (2014) proposed a reference-based NSGA-III algorithm. This algorithm uses a reference-point-based non-dominated sorting to maintain diversity in high-dimensional objective spaces. Development of this method marked a transition from multi-objective optimization to many-objective optimization in construction research. Arya et al. (2024) and Behera et al. (2025) applied NSGA-III to balance time, cost, quality, and sustainability-related objectives, while Zhan et al. (2024) demonstrated its effectiveness in generating well-distributed Pareto fronts for construction project management problems.

Recent studies have explicitly acknowledged the need for five-objective optimization frameworks. Panwar and Jha (2019) formulated a many-objective construction scheduling model integrating time, cost, resources and environment impact and Sharma and Trivedi (2022) integrated time, cost, quality, safety, and resource constraints using NSGA-III-based approaches. Panthi and Hussain (2025) further extended this line of research by proposing an OBNSGA-III framework incorporating time, cost, quality, safety, and environmental sustainability. Similarly, Sharma and Trivedi (2023) modelled trade-offs among time, cost, quality, safety risk, and environmental impact using opposition-based NSGA-III. These studies collectively demonstrate that most existing construction MaOO models converge at five objectives, reflecting both computational complexity and modelling challenges.

Despite of these advances, construction research optimization models are rarely used to optimize six-objectives simultaneously. In particular, the simultaneous integration of time, cost, quality, safety risk, environmental impact, and resource moment within a single optimization framework has received limited attention. Resource moment objectives are often treated separately from safety or environmental criteria, even though all three directly influence production stability and execution reliability. The absence of six-objective formulations restricts the ability of planners to evaluate comprehensive trade-offs inherent in complex construction systems.

Another limitation in literature is the decision-making to select one solution from Pareto-optimal sets. Although NSGA-III-based models generate diverse solution fronts, most studies rely on subjective weighting or visual inspection to select final execution plans. Zhan et al. (2024) explicitly noted that post-optimization decision selection remains weak in construction optimization research. While entropy-based weighting and compromise ranking methods such as VIKOR have been suggested to reduce subjectivity, their systematic integration with many-objective optimization models is still limited.

The majority of existing many-objective construction optimization studies are static and focusing exclusively on the planning stage only. Once project execution starts, the deviations in objective values are typically managed through heuristic rescheduling not from the original optimization logic.

In summary, the literature reveals three interrelated gaps: (1) most construction optimization studies are limited to five objectives, (2) decision-making from many-objective Pareto fronts remains weak and often subjective, and (3) optimization frameworks rarely extend into execution-stage control. The present research addresses these gaps by formulating a six-objective many-objective optimization model, integrating objective decision selection using Entropy–VIKOR, and embedding the model within a dynamic planning-and-control framework suitable for construction production systems. re-optimization.

From a lean construction point of view, construction projects are similar to production systems where planning and control must be continuously aligned to manage variability and improve the workflow reliability. Foundational lean construction studies emphasize the importance of reduction of uncertainty from the production and adapting plans that are based on execution feedback rather than relying on static schedules (Koskela, 1992; Ballard and Howell, 1998). The existing optimization literature remains focused only on planning-stage solutions. The proposed framework responds to this gap by supporting adaptive decision-making during execution phase of project.

Methodology

This research adopts a planning–control perspective to address construction projects as dynamic production systems rather than static optimization problems. The proposed methodology integrates three complementary components:

1. Many objective planning by using OBNSGA-III
2. Objective compromise decision selection using Entropy–VIKOR
3. Dynamic re-optimization during execution.

The proposed framework integrates optimization, decision selection, and re-optimization into a single workflow. The study shows how these can be used together to support planning and control decisions in multimode construction projects.

This model integrates structure existing optimization and decision-making tools into a coherent framework that supports both planning and execution-stage control. This approach aligns with lean construction principles, where plans are continuously optimized and adjusted based on execution feedback instead of being fixed at the outset. This structure explicitly links planning and control, addressing a major limitation of static construction optimization models.

In this study, the VIKOR method is employed to support compromise decision-making under conflicting criteria. The method determines a compromise solution by maximizing group utility and minimizing individual regret, thereby balancing both overall performance and the worst-case criterion. Other techniques, such as TOPSIS, rely on distances to ideal and negative-ideal solutions without considering their relative importance, which may affect the ranking results in certain decision situations (Opricovic, S., and Tzeng, G. H., 2004). As a result, VIKOR provides a more reliable and decision-relevant solution, particularly in situations where poor performance in a critical criterion cannot be ignored.

Many-Objective Optimization Model

Decision Variables and Project Representation

The construction project is modelled as a **precedence-constrained activity network** consisting of n activities. Each activity A has multiple execution modes m, where each mode represents a distinct combination of:

- Activity completion time (ACT_A)
- Activity completion cost (labour, material, equipment) (ACC_A)
- Activity Quality index (AQI_A)
- Activity Safety risk (ASR_A)
- Activity Environmental impact (AEI_A)
- Activity Resource moment (R_t^k)
- A candidate solution is defined as a vector of execution modes, one for each activity.

Objective Function Formulation

The model simultaneously optimizes six project-level objectives, the formulation of objectives are as below:

Objective 1: Minimize Project Completion Time (PCT)

$$PCT = \sum_{A \in CP} A\,CT_A \tag{1}$$

Where, CP denotes the set of activities on the critical path.

Objective 2: Minimize Project Completion Cost (PCC)

$$PCC = \sum_{A=1}^{n} D\,C_A + (IDC \times PCT) \tag{2}$$

Where, DC_A is the direct cost of activity A and IDC is the indirect cost per day.

Objective 3: Maximize Project Quality Index (PQI)

$$PQI = \frac{1}{n}\sum_{A=1}^{n} A\,QI_A \tag{3}$$

Where, AQI_A represents the quality achieved during first-time execution of activity A.

Objective 4: Minimize Project Safety Risk (PSR)

$$PSR = \sum_{A=1}^{n} A\,SR_A \tag{4}$$

Where, ASR_A aggregates safety risks related to labor injury, material wastage, and equipment failure for activity A.

Objective 5: Minimize Project Environmental Impact (PEI)

$$PEI = \sum_{A=1}^{n} A\,EI_A \tag{5}$$

Where, AEI_A is the CO_2-equivalent emission associated with activity A.

Objective 6: Minimize Project Resource Moment (PRM)

$$PRM = \sum_{A=1}^{n} A\,RM_A \tag{6}$$

here, $ARM_A = M_x + M_y$ $M_x = \sum(R_t^k)^2$, $M_y = \sum(R_t^k \cdot t)$

This formulation penalizes resource fluctuation and uneven resource distribution over time.

OBNSGA-III Optimization Procedure

The many-objective optimization problem is solved using Opposition-Based NSGA-III (OBNSGA-III).

The algorithm proceeds as follows:

1. Generate an initial random population of size N based on execution modes.
2. Generate an opposition-based population by computing opposite solutions and selecting the fittest N individuals.
3. Generate structured reference points using: $H = \binom{N+p-1}{p}$; where $N = 6$ objectives and $p = 6$ divisions per objective.
4. Apply simulated binary crossover (SBX) and polynomial mutation to generate offspring. And combine parent and offspring populations and perform non-dominated sorting.
5. Normalize objective values and associate solutions with reference directions.
6. Apply niche-count-based selection using perpendicular distance to reference points.
7. Perform opposition-based generation jumping.
8. Repeat until convergence criteria are met.

This process yields a diverse Pareto-optimal solution set without requiring subjective preferences during optimization.

Entropy–VIKOR Decision-Making Model

To identify a single actionable solution from the Pareto sets, an **Entropy-based VIKOR** model is employed.

Formulation of Entropy–VIKOR Decision-Making Model

Let the Pareto-optimal solution set obtained from OBNSGA-III be denoted as:

$$\mathcal{P} = \{P_1, P_2, \dots, P_m\}$$

Each solution P_i is evaluated against M objectives, forming a decision matrix F:

$$\mathrm{F} = [f_{ij}], \quad i = 1,2, \dots, m; \; j = 1,2, \dots, M$$

where f_{ij} represents the value of the j^{th} objective for the i^{th} Pareto solution.

Entropy-Based Objective Weighting

To eliminate subjective bias in assigning objective importance, entropy theory is applied.

Step 1: Normalization of Decision Matrix

Since the objectives are measured in different units and possess different optimization directions, the decision matrix must be normalized.

For each objective j, the normalized performance value p_{ij} is computed as:

$$p_{ij} = \frac{f_{ij}}{\sum_{i=1}^{m} f_{ij}} \tag{7}$$

Step 2: Entropy Calculation

The entropy value E_j of the j^{th} objective is calculated as:

$$E_j = -\frac{1}{ln(m)} \sum_{i=1}^{m} p_{ij} \, ln(p_{ij}) \tag{8}$$

The entropy value reflects the degree of dispersion of information associated with an objective across Pareto solutions.

Step 3: Objective Weights

The entropy-based weight w_j of the j^{th} objective is then obtained as:

$$w_j = \frac{1 - E_j}{\sum_{j=1}^{M} 1 - E_j} \tag{9}$$

Objectives exhibiting greater variability across Pareto solutions are thus assigned to higher importance.

VIKOR Compromise Ranking

The VIKOR method is employed to rank Pareto-optimal solutions and identify a compromise solution.

Step *1*: Determination of Best and Worst Values

For each objective j, the best and worst values are defined as:

$$f_j^* = \begin{cases} \min\limits_i f_{ij}, & \text{for cost-type objectives} \\ \max\limits_i f_{ij}, & \text{for benefit-type objectives} \end{cases}$$

$$f_j^- = \begin{cases} \max\limits_i f_{ij}, & \text{for cost-type objectives} \\ \min\limits_i f_{ij}, & \text{for benefit-type objectives} \end{cases}$$

Step 1: Group Utility and Individual Regret Measures

The **group utility measure** for solution i is computed as:

$$S_i = \sum_{j=1}^{M} w_j \frac{f_j^* - f_{ij}}{f_j^* - f_j^-} \tag{10}$$

The **individual regret measure** is defined as:

$$R_i = \max_j \left[w_j \frac{f_j^* - f_{ij}}{f_j^* - f_j^-} \right] \tag{11}$$

Step 2: VIKOR Compromise Index

The compromise ranking index Q_i is calculated as:

$$Q_i = v \frac{S_i - S^*}{S^- - S^*} + (1 - v) \frac{R_i - R^*}{R^- - R^*} \tag{12}$$

where:

$S^* = min_i S_i$, $S^- = max_i S_i$

$R^* = min_i R_i$, $R^- = max_i R_i$

v represents the weight of group utility (taken as $v = 0.5$ for this study)

The Pareto solution with minimum Q_i, satisfying acceptable advantage and stability conditions, is selected as the **baseline execution plan**.

Dynamic Many-Objective Optimization Model for project control

To address execution-stage uncertainty, a Dynamic MaOO re-optimization model is introduced. The dynamic MaOO model extends the static planning optimization into the execution phase by incorporating real-time project progress.

At each decision point

- Lock the completed activities and remove from the decision space.
- Partially completed activities contribute only to the remaining work.
- Remaining activities are re-optimized using OBNSGA-III.
- Normalization of new Pareto solutions is done along with the planned solution
- In the normalized plane, calculated the distance between re-optimized solutions.
- The solution having minimum deviation from the planned solution is selected to continue the execution.

This mechanism ensures adaptability while preserving consistency with the original compromise plan.

During project execution, the model incorporates actual obtained objective values for completed activities. If deviations are observed between planned and achieved values, the optimization is re-executed using the updated data. The model then recommends revised execution modes for the remaining activities, ensuring alignment with the original project objectives with minimum deviation from the planned objective values. The model works on OBNSGA-III algorithm with updated input data obtained during execution.

Model Flow Chart is shown in Figure-1.

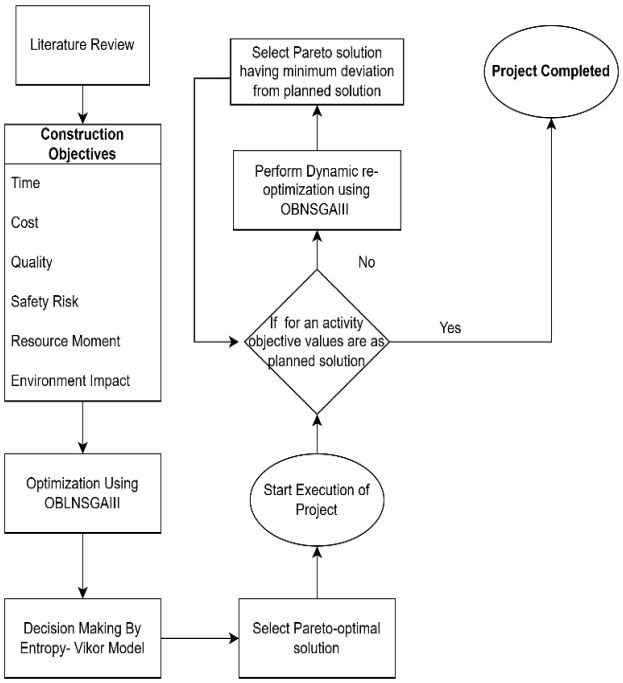


Figure 1: Flow Chart of Proposed Model

Case Study and Result

The case study involves 25 activities with various execution modes, yielding a combinatorial search space of over 53 million alternatives related to a construction project. Optimization of this project is done on OBNSGAIII model, and 78 Nos of pareto-optimal solutions are obtained. The Multi-Objective Trade-off based on normalized values is shown below as Figure 2.

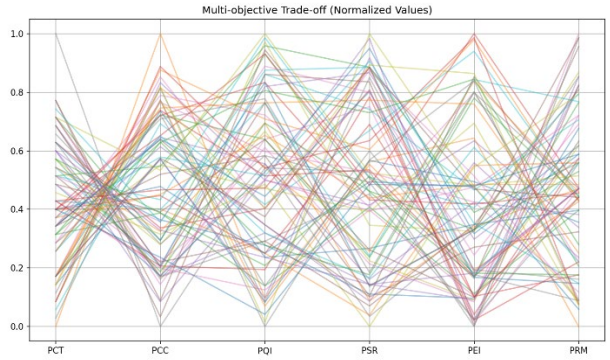


Figure 2 : Multi-Objective Trade-off based on normalized values

After this the multi-criteria decision making was done based on Entropy based VIKOR model to select a solution set for the project execution.

Results of Entropy weight calculated from model for objectives are: PCT: 0.166, PCC: 0.134, PQI: 0.177, PSR: 0.210, PEI:0.160, PRM:0.153. And selected solution set after Entropy VIKOR model- Pareto Index-23, selected modes of execution for 25 activities (A1-A2-…-A25) are: 1-2-1-1-1-1-3-3-2-3-3-1-1-2-1-2-2-1-2-2-1-1-2-1-1 and project objective values are: PCT: 243 days, PCC: 723900 Rs., PQI: 0.86, PSR: 75.22, PEI: 70.92, PRM 1855.

This Pareto solution is used for execution of project. After completion of first two activities, it is found that objective values are deviated as per planned due to resource unavailability at project site. So, dynamic re-optimization model is used to re-optimize the project and 38 pareto solutions are found. The solution having minimum deviation from planned solution, modes are used for re-execution of project.

The achieved values of objectives and deviation from the planned solution are shown in Table-1.

Table 1: Planned vs Achieved Project Performance

Objective	Planned	Achieved	Deviation (%)
PCT	243	230	-5.35
PCC	723900	815700	12.68
PQI	0.8564	0.8676	1.31
PSR	75.22	74.435	-1.04
PEI	70.92	82.295	16.04
PRM	1855	1995	7.55

The dynamic re-optimization results illustrate how planning decisions can be revised in response to resource unavailability without discarding the original planning rationale. This model is similar to lean construction thinking, which emphasizes maintaining plan reliability while accommodating inevitable execution variability through continuous control rather than complete re-planning.

Conclusion and Future Work

This study proposed an integrated many-objective framework for construction planning and control by combining OBNSGA-III optimization, Entropy–VIKOR decision making, and a dynamic re-optimization mechanism. Unlike most existing studies that limit analysis to five objectives and static planning, the proposed approach explicitly considers six project-level objectives—time, cost, quality, safety risk, environmental impact, and resource moment thereby providing a more complete representation of construction production behaviour.

From the case study we can say that OBNSGA-III can effectively manage a large combinatorial decision space and generate a diverse set of Pareto-optimal solutions. However, the results also confirmed that Pareto optimization alone is insufficient for practical decision-making. The Entropy–VIKOR model enabled objective selection of a balanced compromise solution without relying on subjective weighting, producing an execution plan with acceptable performance across all six objectives.

The dynamic re-optimization results further showed that construction plans must adapt during execution. When deviations occurred due to resource unavailability, the proposed dynamic model successfully re-optimized the remaining activities and identified revised execution

modes with minimum deviation from the original plan. This demonstrates that many-objective optimization can function as a control-oriented decision-support tool, rather than a one-time planning exercise.

The explicit inclusion of resource moment as a sixth objective proved critical, as it revealed differences in execution stability that were not visible through time–cost–quality–safety–environment trade-offs alone. Overall, we can say that the proposed framework supports more reliable planning and controlled adaptation during execution of project, and the framework aligns well with lean construction principles.

From a lean construction point of view, the proposed framework contributes to planning–control integration by enabling structured adaptation of production decisions in response to variability, supporting more reliable execution without relying on rigid or static plans.

Future work can also be done on real-time data integration through BIM-enabled progress tracking and site data acquisition systems. By linking this proposed optimization model with live data streams would allow automatic updating of activity status, resource usage, and performance indicators, transforming the framework into a decision-support system that operates throughout the project lifecycle.

References

Abbasnia, R., Afshar, A., and Eshtehardian, E. (2008). Time–cost trade-off problem in construction project management based on fuzzy logic. Journal of Applied Sciences, 8(22), 4159–4165.

Afshar, A., and Dolabi, H. R. Z. (2014). Multi-objective optimization of time–cost–safety using genetic algorithms. International Journal of Optimization in Civil Engineering, 4(4), 433–450.

Arya, A., et al. (2024). NSGA-III-based optimization model for balancing time, cost, and quality in resource-constrained retrofitting projects. Asian Journal of Civil Engineering, 25, 5613–5625.

Ballard, G., and Howell, G. (1998). Shielding production: An essential step in production control.

Behera, A. P., et al. (2025). Optimizing the trade-off between time, cost, and carbon emissions in construction using NSGA-III: An integrated approach for sustainable development. Asian Journal of Civil Engineering, 26, 73–87.

Deb, K., and Jain, H. (2014). An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach: Part I—Solving problems with box constraints. IEEE Transactions on Evolutionary Computation, 18(4), 577–601.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2), 182–197.

Elbeltagi, E. (2016). Overall multiobjective optimization of construction project scheduling using particle swarm optimization. Engineering, Construction and Architectural Management, 23(3), 265–282.

El-Rayes, K., and Kandil, A. (2005). Time–cost–quality trade-off analysis for highway construction. Journal of Construction Engineering and Management, ASCE, 131(4), 477–486.

Fu, F., and Zhang, T. (2016). A new model for solving time-cost-quality trade-off problems in construction. PLoS ONE, 11(12), e0167142.

Ghoddousi, P., et al. (2013). Multi-mode resource-constrained discrete time–cost–resource optimization in project scheduling using non-dominated sorting genetic algorithm. Automation in Construction, 30, 216–227.

Kim, K., et al. (2015). Multi-objective construction schedule optimization using modified niched Pareto genetic algorithm. Journal of Management in Engineering, 32(2), 04015038.

Koskela, L. (1992). Application of the New Production Philosophy to Construction.

Opricovic, S., and Tzeng, G. H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European Journal of Operational Research, 156(2), 445–455.

Ozcan-Deniz, G., Zhu, Y., and Ceron, V. (2012). Time, cost, and environmental impact analysis on construction operation optimization using genetic algorithms. Journal of Management in Engineering, 28(3), 265–272.

Panthi, A., and Hussain, A. (2025). Integrated optimization of construction schedules: A five-objective approach for time, cost, quality, safety, and environmental sustainability using OBNSGA-III. Asian Journal of Civil Engineering. https://doi.org/10.1007/s42107-025-01528-z

Panwar, A., and Jha, K. N. (2019). A many-objective optimization model for construction scheduling. Construction Management and Economics, 37(12), 727–739.

Sharma, K., and Trivedi, M. K. (2022). Latin hypercube sampling-based NSGA-III optimization model for multimode resource-constrained time–cost–quality–safety trade-off in construction projects. International Journal of Construction Management, 22(16), 3158–3168.

Sharma, K., and Trivedi, M. K. (2023). Modelling the resource-constrained time–cost–quality–safety–risk–environmental impact trade-off using opposition-based NSGA-III. Asian Journal of Civil Engineering, 24, 3083–3098.

Singh, G., and Ernst, A. T. (2011). Resource constraint scheduling with a fractional shared resource. Operations Research Letters, 39, 363–368.

Ward, S. C., Curtis, B., and Chapman, C. B. (1991). Objectives and performance in construction projects. Construction Management and Economics, 9(4), 343–353.

Zareei, M., and Hassan-Pour, H. A. (2015). A multi-objective resource-constrained optimization of time–cost trade-off problems in project scheduling. Iranian Journal of Management Studies, 8, 653.

Zhan, Z., Hu, Y., Xia, P., and Ding, J. (2024). Multi-objective optimization in construction project management based on NSGA-III: Pareto front development and decision-making. Buildings, 14, 2112.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

PRINCIPLES FOR THE INDUSTRIALISATION OF THE INFRASTRUCTURE RENOVATION AND REPLACEMENT ASSIGNMENT

Irfan Pottachola, Alexander Bletsis, and Trijntje Verschuuren
Netherlands Organisation for Applied Scientific Research (TNO), Delft, The Netherlands

Abstract

The Netherlands faces an unprecedented infrastructure renewal assignment that cannot be addressed through traditional project-based delivery. Although industrialisation is widely promoted as a solution, its adoption remains limited due to a lack of system-level understanding of the conditions required for scaling. This study positions industrialisation as a coordinated transformation of infrastructure asset management and develops a principles-based perspective to support this transition. Using thematic analysis of literature and expert validation, five principles are identified. Together, these principles clarify why isolated technological initiatives struggle to scale and establishes a foundation for developing action perspectives and policy directions to guide industrialisation.

Introduction

The Netherlands faces an urgent and complex challenge in the renewal (renovation and replacement) of its aging civil infrastructure (TNO, 2023). Approximately 40% of bridges and viaducts are approaching or have exceeded their intended service life, with annual renewal costs projected to rise from €1.1 billion in 2021 to €2.4 billion between 2021 and 2030, and further to €2.9 billion annually between 2031 and 2040 (TNO, 2023). Delivering this scale of intervention requires a significant increase in the pace and efficiency of infrastructure projects.

However, the Dutch construction sector, as currently organised, lacks the capacity and supporting ecosystem to respond effectively (Bletsis, 2020). The sector remains structured around bespoke, one-off projects, characterised by fragmented procurement practices, limited standardisation, and temporary supply chains. This "loosely coupled" structure limits knowledge transfer and learning across projects (Dubois and Gadde, 2002), constraining productivity improvements relative to the wider economy. As a result, existing delivery models remain heavily reliant on labour-intensive, on-site activities, which are associated with inefficiencies such as cost overruns, delays, and waste (Eriksson et al., 2014). At the same time, increasing pressures related to sustainability, resource efficiency, and a shrinking workforce further intensify the need for more productive ways of working (Pottachola, 2024).

Taken together, the scale and urgency of the renewal assignment render the prevailing one-to-one, project-based delivery model increasingly untenable. Without a fundamental shift in how infrastructure renewal is organised and coordinated, the sector will struggle to meet demand (TNO, 2023), making the search for scalable, efficient solutions, and the delivery models that enable them, not only necessary but inevitable.

In response to such systemic challenges, the construction sector has increasingly sought to adopt principles originating from manufacturing, such as standardisation, mechanisation, and automation (Lessing, 2015). These developments are collectively referred to as the industrialisation of construction. Initially, industrialisation in construction was primarily understood as the use of off-site manufacturing as technical means to relocate work away from the unpredictable construction site (Lessing, 2015). Over time, this interpretation has evolved into a more holistic perspective, where industrialisation is framed as a systematic, process-oriented approach characterised by standardisation, continuous improvement, and long-term supply chain integration (van Ee, 2022; van Groesen, 2022; Pottachola, 2024).

Within the Dutch infrastructure sector, industrialisation is widely regarded as a promising solution direction for addressing the renewal assignment. Initiatives such as Industrial, Flexible, and Demountable (IFD) bridge concepts illustrate this potential through reduced project lead times, faster on-site construction with less disruption, and designs that enable disassembly and long-term reuse (EIB, 2022). Feasibility studies in the Dutch context estimate a potential market for IFD of € 3.3 billion for decentralised asset owners between 2025 and 2035, with an associated economic benefit of approximately €300 million (EIB, 2022).

Despite this recognised potential and the availability of supporting technologies, the adoption of industrialised solutions in infrastructure renewal remains limited. This limited uptake cannot be attributed to technological immaturity, but rather to the fragmented way in which industrialisation is pursued, often through isolated instruments or technologies, without a coherent understanding of the organisational conditions required to support industrialised production (Bletsis, 2020). The infrastructure sector operates in an extremely demand-driven environment, reinforcing the central role of asset

owners in enabling or constraining industrialised production. In the Dutch context alone, nearly 370 decentralised public authorities are responsible for infrastructure assets. While efforts on industrialisation heavily focuses on the market side, the role of asset management organisations in this transition is frequently overlooked. Yet, it is precisely the upstream decisions made by asset owners, such as structuring demand to create continuity, defining project requirements that support standardisation, and designing procurement frameworks that incentivise long-term investment, that largely shape the conditions under which industrialised solutions can emerge and scale (EIB, 2022).

The complex nature of infrastructure construction, coupled with the systemic and multifaceted character of industrialisation, requires supporting frameworks capable of structuring and coordinating action across diverse actors (Brouwer, 2024). Although frameworks for industrialised construction have been developed in the housing sector (e.g., Lessing, 2015), these are not fully transferable to the infrastructure domain due to differences in asset characteristics, governance structures, and demand dynamics (Eriksson et al., 2014). Moreover, these frameworks are grounded in an implicit assumption that enabling conditions for industrialised construction such as stable demand and aligned procurement and governance structures are already in place. However, this assumption does not hold in the infrastructure sector, which is characterised by its demand-driven nature and complex product systems. Here, the conditions required for industrialisation are not given, but must be actively created, particularly through asset management practices. Consequently, existing frameworks provide limited guidance for coordinating industrialisation efforts in the infrastructure sector, leaving actors without a shared perspective to align decisions, organise actions, and assess progress.

This study addresses this gap by developing a principles-based framework for the industrialisation of infrastructure asset management, derived through thematic analysis of literature and validated with expert input. By identifying the fundamental principles that characterise industrialised infrastructure systems, the study provides a foundation for aligning and coordinating industrialisation efforts across actors as a system-level transformation of infrastructure asset management.

The remainder of this paper is structured as follows. First, the paper builds its theoretical foundation by drawing on existing industrialisation literature mostly from the building context and adapting these insights to the specific context of infrastructure asset management. Next, the research is described that was conducted to derive principles. The findings section presents the five derived principles. This is followed by a discussion of the implications for asset owners, governance, and sector-wide coordination. Finally, the paper concludes with key insights, limitations, and directions for future research.

Theoretical Foundation – Industrialisation of Asset Management

Industrialisation of construction represents a shift from a traditional design-to-product mindset, where each project is treated as inherently unique, toward a product-to-design perspective that emphasises system-oriented thinking (van Ee, 2022). This shift aims to enable construction companies to achieve economies of scale through integrated value and supply chains across projects (van Groesen, 2022). This transition has been most visibly advanced in the housing sector, where industrially produced modular homes are increasingly regarded as a viable response to the European affordable housing crisis (van Ee, 2022). The growing adoption of prefabricated and modular housing solutions reflects an acceptance of industrial approaches as alternatives to so-called traditional construction methods.

The transition to a product-based mindset in infrastructure sector is not yet advanced. Infrastructure projects are frequently characterised as complex product systems (CoPS) that are unique in nature, which has contributed to a reluctance to adopt standardisation and platform-based approaches due to site-specific conditions and varying client requirements (Stout, 2016). As a result, industrialisation in infrastructure often manifests as isolated technological interventions rather than as a coordinated reorganisation of production and decision-making processes across the value chain. Recent studies, however, demonstrate that even complex civil engineering structures such as bridges can be decomposed into standardised modules and subsystems (EIB, 2022). This perspective is particularly relevant in the context of infrastructure renewal, where the dominant task is no longer greenfield construction but the large-scale renewal of existing assets.

To build on this shift in perspective, industrialisation of construction in this study is positioned not as the application of isolated methods or technologies, but as a coordinated transformation of infrastructure asset management that creates the conditions for the effective adoption and scaling of productivity-enhancing approaches across the value chain (Brouwer, 2024). The definition adopted in this study is synthesised from recurring elements in the literature and subsequently contextualised to the Dutch renewal assignment, where fragmented demand, temporary coalitions, and high transaction costs inhibit productivity improvement. The formulation deliberately emphasises organisational restructuring, scale creation, and the enabling of process-oriented technologies, as these dimensions repeatedly emerged as prerequisites for industrial performance (Lessing, 2015; Stout, 2016; Bletsis, 2020; EIB, 2022; Brouwer, 2024). In line with this perspective, and tailored to the specific dynamics of the Dutch infrastructure renewal assignment, we define industrialisation as:

'The reorganisation of (specific parts of) the infrastructure sector, to reduce its inherent project-based and demand-driven nature, in order to implement and enhance process-oriented technologies, for the

purpose of achieving economies of scale and reducing transaction costs, and eventually to mitigate the productivity challenges in the renewal assignment.'

Industrialisation therefore implies a shift from a purely project-based logic toward a more project-independent or product-oriented approach, in which commonalities across projects are systematically identified and leveraged through product platforms and standardised solution spaces (van Groesen, 2022). Such a shift enables organisations to move from reactive forms of production towards more proactive forms, including mass customisation, where project-specific variation is accommodated within a stable, standardised system (van Groesen, 2022). By aligning supply and demand more effectively, industrialisation creates the conditions under which repetition, learning, and optimisation across projects become possible (Bletsis, 2020).

The reorganisation of the sector is also a prerequisite for the effective implementation of process-oriented technologies, which form the operational backbone of industrialisation (Lessing, 2015). Technologies such as automation, mechanisation, and digital configurators enable the transition from craft-based, labour-intensive practices to manufacturing-led production processes. However, without the process stability and organisational alignment provided by industrialisation, advanced technologies cannot be effectively deployed at scale (Pottachola, 2024). In this sense, industrialisation is not the result of technological adoption, but rather a necessary condition for technologies to deliver productivity gains.

Operationally, industrialisation aims to realise economies of scale and economies of repetition, which are currently difficult to achieve in the infrastructure sector due to fragmented and unpredictable demand (Brouwer, 2024). Standardisation and the aggregation of demand into programs or portfolios enable the production volumes required to justify investments in industrial facilities and tools, thereby reducing unit costs (EIB, 2022). In parallel, a shift from transactional, lowest-price procurement toward long-term supply chain partnerships reduces transaction costs associated with repeated tendering, coordination, and non-value-adding activities, resulting in a more economically viable and resilient value chain (Eriksson et al., 2014).

In this way, industrialisation offers the potential for a necessary productivity leap, enabling the sector to execute extensive renewal programmes within financial, temporal, and environmental constraints, while maintaining quality and safety standards. However, this conceptualisation implies that it cannot be operationalised through isolated methods, technologies, or best practices alone. Given the diversity of asset types, organisational settings, and institutional arrangements in the infrastructure sector, prescriptive solutions are unlikely to be transferable or scalable across contexts (Bletsis, 2020). Therefore, industrialisation requires a framework that articulates the fundamental conditions underpinning such a transition, while allowing for contextual interpretation and implementation.

In this study, these foundational conditions are expressed in the form of principles. Principles serve as high-level propositions that describe essential characteristics of industrialised asset management systems without prescribing specific solutions or technologies. As such, they provide a shared analytical lens for understanding industrialisation across different actors and contexts. They form the basis for systematically analysing how organisations enable or constrain the adoption of productivity-enhancing approaches.

Research Method

This study adopts a qualitative, exploratory research design aimed at defining foundational principles for the industrialisation of infrastructure asset management. The study uses thematic analysis as its primary analytical method to identify, synthesise, and abstract recurring values and patterns into a coherent set of industrialisation principles.

The research process began by formulating a working definition of industrialisation in the context of infrastructure asset management. Industrialisation was conceptualised as a systemic transformation that reduces the inherently project-based nature of infrastructure delivery with scalable, process-oriented approaches. Based on this definition, the purpose of the principles was defined. The principles were positioned as (1) fundamental propositions that provide an action-oriented perspective for stakeholders; (2) a conceptual foundation for defining and guiding industrialisation efforts and for informing policymakers; and (3) a basis for future maturity assessment models. This step established the analytical lens used throughout the study.

A targeted literature review was conducted to explore theoretical and empirical insights related to industrialisation. Search strings combined keywords such as industrialised construction, industrialised building systems, modular construction, prefabrication, and product platforms, together with related concepts including standardisation, DfMA, off-site manufacturing, process integration, supply chain integration, and construction productivity. The initial set of 84 identified publications were screened based on their relevance to the objectives of this study. The primary inclusion criterion concerned the extent to which a study provided insight into why industrialisation-related approaches were adopted and which organisational, institutional, or process conditions were associated with successful implementation and productivity improvement. Following this screening, 39 publications were retained for detailed analysis. Given the limited availability of systemic perspectives on industrialisation within infrastructure, many selected studies originated from the building sector.

Data from literature were analysed using thematic analysis. The analysis proceeded in four iterative steps. First, key concepts and terms related to industrialisation, productivity, and organisational change were identified across data sources. Then, related codes were grouped into broader themes representing recurring values and

patterns. The focus during this analysis was understanding what the characteristics, enabling conditions, and motivations behind various implementation of industrialisation technologies were. The identified themes were further refined and abstracted to capture their underlying fundamentals, resulting in a set of higher-order principles. Each principle represents a necessary condition for enabling industrialised asset management.

The derived principles were subjected to expert validation through a series of semi-structured interviews with domain experts. In total, experts included asset managers, policy specialists, and engineers representing four decentralised authorities, one national agency, and four contractors. In these sessions, the five principles and their preliminary definitions were presented to each organisation individually. Participants were asked to reflect on the extent to which the principles corresponded with their current productivity ambitions, strategic directions, and practical experiences.

The interviews examined the relevance, clarity, and completeness of the principles, as well as their recognisability in ongoing initiatives. Feedback from these validation interviews was used to refine the formulation of the principles, sharpen definitions, and ensure their applicability across both demand- and supply-side contexts.

Findings

The thematic analysis resulted in five principles that characterise industrialisation in infrastructure asset management. A detailed overview of the coding structure linking codes with relevant sources, themes, and principles is provided as supplementary material to this paper (TNO, 2026). Each principle is discussed in detail in the following sections.

Principle 1: Continuity and Repeatability

In any industry, the market operates most efficiently when there is a steady flow of demand (EIB, 2022). In construction, however, a fragmented, project-based approach forces organisations to constantly (re)adjust their resources and operations, leading to inefficiencies, increased costs, and slower delivery (Dubois and Gadde, 2002). A continuous flow of standardised demand allows for a stable and predictable environment where workflows can be optimised and scaled over time (EIB, 2022).

The principle of Continuity and Repeatability describes the extent to which infrastructure renewal activities are organised to enable stable production flows, and reuse across projects. The analysis highlights a transition from fragmented, one-off project delivery towards process-based approaches in which production logic is decoupled from individual project contexts (Eriksson et al., 2014). Central to this principle is the aggregation of demand through programmatic and portfolio-based planning to create predictable workloads that sustain industrial operations (Eriksson et al., 2014; EIB, 2022; van Groesen, 2022). Long-term supply chain integration also emerged as a key theme, reflecting the importance of stable partnerships, team continuity, and synchronised off-site and on-site processes in maintaining steady production flows (Vrijhoef and Koskela, 2005; Söderholm, 2010; Lessing, 2015; Pottachola, 2024).

Along these lines, we define Continuity and Repeatability as *the principle of ensuring continuous operations over time in a way that supports the development of repeatable products and processes.*

In essence, this principle enables scalable and predictable processes, optimises resource allocation, reduces costs and project durations, and fosters innovations and reuse of data, and modular solutions across projects.

Principle 2: Specialisation

Continuity in processes enables organisations to simplify workflows by reducing portfolio diversity and complexity. This, in turn, creates the opportunity to structure operations around a more focused and narrowed scope of products or services, allowing organisations to concentrate their resources and expertise on these defined areas (Vrijhoef and Koskela, 2005). Specialisation refers to the deliberate narrowing of organisational scope to increase efficiency and performance through focused capabilities.

The thematic analysis indicates a shift from generalist, project-oriented organisations towards product- and process-oriented structures. This includes strategic limitation of the solution space (Stout, 2016), market segmentation, and the development of specialised offerings aligned with repeatable production processes (Vrijhoef and Koskela, 2005). At the organisational level, specialisation is supported by permanent, dedicated teams and the distinction between core and non-core competencies (Vrijhoef and Koskela, 2005; Lessing, 2015). At the supply chain level, the analysis points to a transition from project-based subcontracting towards system suppliers and long-term strategic partnerships, enabling cumulative learning and early integration of specialised knowledge (van Ee, 2022; van Groesen, 2022; Pottachola, 2024).

Along these lines, we define Specialisation as *the principle of focusing resources and expertise on a specific, narrowed scope of assets or services by simplifying workflows and reducing portfolio complexity.*

Specialisation enhances the organisation's technical proficiency, simplifies portfolio management, and enables the market to deliver high-quality, tailored solutions for specific assets or types of interventions.

Principle 3: Knowledge Codification

Continuous improvement is necessary in an industrial system (Lessing, 2015). In the fragmented and project-based nature of the construction sector, valuable knowledge is often lost when teams disband after projects (Dubois and Gadde, 2002). Most of this knowledge remains tacit, leading to missed opportunities for improvement and a failure to learn from mistakes (Pottachola, 2024). Furthermore, this knowledge is frequently lost as individuals retire or move to organisations (Pottachola, 2024).

Knowledge codification captures the transformation of tacit, individual expertise into explicit, reusable organisational assets. The analysis reveals recurring themes related to formalising decision logics, and solutions in order to avoid repeated reinvention (Stout, 2016; Andersson and Lessing, 2017). Structuring information for interoperability and machine readability emerged as a critical theme, including the use of standardised data formats, taxonomies, and shared information structures such that knowledge and information from multiple individuals can be aggregated (van Ee, 2022; Pottachola, 2024). Feedback loops for continuous learning and the use of digital tools, such as semantic models, BIM, and design configurators support the retention, application, and evolution of codified knowledge across projects (Cao et al., 2021; van Ee, 2022; Pottachola, 2024).

Along these lines, we define Knowledge Codification as *the principle of formalising and embedding tacit knowledge into structured systems to ensure it is retained, reused, and enables continuous improvement across organisations, projects, and individuals.*

Knowledge codification thus enables organisations to systematically build on past knowledge, improving decision-making, execution, and long-term planning. This ensures the retention of expertise as individuals leave, enables building on experience for better decision-making, facilitates faster onboarding, and prevents repeated errors or rediscovery of solutions.

Principle 4: Pre-Engineering

In the construction industry, resources are often allocated to create bespoke designs and solutions for individual projects (Dubois and Gadde, 2002). While uniqueness is an inherent characteristic of construction projects, many projects share commonalities that are often overlooked, particularly in the infrastructure sector (Bletsis, 2020; EIB, 2022). These overlooked commonalities prevent the industry from utilising standardisation, thereby limiting scalability.

Pre-engineering refers to shifting engineering effort upstream by resolving complexity early and systematically. The analysis highlights the decomposition of complex systems into modular components (Stout, 2016; van Groesen, 2022) and the development of product platforms (Knippenberg et al., 2019; van Ee, 2022) or kits of parts (Cao et al., 2021) as central mechanisms. This enables configuration-based solutions while isolating variability to specific modules.

Front-loading design decisions, including early design freezes and early integration of suppliers (Wuni and Shen, 2020), is identified as a key theme to reduce downstream uncertainty. Further themes include the integration of design and production through Design for Manufacture and Assembly (DfMA) principles and the gradual shift of production logic from on-site construction towards manufacturing-oriented processes, including off-site and hybrid production models (Eriksson et al., 2014; Pottachola, 2024).

Along these lines, we define pre-engineering as *the principle of developing reusable products and processes through standardisation and pre-definition, enabling scalability and reducing the reliance on bespoke, project-specific efforts.*

Pre-engineering accelerates project delivery by reducing bespoke efforts and thus shorter design/ planning lead times, enhances market productivity through reusable and optimised designs and processes, and supports the development of sustainable and circular methods and processes.

Principle 5: Socio-Technical Alignment

The construction industry is inherently fragmented vertically (across project phases), horizontally (across expertise), and longitudinally (across projects) (Adriaanse, 2014). This fragmentation leads to misalignment between social systems (people, contracts, organisations) and technical systems (processes, technologies, tools) (Pottachola, 2024). Industrialisation requires a collective effort toward common goals, and achieving alignment between these systems is important for a full-fledged industrial system.

Sociotechnical alignment captures this need to align contractual, organisational, cultural, and digital dimensions of industrialisation. The analysis shows that industrial objectives are difficult to realise without contractual frameworks that support collaboration, shared risk, and early stakeholder involvement (Vrijhoef and Koskela, 2005; Bletsis, 2020). Traditional transactional and lowest-price contracting approaches frequently conflict with industrialised modes of operation (Eriksson et al., 2014; Bletsis, 2020). Beyond formal arrangements, the analysis highlights the importance of network-centric cultures based on trust, transparency (Stout, 2016), and shared situational awareness (Bletsis, 2020; Pottachola, 2024). Aligning digital systems with organisational workflows also emerges as a theme (Pottachola, 2024), indicating the required social and organisational transformation as a technical one.

Along these lines, we define Sociotechnical alignment as *the principle of achieving the alignment between social systems and technical systems, ensuring that resources are coordinated to function cohesively within a larger operational framework.*

Sociotechnical alignment ensures stakeholder priorities and resources are aligned for effective collaboration on large assignments, such as the infrastructure renewal assignment. Furthermore, this enables the full utilisation of product platforms' potential.

Expert Validation Results

To evaluate these principles, validation interviews were conducted with senior asset managers and project managers from asset owners, contractors and engineering firms. The principles and their definitions were well received, with practitioners confirming that they represent not merely theoretical constructs but active levers they consider for addressing the productivity gap. They were widely recognised, prompting reflection on how they are

already applied in practice while also revealing weaknesses in current approaches and opportunities for improvement.

Continuity & repeatability was consistently recognised as foundational to achieving industrialisation. Asset owners emphasised that creating continuity in demand is increasingly a strategic priority, as it enables more reliable collaboration with the market, lowers inherent variability in tender outcomes, and provides space for innovation. Contractors highlighted that continuity directly reduces transaction costs by allowing reuse of designs, processes, and supply-chain arrangements; several noted that repeatability enables significant learning effects that are currently lost in one-off projects.

Specialisation was seen as a means to allocate resources outside of the project scope and towards more continuous and repeatable operations. Specialisation is relevant at multiple scales within organisations and the sector. For example, on the team level specialisation was related to the ongoing shift of moving from project teams to expertise teams that works across multiple projects often resulting in higher productivity yields.

Knowledge codification was widely recognised by practitioners as essential, yet it remains a difficult principle to operationalise. While organisations acknowledge the need to retain and reuse knowledge, they struggle to translate tacit experience into structured, reusable assets. Challenges arise in transferring knowledge across projects. While lessons learned may be documented through evaluation at project closure, they are rarely converted into systematised and codified forms such as design rules, or configuration logics. In addition, underlying data is often fragmented, stored in inconsistent formats, or confined to individual projects. The experts linked this challenge back to specialisation, as a means to have coherent systems that enable knowledge capture across projects.

Practitioners mentioned a wide variety of activities in asset management where pre-engineering has benefits. Examples included pre-engineered parametric models for the reassessment of bridge typologies, development and use of IFD-standards to stimulate product-platforms for bridges, and detailed designs of specific modular bridge concepts that can be applied across projects. Across these applications, implementing this principle was seen to deliver clear added value.

Sociotechnical Alignment emerged as one of the most challenging and systemically important principles. Practitioners repeatedly noted that even when technical solutions exist (such as standards for IFD bridges) misalignment in contracts, governance structures, and organisational incentives prevents their adoption and potential. Examples included projects where standardisation was technically feasible but not contractually supported, resulting in bespoke solutions despite shared intentions to industrialise. Besides the variability in specifications across projects, current procurement practices often incentivise late-stage design changes thus hindering pre-engineering or the adoption of a pre-engineered concept.

Discussion

While extant literature has catalogued the barriers hindering industrialisation in construction a unified perspective explaining the foundational conditions required to enable this transition has remained largely absent, especially in the infrastructure sector. This study extends the existing body of knowledge by shifting the analytical focus away from individual enabling technologies toward the underlying conditions that shape industrialisation across the entire infrastructure value chain. The five principles identified through the thematic analysis collectively describe how industrialisation restructures the organisational, institutional, and production logic of infrastructure asset management. This supports the ongoing shift in the literature away from viewing industrialisation in construction as a production technique, toward understanding it as a broader organisational and system-level transformation of construction delivery (van Ee, 2022; Pottachola, 2024). The results help explain why the increasing maturity of individual enabling technologies has not translated into widespread adoption. The principles are not independent but forms an interdependent system of conditions. For example, Continuity and Repeatability create the essential market stability required for Specialisation. Without a predictable pipeline of aggregated demand, firms cannot justify the investments in specialised equipment or fixed process crews required for industrial efficiency (Vrijhoef and Koskela, 2005; EIB, 2022). Simultaneously, Knowledge Codification acts as the prerequisite for Pre-Engineering and configuration-based solutions. Shifting from tacit to explicit knowledge is critical for enabling digital configurators and product platforms that facilitate mass customisation (van Ee, 2022; Pottachola, 2024). Socio-Technical Alignment underpins the ability of organisations to sustain these industrialised practices over time by addressing the social gap where implementation fails due to a lack of cultural adaptation or workforce skills (Bletsis, 2020).

Partial implementation, such as adopting isolated technologies or methods without addressing their enabling conditions, is therefore unlikely to result in sustained productivity improvements. The interdependency of principles provides a plausible explanation for why many industrialisation initiatives remain confined to pilots rather than scaling across organisations and portfolios. By making these enabling conditions explicit, the principles offer a diagnostic lens for understanding where and why adoption stalls, and where targeted interventions are required.

These findings have important implications for infrastructure asset management organisations. Importantly, it challenges the implicit assumption that industrialisation is primarily a supply-side responsibility. Asset management activities can broadly be divided into demand-side led planning phases, and supply-side led realisation phases. In practice, industrialisation is

frequently associated with the realisation phases through technologies such as prefabricated systems, with the locus of responsibility placed largely on contractors and manufacturers (EIB, 2022). While these technologies represent concrete applications of industrialisation, they are insufficient in isolation to deliver the required productivity gains. The principles highlight that without the bundling of demand and the harmonisation of technical standards across projects, the market lacks the continuity needed to invest in standardised, repeatable, and scalable solutions. Thus, industrialisation must be reconceptualised not just as a downstream construction tactic, but also as an upstream asset management strategy.

To support the formulation of such upstream strategies, the principles developed in this study provide an actionable reference for both individual asset owners and for coordination across asset owners within the wider infrastructure governance system. At the level of individual asset management organisations, the principles offer a structured way to translate industrialisation ambitions into concrete strategic choices. In practice, this may manifest in portfolio and programme bundling strategies that prioritise repetition and continuity, or in the selection of procurement and contract arrangements that incentivise long-term collaboration, data sharing, and knowledge reuse.

Beyond individual organisations, achieving industrialisation at scale requires coordination across asset owners. In the context of the polycentric Dutch infrastructure sector (Bletsis, 2020), industrialisation cannot be realised through top-down control or isolated organisational change alone. The principles provide a shared coordination mechanism that can support alignment across organisational boundaries. By articulating common conditions for industrialised project delivery, the principles offer a common language through which multiple asset owners can coordinate strategies. In practice, this may manifest in joint project bundling to create market continuity, as well as in the alignment of specifications and contract frameworks to enable standardisation and repeatability. In this sense, the principles enable coordination without centralisation, allowing asset owners to retain autonomy while contributing to a shared industrialisation trajectory.

The principles also complement existing maturity models and transition frameworks by shifting the focus from organisational capability to system-level suitability. Existing maturity models for industrialised construction, such as the Industrialised Construction Maturity Model which evaluates organisational enablers and outcomes (Wang et al., 2020), primarily assess how well an organisation is equipped to develop and deliver industrialised solutions. Implicitly, these models assume that the broader system conditions in which organisations operate are conducive to industrialisation. The principles proposed help to assess whether the foundational conditions across the value chain are aligned to enable industrialisation in the first place. For example, a contractor may develop a technically mature modular bridge concept that scores highly on existing maturity models yet fails to scale due to misalignment with demand-side practices and needs, procurement frameworks, or other sociotechnical conditions. By explicitly accounting for such system-level alignment, particularly between asset owners, market actors, and governance arrangements, the principles enable an assessment of whether organisations are operating within an appropriate playing field for industrialised infrastructure renewal. This system-oriented perspective is largely absent from existing maturity frameworks in the literature.

Conclusions

This study examined why industrialisation, despite recognised potential and technological maturity, has struggled to scale within the Dutch infrastructure renewal context. Through a thematic analysis of literature, complemented by expert validation, the study identified five foundational principles that characterise industrialised infrastructure asset management: Continuity and Repeatability, Specialisation, Knowledge Codification, Pre-engineering, and Sociotechnical Alignment. The findings illustrate that industrialisation cannot be realised through partial or isolated interventions. The principles are inherently interdependent, and this interdependence helps explain why many industrialisation initiatives remain confined to pilots, despite technological maturity. Importantly, the study highlights the decisive role of asset management organisations in shaping the conditions for industrialisation through upstream decisions related to portfolio planning, requirement setting, procurement, and long-term demand aggregation.

Through the principles, this study contributes a missing system-level perspective to the literature on industrialised construction. The principles provide a shared language for aligning stakeholders, structuring decision-making, and creating the conditions necessary for sustained productivity improvements. In doing so, it supports the transition of industrialisation from an abstract ambition to a structured, strategic approach capable of addressing the scale, urgency, and complexity of infrastructure renewal.

This study has several limitations that open avenues for future research. First, the framework is grounded in the Dutch infrastructure renewal context; its applicability to other construction domains with different governance and market structures remains to be tested. Second, the principles are primarily derived from literature and supported through expert validation rather than extensive empirical observation. Future work could therefore enrich the framework through case-based studies that examine how industrialisation conditions emerge and interact in practice. Third, the principles are currently treated as equally important. While their interdependencies are recognised, their relative weight, sequencing, and dominance remain unexplored. Further research could investigate these dynamics and translate the principles into maturity models for transition support.

References

Adriaanse, A. (2014) Bruggen Bouwen met ICT. Universiteit Twente. Enchede. Available at:

Andersson, N. and Lessing, J. (2017) 'The Interface between Industrialized and Project Based Construction', Procedia Engineering, 196, pp. 220–227. Available at: https://doi.org/10.1016/j.proeng.2017.07.193.

Bletsis, A. (2020) Building bridges for bridges: Towards a Participatory Approach for Tackling Large-Scale Societal Assignments in Project-Based Polycentric Systems: The Case of the Large-Scale Rehabilitation Assignment of Civil Engineering Structures in the Netherlands. Master Thesis. TU Delft. Available at: https://repository.tudelft.nl/record/uuid:b65cdc0c-1cde-4e99-8ade-6591c8980c0c.

Brouwer, T. (2024) Exploring the transition to industrialisation in Dutch infrastructure construction. Master Thesis. TU Delft. Available at: https://repository.tudelft.nl/record/uuid:7c559c40-37f1-4669-a04f-a0e0a9d3d640.

Cao, J. et al. (2021) 'Cross-phase product configurator for modular buildings using kit-of-parts', Automation in Construction, 123, p. 103437. Available at: https://doi.org/10.1016/j.autcon.2020.103437.

Dubois, A. and Gadde, L.-E. (2002) 'The construction industry as a loosely coupled system: implications for productivity and innovation', Construction Management and Economics, 20(7), pp. 621–631. Available at: https://doi.org/10.1080/01446190210163543.

van Ee, T. (2022) A Digital product platform for modular and industrial construction: The design of a reference ontology and an application landscape. Engineering Doctorate Thesis. University of Twente. Available at: https://research.utwente.nl/en/publications/a-digital-product-platform-for-modular-and-industrial-constructio/.

EIB (2022) IFD in de infrastructuur. Feasibility Study. The Netherlands: Economisch Instituut voor de Bouw. Available at: https://www.eib.nl/wp-content/uploads/2023/01/EIB-Eindrapport-IFD-in-de-infrastructuur-web.pdf.

Eriksson, P.E. et al. (2014) 'Managing short-term efficiency and long-term development through industrialized construction', Construction Management and Economics, 32(1–2), pp. 97–108. Available at: https://doi.org/10.1080/01446193.2013.814920.

van Groesen, W. (2022) Industrialized Renovations: Towards Mass-Customization in the Construction Industry. Engineering Doctorate Thesis. Technische Universiteit Eindhoven. Available at: https://research.tue.nl/nl/publications/industrialized-renovations-towards-mass-customization-in-the-cons/.

Knippenberg, S.C.M. et al. (2019) 'Development of a Product Platform for Ship Locks Using DSM Methods', Proceedings of the 21st International DSM Conference. The 21st International DSM Conference, The Design Society. Available at: https://doi.org/10.35199/dsm2019.15.

Lessing, J. (2015) Industrialised House-Building - Conceptual orientation and strategic perspectives. Doctoral Thesis. Lund University. Available at: https://lup.lub.lu.se/search/files/3873278/8145690.pdf.

Pottachola, I. (2024) Towards digital twins driven modular construction: Development of a digital twin uses framework & digital twin usage scenarios for modular construction. Engineering Doctorate Thesis. University of Twente. Available at: https://research.utwente.nl/files/481742645/10_Irfan_Pottachola.pdf.

Söderholm, E. (2010) Applicability of continuous improvements in industrialised construction design process. Luleå: Luleå University of Technology.

Stout, A. (2016) Application of mass customisation: standardisation and modularisation, for an infrastructural object: a viaduct. Master Thesis. University of Twente.

TNO (2023) Vernieuwingsopgave infrastructuur: Landelijk prognoserapport. The Netherlands: TNO.

TNO (2026) Supplementary material to Pottachola et. al (2026). TNO-2026-16474. The Netherlands: TNO

Vrijhoef, R. and Koskela, L. (2005) 'Revisiting the Three Peculiarities of Production in Construction'. 13th International Group for Lean Construction Conference, Sydney, Australia.

Wang, G. et al. (2020) 'A Building Project-Based Industrialized Construction Maturity Model Involving Organizational Enablers: A Multi-Case Study in China', Sustainability, 12(10), p. 4029. Available at: https://doi.org/10.3390/su12104029.

Wuni, I.Y. and Shen, G.Q. (2020) 'Critical success factors for modular integrated construction projects: a review', Building Research & Information, 48(7), pp. 763–784. Available at: https://doi.org/10.1080/09613218.2019.1669009.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

MODEL TO PRESS: DIGITAL DATAFLOW FOR THE DESIGN, RATIONALIZATION, AND MANUFACTURE OF FREE-FORM GLULAM BEAMS

Tom Svilans[1], Edgars Rudzītis[2], Jānis Iejavs[3], Edvīns Grants[3], and Kristaps Zīverts[3]
[1]Royal Danish Academy, Copenhagen, Denmark
[2]ZAZA Timber Production Ltd., Jelgava, Latvia
[3]Forest and Wood Products Research and Development Institute (MeKA), Jelgava, Latvia

Abstract

While numerous built examples of free-form glulam structures demonstrate the complexities of creating glulam beams with out-of-plane curvature, the specifics of their modelling, rationalization, and realization have largely remained confined to specialist know-how and internal processes. We present a digital workflow that bridges unpredictable input design geometry and the reconstruction, rationalization, and preparation of free-form glulam beams for industrial production. This effort is part of an innovation project developing an automated press for serial production. It builds upon previous research into early-stage modelling and rationalization, extending it to real-world design contexts and industrial production constraints.

Introduction

While the complexity of timber structures has grown as their potential is discovered, the digital infrastructure required to translate free-form timber designs into robust, mass-produced industrial components remains sparse and fragmented. Techniques of rationalizing and modelling such geometrically complex structures are largely relegated to specialist experience and internal knowledge. The production of large-scale free-form glulam beams, for example, remains a labour-intensive and costly process, tailored to the specific delivery or project. The specifics of translating design geometry into feasible material specifications for such elements are often glossed over in favour of describing the wider design and production process (Stehling et al., 2017). As such, free-form glulam beams remain a largely craft-based product that resists automation and industrialization.

A number of wider efforts have spurred the tight intertwining of digital design and production in contemporary timber research and construction. New modes of relating timber properties and behaviours to architectural models or production environments (Svilans et al., 2019) have sought to imbue a sense of material agency in design and fabrication (Menges et al., 2016; Weinand, 2017). Considering this growing role of computation and automation within the timber manufacturing sector, we seek modelling and analysis methods for deriving production-ready data from coarse design models in an industrial production context, towards an industrialization of geometrically complex timber elements. A primary challenge in this process is the difference in fidelity between early-stage design models – often used for cost estimations and planning – and the necessary production data. Messy design geometry and poor interoperability necessitate interpretation and reconstruction of key manufacturing parameters and specifications. The use of coarse and ad-hoc models for early-stage design exchange often yields geometry that is unstructured and discretized. Further, free-form glulam beams present particular geometric and material constraints arising both from material properties as well as the nature of the pressing process. These must be respected, and the digital model must be validated against them to ensure that the production can be successful. Finally, considering the geometric constraints of bending timber lamellas, the discrepancy between the design geometry of the final beam and the geometry of the glulam blank must be recognized and integrated into the planning process.

We therefore develop a digital workflow for translating highly variable design inputs such as coarse and unstructured mesh objects into materially-responsive production data for pressing large-scale glulam beams under real-world constraints (Fig. 1). We draw from experience working with designers and architects on the delivery of complex timber buildings, and the observations made with respect to typical digital design models.

The presented work demonstrates a path towards the automation and industrialization of free-form glulam beam production from typical real-world design inputs.

Background and related work

Data exchange in free-form timber construction

The importance of robust digital workflows and good interoperability in AEC have been emphasized for a long time (Scheurer, 2012; Stehling et al., 2014), pointing at the wide variety of software used in the design, development, and production of a complex timber building and the lack of a “one size fits all” toolset. The prevalent use of Industry Foundation Classes (IFC) in AEC data exchange is beneficial for the general sharing of design information. However, in practice, the capabilities of the full specification are rarely utilized, especially when it comes to free-form surfaces. This means that the required input data is seldom authored in a

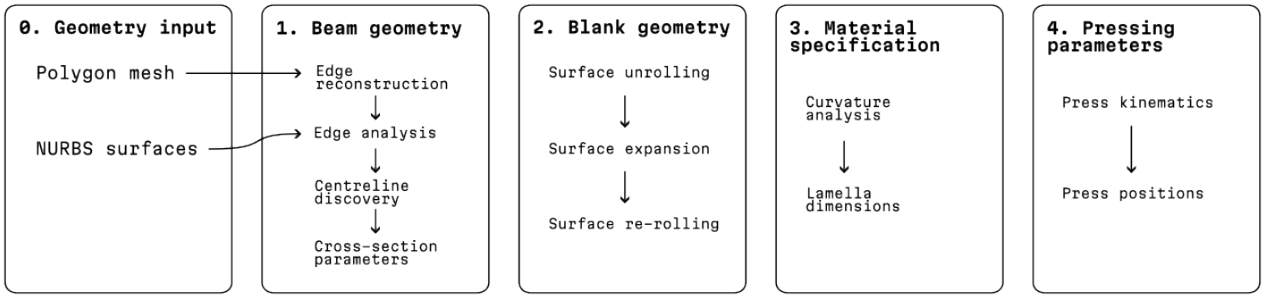


Figure 1 An outline of the digital workflow for automating the translation of coarse design data into production-ready information.

production-ready way. Specifically, the modelling of free-form beams is typically done through ad hoc or non-standardized ways and transferred as polygon meshes or overly detailed NURBS surfaces, creating hurdles for downstream processes. In the case of meshes, important features such as curvature and continuity must be inferred or interpolated. In the case of surfaces, fully detailed models can make it hard to extract precisely the necessary data for producing glulam beams since they might describe the final geometry of the fully detailed element rather than the glulam blank.

For evaluating material specifications and determining the correct press parameters, these geometric properties must be taken into account. Methods for reconstructing surfaces and edges that closely approximate the input design geometry are therefore required.

Single-curved glulam beams are relatively standardized in production, as both their geometry and production parameters are easier to derive. The planar curvature means that most considerations can be relegated to 2D analysis and rationalization. In this way, curvature calculations and lamella thicknesses are relatively straightforward.

For free-form glulam beams with out-of-plane curvature and torsion, previous work has developed modelling methods that result in appropriate material specifications and clean geometry for representing them (Svilans, 2021), however these rely on key data which is often absent in design models – such as a beam axis curve and related cross-sectional data and orientations. These must be introduced towards the beginning of the modelling process and are not always present in design models transferred from outside clients, partners, or collaborators. Additionally, these efforts do not include the translation of glulam blank geometry into specific press parameters for full-scale production.

Therefore, the transfer of information from design to the glulam press is shaky and unpredictable due to the large variance of inputs, necessitating techniques for stabilizing this translation into repeatable and automatable processes.

Industrial process and geometric constraints

Free-form glulam beams are typically fabricated using one of two methods:

- a two-step method where a single-curved blank is first fabricated out of individual *lamellas*, sliced into *layers* in the plane of its curvature, and then bent and laminated out of this plane; and
- a one-step process where approximately square lamellas are assembled in a grid and bent three-dimensionally at once.

We focus on the first method – the two-step lamination process – as it decouples the complex curvature into more manageable steps, albeit with some geometric caveats (Fig. 2).

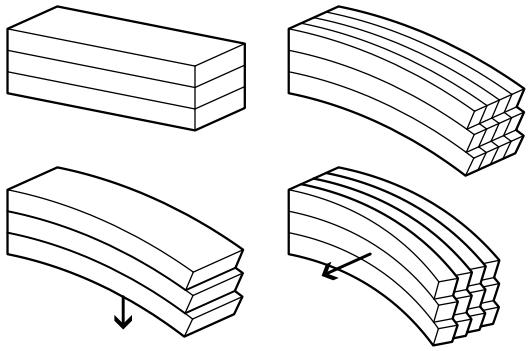


Figure 2 The two-step lamination process begins with straight lamellas (top left) which are bent into a single-curved blank (bottom left). This intermediate blank is resawn into layers (top right) and then bent and laminated in the out-of-plane direction (bottom right).

Double-curved glulam blanks present specific geometric constraints due to their production process. Given lamellas of equal length, as the blank is pressed into a curved shape, the lamellas will slide against one another, meaning that lamellas on the outside of the curvature will appear "drawn back" from the end of the beam and, conversely, lamellas on the interior of the curve will appear to be "pulled forward". While for single-curved glulam blanks this is simple to account for and the cross-section of the blank stays rectangular, out-of-plane curvature causes this slippage to manifest in geometrically complex ways, causing the cross-section to also slip into a trapezoidal shape (Fig. 3). Calculating the final geometry of the double-curved glulam blank therefore requires more advanced methods of analysis and modelling to more accurately reflect the true shape of the glulam blank.

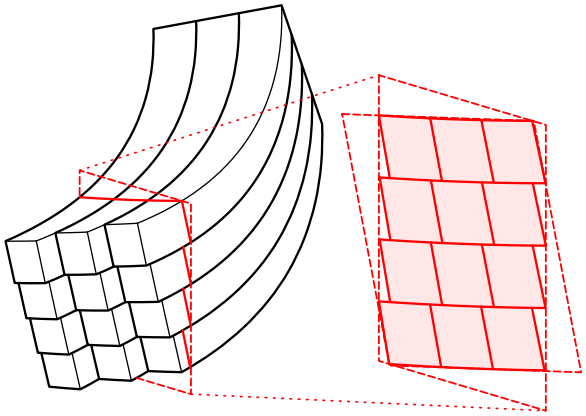


Figure 3 The out-of-plane bending causes the cross-section to become a trapezoid, depending on the degree of curvature.

Similarly, calculating the required lamella thickness for single-curved glulam blanks is straightforward: the lamella thickness is driven by the point of maximum curvature or the smallest radius of curvature in the plane of bending. The largest allowable thickness to avoid overstraining the lamella is defined by the ratio t/R, where t is the lamella thickness and R is the smallest radius of curvature. In Eurocode 5, this ratio must be less than 1/200 for softwoods. In double-curved glulam blanks, the curvature no longer lies in the plane of bending, therefore both the thickness and width of the lamella is affected. The assessment of curvature must therefore be broken down into components to arrive at the correct lamella thicknesses for both steps of the two-step process. To achieve this, the single-curved lamella thickness constraint must be reformulated to work with out-of-plane curvature.

Other challenges are similar for double-curved glulam blanks as for single-curved ones: as curvatures increase and lamella thicknesses get smaller, the spacing of the press pads must decrease to accommodate the sharper changes in geometry. Practically, this becomes an issue when the pressing infrastructure starts to get in its own way and inhibit the ideal distribution of lamination pressure.

Related research in this area has sought to avoid the difficulties of the double-curved lamination process in intriguing ways, such as through novel methods for sawing curved lamellas, thus reducing the bending required to achieve a particular 3d form

(Chai et al., 2021; Zhang et al., 2023). However, these prototypical approaches are not suitable for current industrial contexts that are based on the use of standardized lumber and straight timber lamellas.

Methodology

The project takes places in the factory of ZAZA Timber Production Ltd., a Latvian producer of large-scale timber building components (Fig. 4). An existing single-curved Ledinek press is used as a starting point to implement the two-step lamination process for double-curved glulam blanks. A separate modified press is used to perform the



Figure 4 The industrial production environment at ZAZA Timber Production Ltd.

second pressing step, with the objective to eventually automate both presses and digitize their operation.
The developed workflow is structured into three distinct phases:

- the conversion and reconstruction of design inputs,
- the analysis-driven validation of glulam blank geometry and decision support for material specifications,
- and the extraction of press parameters for production.

The aim is to robustly connect variable design inputs with a large-scale free-form glulam press towards an eventual automation of the preparation and pressing of free-form beams. The digital workflow makes extensive use of the IfcOpenShell (IfcOpenShell, 2025) and RhinoCommon (Baer, 2019) programming APIs.

Geometry reconstruction and rationalization

Early-stage geometry is often ad-hoc, non-standardized, and not appropriate for fabrication, due to uncertainty regarding the final outcome of the design process.

As such, design inputs for the planning and production preparation of glulam beams are usually coarse or otherwise unsuitable, requiring extensive remodelling before manufacturing. Free-form geometry is typically described either as NURBS surface patches or approximated mesh objects. While continuous surface data is ideal for transferring geometric features, limitations of popular exchange formats often hinder their usage, even when supposedly supported by the format. For example, the IFC specification – being related to the STEP format – fully supports structured geometric surface data, however this is rarely fully implemented in the interfaces with different software environments (Fig. 5).

To better understand the nature and variety of input geometry, small batches or individual beams were selected from recent projects. The majority of these were either coarsely modelled meshes with varying resolution – such as an increased density of polygons around details – or full-detailed surface models. Both types of input model benefited from the same approach to estimating the "blank of best fit": the difference chiefly being that the

surface models didn't require reconstructing continuous splines from discretized mesh edges.

To automatically reconstruct a suitable glulam beam model from variable input geometry, an edge-based heuristic method is developed for both mesh-based representations as well as NURBS surface models (Fig. 6). For mesh-based representations, the main additional step is the finding and reconstruction of longitudinal edges of the beam (Fig. 7). The subsequent steps are generally the same for both types of representations.
Reconstructing edges from a mesh model is done as follows:

- mesh edges are found based on their sharpness or angle between corresponding faces
- successive chains of edges within a certain continuity threshold are found
- splines are built through these chains to best approximate them in a curvature-continuous way

Subsequent steps are similar for both mesh and NURBS representations:

- the collection of splines or NURBS edges is evaluated to find the subset that is the most parallel over the longest length, suggesting a common "flow" or axis of movement
- a median curve is generated from this subset, spanning the longitudinal extents of the curve subset
- at regular intervals, a perpendicular plane is constructed on this median curve
- the intersection between the sample plane and the rest of the edge curves is found and the resultant points are used to generate a 2D convex polygon within the sample plane
- a bounding rectangle is fitted to the convex polygon, testing alignment with each edge to find the smallest enclosing boundary
- the resultant rectangle is taken as a cross-section candidate and compared to preceding and succeeding candidates, until the best candidate – i.e., the one that minimizes twist between preceding and succeeding sections – is found
- this step is iteratively refined until some defined threshold is reached
- the axis curve is fitted to the centre points of all resulting cross-sections

The process yields the equivalent of a solid sweep of a rectangle along the axis curve, with a constant cross-section. In similar terms, the resulting geometry creates an enclosing "beam coordinate system" around the input geometry, with one axis being the curved and elongated centreline of the beam. This also ensures that the faces of the resultant beam model are developable.

The resultant *glulam beam model* is an idealized rectangular sweep that must be further processed in order to arrive at the free-form glulam blank model – a more precise geometric representation of the form of the material after pressing, due to the aforementioned geometric constraints and deviations that occur during pressing.

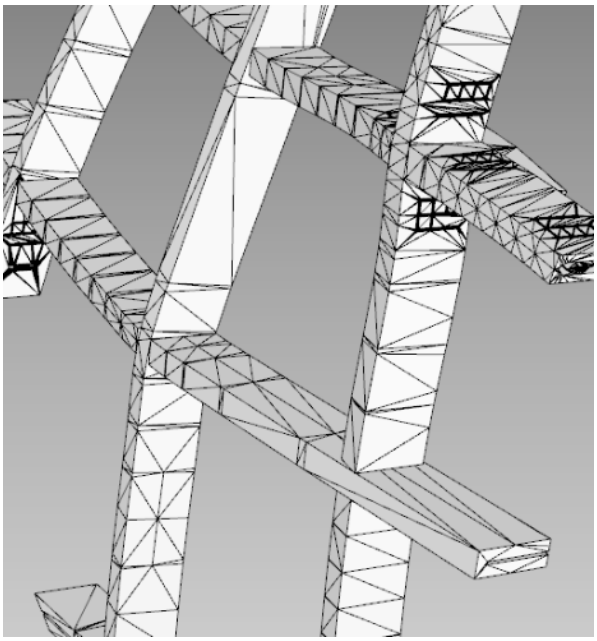


Figure 5 Example of a coarse mesh from an IFC model used for estimation and planning.

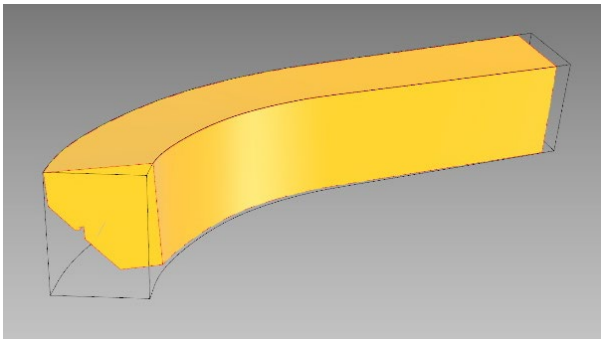


Figure 6 Example of a fitted beam model (black outline) over a detailed NURBS input model.

Once a feasible glulam beam model is acquired, it is decomposed into two phases that correspond to the two pressing steps and their respective constraints. To derive these two phases, the decomposition is done in reverse order of the production process.

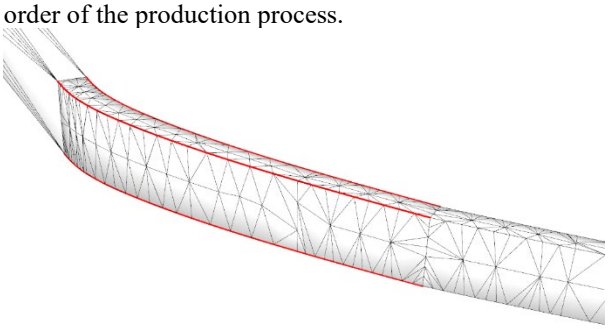


Figure 7 Reconstructed beam edges from a coarse mesh from an IFC model.

Firstly, the pair of beam faces that is least curved is taken as the second pressing step. This is due to the thickness limitations in the resawing of the first pressed form. A median surface between the two faces of the pair is extracted and serves as the neutral axis of the form. Both faces are geodesically unrolled with respect to this median

surface. Overlaying the unrolled faces shows how – due to the double curvature of the beam – achieving a rectangular cross-section is only possible if each layer of the second pressing is different (Fig. 8).

However, since each layer is the result of the same single-curved blank being resawn into layers, each layer is identical and therefore must be expanded so that it envelops both of the unrolled beam faces. This expansion is done by sampling both unrolled faces in relation to the centreline curve of the unrolled median surface to find the farthest point on either side, and offsetting it at the corresponding distance.

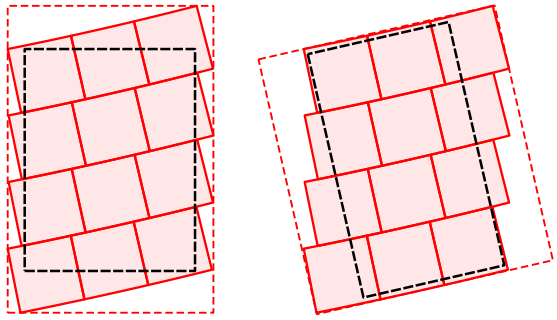


Figure 8 The skewed cross-section means that a rectangular cross-section (black) cannot fit neatly inside of the blank.

The resultant enlarged unrolled shape becomes the required form of the first pressing – in effect, the plan of the single-curved glulam that is to be resawn into layers for the second step (Fig. 9).

The resultant shape is then re-rolled onto the median surface of the glulam beam model, using the relative positions of the original faces. The edges of the resulting faces are connected, forming a closed solid. Depending on the degree of curvature from both pressings, the cross-section of the final glulam blank model becomes trapezoidal (Fig. 10).

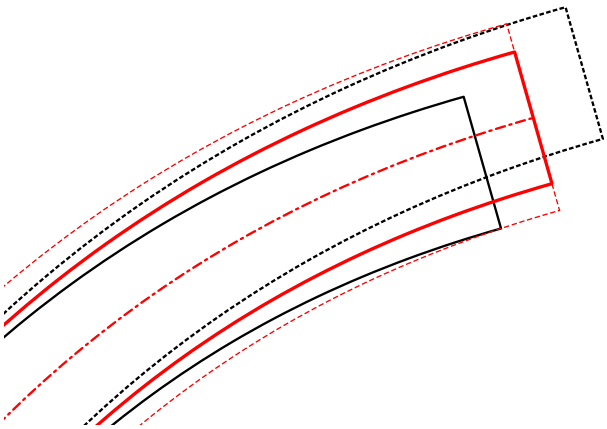


Figure 9 The difference in shape between the top (solid black), middle (solid red), and bottom (dashed black) surfaces when unrolled. A single, widened surface (red dashed) is needed so that the entire beam is encompassed by the re-rolled blank geometry.

Thus, the geometric and process constraints from the two-step pressing result in a particular geometric form that encompasses the original glulam beam model and a more accurate representation of the glulam blank model.

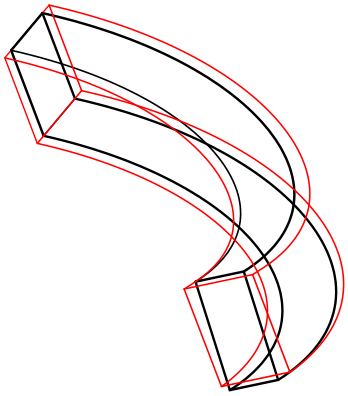


Figure 10 The difference between the beam geometry (black outline) and blank geometry (red outline).

Material parameters and production mechanics

Having arrived at the glulam blank geometry, the next step in the workflow is the selection of material specifications according to performance requirements and material limitations. The relationship between curvature in free-form beams and the lamella dimensions is developed from the typical method used for single-curved beams into a more broadly-applicable simulation-informed validation method. Mechanical tests done in a parallel effort transfer material properties from simpler glulam specimens into generalised design values.

For single-curved glulam beams, lamella thicknesses are related to the maximum curvature the beam, typically expressed as a simple linear ratio between the radius of curvature and the thickness of the lamella. While this provides a straightforward way of arriving at material specifications for planar glulam beams, double-curved – or free-form – beams require additional considerations.

Initially, in the two-step lamination process, curvature was also considered in two parts, leading to a lamella thickness specification for the first pressing and a different one for the second pressing out of plane. While this allows the curvature estimates to be done in a decoupled way, using only the planar curvature at each step, it does not capture the full strain induced on the lamella by the combination of both together, along with torsion.

To overcome this, the curvature limits from single-curved glulam specification are translated into strain coefficients which can then be used as limit values in more complex and three-dimensional scenarios. Bending strain in a beam can be defined by $\varepsilon = \kappa y$ where ϵ is the longitudinal strain, κ is the curvature of the beam (also defined as 1/R, where R is the radius of curvature), and y is the distance from the neutral axis. Since individual lamellas themselves behave like long and slender beams, we assume a Euler-Bernoulli model where bending strain dominates and shear caused by bending is negligible.

Considering the Eurocode ratio of 1:200 between the lamella thickness and the smallest allowable radius of curvature for softwoods (t/R, where t is the lamella thickness and R is the radius of curvature), the maximum induced bending strain for a curved lamella can be 0.0025. Transferring this principle to free-form glulam blanks therefore can be done through a projection of the curvature vector onto the cross-section and its decomposition into width and height values for the lamella dimensions, which is subsequently validated through simulation using the finite element method (FEM). The smallest 3d radius of curvature is found at the innermost lamella edge of the blank. The vector of curvature is decomposed into height and width components relative to the blank cross-section, giving the amount of bending in each direction. The thickness is then found using the established t/R ratio. To introduce a simulation-based validation step in our workflow, finite element analysis is used to simulate the bending of the critical lamella into its final shape in the glulam blank. The maximum longitudinal strain is compared to the strain limit to verify an appropriate dimensioning of the lamella. While this step is still at a preliminary stage in the project, further work will improve both the robustness of the FE model as well as compare the simulation with results of physical testing. This has implications for scenarios where the second pressing step induces substantial torsion on the layers, where shearing across the plane of the layer becomes a significant source of strain. The mechanical testing of lamellas and curved glulam prototypes is an on-going adjacent stream in the project. This parallel effort seeks to also experimentally arrive at generalizable strain limits for different lamella species such as spruce and pine to help with selecting the most appropriate material type and quality for a particular glulam blank.

As a decision-making tool, the lamella material and dimensioning are used to assess the feasibility of a producing a particular glulam blank, as the calculated optimal parameters must nevertheless be compared against process constraints along the glulam production chain. For example, board thickness at the material intake can dictate available lamella sizes, based on whether or not they can be resawn into thinner boards or the degree to which they should be planed. Similarly, the minimum resawing and planing thickness after the first pressing step constrains the possible curvature in the second pressing step, as the size of the first blank is much greater than the individual lamella, necessitating larger resawing and planing equipment. Finally, the spacing of the pressing heads on the glulam blank is strongly related to the thickness of the lamella – thinner lamellas require a closer and more consistent press spacing – however the geometry of the press and its kinematics dictates the possible head layout for a given glulam blank geometry. The ability to confidently identify these constraints and how an incoming glulam beam model stacks up against them is therefore a valuable tool for the assessing and planning of a project.

Production integration and prototyping

The final step in the workflow before production is the generation of positioning data for each of the pressing steps. In an automated workflow, this also includes the position of each linkage in the kinematic chain of the press and the motion planning of the entire assembly.

For the two-step lamination process, the pressing positions need to be generated for the first, single-curved step and then subsequently for the second, out-of-plane bending step. Future work will additionally investigate the feasibility of using the single-step process of bending a free-form glulam directly using smaller, more squarish lamellas.

For the first step, the intermediate or “unrolled” form of the glulam blank model is used. The resultant shape which encloses the unrolled top and bottom faces of the glulam beam model is effectively the 2d representation of the result of the first pressing. This is oriented so that it lies along the longitudinal axis of the single-curved press. For each press track, its linear axis is intersected with the inner – that is, the one closes to the resting position of the press – edge of this shape. The tangent vector of the inner edge at that intersection point yields the angle of the pressing head as well as its position along its track. This therefore gives the full press parameter description for the first step – the single-curved blank.

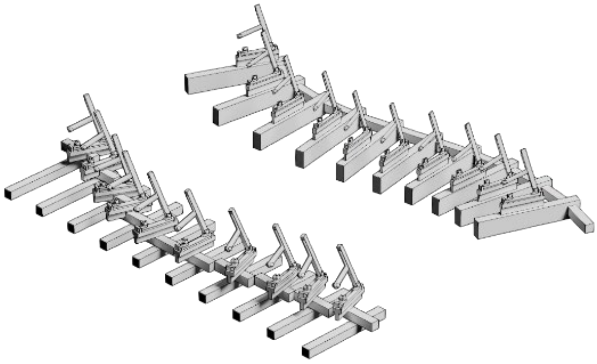


Figure 11 Visualization of the free-form press prototype, showing the modular and articulated pressing heads.

For the second step, a more complex route is required due to the additional degrees of freedom necessary to achieve the out-of-plane curvature. The design of the second press consists of a modified single-curved glulam press which introduces additional press heads that give additional axes of movement and rotation (Fig. 11). For this, a digital kinematic model is constructed consisting of the individual linkages in the press, their axes of rotation or displacement, and relative offsets. The simplicity of the linkage arrangement means that a direct numerical method can be used to solve the inverse kinematic problem instead of an iterative solver, greatly simplifying the computational complexity. For every target press position, the required linkage positions are determined by their offsets and relative rotations. The accuracy of the result is therefore dependent only on the quality of manufacture of the press linkages and the measurement of their true offsets and relative positions. The kinematic

solution is output as a structured dataset consisting of joint rotations and distance values for each press head. These are used to position the press into the required pose for pressing the target form.

The workflow is tested at a reduced scale using the full industrial infrastructure to evaluate the validity of the material specifications and the accuracy of the press data by producing 15 beam prototypes out of spruce lamellas (Fig. 12). Following the methodology, they are first formed as single-curved blanks in a Ledinek press, resawn and planed, and then pressed into the final 3d form with the second modified press. Lamellas thicknesses of 15mm are used for the first step – the minimum thickness driven by the resawing limitations for individual lamellas in the current production line – and 50mm for the second step – limited by the resawing and planing capacity in the current production.



Figure 12 Free-form glulam blank prototypes produced using the digital workflow.

Discussion and future work

The developed workflow demonstrates how inconsistent and coarse design geometry can be translated into a useful glulam blank model so that it can be evaluated against material and process constraints and further used to drive the generation of production data. This effort shows how a typically laborious and craft-based process can be augmented and streamlined through computational methods towards eventually becoming an automated and industrialized process for the construction of curvilinear timber structures. Key data for generating the glulam blank model is extracted or reconstructed from rough design models. Material limits from established standards are integrated into the specification and planning steps, with these being eventually supplemented with experimental data. Production data for the pressing of the glulam blank is derived from the model through geometric and kinematic analysis.

While the accuracy and implementation of the actual physical press infrastructure is outside of the scope of this paper, validation of the physical prototypes is nevertheless seen as a critical path forward in tuning and verifying the digital workflow. To this end, the prototypes have been 3d scanned and compared to their digital beam models to begin preliminary quality assessment of the pressing process (Fig. 13). As a feedback method, this is expected to identify areas of improvement and potentially additional geometric factors to test incoming models against. The result of this validation will primarily affect the generation of press parameters by determining how they must be compensated to better achieve the target form.

Preliminary analyses of the scans show mean deviations of point distance to model geometry between 0.32 and 3.84 mm with standard deviations between 3.42 and 8.14 mm. The mean deviation between beams is 1.29 mm with a standard deviation of 0.92 mm. A greater deviation was observed in the direction of the second pressing. Considering that there were fewer, yet thicker and therefore stiffer, layers in this direction, it is likely that this was caused by greater springback in the material after pressing. The analysis could be improved with more refined registration methods and by observing deviations separately for each step, instead of only for the final result. In 9 of the prototypes, over 90% of scan points were within 10 mm of the beam model, while in 4 prototypes, over 70% of the points were within 5 mm.

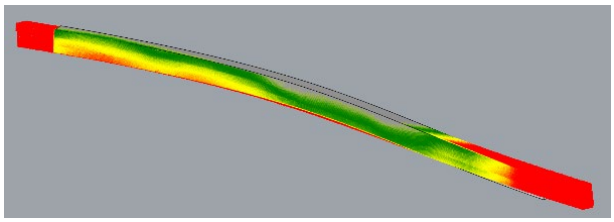


Figure 13 A preliminary comparison between the beam model and scan of a physical prototype, showing areas that lie within a 10mm tolerance (green and yellow) and areas that exceed this tolerance (red), especially at the longer ends.

Future work will also seek to include more nuanced material properties and strain limits based on the results of parallel material testing on both straight lamellas as well as free-form laminated prototypes, to arrive at a reliable material model that is tuned to the particular species and geometry type.

Alongside these plans for the digital workflow from model to press data, we plan for further automation of the press infrastructure itself and therefore using the results of the digital workflow directly in the actuation of motors to drive the press positions. The ability of the second press to achieve three-dimensional forms also shows potential in being used for the one-step lamination process with more squarish lamellas, bypassing the need for two pressing steps. The design and manufacture of free-form glulam elements using either method can therefore be achieved in a seamless manner, avoiding the need for laborious planning and manual work.

Regarding the wider positioning of this effort within the digital design chain, the sussing out of these material specifications and production parameters can serve as feedback mechanisms for evaluating designs at early stages. By relating geometric properties to specific

barriers in production, challenging beams could be identified and revised or else directed towards alternative fabrication methods. For example, pressing constraints – distances between the pressing heads, track sizes, and so on – imply that particular forms might exceed the capacity of a particular setup. This suggests a diversity of press configurations – some finer, some larger and heavier – to deal with a variety of input geometries as opposed to a "one size presses all" solution.

Along the same line of thinking, the issue of tolerances touched upon before also raises the idea of adaptive or performance-driven pressing, where tolerances are tightened around geometrically or structurally critical areas of the beam while relaxed elsewhere. This would enable a more efficient allocation of effort and energy, with implications not only for production time and cost but for material use and sustainability metrics.

Conclusion

We envision a fully automated and industrialized process for fabricating free-form glulam beams from coarse input data in typical architectural workflows.

The described work is positioned as a contribution to digital construction and industrialised timber fabrication, focusing on the design-to-production information workflow required to realise free-form glulam structures under real-world industrial constraints. The work demonstrates how computational modelling, automated rationalisation, and analysis-informed validation can be integrated into a robust, end-to-end digital pipeline that translates imperfect design input into production-ready manufacturing data. It foregrounds automation, data robustness, and scalability, addressing a key challenge in industrialized timber construction: the reliable manufacture of complex architectural intent in free-form structures. In doing so, it demonstrates how materially-aware computation in construction processes can lead to increased industrial efficiency, a greater flexibility and feasibility of manufacturing outcome, and new potentials in achievable architectural expression.

Acknowledgments

In accordance with the contract No. 1.2.1.2.i.2/1/24/A/CFLA/001 between "Forest Sector Competence Centre of Latvia" Ltd. and the Central Finance and Contracting Agency, the study is conducted by "ZAZA TIMBER Production" Ltd. With support from the European Regional Development Fund (ERDF) within the framework of the "Forest Sector Competency Centre" project ZP06 – "Development of bent glued timber manufacturing technology and prototype equipment enabling modelling and design of large-scale free-form timber structures".

References

Baer, S. (2019). What is RhinoCommon? https://developer.rhino3d.com/guides/rhinocommon/what-is-rhinocommon/

Chai, H., So, C., & Yuan, P. F. (2021). Manufacturing double-curved glulam with robotic band saw cutting technique. Automation in Construction, 124(1239), 103571. https://doi.org/10.1016/j.autcon.2021.103571

IfcOpenShell (Version 0.8.4). (2025). [Computer software]. https://ifcopenshell.org/

Menges, A., Schwinn, T., & Krieg, O. D. (Eds). (2016). Advancing wood architecture: A computational approach. Taylor & Francis. https://doi.org/10.4324/9781315678825

Scheurer, F. (2012). Digital Craftsmanship: From Thinking to Modeling to Building. In S. Marble (Ed.), Digital Workflows in Architecture (pp. 110–129). Birkhäuser.

Stehling, H., Scheurer, F., Geglo, H., & Hofmann, M. (2017). From Lamination To Assembly: Modelling the Seine Musicale. In A. Menges, B. Sheil, R. Glynn, & M. Skavara (Eds), Fabricate 2017 (pp. 258–263). UCL Press.

Stehling, H., Scheurer, F., & Roulier, J. (2014). Bridging the gap from CAD to CAM: Concepts, caveats and a new Grasshopper plug-in. In F. Gramazio, M. Kohler, & S. Langenberg (Eds), Fabricate: Negotiating design & making (pp. 52–59).

Svilans, T. (2021). GluLamb: A toolkit for early-stage modelling of free-form glue-laminated timber structures. Proceedings of 2021 European Conference of Computing in Construction (2021 EC3).

Svilans, T., Tamke, M., Ramsgaard Thomsen, M., Runberger, J., Strehlke, K., & Antemann, M. (2019). New workflows for digital timber. In F. Bianconi & M. Filippucci (Eds), Digital Wood Design: Innovative Techniques of Representation in Architectural Design (pp. 93–134). Springer. https://doi.org/10.1007/978-3-030-03676-8_3

Weinand, Y. (Ed.). (2017). Advanced timber structures: Architectural designs and digital dimensioning. Birkhäuser.

Zhang, C., Franco, L., Spencer, L., & Zivkovic, S. (2023). TwistedLam: A Robotic Fabrication Method for Hyperbolic Paraboloid Glulam and Clt Blanks. 281–290. https://doi.org/10.52842/conf.caadria.2023.2.281

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

TELEOPERATION INTERFACES FOR COMPACT DEMOLITION ROBOTS: RESULTS OF A PILOT STUDY WITH CONSTRUCTION WORKERS

Patrick Borges Rodrigues, Zihao Wang, Yiyang Fang, Gale Lucas, Shawn Roll, Burcin Becerik-Gerber, and Lucio Soibelman
University of Southern California, Los Angeles, United States of America

Abstract

As teleoperated robots become more common on construction sites, human-robot interfaces must better support operators' sensory needs. This paper presents results from a pilot study comparing a proposed sensory-enhanced teleoperation interface for compact demolition robots with a standard teleoperation interface. Recommendations from demolition stakeholders informed the design of the sensory-enhanced interface, while the standard interface was based on commercially available teleoperated construction equipment. The goal of this study was to refine the sensory-enhanced interface based on user feedback. Ten construction workers tested the interfaces, and we measured their perceived usability, telepresence, and trust in technology.

Introduction

Increasing numbers of robots with varying levels of autonomy are being deployed at construction sites (Liang et al., 2021). Among these, teleoperated systems are used extensively because they keep workers out of hazardous environments. These technologies fundamentally change construction site dynamics as workers adapt to these machines. However, despite their adoption, many teleoperated systems have created a mismatch between workers' needs and the control requirements of human-machine interfaces (Kim et al., 2026).

This mismatch can result in increased mental workload, decreased situation awareness, ergonomic issues, and compromised performance and safety (Kim et al., 2026). In teleoperated contexts, interfaces aligned with operators' preferences may reduce physical and mental loads, increase operator comfort, performance, and safety. Despite these benefits, teleoperation interfaces still face multiple issues that complicate their broader use (Rea & Seo, 2022), and better user-centered interfaces for construction robots are needed (Rasheed et al., 2024).

In this paper, we compare two teleoperation interfaces for compact demolition robots: a standard and a sensory-enhanced interface. The standard interface was designed based on construction robots' teleoperation interfaces (e.g., Caterpillar Inc. (2022)). The sensory-enhanced interface was designed based on recommendations from demolition stakeholders to extend the capabilities of traditional systems (Rodrigues, Lucas, et al., 2025).

We conducted a pilot study to test the two interfaces and ensure participants could safely complete various tasks in an obstacle course. We gathered participants' subjective feedback on perceived usability, telepresence, and trust in technology. We also conducted semi-structured interviews to collect their opinions and suggestions on interface designs. These results helped refine the interfaces, the obstacle course, and the experimental protocol before the full study with a larger sample.

Literature Review

Teleoperation is widely used in construction robotics (Liang et al., 2021). Applications include teleoperated drones for site inspection, safety surveillance, and surveying (Albeaino et al., 2022); teleoperated heavy equipment (e.g., excavators, bulldozers) for hazardous and remote sites (J. S. Lee et al., 2022); and teleoperated demolition (Rodrigues, Becerik-Gerber, et al., 2025), among others (Hu et al., 2025). This widespread adoption stems from the unstructured and dynamic conditions of most construction sites, which still limit fully autonomous operations and require human perception and decision-making to interpret site conditions, assess risks, and adjust operations in real time (Kim et al., 2026). As a result, the effectiveness of teleoperation in construction depends on the quality of the operator-interface interaction and the availability of adequate sensory feedback.

Commercial teleoperation interfaces in construction commonly replicate manned equipment interfaces (e.g., Caterpillar Inc. (2022)). Operators use seats, joysticks, and pedals modeled after those found on manned machines (e.g., excavators), with camera views replicating cabin perspectives. While this approach leverages operators' familiarity with traditional systems, directly transferring interfaces from manned to teleoperated machines is not straightforward, as physical separation between operators and the site in teleoperation affects the quality, quantity, and type of sensory information operators receive (J. S. Lee & Ham, 2022). For equipment not operated from a cabin, this approach is infeasible, requiring the redesign of the human-machine interfaces rather than transferring traditional control schemes (e.g., H. J. Lee & Brell-Cokcan (2023)).

Over time, however, these interfaces are often improved through trial and error. Advancements in artificial intelligence (AI) and hardware have enabled assistive teleoperation technologies that support higher levels of

robot autonomy, such as robots that adjust behavior to reduce operator workload (e.g., Liu et al. (2021)). AI has also enabled assistive features that fuse and simplify data, identify hazards, and support decision-making (Hong et al., 2020). Similarly, adaptive interfaces adjust the type, amount, and timing of the information operators receive, such as allowing manual or autonomous camera viewpoint switching (Xu et al., 2024). Many of these systems, however, are in early stages of development and have only been tested under laboratory conditions.

Introducing new robotic systems on construction sites has emphasized the need for user involvement in interface design. User Experience (UX) design has been instrumental in this process, supporting ideation, design, and refinement. By recognizing these needs, Rodrigues, Lucas, et al. (2025) proposed a framework to inform the design of sensory-enhanced teleoperation interfaces for construction machines through user engagement and testing. By engaging demolition professionals, the authors identified how operators in traditional demolition tasks use different sensory channels to monitor the robot, task, and environment, and gathered their suggestions for teleoperation interface designs. Drawing on these suggestions, we designed a sensory-enhanced teleoperation interface and tested it against a standard teleoperation interface typical in construction.

Methodology

Robot Description

This study employed a Brokk 120 DII demolition machine equipped with a Brokk BHB 155 hydraulic breaker (Figure 1). It is one of the most compact Brokk Inc. demolition machines, with dimensions of 2030 mm × 780 mm × 1250 mm and a maximum reach of up to 4300 mm.

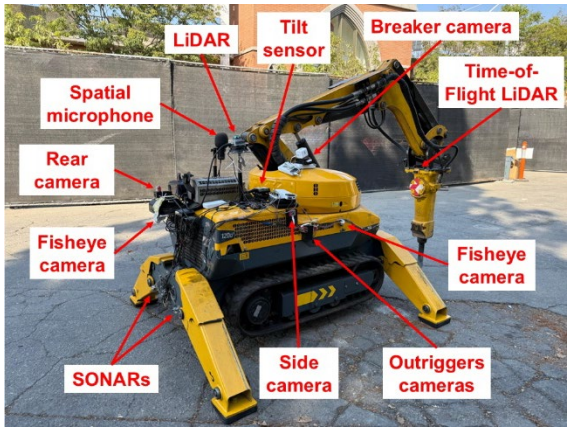


Figure 1. Brokk 120 DII sensor configuration.

Teleoperation System Architecture

Modularity

The teleoperation interfaces used modular designs to provide flexibility for testing and refinement. This approach allowed interface features to be added, removed, or adjusted based on user feedback. Various sensors connected to independent onboard computers and microcontrollers collected operational and site data, which was transmitted to a central computer that aggregated and relayed it via the user interfaces.

Communication Architecture

Communication between the robot and the teleoperation interfaces was enabled via a Wireless Local Area Network (WLAN) established with a single high-performance router in the obstacle course. The system combined the User Datagram Protocol (UDP) and the Web Real-Time Communication (WebRTC) protocol to ensure low-latency communication between the robot's onboard computers and the web-based interfaces.

The robot's onboard system included multiple Raspberry Pi 5 microcontrollers, wirelessly connected to a central minicomputer located inside the construction trailer, which aggregated data from the Raspberry Pi units and generated the web-based interfaces. A separate minicomputer installed on the robot, physically connected to a LiDAR sensor and a spatial microphone, provided real-time point clouds of the environment and environmental audio to the user interfaces. Finally, exocentric cameras installed throughout the obstacle course streamed external visual perspectives of the environment directly to the central minicomputer.

Control

The Brokk 120 DII demolition robot was controlled using the Brokk SmartRemote™ controller (Figure 2). This controller operates via radio communication and can control the robot at distances of up to 300 meters.

A deliberate design decision was made to maintain the robot's original hardware and software. Rather than modifying the robot's control system, an effort that would require an additional investigation, this study focused on augmenting operator situation awareness through external sensors and improved teleoperation interfaces.

Robot Sensors Configuration

Additional sensors were mounted on the robot and integrated with the teleoperation interface (Figure 1). Cameras were installed to provide egocentric views of the robot's front, rear, left, and right sides, its arm, and end-effector. Additionally, a bird's-eye view of the robot was provided using four fisheye cameras, offering operators an overhead perspective of its position and surroundings.

Several sensors were deployed to augment operators' spatial awareness. A tilt sensor measured the robot's pitch and roll. A LiDAR provided real-time point clouds of the environment, improving depth perception. Four robot-mounted sonar sensors measured distances to nearby objects as part of a proximity warning system for collision avoidance. A time-of-flight LiDAR measured distances from the breaker tip to demolition targets. Finally, a spatial microphone transmitted real-time environmental audio from the site to the teleoperation interfaces.

External sensors configuration

Several exocentric cameras were deployed throughout the obstacle course to provide external views of the robot to support localization and situation awareness. Two cameras were installed inside the trailer to capture the operators' postures and their usage of the interface.

User Interfaces

The user interface consisted of multiple TV screens, laptops, the Brokk SmartRemote™ controller, and a VR gaming chair equipped with vibration and rotation capabilities (Figure 2). The specific configuration of interfaces differed between the standard and sensory-enhanced interfaces. The user interface components were installed inside a construction trailer specifically adapted to accommodate the teleoperation interfaces and associated equipment. For the pilot study, this trailer was positioned adjacent to the obstacle course.

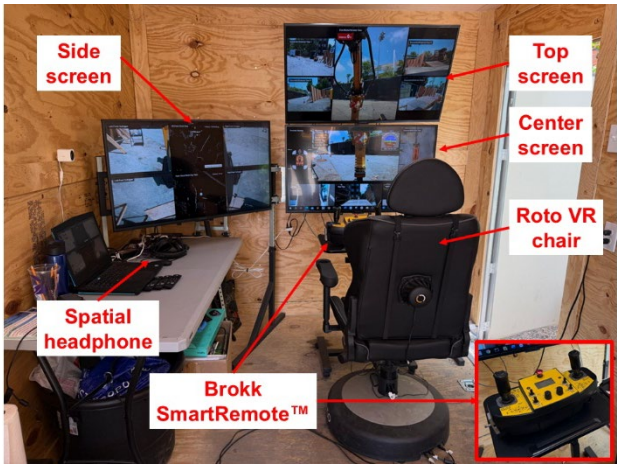


Figure 2. User interfaces inside the construction trailer. The inset shows a close-up of the Brokk SmartRemote™ controller.

Standard Teleoperation Interface

The standard teleoperation interface was designed based on the Caterpillar Cat® Command teleoperation interface (Caterpillar Inc., 2022), commonly used for heavy machinery such as excavators. This interface provided operators with multiple egocentric camera perspectives (i.e., front, rear, left, and right), exocentric camera perspectives, robot status information, environmental sound, and attitude data. In addition to the conventional design of the Caterpillar Cat Command interface, this implementation included a fixed camera to provide a continuous view of the robot's arm and end-effector.

Visual Feedback

Visual feedback (Figure 3) was presented on two vertically stacked 43-inch TVs. The center screen (Figure 3a) displayed four egocentric views: front, rear, left, and right. For the front and rear cameras, the interface also displayed grid lines to help operators estimate the robot's distance to objects. It also included an attitude indicator for the robot's roll and pitch, along with status information such as engine temperature, fuel levels, hydraulic system pressures, and maintenance indicators. The top screen (Figure 3b) presented four exocentric camera perspectives showing the obstacle course and an egocentric view of the robot's arm and breaker. All cameras used in the standard interface were fixed cameras with automatic focus.

Sound Feedback

The sound feedback system consisted of a headset and TV speakers, through which sounds captured by a microphone mounted on the robot were transmitted.



(a)



(b)

Figure 3. Standard teleoperation interface. (a) Center screen. (b) Top screen.

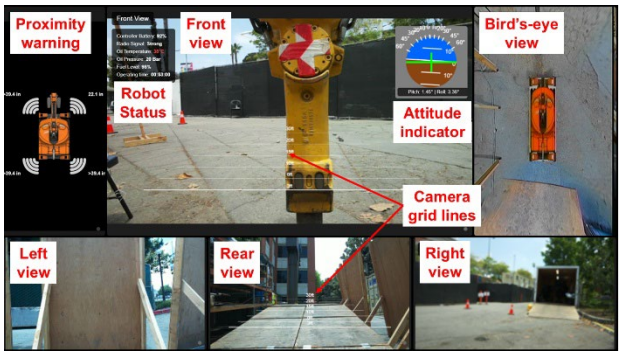
Sensory-Enhanced Teleoperation Interface

The sensory-enhanced teleoperation interface expanded on the standard interface by incorporating enhanced visual, auditory, and haptic feedback, alongside a proximity warning system to improve operator situational awareness. These inclusions were based on recommendations from demolition professionals during focus groups and preliminary testing of VR interfaces in a demolition tradeshow (Rodrigues, Lucas, et al., 2025).

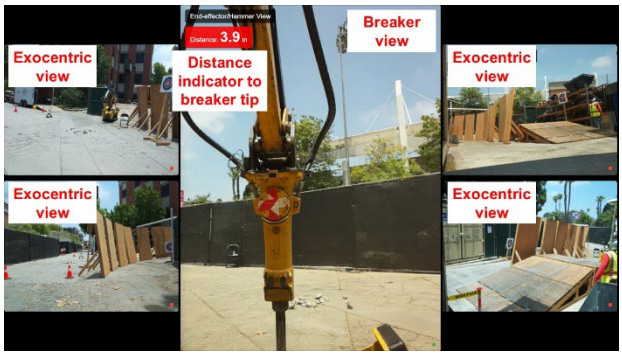
Visual Feedback

The sensory-enhanced interface visual feedback (Figure 4) was presented through three 43-inch screens. The center screen (Figure 4a) displayed four egocentric robot views: front, rear, left, and right. The front and rear views included grid lines to help operators estimate distances. It also showed the robot with proximity warning arcs and distance indicators at each corner, providing real-time visual feedback from four sonar sensors. The screen also included a stitched bird's-eye view of the robot's surroundings from four fisheye cameras placed on the robot's sides, an attitude indicator for pitch and roll, and robot status information, including engine temperature, fuel levels, hydraulic pressures, and maintenance needs.

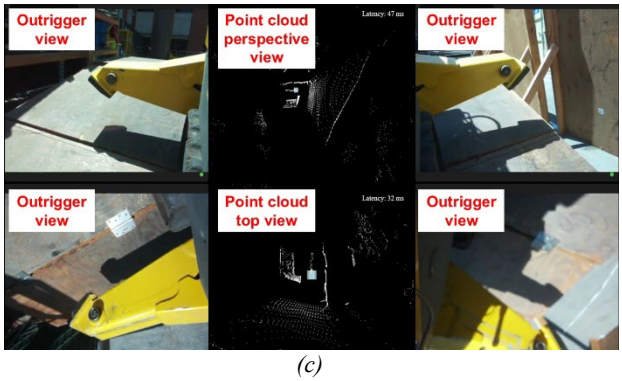
The top screen (Figure 4b) displayed four exocentric views of the obstacle course, an egocentric view of the robot's arm and breaker, and a distance indicator to the breaker tip. Finally, the side screen (Figure 4c) displayed four dedicated camera views, one for each robot outrigger. This screen also displayed two LiDAR point clouds (i.e., top and perspective views) to enhance operator depth perception and situation awareness. All cameras used in the sensory-enhanced teleoperation interface were fixed and equipped with automatic focus.



(a)



(b)



(c)

Figure 4. Sensory-enhanced teleoperation interface. (a) Center screen. (b) Top screen. (c) Side screen.

Sound Feedback

Sound feedback was delivered via a headset and the TV speakers, providing real-time environmental sound captured by a robot-mounted microphone. In addition, the sensory-enhanced teleoperation interface incorporated auditory alerts as part of the proximity warning system.

Haptic Feedback

The sensory-enhanced interface incorporated haptic feedback via a VR gaming chair (Roto VR chair) that vibrated (rumble) and rotated. Haptic feedback was used in the proximity warning system to provide vibratory alerts when the robot approached an obstacle.

Safety Mechanisms

The sensory-enhanced teleoperation interface included a proximity warning system using robot-mounted ultrasonic sensors to alert operators to potential collisions. Visual and auditory feedback were provided when obstacles were within 2 feet, while haptic feedback was activated at distances of 1 foot or less, with warning levels and frequencies increasing as distance decreased.

Additionally, the LiDAR point cloud highlighted moving objects such as people, vehicles, and equipment in red, thus enhancing operator awareness of dynamic elements.

Pilot Study

The objective of the pilot study was to identify opportunities for improvement before a full user study. The pilot also served to test the sequencing and timing of training and experimental tasks, verify the setup and functionality of the teleoperation systems, and ensure the safety of the obstacle course. By identifying potential challenges and opportunities before the user study, we aimed to minimize technical issues, enhance data validity, and avoid mid-study modifications that could introduce bias and compromise data comparability.

After providing informed consent, participants completed intake surveys that included demographics and baseline trust in technology and self-efficacy. We also recorded five-minute baselines for heart rate variability (HRV) and electrodermal activity (EDA) using an Empatica EmbracePlus smartwatch. Participants then underwent VR-based training in the Unity Game Engine, covering the robot's parts and controls, collision avoidance, worker safety, and safe operation on elevated and sloped surfaces. The VR training included three guided exercises focused on driving, arm positioning, and breaking concrete. Next, participants completed pre-experiment surveys assessing trust in technology, self-efficacy, state-trait anxiety, fatigue, musculoskeletal discomfort, and workload.

Next, participants received in-person training with the real robot, covering the Brokk SmartRemote™ controller, safe driving, arm operation, concrete breaking, and robot configuration for operations on ground and elevated surfaces. Then, participants went to the construction trailer to be introduced to their assigned teleoperation interface (standard or sensory-enhanced), with the researchers explaining its functions and features.

The experimental trial involved remotely operating the Brokk robot through a three-task obstacle course (Figure 5). The tasks included (1) driving in a flat open space while avoiding objects such as cones, construction tools, and concrete debris, and positioning the breaker on a target located above the robot, (2) driving in a constrained corridor while avoiding objects such as walls, utility pipes, and concrete beams, positioning the breaker on two targets located at the height of the robot, and demolishing part of a concrete beam, and (3) driving on sloped surfaces (ramps) and elevated surfaces (platform) while avoiding objects such as corridor walls, construction crane, and shelves, and positioning the breaker on a target located above the robot while the robot is on a sloped surface.

Throughout the trial, participants wore the Kexxu Eye eye-tracker to record gaze behavior and the Empatica EmbracePlus smartwatch to record HRV and EDA. After the trial, participants completed post-experiment surveys assessing trust in technology, self-efficacy, state-trait anxiety, fatigue, musculoskeletal discomfort, workload, situation awareness, telepresence, and system usability. Finally, they participated in a semi-structured exit interview to provide feedback on the interfaces.

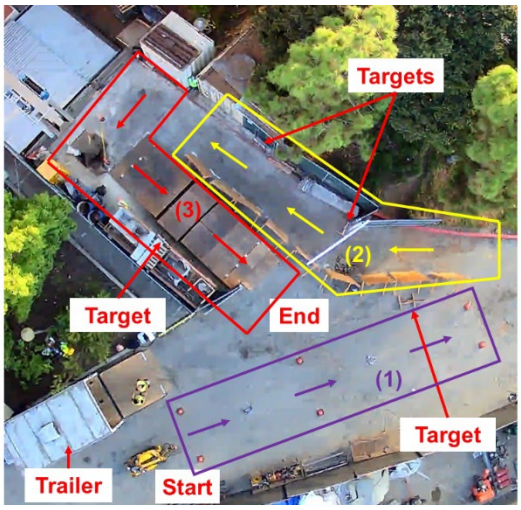


Figure 5. Obstacle course for the pilot study. (1) Open space. (2) Corridor. (3) Ramp.

Participants also completed a survey evaluating workstation features across different site conditions and tasks. Finally, participants assigned to the sensory-enhanced teleoperation interface completed a scenario-based evaluation, in which they read descriptions of demolition scenarios with varying site complexity levels and rated the importance of different sensory features under those conditions. The study protocol was approved by the Institutional Review Board (IRB) at the University of Southern California (USC) (Study ID: UP-24-01122).

Participants

We recruited 12 construction workers to test the two interfaces (*Table 1*). The first two participants were dismissed from the study due to technical issues with the WLAN and were not included in subsequent analysis. Participants were required to be construction workers fluent in either English or Spanish. Recruitment efforts included outreach to construction and demolition industry partners and site visits to construction projects at the USC University Park Campus (UPC). In addition, we employed snowball sampling by asking participants to refer the study to their colleagues.

Participants received compensation for their participation. Those who tested the standard interface completed a session that lasted up to 3.5 hours and received up to $380. Participants who tested the sensory-enhanced teleoperation interface completed a session that lasted up to 4.5 hours and received up to $450. The difference in duration and compensation was due to the scenario-based evaluation at the end of the experiment for the participants assigned to the sensory-enhanced teleoperation interface.

Pilot Study Results

Due to the small sample size of the pilot study, we did not perform formal statistical tests and group comparisons. Rather, we present descriptive statistics on the metrics for the effectiveness and usefulness of the interfaces. We also discuss our observations during the pilot study, the participant recommendations, and how they helped us refine the interfaces, the obstacle course, and the experimental procedures for subsequent studies.

Table 1. Demographic information of the participants (n=10) by experimental condition.

Characteristic	Category	Standard Interface	Sensory-Enhanced Interface
Gender	Male Female	5 0	4 1
Age	18-29 30-39 40-49	2 1 2	2 2 1
Preferred Language	English Spanish	4 1	5 0
Play Video Games	No Yes	4 1	1 4
Work Experience in Construction	None < 5 years 5-10 years	0 1 4	1 3 1
Demolition Experience	None < 5 years	3 2	4 1
Construction Machine Operation Experience	None < 5 years 5-10 years	2 2 1	4 1 0
Construction Machine Type	Excavator Crane Forklift Aerial lift	1 1 2 2	0 0 1 0
Teleoperation Experience	No Yes	5 0	5 0

Note: Totals may not equal 10 because participants could skip questions or select multiple choices.

Trust in Technology

Trust in technology was assessed at three time points: during the pre-VR training session, pre-experiment, and post-experiment. We used a modified version of the Trust Scale with 12 items (Jian et al., 2000), in which participants rated their trust, distrust, and familiarity with their assigned teleoperation interface. First, we reverse-coded the distrust items (items 1-5). Then, we used the average scores across the 12 items of the scale to determine each participant's trust in the teleoperated Brokk robot system. Final scores were in the range of 1 (not at all) through 7 (extremely). Figure 6 presents the means and standard deviations of the trust scores for each interface type and measurement time.

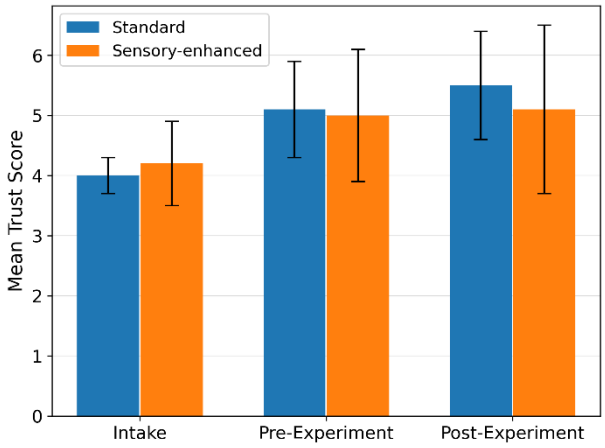


Figure 6. Mean trust scores by interface type and measurement time. Error bars represent ±1 standard deviation.

Usability of the Teleoperation Interfaces

The usability of the two interfaces was assessed using the System Usability Scale (SUS) (Brooke, 1996), a 10-item questionnaire in which participants subjectively rated each interface's usability, yielding final SUS scores ranging from 0 to 100. In usability testing, a commonly used threshold for SUS scores is 68 (Lewis & Sauro, 2018). Scores above 68 indicate above-average usability, while scores below 68 indicate below-average usability. Figure 7 presents the means and standard deviations of SUS scores for each interface type.

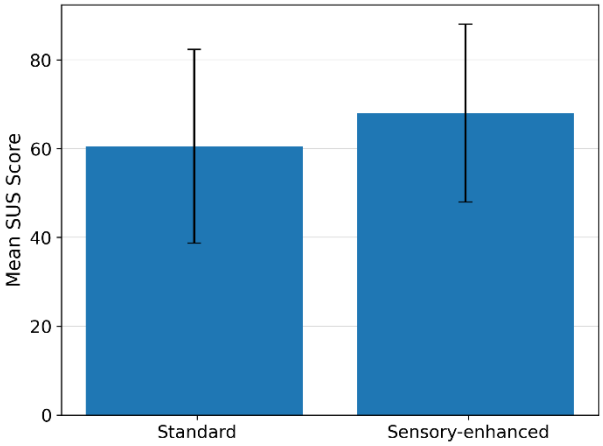


Figure 7. Mean SUS scores by interface type. Error bars represent ±1 standard deviation.

Telepresence

We used the 24-item Presence Questionnaire (PQ) modified by Robillard et al. (2002), based on the original instrument by Witmer & Singer (1998). Of the 24 items, we used 22, excluding the two optional haptic-sensation items, which were not applicable to our interfaces. We computed scores for overall presence (Total Presence) (items 1–19) and for the following presence subscales: Realism (items 3, 4, 5, 6, 7, 10, 13), Possibility to Act (items 1, 2, 8, 9), Quality of Interface (items 14, 17, 18), Possibility to Examine (items 11, 12, 19), Self-Evaluation of Performance (items 15, 16), and Sound (items 20–22). Items 14, 17, and 18 were reverse-coded before analysis. For each variable, scores were computed by summing the responses to the corresponding items and then normalizing by the number of items, so that all variables are in the 1-7 scale. Figure 8 presents the means and standard deviations of total presence scores and presence subscales scores for each interface type.

Refinements of the Interfaces

During the experiment and the exit interviews, participants provided feedback on the interfaces, training, and their experiences. Across both groups, users reported an initial learning curve to get accustomed to the controls. Sensory-enhanced interface users commonly reported improvements in usability over time, with one participant describing the interface as "very self-explanatory" and quick to learn. Conversely, standard interface users frequently reported persistent challenges with depth perception and camera-based navigation, particularly in narrow spaces or when approaching targets.

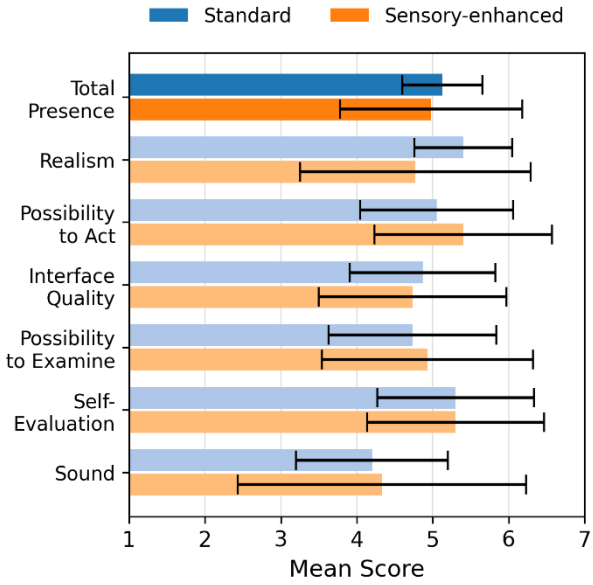


Figure 8. Mean total presence and presence subscale scores by interface type. Error bars represent ±1 standard deviation.

Both groups identified depth perception as a key limitation of the interfaces and suggested additional depth cues and visual overlays. Sensory-enhanced interface users reported positive reactions to the additional sensory feedback (e.g., extra views, sound, vibration), but some found it cognitively demanding to process information from multiple screens simultaneously. Similarly, standard interface users suggested additional visual representation of the arm's position to aid manipulation. Regarding training, participants from both groups valued the VR-based training, noting that it improved confidence during real robot operations.

When asked about interface features, some sensory-enhanced interface users reported ignoring or underutilizing certain features (e.g., point cloud and outrigger views), suggesting that these features require better integration or explanation. Several participants recommended adding automated tilt warnings, collision avoidance, and automated stops, especially for complex arm positioning tasks. While the interface's physical layout was generally well received, some participants from both groups noted that the number of screens could be overwhelming and suggested simplifying the interface to improve operational efficiency. In general, participants felt that the interface improved safety, but noted that its complexity and cognitive demands could reduce productivity and require extensive training due to the number of camera feeds and data streams presented.

Discussion

In this pilot study, we observed that participants' trust in the teleoperation systems increased over time, consistent with expectations as they gained familiarity through VR training and real-world robot operation, such as in Adami et al. (2022). Participants initially reported neutral trust levels prior to interface exposure (intake), which increased following VR-based training (Pre-Experiment) and further after operating the real robots (Post-Experiment). Importantly, this increase was observed for

both teleoperation interfaces. These findings suggest that trust development may be more strongly influenced by interaction time and user experience than by interface design alone, at least in early stages of system exposure.

Regarding SUS scores, both interfaces were relatively close to the industry average of 68 (Lewis & Sauro, 2018), suggesting room for improvement. Another aspect of the pilot study that supports the usability of both interfaces is that all ten participants, five per interface, completed all three obstacle course tasks.

The presence scores suggest that participants experienced moderate to high levels of presence across both interfaces, with mean total presence scores in the upper half of the possible range. The presence subscales showed a similar pattern, with realism and possibility to act having relatively high scores, suggesting that participants felt able to interact with and influence the system. Next, interface quality and possibility to examine were slightly lower, but still above the midpoint of their respective score ranges, suggesting some limitations in these areas. Finally, the sound subscale yielded average scores near the middle of the possible score range, suggesting a moderate contribution of auditory feedback to participants' overall sense of presence for both interfaces.

Participant feedback provided valuable insights that informed refinements to both the teleoperation interfaces and the experimental protocol. As documented in prior teleoperation research, depth perception has been recognized as a persistent challenge, underscoring the need for enhanced visual cues to support spatial awareness (Chen et al., 2007; Luo et al., 2021). The sensory-enhanced interface was generally perceived as more supportive of safety, but some participants also noted it might have induced higher cognitive load, particularly due to the greater number of screens and data streams compared to the standard interface.

Participants also provided valuable suggestions for improving the interface layouts. To better align with participants' needs, we moved the end-effector and exocentric views, which participants identified as more critical for task performance, from the side screen to the top screen for improved visibility. Consequently, the side screen in the sensory-enhanced interface displayed the point cloud and outrigger views. Additionally, the initial laptop screen showing robot status information (temperature, fuel levels, hydraulic pressures, maintenance indicators) and proximity warnings was eliminated, and this information was integrated into the center screen alongside the egocentric and bird's-eye views. These adjustments directly addressed participants' concerns about information overload and interface complexity while preserving situational awareness.

Along with these visual interface changes, we also modified the audio feedback system to improve safety. Originally, both interfaces provided sound feedback through spatial audio headphones. However, the headphones were later replaced with TV speakers because participants struggled to hear the researchers' instructions and safety alerts, which could lead to accidents.

This pilot study had several limitations. First, the sample size was small (N = 10), limiting statistical power to detect differences between interface types and increasing the risk of Type II errors. Second, although all participants were construction workers, their experience with construction machine operations varied widely, potentially influencing their perceptions of usability, trust, and presence, and their task performance. Third, participants used the interfaces for a limited time during the pilot test, and longer-term use may change their perceptions of usability, trust, and presence. Additionally, while eye-tracking data were collected during the study, intermittent internet connection issues caused communication failures between the cellphone app and the eye-tracker, leading to calibration and tracking losses and video corruption. These issues made eye-tracking data unreliable during analysis. However, the issue was resolved by stabilizing the data acquisition setup for subsequent studies. Finally, while this analysis focused on subjective self-reported measures of trust, usability, and presence, further analyses of objective task performance and gaze behavior are necessary to provide a more comprehensive evaluation of interface effectiveness.

To address these limitations, in subsequent full-scale studies, we will recruit more construction workers to robustly analyze how different interface designs affect operators' performance and safety, using both objective and subjective measures. We will also conduct moderation and mediation analyses of the data to examine how demographic characteristics (e.g., age, machine operating experience) and operator responses (e.g., workload) explain observed performance and safety outcomes.

Conclusions

This pilot study provided valuable preliminary insights into usability, trust development, and presence afforded by two teleoperation interfaces for compact demolition robots. Both the standard and sensory-enhanced interfaces supported increasing trust over time, and presence scores were consistent with established reference values in the teleoperation and VR literature. Similarly, SUS scores for the initial designs of both interfaces were rated similarly to those of interfaces in industry and academia.

Next, participants' feedback highlighted specific areas for interface refinement, including improved depth cues and interface layouts. Importantly, iterative adjustments to the interface configuration and experimental protocol were made based on this feedback, highlighting the value of the selected user-centered approach in designing teleoperation systems that better support operator needs.

While results should be interpreted with caution due to sample size and other limitations, this work represents an important first step toward developing teleoperation systems that enhance operator safety, usability, and situation awareness in demolition. Moving forward, integrating objective performance data and involving a broader range of experienced operators will be key to validating and extending these initial findings.

Acknowledgments

This work was supported by the U.S. National Science Foundation (NSF) under Grant No. 2222572 and the U.S. Army Research Office under Grant No. W911NF-20-2-0053. Any opinion, content, or information presented does not necessarily reflect the positions or the policies of NSF or the U.S. Army Research Office, and no official endorsement should be inferred.

References

Adami, P., Rodrigues, P.B., Woods, P.J., Becerik-Gerber, B., Soibelman, L., Copur-Gencturk, Y., & Lucas, G. (2022). Impact of VR-Based Training on Human–Robot Interaction for Remote Operating Construction Robots. Journal of Computing in Civil Engineering, 36(3): 04022006, pp. 1–15.

Albeaino, G., Gheisari, M., & Issa, R.R.A. (2022). Human-Drone Interaction (HDI): Opportunities and Considerations in Construction. In: Automation and Robotics in the Architecture, Engineering, and Construction Industry. Cham, Switzerland, Springer International Publishing, pp. 111–142.

Brooke, J. (1996). SUS: A "Quick and Dirty" Usability Scale. In: Usability Evaluation In Industry (1st ed.). London, England, CRC Press, pp. 189–194.

Caterpillar Inc. (2022). CAT Command: Remote Control Console and Station. Caterpillar Inc.

Chen, J.Y.C., Haas, E.C., & Barnes, M.J. (2007). Human Performance Issues and User Interface Design for Teleoperated Robots. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 37(6), pp. 1231–1245.

Hong, Z., Zhang, Q., Su, X., & Zhang, H. (2020). Effect of virtual annotation on performance of construction equipment teleoperation under adverse visual conditions. Automation in Construction, 118: 103296.

Hu, S., Fang, Y., Feng, Y., & Jiang, T. (2025). Robotic Teleoperation of Construction Machinery: Technologies, Applications, and Future Directions. In: CIB Conferences: Vol. 1 Article 45.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. International Journal of Cognitive Ergonomics, 4(1), pp. 53–71.

Kim, H., Hong, J., Song, E., Hong, T., & Choi, J.-K. (2026). A multi-dimensional failure factor affecting the on-site adoption of construction robots. Developments in the Built Environment, 25: 100841, pp. 1-30.

Lee, H.J., & Brell-Cokcan, S. (2023). Towards controlled semi-autonomous deconstruction. Construction Robotics, 7, pp. 253–263.

Lee, J.S., & Ham, Y. (2022). Exploring Human-Machine Interfaces for Teleoperation of Excavator. In: Construction Research Congress 2022, pp. 757–765.

Lee, J.S., Ham, Y., Park, H., & Kim, J. (2022). Challenges, tasks, and opportunities in teleoperation of excavator toward human-in-the-loop construction automation. Automation in Construction, 135: 104119.

Lewis, J. R., & Sauro, J. (2018). Item benchmarks for the system usability scale. Journal of Usability Studies, 13(3), pp. 158–167.

Liang, C.-J., Wang, X., Kamat, V.R., & Menassa, C.C. (2021). Human–Robot Collaboration in Construction: Classification and Research Trends. Journal of Construction Engineering and Management, 147(10).

Liu, Y., Habibnezhad, M., & Jebelli, H. (2021). Brain-computer interface for hands-free teleoperation of construction robots. Automation in Construction, 123: 103523, pp. 1–16.

Luo, Y., Wang, J., Liang, H.-N., Luo, S., & Lim, E. G. (2021). Monoscopic vs. Stereoscopic Views and Display Types in the Teleoperation of Unmanned Ground Vehicles for Object Avoidance. In: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), Vancouver, BC, Canada, IEEE, pp. 418–425.

Rasheed, U., Cai, J., Xu, X., Hu, Y., & Li, S. (2024). Equipment Teleoperation and Its Impacts on Future Worker and Workforce in Construction: Semi-Structured Interviews. In: Construction Research Congress 2024, pp. 846–855.

Rea, D.J., & Seo, S.H. (2022). Still Not Solved: A Call for Renewed Focus on User-Centered Teleoperation Interfaces. Frontiers in Robotics and AI, 9: 704225.

Robillard, G., Bouchard, S., Renaud, P., & Cournoyer, L.G. (2002). Validation canadienne-française de deux mesures importantes en réalité virtuelle : l'Immersive Tendencies Questionnaire et le Presence Questionnaire. In: 25e congrès de la Société Québécoise pour la Recherche en Psychologie (SQRP).

Rodrigues, P.B., Becerik-Gerber, B., Soibelman, L., Lucas, G.M., & Roll, S.C. (2025). Impact of selective environmental sound attenuation on operator performance, stress, attention, and task engagement in teleoperated demolition. Automation in Construction, 169: 105876, pp. 1-14.

Rodrigues, P.B., Lucas, G. M., Fang, Y., Wang, Z., Roll, S. C., Becerik-Gerber, B., & Soibelman, L. (2025). Understanding Operators' Sensory Needs for Human–Robot Interaction in Teleoperated Demolition. Journal of Construction Engineering and Management, 151(12): 04025207, pp. 1-17.

Witmer, B.G., & Singer, M.J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators and Virtual Environments, 7(3), pp. 225–240.

Xu, J., Dixit, M., & Wang, X. (2024). Adaptive Visual Perception for Robotic Construction Process: A Multi-Robot Coordination Framework. ArXiv:2412.11275.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMIZED PLANNING & DATA EXCHANGE FOR PREFABRICATED FAÇADE RENOVATION PROCESSES

Julius Emig[1], Orjola Braholli[1], Dietmar Siegele[1] and Dominik T. Matt[1,2]
[1]Fraunhofer Italia Research, Italy
[2]Free University of Bozen - Bolzano, Italy

Abstract

Europe's renovation rate remains below what is needed for climate-neutrality targets; scaling façade refurbishment demands standardised, data-driven automation. Current practice relies on bespoke BIM modelling and file-based scan-to-BIM pipelines, which are labour-intensive, sensitive to as-built deviations, and fragmented across design, production, and logistics. We present an end-to-end workflow that ingests 3D point-cloud surveys and performs geometric panelisation to derive storey-scale and small prefabricated façade panels. The system outputs IFC-compliant panel objects with cut-outs and metadata, synchronised via a self-hosted, versioned Speckle hub. A residential case study demonstrates workflow feasibility and reliable handover.

Introduction

Europe's climate-neutrality ambitions, reinforced by energy crises, place immediate pressure on the renovation of the ageing building stock. In the European Union, buildings account for around 40% of total energy consumption and 36% of energy-related greenhouse-gas emissions (European Commission, 2020b). Despite this central role, the annual renovation rate remains persistently low: on average, less than 1% of the building stock is renovated each year (European Commission, 2020a). This is well below the level required to meet European climate targets, which call for at least a doubling of renovation rates by 2030 and the sustained acceleration of renovation activity thereafter to achieve climate neutrality by 2050 (European Commission, 2020b).

At the same time, renovation projects are frequently perceived as slow, costly, and highly disruptive, while the construction sector faces a structural shortage of skilled labour. Together, these factors create a compelling need for a step-change in productivity, predictability, and scalability in renovation processes, particularly for façade retrofits, which offer high energy-saving potential but are still largely executed through bespoke, labour-intensive on-site construction methods.

A widely discussed pathway to address these challenges is the industrialisation of façade refurbishment through prefabricated façade panels or modules. Compared with conventional, largely on-site insulation retrofits, off-site prefabricated façade approaches shift critical activities to factory-controlled environments, enabling shorter on-site phases, improved quality control, and reduced disturbance for occupants during installation (European Commission, 2024; Zimmermann, 2012). In addition, componentised prefabricated assemblies, especially when designed with demountable connections, offer favourable conditions for circular-economy strategies. When combined with digital material passports, such systems can support planned disassembly, higher-value material recovery, and improved material traceability across the building lifecycle (Akinade et al., 2015; ISO, 2020; Çetin et al., 2023).

However, industrialising façade renovation is not merely a matter of developing new construction products. It is fundamentally a process and information-management challenge. Even where technically mature prefabricated solutions exist, their large-scale deployment remains constrained by the high effort required for building surveying, as-is modelling, project-specific design adaptation, multi-stakeholder coordination, and fragmented data handovers across specialised tools and organisational boundaries.

Prefabricated façade renovation

Over the last decade, research and innovation initiatives have advanced prefabricated façade retrofit systems and serial renovation concepts. The TES Energy Façade project established a timber-based modular façade approach for energetic renovation and investigated a survey-to-installation data-flow concept to reduce time and cost (Larsen et al., 2011). EU projects such as SINFONIA implemented prefabricated façade modules in real contexts, albeit sometimes with conventional on-site execution for specific layers (e.g., cladding and sealing) (Andaloro et al., 2019). The iNSPiRe project developed "Industrialised Renovation Kits", including prefabricated timber-frame façade elements with integrated active components to reduce planning and installation effort (Ochs et al., 2015). 4RinEU similarly explored integrated façade modules with active components (e.g., PV, ventilation, shading) as part of deep renovation packages (Noris et al., 2017; Pernetti et al., 2019). On the market-facing side, Energiesprong (n.d.) is often cited as one of the most advanced examples of industrialised renovation combined with an enabling business model, though its roll-out still tends to focus on building typologies that are easier to standardise and replicate. Across these efforts, a

recurring observation is that although technical feasibility is demonstrated, planning and process overhead remain a major driver of cost and a barrier to scaling, particularly when each project requires substantial manual adaptation and coordination effort.

From a computing-in-construction perspective, these observations highlight the central role of digital workflows across the entire renovation process, spanning survey capture, scan-to-model transformation, modular design configuration, performance assessment, production planning, logistics, and installation. Multiple studies report that renovation productivity bottlenecks stem not only from tool-level limitations (e.g., the effort and uncertainty involved in converting survey data into reliable as-built BIM), but also from socio-technical factors such as multi-stakeholder coordination complexity, communication bottlenecks, and information losses/redundancies across phases and organisations (Volk et al., 2014; Prieto et al., 2024; Lupica Spagnolo et al., 2022; Chen & Luo, 2026). In this context, continuous end-to-end digital workflows, linking building survey/scan-to-BIM to downstream design, prefabrication/CAM, and on-site assembly, are increasingly highlighted as key enablers to reduce planning effort and rework, mitigate data loss at interfaces, and support scalable deployment of prefabricated façade renovation solutions (Lupica Spagnolo et al., 2022; Fichera et al., 2025; Iturralde et al., 2025).

Scan-to-BIM & as-is modelling

A reliable, planning-grade as-is description of façade planes and openings is a prerequisite for panelised renovation. Scan-to-BIM / as-built BIM creation is commonly decomposed into (i) reality capture and preprocessing (e.g., registration and cleaning), (ii) geometric and semantic interpretation (segmentation / object recognition), and (iii) generation of parametric BIM/IFC entities and their relationships (Tang et al., 2010; Pătrăucean et al., 2015). Recent work increasingly leverages learning-based segmentation to automate recognition and downstream model creation, as exemplified by SCAN2BIM and Cloud2BIM (Campagnolo et al., 2023; Zbirovský and Nežerka, 2025). However, façade openings remain a hard case: measurement quality deteriorates at edges and window glazing often returns few or no points, while occlusions and missing data still necessitate manual validation/correction in practice (Pu and Vosselman, 2007; Tang et al., 2010; Tang et al., 2009). Consequently, renovation-oriented pipelines often adopt hybrid automation with human-in-the-loop controls, targeting a fit-for-purpose envelope representation (e.g., alignment planes and opening extents) that supports panelisation and prefabricated-panel installation rather than a fully detailed as-built BIM (Fichera et al., 2025; Song et al., 2023).

Research gap & Contribution beyond the state of the art

Across the reviewed literature, a persistent gap can be identified: while numerous methods address individual stages of prefabricated façade renovation, such as reality capture, scan-to-BIM transformation, panelisation strategies, IFC-based exchange, or logistics optimisation, far fewer contributions demonstrate an operationally usable, end-to-end workflow that explicitly treats continuous data flow as a primary research concern. In practice, large-scale deployment is constrained by fragmented information exchange across specialised tools and stakeholders, frequent data loss due to file-based handovers, and the absence of workflow-oriented ecosystems capable of linking survey-derived geometry, planning artefacts, and downstream operational outputs through a shared, versioned backbone.

This paper addresses this gap by presenting an integrated, web-based planning workflow for prefabricated façade renovation that prioritises continuous, interoperable data flow from survey ingestion to panel planning and logistics preparation. Rather than optimising isolated algorithms, the work focuses on the ecosystem-level potential of a modular toolchain coordinated through a central workflow layer. Accordingly, the implemented computational components should be understood as research prototypes for validating end-to-end workflow continuity; in future implementations, individual modules may be replaced by mature third-party solutions where interoperable interfaces and data continuity can be maintained. The implemented workflow combines point cloud ingestion and geometric pre-processing, semi-automatic derivation of façade planes and openings, interactive yet automated panelisation with IFC-compliant outputs, and downstream logistics preparation through panel-to-package transformation and interactive container loading.

Methodology

We propose a structured, workflow-oriented system that prioritises continuous data flow from survey to planning and logistics. The system follows the three-block structure shown in Figure 1: site survey as the data source, a 3D renovation planning web interface as the interaction layer, and a backend that provides custom services and a persistent data layer. The interaction layer is a Three.js-based (three.js, n.d.) web application used for both façade panel planning and the logistics planning. The communication with the backend works through GraphQL- and REST-APIs. The backend stores all project artefacts as versioned objects in a self-hosted Speckle server (Speckle, n.d.).

Overall, the workflow proceeds in four stages. First, façade and window point clouds (see Detection of window opening dimensions) are uploaded through the web interface and validated. Second, façade planes are reconstructed from the point clouds and stored together with the cleaned survey data. Third, façade panelisation is generated and reviewed interactively in the Three.js-based web frontend, and the resulting panel objects are persisted via Speckle. Finally, the panel objects are converted into package objects and transferred to the Three.js-based logistics interface for 3D packing and container loading.

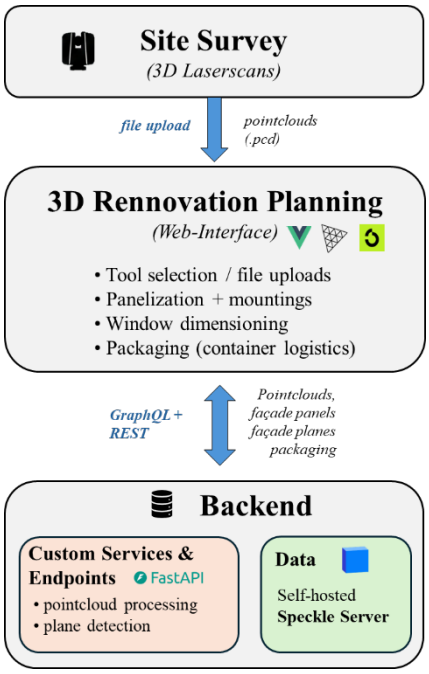


Figure 1 System architecture

Versioned Data Management

When point cloud data are uploaded via the web interface to the backend they get stored as Speckle point cloud. The backend uses the Open3D python library (Zhou et al., 2018) to read the scan data and creates a versioned object graph in the *pointcloud-facade* branch. This gives all downstream tools a consistent and queryable source of truth.

Speckle is used as a versioned common data environment with a deliberate branch/model structure. In Speckle, each project contains named models (branches) that hold a sequence of versions (commits) referencing object graphs. The workflow assigns one model per artefact type so that tools can query the latest version deterministically: *pointcloud-facade* and *pointcloud-windows* store survey data; *facade-planes* stores derived plane objects; and *facade-panels* stores panel layouts in IFC-compliant data format. This convention makes data handovers explicit, reduces ambiguity when multiple tools write to the same project, and enables reproducible retrieval of the latest artefact by model name.

Façade plane reconstruction

A critical process step is the reconstruction of the façade planes from the 3D point cloud data. As discussed already in sub-section Scan-to-BIM & as-is modelling there are existing scan-to-BIM tools. We found however that the results of that tools are not accurate enough to execute a detailed façade planning with prefabricated panels. That is why we implemented a semi-automated approach that lets users define the façade planes directly based on the 3D point cloud data. This ensures a maximum of accuracy and flexibility.

Façade panel generation

The reconstructed façade planes are the basis for the panel generation which follows a parametric grid-based process in the 2D façade plane. The panel system is defined by parameters (standard width & height, minimum width & height, depth, gap width, mounting position, weight) that are iteratively fitted to the façade while respecting openings and façade boundaries. To control where a panel must terminate, the user defines horizontal and vertical break edges (e.g., façade edges and opening boundaries). These break edges induce a rectilinear grid on the façade plane and thus a set of grid fields. Within each grid field, the algorithm first places panels at the standard width/height. If boundary panels would need to be cut, the remaining strip is evaluated against the minimum admissible panel dimension. Whenever the remaining width/height would violate this minimum, the algorithm avoids producing undersized edge pieces by uniformly resizing the panels along the affected direction within that grid field so that all resulting panels satisfy the minimum dimension constraint while maintaining the specified gap. This resizing is applied locally within the affected grid field; standard panel dimensions are preserved in unaffected fields, while only boundary-adjacent panels may deviate where necessary. The resulting panel layout in the 2D façade plane space gets extruded and transformed back to the global 3D coordinate system. The 3D panels are now represented as Three.js geometries with a stable identifier, a façade ID, and dimensional metadata (width, height, depth, gap).

A dedicated IFC writer transforms the generated Three.js panels into IFC-compliant objects (e.g., IfcPlate with swept-area geometry) and attaches property sets containing Width, Height, and Depth. The resulting IFC model is stored on the Speckle server. File-based exports were also implemented (.ifc model with 3D façade panels, .csv lists with panel dimensions and mounting points), but they are meant primarily as documentation features; the operational data for prefabrication and on-site mounting remains available as versioned Speckle objects that can be queried and reused across tools.

Detection of window opening dimensions

Window planning is supported as an optional, separate module. Detailed window design typically requires denser scans than façade-level surveys, so the workflow accepts dedicated window point clouds when they are available. In practice, these local scans are acquired from accessible interior or exterior positions using higher-density close-range measurements, depending on site access and required detail. They are registered to the building-level façade point cloud during survey preprocessing and then uploaded as already aligned window point clouds to the *pointcloud-windows* branch. Then they are processed by a light-weight plane detection routine that produces structured instances with fitted planes which represent the window opening (four reveals and one wall plane). Leveraging Open3D the backend performs a deterministic sequence: The window scan is down-sampled with a voxel size of 0.02, normals are estimated and oriented

consistently, and planar patches are detected using normal variance 60°, coplanarity 75°, outlier ratio 0.75, and a k-nearest-neighbour search of k=30. The largest detected plane is interpreted as wall and the remaining planes as window reveal candidates. Each candidate is converted to a rectangular plane using its oriented bounding box and stored alongside the raw point cloud for detailed window planning.

Container packaging

A logistics planning feature was developed and implemented to calculate the number and size of shipping containers required for transporting the planned façade panels. It consists of an interactive 3D environment based on Three.js and ThatOpen bim-components (ThatOpen, n.d.) with a packing solver that is based on 3D-bin-picking algorithm from (Aoulad Lahcene, 2023).

The backend resolves the latest *facade-panels* model through the GraphQL API, loads the referenced object graph, and traverses it to identify panel candidates. Dimensions are taken from attached property sets when available. The backend emits a package list with dimensions, quantity, priority, stacking capacity, and allowed rotations, which is consumed by the logistics tool for packing.

The packing solver uses a greedy heuristic with priorities and open-point placement. Packages are expanded by quantity, grouped by priority, and sorted by volume within each priority. Placement proceeds by testing rotations at available open points, checking container bounds and collision constraints, and honoring stacking limits. New open points are generated on the front, right, and top faces of each placed package, while invalid diagonal points are removed to stabilize the search. A manual drag-and-drop mode remains available for operational corrections that are not captured by the heuristic.

Results & Discussion

The workflow was validated on a representative residential building using terrestrial laser-scan (TLS) data captured with a FARO Focus M70. Figure 2 shows the façade point cloud used as survey input. To anchor the digital planning workflow in a manufacturable solution, a physical prototype of the prefabricated wood-fibre insulation panel was used as planning reference (Figure 3), confirming the feasibility of the intended panel geometry and joint concept for off-site fabrication and on-site installation.

Across the end-to-end validation, three main outcomes were observed. First, the semi-automated façade-plane reconstruction provided planning-grade reference geometry from real scan data, enabling downstream panelisation without relying on fully automatic scan-to-BIM outputs. Second, automated panelisation generated façade module layouts that are directly usable for downstream BIM and fabrication preparation through IFC-compliant objects and associated dimensional metadata. Third, the versioned Speckle-based data backbone enabled a consistent handover from survey-derived façade modules to logistics planning, avoiding repeated file exports/imports and reducing data fragmentation between tools. Given the single-case validation, these results should be interpreted as evidence of workflow feasibility and reduced data-fragmentation effort rather than as a benchmarked productivity gain.



Figure 2 Point cloud of residential building façade



Figure 3 Prototype of prefabricated wood-fiber insulation panels mounted on lab test wall

Based on the survey data, façade planes were reconstructed in a human-in-the-loop step by selecting a small number of corner points directly in the 3D point cloud. In practice, this interaction step acted as a controlled “alignment and interpretation” layer: it resolved ambiguities caused by occlusions and scan noise and ensured that the resulting planes represent the reference geometry required for fabrication-oriented planning. Human input is retained where survey ambiguity or design intent materially affects manufacturability, whereas steps governed by explicit geometric rules and machine-checkable constraints are automated and subsequently interactively reviewed.

Figure 4 exemplifies the resulting panel layout for a reconstructed façade plane. In this validation, user-defined break edges (e.g., storey lines, façade boundaries, opening extents) proved effective to enforce architecturally meaningful alignments while keeping the layout manufacturable. The panelisation outcome demonstrates that minimum-dimension constraints can be respected without producing impractically small boundary pieces; where the grid would otherwise create undersized edge panels, the layout adapts by distributing the residual dimension within the affected grid field. As output, the workflow produces an IFC-compliant façade panel model (IfcPlate entities with dimensional property sets) and a fabrication/installation part list containing panel dimensions and mounting positions.

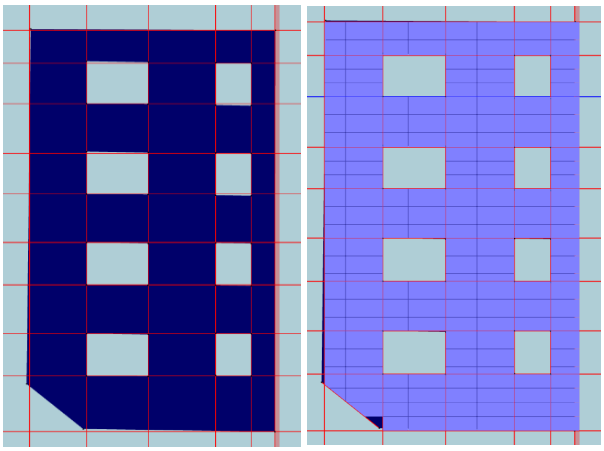


Figure 4 Façade plane (left) and generated panels (right) with grid pattern composed out of break edges (red)

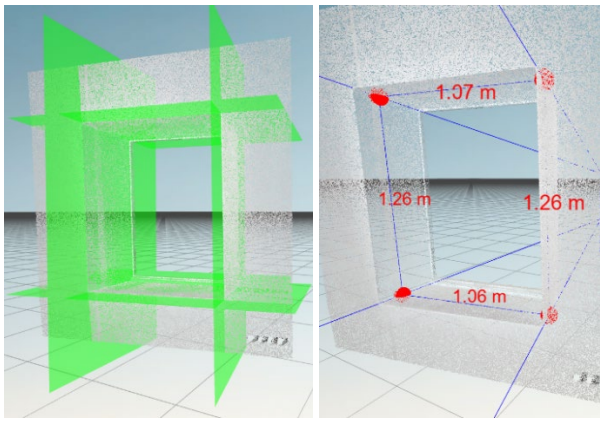


Figure 5 Detected window reveal planes (left) and dimensioning of window opening (right)

For optional window replacement, dedicated window scans were processed to detect planar patches corresponding to the wall plane and window reveals (Figure 5, left). The fitted planes enable robust opening dimensioning. In the shown example, the extracted opening measures approximately 1.26 m in height and 1.06–1.07 m in width (Figure 5, right). These dimensions can be re-used to parametrically define window objects and to validate panel cut-outs and tolerances at the interface between façade modules and window frames.

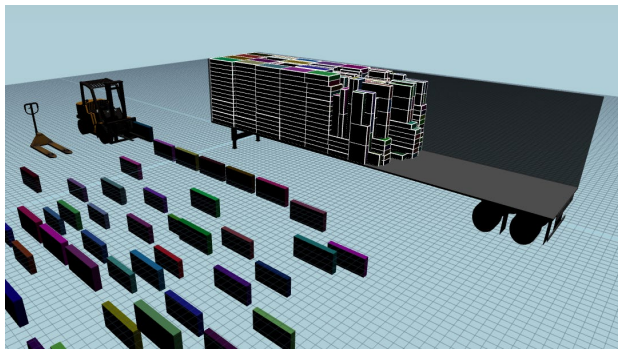


Figure 6 Logistics tool for container shipping with IFC and Speckle APIs

The generated panel objects were subsequently transferred to the logistics planning tool without file-based re-entry. By resolving the latest façade-panels version from Speckle and reading panel dimensions from their attached properties, the backend automatically emitted a package list that was consumed by the Three.js-based packing interface. Figure 6 shows the interactive container-loading environment, where a 3D bin-packing heuristic proposes an initial packing based on priority, rotations and stacking constraints. Planners can apply drag-and-drop corrections when operational considerations require it, i.e., the packing logic functions as decision support rather than a fully automatic optimiser. Overall, this end-to-end handover demonstrates that a versioned, interoperable data backbone can connect survey-derived planning artefacts to downstream logistics decisions in a consistent way.

Encountered limitations

- Operator input remains a dependency for "planning-grade" geometry. While the human-in-the-loop plane reconstruction improved robustness on real scan data, it introduces operator effort and potential variability (e.g., depending on how corner points are selected under occlusion/noise).
- Break-edge definition is a knowledge bottleneck. The quality of panel alignments and module boundaries strongly depends on correctly specified break edges. This provides necessary design control, but it also limits full automation and requires consistent modelling conventions.
- Standardisation vs. manufacturability trade-off. Avoiding undersized boundary panels through layout adaptation improves feasibility, but it can increase the number of non-standard panel dimensions near boundaries/openings and thus may reduce production economies unless size-variant limits are introduced. This effect is expected to become more pronounced for more geometrically complex façades with a higher density of break edges and local exceptions.
- Window dimensioning requires dedicated scan coverage. Reliable opening extraction benefits from denser, targeted scans. Where such data is unavailable or incomplete, window replacement support may be partial and cut-out validation less robust.

- Packing is not fully automatable in operational contexts. Even with a heuristic initial packing, manual adjustments were necessary to reflect practical constraints (handling preferences, sequencing, transport considerations), implying that packing quality should be assessed by "solution usefulness + edit effort" rather than by optimality alone.
- Limited external validity and limited quantitative evaluation. The validation was conducted on a single representative case; additional façade typologies (e.g., highly articulated/non-planar façades, heavier occlusion) and systematic metrics (time-on-task, number of panel variants, amount of manual correction, downstream IFC usability checks) are needed to quantify scalability and robustness.

Conclusions & Outlook

This paper presented an integrated, web-based planning workflow for prefabricated façade renovation that connects survey ingestion, façade and window geometry extraction, automated panel planning, and logistics preparation through a shared, versioned data backbone. In contrast to tool-isolated solutions, the contribution is workflow-oriented: a central object graph provides a consistent "source of truth" across planning stages, while IFC-compliant semantics expose façade modules in a form that is directly consumable by downstream BIM and production environments.

Validation on a residential façade scan demonstrated that a semi-automated, human-in-the-loop reconstruction of façade planes provides the accuracy required for fabrication-oriented module planning, motivated by the observed insufficiency of fully automatic scan-to-BIM outputs for detailed panel planning. On top of this reference geometry, grid-driven panelisation yields IFC-compliant panel objects and installation metadata suitable for fabrication documentation and on-site mounting preparation. Optional window opening dimensioning further supports coordinated window replacement when dedicated, denser window scans are available. Finally, the same panel objects can be transformed into logistics packages and transferred to an interactive container-loading environment without file-based re-entry, demonstrating continuous data flow from survey-derived planning artefacts to downstream logistics decisions.

The validation also highlights practical limitations that shape industrial applicability. First, "planning-grade" envelope geometry currently depends on operator input for plane definition, reflecting a deliberate accuracy–automation trade-off. Second, panel layout quality depends on consistently defined break edges (storey lines, façade boundaries, opening extents), which provide necessary design control but remain a manual knowledge bottleneck. Third, robust window dimensioning depends on scan coverage quality and the availability of dedicated window scans. Fourth, logistics packing relies on a heuristic with an explicit manual correction mode for operational constraints not captured algorithmically. Finally, results are currently based on a single representative case study, motivating broader evaluation across building typologies.

Future work will therefore focus on closing remaining process gaps, strengthening downstream integration, and quantifying robustness and productivity gains under realistic constraints, in particular:

- Socio-technical workflow integration and praxis adoption, including stakeholder roles, validation responsibilities, trust in semi-automated outputs, and acceptance in multi-actor renovation practice.
- Coupling to LCA/LCC assessment to provide early-stage environmental and cost feedback for alternative panel configurations.
- Integration of construction-site simulation and 4D time planning to link panel layouts and logistics decisions to installation sequences, crane operations, and site constraints.
- A production link to prefabrication/CAM systems (e.g., CNC cutting, drilling templates, labelling) to translate panel and mounting-point metadata into shop-floor instructions.
- Refinement of joint/gap modelling and tolerance handling, including explicit as-built deviation compensation strategies, to improve robustness for industrial roll-out.
- Further automation of façade feature extraction (openings, storey lines) and systematic evaluation on additional building typologies, including metrics such as time-on-task, number of panel variants, required manual corrections, and downstream interoperability performance, to quantify time savings and error reduction.

Acknowledgments

This work was supported by the European Regional Development Fund of the Autonomous Province of Bolzano - Investments in growth and employment FESR 2020-2027, within the FESR project 1015 'DIAMANT' [CUP: B57H23003880007]

References

Akinade, O.O., Oyedele, L.O., Bilal, M., Ajayi, S.O., Owolabi, H.A., Alaka, H.A. and Bello, S.A. (2015) 'Waste minimisation through deconstruction: A BIM based Deconstructability Assessment Score (BIM-DAS)', Resources, Conservation and Recycling, 105, pp. 167–176. https://doi.org/10.1016/j.resconrec.2015.10.018.

Andaloro, A., Gasparri, E., Avesani, S. and Aitchison, M. (2019) 'Market survey of timber prefabricated envelopes for new and existing buildings', in PowerSkin Conference 2019: Proceedings. Munich, Germany, 17 January 2019, pp. 123–132.

Aoulad Lahcene, O. (2023) '3D-bin-packing: This project solved the problem of the 3D bin packing using JavaScript/Threejs', GitHub repository.

https://github.com/AouladLahceneOussama/3D-bin-packing

Campagnolo, D., Camuffo, E., Michieli, U., Borin, P., Milani, S. and Giordano, A. (2023) 'Fully automated scan-to-BIM via point cloud instance segmentation', in 2023 IEEE International Conference on Image Processing (ICIP), pp. 286–290. https://doi.org/10.1109/ICIP49359.2023.10222064

Çetin, S., Raghu, D., Honic, M., Straub, A. and Gruis, V. (2023) 'Data requirements and availabilities for material passports: A digitally enabled framework for improving the circularity of existing buildings', Sustainable Production and Consumption, 40, pp. 422–437. https://doi.org/10.1016/j.spc.2023.07.011

Chen, Y. and Luo, X. (2026) 'Indoor scan-to-BIM workflows: Progress, challenges, and future directions (2014–2024)', Automation in Construction, 181(Part A), 106578. https://doi.org/10.1016/j.autcon.2025.106578

Energiesprong (n.d.) What is Energiesprong? https://www.energiesprong.org/what-is-energiesprong/

European Commission (2020a) Energy efficiency in buildings. In focus (Directorate-General for Energy), Brussels, 17 February 2020. https://commission.europa.eu/system/files/2020-03/in_focus_energy_efficiency_in_buildings_en.pdf

European Commission (2020b) A Renovation Wave for Europe, greening our buildings, creating jobs, improving lives. Communication from the Commission, COM(2020) 662 final, Brussels, 14 October 2020. https://energy.ec.europa.eu/system/files/2020-10/renovation_wave_strategy_0.pdf

European Commission (2024) Research Note on Offsite Construction. Prepared by Ecorys under the Technical Secretariat of the High Level Construction Forum, manuscript completed December 2024. https://build-up.ec.europa.eu/system/files/2025-01/RB385Rpj2z_29_01_2025_083227.pdf

Fichera, G., Guardo, V., Distefano, D.L., Rodonò, G., Tardo, C. and Margani, G. (2025) 'Scan-To-BIM Methodology for Installation of Prefabricated Panels on Existing Building Façades: From Point Cloud Data to Native BIM Categories', in Albatici, R. et al. (eds) Envisioning the Futures, Designing and Building for People and the Environment (Colloqui.AT.e 2025), Lecture Notes in Civil Engineering, vol. 766. Springer, Cham. https://doi.org/10.1007/978-3-032-06993-1_22

ISO (2020) ISO 20887:2020 Sustainability in buildings and civil engineering works , Design for disassembly and adaptability , Principles, requirements and guidance. International Organization for Standardization. https://www.iso.org/standard/69370.html

Iturralde, K., Shen, W., Bazan, R., Amo-Grau, O., Liu, D., Srinivasaragavan, A., Das, S., Bock, T. and Holst, C. (2025) 'Integrated workflow for a semi-automated and robotic facade renovation', Frontiers in Built Environment, 11:1649278. https://doi.org/10.3389/fbuil.2025.1649278

Larsen, K.E., Lattke, F., Ott, S. and Winter, S. (2011) 'Surveying and digital workflow in energy performance retrofit projects using prefabricated elements', Automation in Construction, 20(8), pp. 999–1011. https://doi.org/10.1016/j.autcon.2011.04.001

Lupica Spagnolo, S., Signorini, M., Vesanen, T., Pavan, A. and Kousouris, S. (2022) 'Information Requirements for an Efficient Renovation Process', in Daniotti, B. et al. (eds) Innovative Tools and Methods Using BIM for an Efficient Renovation in Buildings. SpringerBriefs in Applied Sciences and Technology. Springer. https://doi.org/10.1007/978-3-031-04670-4_1

Noris, F., Pernetti, R., Lennard, Z., Signore, G. and Lollini, R. (2017) '4RinEU: Robust and Reliable Technology Concepts and Business Models for Triggering Deep Renovation of Residential Buildings in EU', Proceedings, 1(7), 661. https://doi.org/10.3390/proceedings1070661

Ochs, F., Siegele, D., Dermentzis, G. and Feist, W. (2015) 'Prefabricated timber frame façade with integrated active components for minimal invasive renovations', Energy Procedia, 78, pp. 61–66. https://doi.org/10.1016/j.egypro.2015.11.115

Pătrăucean, V., Armeni, I., Nahangi, M., Yeung, J., Brilakis, I. and Haas, C. (2015) 'State of research in automatic as-built modelling', Advanced Engineering Informatics, 29(2), pp. 162–171. https://doi.org/10.1016/j.aei.2015.01.001

Pernetti, R., Pinotti, R., Babich, F. and Lollini, R. (2019) 'Towards a Systemic Approach For Supporting Design Process Of Deep Renovation Packages Based On Prefabricated Façade Technologies', in Proceedings of Building Simulation 2019: 16th Conference of IBPSA, pp. 1552–1558. https://doi.org/10.26868/25222708.2019.211214

Prieto, A., Armijos-Moya, T. and Konstantinou, T. (2024) 'Renovation process challenges and barriers: addressing the communication and coordination bottlenecks in the zero-energy building renovation workflow in European residential buildings', Architectural Science Review, 67(3), pp. 205–217. https://doi.org/10.1080/00038628.2023.2214520

Pu, S. and Vosselman, G. (2007) 'Extracting windows from terrestrial laser scanning', in Rönnholm, P., Hyyppä, H. and Hyyppä, J. (eds) ISPRS 2007: Vol. XXXVI, Part 3/W52, Proceedings of the ISPRS Workshop Laser Scanning 2007 and SilviLaser 2007, Espoo, Finland, 12–14 September 2007, pp. 320–325. Espoo: International Society for Photogrammetry and

Remote Sensing (ISPRS). https://www.isprs.org/proceedings/XXXVI/3-W52/final_papers/Pu_2007.pdf

Song, W., Luo, J., Zhao, D., Fu, Y., Cheng, C.-Y. and Furukawa, Y. (2023) 'A-Scan2BIM: Assistive scan to building information modeling', in 34th British Machine Vision Conference 2023 (BMVC 2023), Aberdeen, UK, 20–24 November 2023. Durham: BMVA. https://papers.bmvc2023.org/0014.pdf

Speckle (n.d.) speckle-server. GitHub repository. https://github.com/specklesystems/speckle-server

Tang, P., Akinci, B. and Huber, D. (2009) 'Quantification of edge loss of laser scanned data at spatial discontinuities', Automation in Construction, 18(8), pp. 1070–1083. https://doi.org/10.1016/j.autcon.2009.07.001.

Tang, P., Huber, D., Akinci, B., Lipman, R. and Lytle, A. (2010) 'Automatic reconstruction of as-built building information models from laser-scanned point clouds: A review of related techniques', Automation in Construction, 19(7), pp. 829–843. https://doi.org/10.1016/j.autcon.2010.06.007.

ThatOpen (n.d.) engine_components. GitHub repository. https://github.com/ThatOpen/engine_components

three.js (n.d.) Three.js – JavaScript 3D Library. https://threejs.org/

Volk, R., Stengel, J. and Schultmann, F. (2014) 'Building Information Modeling (BIM) for existing buildings, Literature review and future needs', Automation in Construction, 38, pp. 109–127. https://doi.org/10.1016/j.autcon.2013.10.023

Zbirovský, S. and Nežerka, V. (2025) 'Open-source automatic pipeline for efficient conversion of large-scale point clouds to IFC format', Automation in Construction, 177, 106303. https://doi.org/10.1016/j.autcon.2025.106303

Zhou, Q.-Y., Park, J. and Koltun, V. (2018) 'Open3D: A Modern Library for 3D Data Processing', arXiv preprint arXiv:1801.09847. https://arxiv.org/abs/1801.09847

Zimmermann, M. (2012) Prefabricated Systems for Low Energy Renovation of Residential Buildings: ECBCS Annex 50, Project Summary Report. International Energy Agency (IEA), Energy Conservation in Buildings and Community Systems Programme. https://www.iea-ebc.org/Data/publications/EBC_PSR_Annex50.pdf

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

EVALUATION OF MOBILE PLATFORMS FOR AUTOMATION OF METRO TUNNEL INSPECTIONS

Timur Kuzu, Katharina Klemt-Albert
Chair and Institute of Construction Management, Digital Engineering and Robotics in Construction,
RWTH Aachen University, Aachen, Germany

Abstract

Metro tunnel inspections face growing demands from urbanization and tight maintenance windows, yet manual and time-consuming inspection methods remain. While rail-bound vehicles, wheeled and tracked robots, quadruped robots, and Unmanned Aerial Vehicles (UAVs) show promise for inspection tasks, no systematic evaluation addresses metro tunnel-specific constraints like track geometries, tunnel access and inspection windows for mobile platform choice. This paper reviews inspection platforms, derives evaluation criteria from literature, operator interviews, and observations, and quantitatively compares platforms across sensing, payload, speed, runtime, deployment, and locomotion. Results identify optimal use-cases for each system, guiding deployment strategies for automated metro tunnel inspections.

Introduction

Motivation

Tunnel infrastructure, including road tunnels, railway tunnels, utility corridors and metro systems, plays a crucial role in modern transportation networks, particularly in regions with complex topography or high population density. Urbanization and the resulting densification of cities are driving the need to maintain and expand this underground infrastructure. In the past 75 years, the urban population has increased from 20% of the global population (500 million people) in 1950 to 45% of the population in 2025 (3.69 billion people) (United Nations, 2025). During a similar period of time, from 1950 to 2015, the cumulative global length of metro systems has increased more than tenfold from 1250 km to 13000 km through a combination of extending old networks and the construction of new metro lines (Bono et al., 2022). As such, metro systems have established themselves as a substantial contributor to inner-city passenger transport with a total annual ridership count of 58 billion in 2023 (International Association of Public Transport, 2024), accounting for up to 33% of the modal share in major cities like Tokyo (Le et al., 2022).

Like other infrastructure buildings, tunnels inevitably deteriorate over time due to the influence of material aging, heavy loads, environmental conditions, and irregular events such as earthquakes. In combination with an increasing average service life of tunnels due to the demand for underground infrastructure, structural damage can occur in the form of cracks, leakages and cross-section deformations (Bao et al., 2026). From a safety perspective, defects in tunnel linings and surfaces, such as water ingress and spalling, pose a potential threat to passengers and workers within them. Regular building inspections are thus required to ensure the safe, fully-functional operation of tunnels. Currently, manual on-site inspections are performed by certified inspectors and structural engineers at regular inspection intervals (routine inspection) or after unforeseeable events (post-disaster inspection). These inspections are used to assess and predict future building conditions, plan investments and allocate maintenance and repair resources, however the manual inspection process is time-consuming and often suffers from poor visibility and subjective judgment, expressing the demand for more intelligent and automated approaches (Bao et al., 2026).

Metro tunnels impose particularly stringent operational constraints that further demonstrate the need for automated inspections. As metro services operate throughout the day, maintenance windows are timely restricted and typically limited to a maximum of 4-6 hours, including time required for safety checks, resulting in an effective inspection window of approximately 3 hours during nightly shutdowns (Chen et al., 2023). Moreover, the metro-specific environment, including central rails, ballast, crossties, and frequent stations create fundamentally different requirements for automated inspection systems compared to road tunnels. Given the unique combination of metro tunnel-specific inspection constraints, a comparative assessment of mobile inspection platforms is necessary to guide future developments of inspection systems and inform metro operators looking to automate their inspection processes.

Contributions and Outline

While previous work mainly focused on the design and development of individual inspection systems for tunnel environments, this work contributes by evaluating different mobile platforms for automated metro tunnel inspections, specifically focusing on the requirements and operating conditions therein. By doing so, this paper addresses the research gap of systematically evaluating mobile inspection platforms such as rail-bound vehicles, wheeled and tracked robots, quadruped robots and Unmanned Aerial Vehicles (UAVs) against metro tunnel-specific environments and operating constraints.

Including the Abstract and Introduction, the paper is structured into seven sections. Section 3 presents the methodology for deriving metro tunnel-specific evaluation criteria from literature, operator interviews, and in-tunnel observations. Section 4 reviews literature on developed mobile inspection systems for tunnel environments and identifies the main platform classes analyzed in this study. Section 5 provides the core contribution, an evaluation of mobile platforms including rail-bound vehicles, wheeled and tracked robots, quadruped robots, and UAVs. Sections 6 and 7 discuss the implications of the findings and present the conclusions, with recommendations for future work.

Methodology

The evaluation of mobile platforms presented in this paper is based on a mixed-method approach combining a literature review, semi-structured interviews with German metro tunnel operating companies, and qualitative in-tunnel observations. The following section explains how these three data sources were collected and analyzed, and how they informed metro tunnel-specific evaluation criteria.

First, a literature review was conducted to identify existing mobile inspection systems for tunnel environments and to extract key technical parameters. The review focused on robotic and automated inspection systems used in road tunnels, railway tunnels, and metro systems. Relevant publications were identified through the Scopus and Google Scholar databases using a keyword-based search approach, with combinations of terms related to automated tunnel and subway tunnel inspections. From the resulting set of publications, studies were selected based on their relevance to tunnel inspection tasks and the maturity of the described systems. For each platform class, typical performance ranges of key system characteristics were consolidated.

Semi-structured interviews were carried out with multiple representatives from German metro tunnel operating companies. In total, six participants from maintenance and infrastructure departments participated, covering a range of responsibilities from planning and scheduling to on-site inspection and asset management. The main focus of the interviews was on existing maintenance approaches for metro tunnels and the derivation of requirements for a digital, predictive maintenance system. The interview guideline addresses several thematic blocks, including the organizational processes of tunnel inspections from planning to execution, current maintenance strategies, handling of inspection data and reports, and operational constraints in the maintenance process.

The interviews were recorded and transcribed, and the textual data were analyzed using qualitative content analysis according to Mayring's approach. The analysis aimed to identify recurring themes and explicitly stated operational constraints relevant to metro tunnel inspections, such as the duration of night-time inspection windows, track-related access restrictions, and safety procedures. The main outcome for this paper was the extraction of key operational constraints that later served as core criteria in the platform evaluation.

Third, qualitative in-tunnel observations were conducted inside a German metro tunnel during regular inspection and maintenance windows. These observations focused on the physical environment and infrastructural conditions inside metro tunnels. The observations were documented through photos and videos, and the collected material was subsequently analyzed to identify requirements that were not explicitly mentioned in academic literature. The resulting insights were translated into additional evaluation criteria and integrated into the final comparison of mobile platforms.

Tunnel inspection systems

The following section presents existing mobile systems developed for tunnel infrastructure inspections. Since the focus of this work is on the mobile platform of inspection systems, they are categorized according to four main locomotion types: rail-bound vehicles, wheeled and tracked robots, quadruped robots and UAVs. These platforms were selected as they represent the largest share of used platforms for developed systems, including the most mature commercial technologies, balancing trade-offs in characteristics such as payload and terrain adaptability. Figure 1 shows an inspection system based on each locomotion type.

Rail-bound vehicles

Rail-bound inspection vehicles represent an established approach for railway and metro tunnel maintenance, leveraging existing rail infrastructure for stable, high-capacity operation. The Metro Tunnel Surface Inspection System (MTSIS) (Li et al., 2021) deploys a rail-mounted platform equipped with eight high-resolution line-scan cameras and linear lights, capturing tunnel lining images at speeds up to 15 km/h with 0.15 mm/pixel resolution during 3-hour nightly windows. Covering 280° of the tunnel circumference, MTSIS enables automated defect detection via deep learning, achieving 88.5% mean average precision while reducing inspection time by 40% compared to manual methods, which has been evaluated in a series of experiments within different metro tunnels (Li et al., 2021).

Similarly, Zhou et al.'s mobile laser scanning rail trolley (Zhou et al., 2021) uses a Z+F PROFILER 9012 laser scanner on a motorized rail cart, generating intensity/depth point clouds at 0.5 m/s (100 Hz scan rate) for 3D spalling quantification in railway tunnels. The SIDNet model segments defects with 92.5% mean intersection over union (MIoU), enabling volume calculation of spalling via depth analysis. The system was experimentally tested in several railway tunnels in China, proving the validity and feasibility (Zhou et al., 2021).

Specifically targeting crack detection, Liu et al. (Liu et al., 2022) present a system mounted on a Metro Intelligence Inspection Vehicle (MIIV) for subway tunnels. The MIIV

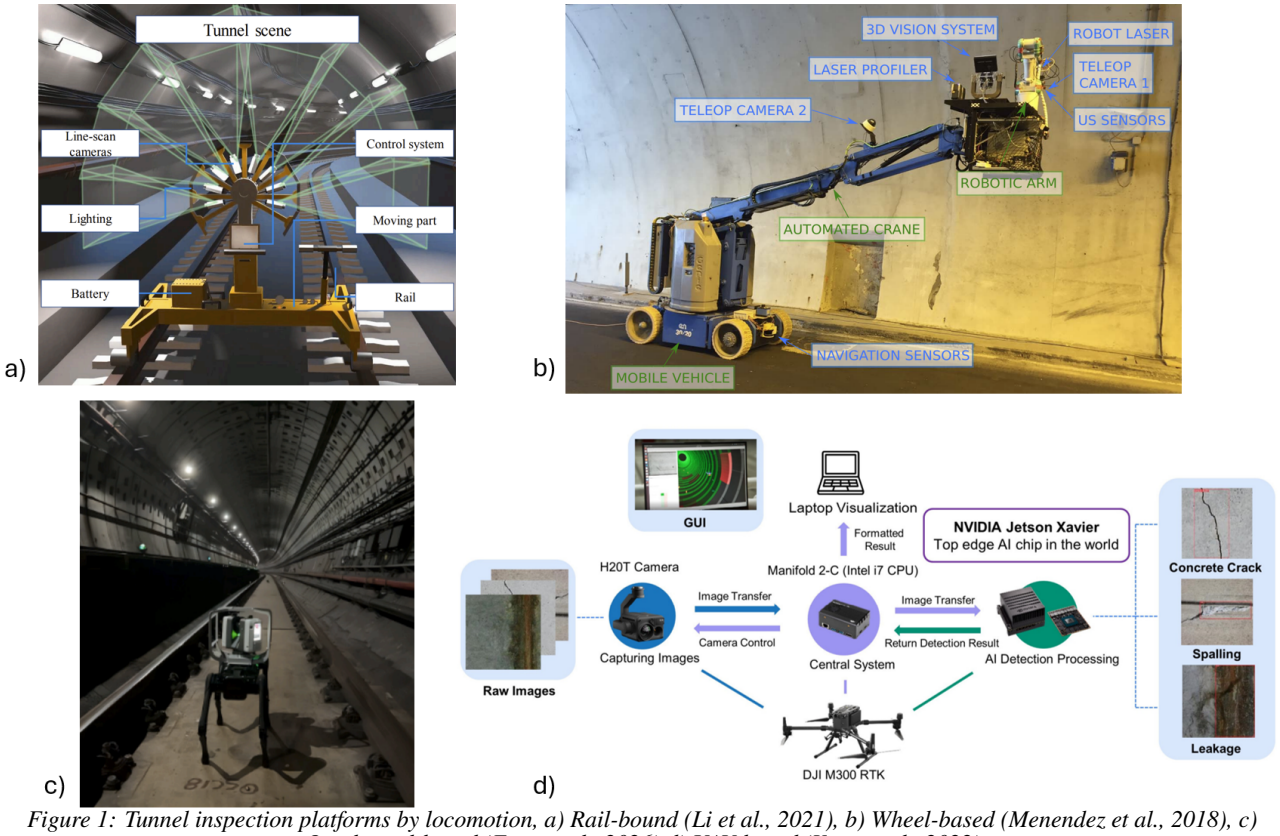


Figure 1: Tunnel inspection platforms by locomotion, a) Rail-bound (Li et al., 2021), b) Wheel-based (Menendez et al., 2018), c) Quadruped-based (Feng et al., 2026) d) UAV-based (Yang et al., 2022)

carries three high-speed cameras on a circular frame, synchronized to capture 150° of inspection range at an optimal inspection speed of 18 km/h. An improved image processing pipeline, including illumination correction, adaptive thresholding, structural noise removal and pixel-level crack width estimation, achieves a crack recognition rate of 91.15%, evaluated in railway tunnel environments (Liu et al., 2022).

Wheeled and tracked robots

The ROBO-SPECT project presents a highly integrated road-tunnel inspection robot composed of a self-navigating mobile vehicle, an automated crane, a 7-degree of freedom (DOF) robotic arm carrying an ultrasonic sensor, stereo cameras, and a 3D laser profiler. The wheeled platform autonomously follows the tunnel wall using laser-based navigation and Simultaneous Localization and Mapping (SLAM) with reflective beacons, performs visual inspection for cracks and other defects, and when a crack is detected, repositions the crane and arm to place the ultrasonic sensor on the lining for width and depth measurements, while the laser profiler captures 3D deformation with 2 mm accuracy. Field trials in motorway tunnels demonstrate the complete inspection workflow and show reduced inspection times compared to manual methods while delivering comparable inspection quality (Menendez et al., 2018).

Earlier work by Yu et al. in 2007 presents a mobile robot-based crack inspection system for concrete tunnels that combines a two-wheel differential-drive platform with a line-scan camera, high-power illumination, and an image-processing pipeline for semi-automatic crack detection and measurement. The robot maintains a constant distance to the tunnel wall, acquires line-scan imagery at up to 5 km/h with a spatial resolution of 0.3 mm/pixel, and applies contrast enhancement, edge detection, graph-based crack tracing, and feature extraction to estimate crack length, orientation, and width, yielding measurement errors below 10%. Although the system focuses on 2D vision rather than full 3D evaluation, it demonstrates the feasibility of wheel-based tunnel inspection with higher speed and more objective results than manual inspections (Yu et al., 2007).

Quadruped robots

Quadruped robots are emerging as a promising mobile platform for autonomous tunnel inspection because they decouple the inspection system from the rail infrastructure while maintaining high mobility on uneven terrain. Feng et al. integrate a laser scanner with a legged inspection robot to realize rail-independent deformation inspection of metro tunnels, enabling dense point cloud acquisition in sections where trolley-based mobile laser scanners are impractical due to track layout. In their system, the quadruped robot carries a laser scanner through the tunnel and, together with terrestrial laser scanners, provides a point cloud dataset containing more than 574 mil-

lion points to support segmentation of tunnel components and deformation analysis. The approach achieves submillimeter accuracy (average error 0.60 mm) comparable to total stations while overcoming the low efficiency of static terrestrial laser scanning and was tested in various metro tunnel sites in Shanghai, China. (Feng et al., 2026).

Beyond pure defect inspections, quadruped robots are also being used as carriers in heterogeneous inspection approaches that combine legged mobility with aerial close-up sensing. Chu et al. propose a collaborative system composed of a quadruped robot and a collision-resilient micro-UAV for crack inspection in road tunnels, explicitly addressing GPS-denied navigation, poor lighting, and wet, obstacle-rich road surfaces that challenge wheeled or tracked platforms. The quadruped robot is equipped with stereo and panoramic cameras plus a VLP-16 LiDAR scanner to autonomously follow a pre-planned path to perform global scanning and patch-based CNN crack localization, while serving as a mobile helipad and communication relay for the UAV. Once potential damage regions are detected, the caged UAV is launched to perform detailed crack segmentation, allowing tiny cracks to be identified with over 93% segmentation accuracy in field tests (Chu et al., 2023).

Unmanned Aerial Vehicles (UAVs)

UAV tunnel inspection is increasingly defined by full-stack, field-tested systems that tightly couple navigation, sensing, and defect analytics in Global Navigation Satellite System (GNSS)-denied, low-light environments. Zhang et al. propose a Proximity-Pause-Photo for Surface Defect Inspection (PMPP-SDI) method, which uses a DJI Mavic 2 Enterprise with onboard lighting and omnidirectional obstacle sensing to execute a reactive grid-scan of tunnel walls. Instead of relying on pre-built maps, the UAV regulates lateral position via vision, height via infrared ranging, and axial motion via front/back obstacle distances, moving in short steps (1 m - 1.5 m) and pausing to acquire high-overlap image strips that later feed a structure-from-motion photogrammetry pipeline. A field test in a railway tunnel demonstrates stable navigation in darkness, high-resolution 3D reconstruction, and accurate surface defect localization, while highlighting remaining needs in higher-level autonomy, data quality control and integrated computer-vision-based defect analysis (Zhang et al., 2024).

Yang et al. developed a real-time tunnel lining inspection system that embeds an edge-AI stack and a dual RGB/infrared payload on a DJI Matrice 300 RTK platform to detect cracks, spalling, and leakages in TBM-driven tunnels. Positioning and tunnel coverage are enforced by AprilTag-based relative pose estimation, with an automated routine that captures 12 RGB and 12 infrared images per tunnel segment. Subsequently, the on-board YOLOv3-inference evaluates the data and returns the defect results, achieving the highest precision when combining RGB and thermal channels during field tests (Yang et al., 2022).

Evaluation of Mobile Platforms

Metro tunnel operating constraints

While robotic tunnel inspection literature often assumes highway or generic underground conditions, metro tunnels present distinct operational challenges that fundamentally constrain mobile platform selection. The following aspects are derived from the three complementary data sources (literature, interviews, observations) and later serve as an underlying basis for the evaluation of different mobile platforms.

Literature-based constraints:

- **GNSS-denied environment:** Absence of GPS demands robust onboard localization, while tunnel feature scarcity and repetitive geometries challenge SLAM performance (Schichler et al., 2025).
- **Low illumination:** Low-light tunnel environment requires the addition of external light sources for sufficient illumination and high quality data acquisition (Liao et al., 2022).
- **Dust pollution:** Metro tunnels show increased particulate matter concentrations (Smith et al., 2020), demanding sealed electronics, filtered intakes and protective covers for reliable sensor operation.

For the low illumination and dust pollution constraints, while they are derived from literature, in-tunnel observations confirmed the presence and operational impacts of these conditions.

Interview-based constraints:

- **Limited inspection time:** Interviewed metro operators in Germany reported that daily inspection windows during nighttime shutdowns are limited to approximately 3 hours of effective inspection time.
- **Railway track maintenance:** Major maintenance and repairs of railway tracks require temporary track closures, resulting in sections periodically unavailable for continuous track access (Dao et al., 2018). Interviewed metro operators confirmed that individual track sections are regularly unavailable for maintenance.

Observation-based constraints:

- **Railway track geometry:** Tracks, ballast and crossties require obstacle-straddling capability for non rail-bound platforms, ensuring reliable navigation and freedom of movement.
- **Metro tunnel access:** Entry and exit to metro tunnels can involve overcoming elevation changes such as stairs.

Platform Evaluation

Building on the general constraints of tunnel environments and the metro-specific constraints outlined previously, Table 1 evaluates four mobile platforms against criteria including payload capacity, speed, system runtime, deployment time, locomotion, and general considerations, de-

rived from inspection requirements.

Payload Capacity

Payload capacity directly determines sensor quality and availability in metro tunnel inspection. Rail-bound platforms excel with high possible payloads, supporting extensive sensor suites for data collection. Depending on the system size, wheeled and tracked platforms can achieve around 20 - 100 kg of payload capacity, offering potential for comprehensive multi-sensor setups, while quadruped robots are usually limited to around 10 - 40 kg. UAVs remain with the lowest payload capabilities of up to 2.7 kg for commercial drone systems, usually restricting them to camera setups.

Inspection Speed

Inspection speed reflects the traversal efficiency across tunnel geometries. Rail-bound platforms can achieve fast inspection speeds of up to 4.2 m/s due to their ease of locomotion in subway environments, while wheeled and tracked platforms reach a velocity of 2 m/s in non-track areas, further limiting the resulting inspection speed due to track-crossing and ballast navigation. Quadruped systems can maintain a speed of 0.75 - 1.6 m/s in non-track environments, however they are able to maintain more of that speed due to their obstacle avoidance capabilities. UAVs attain up to 20 - 23 m/s in ideal outdoor conditions, but tunnel constraints reduce practical speeds significantly.

System Runtime

System runtime depends on energy capacity and operational demands in tunnel conditions. Since inspection windows are usually limited to around 3 hours per day, this reflects the target system runtime for inspection operations without battery changes. As such, rail-bound platforms, wheeled and tracked systems and quadruped robots can reach system runtimes of up to 4 - 8 hours depending on the installed sensor systems, resulting in sufficient operation time for metro tunnel inspection windows. UAVs are usually limited to a shorter runtime due to limited battery-to-weight ratios, leading to required battery changes and inspection stops.

Deployment Time

Deployment time critically impacts the 3-hour inspection window. Developed rail-bound platforms require 3 - 5 minutes for assembly and disassembly tasks on tracks. Wheeled and tracked systems are estimated for around 10 - 15 minutes for deployment due to their bulky nature and difficult tunnel access. Quadruped robots, usually lighter and more flexible in motion capabilities, are estimated to require around 5 - 10 minutes, while UAVs excel with the fastest deployment due to their low size and weight.

Locomotion

Locomotion determines the access to diverse tunnel locations, while considering the complexity of system control. Rail-bound platforms use available rail systems with no possible deviation from the track layout, however they require very simple control to move along the tunnel. Wheeled and tracked platforms feature medium control complexity, while being able to navigate floors effectively but struggling with track straddling and poor object avoidance. Quadruped robots depend on complex system control, but reward high flexibility, being able to efficiently navigate around obstacles and provide large tunnel access including side tunnels not connected by tracks. UAVs require complex control due to aerial collision avoidance, but benefit by being able to inspect hard-to-reach locations that would be inaccessible by other systems.

General Considerations

General considerations highlight platform-specific limitations for metro tunnel operations. Rail-bound systems depend entirely on available tracks, excluding any areas currently under track maintenance. Wheeled and tracked systems suffer from poor obstacle avoidance and track-straddling features in complex metro environments. Quadruped robots benefit from stair-walking capabilities, while large UAVs face the risk of propeller-induced dust swirling, reducing visibility and damage detection potential.

Discussion

The comparative evaluation of mobile platforms reveals distinct strengths and weaknesses emerging from metro tunnel-specific inspection conditions, guiding deployment decision based on environment, access, data requirements and mission duration. Each platform benefits in particular aspects, which promotes scenario-specific deployment strategies rather than generic platform recommendations.

Rail-bound platforms excel in linear inspection with substantial payload capacity for comprehensive sensor suites, achieving fast traversal and extended runtime on established rails. They therefore perform well in long, straight sections with uninterrupted track access and limited track maintenance activity. Thus, their ideal use case involves track-maintenance free tunnels with uninterrupted track access, where large amounts of the tunnel surface can be inspected rapidly once the system is established.

Wheeled and tracked platforms offer balanced performance across payload, speed, and runtime, with robust sensor integration but limitations in obstacle avoidance and reliable track straddling. They perform best in relatively flat metro tunnel areas, where the ground surface is continuous and track-crossing poses minimal issues. In such scenarios, their locomotion simplicity supports efficient coverage of extensive tunnel sections, prioritizing horizontal mapping and data acquisition over vertical or complex geometries. However, they lag in environments with frequent ballast sections, crossties, and stations, where their limited obstacle-navigation reduces inspection capability.

Quadruped robots provide versatile and flexible locomotion with good stair-climbing and obstacle avoidance.

Table 1: Comparison of mobile platforms for automated metro tunnel inspection.

Criteria	**Rail-bound** (Huang et al., 2017) (Zhou et al., 2021) (Li et al., 2021)	**Wheeled and tracked** (Clearpath Robotics, 2020) (Clearpath Robotics, 2024)	**Quadruped** (ANYbotics AG, 2022) (Boston Dynamics, 2024) (Unitree Robotics, 2023)	**UAV** (DJI, 2021) (DJI, 2024)
Payload Capacity	high (total system weight 65 - 140 kg)	20 - 100 kg	10 - 40 kg	<0.2 - 2.7 kg
Speed	0.5 - 4.2 m/s	~ 2.0 m/s, in non-track environment	0.75 - 1.6 m/s, in non-track environment	20 - 23 m/s, outdoor and no-wind conditions
System Runtime	3 - 4 hours	2 - 8 hours	1.5 - 4 hours	30 - 55 min
Deployment Time	3 - 5 min, for each assembly and disassembly	10 - 15 min, bulky and difficult tunnel access	5 - 10 min, fast tunnel access and mid-range weight	<3 min, fast setup due to small size and weight
Locomotion	simple, no deviation from track possible	medium complexity, struggles with crossing tracks	complex, high flexibility	complex, able to inspect hard-to-reach locations
General Considerations	dependent on available tracks	poor 3D obstacle avoidance	stair-walking capability	risk of dust swirling from large drone propellers

They support comprehensive sensor configurations with sufficient system runtime for typical inspection windows available for metro tunnels, while excelling in complex-geometry environments such as station-adjacent tunnels, cross-passages, and elevated areas reachable by stairs. Their main drawbacks are slower inspection speeds compared with rail-bound platforms and higher control complexity. Nevertheless, they benefit in complex-geometry environments such as direct track proximity and inspection routes that require climbing stairs or navigating uneven terrain.

UAVs deliver fast deployment and unique aerial reach for otherwise inaccessible locations such as overhead and ceiling inspections, despite short runtime and payload constraints that are mostly limited to camera setups. They are therefore superior in confined, vertical, or hard-to-reach areas where ground-based access is limited. As such, they pose most valuable for quick vertical scans in confined areas or for temporary assessments and spot checks in areas not reachable by other platforms.

Conclusions and Outlook

This work evaluates mobile platforms for metro tunnel inspections, demonstrating how rail-bound, wheeled and tracked, quadruped, and UAV systems address distinct operational constraints through tailored strengths in system payload, mobility and deployment. The analysis identifies optimal deployment strategies matching platform capabilities to tunnel geometries, access challenges and time windows, advancing automated inspection and monitoring of metro tunnel infrastructure. In summary, the evaluation confirms rail-bound platforms suit long linear rail surveys, aiming to cover large portions of the tunnel surface, while wheeled and tracked systems excel in flat-floor mapping due to their balanced performance and ease of control. Additionally, quadruped robots enable access to complex areas with their high locomotive flexibility and stair-climbing ability, while UAVs offer quick system deployment, targeting short overhead inspections in hard-to-reach areas.

Future research can be conducted into hybrid inspection fleets combining multiple platforms for entire metro tunnel coverage under efficient system usage. Thus, the benefits and drawbacks of single systems can be compensated to allow for comprehensive metro tunnel inspections under varying conditions. Additional emerging research fields include multi-platform sensor fusion, enabling the data of different platforms and sensors to merge into single

evaluation models, which could leverage complementary strengths such as UAV overhead imagery with quadruped ground-level LiDAR for advanced defect mapping.

Acknowledgments

The authors would like to thank the German Federal Ministry of Transport (BMV) for funding this research within the mFUND initiative as part of the RoboTUNN project (funding code 01F2280A), as well as the associated project partners for their valuable collaboration and support throughout the research.

References

ANYbotics AG (2022). Anymal technical specifications: Datasheet.

Bao, X., Zheng, B., Shen, J., Chen, X., Kong, H., Li, J., Wang, X., and Cui, H. (2026). Intelligent technologies for tunnel construction and maintenance: A state-of-the-art review of methods and supporting platforms. Tunnelling and Underground Space Technology, 168:107207.

Bono, P.-H., David, Q., Desbordes, R., and Py, L. (2022). Metro infrastructure and metropolitan attractiveness. Regional Science and Urban Economics, 93:103757.

Boston Dynamics (2024). Spot technical specifications: Datasheet.

Chen, X., Zhang, Q., Liu, R., Wang, X., and He, W. (2023). Maintenance strategies and life-cycle cost analysis of inspection robots in metro tunnels. Tunnelling and Underground Space Technology, 140:105270.

Chu, H. H., Cao, R., and Deng, L. (2023). A collaborative inspection system composed of quadruped and flying robot for crack segmentation in tunnel environment. In Anagnostou, G., Benardos, A., and Marinos, V. P., editors, Expanding Underground - Knowledge and Passion to Make a Positive Impact on the World, pages 2406–2413. CRC Press, London.

Clearpath Robotics (2020). Jackal datasheet.

Clearpath Robotics (2024). Husky a300 datasheet.

Dao, C., Basten, R., and Hartmann, A. (2018). Maintenance scheduling for railway tracks under limited possession time. Journal of Transportation Engineering, Part A: Systems, 144(8).

DJI (2021). Mavic 2 enterprise user manual.

DJI (2024). Matrice 350 rtk user manual.

Feng, Y., Feng, S.-J., Zhang, X.-L., Kong, Q.-Z., and Zhao, Y. (2026). Automatic deformation detection of metro tunnels via point cloud segmentation and geometric analysis. Automation in Construction, 181:106657.

Huang, H., Sun, Y., Xue, Y., and Wang, F. (2017). Inspection equipment study for subway tunnel defects by grey-scale image processing. Advanced Engineering Informatics, 32:188–201.

International Association of Public Transport (2024). Global metro figures 2024.

Le, Y., Oka, M., and Kato, H. (2022). Efficiencies of the urban railway lines incorporating financial performance and in-vehicle congestion in the tokyo metropolitan area. Transport Policy, 116:343–354.

Li, D., Xie, Q., Gong, X., Yu, Z., Xu, J., Sun, Y., and Wang, J. (2021). Automatic defect detection of metro tunnel surfaces using a vision-based inspection system. Advanced Engineering Informatics, 47:101206.

Liao, J., Yue, Y., Zhang, D., Tu, W., Cao, R., Zou, Q., and Li, Q. (2022). Automatic tunnel crack inspection using an efficient mobile imaging module and a lightweight cnn. IEEE Transactions on Intelligent Transportation Systems, 23(9):15190–15203.

Liu, X., Zhu, L., Wang, Y., and Yu, Z. (2022). A crack detection system of subway tunnel based on image processing. Measurement and Control, 55(3-4):164–177.

Menendez, E., Victores, J. G., Montero, R., Martínez, S., and Balaguer, C. (2018). Tunnel structural inspection and assessment using an autonomous robotic system. Automation in Construction, 87:117–126.

Schichler, L., Festl, K., Solmaz, S., and Watzenig, D. (2025). Thermal-lidar fusion for robust tunnel localization in gnss-denied and low-visibility conditions.

Smith, J. D., Barratt, B. M., Fuller, G. W., Kelly, F. J., Loxham, M., Nicolosi, E., Priestman, M., Tremper, A. H., and Green, D. C. (2020). Pm2.5 on the london underground. Environment international, 134:105188.

United Nations (2025). World urbanization prospects 2025: Summary of results.

Unitree Robotics (2023). Unitree b2: Product page.

Yang, Z., Lee, W. C., Chan, H. N., and Ge, M. (2022). A real-time tunnel surface inspection system using edge-ai on drone. In 2022 IEEE International Conference on Mechatronics and Automation (ICMA), pages 749–754. IEEE.

Yu, S.-N., Jang, J.-H., and Han, C.-S. (2007). Auto inspection system using a mobile robot for detecting concrete cracks in a tunnel. Automation in Construction, 16(3):255–261.

Zhang, R., Hao, G., Zhang, K., and Li, Z. (2024). Reactive uav-based automatic tunnel surface defect inspection with a field test. Automation in Construction, 163:105424.

Zhou, M., Cheng, W., Huang, H., and Chen, J. (2021). A novel approach to automated 3d spalling defects inspection in railway tunnel linings using laser intensity and depth information. Sensors (Basel, Switzerland), 21(17).

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

OPTIMIZING LINEAR CONSTRUCTION PROCESSES: A SIMULATION-BASED APPROACH TOWARDS AI-SUPPORTED DECISION MAKING

Anna-Maria Mehringer[1,2], Simon K. Hoeng[1], Jonas Wiederer [1], and Mathias Obergrießer[1]
[1]Ostbayerische Technische Hochschule Regensburg, Regensburg, Germany
[2]Firmengruppe Max Bögl, Sengenthal, Germany

Abstract

Large-scale linear construction projects involve complex interdependencies, extended sites, and numerous execution variants, while existing approaches lack personalized recommendation systems for trenching applications. This paper presents a simulation-based framework conceptually enabling AI-driven optimization methods to support data-informed planning under limited historical data availability. A bottom-up simulation model captures process-specific dynamics and enables synthetic data generation, forming the basis for proposing focused parameter ranges with promising execution variants. Parametric validation across 3,200 simulation runs confirms realistic system behavior. The results indicate that AI-integrated simulation frameworks represent a viable and transferable approach for optimization in data-scarce linear construction environments.

Introduction

The European Union has established comprehensive guidelines for trans-European energy infrastructure to achieve climate neutrality by 2050 while ensuring energy security and market integration across Member States (European Parliament and Council, 2022). Priority electricity transmission corridors have been defined, particularly North–South interconnections in Central and South Eastern Europe, to support renewable energy integration into the internal market. To support this European objective, the Federal Requirements Plan Act has promoted 35 projects totaling approximately 14,000 km since 2013, yet by 2023 only about 4,000 km had been approved or completed (enercity, 2025). This discrepancy reflects persistent challenges in the planning, approval, and execution of large-scale linear infrastructure projects.

The scale and complexity of these projects limit the availability of robust empirical reference data and standardized methods for performance assessment and optimization. Construction processes are highly dependent on site-specific conditions such as geology, logistics, and resource availability, complicating systematic evaluation of alternative construction strategies. Simulation-based methods provide a promising means to address these challenges. By representing construction processes as parameterized models, simulation enables systematic analysis of construction sequences, resource interactions, and logistical dependencies. When combined with optimization methods, such approaches can support informed decision-making and contribute to the timely implementation of projects critical to achieving national and European climate and energy objectives.

Given the complexity of large-scale construction sites operated by numerous crews simultaneously, the integration of Artificial Intelligence (AI) for decision support represents a promising extension of simulation-based planning. Recent surveys indicate growing industry interest in AI applications for optimization tasks. While German construction companies increasingly employ data analysis and predictive modeling, personalized recommendation systems remain largely unimplemented despite planned investments by over one-third of surveyed firms with more than 250 employees (Engels et al., 2025). This indicates substantial development potential and demand for AI-supported optimization tools. Consequently, this work investigates whether AI-integrated simulation can effectively address the variant generation and optimization challenges inherent in large-scale linear construction projects.

State of the Art

In order to contextualize this paper within the existing body of research, relevant literature on simulation-based decision support and AI-integrated optimization methods is reviewed. Particular emphasis is placed on applications in industrial manufacturing and their transferability to construction domains. While construction projects present distinct challenges due to site-specific variability and data scarcity, methodologies proven in manufacturing environments provide valuable foundations for developing simulation-based planning approaches in linear infrastructure construction.

Relevant Related Work

In industrial manufacturing, the discrete event simulation (DES) is a well-established tool for decision support, system analysis, and performance evaluation. Negahban and Smith (2014) identify three primary areas of application:

- Manufacturing system design
- Manufacturing system operations
- Simulation language/package development

These categories can also be applied in the construction industry, as relevant construction literature subsumed under

these categories can be found. With respect to manufacturing system design, simulation-based approaches have been adopted to address construction site layout planning. This aspect is particularly critical for linear construction projects, where site layouts continuously evolve along the project corridor and material transportation distances represent a major cost and performance driver. Petroutsatou et al. (2021) proposed an algorithm for optimizing construction site layouts by dynamically relocating site facilities and storage areas, demonstrating reductions in transportation effort and overall operating costs. Similarly, RazaviAlavi and AbouRizk (2021) developed a simulation-based decision-support tool to evaluate alternative site layouts based on on-site transportation, construction logistics, and safety considerations.

Analogous to manufacturing system operations, simulation has also been applied to the analysis and optimization of construction processes themselves. Typical approaches vary equipment capacities, crew sizes, or task sequences and analyze resulting impacts on process performance through repeated simulation runs. König et al. (2007) applied this principle to drywall construction by simulating processes to analyze the impact of resource constraints and scheduling on efficiency.

More recent work extends this paradigm by systematically exploring large sets of feasible construction schedules instead of isolated scenarios. Morkos (2014) presents an approach that automatically generates schedules while enforcing precedence, resource, and spatial constraints. Rather than identifying a single optimal solution, the method enables the exploration of the feasible solution space and the analysis of trade-offs across multiple alternatives.

Within the category of simulation language and package development, a trend toward hybrid simulation frameworks can be observed, combining generic models, comprehensive system architectures, and integrated optimization methods. A similar development can be observed in construction-related research. Early work by Weber (2007) focused on the evaluation of logistics and site layout strategies for material flow in building construction. The developed tool, *SIMUBAU*, utilized CAD models to derive geometric and semantic model components and integrated these with databases and external programs for logistics and planning data. Within the research project *ForBAU*, a further framework was developed for earthworks in road construction. This approach employed CAD-based tools to generate voxel-based earthwork models and linked them with external databases providing simulation parameters, thereby improving both visualization and result analysis capabilities (Günthner et al., 2013).

More recent research extends these frameworks toward data-driven and continuously updated simulation environments. Jungmann (2025) proposes a digital twin-based approach in which real-time data collected on construction sites are processed to extract activity durations, calibrate stochastic simulation models, and generate key performance indicators (KPIs) for decision support. The approach further integrates machine learning-based activity recognition, stochastic productivity modeling, and DES to enable continuous model calibration and more accurate forecasting of construction processes.

While such approaches significantly enhance the realism and timeliness of simulation models, they primarily focus on improving model accuracy and enabling data-driven evaluation of predefined decision alternatives. Although multi-criteria optimization techniques are applied to identify promising solutions, these approaches do not inherently generate or adapt decision alternatives through learning-based mechanisms.

Addressing this gap as well as the third area of application, research has increasingly focused on integrating optimization methods with simulation frameworks. In construction, hybrid approaches combining simulation with genetic algorithms have been applied particularly for site layout planning, jointly addressing cost efficiency, safety, and productivity objectives (Said and El-Rayes, 2013; Papadaki and Chassiakos, 2016). However, these approaches remain largely static and do not adapt to changing site conditions. Evidence from related fields suggests that a more dynamic integration is feasible. In industrial manufacturing, Choi et al. (2024) and Klar et al. (2024) demonstrate that combining simulation with learning-based optimization enables continuous performance improvement. Kandemir et al. (2025) on the other hand, integrates DES with AI methods to optimize job scheduling under uncertainty. These results highlight the broader potential of adaptive, data-driven optimization, which remains largely untapped in construction site management.

Despite considerable progress in simulation-based decision support, the reviewed literature reveals limitations that need further investigation. Particularly in the construction domain, the integration of optimization methods with simulation frameworks remains underexplored compared to manufacturing applications. Moreover, recent advances in AI-driven optimization show promising results, yet their practical application in construction site management, particularly for personalized recommendation systems, remains limited.

Limitations and Open Directions in the Current State of the Art

A review of existing approaches indicates limitations concerning practical applicability, integration challenges, and data availability.

While numerous simulation frameworks have been developed for construction site management, their practical deployment remains limited. The *ForBAU* project's demonstrator exemplifies this challenge: despite representing a powerful approach, Wimmer et al. (2010) explicitly stated that expert knowledge is necessary to utilize the tool effectively. Most existing systems are developed primarily in research environments rather than for practical use by site managers, who would benefit most from them in daily

operations. Future research should therefore emphasize user-friendly interfaces and intuitive workflows that enable construction professionals without extensive expertise to apply these tools effectively for decision-making.

Despite significant advances in AI-driven optimization and their successful application in industrial manufacturing, the integration of these methods with construction site simulation frameworks appears to be an emerging area of research. While several studies have explored AI-based optimization using parametric BIM models (Hoeng et al., 2025; Wiederer et al., 2025), these primarily focus on optimizing building geometry, site layout, or isolated aspects such as resource allocation and scheduling. To the best of the authors' knowledge, comprehensive frameworks systematically integrating AI-based optimization with construction execution processes within a simulation environment are scarce in the current literature. Such frameworks would need to simultaneously consider dynamic scheduling, task allocation, resource constraints, logistical dependencies, and overall project costs while accounting for the strong interdependencies between site layout, transportation processes, and resource utilization.

A key barrier to AI adoption in construction is the lack of structured historical data for training machine learning models. Unlike manufacturing, where data is consistently captured in standardized environments, construction projects involve heterogeneous processes, site-specific conditions, and high contextual variability. This challenge is especially pronounced in linear infrastructure projects due to long durations, spatially distributed work fronts, and continuously changing site conditions.

Simulation-based approaches provide a way to potentially overcome this challenge through synthetic data generation. By simulating diverse construction scenarios with varying parameters and constraints, researchers can create comprehensive training datasets that capture complex interdependencies (Chan et al., 2022). Future research should investigate methodologies for generating high-quality synthetic construction data based on simulations that accurately reflect real-world variability, including validation frameworks and transfer learning approaches that effectively leverage synthetic data alongside limited real-world observations.

Within the specific domain of resource allocation optimization for linear construction sites using simulation-based approaches, several research opportunities emerge: (i) development of user-friendly simulation tools tailored to construction practitioners; (ii) integration of AI-based optimization methods with comprehensive construction process simulation; (iii) systematic approaches to synthetic data generation and validation; (iv) holistic frameworks capturing interdependencies across planning, scheduling, and execution; and (v) real-time adaptive optimization capabilities for dynamic construction environments.

In addressing the aforementioned opportunities, this work presents a framework that enables the concept of AI-based optimization in conjunction with construction site simulation by generating synthetic training data.

Methodology: Integrated Simulation Framework for Process Optimization

The methodology presented in the following is built upon a modular, process-oriented simulation framework specifically designed for open-cut trench construction in energy transmission projects. It provides the foundation for an AI-supported decision-making approach aimed at improving the allocation and coordination of earthwork resources. Figure 1 illustrates the overall structure and workflow of the proposed framework. The process begins with the collection and preprocessing of project-specific input data provided by the user.

The framework follows a hierarchical modeling strategy in which individual subprocesses are initially represented in a structured and modular form. These representations are used to derive simplified performance estimates that capture the essential behavior of each process step. The resulting information is then aggregated to form an abstracted representation of the overall construction workflow, linking the subprocesses in a sequential manner. Based on this abstraction, an AI-based component defines a constrained parameter space that focuses on promising execution strategies. Instead of evaluating isolated process steps, the simulation is applied at the level of the integrated system, allowing interactions between subprocesses to be considered. Within this reduced solution space, parametric simulations are conducted to assess different resource allocation strategies and their system-wide effects, from which KPIs are derived.

Finally, an iterative validation and refinement loop involving the user ensures the resulting scenarios are realistic, comparable, and optimized.

Input Information and Preprocessing

The proposed framework requires a set of project-specific input information provided by the user, which is accessible through an intuitive graphical user interface. The project's geometric and spatial characteristics must be specified, including segment cross-sections, lengths, radii, and gradients, as well as the intermediate storage location. Project-specific construction alternatives may also be considered, such as whether to improve existing soil in place or replace it entirely. Additional restrictions, such as the availability of machinery and labor, can likewise be incorporated. The user is further able to specify overarching project goals, for example maximum allowable cost or target construction duration. On the basis of these inputs, the simulation model as well as the appropriate parameter ranges for an optimized execution are subsequently generated.

Generation of the Parameter Space

Large scale linear construction projects are characterized by a high number of interdependent decisions. Even individual construction steps provide numerous configuration

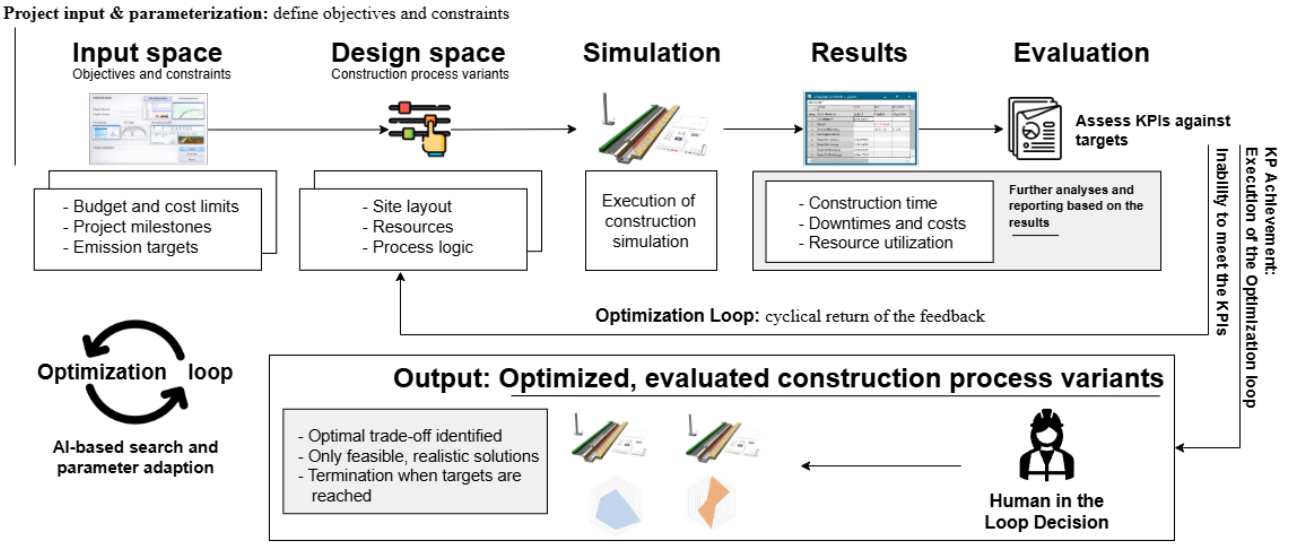


Figure 1: Holistic framework for AI-based parameter space generation and parametric simulation

options depending on equipment selection and logistical setup. When multiple crews operate simultaneously, identifying efficient task allocation and coordination strategies becomes increasingly difficult, making a comprehensive exploration of all possible configurations impractical.

To address this challenge, the proposed approach combines data-driven prediction with hierarchical aggregation to reduce the solution space. At its core, the method follows a bottom-up principle, first evaluating individual process steps independently. For each subprocess, a regression-based model is trained to predict the duration and cost based on the given input parameters. Processes that do not require logistical coordination can be evaluated through direct performance predictions. In contrast, logistics-dependent processes are further refined by optimizing relevant parameters using a differential evolution algorithm. This enables the identification of efficient equipment configurations and transport strategies.

This step provides KPIs for each individual process, including estimated duration and cost factors. Where applicable, it also provides optimized logistical configurations, all without requiring simulation at the subprocess level.

A higher-level learning component then aggregates these subprocess KPIs to derive system-level predictions, including overall project duration, cost development, task allocation, machinery utilization, and process synchronization. Although subprocess characteristics such as duration may initially serve as input variables, the model is also intended to capture inverse relationships. This indicates that adjustments at the subprocess level, such as longer execution times, may improve overall system performance. One example of this would be stabilizing process synchronization. In such cases, sub-processes can be re-evaluated in light of these updated objectives. This results in an abstract yet consistent representation of the overall construction workflow and defines a constrained parameter space that is likely to contain promising execution variants. Based on this reduced solution space, the simulation model is applied exclusively at the level of the integrated system to evaluate and compare these variants in detail.

In order to enable this predictive capability, the AI system must learn the complex interdependencies inherent to large-scale linear construction processes, which require sufficient high-quality training data. However, as previously mentioned, adequate volumes of such data are not available in a structured form. Therefore, synthetic training data is generated by systematically varying project and execution parameters in order to capture the underlying process interactions. This enables the AI models to generalize from limited empirical data and supports robust and optimized decision making.

Parametric Simulation for Energy Transmission Lines

To evaluate the parameter space proposed by the AI system and derive KPIs for execution variants, a parametric simulation model for energy transmission line construction is employed. The simulation follows a bottom-up approach and comprises three main components: a parametric representation of the working corridor, detailed models of individual process steps, and a holistic model of the entire open-cut construction process.

The working corridor model forms the foundation of the simulation and includes the trench cross-section, construction road, and soil storage areas. Key parameters such as trench depth, width, slope angles, layer heights, and number of pipes can be adjusted, ensuring broad applicability across different scenarios and projects.

This geometric representation provides the basis for calculating earthwork volumes and material quantities. Rather than deriving soil conditions from continuously updated 3D models, as demonstrated by Ji et al. (2009), they are provided directly by the user as input parameters. To ensure practical applicability, subsoil categories have been developed in collaboration with domain experts based

on the traditional German classification system for earthworks (DIN 18300, 2019). The categories are designed for rapid and reliable field assessment by construction personnel. Based on user-specified soil class percentages, stochastic models simulate the spatial distribution and behavior of different soil types within the trench.

Based on the project-specific working corridor, individual processes of open-cut construction are modeled separately. Five main processes have been identified: (1) topsoil removal, (2) trench excavation, (3) pipe zone construction, (4) trench backfilling, and (5) topsoil replacement. Each process model centers on an excavator that traverses the trench and moves material to its designated location. During (2) trench excavation and (3) pipe zone construction, additional transport vehicles are required to remove or deliver materials, as exemplarily shown for process (3) in Figure 2. Various logistic parameters can be adjusted within these individual process simulations. To ensure realistic representation, machinery does not operate at full capacity continuously. Instead, randomly distributed disruptions are integrated to reflect actual field conditions. Each process is simulated independently, allowing process duration and downtime to be determined individually based on the input parameters of each section.

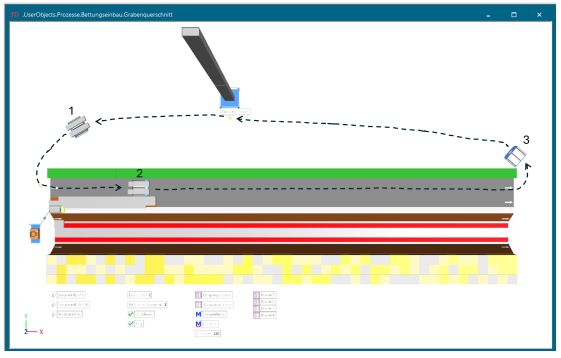


Figure 2: Parametric simulation of the pipe zone construction

These detailed process simulations provide the foundation for a holistic model that analyzes the interdependencies between different execution steps. This higher-level model is intentionally kept at a greater level of abstraction, using effort values and average durations derived from the process simulations as input parameters rather than replicating their logistical detail. At this level, tasks can be distributed across varying numbers of excavators and scheduled with different timing patterns, enabling identification of the fastest and least disruption-prone construction sequence. This combination of detailed process-level and abstract system-level simulation has proven effective for optimizing earthwork processes, as Ji et al. (2011) have similarly shown.

Simulation-Based Performance Metric Derivation

Each simulation yields performance metrics related to time, economics, and sustainability. Following the framework's bottom-up structure, these metrics are first determined for each process step based on the deployed equipment and project parameters. Then, they are aggregated at the system level. The calculations are based on a comprehensive database of effort values, cost factors, and machinery emission data provided by industry partners. This allows for the systematic comparison and evaluation of different execution variants.

Human Validation, Refinement and Final Output

The derived KPIs are evaluated based on predefined project goals by combining quantitative criteria with qualitative user judgment about the plausibility of each variant. If the resulting variants do not satisfy the specified criteria, the relevant parameters are adjusted iteratively, and a new optimization loop is initiated. This process continues until either a sufficient number of viable execution variants have been generated or the predefined termination criteria have been met. The result is a set of feasible, optimized construction scenarios that align with project-specific requirements.

Verification of the Simulation

The simulation model was developed using *Tecnomatix Plant Simulation*, with most methods implemented in the platform's native programming language, *SimTalk*. Technical verification was conducted iteratively throughout the development process using integrated debugging tools. This iterative verification process ensured that the model implementation accurately reflected the intended conceptual design.

Structured historical performance data for linear construction processes is scarce, and existing documentation is influenced by numerous confounding factors, such as weather conditions, regulatory delays, and crew-specific practices. This makes it difficult to isolate baseline process performance. Consequently, traditional validation approaches relying on historical performance data were not fully applicable. To address these limitations, the simulation model was developed in close collaboration with construction personnel, ensuring alignment with practical experience. Performance factors were verified through spot checks on active sites, and experienced professionals confirmed that the simulation produces plausible results, including realistic daily work targets.

Table 1: Simulation parameters and their value ranges

Logistic Parameter	Lower Limit	Upper Limit	Steps
Vehicle Count	1	10	1
Cycle Time [min]	15	120	15
Distance to Storage [km]	5	40	5

To further validate the realism of the simulation, systematic experiments were conducted by varying key input parameters. At this stage, only the temporal output is fully implemented, therefore, the experiments focus exclusively on construction time as the performance metric. The input parameters including vehicle count, cycle time, and distance to the intermediate storage facility were varied across the ranges shown in Table 1. These were defined by experienced site personnel to represent realistic conditions. Each parameter combination was simulated with five different random seed values to account for stochastic variability in simulation outcomes. This yielded a total of 3,200 simulation runs, covering a comprehensive and structured parameter space. The resulting simulation outputs were subjected to comparative sensitivity analyses to verify whether parameter variations produced effects on construction time and resource utilization consistent with empirical expectations and practical construction experience.

Due to their independence from logistic parameters, the processes of (1) topsoil removal, (4) trench backfilling, and (5) topsoil replacement displayed only limited temporal variability, driven primarily by randomly distributed disruptions. The resulting standard deviation was approximately 20 minutes for trench backfilling and about 9 minutes for topsoil removal and replacement. In contrast, (2) trench excavation and (3) pipe zone construction exhibited substantially higher temporal variability due to multiple logistical influences, with a standard deviation of approximately 3.5 hours for pipe zone construction and around 21 hours for trench excavation. Consequently, trench excavation emerges as the process with the greatest influence on overall schedule variability.

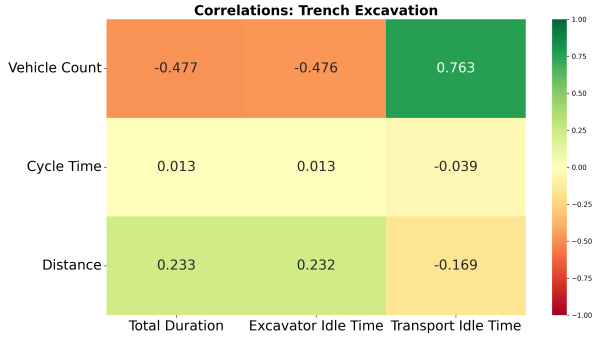


Figure 3: Heatmap of the trench excavation parameter space

To reveal the influence of different logistic parameters more clearly, a correlation analysis was conducted. The results are presented in Figure 3 and Figure 4.

The vehicle count shows a strong negative correlation with total duration (r = -0.477 for trench excavation, r = -0.464 for pipe zone) and excavator idle time (r = -0.476 and r = -0.464), and a strong positive correlation with transport idle time (r = 0.763 and r = 0.751). The parameter cycle time demonstrates minimal correlation with all output metrics in trench excavation (|r| < 0.04). In the pipe zone process, correlations are slightly higher (r ~ 0.09 for duration and excavator idle time, r = -0.312 for transport idle time). Distance to storage location shows moderate positive correlations with total duration and excavator idle time (r = 0.23 for trench excavation and r = 0.454 for pipe zone), and negative correlations with transport idle time (r = -0.169 and r = -0.179).

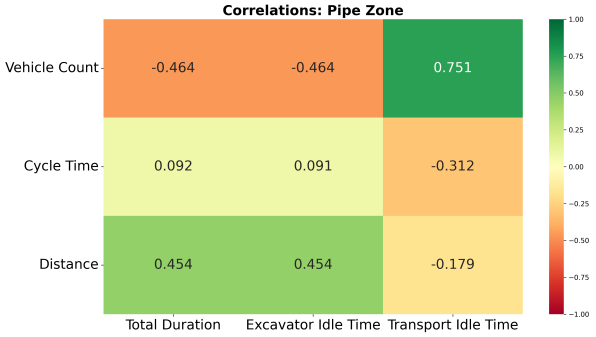


Figure 4: Heatmap of the pipe zone parameter space

Discussion and Result Analysis

The parametric validation confirmed that the modeled relationships between parameters and system behavior are plausible and aligned with observed construction practices. Increasing vehicle count effectively reduces construction time and excavator idle time by ensuring continuous excavator operation. However, beyond a certain threshold, adding more vehicles results in smaller reductions in duration while substantially increasing idle time for the vehicles. This reflects a fundamental trade-off between excavator utilization and transport fleet efficiency.

The analysis of cycle time and distance to storage reveals distinct operational characteristics between trench excavation and pipe zone construction. In trench excavation, transport vehicles travel alongside the excavator during continuous loading, making the excavator's digging speed the limiting factor regardless of vehicle timing or transport distance. In contrast, pipe zone construction involves rapid unloading where vehicles simply discharge embedment material next to the trench and depart, operating independently of the excavator. Poorly synchronized cycle times in relation to storage distance can cause multiple vehicles to arrive simultaneously, creating bottlenecks and increased idle time. Nevertheless, the influence of the cycle time remains largely negligible except for its impact on transport vehicle idle time. This is particularly favorable from a practical standpoint, as cycle time is difficult to control in real construction scenarios due to synchronized work schedules and operational constraints.

The reliance on expert judgment and parametric validation does highlight limitations imposed by the absence of comprehensive historical performance data. Ongoing efforts focus on establishing structured documentation protocols for future construction phases, enabling more rigorous quantitative validation and supporting continuous model refinement.

Table 2: Comparison with existing tools

Criterion	proposed tool	ForBAU	Jungmann
Framework type	AI-integrated simulation-based optimization	3D simulation and decision support	Digital twin-based decision support
Domain focus	Linear construction sites (trenching)	Earthwork (road construction)	Repetitive construction site processes
Primary objective	automated generation, evaluation and optimization of resource- and schedule-related execution variants	integrated data management, simulation-based process optimization, dynamic 4D construction site information modeling	real-time data-driven simulation, KPI generation and decision support through continuous model calibration
Key innovation	Automated AI-supported variant generation and optimization	Resistance coefficient-based process modeling, holistic machine fleet interaction simulation	ML-based activity recognition and real-time model calibration
Spatial resolution	process - / segment-based	voxel-based	process-based

Conclusions and Scientific Contribution

This work introduces a framework combining parametric simulation with AI-enhanced parameter optimization to support decision-making in large-scale linear trenching projects. By enabling the automated generation and evaluation of execution variants, the framework addresses a key limitation of existing decision support tools, which typically rely on manually defined scenarios.

Compared to existing approaches such as ForBAU and the digital twin-based framework of Jungmann (2025), the proposed framework is specifically tailored to linear construction sites. While ForBAU applies voxel-based terrain models for road earthworks and Jungmann focuses on continuous model calibration from real-time sensor data, the proposed approach adopts a process- and segment-based spatial resolution aligned with the characteristics of trenching projects. This reduces data requirements and model complexity while maintaining practical relevance for construction planning.

A central contribution of the framework is the generation of synthetic training data through systematic simulation of diverse construction scenarios, addressing the lack of structured historical data. This potentially enables the AI component to learn process interdependencies and identify promising parameter ranges. Parametric analyzes show that the simplified simulation captures essential dynamics, while a user-oriented design based on familiar soil classifications ensures practical applicability without requiring detailed geotechnical input.

Despite promising results, several limitations remain. Cost and sustainability indicators are only partially integrated, and validation relies primarily on expert assessment and sensitivity analysis due to limited empirical data availability. Rather than adopting a sensor-based approach as proposed by Jungmann (2025), this work relies solely on simulated synthetic data, for which initial tests deliver encouraging results. Terrain and soil modeling remains simplified, restricting the representation of spatial heterogeneity, and the framework currently covers only the open-cut construction phase.

Two directions for future work deserve particular attention. First, the AI models must be trained across a sufficiently diverse set of cross-sections and project configurations to ensure robust generalization. Should the underlying construction processes or project characteristics change significantly, retraining may be required. Ongoing efforts therefore focus on making the data generation and training workflow as reproducible as possible to minimize the overhead of future retraining cycles. Second, the current workflow depends on licensed simulation software, which may limit accessibility. Once AI models are sufficiently trained, bypassing the simulation for routine optimization tasks represents a promising step toward broader practical applicability.

Overall, the results indicate that AI-integrated methods based on simulation frameworks offer a viable path toward data-driven optimization in linear construction, even under significant data scarcity. The proposed methodology for simulation-based synthetic data generation is transferable to other construction or production processes with similarly definable constraints, providing a solid foundation for more informed planning of large-scale linear infrastructure projects.

Acknowledgments

This research was funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy under the Bavarian Collaborative Research Program (BayVFP) as part of the KIREL project.

References

Chan, K. C., Rabaev, M., and Pratama, H. (2022). Generation of synthetic manufacturing datasets for machine

learning using discrete-event simulation. Production & Manufacturing Research, 10(1):337–353.

Choi, H., Yu, S., Lee, D., Noh, S. D., Ji, S., Kim, H., Yoon, H., Kwon, M., and Han, J. (2024). Optimization of the factory layout and production flow using production-simulation-based reinforcement learning. Machines, 12:390.

DIN 18300 (2019). Erdarbeiten. Allgemeine Technische Vertragsbedingungen für Bauleistungen (ATV), VOB/C.

enercity (2025). Netzausbau 2024: Aktuelle zahlen und entwicklungen.

Engels, B., Scheufen, M., and Schmitz, E. (2025). Künstliche intelligenz als wettbewerbsfaktor für die deutsche wirtschaft: Empirische befunde und handlungsempfehlungen zum einsatz von ki in deutschen unternehmen.

European Parliament and Council (2022). Regulation (eu) 2022/869 on guidelines for trans-european energy infrastructure.

Günthner, W. A., Kessler, S., Frenz, T., and Wimmer, J. (2013). Transportlogistikplanung im erdbau: Entwicklung eines simulationsverfahrens zur optimierung der transportlogistischen ablaufplanung im erdbau basierend auf dem ansatz der berechnung von stetigfördern im bereich der fördertechnik. Technical report, Technische Universität München. Forschungsbericht.

Hoeng, S. K., Wiederer, J., Eder, F., Obergriesser, M., and Linner, T. (2025). Towards AI-enhanced Facade Planning: Integrating Human Expertise with Machine Learning-driven Parametric Modeling. In Proceedings of the 2025 European Conference on Computing in Construction.

Ji, Y., Bormann, A., Rank, E., and Wimmer, G. (2009). An integrated 3d simulation framework for earthwork processes. In Managing IT in Construction / Managing Construction for Tomorrow, pages 47–54. CRC Press.

Ji, Y., Borrmann, A., and Wimmer, J. (2011). Coupling microscopic simulation and macroscopic optimization to improve earthwork construction processes. In Proc. of the EG-ICE Workshop on Intelligent Computing in Engineering.

Jungmann, M. (2025). Exploring construction process simulation as part of the digital twin concept in real-world contexts. PhD thesis, Technische Universität Berlin.

Kandemir, C., Herod, S., Ghojogh, M., Naraghipour, A., Franchino, M., and Soulatiantork, P. (2025). Discrete event simulation with reinforcement learning for advanced manufacturing decision-making. In Proceedings of the Operational Research Society Simulation Workshop 2025 (SW25), pages 266–274. Operational Research Society.

Klar, M., Schworm, P., Wu, X., Simon, P., Glatt, M., Ravani, B., and Aurich, J. C. (2024). Transferable multi-objective factory layout planning using simulation-based deep reinforcement learning. Journal of Manufacturing Systems, 74:487–511.

König, M., Beißert, U., Steinhauer, D., and Bargstädt, H.-J. (2007). Constraint-based simulation of outfitting processes in shipbuilding and civil engineering. In 6th EUROSIM Congress on Modelling and Simulation, volume 2, Ljubljana, Slovenia. ARGESIM. Full Papers (CD).

Morkos, R. (2014). Visualizing, Manipulating, and Navigating the construction scheduling state space with precedence, discrete, and disjunctive constraints. PhD thesis, Stanford University.

Negahban, A. and Smith, J. S. (2014). Simulation for manufacturing system design and operation: Literature review and analysis. Journal of Manufacturing Systems, 33(2):241–261.

Papadaki, I. N. and Chassiakos, A. P. (2016). Multi-objective Construction Site Layout Planning Using Genetic Algorithms. Procedia Engineering, 164:20–27.

Petroutsatou, K., Apostolidis, N., Zarkada, A., and Ntokou, A. (2021). Dynamic Planning of Construction Site for Linear Projects. Infrastructures, 6(2):21.

RazaviAlavi, S. and AbouRizk, S. (2021). Construction Site Layout Planning Using a Simulation-Based Decision Support Tool. Logistics, 5(4).

Said, H. and El-Rayes, K. (2013). Performance of global optimization models for dynamic site layout planning of construction projects. Automation in Construction, 36:71–78.

Weber, J. (2007). Simulation von Logistikprozessen auf Baustellen auf Basis von 3D-CAD Daten. PhD thesis, Universität Dortmund.

Wiederer, J., Höng, S., Görgner, M., and Mehringer, A.-M. (2025). End-to-end parametric workflow for bridge design with automated cost analysis and generative modeling. In Tagungsband des 36. Forum Bauinformatik: 24.-26. September 2025.

Wimmer, J., Horenburg, T., Günthner, W. A., Ji, Y., Borrmann, A., and Rank, E. (2010). Evaluierung einer 3d-modell-basierten ablaufsimulation von erdbauprozessen in der praxis. In Integrationsaspekte der Simulation: Technik, Organisation und Personal, pages 142–148. KIT Scientific Publishing.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AN INTEGRATED SENSOR DATA AND MACHINE LEARNING SYSTEM FOR WORK PROGRESS MONITORING IN INDUSTRIALIZED CONSTRUCTION

Eitan Ballivián[1], Tomás Guzmán[1], Caroline Araújo[1,2], Pablo Martinez[3], Tito Arévalo[1], Beda Barkokebas[1]
[1]Pontificia Universidad Católica de Chile, Chile
[2]Universidade Estadual de Feira de Santana, Brazil
[3]Northumbria University, England

Abstract

Industrialized Construction (IC) is an approach to address ongoing challenges in the construction industry. Barriers affecting IC adoption include manufacturing inefficiencies caused by inaccurate production monitoring on shop floors. This paper presents the integrated use of machine learning and sensor data collected from IMUs and load cells installed on manufacturing platforms to monitor work progress. The research method includes the development of data collection, processing, and data analysis methods in a controlled laboratory, followed by validation on a real-world shop floor. Results present an accurate model robust against stochastic noise, achieving 77% F1-score monitoring through temporal windows and weighted deltas.

Introduction

The construction industry faces several challenges related to low productivity on-site (Kauppinen et al., 2024). The traditional approaches to monitoring production in on-site construction rely on error-prone and time-consuming manual operations, which lead to ineffective resource allocation, quality failures, high volumes of waste, rework, delays, and budget overruns (Hasan and Sacks, 2023).

Considering this context, Industrialized Construction (IC) has been pointed out as a viable alternative to address these gaps, adapting traditional construction methods through the integration of manufacturing-oriented design and optimization tools (Qi et al., 2021). IC involves performing activities and producing construction components in a controlled shop floor environment before assembling on-site (Melgar et al., 2025).

Despite its potential, current obstacles affecting IC performance include the difficulty in monitoring the production and identifying bottlenecks on the shop floor in a timely manner, thus requiring a change from manual methods (e.g., stopwatches, visual observation, or analysis of pre-recorded video footage) to automated methods (Rashid and Louis, 2020). Indeed, manual methods lack accuracy in representing actual working conditions (Chen et al., 2024). Besides, Barkokebas et al. (2023) argue that current strategies and decision-making processes in IC shop floors are largely based on workers' personal experience rather than on empirical and real-time data, which hinders process improvement. Moreover, Qi et al. (2021) argue that the adoption of emerging digital technologies can address current obstacles and thereby unlock the full potential of IC.

A step toward increasing productivity in IC is the implementation of automated work monitoring and tracking techniques, enabling managers to determine what and when work is being performed on the shop floor. Despite recent advances, existing techniques exhibit significant room for improvement. Authors such as Panahi et al. (2022), Chen et al. (2024), and Martinez et al. (2021) have applied computer vision (CV) in IC for progress monitoring and productivity measurement. However, this technique presents limitations due to general occlusions, illumination issues, and limited field of view (Alsakka et al., 2023). Although static cameras can provide an adequate viewing angle during the fabrication of linear or 2D components (i.e., wall and floor panels), it is challenging to follow the manufacturing of volumetric modules, considering that the components closer to the camera can occlude those positioned behind them (Rashid and Louis, 2020; Alsakka et al., 2023), resulting in a "black box" effect.

Other authors, such as Demiralp et al. (2012), Altaf et al. (2018), and Li et al. (2017), have proposed the use of tracking solutions based on radio frequency identification (RFID) and real-time location systems (RTLS). However, despite the capability of tracking the location of different elements, these techniques have limitations in interpreting the context and identifying the activities performed. In light of this, Rashid and Louis (2020) proposed a method based on machine learning (ML) that analyzes the sound of the shop floor environment to automatically identify activities, such as hammering and nailing. The main limitation is the difficulty in isolating and classifying only the sounds of interest in real-world conditions. Rashid and Louis (2020) also highlighted the advantages of inertial measurement unit (IMU) techniques, which can identify activities through parameters, such as motion and vibration. However, they argue that in previous applications, each sensor is attached individually to each element of interest, causing practical difficulties due to the multiple components and elevated costs.

Rashid and Louis (2021) proposed a framework for modular construction to automatically measure active and idle times at workstations using IMU vibration data and a

deep learning network. However, the framework limitations include issues related to data resolution, sampling frequencies, ground-truth labeling, and the lack of clearly defined methodological steps.

Considering the gaps and limitations highlighted in the literature, this paper proposes a novel system combining ML models with data acquired from sensors to monitor manufacturing activities of volumetric modules on IC shop floors. Real time data is collected from IMUs and load cells installed beneath the monitored IC workstations to identify whether work is being actively performed at workstations and therefore determine work progress. This paper's contribution is bi-fold: (1) it proposes ML-based methods to monitor manufacturing IC activities in real-time, and (2) it demonstrates their application.

Research methods

This paper presents the development of a system in a laboratory setting, followed by its final testing and validation on a shop floor in a bathroom pod in the Los Lagos Region, Chile. By applying both experiments, this research aims to validate the following hypothesis: "*The integrated use of sensors, such as IMUs and load cells, coupled with ML models, enables activity classification and work progress monitoring for manufacturing operations of volumetric modules on IC shop floors.*" Additionally, this research aims for the proposed system to achieve a technology readiness level (TRL) 5, in which the system is validated in a relevant environment (Kimmel et al., 2020).

As illustrated in Figure 1, the research steps and its workflow are divided into Laboratory and Shop Floor phases. Following its initial development in the Laboratory, the proposed method is then adapted and validated on the Shop Floor. The Laboratory phase follows a controlled experiment design as per the steps below:

- **Data acquisition:** Design of the experimental setup and data acquisition.
- **Data processing and algorithm development:** Conducting an exploratory data analysis (EDA) for ML model development and refinement.
- **Internal validation:** Checkpoint to assess performance within Laboratory experiments.

If the internal validation criteria are not met, the workflow loops back to the algorithm development stage for further refinement. Conversely, successful validation allows for further development of the proposed system. Following a quasi-experiment research design, the Shop Floor phase consists of the following stages:

- **System's adaptation:** Adapting the proposed system to shop floor conditions, including data acquisition procedures and the ML model.
- **External Validation:** Assessment of the proposed system in real-world operating conditions.

To verify the presented hypothesis, the proposed system relies on the monitored workstation being transformed into an instrumented platform (i.e., physical platform with sensors placed under it) so the manufacturing operations can be monitored as per Figure 2.

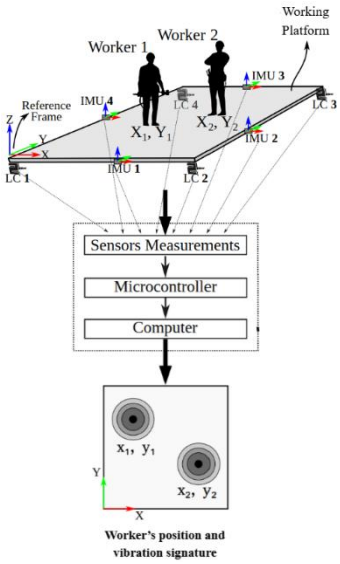


Figure 2: Graphical representation of the experimental setup.

The specific technical details of each environment are elaborated upon in the following subsections.

Laboratory experiment

A timber frame platform was built in the laboratory to replicate conditions in the shop floor and install sensors underneath it. The platform built measured $2{,}40 \times 2{,}40\ m$ and was designed to maintain structural stability during experiments.

The primary objective of this phase was to generate a high-fidelity dataset that captures the distinct kinematic and weight patterns of typical production activities as demonstrated in Figure 3a. This controlled environment allowed for the development and refinement of the machine learning pipeline, which was executed using the RapidMiner platform to evaluate model performance before industrial deployment.

Data acquisition

The data acquisition (DAQ) network comprised the following components:

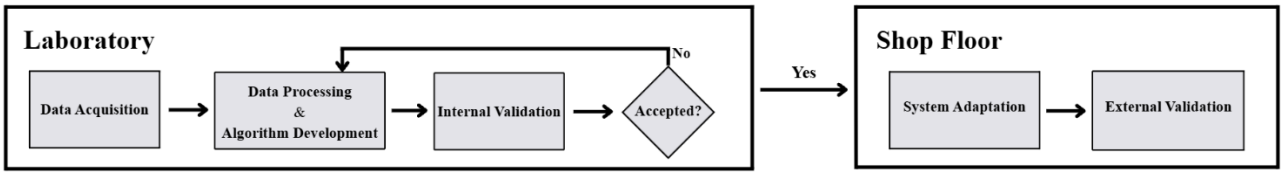


Figure 1: Research Methodology for the 2-Phase System Development and Testing.

- **Load cells:** Four industrial TRB load cells (15-ton capacity, 2.85 mV/V rated output) were installed under the platform at each corner. The weight signal (in kg) was acquired using HX711 ADC modules.
- **IMUs:** Four MPU-6050 3-axis IMUs were equidistantly mounted at the center of each lateral side of the base. The data was recorded in m/s^2.
- **Data management:** The hardware architecture was governed by an ESP32 microcontroller for each load cell and IMU pair, which facilitated data sampling at a constant Hz. Raw data transmission was executed via MQTT protocol and consolidated into temporal blocks for subsequent analysis, with each block being transmitted every second.

To replicate conditions typical of IC shop floors, several activities, such as "None", "Walking", "Hammering", and "Drilling" were executed in the laboratory. These were all recorded by the proposed system and video cameras. Furthermore, composite activities (i.e., two activities performed at the same time) were also performed and recorded. Recording manufacturing operations from both the proposed system and video cameras allows researchers to train and validate ML models to classify whether work is being performed or not under the platform, as demonstrated in Figure 3b.



Figure 3A Figure 3B

Figure 3: Images of the laboratory's experimental setup.

Data from the proposed system was acquired via the load cells and IMUs, transmitted by the ESP32 controllers and subsequently concatenated using Python scripts. The final data structure was exported as a comma-separated values (CSV) file containing nine primary features: a timestamp, four independent weight readings (w_1, w_2, w_3, w_4), and four IMU sensor streams (s_1, s_2, s_3, s_4). Tri-axial IMU measurements within each stream were combined using a semicolon (;) as the delimiter. Recordings from the video cameras were also stored locally and later processed to train the ML classification models.

Data processing and Algorithm Development

The raw CSV dataset was processed through a Python pipeline to enhance feature compatibility for the ML model. In that spirit, each rotational velocity column was decomposed into its tri-axial components (x, y, z), generating a total of 12 inertial features. This decomposition provides a higher dimensional representation of the kinematic signatures inherent to each activity.

To handle temporal resolution, a temporal alignment procedure was implemented. Given the 10 Hz sampling rate, 10 discrete rows were generated for every second of data, ensuring each IMU measurement was mapped to its specific time slot.

Final synchronization of the temporally aligned data relied on a multi-modal verification process. The documented activities timestamps were validated against the video recordings to guarantee ground-truth accuracy. This processed dataset was then organized into a matrix comprising seventeen variables, including: four weight readings, twelve inertial axes, and the target label, as detailed in Table 1.

Table 1: Initial Dataset: first three observations

Obs.	Weights (kg)	Acc. (m/s^2)	Target Variable
1	$w_{1,(1:4)}$	$a_{1,(1:4)}$	1_{st} value
2	$w_{2,(1:4)}$	$a_{2,(1:4)}$	2_{nd} value
3	$w_{3,(1:4)}$	$a_{3,(1:4)}$	3_{rd} value

An exploratory data analysis was then conducted on it to evaluate data quality and distribution before modelling. The primary objectives of this phase included:

- Analyzing the class distribution and balance across the dataset.
- Implementation of noise reduction and data cleansing protocols.

Internal Validation

The target variable was defined as a multiclass Activity feature. In instances of concurrent activities, a hierarchical labeling priority was implemented ("Walking" > "Hammering" > "Drilling") to maintain a mutually exclusive, single-label-per-timestamp structure.

To evaluate the contribution of each sensor modality, the models were subjected to an ablation study involving three distinct configurations: (1) load cells and IMU data, (2) load cells data exclusively, and (3) IMU data exclusively. The macro-averaged F1-score performance metric for each configuration was recorded and compared to assess the relative predictive power of each sensor type.

This metric was selected to mitigate the risk of an "Accuracy Paradox", "Walking" and "None" activities were more present in the dataset than others. This allowed that the high frequency of "None" and "Walking" couldn't mask poor model performance on shorter, more high-intensity tasks like "Hammering" or "Drilling".

The development of this phase successfully places the system in a TRL 4. By validating the technical integration of the hardware components and the initial classification models, this phase demonstrates that the basic technological elements can function together as a cohesive system in a laboratory environment, establishing a solid proof-of-concept before subjecting the system to the complexities of a real-world IC shop floor environment.

Shop floor experiment

System's adaptation

The experiment was conducted at an active bathroom module production facility using a $2{,}40 \times 1{,}80\ m$ mobile industrial platform. The platform featured a wooden deck integrated into a steel frame and, while equipped with casters for transit between production stages, it was secured in a stationary position for the duration of all data acquisition sessions, which were also recorded for ground-truth synchronization.

The experimental protocol mirrored the laboratory phase, with the only difference being that the sampling frequency was increased from 10 Hz to 90 Hz.

However, the dataset itself was processed through three extra stages to transition from the raw data-driven approach used in the Internal Validation to a robust supervised learning architecture. This progression was designed to better interpret the more complex physical signatures within the high-noise environment of industrial production facilities.

External validation

To isolate the sensing system from environmental noise, the platform was elevated by the load cells themselves, effectively bypassing the casters. This "floating", configuration visible in Figure 4**,** ensured that all vertical loads and mechanical vibrations were transmitted exclusively through the load cells and the DAQ system, replicating the controlled environment of the IMA laboratory.

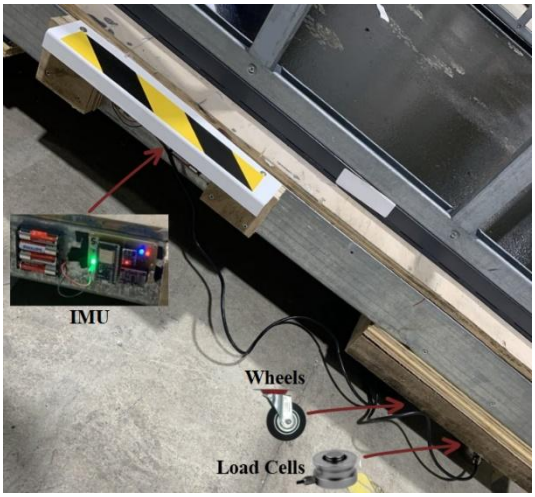


Figure 4: Shop Floor's experimental setup.

Operational activities (categorized as Working) were defined based on the assembly protocols established in the manufacturer's technical manuals, which were comprised of sub-activities like the "Walking", "Hammering", and "Drilling" activities tested in the laboratory. This label was assigned to any Value-Added (VA) construction task performed on the module structure, regardless of whether the operator was positioned on the platform or the factory floor.

Conversely, non-contributing activities were classified as Not Working. This category encompasses Non-Value-Added (NVA) tasks such as platform cleaning, the handling of secondary materials (e.g., relocating toolboxes), or worker random movement across the platform.

The development of this phase represents an advancement beyond laboratory experiments by exposing the system to the operational constraints inherent to an active IC shop floor. It marks the beginning of the transition towards TRL 5, allowing for verification of whether the ML model architecture and the system in general maintain their predictive capacity and robustness in a relevant environment.

Results and discussion

The performance of the proposed classification architecture was evaluated through a two-stage validation process. First, the laboratory dataset was used to develop and benchmark the algorithm performance and select a lead architecture. Second, the model was subjected to an external validation phase using the 90 Hz industrial shop floor dataset, where the impact of feature engineering and temporal windows was analyzed. This dual approach ensures that the final model is not only statistically accurate but also robust against stochastic noise, inherent to IC environments.

Laboratory experiment results

The Temporal Alignment procedure for the IMA dataset resulted in 28,120 observations, representing approximately 52 minutes of operational data. The class distribution was intentionally balanced across seven recorded activities to ensure the model could distinguish between similar kinematic signatures.

Five machine learning algorithms were implemented in RapidMiner to establish a performance baseline. As demonstrated in Table 3, the Random Forest (RF) ensemble emerged as the superior model with 86.77% accuracy. This performance is attributed to the RF's inherent bagging approach, which offered significant robustness against the inherent sensor noise and local outliers' data. Finally, the RF's built-in feature importance provided a layer of interpretability, allowing for an internal validation of which specific inertial signals were most indicative of the activities monitored.

Table 3: RapidMiner ML models average accuracy of the ablation study.

Model	Parameters	Accuracy
Random Forest	100 trees, max depth 15	86,77%
Gradient Boosted Tree	50 trees, max depth 7	83,81%
k-NN	k = 5	78,92%
SVM	RBF kernel	75,61%
Naive Bayes	10 Kernels	73,3%

Table 2: Random Forest's ablation study results

Activity	IMU and Load Cells		Load Cells Only		IMUs Only	
	Precision	Recall	Precision	Recall	Precision	Recall
Walking	89.56%	96.47%	97.95%	96.29%	71.34%	94.01%
Hammering	94.53%	72.49%	78.03%	98.24%	97.31%	50.70%
Drilling	91.06%	87.45%	91.38%	98.54%	91.32%	61.77%
None	91.89%	91.07%	97.23%	93.03%	86.47%	79.87%
F1-Score	88.91%		93.50%		76.13%	

Ablation Study

To further understand the contribution of each data source (IMU and load cells), an ablation study was conducted using the RF algorithm. The results, visible in Table 2, reveal a significant disparity between sensor modalities.

The Load Cells Only configuration achieved the highest overall performance with an F1-score of 93.50%. Specifically, the Recall for "Hammering" (98.24%) and "Drilling" (98.54%) indicate that the kinetic energy and vibration patterns transferred to the platform are highly distinct. This suggests that force-based sensing captures the "mechanical signature" of industrial tasks more reliably than human kinematics.

Conversely, the IMUs Only model yielded the lowest F1-score (76.13%), struggling particularly with "Hammering" (Recall 50.70%). This suggests that while inertial sensors capture the worker's gross motor movements, the high-frequency vibrations of manual tools are likely aliased or obscured at lower sampling rates. This implies that capturing such work signatures via kinematics would require a higher sampling frequency (Hz) to move beyond the noise floor.

Shop Floor Experiment Results

To capture higher-frequency mechanical transients, the sampling frequency in this experiment increased from 10 Hz to 90 Hz. Thus, the temporal alignment process done on the Shop Floor dataset yielded an initial 491,760 observations (61 minutes) with 58.6% of it labeled as Working and 42.4% as Not Working.

Model adaptation and Feature Engineering

Stage I (Baseline Configuration and Bias Identification): An initial Random Forest model was established using a standard 80/20 random split and optimized via *RandomizedSearchCV*. While this model achieved an F1-score of 0.97, a subsequent feature importance analysis revealed a structural flaw: static weight features accounted for 86% of the decision-making weight. This indicated that the model was functioning as a "digital scale" (identifying that a worker or object was on the platform), rather than recognizing the assembly processes.

Stage II (Feature Sensitivity, Class Balancing, and Temporal Integrity): To transition from presence detection to activity recognition, the raw weight columns were temporarily excluded. The goal was to evaluate the predictive power of the IMU data independently. This analysis yielded two critical architectural adjustments:

- A Signal Magnitude Vector (SMV) was calculated to ensure the model's robustness against sensor mounting variations. By transforming tri-axial data into a rotation-invariant magnitude, the model is prevented from overfitting to specific hardware orientations, mitigating the risk of memorizing axial-specific patterns.
- The standard random split was found to induce data leakage. The high-frequency (90 Hz) sampling rate caused adjacent samples to be highly similar, meaning that random shuffling placed temporally correlated fragments into both training and testing sets, resulting in inflated performance metrics. To rectify this, the protocol transitioned to a 73/27 Chronological Split (*TimeSeriesSplit*).

The 73/27 split was used to maintain a consistent class distribution across the Training and Test subsets, since the Working state was mostly concentrated in the initial portion of the dataset, while Non-working intervals were more prevalent in the latter half. This partitioning strategy ensured a more balanced representation in both the training and validation sets.

Stage III (Feature Aggregation and Temporal Context): Rolling Windows (RW) of one and three seconds were implemented, providing the model with a "short-term memory". This allowed the algorithm to filter transient noise and prioritize the sustained vibration signatures from Working activities. Finally, the raw weight data was replaced with a single 3-second Weight Delta (ΔW) feature. By calculating the change in weight across the platform, this feature emphasizes loading transitions and dynamic events, rather than static mass. The final dataset variables are summarized in Table 4.

Table 4: Final dataset: First three observations

Obs.	1s RW	3s RW	ΔWeight	Target
1	$w_{1,(1:4)}$	$a_{1,(1:4)}$	Δw_1	1_{st} value
2	$w_{2,(1:4)}$	$a_{2,(1:4)}$	Δw_2	2_{nd} value
3	$w_{3,(1:4)}$	$a_{3,(1:4)}$	Δw_3	3_{rd} value

This final model's *RandomizedSearchCV* reached an optimized configuration of 100 estimators, a maximum depth of 8, and a minimum sample split of 5.

Table 5: Final model Classification Report

	Precision	Recall	F1-Score
Not Working	0.71	0.73	0.72
Working	0.81	0.79	0.80
Weighted Average	0.77	0.77	**0.77**

As detailed in Table 5, this configuration achieved a Weighted F1-score of 0.77. Notably, the model demonstrates a sensitivity (Recall) of 0.79 for the Working class. Given the stochastic nature of the shop floor, this result suggests that the feature engineering process successfully filtered out ambient mechanical noise, typical of industrial settings.

Analyzing feature importance analysis, illustrated in Figure 5 identifies the 3-second vibration windows as the primary predictor for the final classifier. This confirms that temporal signal stability offers superior predictive power compared to instantaneous, high-frequency movement. Notably, weight variables were successfully attenuated to secondary support features. This shift transitioned the system from a presence detector to an activity monitor, capable of identifying true production cycles.

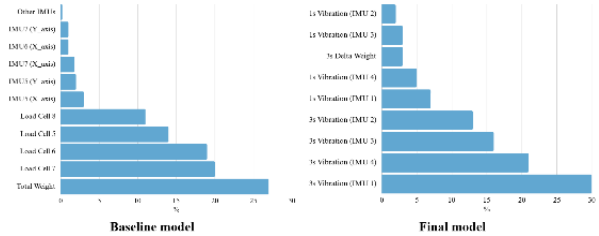


Figure 5: Baseline and Final models' most important features

The final normalized confusion matrix (Figure 6) identifies two primary error types: 12.37% (11,039) false negatives, where active work was classified as Not Working, and 11.15% (9,650) false positives, where NVA tasks (mostly walking on the platform) were misidentified as Working.

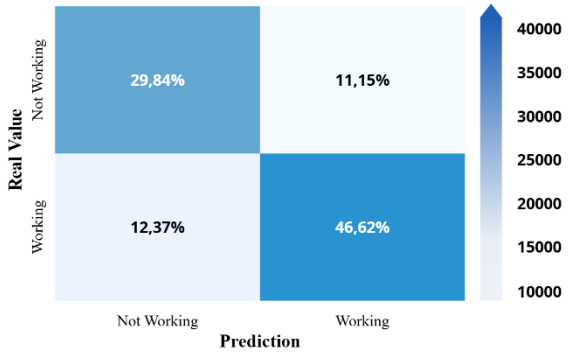


Figure 6: Final model's normalized confusion matrix

Key Differences Between Laboratory and Shop Floor Results

This section discusses the key differences between the laboratory and shop floor experiments, including variations in environmental conditions, data acquisition, and operational constraints that influenced the model performance.

In the laboratory, the Load Cells Only configuration achieved the superior F1-score, by capturing the isolated mechanical signatures of "Hammering" and "Drilling". However, it was later revealed that the model had become a "digital scale", meaning that, while load cells are excellent at detecting high-energy impulses in controlled settings, they can overshadow kinematic data in complex environments if not properly regulated. By replacing raw weight with the ΔW, the bias was mitigated, effectively forcing the model to prioritize dynamic events over static mass.

Another notable difference was observed in the classification of "Walking". In the laboratory, "Walking" was identified with high precision (only 3,05% false positives), but in the shop floor model there were 11,15% False Positives. This error is attributed to the environmental constraints of the active production site, since, unlike the spacious and relatively unobstructed laboratory platform, the shop floor required workers to perform short, irregular steps and weight shifts while reaching for tools or standing still during inspections.

On the other hand, the implementation of 1-second and 3-second temporal windows (Stage III) represented a shift towards a more human-centric monitoring logic. Just as a supervisor does not judge productivity based on a single millisecond of movement, the rolling windows allowed the model to account for "micro-rests" (i.e., the brief pauses when a worker catches their breath or adjusts a tool), which the Laboratory model could not do with its millisecond-scaled observations.

Ultimately, the integrated use of SMV and ΔW successfully enabled activity classification validating the hypothesis. However, the integration revealed a "weighted partnership" where the ΔW contributed only 3% to the model's predictive power. While this confirms the system is a functional activity monitor, it suggests that the weight sensors' primary value may lie in secondary applications.

Conclusions

The proposed approach integrated weight sensor data with IMU data, introduced a Random Forest model, and detailed the complete data flow and feature engineering process from raw data to the final model. By capturing weight shifts and mechanical vibrations, the system can infer internal activities, identifying internal assembly tasks without the need for line-of-sight or invasive wearable sensors on the workers. Since the system captures data directly through the volumetric module platform, it is not affected by the 'black-box' effect caused by occlusions, which commonly occurs in camera-based data capture. In addition, general technical advances include a more accurate distinction between working and idle states, higher data resolution, and customized ground-truth labeling based on exact activity execution times rather than fixed windows.

Building on these technical foundations, this research validates the system's effectiveness within the specific context of volumetric module production (IC) shop floors. During experimental testing in a bathroom pod facility, the methodology achieved a 77% Weighted F1-score, effectively elevating the technology to TRL 5. This successful deployment under real-world operating conditions confirms the system's feasibility as a monitoring solution for relevant complex industrial environments.

While initial laboratory results suggested that load cells alone could achieve a 93.50% F1-score, the shop floor validation revealed that such models are prone to structural bias toward simple presence detection. The implementation of feature engineering, specifically the use of Temporal Rolling Windows, allowed transitioning from a "digital scale" to a true activity recognition system.

Nevertheless, this paper also identifies different limitations such as the need for experiments to make ML models more robust, improvements in hardware to guarantee the integrity of sensors in the instrumented platform, and the difficulty in distinguishing between the high-frequency mechanical transients of manual tools and the erratic movements or weight shifts of workers in confined spaces. Furthermore, the current reliance on fixed platform locations suggests that future iterations should explore configurations that adapt to more complex, mobile, or dynamic shop floor layouts.

Several avenues for future research also arise from the experimental findings. First, the impact of extending the temporal windows to better capture the context needed to distinguish between active work and brief pauses should be studied. In addition to that, the ΔW feature's low contribution to the classification model suggests its utility may lie elsewhere. In that sense, future studies could leverage weight sensor data to detect the placement of heavy components on the platforms, providing a more robust mechanism for tracking production progress by creating an automated validation process of assembly sequences. In practice, a rule-based system could be developed to connect, in real time, the measured weight of heavy components to their corresponding positions on the platform according to the module plans. Thus, the final position of each component, as confirmed by sensor data, could indicate in real time the progress of module completion.

Finally, the study concludes that while human movement on confined IC platforms is inherently random and interwoven with production, it can be mathematically distinguished through the integrated use of IMUs and load cells, coupled with ML models. Ultimately, this system empowers IC managers with empirical data to identify bottlenecks and optimize shop floor workflows, unlocking the productivity potential inherent in industrialized methods.

Acknowledgments

The authors would like to express their gratitude to CIPYCS (Centro Interdisciplinario para la Productividad y Construcción Sustentable) and Axis Desarrollos Constructivos for providing laboratory and shop floor space, respectively. Moreover, the authors would also like to express their gratitude to Enrique Loeser, Claudio Inostroza, and Ignacio Peña for their continuous support. Also, the authors acknowledge the support of ANID Fondecyt Iniciación 11241027.

References

Alsakka, F., Assaf, S., El-Chami, I., and Al-Hussein, M. (2023). Computer vision applications in offsite construction. Automation in Construction, 154:104980.

Altaf, M. S., Bouferguene, A., Liu, H., Al-Hussein, M., and Yu, H. (2018). Integrated production planning and control system for a panelized home prefabrication facility using simulation and rfid. Automation in Construction, 85:369–383.

Barkokebas, B., Martinez, P., Bouferguene, A., Hamzeh, F., and Al-Hussein, M. (2023). Digitalization-based process improvement and decision-making in offsite construction. Automation in Construction, 155:105052.

Chen, X., Wang, Y., Wang, J., Bouferguene, A., and Al-Hussein, M. (2024). Vision-based real-time process monitoring and problem feedback for productivity oriented analysis in off-site construction. Automation in Construction, 162:105389.

Demiralp, G., Guven, G., and Ergen, E. (2012). Analyzing the benefits of rfid technology for cost sharing in construction supply chains: A case study on prefabricated precast components. Automation in Construction, 24:120–129.

Hasan, S. and Sacks, R. (2023). Integrating bim and multiple construction monitoring technologies for acquisition of project status information. Journal of Construction Engineering and Management, 149(7):04023051.

Kauppinen, L., Annunen, P., and Haapasalo, H. (2024). Systematic literature review of themes and definitions of industrialized construction. Smart and Sustainable Built Environment. Ahead-of-print.

Kimmel, W. M., Beauchamp, P. M., Frerking, M. A., Kline, T. R., Vassigh, K. K., Willard, D. E., Johnson, M. A., and Trenkle, T. G. (2020). Technology readiness assessment: Best practices guide. Technical Report NASA/SP-20205003605, National Aeronautics and Space Administration (NASA). NASA Headquarters, Office of the Chief Engineer.

Li, C. Z., Zhong, R. Y., Xue, F., Xu, G., Chen, K., Huang, G. G., and Shen, G. Q. (2017). Integrating rfid and bim technologies for mitigating risks and improving schedule performance of prefabricated house construction. Journal of Cleaner Production, 165:1048–1062.

Martinez, P., Barkokebas, B., Hamzeh, F., Al-Hussein, M., and Ahmad, R. (2021). A vision-based approach for automatic progress tracking of floor paneling in

offsite construction facilities. Automation in Construction, 125:103620.

Melgar, M., Miranda, Y., Al-Hussein, M., and Barkokebas, B. (2025). Incorporation of context-based data to refine machine learning models in industrialized construction. Working Paper / To appear.

Panahi, R., Louis, J., Podder, A., and Swanson, C. (2022). Tracking volumetric units in modular factories for automated progress monitoring using computer vision. In Construction Research Congress 2022, pages 822–829.

Qi, B., Razkenari, M., Costin, A., Kibert, C., and Fu, M. (2021). A systematic review of emerging technologies in industrialized construction. Journal of Building Engineering, 39:102265.

Rashid, K. M. and Louis, J. (2020). Activity identification in modular construction using audio signals and machine learning. Automation in Construction, 119:103361.

Rashid, K. M. and Louis, J. (2021). Automated active and idle time measurement in modular construction factory using inertial measurement unit and deep learning for dynamic simulation input. In Kim, S., Feng, B., Smith, K., Masoud, S., Zheng, Z., Szabo, C., and Loper, M., editors, Proceedings of the 2021 Winter Simulation

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DESIGNING HUMAN–ROBOT COLLABORATION AS A JOINT COGNITIVE SYSTEM: A SOCIOTECHNICAL FRAMEWORK FOR MEANINGFUL WORK

Angelina Kozhevnikova and Daniel M. Hall
Delft University of Technology, Delft, Netherlands

Abstract

Despite decades of research, construction robotics adoption remains limited due to unstructured environments, task complexity, and misaligned automation models adapted from manufacturing. Integrating insights from Joint Cognitive Systems (JCS), Meaningful Work research and Human-Robot Collaboration in construction studies, we propose an alternative, sociotechnical framework for robotic system design structured around three interdependent dimensions: (1) Task and Agent Design, which supports dynamic structuring, agent configuration, and time-based collaboration; (2) Coordination and Communication, which facilitates communication, coordination, shared mental models and situational awareness; (3) Worker Well-Being, which embeds physical and psychological safety, relatedness, competence and autonomy considerations.

Introduction

Research in construction robotics dates back to the 1980s and earlier, yet adoption still remains limited. This may suggest a mismatch between dominant automation models, largely derived from manufacturing, and key characteristics of construction work, including unstructured environments, high variability in tasks and materials, and reliance on situational judgment and embodied skill (Ardiny, Witwicki & Mondada 2015; Cai et al. 2019; Liu & Jebelli 2022; Saidi, Bock & Georgoulas 2016). As a result, models developed for standardised and predictable settings are difficult to transfer to construction contexts, that are difficult to formalise and therefore automate with the dominant model (Liu et al. 2024; Ardiny, Witwicki & Mondada 2015; Cai et al. 2019; Saidi, Bock & Georgoulas 2016).At the same time, labour shortages, safety risks, and an ageing workforce increase pressure for automation while complicating its implementation (Ranasinghe et al. 2023; Schwatka, Butler & Rosecrance 2012; Yamani Bin Yahya et al. n.d.).

Human–Robot Collaboration (HRC) has recently gained attention as a way to enable robotic systems to operate in dynamic conditions through maintaining human involvement (Ma, Mao & Liu 2022; Onososen, Musonda & Ramabodu 2022; Zhang et al. 2023). However, many current frameworks still reproduce manufacturing-based models, such as levels of automation (LoA) taxonomies and function allocation schemes derived from Fitts' List (Yang et al. 2024; Liang et al. 2021). This approach shows limited adequacy for the adoption in the industry (Du et al. 2025; Hancock & Scallen 1996) and has been linked to broader negative implications on societal well-being (Constantinides & Quercia 2025; Hancock 2014; Kaur, Gautam & Verma 2025). In other words, our current technology-centric mindset restricts the full potential of HRC and risks repeating earlier failures of construction robotics.

In this paper, we offer a way to reframe human-robot collaboration in the context of construction. We suggest the need to define HRC more broadly as "*humans working jointly with robots or together in a specific construction task, and each role must be capable of making contributions to the overall system.*" (Zhang et al. 2023) and shift towards a sociotechnical framework that acknowledges the complexity of the process and supports safe, meaningful work integrated into the realities of the industry.

To do this, our paper proceeds as follows. First, the background section introduces two key theoretical foundations: the Impact of Automation on Work Design and the Joint Cognitive System as a design unit. Second, the methods section introduces Critical Interpretive Synthesis, a methodology used. The results section presents this framework, structured around three interrelated HRC system design dimensions: task and agent design, coordination and communication, and worker well-being, and identifies key considerations for each based on Joint Cognitive Systems theory, meaningful work literature, and current HRC practices. Finally, the discussion and conclusion outline the framework, and discuss how it can be further developed and translated into practical tools for research and application.

Background

Impact of automation on work design

Understanding how automation reshapes work is key to designing human–robot collaboration. Each new technology redefines work processes yet many approaches still rest on the *Substitution Myth*, the belief that machines can take over human tasks without broader consequences for the system (Hollnagel & Woods 2005).

This view treats automation as a neutral replacement rather than a transformation of work. It appears mainly in two forms: the *Left-Over Principle* and the *Compensatory Principle*.

Left-Over Principle. The focus is on automating as much as possible, leaving humans with tasks that are either too trivial or too complex, rare, and irregular, often leading to cognitive overload (Woods & Hollnagel 2006). Since it's mainly developed without users in mind, it can also compromise safety, as workers develop workarounds that introduce new physical or organisational risks (Goštautaitė et al. 2024).

Compensatory Principle. This principle is based on juxtaposition; it is often referred to as *"Men-Are-Better-At, Machines-Are-Better-At"* or *MABA–MABA*, also known as *Fitts' List* (Fitts 1951; Woods & Hollnagel 2006). It compares human and machine capabilities across fixed dimensions a-priori, assuming minimal variability and treating both as predictable systems. This approach also has safety implications. It often reduces human autonomy (Sadeghian & Hassenzahl 2022) and blurs responsibility, creating accountability gaps (Sadeghian, Uhde & Hassenzahl 2024). It also typically confines workers to supervisory or advisory roles with limited control, leading to alienation and difficulties in justifying the value of their work (Staaby, Hansen & Grønli 2021).

The question of meaningful human work design has become increasingly important with the growing automation of work. Organisational research on meaningfulness has developed two main perspectives: realisation and justification (Lepisto & Pratt 2017; Goštautaitė et al. 2024).

Realisation Perspective. The main focus is on fulfilling psychological needs such as autonomy, competence, and belonging. Meaningful work arises when individuals can express themselves through job design or job crafting, as described in Self-Determination Theory (Ryan & Deci 2000; Peters & Calvo 2023; Goštautaitė et al. 2024).

Justification Perspective. Here, meaningfulness is emerging through sense-making and reaffirming the significance of human contribution. It suggests that meaningfulness arises when people confront uncertainty about the value of their work, prompting sense-making processes shaped by leadership, organizational practices, and how performance is evaluated (Goštautaitė et al. 2024; Lepisto 2022; Beer, Micheli & Besharov 2022; Grant 2012). It draws on Honneth's recognition theory (Honneth 1996), where recognition at work means acknowledging individuals as skilled contributors in cooperative activity, fostering self-confidence, self-respect, and self-esteem (Hirvonen & Breen 2020; Mardosas 2024).

Complementarity Principle. These psychological perspectives resonate with a third view on automation called the *Complementarity Principle*, introduced in the 1990s (Grote et al. 1995; Wafler et al. 2003). This principle examines work as a dynamic system that develops over time and aims to enhance human performance and learning instead of replacing it. Often described as function congruence or function matching (Hollnagel 2018), it emphasises how work routines adapt to situational changes and seeks to maintain system performance across varying conditions (Woods & Hollnagel 2006). This principle becomes very useful for safety-critical industries like construction, which heavily rely on craft-based practice with workers at the center of decision-making.

Joint Cognitive System as a Design Unit

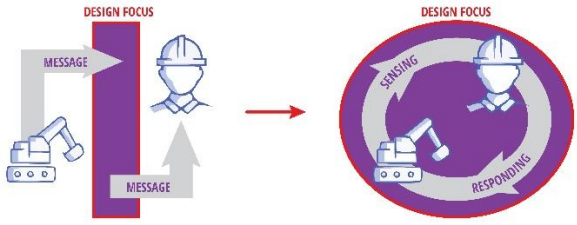


Figure 1: A figure contrasts the classical human–robot interaction view with the co-agency view across system boundary, temporal structure of action, and design focus

Joint Cognitive Systems (JCS) build on the complementarity approach and adopt a sociotechnical perspective, treating the practitioner as an integral part of the system. Developed in safety-critical domains such as aviation, nuclear power, and healthcare, JCS address the limitations of information-processing models by focusing on how systems maintain performance under changing conditions (Woods & Hollnagel 2006; Hollnagel & Woods 2005).

Classical Human-Robot Interaction (HRI) View. In the classical human–robot interaction view, humans and machines are treated as separate components connected through an interface. Interaction is defined as the exchange of inputs and outputs, as illustrated on the left side of Figure 1, where communication occurs through a bounded interface layer. Cognition is modeled as a response to discrete stimuli, and action is understood as a sequence of stepwise reactions. The analytical focus remains structural, describing system components and their interactions rather than how the system performs as a whole.

Limitations of the Classical View. This representation abstracts away key characteristics of real-world work. It treats action as discrete rather than continuous, and behavior as reactive rather than anticipatory. Context is reduced to inputs, rather than directly shaping action, and cognition is effectively placed outside of its situated environment. It also assumes isolated actors, while in practice human activity is embedded in broader systems of coordination. As a result, the design focus remains on interfaces and components rather than system-level performance.

Co-Agency View (JCS). JCS shifts the focus from interaction to co-agency, understood as joint agency within a single system. As shown on the right side of Figure 1, humans and machines are physically separate, but physical separateness does not imply functional

separateness. The unit of analysis becomes the joint system, engaged in continuous cycles of sensing and responding. The focus moves to what the system achieves as a whole, and how it adapts under changing conditions. Behavior is continuous and context-dependent, and agency is distributed across human and machine.

Critical interpretive synthesis (CIS)

This study employs Critical Interpretive Synthesis (CIS), a method first developed in medical research (Dixon-Woods et al. 2006) and later adapted for the built environment and construction automation (Bucher et al. 2024; Pan et al. 2018). CIS combines review and interpretation, using induction and critical reflection to identify conceptual patterns, making it suitable for developing theoretical frameworks that connect diverse perspectives.

After preliminary scoping, we conducted three targeted Scopus searches: (1) human–robot collaboration (HRC) in construction (272 records), (2) Joint Cognitive Systems (JCS) (144 records), and (3) meaningful work (MW) in automation contexts (45 records). After screening, studies without a human-in-the-loop or not related to construction were excluded from the HRC set, overly technical papers from the JCS set, and superficial human factors studies from the MW set. This resulted in 74 HRC, 39 JCS, and 21 MW papers, further reduced through full-text review to 38 HRC, 22 JCS, and 14 MW studies. Two foundational JCS books were also included (Hollnagel & Woods 2005; Woods & Hollnagel 2006).

The synthesis proceeded iteratively in three steps. First, the HRC literature was analysed inductively to identify recurring patterns in how collaboration is studied in construction. The review revealed a fragmented, largely technology-centric field dominated by experimental studies and demonstrators, with limited attention to real-world adoption and an underdeveloped account of the human role. This lack of implementation-oriented research motivated the search for a stronger theoretical foundation.

Second, these limitations led to the selection of Joint Cognitive Systems (JCS) as the primary theoretical foundation, as it provides a systems perspective grounded in safety-critical domains with established real-world use.

Third, meaningful work research was incorporated to extend this foundation with a more direct account of the human dimension, particularly longer-term work design and its impact on workers.

The framework was then defined by adapting JCS to the context of human–robot collaboration in construction, with meaningful work as a complementary perspective. The construction HRC literature was used to ground this adaptation, guiding the selection and alignment of concepts based on current developments, observed gaps, and practical constraints in the field.

Dimensions of human-robot collaboration (HRC) system design

This framework is a conceptual, high-level structure for human–robot collaboration system design in construction, intended for system designers, engineers, and researchers working at the level of system configuration and integration. It structures how work, coordination, and human factors are addressed together when designing joint human–robot work systems at the level of sociotechnical system architecture.

The framework is organised around the core components of human–robot collaboration: humans, robots, and their collaboration. Task and Agent Design focuses on work, addressing task structure and role configuration. Coordination and Communication focuses on the robotic system and its role in enabling joint activity, addressing sensing, interaction, and control. Worker Well-Being focuses on the human, addressing longer-term work conditions and their impact.

The dimensions are tightly interconnected and operate as system-level design lenses. Figure 2 shows their cyclical relationship: they are not ontologically separate components, but purpose-driven perspectives used to guide system design. Coordination and Communication is positioned centrally because the robotic system acts as the mediating and coordinating layer through which joint activity is organised, linking work processes and worker conditions.

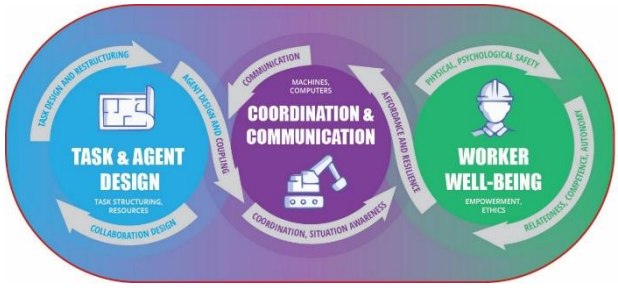


Figure 2: *Three main interrelated human-robot collaboration (HRC) system design dimensions.*

Task and agent design

Task design. To design a system, its boundaries must first be defined by distinguishing between the system and its environment. Since control, both in goals and variability, occurs at multiple levels, several joint cognitive systems can coexist (Hollnagel & Woods 2005). In construction, task design concerns what needs to be built and how it is divided into levels or subtasks based on granularity and scope. For example, scaffolding construction can be broken down into subtasks, steps, activities, and fine-grained task primitives (Pan & Yu n.d.). Integrating these task hierarchies into management workflows and related cognitive systems is essential. It is also important to consider not only direct but also indirect and off-task work (Van Diggelen & Johnson 2019), which often increases with automation, and to allow for dynamic adaptation when conditions change.

Dynamic task restructuring. Adaptation in a successful HRC system goes beyond allocating tasks based on human and robotic skills (Amtsberg et al. 2021). It also involves restructuring tasks and goals. Shifting objectives, refining standards, and enabling workers to shape tasks are central to meaningful work. This approach fosters ownership, responsibility, and helps prevent accountability gaps and alienation (Parmer 2024; Sadeghian, Uhde & Hassenzahl 2024; Staaby, Hansen & Grønli 2021; Mardosas 2024; Bielskis 2024; Smids, Nyholm & Berkers 2020; Celentano 2024). JCS studies highlight the importance of directability, the ability to assess and modify the actions of others in a joint activity as conditions evolve (Klein et al. 2004). In construction, this remains underdeveloped. One study addressed human control through a digital twin that visualises both the work object and the robot's motion plan (Amtsberg et al. 2021). The craftsman in the study can take over a task at any time in case of material inconsistencies, path collisions, or other unforeseen issues. Bidirectional communication and real-time visualisation support informed decision-making for timber handling.

Time-based collaboration design. Human–robot collaboration is inherently time-based, and construction management traditionally depends on distinct phases and scheduling. HRC design should acknowledge this temporality and the continuously shifting relationships between human and robotic actors. Instead of fixed roles, collaboration can be viewed as a sequence of shifting team patterns depending on task, state, and context. Common team patterns between humans and intelligent systems (Van Diggelen & Johnson 2019) include synchronous joint work, synchronous asymmetric work such as monitoring or teleoperation, and mixed modes that combine synchronous pre-planning with asynchronous execution. These patterns involve transitions and handoffs as activities progress and are interdependent with neighbouring ones. In practice, even highly autonomous modes remain embedded within broader human oversight and system interdependencies. The highly autonomous pattern is one among several interconnected patterns, implying a human backup role that requires continuous monitoring and readiness to intervene. In construction, this aspect remains underdeveloped as well. Most HRC studies prototype only one activity type, usually synchronous, focusing on physical interaction or teleoperation.

Agent design. In order to stay purpose-driven and goal-oriented, robotic agents should be seen as functionally integrated extensions of human capabilities rather than separate tools (Hollnagel & Woods 2005). From a JCS perspective, all work is cognitive, but for practical design, it is helpful to view the robotic agent as combining a machine and a computer, which are tightly coupled. This corresponds to the dual nature of construction work, where physical activity involves manipulating tangible materials and cognitive activity involves perception and information processing. In timber prefabrication, human and robotic skills can be grouped into manipulation and mobility, with humans contributing dexterity and robots providing precision and strength (Yang et al. 2024). Cognitive skills include situation awareness, decision-making, motion planning, and adaptation. In masonry, robots can take over repetitive lifting and bricklaying, allowing workers to focus on detailed finishing tasks (Baek et al. 2024). These examples show that agent selection must always be grounded in the specific capabilities required by each construction activity and seen as evolving over time.

Coordination and communication

Communication channels. An important part of human–robot collaborative co-agency is communication. Instead of viewing the interface as a piece of technology, the system can be understood through coordination and the flow of information between agents as the basis for effective collaboration. Human-to-robot communication can rely on gesture, voice, gaze, bio-signal, and vision-based commands or manual controllers, including haptic master devices, AR, and multiple sensors such as EEG (Liu & Jebelli 2022; Zhang et al. 2023). For other communication channels, the technology should be chosen or designed accordingly. For example, in on-site construction scenarios, robot-to-human communication can be embedded in wearables such as wrist-mounted displays or helmet indicators, and robot-to-environment communication can use simple machine-based indicators (Frank et al. 2019). These channels together provide essential infrastructure for coordination, but do not replace coordination design itself.

Multi-modal coordination. Support for effective coordination is one of the generic requirements for designing a joint cognitive HRC system. It involves synchronising agents in changing conditions, creating and updating a shared mental model, balancing goals, and ensuring observability (Woods & Hollnagel 2006). In practice, coordination should provide not only feedback, which is common, but also feedforward and feednow. In the Extended Control Model (ECOM), these correspond to control layers such as tracking, regulating, monitoring, and targeting (Hollnagel & Woods 2005). Integration with Building Information Modelling (BIM), the Internet of Things (IoT), digital twins, and other advanced technologies provides an additional coordination and data-processing layer within the joint cognitive system. (Amtsberg et al. 2021; Zhang et al. 2023).

Effective coupling. In HRC new qualities emerge through human-robot coupling or their interrelations (Hollnagel & Woods 2005). Human-robot coupling is crucial since inadequate cognitive coupling results in poor user acceptance (Woods 1985). Effective coupling and systems that remain viable over time require involving construction workers early in system design, well before implementation (Fraboni, Brendel & Pietrantoni 2023; Staaby, Hansen & Grønli 2021; Celentano 2024). This can be achieved through prototyped VR-based environments that provide low-cost, safe tools integrating simulation, ergonomic analysis, and embodied feedback to refine interaction, coordination, and workflow design

(Onososen & Dzuwa n.d.; Saffert et al. 2024; Onososen et al. 2023). These tools also act as methodological environments for exploring team dynamics with continuous human feedback.

Shared mental models and situation awareness. Effective coupling can be further improved through shared mental models, system transparency, and mutual observability. System transparency enhances explainability and helps prevent accountability gaps (Sadeghian, Uhde & Hassenzahl 2024). Maintaining shared mental models and situational awareness requires integrating multiple communication and processing threads. For dynamic environments such as on-site construction, integrating multi-sensory channels has shown promising results for human–robot collaboration (Zhang et al. 2023). This reflects real construction situations where workers routinely use multiple modes of communication, such as eye contact, gestural confirmation, and verbal exchange, especially before entering potentially hazardous areas near equipment (Frank et al. 2019).

Affordance and resilience. Beyond coordination and miscoordination, affordance and resilience are two main patterns that emerge at the intersection of people, technology, and work (Woods & Hollnagel 2006), and therefore should be explicitly addressed in HRC system design. In a study from Switzerland, construction and mining workers expressed similar views on what future automation should provide. They emphasised adaptation to changing situations and the facilitation of collaboration in all work contexts as fundamental requirements (Frank et al. 2019).

Worker well-being

Physical and psychological safety. Worker well-being is treated not as an external ethical constraint but as a constitutive condition and affordance for joint system performance. This includes strategic choices such as using teleoperation in high-risk scenarios (Zhang et al. 2023) or adjusting robot speed and working parts based on a worker's cognitive load (Liu & Jebelli 2022). Good practices include demonstrating safety and reliability, involving workers in hazard analysis, and applying explainable mechanisms that clarify system goals and reasoning (Fraboni, Brendel & Pietrantoni 2023; Sadeghian, Uhde & Hassenzahl 2024). Avoiding unrealistic expectations about communication or capability is equally important, as unmet expectations can harm team dynamics (Sadeghian & Hassenzahl 2022).

Relatedness and team composition. The question of socio-cognitive misalignment also arises when designing human-robot collaboration in relation to the need for human connection. Robots can successfully cooperate to achieve shared goals, yet they are not equivalent to human colleagues due to differences in social interaction, informal communication, and behaviour in safety-critical situations (Nyholm & Smids 2020). Social relationships remain fundamental for meaningful work and self-development (Sadeghian & Hassenzahl 2022; Staaby, Hansen & Grønli 2021; Smids, Nyholm & Berkers 2020; Nikolova, Cnossen & Nikolaev 2024). Interviews with timber construction workers consistently highlighted three key motivators: the joy of building, task variety, and teamwork (Yang et al. 2024). In the HRC system design, social connection must be explicitly supported.

Competence and expertise creation. A further human-centered goal in HRC system design is continuous learning for both human and robotic agents. This includes the refinement of standards, open-ended rather than predefined automation loops, and the active development of skills and self-growth (Parmer 2024; Sadeghian & Hassenzahl 2022; Staaby, Hansen & Grønli 2021; Smids, Nyholm & Berkers 2020; Celentano 2024; Graham & Jones n.d.). The system itself can learn and adapt through methods such as Robot Learning by Demonstration, Imitation Learning, or programming by demonstration, which enable robots to acquire complex construction skills (Liang et al. 2021). At the same time, embedded training modules and interactive systems can help workers strengthen their expertise and develop new interaction skills (Onososen & Dzuwa n.d.; Saffert et al. 2024; Shayesteh et al. 2023). This fluid, adaptive approach to expertise creation within co-agency can address workforce challenges in the construction industry.

Autonomy and job crafting. Continuous learning and adaptation also strengthen personal autonomy and its recognition in work design. Limiting human roles to supervision or advisory functions can fragment ownership, reinforce alienation (Staaby, Hansen & Grønli 2021), and create authority–responsibility double binds, where workers are accountable but unable to exercise real autonomy (Sadeghian, Uhde & Hassenzahl 2024; Woods & Hollnagel 2006). HRC systems should enable agency, feedback, and flexibility, fostering a sense of ownership and purpose (Parmer 2024; Sadeghian, Uhde & Hassenzahl 2024; Sadeghian & Hassenzahl 2022; Staaby, Hansen & Grønli 2021; Mardosas 2024; Bielskis 2024; Smids, Nyholm & Berkers 2020; Nikolova, Cnossen & Nikolaev 2024; Graham & Jones n.d.). This aligns with ethical design principles that emphasise privacy, autonomy, dignity, reliability, inclusion, and the right to choose one's own way of working (Kaasinen et al. 2022). HRC systems need to maintain open-ended structures that allow roles and tasks to be shaped dynamically. Job crafting should be treated as an active process of person–job fit, with explicit goals to support meaningful work and to avoid the Psychologist's Fallacy, which is the mistake of confusing the designer's or observer's vision of work with the lived experience of those who actually perform it (Staaby, Hansen & Grønli 2021; Woods & Hollnagel 2006).

Discussion

Prior human-robot collaboration (HRC) frameworks in construction (Yang et al. 2024; Liang et al. 2021) build on concepts from manufacturing and industrial automation, particularly Levels of Automation, function allocation, and Human–Robot Interaction (HRI) taxonomies, providing a structured basis for describing roles,

interactions, and system configurations under relatively stable, fixed, a priori conditions. This paper reframes HRC as a sociotechnical and temporal process in which human and robotic capabilities, fragmented organisational structures, shifting teams, and evolving site conditions co-adapt. It considers how responsibility shifts, how expertise is negotiated, and how coordination can be maintained under changing conditions.

It does so by redefining the unit of design from interfaces or separate human and robot agents to a joint cognitive system. This shifts the focus from task allocation to coordination of work, from static levels of autonomy to continuous co-adaptation, from isolated interaction variables to system-level performance, from describing and classifying parts or configurations to designing purpose-driven systems oriented to performance, adoption, and meaningful work.

From this perspective, construction robotics can move from being "robot-ready", organised around machine capabilities, to "system-ready", where task structures, communication flows, and organisational roles support joint performance.

Future research should further develop this framework into more explicit design-oriented structures that can guide system design. The current considerations define key domains and relationships but remain conceptual. They need to be translated into forms that inform task decomposition, role configuration, coordination, and system design. Empirical work in real construction settings is required to examine how these considerations support implementation, adaptation, and adoption in practice.

Conclusions

This paper synthesises insights from Human–Robot Collaboration (HRC) in construction, Joint Cognitive Systems (JCS), and Meaningful Work (MW) research to propose an integrated framework for HRC system design. The framework is structured around three interdependent dimensions: (1) Task and Agent Design, which addresses multilevel task design, dynamic task restructuring, agent design, and time-based collaboration design; (2) Coordination and Communication, which focuses on communication channels, multi-modal coordination, effective coupling, shared mental models, situational awareness, affordance, and resilience; and (3) Worker Well-Being, which emphasises physical and psychological safety, relatedness and team composition, competence and expertise creation, autonomy, and job crafting.

Grounded in Joint Cognitive Systems (JCS) theoretical framework with added perspective from Meaningful work (MW) scholarship, the proposed conceptual framework reframes construction Human-Robot Collaboration (HRC) as a joint cognitive system in which performance, safety, learning, and accountability emerge from the quality of coupling across people, robots, and work over time. The paper shifts the design problem from "who does which task" to "how joint work is structured, coordinated, and sustained without degrading human work." Future research will extend the framework into HRC system design and engineering tools for construction robotics projects.

Acknowledgments

Funding for this research is provided by the Swiss National Center for Competence (NCCR) Digital Fabrication in Architecture.

References

Ajoudani A, Zanchettin AM, Ivaldi S, Albu-Schäffer A, Kosuge K & Khatib O (2018) Progress and prospects of the human–robot collaboration, Autonomous Robots, vol. 42, no. 5, pp. 957–975

Amtsberg F, Yang X, Skoury L, Wagner HJ & Menges A (2021) iHRC: An AR-Based Interface for Intuitive, Interactive and Coordinated Task Sharing Between Humans and Robots in Building Construction, in Proceedings of the International Symposium on Automation and Robotics in Construction, International Association for Automation and Robotics in Construction (IAARC), pp. 25–32

Ardiny H, Witwicki S & Mondada F (2015) Construction automation with autonomous mobile robots: A review, in International Conference on Robotics and Mechatronics, ICROM 2015, Institute of Electrical and Electronics Engineers Inc., pp. 418–424

Baek F, Kim D, Lee G, Choi B & Lee SH (2024) Effect of human emotional responses on human-robot team pty in construction, Automation in Construction, vol. 166

Beer H, Micheli P & Besharov M (2022) Meaning, Mission, and Measurement: How Organizational Performance Measurement Shapes Perceptions of Work as Worthy, Academy of Management Journal, vol. 65, no. 6, pp. 1923–1953

Bielskis A (2024) Theorising meaningful non-alienated labour, Critique (United Kingdom), vol. 52, no. 1, pp. 41–57

Bucher DF, Hunhevicz JJ, Soman RK, Pauwels P & Hall DM (2024) From BIM to Web3: A critical interpretive synthesis of present and emerging data management approaches in construction informatics, Advanced Engineering Informatics, vol. 62, p. 102884

Cai S, Ma Z, Skibniewski MJ & Bao S (2019) Construction automation and robotics for high-rise buildings over the past decades: A comprehensive review, Advanced Engineering Informatics, vol. 42

Celentano D (2024) Labor automation for fair cooperation: Why and how machines should provide meaningful work for all, Journal of Social Philosophy, vol. 55, no. 1, pp. 25–43

Constantinides M & Quercia D (2025) AI, Jobs, and the Automation Trap: Where Is HCI?, http://arxiv.org/abs/2501.18948

Van Diggelen J & Johnson M (2019) Team design patterns, in HAI 2019 - Proceedings of the 7th International Conference on Human-Agent Interaction, Association for Computing Machinery, Inc, pp. 118–126

Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, Hsu R, Katbamna S, Olsen R, Smith L, Riley R & Sutton AJ (2006) Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups, BMC Medical Research Methodology, vol. 6, no. 1, p. 35

Du L, Li Y, Wang X, Wu Y, Li M, Wen T, Wei Q, Liu S, Ji M, Li Y & You X (2025) The Impact of Automation Systems on Commercial Aircraft: A Human-Machine Interaction Perspective,

Fitts PM (Ed.) (1951) Human engineering for an effective air-navigation and traffic-control system, Washington, D.C., accessed 30 November 2025.

Fraboni F, Brendel H & Pietrantoni L (2023) Evaluating Organizational Guidelines for Enhancing Psychological Well-Being, Safety, and Performance in Technology Integration, Sustainability (Switzerland), vol. 15, no. 10

Frank M, Ruvald R, Johansson C, Larsson T & Larsson A (2019) Towards autonomous construction equipment-supporting on-site collaboration between automatons and humans, International Journal of Product Development, vol. 23, no. 4, pp. 292–308

Goštautaitė B, Liubertė I, Parker SK & Bučiūnienė I (2024) CAN YOU OUTSMART THE ROBOT? AN UNEXPECTED PATH TO WORK MEANINGFULNESS, Academy of Management Discoveries, vol. 10, no. 3, pp. 416–437

Graham CM & Jones N Information Systems and the Problem of Work Information Systems and the Problem of Work: Protocol for a Systematic Review

Grant AM (2012) Leading with Meaning: Beneficiary Contact, Prosocial Impact, and the Performance Effects of Transformational Leadership, Academy of Management Journal, vol. 55, no. 2, pp. 458–476

Grote G, Weik S, Wäfler T & Zölch M (1995) Complementary allocation of functions in automated work systems, in pp. 989–994

Hancock PA (2014) Automation: How much is too much?, Ergonomics, vol. 57, no. 3, pp. 449–454

Hancock PA & Scallen SF (1996) The-future-of-function-allocation, Ergonomics in Design,, pp. 24–29

Hirvonen O & Breen K (2020) Recognitive arguments for workplace democracy, Constellations, vol. 27, no. 4, pp. 716–731

Hollnagel E (2018) From Function Allocation to Function Congruence, in Coping with Computers in the Cockpit, Routledge, pp. 29–54

Hollnagel E & Woods DD (2005) Joint Cognitive Systems, CRC Press

Honneth Axel (1996) The struggle for recognition : the moral grammar of social conflicts, MIT Press

Kaasinen E, Anttila AH, Heikkilä P, Laarni J, Koskinen H & Väätänen A (2022) Smooth and Resilient Human–Machine Teamwork as an Industry 5.0 Design Challenge, Sustainability (Switzerland), vol. 14, no. 5

Kaur M, Gautam S & Verma R (2025) The Social Implications of Self-Serving Technology, in Navigating the Self-Service Revolution With Smart Machines, IGI Global, pp. 133–154

Klein G, Woods DD, Bradshaw JM, Hoffman RR & Feltovich PJ (2004) Ten Challenges for Making Automation a 'Team Player' in Joint Human-Agent Activity, IEEE Intelligent Systems, vol. 19, no. 06, pp. 91–95

Lepisto DA (2022) Ritual Work and the Formation of a Shared Sense of Meaningfulness, Academy of Management Journal, vol. 65, no. 4, pp. 1327–1352

Lepisto DA & Pratt MG (2017) Meaningful work as realization and justification, Organizational Psychology Review, vol. 7, no. 2, pp. 99–121

Liang C-J, Wang X, Kamat VR & Menassa CC (2021) Human–Robot Collaboration in Construction: Classification and Research Trends, Journal of Construction Engineering and Management, vol. 147, no. 10

Liu Y, A.H A, Haron NA, N.A B & Wang H (2024) Robotics in the Construction Sector: Trends, Advances, and Challenges, Journal of Intelligent and Robotic Systems: Theory and Applications, vol. 110, no. 2

Liu Y & Jebelli H (2022) Worker-Aware Robotic Motion Planner in Construction for Improved Psychological Well- Being during Worker-Robot Interaction, in Construction Research Congress 2022: Computer Applications, Automation, and Data Analytics - Selected Papers from Construction Research Congress 2022, American Society of Civil Engineers (ASCE), pp. 205–214

Ma X, Mao C & Liu G (2022) Can robots replace human beings? —Assessment on the developmental potential of construction robot, Journal of Building Engineering, vol. 56

Mardosas E (2024) Automation and Workplace Democracy: Autonomy, Recognition, and Meaningful Work, Filosofija, Sociologija, vol. 35, no. 3, pp. 348–355

Nikolova M, Cnossen F & Nikolaev B (2024) Robots, meaning, and self-determination, Research Policy, vol. 53, no. 5

Nyholm S & Smids J (2020) Can a Robot Be a Good Colleague?, Science and Engineering Ethics, vol. 26, no. 4, pp. 2169–2188

Onososen A & Dzuwa C Enhancing Human-Robot Teaming in Construction through the Integration of Virtual Reality-Based Training in Human-Robot Collaboration,

Onososen A, Musonda I, Ramabodu M & Dzuwa C (2023) Safety and Training Implications of Human-Drone Interaction in Industrialised Construction Sites, in Lecture Notes in Civil Engineering, Springer Science and Business Media Deutschland GmbH, pp. 281–295

Onososen AO, Musonda I & Ramabodu M (2022) Construction Robotics and Human–Robot Teams Research Methods, Buildings, vol. 12, no. 8

Pan M, Linner T, Pan W, Cheng H & Bock T (2018) A framework of indicators for assessing construction automation and robotics in the sustainability context, Journal of Cleaner Production, vol. 182, pp. 82–95

Pan Z & Yu Y Learning multi-granularity task primitives from construction videos for human-robot collaboration,

Parmer WJ (2024) Meaningful Work and Achievement in Increasingly Automated Workplaces, Journal of Ethics, vol. 28, no. 3, pp. 527–551

Peters D & Calvo RA (2023) Self-Determination Theory and Technology Design, in The Oxford Handbook of Self-Determination Theory, Oxford University Press, pp. 978–999

Ranasinghe U, Tang LM, Harris C, Li W, Montayre J, de Almeida Neto A & Antoniu M (2023) A systematic review on workplace health and safety of ageing construction workers, Safety Science, vol. 167

Ryan RM & Deci EL (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being., American Psychologist, vol. 55, no. 1, pp. 68–78

Sadeghian S & Hassenzahl M (2022) The 'Artificial' Colleague: Evaluation ofWork Satisfaction in Collaboration with Non-human Coworkers, in International Conference on Intelligent User Interfaces, Proceedings IUI, Association for Computing Machinery, pp. 27–35

Sadeghian S, Uhde A & Hassenzahl M (2024) The Soul of Work: Evaluation of Job Meaningfulness and Accountability in Human-AI Collaboration, Proceedings of the ACM on Human-Computer Interaction, vol. 8, no. 1

Saffert AS, Schmailzl M, Spitzhirn M & Linner T (2024) A Cyber-Physical Toolbox for Teaching Digital Construction – Technical Configuration, Learning Tactics and Hands-On Testing and Evaluation in Dedicated Courses, in Lecture Notes in Networks and Systems, Springer Science and Business Media Deutschland GmbH, pp. 83–91

Saidi KS, Bock T & Georgoulas C (2016) Robotics in Construction, in Springer Handbooks, Springer Science and Business Media Deutschland GmbH, pp. 1493–1520

Schwatka N V., Butler LM & Rosecrance JR (2012) An aging workforce and injury in the construction industry, Epidemiologic Reviews, vol. 34, no. 1, pp. 156–167

Shayesteh S, Ojha A, Liu Y & Jebelli H (2023) Human-robot teaming in construction: Evaluative safety training through the integration of immersive technologies and wearable physiological sensing, Safety Science, vol. 159

Smids J, Nyholm S & Berkers H (2020) Robots in the Workplace: a Threat to—or Opportunity for—Meaningful Work?, Philosophy and Technology, vol. 33, no. 3, pp. 503–522

Staaby A, Hansen K & Grønli TM (2021) Automation of routine work: A case study of employees' experiences of work meaningfulness, in Proceedings of the Annual Hawaii International Conference on System Sciences, IEEE Computer Society, pp. 156–165

Wafler T, Grote G, Windischer A & Ryser C (2003) Kompass, in pp. 477–502

Woods DD (1985) Cognitive Technologies: The Design of Joint Human-Machine Cognitive Systems Human-Machine Cognitive Systems,

Woods DD & Hollnagel E (2006) Joint cognitive systems: Patterns in cognitive systems engineering,

Yamani Bin Yahya M, Hui Yin L, Binti Md Yassin A, Omar R, Orbintang anak Robin R & Kasim N The Challenges of the Implementation of Construction Robotics Technologies in the Construction, https://doi.org/10.1051/matecconf/2019

Yang X, Amtsberg F, Sedlmair M & Menges A (2024) Challenges and potential for human–robot collaboration in timber prefabrication, Automation in Construction, vol. 160

Zhang M, Xu R, Wu H, Pan J & Luo X (2023) Human–robot collaboration for on-site construction, Automation in Construction, vol. 150

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A THREE-LAYER TAXONOMY AND PREDICTIVE DIGITAL TWIN FOR ROLE-AWARE HUMAN–ROBOT COLLABORATION IN 3D CONCRETE PRINTING

Ibrahim Chahbaz[1], Reina Chmayssany[1], Jiany Samara[1], Noha Ramadan[1], and Hiam Khoury[1]
[1]American University of Beirut, Beirut, Lebanon

Abstract

3D concrete printing (3DCP) has reached early industrial deployment, yet current standards exclude environmental, health, and safety aspects of robotic operation, and existing collaborative-robot frameworks do not address process-coupled hazards or multi-role construction environments. This paper proposes a three-layer Human-Robot-Site taxonomy and a site-wide predictive digital twin with five-second update cadence and short-horizon forecasting. The framework is operationalized through seven role-aware interaction patterns aligned with ISO 10218. A scenario-based evaluation of the Human-layer physiological pipeline using a proxy dataset demonstrates the feasibility of stress classification and calibration. The work provides a reference architecture for human–robot–site integration in 3DCP.

Introduction

3D concrete printing (3DCP) has crossed an inflection point. Recently, Harcourt Technologies completed the first social-housing project compliant with ISO/ASTM 52939:2023, the new qualification standard for additive construction, using a gantry printer to deliver three terraced units in 132 working days; a 35% reduction relative to conventional methods (COBOD International, 2024). The platform is no longer experimental.

What has not crossed that inflection point is the safety framework around it. For instance, Section 1 of ISO/ASTM 52939:2023 is unambiguous: the standard *"does not cover environmental, health* and *safety aspects that apply to printing facility setup, material handling, operating of robotic equipment, and packing of equipment and/or elements for shipping"* (ISO/ASTM, 2023). The standard governing print quality says nothing about how to keep workers safe alongside the machine that produces it. The companion standards available to fill this gap (ISO 10218-1:2025 and ISO 10218-2:2025) integrate the former ISO/TS 15066 collaborative requirement and were written for stationary articulated industrial robots in fixed cells (ISO, 2025). They define collaborative applications around four operational types (e.g. monitored standstill, hand guiding, speed and separation monitoring, and power and force limiting) but presuppose a workspace topology and hazard model that 3DCP does not match.

3DCP differs from cell-based collaborative robotics in four ways that matter for human–robot collaboration (HRC). First, the dominant on-site platform is a gantry: large, slow, with high inertia, traversing an open construction site rather than operating within a guarded cell. Second, the robot is tightly coupled to a continuous material process (e.g. pump pressure, nozzle flow, fresh concrete rheology) whose state cannot be paused without quality consequences such as cold joints, where bond strength reduces substantially with delay (Panda et al., 2018). Third, multiple worker roles operate concurrently in or near the work envelope. In other words, print operators monitor the system from a console, rebar placers physically intervene between layers, material handlers tend the pump and supply, and supervisors and other construction personnel occupy peripheral zones. These roles bear different physical and cognitive risks. Fourth, the work product is itself a hazard: a partially built and cured wall whose strength evolves over hours and where workers must approach for inspection, finishing, or steel placement.

The empirical work that does exist confirms the relevance of these factors but does not resolve them. Sawicki et al. (2026) presented the first quantitative comparison of human–robot collaborative shotcrete printing against manual cast concrete in equivalent reinforced beams, observing roughly threefold higher task-level productivity and reduced physical strain (carried weight at 44%, covered distance at 37%, perceived exertion at 63% of manual reference values) . However, it also flagged the high mental demand of the operator of the robotic system as a key finding, with heart rate and blood lactate unchanged between processes. In this case, strain shifted from the body to the mind and the framework that would catch that shift does not yet exist for 3DCP.

Generic collaborative robot (i.e. cobot) frameworks have made progress on adjacent problems. For instance, PRO-MIND (Lagomarsino et al., 2024) adapts robot trajectories using heart-rate variability and frantic-movement estimates of operator stress. The largest closed-loop perceived-safety study to date, with 48 subjects on a Kinova Gen3 cobot, found that heart-rate-based speed modulation reduced productivity but did not improve perceived safety (Tusseyeva et al., 2022). In addition, continuous physiological stress models such as cStress demonstrate that lab-quality recall (88.6% on lab data) drops substantially in the field (median 72%, an 18-percentage-point gap) (Hovsepian et al., 2015). These

findings transfer in principle, but the platforms, the worker roles, and the process do not.

This paper addresses these limitations by proposing a reference architecture that explicitly accounts for platform characteristics, worker roles, and process coupling in 3DCP. The design space of 3DCP HRC is organized into a three-layer taxonomy (Human, Robot, and Site) and instantiate that taxonomy as a site-wide predictive digital twin. This, in turn, acts as a continuously updated, role-aware, forward-simulating model of the print site whose outputs drive the robot's interaction with each worker present.

Related work

Standards landscape

Three regulatory frameworks bear directly on 3DCP HRC. ISO/ASTM 52939:2023 sets quality assurance requirements for additive construction but as noted, excludes EHS aspects of robotic operation from its scope (ISO/ASTM, 2023). ISO 10218-1:2025 and ISO 10218-2:2025 specify safety requirements for industrial robots and integrators respectively, with collaborative-application content (formerly ISO/TS 15066) integrated into the series. The 2025 revisions also clarify functional safety requirements and add cybersecurity provisions (ISO, 2025). The latest European Union (EU) regulation 2023/1230 on machinery, set to apply starting January 2027, introduces specific requirements for AI-based safety functions and treats software performing a safety function as a safety component subject to conformity assessment (European Parliament and Council, 2023). Together, these instruments define the regulatory envelope into which any 3DCP HRC architecture must fit, but none specifies how or which observable signals should drive that fit.

Beyond regulatory frameworks, a substantial body of research has explored the use of wearable and physiological sensing for estimating human affective and cognitive states in collaborative environments.

Psychophysiological monitoring in human–robot collaboration

The WESAD dataset (Schmidt et al., 2018) enabled standardized evaluation of multimodal stress detection from chest- and wrist-worn devices. The original benchmark reported up to 80% accuracy on a three-class problem (baseline, stress, amusement) and up to 93% on the binary stress versus non-stress task. Field studies routinely report accuracy losses relative to lab conditions. For instance, cStress, trained on a rigorous laboratory protocol with 21 participants and validated in a week-long field study with 20 participants, reported median field accuracy of 72% against self-reports compared with 90% in the laboratory (Hovsepian et al., 2015). These ceilings shape what any psychological-state-aware safety system can realistically claim.

In collaborative robotics, two threads matter for the present work. Trajectory-adaptive frameworks such as PRO-MIND (Lagomarsino et al., 2024) close a loop between operator state and robot motion. In this case, Heart Rate Variability (HRV) and frantic-movement features estimate stress and B-spline trajectories are re-optimised to balance smoothness against execution time. The empirical perceived-safety literature provides cautions. For instance, the 48-subject study by Tusseyeva et al. (2022) found that real-time speed reduction based on operator heart rate did not improve perceived safety but did reduce productivity, and that perceived safety improved with experimental habituation regardless of algorithm. The lesson is that physiological closed-loop control is not automatically safer in the user's perception and predictability matters as much as responsiveness.

Although psychophysiological monitoring enables estimation of human state in human–robot collaboration, its interpretation becomes more complex in construction environments where cognitive workload, multi-role interaction, and process constraints dominate human factors.

Human factors in construction robotics

Human factors research in construction robotics remains limited, with the field largely dominated by mechanical, materials, and process-oriented studies (Onososen and Musonda, 2023). In 3D concrete printing, Sawicki et al. (2026) show that collaborative shotcrete printing increases operator mental demand despite reduced physical strain, highlighting cognitive load as a key safety factor even when traditional physiological signals remain unchanged. Complementary simulation-based studies have explored spatial and cognitive aspects of human–robot interaction in construction, including game-based environments for analyzing interaction dynamics (Mansour et al., 2025).

Gaps and contributions

Existing cell-based collaborative robot frameworks do not transfer directly to 3D concrete printing. They assume a single operator role, a fixed and guarded workspace, localized robot hazards, and interruptible processes. On the other hand, 3D concrete printing involves multiple concurrent roles, an evolving open workspace, gantry-scale hazard envelopes, and continuous material deposition where interruptions can compromise structural integrity. Existing regulatory frameworks or data-driven HRC approaches in human–robot collaboration do not resolve these structural mismatches. Despite advances in regulatory frameworks, human state sensing, and HRC methods, there remains no unified representation that integrates human, robot, and site-level dynamics for real-time decision-making in construction environments. Physiological monitoring provides noisy estimates of cognitive and affective states that degrade in field conditions, while existing HRC approaches remain largely reactive and do not incorporate broader contextual or multi-actor site information. In parallel, construction robotics research highlights increasing cognitive workload and complex role interactions that are not captured by conventional sensing or control frameworks.

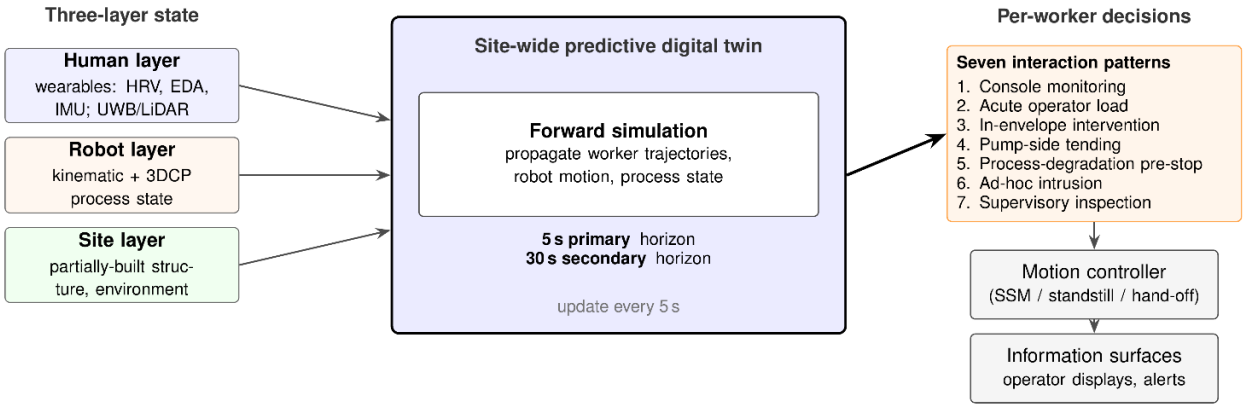


Figure 1: Proposed framework

These limitations reveal a gap between human state estimation, robot control strategies, and the operational realities of 3D concrete printing.

As such, this paper makes three contributions. First, it introduces a three-layer Human–Robot–Site taxonomy defining the required observables and state variables. Second, it develops a predictive digital-twin architecture that instantiates this taxonomy as a site-wide, continuously updated model with short-horizon forward simulation. This explicitly addresses gaps in existing safety provisions not covered by ISO/ASTM 52939 and extends beyond the assumptions of ISO 10218 collaborative robot standards. Third, it proposes a set of seven role-aware interaction patterns that map system states into actionable, standards-aligned robot behaviors for safe and auditable collaboration in 3D concrete printing. Figure 1 illustrates the proposed framework and highlights its three core components, which correspond to the contributions outlined above.

Component 1: The three-layer taxonomy

The proposed taxonomy organizes HRC-relevant observables and state variables into three layers reflecting what must be twinned to support safe collaboration. The three layers correspond to the three independently instrumented domains of the 3DCP print site and include the human counterparts, the robotic system, and the site they share, each with distinct sensing modalities, update cadences, and prediction horizons (Figure 2).

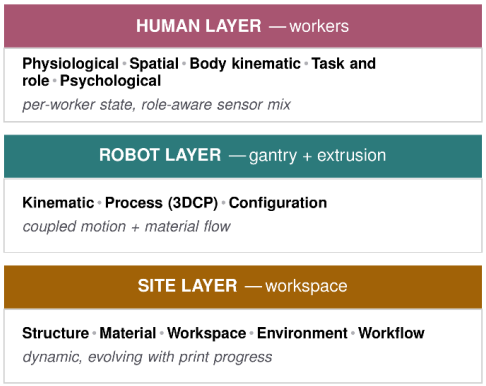


Figure 2: Proposed three-layer taxonomy

Human layer

The Human layer represents each worker present in the work envelope as an instance with five categories of state.

More specifically, the *physiological state* captures cardiovascular and electrodermal indicators relevant to acute stress and cognitive load: heart rate, heart-rate variability (RMSSD, SDNN, low-/high-frequency ratios), electrodermal activity, respiratory rate, and skin temperature. Where consented and operationally feasible, electroencephalography contributes a complementary cognitive-load and fatigue channel. In this case, industrial-grade in-helmet EEG products have demonstrated practicality at scale (e.g., over 5,000 deployed users in mining and heavy-vehicle operation), making it a credible addition to the sensor suite for operator stations rather than for all roles.

The spatial state captures three-dimensional position, velocity vector, body orientation, and a short-horizon predicted trajectory of the worker's torso and effective-arm-reach envelope. This is what enables proximity-based safety to function.

The body kinematic state, drawn from a small set of inertial measurement units (IMUs), captures three-axis acceleration, angular velocity, posture classification (standing, bent, kneeling, prone), activity classification, sudden-event detection (falls, slips, startles), and cumulative physical exertion. This layer differentiates a stationary supervisor from a rebar worker working between layers. Without it, role inference is incomplete.

On the other hand, *the task and role context* is a discrete attribute attached to each worker namely: role (operator, rebar worker, material handler, passer-by, and supervisor), current task, and experience level. More specifically, role determines which of the other state categories matter most for a given worker. For instance, a passer-by needs accurate spatial state but no physiological monitoring whereas an operator needs cognitive-load monitoring but their physical exertion is low. It is worth noting as well that the five-role list reflects the differentiation observed across current 3DCP deployments (Sawicki et al., 2026) and the framework accepts additional roles via the same role-context tag without structural change.

The inferred psychological state is a derived layer producing scalar or categorical estimates of stress, cognitive load, fatigue, and acute fear, computed from the physiological and kinematic channels above. The inference quality is constrained by the performance limits observed in field deployments, as reported in the previous section.

Robot layer

The Robot layer represents the printer as a coupled kinematic-process entity and includes several states

The kinematic and dynamic state includes: (1) position, velocity, and acceleration along the principal axes, (2) predicted trajectory along the planned toolpath, (3) current mass distribution (which on a gantry varies during extrusion as material moves through the system), (4) stopping distance under emergency braking, (5) vibration profile, and (6) instantaneous power and reaction forces. These are needed to compute speed and separation monitoring (SSM) envelopes that respect the platform's actual dynamics rather than an idealized cobot.

The process state is what makes 3DCP distinct. It includes the pump's commanded versus measured flow rate, pump-to-nozzle hose delay and rheological transit time, rheological state at the nozzle (yield stress, viscosity, thixotropic build-up), current toolpath segment, bead, and layer index, and the running gap between commanded velocity and feed rate. A robot whose kinematic state is nominal but whose process state is degrading is approaching a quality failure that may force an unplanned stop, and the safety system needs to know that.

Configuration captures setup-time invariants: bead height (H) and width (W), mix identification, and configured speed limits. These define the operating envelope within which kinematic and process state are interpreted.

Site layer

The Site layer represents the shared workspace as a dynamic structure rather than a static enclosure and includes several components.

The partially-built structure, unique to additive construction, is the most consequential novelty. It includes per-bead geometry, time since deposition (distributed across the structure as different elements were laid at different times), distributed curing state, and the nozzle's current position within the structure. A wall section deposited eight minutes ago is structurally weaker than one deposited eighty minutes ago and a worker leaning on the former is in a different hazard regime than one leaning on the latter. *On the other hand, the material delivery* tracks the pump-to-nozzle hose path and length, batch-plant state, mix supply, and material temperature. These determine when the process state at the nozzle will diverge from commanded values.

The workspace geometry tracks gantry-frame boundaries, hazard zones, sensor coverage maps (where can the proximity layer see, and where are blind spots), and the positioning of all workers and equipment within the envelope.

The *environmental state* includes ambient temperature, humidity, wind, and lighting, all of which influence both worker physiological baselines and concrete curing rates.

The *workflow and scheduling* captures the print-phase index, the schedule against which it is running, and the assignment of personnel to tasks over time. This is what enables the safety system to anticipate, rather than merely react to, role-aware interactions.

In summary, the three human, robot, and site layers are domains of content and are orthogonal to the functional decomposition of the safety controller (typically a fast proximity-tracking pipeline operating in tens of milliseconds and a slower physiological-state pipeline operating in seconds), which is well-established in the literature (Lagomarsino et al., 2024). The novelty here is not the functional split but the content schema that both pipelines need to reference.

Component 2: Predictive digital twin instantiation

A taxonomy describes but an architecture must compute. In this case, the three-layer taxonomy is instantiated as a site-wide predictive digital twin (Figure 1). The three taxonomy layers feed a continuously updated twin state, which forward-simulates worker trajectories, robot motion, and process state at 5 s primary and 30 s secondary horizons, emitting per-worker interaction-pattern decisions every 5 s. Additional details are provided in the following subsections.

Twinning fidelity across layers

Each layer requires a different update cadence and prediction horizon. Human-layer spatial and kinematic states must update at the perception subsystem's native rate (typically 10–30 Hz from UWB or LiDAR) to support proximity tracking. Human-layer physiological state updates at a slower pace consistent with the underlying signal physiology and feature-extraction windows. In this case, five seconds is a defensible primary cadence, matching the shortest meaningful HRV-feature horizon and giving the inference subsystem a stable estimate to publish. On the other hand, the robot-layer kinematic state updates at the controller's native rate (kHz at the joint level, but tens of Hz at the supervisory level relevant for HRC). The robot-layer process state updates at the pump-to-nozzle latency timescale (seconds). The site-layer partially-built structure state updates per layer or per bead, with curing models running on a slow background timer. The environmental state updates on tens-of-seconds intervals.

Developing the forward simulation mechanism

A reactive safety system catches a hazard when it has happened. A predictive system catches it before it happens which is of utmost importance in this study. Accordingly, the predictive digital twin runs a short-horizon forward simulation that propagates worker trajectories, robot motion, and process state forward under their current commands. There is a primary horizon of five seconds and another secondary horizon of thirty

seconds. The horizon is bounded by the prediction error of the spatial models. Beyond five seconds, intent becomes the dominant uncertainty rather than dynamics. The simulation produces, per worker, a probability surface over near-future positions and a corresponding intersection probability with the gantry envelope and process-active zone. Figure 3 illustrates the forward simulation across the primary (5 s) and secondary (30 s) horizons. The supervisory cycle issues a new pattern decision every 5 s based on predicted intersection between worker trajectories and the gantry envelope.



Figure 3: Forward simulation across the primary (5 s) and secondary (30 s) horizons.

Updating cadence and decision cycle

The supervisory decision cycle runs at five-second cadence. At each cycle, the digital twin (1) ingests the latest layer states, (2) computes per-worker forward simulations to the primary and secondary horizons; (3) computes interaction-pattern decisions for each worker, and (4) emits supervisory references to the underlying motion controller. Faster reflexes such as emergency stops and immediate proximity violations remain on a short-loop pipeline and are not gated on the supervisory cycle.

Critically, the digital twin does not collapse its outputs into a single site-wide safety scalar. Roles have different risk profiles and different optimal interactions. A single-scalar output erases that structure. Instead, the digital twin outputs a discrete interaction pattern per worker, drawn from the seven-pattern set described in the following section.

Component 3: Role-aware interaction patterns

Seven interaction patterns were identified, each defined by who the worker is, what state the digital twin estimates, and what response the robot adopts. The seven patterns emerge from cross-tabulating the worker roles observed in current 3DCP deployments (i.e. operator, rebar worker, material handler, supervisor, passer-by) with the dominant trigger types (spatial, physiological, process-state); together they cover the operationally common cases for gantry-based 3DCP, and a deployment may add or specialize patterns without breaking the schema. Table 1 summarizes the seven patterns. Each pattern combines a role qualifier, a state qualifier from one or more taxonomy layers, and a response specification that can be mapped to an ISO 10218-1/2:2025 collaborative-application mode. The advantage of expressing safety responses as a finite pattern set, rather than as a continuous function of state, is auditability. In this case,

Table 1: Seven role-aware interaction patterns and associated information

Pattern	Role	Triggering state	Digital Twin response	ISO 10218 mode
1. Console monitoring	Operator	Cognitive-load and fatigue estimates within nominal range; large spatial separation	Adjust information presentation (alert pacing, notification suppression); no motion change	Automatic operation
2. Acute operator load	Operator	Spike in cognitive-load or fear estimate from console-side wearables	Precautionary slowdown of next path segment; pause before high-consequence action	Speed and separation monitoring (SSM)
3. In-envelope intervention	Rebar worker	Spatial intrusion within envelope during scheduled inter-layer pause	Hand-off mode: motion constrained; re-engagement gated on task-completion signal or timeout	Hand guiding / SSM
4. Pump-side tending	Material handler	Moderate proximity; process-state divergence (commanded vs measured flow)	Monitor degradation; pre-position transition to Pattern 5 if divergence persists	SSM
5. Process-degradation pre-stop	All present	Imminent quality-driven stop predicted from process state	Extend path to graceful transition (end of bead or layer); notify Patterns 3 and 4 of pending pause	Monitored standstill (planned)
6. Ad-hoc intrusion	Passer-by	Spatial intrusion into envelope; no role data available	Standard SSM response; no role-specific or physiological gating	SSM
7. Supervisory inspection	Supervisor	Permitted in-envelope presence with role qualifier; proximity to recently-deposited elements	Moderate motion; consult Site layer to flag low-strength elements on supervisor's path	SSM with site-aware caution

each pattern can be traced to the standards-defined collaborative-application types (ISO, 2025) and to a specific risk hypothesis.

Scenario-based evaluation and discussion

Given the modular structure of the proposed framework (Figure 1), the evaluation focuses on the Human-layer physiological-state inference pipeline. This includes the feasibility of end-to-end signal processing, stress classification, and calibration mechanisms using a proxy multimodal dataset. This choice is motivated by the availability of suitable datasets and the ability to conduct controlled, reproducible testing in isolation. The Robot and Site layers, as well as their integration within the predictive digital twin and interaction patterns, require system-level deployment in construction environments and are therefore reserved for future validation.

Rebar worker during inter-layer pause scenario

Consider a rebar worker who enters the print envelope at t = 0 to install reinforcement during a scheduled inter-layer pause. At the next supervisory cycle (t = 5 s), the spatial subsystem flags intrusion within the envelope. The role-context tag identifies the worker as a rebar worker and the workflow-scheduling layer confirms an inter-layer pause is in progress. The digital twin classifies this configuration as Pattern 3 (i.e. in-envelope intervention) and transitions the gantry to a hand-off mode. Motion is constrained to the SSM-compliant envelope and the supervisory cycle permits re-engagement only upon either an explicit task-completion signal from the rebar worker or the expiration of a configured timeout. The process-state subsystem continues independently. If rheology at the nozzle degrades beyond threshold during the hand-off, Pattern 5 (i.e. process-degradation pre-stop) supersedes Pattern 3, and the gantry extends path completion to a graceful transition (end of bead) rather than executing an immediate halt. After the placer signals completion, the cycle returns to Pattern 1 (i.e. console monitoring) and motion resumes. The example illustrates two framework properties: pattern selection is per worker rather than global, and pattern transitions can be process-driven rather than only space-driven.

Preliminary feasibility check

To verify that the physiological-state pipeline of the Human layer is implementable end-to-end, a preliminary benchmark was conducted using the public WESAD dataset (Schmidt et al., 2018), which provides clean multimodal physiological signals (ECG, EDA, and skin temperature) with standardized stress labels. Although the dataset was not originally collected for construction applications, its multimodal structure provides sufficient signal richness to support a proxy-based evaluation of the proposed framework. In this study, it serves as a controlled proxy for validating preprocessing, windowing, model training, threshold selection, and evaluation procedures prior to the availability of construction-specific data.

Two model classes were evaluated under a strict subject-holdout protocol (train: S2–S11, validation: S13–S14, test: S15–S17) on the binary stress versus non-stress task. The models were (1) XGBoost (Chen and Guestrin, 2016), trained on engineered window-level features (summary statistics and temporal trends), and (2) a Temporal Fusion Transformer (Lim et al., 2021), trained on sequence windows. Signals were segmented into fixed-length windows with 50% overlap. They were standardized per subject using training-set statistics. The decision threshold was selected on the validation set by maximizing the F1 score. It was then applied unchanged to the test set. Each model was evaluated under two settings: with and without per-subject calibration profiles. These profiles consist of short resting segments appended as static covariates. They correspond to the calibration mechanism prescribed by the proposed framework for operator stations. Experiments used a deliberately reduced configuration, including downsampling and a capped number of windows per subject. This configuration was chosen to support rapid iteration. The reported results therefore demonstrate pipeline functionality rather than state-of-the-art performance.

Table 2 reports the Area Under the Receiver Operating Characteristic curve (AUROC), Area Under the Precision–Recall Curve (AUPRC), and F1 on the held-out test split (11,988 windows; stress prevalence ≈ 0.31). AUROC measures the model's ability to distinguish between classes across all decision thresholds. AUPRC is more informative under class imbalance, as it focuses on performance for the positive class. F1 score is the harmonic mean of precision and recall and reflects the balance between false positives and false negatives at a chosen threshold. With calibration profiles enabled, the Temporal Fusion Transformer outperforms the tree-based baseline. With profiles disabled, the sequence model performance collapses while XGBoost remains stable. This indicates stronger reliance of the sequence model on the personalization pathway prescribed by the framework. To ensure comparability, train/validation/test splits and class distributions were kept identical across configurations. The profiles-off configuration was rerun and reproduced identical metrics. Threshold selection was performed exclusively on the validation set.

Table 2: WESAD pipeline feasibility check on the held-out test split

Setting	Model	AUROC	AUPRC	F1
Profiles on	XGBoost	0.712	0.542	0.498
Profiles on	TFT	0.768	0.683	0.576
Profiles off	XGBoost	0.712	0.542	0.498
Profiles off	TFT	0.548	0.354	0.302

These results demonstrate that the Human-layer inference pipeline can support stable stress classification and calibration under controlled conditions, representing an

initial step toward operationalizing the proposed framework.

The seven-pattern model, role-aware response logic, and process-coupled patterns (4 and 5) remain to be evaluated in deployment settings.

Conclusion, limitations, and future work

This paper has presented a three-layer taxonomy (Human, Robot, Site), an instantiating site-wide predictive digital twin for human–robot collaboration in 3D concrete printing, seven role-aware interaction patterns, and a scenario-based evaluation of the Human-layer physiological-state inference pipeline. The proposed taxonomy structures the design space of observables and state variables, while the digital twin operationalizes this structure into a continuously updated model. This model supports seven role-aware interaction patterns mapped onto ISO 10218-1/2:2025 collaborative-application modes. The framework is platform-agnostic across gantry, robot-arm, and crane-based 3DCP systems and is intended as a reference architecture for future deployment and standardization. The work is motivated by a regulatory gap, as ISO/ASTM 52939:2023 excludes the EHS aspects of robotic operation, and by an empirical gap in existing collaborative-robot frameworks, which do not account for process-coupled hazards or multi-role construction environments.

While the paper presents an innovative framework, it does not constitute a deployed system. The main limitation is that the proposed interaction patterns and layered digital twin have not yet been validated in integrated real-world 3D concrete printing deployments. The Human-layer evaluation demonstrates feasibility under controlled conditions, but full system-level validation across Human, Robot, and Site layers remains to be conducted.

Future work will proceed in three directions. First, system-level deployment and validation will be carried out on an operational gantry-based 3DCP platform, including quantitative assessment of false positives and false negatives under scripted hazard scenarios. Second, role inference mechanisms will be developed to enable automatic classification of worker roles from kinematic and contextual signals, reducing reliance on manual role assignment. Third, the interaction pattern set will be formalized as a normative contribution aligned with the ISO 10218 series, alongside engagement with ISO/ASTM TC261/F42 to support future revisions of ISO/ASTM 52939.

Acknowledgements

The presented work was supported by the American University of Beirut (AUB) Vertical Integration Projects (VIP) Program. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the VIP Program or its stakeholders.

References

Bartneck, C., Kulić, D., Croft, E. and Zoghbi, S. (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics, 1(1), pp.71–81.

Buswell, R.A., Leal de Silva, W.R., Jones, S.Z. and Dirrenberger, J. (2018) 3D printing using concrete extrusion: A roadmap for research. Cement and Concrete Research, 112, pp.37–49.

Chen, T. and Guestrin, C. (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA, ACM, pp.785–794.

COBOD International (2024) First 3D Printed Social Housing Project Built to New ISO/ASTM Standard Completed 35% Faster Than With Conventional Methods. Press release, 17 December 2024.

European Parliament and Council (2023) Regulation (EU) 2023/1230 of 14 June 2023 on machinery and repealing Directive 2006/42/EC. Official Journal of the European Union, L 165, 29 June 2023, pp.1–102.

Hovsepian, K., al'Absi, M., Ertin, E., Kamarck, T., Nakajima, M. and Kumar, S. (2015) cStress: Towards a gold standard for continuous stress assessment in the mobile environment. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15). New York, NY, USA, ACM, pp.493–504.

ISO (2025) ISO 10218-1:2025 — Robotics — Safety requirements — Part 1: Industrial robots and ISO 10218-2:2025 — Robotics — Safety requirements — Part 2: Industrial robot applications and robot cells. Geneva, International Organization for Standardization.

ISO/ASTM (2023) ISO/ASTM 52939:2023 — Additive manufacturing for construction — Qualification principles — Structural and infrastructure elements. Geneva, International Organization for Standardization.

Lagomarsino, M., Lorenzini, M., De Momi, E. and Ajoudani, A. (2024) PRO-MIND: Proximity and Reactivity Optimisation of robot Motion to tune safety limits, human stress, and productivity in INDustrial settings. arXiv preprint arXiv:2409.06864. (Forthcoming, IEEE Robotics and Automation Letters.)

Lim, B., Arık, S.Ö., Loeff, N. and Pfister, T. (2021) Temporal fusion transformers for interpretable multi-horizon time series forecasting. International Journal of Forecasting, 37(4), pp.1748–1764.

Mansour, A., Abou Khalil, A. & Khoury, H.M. (2025) Toward spatial and cognitive human–robot interaction: A game-based simulation for concrete 3D printing in construction. In: Proceedings of the Creative Construction Conference (CCC). Zadar, Croatia, IAARC, pp. 750–755.

Marvel, J.A. (2013) Performance metrics of speed and separation monitoring in shared workspaces. IEEE Transactions on Automation Science and Engineering, 10(2), pp.405–414.

Onososen, A.O. and Musonda, I. (2023) Research focus, trends and outlook on construction robotics for human-robot teams: A scientometric review. Journal of Engineering, Design and Technology, 21(2), pp.502–526.

Panda, B., Paul, S.C., Mohamed, N.A.N., Tay, Y.W.D. and Tan, M.J. (2018) Measurement of tensile bond strength of 3D printed geopolymer mortar. Measurement, 113, pp.108–116.

Roussel, N. (2018) Rheological requirements for printable concretes. Cement and Concrete Research, 112, pp.76–85.

Sawicki, B., Düking, P., Placzek, G., Masur, L., Dörrie, R., Schwerdtner, P. and Kloft, H. (2026) Human–robot collaboration in digital fabrication with concrete: quantifying productivity and psychophysiological strain of human workers. Construction Robotics, online first.

Schmidt, P., Reiss, A., Duerichen, R., Marberger, C. and Van Laerhoven, K. (2018) Introducing WESAD, a multimodal dataset for wearable stress and affect detection. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI). Boulder, CO, USA, ACM, pp. 400–408.

Decentralized Data, Cybersecurity & Trustworthy Automation

Sub-Tracks:

Cybersecurity & Data Integrity

Blockchain + BIM Integration

Smart Contracts in Construction

Cryptoeconomics & Governance / DAOs

Transparency, Explainable AI & Accountability

Ethical Automation & Trust Mechanisms

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

SUPPORTING CONSTRUCTION CYBERSECURITY DECISION-MAKING THROUGH A CROWDSOURCING PLATFORM

Muammer Semih Sonkor[1], and Borja García de Soto[1]
[1]S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates

Abstract

Construction is increasingly digitalized, and organizations are integrating operational technology, such as robots, into site operations, introducing cyber risks. This paper presents a web-based crowdsourcing platform to collect and analyze human knowledge and to provide recommendations to guide cybersecurity decision-making in the construction sector. Collected insights on risks and mitigations via open-ended questions are converted into recommendations using semantic segmentation, topic modeling, and large language models. The platform was implemented in collaboration with a UAE-based contractor on a lift-assist robot. The resulting recommendations were mapped to the foundational requirements of ISA/IEC 62443, showing alignment with six of the seven requirements in this case study.

Introduction

The construction sector has been undergoing a digital transformation, affecting site tasks. Besides the digitalization of documents and the increasing use of common data environments for data exchange and coordination, construction sites are integrating operational technology (OT), such as robotics, connected equipment, and IoT devices (García de Soto et al., 2020). While these technologies aim to improve productivity and quality, they also introduce new cyber risks into construction environments that are dynamic and often involve close interaction between humans and machines (García de Soto et al., 2020). Therefore, cybersecurity in construction is not only a concern related to confidentiality, integrity, and availability of information and resources, but also a matter of safety.

While cybersecurity emerged as a research topic in the 1970s (Lipner, 2015) and has been widely studied over time, it has not yet received enough attention in the construction domain, leaving essential areas to be explored (Salami Pargoo and Ilbeigi, 2023). This can be explained by the relatively lower maturity of digitalization in the sector (Tambwe et al., 2025) compared to more technologically advanced ones, such as energy and manufacturing. However, the increase in cyberattacks (especially ransomware) targeting construction companies (Check Point, 2025) necessitates investigating approaches suitable for the characteristics of construction projects.

Crowdsourcing is one of the effective mechanisms deployed by companies to address the lack of in-house cybersecurity expertise (Zhao et al., 2023). It has mostly been associated with bug bounty programs in the cybersecurity context, where organizations invite external security experts through an open call to identify and disclose vulnerabilities in defined systems and code bases before they are exploited by malicious actors (Malladi and Subramanian, 2020). Another common application is crowdsourced cyber threat intelligence (CTI), which involves aggregating information from open unstructured sources (e.g., social media) and structured repositories (e.g., vulnerability databases) to detect emerging threats and vulnerabilities (Sun et al., 2020).

While these approaches are highly effective in identifying isolated software vulnerabilities, they do not yield holistic, system-level guidance. Construction sites that integrate complex technologies such as robotics and connected equipment, along with building information modeling (BIM)-enabled information workflows, require deployment-oriented recommendations rather than lists of code-level findings. Such guidance is typically provided by cybersecurity consultants (often at high cost) or by in-house experts who are scarce in the construction sector, according to a recent report (Sophos, 2025). The report involved a survey of 3,700 information technology (IT) professionals from 17 countries, in which 41% of the participants from the construction and property sectors indicated that the lack of expertise was the root cause of falling victim to ransomware attacks (Sophos, 2025).

This study presents a crowdsourcing platform developed for the construction sector that gathers cybersecurity knowledge through structured open-ended questions and converts it into actionable, management-ready recommendations using a reproducible data analysis pipeline. It addresses the need for cost-effective, technology-specific cybersecurity guidance for increasingly digitized and autonomous construction sites. The platform was implemented in collaboration with a UAE-based contractor focusing on a lift-assist robot designed to support blocklaying.

In this study, the input was collected from the contractor's employees, and the resulting recommendations were mapped to ISA/IEC 62443 foundational requirements (FRs) to support systematic cybersecurity decision-making. ISA/IEC 62443 is an international cybersecurity

standard series developed for industrial automation and control systems (IACS) and broader OT environments, defining structured security requirements throughout their lifecycles.

Against this background, this study aims to address the following research questions (RQs):

- (RQ1) How should a crowdsourcing platform and analysis pipeline be structured to convert employee cybersecurity knowledge into actionable, decision-support recommendations for construction management?
- (RQ2) To what extent can internal crowdsourcing capture OT-related cybersecurity needs in construction?

RQ1 is addressed through the proposed platform architecture and the end-to-end analysis workflow presented in the Methodology section. Moreover, RQ2 is addressed by showing the extent to which the identified recommendations cover the ISA/IEC 62443 foundational requirements.

Background and Related Work

This section reviews literature identified through a targeted review of peer-reviewed studies and industry reports on construction cybersecurity and crowdsourcing.

Cybersecurity in Construction

The digital transformation of the construction industry, accelerated by the convergence of information technology (IT) and OT (e.g., autonomous machinery, IoT devices), has significantly expanded the sector's cyberattack surface (Sonkor and García de Soto, 2021). In addition to the increase in ransomware attacks targeting the sector (Check Point, 2025) in recent years, high-profile incidents such as the $25 million deepfake fraud targeting Arup (Elliott, 2025) have shown that the construction industry is on hackers' radar. The growing interest from malicious actors can be explained by the low cybersecurity preparedness among construction companies (QBE, 2024), which makes them easy targets.

To address these challenges, recent research emphasized the need for tailored vulnerability assessment methods and governance frameworks that account for the unique, dynamic nature of construction sites and their complex supply chains (Mantha et al., 2024). Moreover, Salami Pargoo and Ilbeigi (2023) indicated the need for research to develop methodologies to safeguard automated and intelligent systems used in construction projects. In addition to research, several standards and guidelines have been published to address the increasing cyber risks in the construction industry. Some examples are ISO 19650-5, which focuses on security-minded information management in BIM-enabled projects, and the code of practice "Cyber Security in the Built Environment" by the Institution of Engineering and Technology (IET). These documents provide high-level guidance to construction companies and do not focus on specific construction technologies.

Internal Crowdsourcing as a Knowledge Mobilization Mechanism

Crowdsourcing is a collaborative method for solving specific problems by leveraging a dynamic group of people formed through an open call (Pedersen et al., 2013). In cybersecurity, it is commonly utilized through programs such as bug bounties, which involve independent external experts. Although effective in many software contexts, such approaches are not always suitable for OT, where safety and availability considerations come first (Malladi and Subramanian, 2020). These concerns also apply to construction projects that use OT, such as robotic systems, IoT devices, and autonomous machinery, as they interact with the physical environment. In addition, construction systems often involve mixed ownership and responsibility (e.g., vendor-managed remote connectivity, site networks managed by the IT department, operational procedures managed by site teams), complicating external engagement.

An alternative to traditional crowdsourcing is internal crowdsourcing, where companies mobilize distributed employee knowledge for problem-solving or ideation (Ulbrich et al., 2021). One advantage of internal crowdsourcing is employees' better understanding of their organization and its operational conditions (Malhotra et al., 2017). Moreover, intellectual property and confidentiality issues are eliminated because internal participants' employee contracts already address property rights (Malhotra et al., 2017). However, companies often struggle to motivate employees to participate in internal crowdsourcing programs, and monetary rewards alone may not always be sufficient, requiring additional incentives such as recognition (Beretta et al., 2023).

The approach proposed in this study differs from bug bounties in both its participation model and intended output. Rather than expecting the participants to identify isolated code-level vulnerabilities and provide structured technical reports, it collects knowledge through open-ended questions to derive deployment-oriented, technology-specific recommendations for construction management teams. Moreover, the platform introduced in this study was not developed solely for internal crowdsourcing purposes. However, given the challenges of engaging external participants in a cybersecurity case study involving safety-critical technology, participation was limited to the contractor's employees. Therefore, the case study in this paper explores how internal crowdsourcing can be used to support cybersecurity decision-making in construction companies.

Methodology

This study adopts a design science research approach (Hevner et al., 2004), which is characterized by the creation of purposeful and innovative artifacts (in this study, a crowdsourcing platform) to solve identified practical problems.

The research was implemented through an exploratory single-case study (Yin, 2009). This methodological choice is appropriate for investigating the efficacy of new technological interventions in their real-world context, particularly in specialized and complex domains such as construction robotics.

Overview of the Crowdsourcing Platform

The proposed platform is a web-based system that enables construction companies to define a technology to be investigated, provide contextual documentation and technical details, and collect employee or external participant responses to structured open-ended questions. Submissions are then analyzed through a reproducible data analysis pipeline that identifies the common themes and derives actionable recommendations for management.

As shown in Figure 1, the high-level process flow of the proposed crowdsourcing platform involves four steps: (1) challenge creation, (2) response collection, (3) data analysis pipeline ((3a) threats and vulnerabilities and (3b) mitigations), and (4) recommendation synthesis.

System Architecture

The platform follows a standard web application structure, composed of the following main components:

- **Frontend/User Interface (UI):** Django templates (using HTML, CSS, JavaScript) implement (i) a submission dashboard for contributors to see the technical details of the investigated technology and submit free-text responses to open-ended questions, and (ii) a management dashboard for decision-makers to view the recommendations derived from the analysis.
- **Backend data handling:** Django handles HTTP requests and response processing and manages data exchange with the database.
- **Database:** MySQL stores user submissions, metadata, and analysis artifacts (e.g., derived recommendations).
- **Data Analysis Pipeline:** Python modules (separate for "threats and vulnerabilities" and "mitigations") execute the data analysis pipeline that involves data preparation, coherence-based semantic segmentation, topic modeling, and LLM-based topic labeling and summarization. The whole pipeline, including the LLM-based topic labeling and summarization, is executed locally.
- **Recommendation Synthesis:** A Python module performs LLM-based synthesis to fetch the analysis outputs from the database and convert them into a set of cybersecurity recommendations tailored to the investigated technology. The resulting recommendations are stored in the database and displayed on the management dashboard.

Overview of the Case Study

The platform was implemented in collaboration with a UAE-based contractor focusing on the MULE MZ100 (Figure 2) robot from the manufacturer Construction Robotics. The MZ100 is a lift-assist robot designed to support block laying by enabling operator-guided semi-automated lifting and positioning. The system includes an operator control panel at the end effector, load sensors intended to prevent overloads and unsafe movements, and built-in cellular connectivity that provides real-time status data and enables remote service functions (e.g., troubleshooting, diagnostics, software updates). Firmware updates require an on-site technician who connects a dedicated service cable between the machine and a laptop. These features create a mixed cyber-physical exposure profile with both physical interfaces and network-enabled functionality.

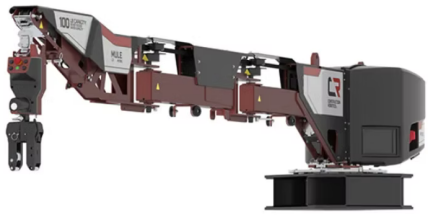


Figure 2. MULE MZ100 by Construction Robotics (Alley-Cassetty, 2026)

The case study was announced as an open call for participation within the company, targeting employees in the Design, Innovation, Digital Construction, and IT departments, as identified by the contractor. Participation was voluntary, and the analysis reported in this paper is based on responses from 11 employees who chose to contribute. The pilot ran asynchronously over two months, and participants were incentivized through the contractor's internal innovation points system managed by the Innovation department. The case study included two thematic lanes and a set of questions for each, presented on the platform: Lane A (Threats and Vulnerabilities; Q1-

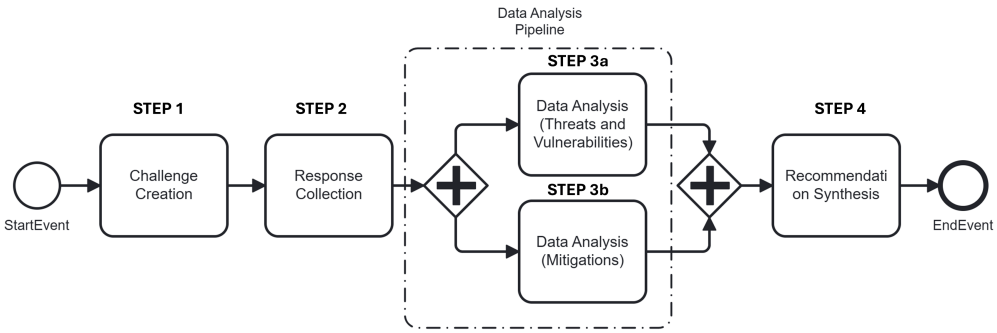


Figure 1. High-level process flow of the proposed crowdsourcing platform

Q4) and Lane B (Mitigations; Q5-Q6). All questions are presented in Table 1.

Table 1. Crowdsourcing questions used in the case study

ID	Questions
Q1	Considering the provided robotic system, what could be done to compromise the data collected? (e.g., sensor readings)
Q2	Considering the provided robotic system, what could be done to disrupt or manipulate its operation?
Q3	In your opinion, what are the most vulnerable components from the cybersecurity perspective? Why?
Q4	If you were an attacker, what would be your motivation to compromise the provided robotic system?
Q5	What would you suggest to make the system less vulnerable?
Q6	What questions should be asked to the vendor before deploying such equipment on site to identify potential cybersecurity issues?

Data Analysis Pipeline

Algorithm 1 summarizes the data analysis pipeline that converts free-text responses collected from the contributors into cybersecurity topics.

Algorithm 1. Text-to-topics data analysis pipeline

// Step 1: Fetch and Prepare the Data
1: Fetch responses from the database:
2: If Lane A: use Q1-Q4 (concatenated)
3: If Lane B: use Q5-Q6 (concatenated)
4: Clean the raw text (e.g., normalize whitespace, remove non-informative artifacts).

// Step 2: Semantic Segmentation (cosine similarity)
5: Split responses into sentences and embed them using SBERT.
6: Generate segmented corpora S_τ by applying cosine similarity thresholding to consecutive sentence embeddings over τ values (Table 2).

// Step 3: Coherence-guided selection of τ^* and k^*
7: For each candidate (τ, k) combination (Table 2), fit BERTopic on S_τ and compute c_v coherence on S_τ.
8: Select (τ^*, k^*) that maximizes c_v.

// Step 4: UMAP Hyperparameter Search (Select u^*)
9: With S_{τ^*} and k^* fixed, fit BERTopic using candidate UMAP settings (Table 2), evaluate with c_v, and select u^*.

// Step 5: HDBSCAN Hyperparameter Search (Select h^*)
10: With S_{τ^*}, k^*, and u^* fixed, fit BERTopic using candidate HDBSCAN settings (Table 2), evaluate with c_v, and select h^*.

// Step 6: Final Topic Model Fitting
11: Fit the final BERTopic model on S_{τ^*} using (k^*, u^*, h^*).
12: Extract topic keywords and representative segments for each non-outlier topic.

// Step 7: Topic Labeling and Summarization
13: Use a locally deployed LLM to generate a short label and summary for each topic (model and configuration reported in Table 2).
14: Store topic labels and summaries in the database.

The whole pipeline presented in Algorithm 1 is executed in parallel for both lanes (1) Lane A: Threats & Vulnerabilities (Q1-Q4) and (2) Lane B: Mitigations (Q5-Q6) to produce two sets of topics, including their summaries and labels. These topics are then synthesized into recommendations for management. In Algorithm 1, S_τ denotes the segmented corpus produced using threshold τ, k denotes the target topic count nr_topics, u^* denotes the selected UMAP configuration (n_neighbors, n_components) that maximizes coherence, and h^* denotes the selected HDBSCAN configuration (min_cluster_size, min_samples) that maximizes coherence.

In all search stages (i.e., Steps 3-5 in Algorithm 1), candidate configurations were evaluated using the c_v coherence metric (Röder et al., 2015) computed on the segmented corpus, and the best configuration was selected using an argmax rule. The options that produced no valid non-outlier topics were assigned a worst-case score (i.e., $-\infty$) to prevent selection. Table 2 presents the thresholds and grid-search values used in Algorithm 1.

Table 2. Values tested for each parameter in Algorithm 1

Step	Component	Parameter(s)	Values tested
Step 2	Semantic segmentation	Similarity threshold (τ)	$\{0.10, 0.15, \ldots, 0.90\}$
Step 3	Initial topic modeling search	nr_topics (k)	$\{5, 6, 7\}$
Step 4	UMAP search	n_neighbors	$\{5, 10, 15, 20, 25, 30, 35\}$
Step 4	UMAP search	n_components	$\{2, 5, 10, 15\}$
Step 5	HDBSCAN search	min_cluster_size	$\{5, 10, 15, 20, 25\}$
Step 5	HDBSCAN search	min_samples	derived from min_cluster_size (see note*)
Step 7	Topic labeling & summarization	Model; temperature	Gemma 3 27B (self-hosted via Ollama); temperature = 0

***Note:** p ∈ $\{0.10, 0.50, 1.00\}$;
min_samples = $max(1, int(p \times min_cluster_size))$

To identify semantic boundaries between consecutive sentences, the pipeline computes cosine similarity between their SBERT embeddings and starts a new segment when the similarity falls below the threshold τ (Steps 2 in Algorithm 1). Cosine similarity was selected since it is independent of vector magnitude, focusing strictly on the semantic orientation of the text. This choice is consistent with the original SBERT study (Reimers and Gurevych, 2019), which used cosine similarity to compare sentence embeddings. Cosine similarity between two embedding vectors (e_i) and (e_{i+1}) is calculated as follows:

$$\mathrm{sim}(\mathrm{e}_i, \mathrm{e}_{i+1}) = \frac{\mathrm{e}_i \cdot \mathrm{e}_{i+1}}{\|\mathrm{e}_i\| \|\mathrm{e}_{i+1}\|} \quad (1)$$

In Step 7 of Algorithm 1, a structured prompt template was used for LLM-based topic labeling and summarization, organized into instruction, context, input data, and output indicator, as suggested by Giray (2023). Lane-specific templates were used to maintain focus on either "Vulnerabilities and Threats" or "Mitigations". The prompt templates are not included in this paper due to page limitations; however, they can be provided upon reasonable request. Moreover, this study does not claim novelty in the individual analytical components of

Algorithm 1, since they are established techniques. Instead, the contribution lies in the proposed coherence-guided pipeline, which iteratively adjusts the semantic segmentation threshold (τ) and topic modeling configuration to maximize the coherence of the final topic-modeling output.

Recommendation Synthesis

To identify actionable recommendations for construction management, the outputs of the two analysis lanes (i.e., topic labels and summaries) were synthesized using a locally deployed LLM. The LLM (i.e., Gemma 3 27B via Ollama) was prompted to generate 5-7 actionable recommendations for construction decision-makers. The pseudocode presented in Algorithm 2 summarizes this procedure.

Algorithm 2. Recommendation synthesis workflow

1: Fetch topic labels and summaries for Lanes A and B from the database:
2: Create a synthesis prompt that instructs the model to:
- act as a Chief Information Security Officer (CISO)
- generate 5-7 actionable recommendations for a construction management team
- use only the provided topic labels and summaries as input (no other assumptions)
- write implementable and technical guidance without extra narration
3: Query the locally deployed LLM.
4: Store the resulting recommendations in the database.

This section (Methodology) addresses RQ1 by describing the proposed platform design and analysis workflow, specifying how employee knowledge was collected, processed, and converted into cybersecurity recommendations for construction management teams.

Results and Discussion

Overview of the Participants and Collected Responses

In this case study, the crowdsourcing platform received contributions from 11 employees across four departments within the collaborating construction company. Employees' departments and the top five keywords extracted from their responses (stop words removed) are presented in Table 3. Across all submissions, the Threats and Vulnerabilities lane (Q1-Q4) contained 1,183 words (108 words per employee submission, on average), whereas the Mitigations lane (Q5-Q6) contained 534 words (49 words per employee submission, on average). This difference is expected given the larger number of questions in the Threats and Vulnerabilities lane.

Platform Outputs: Actionable Recommendations for Construction Management

For traceability, Table 4 and **Error! Reference source not found.** present the topics identified by the data analysis pipeline (Figure 1) for the (i) Threats and Vulnerabilities and (ii) Mitigations lanes, respectively. These topics reflect the main themes that emerged from employees' contributions (e.g., concerns about remote access, data integrity, and physical access points). Topic summaries are not included in this paper due to page limitations; however, they are available in a repository (Sonkor and García de Soto, 2026).

Table 3. List of the contributing employees and top keywords

Employee #	Department	Top 5 Keywords
Employee 1	Design	software, data, network, communication, secure
Employee 2	Design	manipulated, opportunity, troubleshooting, telemetry, firmware
Employee 3	Design	access, program, unauthorized, hacking, mule
Employee 4	Design	remote, connectivity, connect, remotely, unsafe
Employee 5	Design	sensors, data, sensor, signals, confuse
Employee 6	Design	access, data, robot, power, control
Employee 7	Design	data, access, physical, updates, robotic
Employee 8	Design	data, equipment, connected, device, access
Employee 9	Innovation	data, security, access, incorrect, training
Employee 10	Digital Construction	access, data, machine, think, physical
Employee 11	IT	access, unauthorized, public, addresses, system

Table 4. Topics identified for threats and vulnerabilities

ID	Topic Label	Top keywords (weight)
T1	Robotic System Manipulation & Unauthorized Access	robot (0.07), sensor (0.05), weak (0.05), tampering (0.05), access (0.05), power (0.04), command (0.04), manipulation (0.04), input (0.04), feeding false (0.04)
T2	Remote Access & Physical Security Risks	access (0.07), remote (0.07), malicious (0.06), update (0.05), security (0.05), cause (0.04), control (0.04), data (0.04), physical (0.04), remote connectivity (0.03)

Table 5. Topics identified for mitigations

ID	Topic Label	Top keywords (weight)
M1	Robotic Equipment Cybersecurity: Mitigation Strategies	security (0.09), data (0.06), secure (0.06), communication (0.05), vendor (0.05), cyber (0.05), monitoring (0.05), encryption (0.04), defense (0.04), standard (0.04)
M2	Robotic Equipment Access Control & Remote Connectivity	access (0.13), remote (0.07), device (0.04), maybe (0.04), physical (0.04), robot (0.04), remote access (0.04), unauthorized (0.04), authentication (0.04), data (0.03)

The platform's recommendation synthesis step (Figure 1) then identified management-ready controls by using the identified topics as input and an LLM prompt template, ensuring that the recommendations are grounded in crowd-derived themes rather than relying on the LLM's

parametric knowledge. A list of the identified recommendations is presented in Table 6.

Table 6. Recommendations identified by the crowdsourcing platform

ID	Recommendation
R1	Implement physical access controls to the MULE MZ100, including lockable panels covering the service port and a logged, supervised access procedure for any physical interaction with the robot's internal components.
R2	Enforce multi-factor authentication (MFA) for all remote access connections to the MULE MZ100, utilizing a strong authentication method beyond passwords, such as time-based one-time passwords (TOTP) or hardware tokens.
R3	Restrict remote access to the MULE MZ100 to a pre-approved list of authorized personnel and block all incoming connections originating from public IP addresses using firewall rules.
R4	Conduct thorough vendor vetting of Construction Robotics, including a review of their security practices, compliance certifications (e.g., SOC 2), and a detailed questionnaire regarding potential vulnerabilities and incident response plans.
R5	Implement encrypted communication channels for all data transmitted to and from the MULE MZ100, utilizing TLS 1.3 or higher, and verify certificate validity regularly.
R6	Establish a logging system to record all physical and remote access attempts to the MULE MZ100, including timestamps, user IDs, and actions performed, and review these logs regularly for anomalies.
R7	Contractually obligate remote support vendors providing access to the MULE MZ100 to adhere to specific security standards and assume responsibility for any security vulnerabilities introduced through their access or actions.

The recommendations listed in Table 6 are presented as controls based on the identified topics that a construction management team can practically implement. Each recommendation suggests technical actions to safeguard the deployed technology (i.e., MULE MZ100) from different aspects (e.g., integrity, availability) and monitor it to detect potential attacks in a timely manner. Moreover, the recommendations include vendor governance actions (i.e., R4, R7), which are critical in construction OT contexts since vendors often control parts of the system lifecycle (e.g., remote support, update mechanisms) while the construction IT team might be responsible for other parts, such as deployment on site and connectivity. This might make ownership and responsibility boundaries ambiguous in projects.

Mapping Recommendations to ISA/IEC 62443 Foundational Requirements

As mentioned in the Introduction section, ISA/IEC 62443 is an international standards series that establishes cybersecurity requirements for securing IACS. IEC 62443-4-2—a part of this series—structures these requirements around seven component-level FRs (Table 7) to support holistic protection against threats. Although construction OT differs from traditional IACS, robotic equipment such as the MULE MZ100 exhibits exposure patterns comparable to those of traditional IACS (e.g., remote connectivity, maintenance interfaces) and may have operational and safety implications if compromised. In this study, the recommendations were mapped to the ISA/IEC 62443 FRs to provide a standardized way to present which security objectives are addressed, highlight potential gaps, and support interpretation. This mapping is intended to illustrate the alignment of the recommendations with the FRs and does not claim full satisfaction of any of them. The mapping is shown in Table 8.

Table 7. Foundational requirements from ISA/IEC 62443

Foundational Requirement	Description
FR1 - Identification & Authentication Control (IAC)	Ensure all users, services, and devices are uniquely identified and strongly authenticated before any system access is granted.
FR2 - Use Control (UC)	Limit authenticated entities to the minimum necessary actions via role-based authorization and enforceable policies.
FR3 - System Integrity (SI)	Preserve the integrity of software, firmware, and configurations by preventing, detecting, and recovering from unauthorized changes or code execution.
FR4 - Data Confidentiality (DC)	Safeguard data in transit and at rest through robust cryptography, sound key management, and secure handling practices.
FR5 - Restricted Data Flow (RDF)	Constrain communications by segmenting networks and enforcing least-route, allowlisted flows across defined trust boundaries.
FR6 - Timely Response to Events (TRE)	Detect, record, analyze, and act on security-relevant events promptly to contain incidents and support recovery.
FR7 - Resource Availability (RA)	Maintain essential functions during faults or attacks by resisting resource exhaustion, providing resilience/redundancy, and enabling rapid restoration.

Table 8. Mapping of recommendations to ISA/IEC 62443 FRs

Recommendation ID	Mapped Foundational Requirement(s)
R1	• FR2 - Use Control (UC)
R2	• FR1 - Identification & Authentication Control (IAC)
R3	• FR2 - Use Control (UC) • FR5 - Restricted Data Flow (RDF)
R4	• FR3 - System Integrity (SI) • FR6 - Timely Response to Events (TRE)
R5	• FR4 - Data Confidentiality (DC)
R6	• FR6 - Timely Response to Events (TRE)
R7	• FR2 - Use Control (UC) • FR3 - System Integrity (SI)

As shown in Table 8, the identified recommendations collectively address six of the seven ISA/IEC 62443 FRs

(Table 7), indicating alignment with key security objectives and thus addressing RQ2. FR7 (i.e., Resource Availability) is not addressed by any recommendations, suggesting that availability-focused controls were less prominent in employee contributions. This can be explained by both the question design and the participants' lack of cybersecurity expertise. In future case studies, this can be addressed by adding availability-oriented questions during the submission phase and involving participants with deeper knowledge of different facets of cybersecurity.

Discussion in the Construction Context

A key value of the crowdsourcing approach in this study is that it surfaced context-dependent issues that are difficult to derive from generic control catalogs alone. Although the resulting recommendations are preliminary and specific to this case study, they illustrate how employee knowledge can be translated into deployment-oriented site- and technology-specific actions for construction management. The identified recommendations can be grouped into four operational categories relevant to construction decision-making:

- **Physical access management (R1):** Construction equipment is often operated in environments where physical access is difficult to control completely. The MZ100's firmware update process involves on-site connection to a laptop via a dedicated service cable, indicating the existence of service access pathways that should be protected through physical safeguards and clear site procedures.
- **Remote access hardening (R2, R3):** Unlike most industrial facilities, construction sites mostly rely on temporary connectivity (e.g., site Wi-Fi, cellular networks), which increases the likelihood of inconsistent configurations and shared credentials. In this case study, cellular connectivity enables efficient remote software updates and troubleshooting, given the robotic system's potentially changing location as the project progresses; however, it also creates a high-impact threat surface. Safeguarding remote access through MFA and restricting it to a pre-approved list of personnel can reduce the risk during site transitions.
- **Secure communication and monitoring (R5, R6):** Construction projects generate operational data that can influence production planning and site decision-making. Therefore, telemetry and real-time status data introduce integrity concerns. Encryption and certificate validation reduce interception risk, while logging and anomaly review support detection and response.
- **Vendor governance and responsibility allocation (R4, R7):** Construction OT deployments often involve multi-party responsibility, where the vendor may handle remote support and updates, the contractor's IT team may manage connectivity, and site teams may control day-to-day operation and physical access. Due to this fragmented governance, vendor vetting and contractual obligations are essential to clearly define roles, especially when scaling OT deployment across projects and sites.

Limitations

This study has several limitations. First, the proposed platform and data analysis flow were tested on a single case study involving contributions from 11 employees. More case studies are needed to further evaluate the platform across different scenarios. Second, although an open call was made to four departments, including IT, only one IT employee responded. The remaining employees were mostly from the Design department, suggesting lower cybersecurity familiarity than IT personnel. Moreover, the responses were shorter than expected, limiting their breadth. These aspects should be kept in mind while judging the potential of the proposed methodology in this study. Finally, the mapping of the recommendations to ISA/IEC 62443 FRs (Table 8) has not yet been reviewed by external cybersecurity experts. The results should be interpreted as decision-support insights derived from employee knowledge rather than verified statements about exploitable vulnerabilities or compliance.

Conclusions and Future Work

This paper presents a web-based crowdsourcing platform that supports construction cybersecurity decision-making by gathering human knowledge through free-text contributions and converting it into recommendations for management teams. The platform was implemented in collaboration with a UAE-based contractor, and the MULE MZ100 lift-assist robot was selected as the investigated technology. A reproducible data analysis pipeline produced four topics. These topics were synthesized into seven actionable recommendations using a local LLM. The recommendations were mapped to ISA/IEC 62443 FRs to support systematic interpretation. The methodology presented in this paper demonstrates a reproducible pathway from employee inputs to decision-support recommendations, addressing RQ1. Moreover, the mapping of recommendations, with six of seven ISA/IEC 62443 FRs captured, indicates alignment with OT security objectives, addressing RQ2, and highlights Resource Availability (FR7) as a gap to target in future case studies.

Future work will focus on addressing the limitations. First, the mapping of the identified recommendations to ISA/IEC 62443 FRs will be validated by cybersecurity experts. Second, the platform will be deployed across different construction OT scenarios, and a larger participant group with more advanced cybersecurity knowledge will be targeted. Finally, the platform will be expanded to include a retrieval-augmented generation (RAG) engine that leverages relevant cybersecurity documents (e.g., standards, guidelines) to support the identified crowdsourced recommendations.

Acknowledgments

This research was partially supported by the Center for Cyber Security at New York University Abu Dhabi (CCS-AD), funded by Tamkeen under the NYUAD Research Institute Award G1104. This research was partially supported by the Center for Interacting Urban Networks

(CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001. This research was partially supported by the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES), funded by Tamkeen under the NYUAD Research Institute Award CG013.

References

Alley-Cassetty, 2026. MULE MZ100. Alley-Cassetty. URL https://www.alley-cassetty.com/mulemz100 Accessed: Jan. 29, 2026.

Beretta, M., Frederiksen, L., Wallin, M., Kulikovskaja, V., 2023. Why and How Firms Implement Internal Crowdsourcing Platforms. IEEE Transactions on Engineering Management 70, 3036–3049. https://doi.org/10.1109/TEM.2020.3045118

Check Point, 2025. Global Cyber Threats September 2025. Check Point Blog. URL https://blog.checkpoint.com/security/global-cyber-threats-september-2025-attack-volumes-ease-slightly-but-genai-risks-intensify-as-ransomware-surges-46/ Accessed: Nov. 17, 2025.

Elliott, D., 2025. Cybercrime: Lessons learned from a $25m deepfake attack. World Economic Forum. URL https://www.weforum.org/stories/2025/02/deepfake-ai-cybercrime-arup/ Accessed: June 6, 2025.

García de Soto, B., Georgescu, A., Mantha, B.R.K., Turk, Ž., Maciel, A., 2020. Construction Cybersecurity and Critical Infrastructure Protection: Significance, Overlaps, and Proposed Action Plan. Preprints 2020. https://doi.org/10.20944/preprints202005.0213.v1

Giray, L., 2023. Prompt Engineering with ChatGPT: A Guide for Academic Writers. Annals of Biomedical Engineering 51, 2629–2633. https://doi.org/10.1007/s10439-023-03272-4

Hevner, A., March, S., Park, J., Ram, S., 2004. Design Science in Information Systems Research. Management Information Systems Quarterly 28.

Lipner, S.B., 2015. The Birth and Death of the Orange Book. IEEE Annals of the History of Computing 37, 19–31. https://doi.org/10.1109/MAHC.2015.27

Malhotra, A., Majchrzak, A., Kesebi, L., Looram, S., 2017. Developing Innovative Solutions Through Internal Crowdsourcing. MIT SMR.

Malladi, S.S., Subramanian, H.C., 2020. Bug Bounty Programs for Cybersecurity: Practices, Issues, and Recommendations. IEEE Software 37, 31–39. https://doi.org/10.1109/MS.2018.2880508

Mantha, B.R.K., Sonkor, M.S., García de Soto, B., 2024. Investigation of the cyber vulnerabilities of construction networks using an agent-based model. Developments in the Built Environment 18, 100452. https://doi.org/10.1016/j.dibe.2024.100452

Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., Xiong, J., Deng, W., Oh, O., de Vreede, G.-J., 2013. Conceptual Foundations of Crowdsourcing: A Review of IS Research, in: 46th Hawaii International Conference on System Sciences. pp. 579–588. https://doi.org/10.1109/HICSS.2013.143

QBE, 2024. Commercial Construction Risk Report. QBE.

Reimers, N., Gurevych, I., 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics.

Röder, M., Both, A., Hinneburg, A., 2015. Exploring the Space of Topic Coherence Measures, in: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. pp. 399–408. https://doi.org/10.1145/2684822.2685324

Salami Pargoo, N., Ilbeigi, M., 2023. A Scoping Review for Cybersecurity in the Construction Industry. Journal of Management in Engineering 39, 3122003. https://doi.org/10.1061/JMENEA.MEENG-5034

Sonkor, M.S., García de Soto, B., 2026. Data for the Construction Cybersecurity Crowdsourcing Platform. https://doi.org/10.7910/DVN/KNA5VS

Sonkor, M.S., García de Soto, B., 2021. Operational Technology on Construction Sites: A Review from the Cybersecurity Perspective. Journal of Construction Engineering and Management 147, 4021172. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002193

Sophos, 2025. The State of Ransomware 2025. Sophos.

Sun, N., Zhang, J., Gao, S., Zhang, L.Y., Camtepe, S., Xiang, Y., 2020. Data Analytics of Crowdsourced Resources for Cybersecurity Intelligence. Presented at the Lect. Notes Comput. Sci., pp. 3–21. https://doi.org/10.1007/978-3-030-65745-1_1

Tambwe, O.T., Aigbavboa, C.O., Akinradewo, O.I., Adekunle, P.A., 2025. Measures to Address Cyber-Attacks in Construction Project Data Management Processes: A Cybersecurity Perspective. IET Information Security 2025, 7398742. https://doi.org/10.1049/ise2/7398742

Ulbrich, H., Wedel, M., Dienel, H.-L., 2021. Introduction to Internal Crowdsourcing: Theoretical Foundations and Practical Applications. Springer International Publishing, Cham, pp. 1–14. https://doi.org/10.1007/978-3-030-52881-2_1

Yin, R.K., 2009. Case Study Research: Design and Methods. SAGE.

Zhao, L., Yu, X., Zhou, X., 2023. The impact of regulatory mechanisms on vulnerability disclosure behavior during crowdsourcing cybersecurity testing. Math. Biosci. Eng. 20, 19012–19039. https://doi.org/10.3934/mbe.2023841

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

TRUST LAYERS FOR AUTOMATED BRIDGE MANAGEMENT USING BLOCKCHAIN AND SMART CONTRACTS

Leonardo Zunino[1], Valentina Gatteschi[1], Marco Domaneschi[1], Joan Ramon Casas[2], and Valentina Villa[1]
[1]Politecnico di Torino, Turin, Italy
[2]Universitat Politècnica de Catalunya, Barcelona, Spain

Abstract

Bridge management is increasingly automated through continuous monitoring, AI-based diagnostics, and digitally coordinated interventions. Yet, current workflows rarely provide end-to-end verifiability: sensor datasets, off-chain computations, and resulting maintenance decisions are difficult to audit across multiple stakeholders. This paper proposes a trust-layered architecture that combines decentralized storage, verifiable computation, and blockchain governance to make monitoring-to-action pipelines tamper-evident and reproducible. The paper discusses how smart contracts can orchestrate inspection and maintenance playbooks, while computation integrity can be certified via Trusted Execution Environments or zero-knowledge proofs. Trade-offs in scalability, privacy, and compliance are analyzed, outlining practical deployment patterns for bridge lifecycles.

Introduction

The ongoing digital transformation of civil infrastructure is reshaping how bridges are monitored, assessed, and maintained over their service life. In recent years, Structural Health Monitoring (SHM) systems have evolved from isolated sensing deployments into data-centric platforms capable of continuously observing structural response under real traffic and environmental variability. In parallel, advances in signal processing and Artificial Intelligence (AI) have increased the ability to extract damage-sensitive features, detect anomalies, and support condition-based and predictive maintenance strategies (An et al., 2019; Malekloo et al., 2022; Zinno et al., 2022; Jiménez Rios et al., 2023). Nevertheless, despite the growing availability of monitoring data and diagnostic tools, a persistent gap remains between ”sensing” and ”action”: monitoring outcomes do not automatically translate into traceable engineering decisions or timely interventions.

This gap is not only technical; it is also organizational and procedural. Bridge management workflows involve multiple stakeholders (owners, operators, engineers, authorities, contractors, technology providers) and multiple data sources (sensors, inspections, numerical models, maintenance records). In this multi-actor setting, it becomes difficult to ensure that the data used for diagnosis have not been manipulated, the diagnostic pipeline has been executed on the declared inputs, the decision logic applied to diagnostic results is consistent and auditable, and subsequent actions and approvals remain accountable years after the event. Conventional centralized data management solutions struggle to provide long-term integrity, verifiable provenance, and non-repudiation, especially when information must be exchanged across institutional boundaries and justified retrospectively.

Blockchain technology has been proposed as an enabling layer for tamper-resistant data governance and transparent process automation. Its core properties, immutability, distributed consensus, and the ability to execute smart contracts, provide mechanisms to register evidence, log events, and enforce pre-defined rules without relying on a single trusted intermediary (Christidis and Devetsikiotis, 2016; Casino et al., 2019; Khan et al., 2021; Liu et al., 2024). In the Architecture, Engineering, and Construction domain, blockchain-based approaches have been explored for documentation, contractual automation, and lifecycle information management (Turk and Klinc, 2017; Plevris et al., 2022; Christodoulou and Kalergis, 2024; Zhang et al., 2022, 2023; Adel et al., 2023; Villa et al., 2025; Gioberti et al., 2024; Romani et al., 2025), while, in the SHM field, several studies have considered blockchain to timestamp measurements, certify alarms, or coordinate stakeholders (Xu et al., 2021; Gordan et al., 2022; Gigli et al., 2022; Mariniello et al., 2025). These contributions confirm the potential of blockchain to improve trust and transparency; however, they also reveal a critical limitation: most frameworks implicitly assume that the off-chain computations that produce alarms and indicators are correct and faithfully executed. In detail, blockchain records can certify that an alert was registered, but they do not, by themselves, prove that the alert was obtained by executing the declared algorithm on the declared dataset.

This paper addresses the above limitation by proposing a methodological framework for trust-aware automated bridge management that spans the full path from sensing to intervention. The framework integrates: off-chain monitoring and diagnostics, decentralized or content-addressed storage for scalable data management, blockchain and smart contracts for governance and audit trails, and optional mechanisms for certifying algorithm execution (e.g., Trusted Execution Environments or cryptographic proofs). The objective is not to replace engineering judgment with automation, but to embed verifiability into the digital workflow so that data, computation, decisions, and actions

can be independently checked and retrospectively reconstructed. In this sense, the contribution of the paper is primarily architectural and methodological: it defines a design space of "trust layers" that can be combined to achieve different levels of accountability, performance, and decentralization depending on stakeholder needs and operational constraints.

The remainder of the paper is organized as follows. The next section introduces the end-to-end methodological framework connecting SHM evidence to intervention governance. Then, the core section discusses alternative trust layers and digitalization levels, comparing technologies and architectural options through a structured trade-off analysis and a summary table. Finally, discussion and conclusions outline practical implications and directions for future work.

Methodological Framework: From Sensors to Intervention

The proposed framework follows a design-oriented research approach aimed at structuring a conceptual solution to a recognized gap in current bridge management practice, rather than validating a specific algorithm or implementation. Its development builds on both the existing literature on SHM, digital twins, and blockchain-based infrastructures, and the authors' ongoing research on SHM-driven bridge management.

The resulting framework can therefore be interpreted as a synthesis of established knowledge and research-driven insights, organized into a layered architecture that maps identified trust gaps to corresponding technological mechanisms. This section introduces the resulting end-to-end framework as a technology-agnostic reference process from sensing to intervention, clarifying where trust issues typically arise and how they can be progressively addressed.

The framework starts with sensing and data acquisition. A bridge is monitored through a network of structural and environmental sensors, often complemented by inspections and operational records. These sources produce heterogeneous data. Raw signals, processed time series, extracted features, and supporting documents are generated at different rates and with different formats. For this reason, data are usually managed off-chain using scalable storage. A key point is that data must remain uniquely identifiable over time. In practice, this is achieved by associating each data batch and each derived artefact with a cryptographic fingerprint, so that the same content can always be referenced and checked.

After acquisition, data move through a diagnostic pipeline. Typical steps include cleaning, synchronization, normalization, feature extraction, and anomaly detection. The pipeline produces indicators and alerts that summarize the structural condition. At this stage, automation becomes attractive, but also fragile. If the input data are not reliable, or if the computation is not transparent, then the alert can be disputed and the whole chain loses credibility. This is one of the main motivations for introducing trust layers: not only to protect the final output, but to make the pathway from data to results verifiable.

Diagnostic outputs are then interpreted within a management workflow. The goal is not to turn algorithms into decision-makers. The goal is to connect monitoring evidence to consistent actions. A practical approach is to adopt clear response policies that map alerts to intervention intents, such as enhanced monitoring, rapid assessment, or emergency safeguards. These policies can be prepared offline and updated when needed. When a new alert arrives, the workflow selects the appropriate response and starts the corresponding process. The process includes notifications, deadlines, responsibilities, and approvals. The technical assessment and the final engineering judgement remain off-chain and under human control, especially for safety-critical cases.

Finally, the response is executed and documented. Inspection reports, images, calculations, and contractor deliverables are produced. These artefacts are typically too large and too diverse to be stored on a ledger. However, the framework assumes that they are still linked to the workflow through immutable references. In this way, the system can preserve an auditable trail that connects monitoring evidence, diagnostic outputs, response activation, approvals, and completion proofs.

Overall, the framework provides a simple reference model that separates computation and storage from governance and accountability. It also makes explicit where trust must be reinforced: data provenance, algorithm execution, decision logic, and action traceability. In the next section, these weak points are addressed through a set of digital trust layers that can be combined depending on cost, performance, privacy, and governance requirements. An overview of the framework and the position of the different trust layers along the monitoring-to-intervention pipeline is shown in Figure 1.

Trust Assumptions and Threat Model

The increasing digitalization of bridge monitoring and management implicitly introduces a set of trust assumptions that are often left unspoken. In traditional SHM-based workflows, trust is typically placed in data providers, algorithm developers, system operators, and centralized data platforms. While this model may be sufficient in controlled or single-organization settings, it becomes fragile when monitoring results are used to justify high-impact decisions, involve multiple stakeholders, or must be audited retrospectively.

One of the main risks concerns data integrity. Sensor measurements, processed signals, and derived indicators are usually stored and handled off-chain, often within proprietary platforms. Without explicit integrity guarantees, data can be altered, selectively filtered, or replaced without leaving clear evidence. Even unintentional errors, such as data loss, version mismatches, or undocumented preprocessing changes, can undermine confidence in the moni-

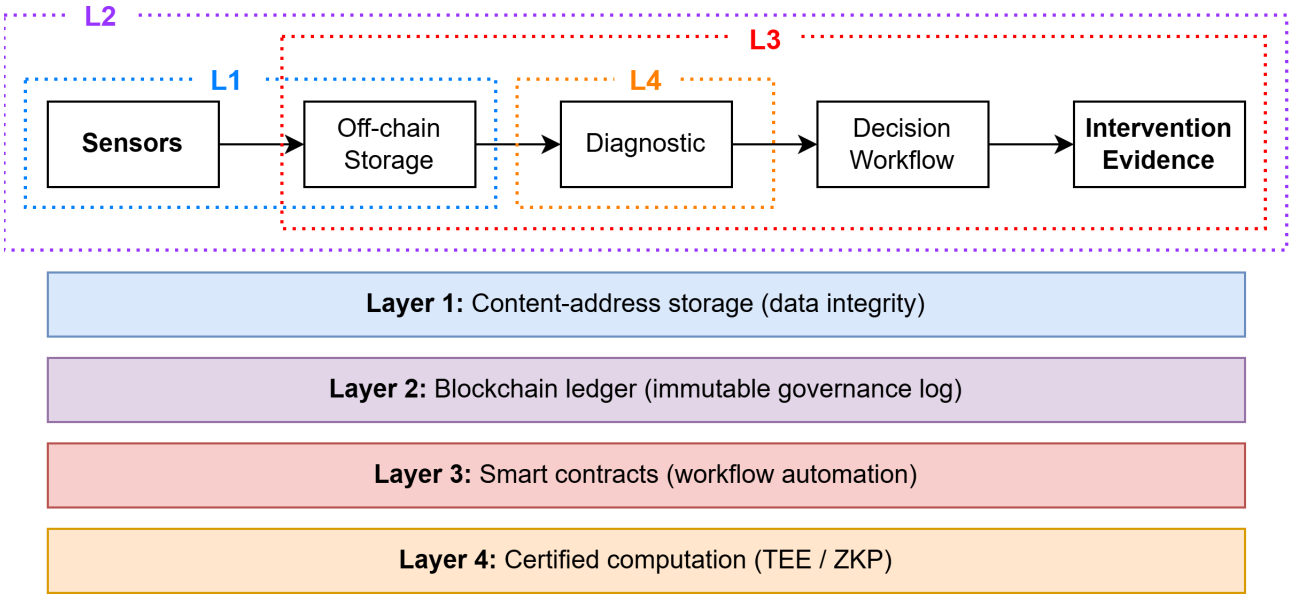


Figure 1: Monitoring-to-intervention workflow and associated trust layers.

toring results when decisions are questioned years later.

A second critical risk is related to opaque computation. Modern SHM pipelines increasingly rely on complex algorithms, including machine learning models and hybrid data-driven approaches. In many cases, the execution of these algorithms is not externally verifiable. Stakeholders must trust that the declared algorithm has been executed on the declared data, following the expected logic. When this assumption is violated, either by misconfiguration, software updates, or deliberate manipulation, the resulting alerts and indicators cannot be independently validated.

Disputes across stakeholders represent another relevant threat. Bridge management typically involves owners, operators, inspectors, engineers, contractors, and authorities, each with different responsibilities and incentives. When an alert triggers inspections, restrictions, or costly interventions, disagreements may arise regarding the correctness of the diagnosis, the timing of the alert, or the appropriateness of the response. In the absence of a shared and immutable audit trail, responsibility attribution becomes difficult and decisions may be contested.

Finally, retrospective accountability poses a long-term challenge. Bridge management decisions are often reviewed long after their execution, for example following incidents, legal disputes, or changes in ownership. At that point, personnel, software, and data platforms may have changed. Traditional monitoring systems rarely provide the guarantees needed to reconstruct the full chain from raw data to final action in a reliable and verifiable manner.

The trust layers introduced in the next section directly address these risks. Rather than assuming fully trusted actors or platforms, the proposed approach progressively reduces trust requirements by making data, computation, decision logic, and actions tamper-evident and auditable. Each layer targets specific weaknesses of traditional monitoring workflows and can be adopted independently or in combination, depending on the required assurance level.

Trust Layers and Digitalization Levels

Building on the reference workflow shown in Figure 1, this section discusses how different trust layers can be combined to strengthen specific stages of the process.

The central idea of the proposed approach is that bridge management digitalization can be equipped with a set of modular ”trust layers”, each providing additional guarantees in terms of integrity, transparency, and accountability. These layers should not be interpreted as mandatory components of a single fixed architecture, but as optional mechanisms that can be selectively adopted and progressively combined to achieve increasing levels of assurance.

In this sense, the framework is both modular and progressive: stakeholders may choose minimal configurations based on off-chain integrity and logging, or adopt more advanced layers (e.g., blockchain governance, smart contracts, or certified computation) depending on performance, privacy, interoperability, and governance requirements. This section introduces the resulting design space and discusses the associated trade-offs.

Layer 1: Content-addressed storage and integrity anchors (e.g., IPFS-like approach)

A first trust layer is achieved by separating data storage from trust in the storage provider. Content-addressed storage assigns a stable identifier to each dataset based on its content digest, enabling any party to verify that retrieved data match the declared identifier. This approach improves long-term traceability and supports distributed access while keeping heavy data off-chain. Limitations include data availability management (pinning/replication), access control design, and operational overhead when integrating with legacy data pipelines.

Layer 2: Blockchain ledger for immutable logs (public vs. consortium vs. private)

A blockchain ledger provides a tamper-resistant timeline of events, data anchors, and governance actions. From an architectural perspective, blockchain deployment can be interpreted along a single dimension with three main configurations: public, consortium, and private networks. Each configuration reflects a different trade-off between decentralization, performance, and governance.

- **Public blockchains** maximize transparency and third-party verifiability, but may raise concerns on transaction costs, throughput, and governance externalities.
- **Consortium / semi-private** networks balance decentralization and performance by restricting validators to identified organizations, improving scalability while preserving shared governance.
- **Private ledgers** provide high performance and administrative control but shift trust toward a single operator and may weaken independent verifiability.

In all cases, practical designs store only compact evidence on-chain (hashes, identifiers, proofs), while keeping datasets and detailed documents off-chain.

Layer 3: Smart contracts for automated and auditable workflows

Smart contracts implement transparent and deterministic rules that are executed on-chain and recorded immutably. In the proposed framework, they do not perform structural analysis or SHM computations, but act as a governance and orchestration layer connecting data, diagnostics, and management actions.

More specifically, different types of smart contracts can be introduced to support distinct functions along the workflow (Xu et al., 2021; Mariniello et al., 2025). These include:

- **Data validation and provenance contracts**, which register dataset identifiers and verify data integrity and quality before use;
- **Diagnostic certification contracts**, which record and validate the outputs of anomaly detection algorithms;
- **Response or workflow contracts**, which activate predefined management procedures based on alert conditions.

Within bridge management, these contracts enable the automation of governance logic, such as classification rules, response activation, deadlines and escalation paths, multi-signature approvals, and auditable milestone recording. For example, when an alert exceeds a predefined threshold, a smart contract may trigger a corresponding response playbook, notify responsible stakeholders, enforce time constraints, and require formal approvals before progressing to subsequent stages.

Importantly, the framework advocates *automation with oversight*: smart contracts formalize and enforce the decision process, but do not replace engineering judgement. All computationally intensive tasks and technical assessments remain off-chain, while smart contracts ensure that actions are consistent, traceable, and accountable across stakeholders.

Layer 4: Certified algorithm execution (TEE and ZKP as alternatives)

Even with data anchors and on-chain governance, a major trust gap remains if diagnostic computation is performed off-chain without verifiable execution. To reduce this gap, the framework introduces optional mechanisms to certify that a declared algorithm was executed on declared data:

- **Trusted Execution Environments (TEE)** provide hardware-isolated execution and remote attestation reports. They offer near-native performance and practical deployability, at the price of relying on hardware vendors and being potentially exposed to implementation and side-channel risks (Sabt et al., 2015).
- **Zero-Knowledge Proofs (ZKP)** can, in principle, provide cryptographic evidence that a computation was executed correctly with respect to its formal semantics, potentially without revealing proprietary details. They offer strong verifiability but may introduce substantial computational overhead and engineering complexity, especially for floating-point intensive pipelines and large datasets (Sun et al., 2021).

In this paper, these mechanisms are treated as part of a design space rather than a mandatory requirement: stakeholders may choose different layers based on the required assurance level and operational constraints.

Summary and trade-off table

The layered structure presented in this section directly reflects the methodological approach described above, where each trust layer corresponds to a specific class of trust gaps identified in the monitoring-to-intervention workflow. Table 1 summarizes the role of each technology and highlights typical advantages and limitations. The table is intended as a practical decision aid when assembling a trust-aware bridge management architecture.

Discussion

The proposed framework frames bridge management as a verifiable digital workflow rather than as a collection of disconnected tools. A key implication is that trust is not a binary property, but a configurable design choice. Different stakeholders and decision contexts require different assurance levels, and the framework allows these to be addressed without enforcing a single technological solution. For example, routine monitoring activities within a single organization may only require basic data integrity and logging mechanisms. In contrast, safety-critical decisions, contractual disputes, or regulatory reporting may benefit from stronger guarantees, such as immutable governance logs or certified computation. By separating the functional workflow from the trust layers, the framework supports incremental adoption and avoids unnecessary overhead where it is not justified.

The discussion also highlights a fundamental tension be-

Table 1: Trust layers and digitalization options: role, benefits, and limitations.

Layer / Technology	Primary contribution	Main limitations / trade-offs
Content-addressed storage (e.g., IPFS-like)	Integrity-by-design through content hashing; scalable off-chain storage; persistent references via content identifiers; facilitates distributed access and reproducible evidence chains.	Data availability management (pinning/replication) required; access control and confidentiality must be handled outside the storage layer; integration effort with existing pipelines.
Blockchain ledger (public)	Strong immutability and broad third-party verifiability; ideal for public audit trails and long-term non-repudiation.	Higher transaction costs and limited throughput; potential privacy concerns for metadata; governance and upgradeability depend on public ecosystem.
Blockchain ledger (consortium / semi-private)	Shared governance among identified participants; improved performance and easier alignment with regulatory/compliance constraints.	Less decentralization than public chains; requires formal consortium arrangements and operational coordination among validators.
Blockchain ledger (private)	High throughput and administrative control; straightforward integration inside a single organisation.	Trust is concentrated in the operator(s); weaker independent verifiability and reduced external auditability.
Smart contracts	Deterministic, auditable workflow logic: automate notifications, deadlines, escalation, multi-signature approvals and immutable event logs.	Correctness/bugs and upgrade governance are critical; legal and organisational alignment needed; smart contracts should orchestrate processes, not replace engineering judgement.
TEE-based execution certification	Hardware-isolated execution with remote attestation; near-native runtime performance; suitable for complex, floating-point-intensive pipelines.	Relies on hardware trust (vendor certificates); vulnerable to side-channel attacks and implementation flaws; attestation proves platform+binary, not full semantic correctness of the algorithm.
ZKP-based computation proofs	Cryptographic binding of input, computation and output; strong mathematical verifiability and optional privacy-preserving proofs.	Potentially very high proof-generation cost for large or floating-point workloads; engineering complexity and immature tooling for some classes of SHM algorithms.

tween transparency and operational constraints. While strong auditability improves accountability, it may conflict with privacy requirements, proprietary algorithms, or regulatory restrictions on data sharing. The proposed architecture mitigates this tension by relying on off-chain storage and selective on-chain anchoring, but careful system design remains essential to prevent sensitive information leakage while preserving verifiability.

Another important aspect concerns the role of automation. Although smart contracts enable deterministic and auditable workflows, they are not intended to replace engineering judgement. In the proposed framework, automation is used to enforce procedures, deadlines, and traceability, while technical assessments and final decisions remain under human control. This distinction is particularly important in bridge management, where safety considerations require conservative and accountable decision-making.

Finally, the comparison between Trusted Execution Environments and Zero-Knowledge Proofs illustrates that certified computation is technically feasible, but context-dependent. Hardware-based approaches currently offer a pragmatic balance between performance and assurance for complex SHM pipelines, while cryptographic proofs provide stronger theoretical guarantees at the cost of higher computational effort. The framework accommodates both approaches, allowing stakeholders to select the most appropriate option for their use case.

Conclusions

This paper proposed a trust-aware methodological framework for automated bridge management that spans the full path from sensing to intervention. The approach integrates scalable off-chain data handling with blockchain-based governance and introduces a layered view of digitalization, where additional "trust layers" can be progressively added to improve integrity, transparency, and accountability.

The core contribution of this paper is the definition of a structured design space and a comparative analysis of the main technologies that enable trust-aware digital workflows. These include content-addressed storage for data integrity, different blockchain deployment models, smart contracts for workflow governance, and mechanisms for certifying algorithm execution such as Trusted Execution Environments and Zero-Knowledge Proofs. By combining these components appropriately, bridge management systems can embed verifiability into digital processes while remaining adaptable to performance, privacy, and organizational constraints.

Future work will focus on several complementary directions. A first line of development concerns the extension of trust layers to additional stages of the infrastructure lifecycle, including sensor-level validation, inspection data certification, and the integration of digital twin updates within verifiable workflows.

A second research direction involves hybrid verification strategies. As zero-knowledge technologies mature and

hardware-based trusted execution evolves, combinations of TEE and cryptographic proofs may reduce trust assumptions while maintaining acceptable performance for data-intensive SHM applications. Evaluating these hybrid approaches in realistic operational settings remains an open challenge.

Further work is also needed to address interoperability and standardization. Integrating trust-aware workflows with existing monitoring platforms, asset management systems, and regulatory reporting processes requires well-defined interfaces and shared data models. Progress in this direction is essential to enable adoption beyond isolated pilots.

Finally, scaling the proposed framework from single assets to bridge networks represents an important step. Network-level deployment would enable coordinated risk assessment, prioritization of interventions, and transparent allocation of resources across multiple infrastructures, aligning the framework with the needs of infrastructure owners and public authorities.

Acknowledgments

Giorgio Marengo contributed to the conceptual development of this work in partial fulfilment of the requirements for his MSc degree in Industrial Production and Technological Innovation Engineering at Politecnico di Torino, under the supervision of the authors. His contribution is gratefully acknowledged.

This work was supported by MIUR - PRIN 2022 "BIO-RESTORE" [Prot. 202234HM8J, CUP E53C24002680006]; and it is part of the project "National Recovery and Resilience Plan-Next Generation EU" (PNRR-NGEU) which has received funding from the Italian Ministry of University and Research (MUR) – DM 629/2024.

References

Adel, K., Elhakeem, A., and Marzouk, M. (2023). Decentralized system for construction projects data management using blockchain and ipfs. Journal of Civil Engineering and Management, 29(4):342–359.

An, Y., Chatzi, E., Sim, S.-H., Laflamme, S., Blachowski, B., and Ou, J. (2019). Recent progress and future trends on damage identification methods for bridge structures. Structural Control and Health Monitoring, 26(10):e2416.

Casino, F., Dasaklis, T. K., and Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues. Telematics and informatics, 36:55–81.

Christidis, K. and Devetsikiotis, M. (2016). Blockchains and smart contracts for the internet of things. IEEE access, 4:2292–2303.

Christodoulou, D. and Kalergis, D. (2024). Leveraging blockchain for smart city resilience: Strengthening prevention and civil protection systems. Journal of Sustainable Development, Culture, Tradition, 3.

Gigli, L., Sciullo, L., Montori, F., Marzani, A., and Di Felice, M. (2022). Blockchain and web of things for structural health monitoring applications: A proof of concept. In 2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC), pages 699–702. IEEE.

Gioberti, L., Domaneschi, M., Villa, V. B., Chiaia, B. M., and Catbas, F. N. (2024). Blockchain technology conception in digital data collection for damage assessment and l.c.a. pages 1575 – 1582.

Gordan, M., Siow, P. Y., Deifalla, A. F., Ong, Z. C., Ismail, Z., and Yee, K. S. (2022). Implementation of a secure storage using blockchain for pca-frf sensor data of plate-like structures. Ieee Access, 10:84837–84852.

Jiménez Rios, A., Plevris, V., and Nogal, M. (2023). Bridge management through digital twin-based anomaly detection systems: A systematic review. Frontiers in Built Environment, 9:1176621.

Khan, S. N., Loukil, F., Ghedira-Guegan, C., Benkhelifa, E., and Bani-Hani, A. (2021). Blockchain smart contracts: Applications, challenges, and future trends. Peer-to-peer Networking and Applications, 14(5):2901–2925.

Liu, Y., He, J., Li, X., Chen, J., Liu, X., Peng, S., and Wang, Y. (2024). An overview of blockchain smart contract execution mechanism. Journal of Industrial Information Integration, 41:100674.

Malekloo, A., Ozer, E., AlHamaydeh, M., and Girolami, M. (2022). Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. Structural Health Monitoring, 21(4):1906–1955.

Mariniello, G., Gragnaniello, C., and Asprone, D. (2025). Enhancing structural health monitoring data management and damage detection in digital twins with blockchain and smart contracts. Automation in Construction, 180:106558.

Plevris, V., Lagaros, N. D., and Zeytinci, A. (2022). Blockchain in civil engineering, architecture and construction industry: state of the art, evolution, challenges and opportunities. Frontiers in Built Environment, 8:840303.

Romani, N., de Caria, R., Restori, F., Domaneschi, M., Gatteschi, V., and Villa, V. B. (2025). A blockchain-based platform for real estate management. pages 534 – 542.

Sabt, M., Achemlal, M., and Bouabdallah, A. (2015). Trusted execution environment: What it is, and what it is not. In 2015 IEEE Trustcom/BigDataSE/Ispa, volume 1, pages 57–64. IEEE.

Sun, X., Yu, F. R., Zhang, P., Sun, Z., Xie, W., and Peng, X. (2021). A survey on zero-knowledge proof in blockchain. IEEE network, 35(4):198–205.

Turk, Ž. and Klinc, R. (2017). Potentials of blockchain technology for construction management. Procedia Engineering, 196:638–645.

Villa, V. B., Gioberti, L., Domaneschi, M., and Catbas, N. (2025). Conceptual advancements in infrastructure maintenance and management using smart contracts: Reducing costs and improving resilience. Buildings, 15(5).

Xu, J., Liu, H., and Han, Q. (2021). Blockchain technology and smart contract for civil structural health monitoring system. Computer-Aided Civil and Infrastructure Engineering, 36(10):1288–1305.

Zhang, X., Liu, T., Rahman, A., and Zhou, L. (2023). Blockchain applications for construction contract management: A systematic literature review. Journal of Construction Engineering and Management, 149(1):03122011.

Zhang, Y., Wang, T., and Yuen, K. V. (2022). Construction site information decentralized management using blockchain and smart contracts. Computer-Aided Civil and Infrastructure Engineering, 37(11):1450–1467.

Zinno, R., Haghshenas, S. S., Guido, G., and VItale, A. (2022). Artificial intelligence and structural health monitoring of bridges: A review of the state-of-the-art. IEEE Access, 10:88058–88078.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

USAGE POLICIES FOR FRAGMENTED DIGITAL PRODUCT PASSPORT DATA RETRIEVAL IN AECO DATA SPACES

Krista van Zandwijk, Ekaterina Petrova, Rik Eshuis and Pieter Pauwels
Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract

Decentralized digital product passports (DPPs) retrieval via data spaces are emerging as a promising approach to enable material reuse in the Architecture, Engineering, Construction, and Operations (AECO) industry. Yet, no DPP-specific operational data space currently exists, and usage policy implementation remains unexplored. Building on Eclipse, this paper develops a proof-of-concept DPP data space, implementing asset registration, usage policy definition, contract negotiation, and authenticated data retrieval. Demonstrated through a suspended ceiling element, the research shows how multiple data providers can offer modular DPP data under differentiated usage policies, offering the first technical validation for decentralized DPP exchange in the AECO industry.

Introduction

To stimulate construction and demolition waste (CDW) reuse, European policies, such as the Construction Products Regulation (CPR) and the Ecodesign for Sustainable Products Regulation (ESPR), mandate the development of Digital Product Passports (DPPs) (European Parliament and European Union Council, 2024b; European Commission, 2025). DPPs are defined as *'a set of data specific to a product that includes information specified in the applicable delegated act and that is accessible via electronic means through a data carrier'* (European Parliament and European Union Council, 2024a, p.28) and aim to support the reuse of CDW by enabling more informed decision-making throughout the value chain (Kovacic et al., 2024). However, their current implementations fall short. In the absence of a common data standard, DPPs are developed as client-specific, non-interoperable formats, and are typically managed as static documents, resulting in incomplete and quickly outdated DPPs (Zhang et al., 2026).

Addressing these limitations requires DPPs to function as dynamic records that continuously reflect up-to-date lifecycle information (Çetin et al., 2023). This, in turn, calls for a shift from centralized data management toward decentralized frameworks, where data remains with the stakeholders who generate and maintain it (Bucher et al., 2025). A promising solution is the data space framework as proposed by the International Data Spaces Association (IDSA), which is *'a data integration concept which does not require common database schemas and physical data integration, but is rather based on distributed data stores and integration on an "as needed" basis on a semantic level'* (Nagel and Lycklama, 2021, p. 7). Data spaces enable federated, secure, and interoperable data retrieval while preserving data sovereignty (International Data Space Association, 2024).

A central mechanism for ensuring sovereignty in data spaces is the specification of usage policies, which define how, by whom, and under what conditions data may be used. Steinbuss et al. (2021, p.11) define usage control as *'specification and enforcement of restrictions regulating what must (not) happen to data'*. Usage control extends traditional access control by enforcing these policies continuously, regulating how data can be processed, aggregated, or shared even after access has been granted (Steinbuss et al., 2021). For DPPs, this would, for example, enable stakeholders to disclose only the information necessary for circular practices, such as material composition, demountability instructions, or maintenance history, while restricting the use of commercially sensitive data, preventing unauthorized redistribution.

However, no dedicated data space exists for DPP creation and management in the Architecture, Engineering, Construction, and Operations (AECO) industry, and consequently, no usage policies have been established to govern DPP data retrieval. Prior research has not investigated the implementation of these policies, leaving it unclear how an end-to-end technical pipeline can operationalize usage policies to allow multiple data providers with differing data types and related protection needs to share DPP data securely and effectively.

To address this, this paper explores the first domain-specific investigation of usage policy formulation for AECO DPP data retrieval. It demonstrates a proof-of-concept DPP data space that implements the full technical pipeline, including asset registration, access and usage policy definition, contract negotiation, and authenticated data retrieval. The approach is validated in a real construction case, where three distinct data providers, each with differentiated and stakeholder-motivated usage policies, offer their data and one data consumer retrieves it.

The remainder of the paper is structured as follows. The next Section presents a systematic literature review following the PRISMA methodology (Page et al., 2021), covering related work on data spaces, usage policies, and DPPs. The Methodology Section then describes the development

of the end-to-end technical pipeline. The prototype is tested in a real construction environment, illustrating the data offering and retrieval process for a suspended ceiling component from a demonstrator project in the Netherlands, with the results presented in the Results Section. Finally, the last Section concludes with a discussion of the findings and outlines directions for future research.

Related Work

Data Space Architecture

The International Data Space (IDS) Reference Architecture Model (RAM) 4, developed by the International Data Spaces Association (IDSA), is the most widely adopted system architecture to develop data spaces (Gabellini et al., 2025). As shown in Figure 1, each data space participant has a verified digital identity and an IDS Connector, which form the core of the data space. The verified digital identity, authenticated and authorized by the identity provider, ensures trust and accountability, while the connector acts as a secure gateway, enabling controlled data access while keeping data stored at its source. Data providers define usage policies that specify how their data may be accessed and used. The data space itself does not store the actual data; it only exchanges metadata and enforces policies, while the real data flows directly between connectors once all requirements are met (International Data Space Association, 2024). Beyond these core components, the IDS framework includes several federated services that support data space operations. A Broker enables the publication and discovery of dataset metadata, while the Vocabulary Hub ensures semantic interoperability through shared ontologies. The App Store offers certified applications for data processing, transforming, or integration, and a Clearing House ensures accountability by monitoring, logging, and verifying data-sharing activities.

Few studies have examined the application of data spaces in the AECO industry. Cuno et al. (2019) proposed an urban data space model based on a German municipal case, while Gil et al. (2024) developed a regulatory framework that uses a building data space to automate and digitize building permit processes. Demonstrating its applicability in circular construction, Hagedorn et al. (2025) validated the integration of data space principles with Information container for linked document delivery (ICDD). Complementing these technical contributions, van Zandwijk et al. (2025) examined how data spaces should be designed at the User Interface (UI) level to meet user needs and align with existing AECO workflows.

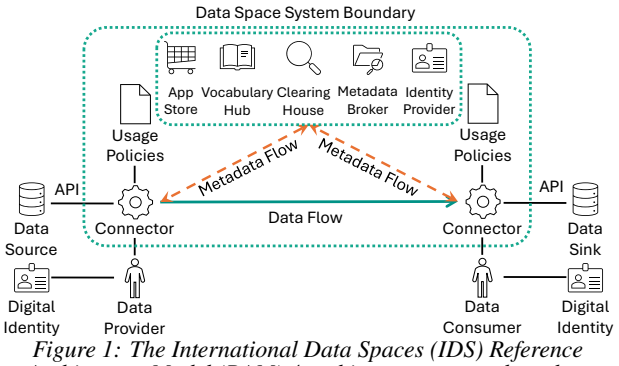


Figure 1: The International Data Spaces (IDS) Reference Architecture Model (RAM) 4 and its components, based on International Data Space Association (2024).

Usage Policies

Usage policies and contract negotiation lie at the core of data offers and retrieval within a data space. When offering data, data providers attach a usage policy specifying what data consumers are allowed to do with the data once they have authorized access. In IDS, these policies are described in JSON-based contracts with a fixed structure, consisting of contract metadata and a set of usage control policy rules that define permissions, prohibitions, and obligations for data usage, as shown in Figure 2. Steinbuss et al. (2021) propose 21 classes of usage control rules, which can be combined to create tailored policies covering a wide range of mechanisms, including constraints on users, systems, locations, purposes, time, or security level, as well as rules for modification, logging, notifications, redistribution, encryption, and contractual conditions.

The usage control policy can be specified using the IDS Usage Control Language (UCL), a technology-independent, machine-readable language based on W3C's Open Digital Rights Language (ODRL) and represented in JSON-LD (Steinbuss et al., 2021). While ODRL can express usage rules, it lacks built-in enforcement capabilities, requiring external systems or additional machine-readable policy languages to implement and monitor compliance (Petrova-Antonova et al., 2026). Although enforcement is essential for a trusted data space where usage policies are effective, it is beyond the scope of this study and remains an important topic for future research. Existing research on usage control in data spaces focuses predominantly on the technical aspects of usage policy creation, negotiation, and enforcement. For example, Yang et al. (2022) propose a negotiation scheme for IDS usage policy generation, while Munoz-Arcentales et al. (2020) implement and validate a FIWARE-based architecture that enforces usage control across multi-organization environments. Similarly, Dam et al. (2023) formalize domain-specific usage-control patterns in ODRL for mobility data spaces.

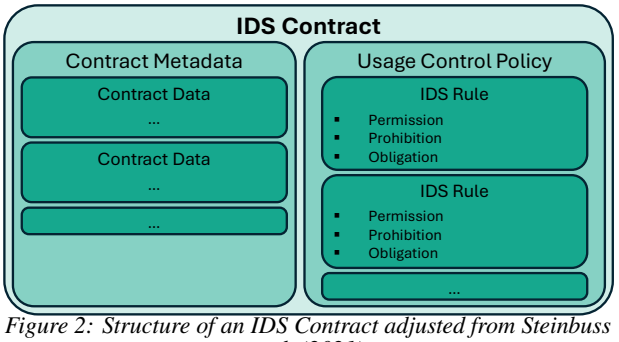


Figure 2: Structure of an IDS Contract adjusted from Steinbuss et al. (2021).

However, to the best of the authors' knowledge, no prior work investigates the implementation and negotiation of usage policies for DPP data retrieval in the AECO industry. AECO data is often complex, multi-stakeholder related, and commercially sensitive, making dedicated data spaces and tailored usage policies particularly important for this domain. The absence of domain-specific studies shows the need to explore how usage policies can support trustworthy, compliant, and efficient DPP data retrieval in circular construction contexts.

Digital Passports

Digital passports, first mentioned in 1982, gained momentum after the 2017 Buildings as Material Banks (BAMB) report (van Capelleveen et al., 2023), with the ESPR and revised CPR establishing EU-wide frameworks and registries for construction products and materials. Despite this progress, the implementation of DPPs is still evolving due to two key challenges. First, the absence of a fully standardized regulatory framework creates uncertainty about what information a DPP should contain. Second, there is no standardized digital infrastructure to support the generation, retrieval, and maintenance of DPP data across stakeholders.

Several studies have investigated the information requirements for digital passports. For example, Çetin et al. (2023) investigate data needs and actual data availability in the existing building stock, identifying eight relevant categories, including building general information, product general information, product properties, product safety, health and environment, product operational aspects, and product end-of-life aspects. Similarly, Akbarieh (2023) conducts a literature review of 146 publications and groups information needs into six categories, ranging from general information to properties requirements for safe transformation, intervention, circular information, to structural considerations. Complementing these perspectives, Byers et al. (2024b) propose a taxonomy for information acquisition to support material reuse, distinguishing between material reuse actions, information properties, and information sources.

Next to that, a variety of infrastructures have been proposed to enable DPP data exchange and retrieval. Based on a cross-domain review of 186 publications, Zhang et al. (2026) identify Distributed Ledger Technologies (DLT), ontology and knowledge-graph–based approaches, Digital Twins, and sector-specific systems such as BIM and Digital Building Logbooks (DBLs) as the most common technologies used in the construction industry. Centralized platforms, as discussed in Kovacic et al. (2024), offer structured, updatable repositories but compromise data sovereignty and frequently yield incomplete datasets due to interoperability limitations between these platforms. DLT-based systems, as proposed by Byers et al. (2024a), support decentralized trust and immutable provenance, yet they do not provide fine-grained, dynamic, and enforceable usage policies without additional off-chain mechanisms (Rohrschneider et al., 2026). Linked Data and ontology-driven frameworks, such as the work of Kebede et al. (2024), offer machine-readable semantics and flexible querying, but similarly lack mechanisms for enforcing peer-to-peer usage rules during cross-organizational exchanges (Zhang et al., 2026). Emerging work on dataspace based DPPs by Gallina et al. (2025) and Belova et al. (2025) illustrate the potential of data spaces to overcome these limitations. Gallina et al. (2025) developed a data space prototype for dynamic DPP management in an industrial case of a furniture dealer. They explores various possibilities enabled by data spaces and ontologies while identifying several challenges, such as missing standards and procedures concerning trust and data usage control. Belova et al. (2025) propose an initial architecture of a DPP system using data spaces and Knowledge Graphs, using carbon foot printing in the wood processing industry as context. Their focus lie in the proposal of a conceptual architecture, and the specific usage policies are outside the scope. Zhang et al. (2026) identify a persistent privacy–security trade-off in current DPP implementations, recognizing that additional technologies are often needed to achieve the needed privacy and security levels. However, this typically address only secure access control, not usage control. Data spaces address this gap by providing both fine-grained access control and enforceable usage control.

Methodology

To explore how usage policies can support DPP data retrieval in an AECO data space, this research simulates the retrieval of DPP information from a real-life demonstrator project using a suspended ceiling element. The simulation evaluates how modular DPPs can be shared while respecting data providers' data sharing constraints. This Section introduces the case study, the modular DPP structure, the data space prototype, and the simulation setup.

Case Introduction

The research builds on the Foppingadreef project, a renovation project of a Dutch bank's headquarters. Originally constructed in 1985, the building underwent renovation and expansion in 2022, unconsciously aligning with the three step process for reuse proposed by Bellini et al. (2024), which consists of collecting information, performing information-driven evaluations, and planning for reuse. Although the renovation comprises thousands of construction elements, this study specifically focuses on the reuse of a suspended ceiling. At the start of the renovation, the building contained an estimated 82,665 ceiling panels, which were in good condition and easily demountable. During the renovation, one-third of these panels were reused on-site, while the remaining two-thirds were repurposed in another building. This requires DPP data sharing in a complex multi-actor environment where well-defined usage policies are essential to enable secure and dynamic DPP data retrieval.

DPP creation

Since the Foppingadreef was built in 1985, no DPPs or BIM models existed at the start of renovation, an issue also identified by Çetin et al. (2023). Therefore, a material inventory was created listing all visible building elements and their properties, including suspended ceilings. For these elements, the collected data was converted into a modular DPP using the networked DPP ontology by Jansen et al. (2024). Modular DPPs are necessary to achieve scalable and cost-effective DPP solutions in complex, multi-actor environments, such as the AECO industry (Zhang et al., 2026).

The DPP ontology network developed by Jansen et al. (2024) is sector-agnostic, meaning it is not tailored specifically to construction materials and may therefore be too generic to capture sector-specific nuances. Nevertheless, Jansen et al. (2024) aligned their ontology with the requirements of the ESPR, ensuring compliance with emerging regulatory requirements. Another suitable ontology is the modular DPP ontology proposed by Kebede et al. (2024), which is explicitly designed for products in the built environment. However, as this ontology is not published online, the present study adopts the published ontology of Jansen et al. (2024).

Although this study adopts the DPP ontology network for demonstration, data spaces support heterogeneous data models, as long as each dataset is accompanied by the appropriate documentation. In practice, various data standards can coexist within the same data space. The DPP ontology network is used in this research to illustrate how a complete end-to-end pipeline operates, but alternative data models or ontologies can also be used. Nevertheless, mapping different ontologies for an interoperable, scalable DPP system is beyond the scope of this research.

The DPP ontology network comprises five interconnected parts; The *DPP Ontology Design Pattern (ODP)*, which introduces the core concepts *Product* and *DPP* and their part–whole relations; The *DPP Core*, which refines the *Product* concept into subclasses such as components, materials, and substances in line with EU directives like the ESPR; The *DPP Information* module, which models the various types of information a DPP may contain through reified information objects, such as characteristics, product quality, and certification data; The *Product Composition* module, which represents detailed part–whole relationships, including substances of concern; And finally, the *DPP Provenance* module, which supports tracking of actors and temporal validity associated with the DPP data.

To create this modular DPP, the data from the material inventory is supplemented with mock-up data, as the inventory only contains information obtainable through visual inspection and therefore lacks essential elements such as certificates, repair instructions, and fire safety documentation. Since this research focuses on retrieving DPP data via usage policies rather than creating a complete and fully accurate DPP, the enriched, partially synthetic DPP is considered sufficiently robust to demonstrate and validate the proposed concept.

Data Space Environment

A small-scale data space is created using the Eclipse Minimum Data Space (MVDS) framework[1]. The MVDS code enables the creation of a small-scale demo data space compliant with current standards, including the IDS Dataspace Protocol and Gaia-X–aligned trust and identity mechanisms. The Eclipse MVDS is chosen over IDSA's MVDS[2] as it represents the state-of-the-art in open-source data space implementations (Gabellini et al., 2025).

As shown in Figure 3, the MVDS consists of only the core components required to establish a functional data space: a connector and an identity hub for both data providers and consumers. Each connector includes an internal catalog that describes the metadata of the provided data and the associated policies. The provider's connector has two catalogs: a private catalog, accessible only to the data provider, and a shared catalog, which is accessible to other participants who meet the access policies defined for the catalog. The identity hubs use decentralized identifiers (DIDs), but in this demo, identities and credentials are pre-created and manually seeded. The dynamic issuance of verifiable credentials by trusted authorities lies beyond the scope of this study.

The MVDS code is run locally on Linux via IntelliJ, which provides configurations for consumer and provider connectors, identity hubs, the provider catalog server, and a compound 'dataspace' setup to start all runtimes simultaneously. A NGINX instance hosts the issuer's DID document, with port mapping aligned to the expected DID. Once the runtimes are running, the data space is initialized using the provided seed.sh script to set up the Participant Contexts and enable connector-to-connector communication. REST API interactions can then be executed via the included Postman collection.

Scenario simulation & Policy formulation

To demonstrate the full technical pipeline, the research build upon a scenario in which multiple data providers offer different parts of a modular DPP with a single data consumer, conceptually shown in Figure 4. Three actors

[1] https://github.com/eclipse-edc/MinimumViableDataspace

[2] https://github.com/International-Data-Spaces-Association/IDS-testbed

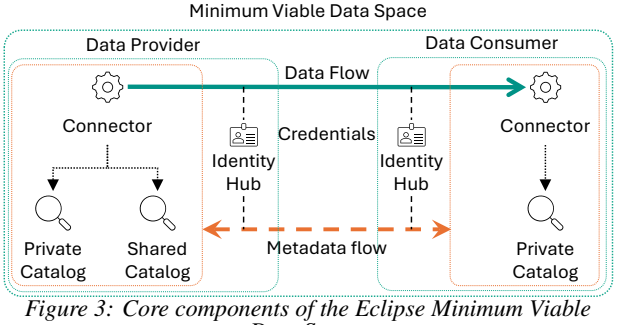


Figure 3: Core components of the Eclipse Minimum Viable Data Space.

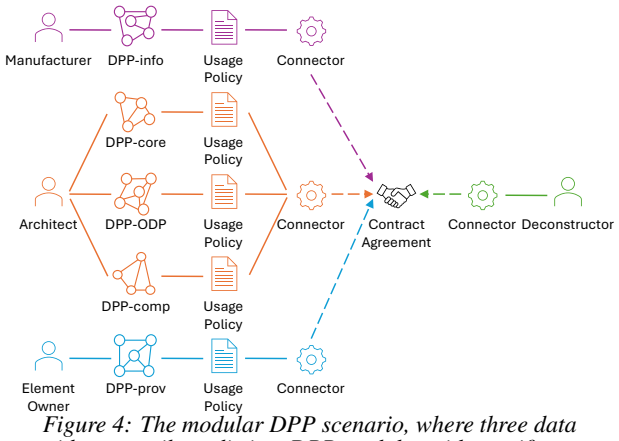


Figure 4: The modular DPP scenario, where three data providers contribute distinct DPP modules with specific usage policies, all accessed by a single data consumer.

act as data providers: the manufacturer, the architect, and the element owner. Each contributes specific DPP modules: the manufacturer supplies DPP-info; the architect provides DPP-core, DPP-ODP, and DPP-comp, and the element owner supplies DPP-prov. Verstraete et al. (2025) argues that data providers are reluctant to share data with other stakeholders, and Dam et al. (2023) argues that different data providers may have different data sharing constraints depending on their role and data offer. Hence, each data offer has its own sharing constraints and therefore usage policy. The manufacturer protects competitive information and limits liability by allowing DPP-info access only for safety, planning, and compliance during the deconstruction project. The architect safeguards intellectual property by permitting read-only access while prohibiting modification, redistribution, and printing. The element owner prioritizes privacy, requiring one-time read access, mandatory deletion within 24 hours, and notifications upon access. The usage policy for the architect is shown in Listing 1. This IDS contract permits the data consumer to read and display the specified dataset, while prohibiting modification, distribution, or printing.

These policies and constraints are hypothetical mock-up scenarios designed to demonstrate the proof of concept. However, the specific differences in usage policies per stakeholder group and data type for DPP data in the AECO industry is unexplored. Therefore, this research builds upon these fictitious scenarios to illustrate the modular DPP approach, but future work should explore specific differences in usage policies per stakeholder groups and data types.

Listing 1 Example of an IDS Usage Policy describing permitted and prohibited actions for using the DPP-core asset.

```
{  "@context": [
     "https://w3id.org/edc/connector/management/v0.0.1"
   ],
   "@type": "ContractRequest",
   "counterPartyAddress": "{{PROVIDER_DSP_URL}}/api/dsp",
   "counterPartyId": "{{PROVIDER_ID}}",
   "protocol": "dataspace-protocol-http",
   "policy": {
     "@type": "Offer",
     "@id": "{{POLICY_ID_DPP-core}}",
     "assigner": "{{PROVIDER_ID}}",
     "permission": [
       {"action": "use",
         "constraint": [
           {"leftOperand": "idsc:ALLOWED_ACTION",
             "operator": "eq",
             "rightOperand": ["idsc:READ", "idsc:DISPLAY"] }
         ]
       }],
     "prohibition": [
       {"action": "use",
         "constraint": [
           { "leftOperand": "idsc:PROHIBITED_ACTION",
             "operator": "eq",
             "rightOperand": ["idsc:MODIFY", "idsc:DISTRIBUTE", "idsc:PRINT"] }
         ]
       }],
     "obligation": [],
     "target": "DPP-core"},
   "callbackAddresses": []
}
```

Results

The Eclipse MVDS provides a small-scale data space where participants can offer and retrieve data through a six-step process, as illustrated in Figure 5. This Section describes each step of this process in detail. Sequence diagrams for all steps are available on GitHub[3], but are not shown here due to space constraints.

Creating the Asset and the Usage Policies

First, the data offer must be registered in the connector by the data provider. This registration should include a short description of the dataset and the base URL that can be used to access the data. Next, several policies need to be defined. The first is an access policy, which ensures that only participants with an active membership credential are allowed to use the asset. The second is a data usage policy, which specifies how the data can be processed and allows the data provider to enforce different usage rules

[3]https://github.com/KristavanZandwijk/MVDS

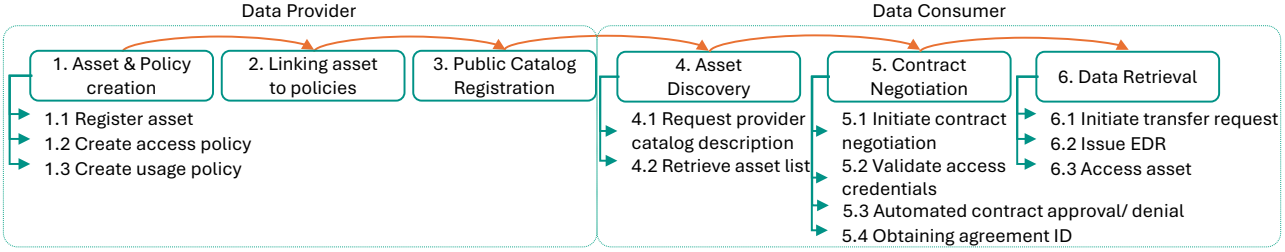


Figure 5: Six-step process outlining how data providers and consumers offer and retrieve DPP data within the data space.

depending on the type of data or the requirements of the data provider. The usage policy for the architect is already shown in Listing 1. Finally, these policies are linked to the specific asset through a contract definition, ensuring that the rules apply whenever the asset is accessed.

Public Catalog Registration

The data now has a dedicated policy but is not yet discoverable in the data space. To make it visible, the asset must be registered in the shared catalog of the provider's connector and linked with its access and usage policies. Only participants who satisfy the access policy can discover the asset, and those meeting the usage policy can access and use it under the defined conditions.

Asset Discovery

Next, the data consumer can explore available data offers. Assuming that they meet the access policy, they request the catalog from the data provider's connector, retrieving the list of exposed assets. The data consumer then queries the cached catalogs to identify the applicable policies for contract negotiation, enabling an informed decision on which data to request.

Contract Negotiation

Based on this information, the consumer initiates a contract negotiation, of which the process is shown in Figure 6. The data provider's connector verifies the data consumer's DID against the access policies. If the data consumer complies with the predefined terms, the negotiation is automatically accepted, and rejected otherwise. For now, this means that the negotiation process functions as a contract validation mechanism rather than an interactive bilateral negotiation, which is something that should be explored in future research. Once approved, the consumer receives a contract agreement ID to initiate the data transfer.

Data Retrieval

Using the agreement ID, the data consumer initiate the data transfer. The transfer is tracked via transfer processes, and the consumer retrieves the corresponding Endpoint Data Reference (EDR) to get the data address and authorization token. Finally, the data consumer uses this information to access the data from the data provider.

Following this process, the manufacturer, architect, and element owner each provide their respective DPP modules for the suspended ceiling element. The deconstructor retrieves each module in accordance with the corresponding usage policies, ensuring compliance with each provider's constraints. By repeating this process for all relevant assets, the deconstructor obtains access to the modular DPP data for the ceiling element.

Conclusion

This paper demonstrates how usage policies can enable modular and decentralized DPP data provision and retrieval within an AECO data space, using a suspended ceiling element as example. The research shows how building-product data supplied by multiple data providers can be offered and retrieved using differentiated usage policies. In doing so, the study contributes an exploratory proof-of-concept DPP data space that implements the complete technical workflow, including asset registration, access and usage policy definition, contract negotiation, and authenticated data retrieval. Because this research is based on hypothetical scenarios and a limited set of generic policy templates, it should be considered as an initial exploration rather than a mature implementation of usage poli-

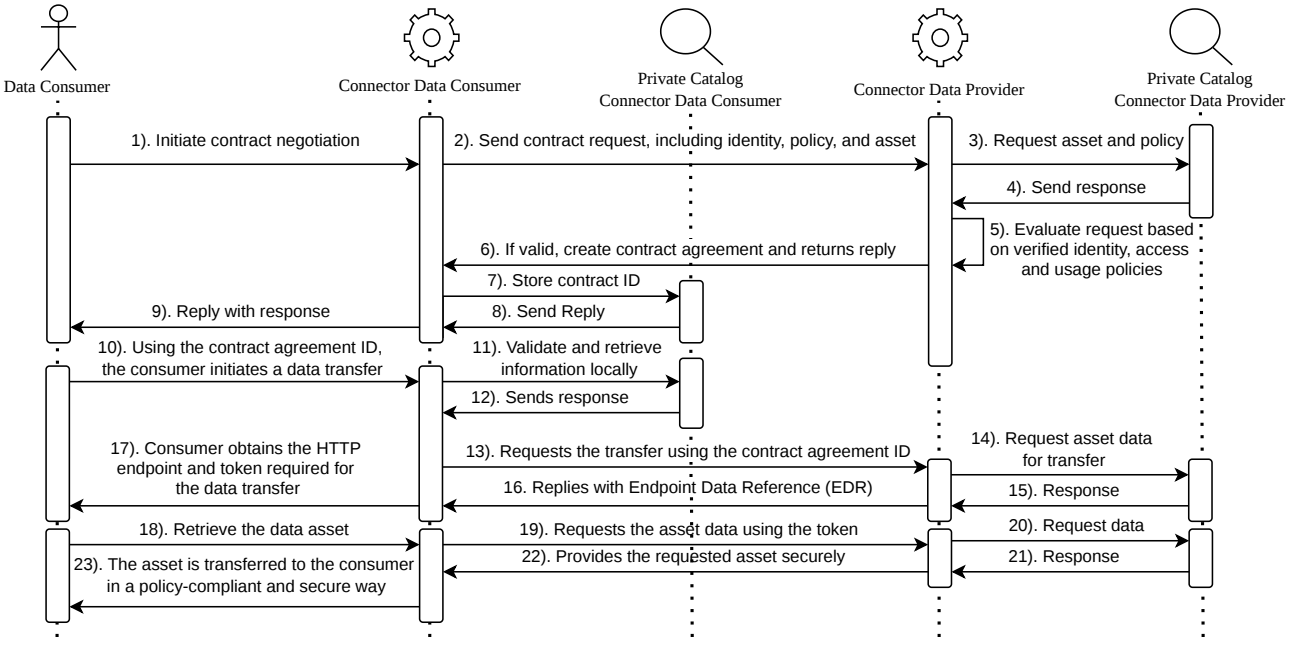


Figure 6: Sequence diagram showing contract negotiation and data retrieval between data provider and consumer.

cies in the AECO domain. Advancing toward operational deployment will require future research in several directions.

First, further work is needed to understand which specific usage-policy requirements AECO stakeholders have when sharing product-related data. Real-world needs must be validated to determine whether existing IDS policy templates are sufficient or require domain-specific extensions. As also highlighted by Bastiaansen et al. (2020), future research should therefore examine stakeholder- and data-specific policy needs while also developing more efficient negotiation mechanisms that prevent repetitive, module-by-module contracting and thus support both scalable and flexible DPP retrieval for data consumers.

Second, defining a usage policy is only one part of effective data governance; enforcing and monitoring compliance is another issue, as demonstrated in Cimmino et al. (2025). Future research should investigate mechanisms to verify policy adherence both within and beyond the boundaries the data space.

Third, as also highlighted by Sołtysiński et al. (2025), usage policies rely on the trustworthiness of digital identities associated with both data providers and consumers. Future work should explore how self-sovereign decentralized identities and trust frameworks interact with usage control and how they can be integrated into AECO data spaces.

Acknowledgments

This work is supported by the UPSCALE project, funded by Dutch Organization for Scientific Research (NWO).

Declaration of AI-generated text

During the preparation of this work, the authors used generative AI to improve the grammar and readability of the text. After using this, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Akbarieh, A. (2023). Information Modelling for Circular Transformation of Built Assets' End-Of-Life. PhD thesis, Unilu - Université du Luxembourg, Luxembourg.

Bastiaansen, H., Dalmolen, S., Kollenstart, M., and Engers, T. (2020). User-Centric Network-Model for Data Control with Interoperable Legal Data Sharing Artefacts. In PACIS 2020 Proceedings.

Bellini, A., Andersen, B., Klungseth, N. J., and Tadayon, A. (2024). Achieving a circular economy through the effective reuse of construction products: A case study of a residential building. Journal of Cleaner Production, 450:141753.

Belova, A., Gallina, V., Revenko, A., and Ahmeti, A. (2025). Digital Product Passport: Initial System Architecture with Knowledge Graphs and Data Spaces. In CEUR Workshop Proceedings, Portorož, Slovenia.

Bucher, D. F., Hunhevicz, J. J., Byers, B. S., Honic, M., De Wolf, C., and Hall, D. M. (2025). Decentralized Data Networks for Lifecycle Management in the Built Environment. ITcon, 30:826–851.

Byers, B., Hunhevicz, J. J., Honic-Eser, M., and De Wolf, C. (2024a). Exploring Tokenized Product Passport for Circular Construction Supply Chains. In 2024 European Conference on Computing in Construction. ETH Zurich.

Byers, B. S., Raghu, D., Olumo, A., De Wolf, C., and Haas, C. (2024b). From research to practice: A review on technologies for addressing the information gap for building material reuse in circular construction. Sustainable Production and Consumption, 45:177–191.

Çetin, S., Raghu, D., Honic, M., Straub, A., and Gruis, V. (2023). Data requirements and availabilities for material passports: A digitally enabled framework for improving the circularity of existing buildings. Sustainable Production and Consumption, 40:422–437.

Cimmino, A., Cano-Benito, J., and García-Castro, R. (2025). Open Digital Rights Enforcement framework (ODRE): From descriptive to enforceable policies. Computers & Security, 150:104282.

Cuno, S., Bruns, L., Tcholtchev, N., Lämmel, P., and Schieferdecker, I. (2019). Data Governance and Sovereignty in Urban Data Spaces Based on Standardized ICT Reference Architectures. Data, 4(1):16.

Dam, T., Krimbacher, A., and Neumaier, S. (2023). Policy Patterns for Usage Control in Data Spaces. In SEMANTiCS Workshops Proceedings, volume 3510, Leipzig, Germany.

European Commission (2025). (First) CPR Working Plan for 2026-2029. Technical report, European Union, Brussels.

European Parliament and European Union Council (2024a). Establishing a framework for the setting of ecodesign requirements for sustainable products, amending Directive (EU) 2020/1828 and Regulation (EU) 2023/1542 and repealing Directive 2009/125/EC. Technical report, European Union, Brussels.

European Parliament and European Union Council (2024b). Regulation (EU) 2024/1781 establishing a framework for the setting of ecodesign requirements for sustainable products, amending Directive (EU) 2020/1828 and Regulation (EU) 2023/1542 and repealing Directive 2009/125/EC. Technical report, European Union.

Gabellini, M., Civolani, L., Ronchi, M., and et al. (2025). Data Spaces in Manufacturing and Supply Chains: A Review and Insights from European Initiatives. Applied Sciences, 15(11).

Gallina, V., Lanbach, E., Ahmeti, A., and et al. (2025). Creating a Digital Product Passport Using Data Spaces and Ontologies: A Case Study at a Furniture Dealer. IFAC-PapersOnLine, 59(10):643–648.

Gil, G., Diez, F., Romero, R., Lazaro, I., and Fernandez, I. (2024). Construction Data Space for Building Permit Management. In Proceedings of the European Conference on Computing in Construction, volume 2024, pages 191–197.

Hagedorn, P., Donkers, A. J., Petrova, E., and König, M. (2025). Decentralised Data Exchange in Construction Data Spaces using Information Containers. In Proceedings of the 2025 European Conference on Computing in Construction and the 42nd CIB W78 Conference on IT in Construction, pages 566–573, Porto, Portugal.

International Data Space Association (2024). IDS RAM 4. https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0?tab=readme-ov-file.

Jansen, M., Blomqvist, E., Keskisärkkä, R., and et al. (2024). Modelling Digital Product Passports for the Circular Economy. In Proceedings of the The 2nd International Workshop on Knowledge Graphs for Sustainability (KG4S 2024) Colocated with the 21st Extended Semantic Web Conference (ESWC 2024), pages 14–28, Hersonissos, Greece.

Kebede, R., Moscati, A., Tan, H., and Johansson, P. (2024). A modular ontology modeling approach to developing digital product passports to promote circular economy in the built environment. Sustainable Production and Consumption, 48:248–268.

Kovacic, I., Honic, M., and Sreckovic, M. (2024). Digital Platform for Circular Economy in AEC Industry. EPOJ, 9(1):16.

Munoz-Arcentales, A., López-Pernas, S., Pozo, A., and et al. (2020). Data Usage and Access Control in Industrial Data Spaces: Implementation Using FIWARE. Sustainability, 12(9).

Nagel, L. and Lycklama, D. (2021). Design Principles for Data Spaces. Technical report, Zenodo.

Page, M. J., Moher, D., Bossuyt, P. M., and et al. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. BMJ, 372:n160.

Petrova-Antonova, D., Pareti, P., Tomov, P., and et al. (2026). Knowledge graph-driven policy enforcement in urban dataspace. In Knowledge Graphs, pages 406–421, Singapore. Springer Nature Singapore.

Rohrschneider, D., Pehlke, M., Handmann, U., and Jansen, M. (2026). LLM-based JSON Mapping and Blockchain Integration for Digital Product Passports. Digital Business, 6(1):100167.

Sołtysiński, T., Niederhausen, J., and Eichstädt, S. (2025). Digital Calibration Certificate in a trusted quality infrastructure federated data space: A proof of concept. Measurement: Sensors, 38:101484.

Steinbuss, S., Eitel, A., Jung, C., and et al. (2021). Usage Control in the International Data Spaces. Technical report, International Data Space Association.

van Capelleveen, G., Vegter, D., Olthaar, M., and van Hillegersberg, J. (2023). The anatomy of a passport for the circular economy: A conceptual definition, vision and structured literature review. Resources, Conservation & Recycling Advances, 17:200131.

van Zandwijk, K., Donkers, A. J., Petrova, E., and Pauwels, P. (2025). From Siloed Centralized Information Systems Towards Federated Data Ecosystems: Developing AECO Data Spaces. In Proceedings of the 2025 European Conference on Computing in Construction and the 42nd CIB W78 Conference on IT in Construction.

Verstraete, M., D'Hauwers, R., de Mildt, M., Colle, D., and Verbrugge, S. (2025). Assessing the value of decentralized and interoperable data storage for service providers. Inf Syst E-Bus Manage, 23(2):467–488.

Yang, C., Guo, Y., Yu, M., and Zhang, L. (2022). Truthfully Negotiating Usage Policy for Data Sovereignty. In 2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pages 20–27.

Zhang, Z., Gu, J., Li, R., Lian, W., Ma, C., and Tian, J. (2026). Bridging digital and green transitions: A literature review of digital product passport systems. Renewable and Sustainable Energy Reviews, 228:116600.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

SOCIAL MEDIA PUBLIC RESPONSE TO CRITICAL INFRASTRUCTURE CYBER DISRUPTIONS: A REDDIT BASED MULTI-DIMENSIONAL SENTIMENT ANALYSIS

Bharadwaj R. K. Mantha[1,2], Balaji Kesavan[3], and Johan Ninan[4]
[1]University of Sharjah, Sharjah, United Arab Emirates
[2]Indian Institute of Technology Madras, Chennai, India
[3]Independent Freelancer, Beaverton, United States of America
[4] Delft University of Technology, Delft, The Netherlands

Abstract

This study examines social media public response and discourse patterns to critical infrastructure cyber disruptions through a reddit based multi-dimensional sentiment analysis of the Florida water treatment plant incident. Results from 300 most relevant posts indicate that discourse is highly time-sensitive (e.g., 90% of total engagement occurred within first six hours). Security concern emerged as the dominant sentiment (~50%), while humor and sarcasm (~26%) reflected social coping dynamics. The findings highlight how online community signals can inform crisis communication and public trust strategies, rather than technical cybersecurity management, to support more transparent, trust-oriented CI protection efforts.

Introduction

Critical Infrastructure (CI) constituting of facilities such as roads, bridges, power grids, nuclear facilities, water and wastewater distribution systems, and flood protection mechanisms are vital for any country's economy, security, and societal well-being. These are considered critical because of their functional significance, whose disruption could have severe consequences on public safety, health, and economic prosperity. For instance, the U.S. Department of Homeland Security (DHS, 2020) defines CI as "*the physical and cyber systems and assets so vital that their incapacitation or destruction would have a debilitating impact on our physical or economic security or public health or safety*". Similarly, the European Programme for Critical Infrastructure Protection (EPCIP) highlights the need for uniform protection levels, minimal single points of failure, and rapid recovery arrangements across the European Union (EC, 2006).

Traditionally, these systems and processes were operated, maintained, monitored, and inspected within their respective functional capabilities with little to no overlap among different CIs. For example, transportation network had no direct linkages on their operational synergies when compared to flood protection mechanisms. Technological advancements enabled synergistic, strategic, and efficient integration of different CIs. The growing interdependencies and technological integrations among CIs amplify their complexity, as disruptions in one sector can cascade into others, exacerbating the impact of failures (Mantha et al., 2024). Moreover, the increasing reliance on cyber-physical integration for real-time monitoring and control, where physical systems are interconnected with cyber systems, introduces vulnerabilities that make these infrastructures susceptible to cyberattacks. In an era of human-centred digital transformation, social media platforms serve as a crucial medium for connecting end users with policymakers, offering a dynamic space to share concerns, disseminate information, and respond to crises in real time. Such vulnerabilities can compromise the physical systems they govern, potentially leading to catastrophic failures (Palleti et al., 2021). This study critically examines these vulnerabilities of CIs to cyberattacks and explores the implications to ensure the resilience of such essential systems. Understanding public responses through social media can provide actionable insights that incorporate public feedback, fostering a more adaptive and resilient approach to managing cyber crises. Cyberattacks on these CI assets pose a severe threat to physical systems, potentially rendering them inoperable, delegating control to unauthorized entities, or compromising sensitive data (Pacheco et al., 2019). The interconnected nature of CI, while essential for meeting performance and design specifications, exacerbates these risks. Interdependencies among infrastructures mean that faults or attacks on one CI can cascade into others, leading to complex and unforeseen disruptions (Rinaldi et al., 2002). Such escalation can disrupt operations, initiate feedback loops, and amplify disturbances across interconnected systems (Palletti et al., 2021).

Despite significant advancements in securing networked control systems, existing technologies remain inadequate to address the fast-evolving threat landscape (Stellios et al., 2018). Furthermore, as technology advances, malicious actors continue to exploit vulnerabilities, using increasingly sophisticated tools to compromise systems (Tounsi and Rais, 2018). These disruptions inconvenience the public, disrupting daily life and eroding trust in essential services. Efforts to address these cybersecurity challenges vary internationally. For instance, the European Union Agency for Cybersecurity (ENISA) promotes a collaborative, multi-stakeholder approach to securing CI, with an emphasis on public-private partnerships and citizen engagement (Bossong and Wagner, 2017). In contrast, countries such as Japan and South Korea emphasize top-down regulatory frameworks with centralized response mechanisms (Aggarwal and

Reddie, 2018). These diverse models of policy and engagement reflect contextual differences in governance, digital infrastructure maturity, and cultural attitudes toward public trust in technology (Krishna et al., 2023). However, across these diverse international approaches it is important to understand community responses to cyberattacks necessitating the need for this research.

Social media analysis can serve as an early detection mechanism, offering valuable insights into public sentiment, concerns, and evolving perceptions of cybersecurity incidents, thus playing a pivotal role in shaping adaptive cybersecurity frameworks within the architecture engineering construction and facility management (AECFM) industry. In this context, understanding public reactions to cyberattacks can offer valuable insights for enhancing cybersecurity strategies and future infrastructure resilience. This study seeks to answer two research questions, 1) How does the public respond on social media during a cyberattack on critical infrastructure assets, and what patterns emerge in their engagement? 2) How can insights from non-expert public responses inform crisis communication and public trust strategies for future infrastructure projects?

To address these research questions, literature review section critically explores existing research, methodology section details the research methodology and case selection strategy, in which a case study approach is used focusing on the Florida water plant hack (CISA, 2021). Specifically, this case study was chosen due to its public involvement and reaction which is critical in the context of this study. Following this, the results and discussion section discusses the findings on how people react to the identified cyberattack. Finally, the conclusions and future work section details the contributions to theory, practice, study limitations and future research directions.

Literature Review

Cybersecurity Challenges of Critical Infrastructure

Critical infrastructure (CI) encompasses systems and functions essential for society's uninterrupted functioning, such as energy grids, transportation networks, water supply systems, and smart city technologies (Lehto, 2022). In recent years, cyberattacks targeting CI and critical information infrastructures have become increasingly frequent, complex, and targeted, as attackers grow more professional and sophisticated (Tounsi and Rais, 2018). These cyberattacks can disrupt or damage physical infrastructure by infiltrating digital systems controlling critical processes, leading to catastrophic consequences without requiring a physical assault. Such attacks may not only damage physical systems but also pollute environments, disrupt vital services, and compromise public safety.

The evolving nature of cyber threats compounds the challenges of defending against these attacks. Identifying the source, motive, and trajectory of an attack is notoriously difficult due to the global scope of cyberspace and blurred boundaries between national, international, public, and private interests (Lehto, 2013). Rapid technological advancements further complicate this landscape (Garcia de Soto et al., 2022), making it difficult to anticipate and mitigate new forms of threats. While existing literature emphasizes intrusion detection and prevention methods (Baykara and Das, 2018), these measures are often inadequate to address the dynamic and pervasive nature of modern cyber threats.

Human factors play a significant role in cyberattacks, often serving as the weakest link in cybersecurity. Reports indicate that employee errors account for a majority of data breaches, frequently exploited through social engineering tactics (Romanosky, 2016). At the same time, attackers leverage sophisticated techniques to exploit cyber-physical systems in critical sectors such as energy, water, and transportation, causing disruptions with far-reaching implications for public safety, economic stability, and trust (Stellios et al., 2018). For instance, cyberattacks have the potential to contaminate water supplies, disrupt electricity grids, or compromise transportation systems, highlighting the severe societal impacts of such crises (George et al., 2024).

Technical vulnerabilities in CI systems commonly stemming from legacy hardware and software, insufficient authentication protocols, and poor segmentation between IT and operational technology networks provide additional context for public responses (Sonkor and Garcia de Soto, 2021). Systems such as programmable logic controllers, SCADA platforms, and human–machine interfaces often lack robust cybersecurity measures because they were not originally designed with digital threats in mind (Stellios et al., 2018). These technical weaknesses can be considered important contextual factors affecting public perception and emotional response. Thus, the growing overlap of technical and social vulnerabilities, increased attack frequency, and severity of cyber incidents necessitates framing cyberattacks on CI as socio-technical crises.

Cyber Attacks as Socio-Technical Crises

Crisis, as a concept, encapsulates human reactions to significant disruptions or misfortunes, ranging from personal losses to large-scale societal events. Multiple frameworks help understand crises, including the Protective Action Decision Model (Heath et al., 2018), sensemaking theory (Maitlis and Christianson, 2014), and Situational Crisis Communication Theory (Coombs, 2007). Prior work has applied structured emotional trajectory models to interpret public reactions during disruptive events. While qualitative interpretations of emotional evolution remain useful, recent research increasingly treats crises as socio-technical phenomena observable through large-scale digital traces and measurable public discourse patterns.

The growing availability of social media data has enabled quantitative analysis of public responses to crises using natural language processing (NLP), network analysis, and engagement metrics. Social media platforms such as Reddit function as real-time participatory environments where users interpret events, assign accountability, and construct risk narratives during unfolding crises.

Comparative platform studies show that different social media environments generate distinct thematic and emotional patterns due to their structural affordances and communication norms (Kwon and Park, 2023). Reddit supports longer-form, focused and threaded discussions that enable users to express nuanced opinions and collectively interpret uncertain or controversial events, like group discussions and focused interviews. Quantitative social media research demonstrates that online discussions can reveal hidden facets of societal reactions, identify topic-aligned communities, and track the evolution of narratives (Goel and Sharma, 2020).

Network-based and linguistic analyses further show that language characteristics and engagement patterns influence how cohesive online communities form and how discourse evolves over time (Arazzi et al., 2023). Prior studies indicate that discussions of vulnerabilities and threats often emerge earlier or with deeper engagement on Reddit compared with other platforms, highlighting its potential as an early indicator of public awareness and concern (Horawalavithana et al., 2019).

These advances position social media as both a communication channel and a measurable socio-technical sensor for crisis dynamics. Quantitative analyses using sentiment scoring, topic modelling, and engagement metrics allow researchers to capture temporal shifts in public perception and identify dominant narratives surrounding risk, trust, and accountability. In addition, systematic reviews of Reddit-based research emphasize the importance of considering platform-specific norms, subreddit cultures, and ethical considerations when interpreting findings (Proferes et al., 2021).

Within this study, cyberattacks on CI are conceptualized as socio-technical crises that unfold through both technical disruption and public interpretation across decentralized online communities. Quantitative, thread-level analysis of Reddit discussions provides a means to observe how users collectively interpret infrastructure vulnerabilities, assign responsibility, and express trust or concern in real time. This framing integrates crisis theory with computational social media analytics, enabling the measurement of public sentiment trajectories and engagement patterns during cyber incidents.

Research Gaps and Contributions

The following are the three main research gaps based on the aforementioned discussion.

1)Human trust dynamics gap: There is limited evidence-based understanding of how decentralized online communities shape public trust, risk perception, and accountability narratives following cyberattacks on CI.

2)Methodological gap: Most existing studies focus on technical threat detection or rely on single-platform, polarity-based sentiment analyses, offering only partial insights into how public discourse and crisis narratives evolve in real time.

3)Analytical gap: Quantitative social media research has largely emphasized topic detection, vulnerability mentions, or information diffusion patterns, with limited use of multi-dimensional sentiment, engagement-aware metrics, and thread-level analysis to capture nuanced public reactions.

This study therefore contributes to advancing empirical understanding of public response dynamics during CI cyber disruptions by treating social media discourse as a measurable feedback signal for cybersecurity communication and resilience. The aims are to: (1) quantify key sentiment distributions in early-stage discussions, (2) examine thread-level and engagement dynamics shaping trust and risk narratives through the lens of Situational Crisis Communication Theory (SCCT), and (3) identify dominant discourse patterns that can inform more transparent and trust-oriented cybersecurity communication.

Methodology

The following interconnected subsections provide a coherent approach for quantifying public response patterns and addressing the study's objectives.

Research Design and Case Context

This study adopts a quantitative social media analysis approach to examine public responses to a CI cyber disruption through Reddit discourse. Cyberattacks on infrastructure systems often generate rapid and emotionally varied public reactions that unfold in real time across digital platforms. Social media platforms, particularly Reddit, provide a valuable data source for capturing these reactions because of their threaded discussion structure, relative anonymity, and capacity for detailed user commentary (Proferes et al., 2021). Unlike survey-based approaches that capture retrospective perceptions, Reddit discussions enable observation of emergent public sentiment during the early stages of a crisis, making them suitable for examining evolving trust, accountability, and risk narratives.

Building on prior qualitative work that examined emotional trajectories in cyber crisis discourse, this study extends the analysis through a quantitative, thread-level, multi-dimensional sentiment framework. The goal is to empirically capture how decentralized online communities interpret and respond to cyber disruptions affecting CI, with a particular focus on measurable sentiment patterns and engagement dynamics. While this study relies on a single-incident design, this approach aligns with recent sentiment literature that leverages single-case events to establish deep, thread-level analytical baselines (Goel & Sharma, 2020; Kahlawi et al., 2025; Ninan et al., 2025). Furthermore, it has to be noted that Reddit discussions enable the observation of emergent public sentiment during the early stages of a crisis as a first step, making them suitable for examining evolving trust, accountability, and risk narratives.

Data Source and Collection

Reddit was selected as the primary data source because its subreddit-based structure facilitates focused, topic-specific discussions and supports longer-form discourse compared with microblogging platforms. Previous

research shows that Reddit often hosts deeper and more sustained conversations on cybersecurity and infrastructure-related issues, making it a suitable platform for capturing nuanced public reactions (Horawalavithana et al., 2019). Furthermore, Reddit's threaded architecture enables analysis of conversational dynamics across original posts, comments, and replies.

Data were collected from relevant public subreddits discussing the selected cyber incident within the first week following the event. Figure 1 shows the pseudo code developed to achieve the same. The collection window was chosen to capture early public reactions, which are often critical in shaping trust and risk perception narratives during crises. Prior research suggests that the initial hours and days following a disruptive event are particularly influential in establishing dominant public narratives (Maitlis & Christianson, 2014).

The final reviewed, filtered, and cleaned dataset consisted of 300 posts spanned across 15 total subreddits, 20 original posts (OP), 66 comments, 214 replies. Data cleaning involved removing duplicate posts, irrelevant discussions, spam, and posts lacking substantive textual content. Only posts directly related to the incident and public reaction were retained. All data were drawn from publicly accessible Reddit threads and anonymized prior to analysis.

```
BEGIN PROCEDURE
  Load structured CSV containing Reddit posts and comments

  FOR EACH post/comment in dataset:
    Build thread context (chain parent comments for replies)

    Analyze relevance using OpenAI gpt-5.2-2025-12-11
      // Outputs: Relevant_to_attack, Relevant_to_thread + explanations

    Analyze sentiment tags using OpenAI gpt-5.2-2025-12-11
      // Outputs: 20 boolean tags + explanation
      // Tags: Denial, Anger, Bargaining, Acceptance, Anxiety, etc.

    Record results (relevance + tags + explanations) to dataset
  END FOR

  Save enhanced CSV with analysis results
END PROCEDURE
```

Figure 1: Pseudo code developed to pull the Redditt Data

Unit of Analysis and Data Preparation

The unit of analysis in this study is the individual Reddit post, defined as any original post (OP), comment, or reply within a discussion thread. This approach allows for capturing both primary narratives and conversational responses, providing a more comprehensive understanding of discourse dynamics. Previous research highlights that analyzing threaded discussions rather than isolated posts yields richer insights into collective sensemaking processes during crises (Kahlawi et al., 2025).

Each post was assigned a unique identifier and categorized by type (OP, comment, reply), subreddit, and time of posting. Basic engagement metrics including upvotes and interaction structure were recorded to support engagement-weighted analysis. Textual data were preprocessed by removing URLs, emojis where necessary for clarity, and non-relevant formatting while preserving semantic meaning. Posts were retained in their original wording to maintain contextual integrity for sentiment tagging.

Multi-Dimensional Sentiment Framework

To capture the emotional and interpretive dimensions of public response, this study applies a multi-dimensional sentiment framework consisting of eight categories namely 1) Denial, 2) Anger, 3) Bargaining, 4) Depression, 5) Acceptance, 6) Security Concern, 7) Sarcasm, and 8) Humor. Five of these categories (i.e., 1 to 5) were derived from the Kübler-Ross model of emotional response to crisis (Kübler-Ross, 1969). This model has been widely applied beyond grief contexts to understand reactions to disruptive societal events and organizational crises (Ninan et al., 2025). Its structured progression provides a useful lens for examining how individuals process shock, assign blame, and eventually move toward adaptation following a disruptive event.

However, applying the Kübler-Ross model alone is insufficient for analyzing social media discourse. To address this, these methods were calibrated for a sentiment analysis of a crisis response which augments the baseline emotional framework with interpretive and coping mechanisms such as humor and sarcasm, which function as social responses to uncertainty and stress. Including humor, sarcasm, and security concern allows for capturing these communicative behaviors that may not fit neatly within traditional emotional stages, making the framework defensible for digital discourse. Humor has been shown to act as a psychological coping strategy during crises, enabling individuals to process fear and anxiety collectively (Martin and Ford, 2018). Similarly, sarcasm is frequently used in online discourse to express skepticism, critique institutions, or signal distrust (Filatova, 2012). Including humor and sarcasm as distinct categories allows for capturing these communicative behaviors that may not fit neatly within traditional emotional stages. The category security concern was added to explicitly capture expressions of risk awareness, vulnerability perception, and calls for cybersecurity improvements. Prior research indicates that public discourse following cyber incidents often centers on perceived system weaknesses and demands for accountability, making this a distinct and analytically relevant category (Tounsi & Rais, 2018). Together, these eight categories provide a comprehensive framework for analyzing both emotional and interpretive dimensions of public response.

Sentiment Tagging Procedure

Sentiment tagging was conducted using a large language model (LLM) assisted classification process supported by manual verification through the algorithm logic represented as pseudocode in Figure 2. Each post was assigned sentiment categories based on its tone and intent. Where posts contained multiple sentiments, the most salient sentiment was selected to maintain consistency, but the others were preserved and allocated in quantitative aggregation. To improve tagging reliability, a subset of posts (e.g., 10% randomly sampled) were manually reviewed to confirm alignment between automated classification and contextual interpretation. While identifying sarcasm computationally has historically been

a significant challenge for traditional algorithms, modern Large Language Models have demonstrated high accuracy in emulating human experts' evaluation of complex public sentiments on social media (Kim & Kim, 2025). Furthermore, recent NLP literature confirms that advanced LLMs are highly reliable in detecting contextual sarcasm and irony in digital discourse. Consequently, verification threshold is also sufficient to confirm alignment. This hybrid approach effectively combines computational efficiency with human interpretive oversight, which is the recommended standard in current social media sentiment analysis research (Kim & Kim, 2025). Tagged data were then further analyzed to identify overall distribution patterns and dominant narrative themes

To improve tagging reliability, a subset of posts (e.g., about 10% randomly sampled) were manually reviewed to confirm alignment between automated classification and contextual interpretation. This hybrid approach combines computational efficiency with human interpretive oversight, which is recommended in social media sentiment analysis research (Kim & Kim, 2025). Tagged data were then further analyzed to identify overall distribution patterns and dominant narrative themes.

Engagement Metrics and Quantitative Analysis

Quantitative analysis focused on both frequency based and engagement-weighted metrics. Frequency counts were used to determine the proportion of posts associated with each sentiment category. Engagement-weighted metrics considered upvotes and reply structures to account for the visibility and influence of posts within discussions (Fottner et al., 2022). Engagement weighting is important because highly endorsed posts often shape perceived consensus within online communities (Arazzi et al., 2023). By examining both raw counts and engagement-adjusted measures, the analysis captures not only what sentiments are expressed but also which sentiments gain traction within the community. Descriptive statistics were used to summarize sentiment distributions, while cross tabulations examined variations across post types such as OP and comments.

Temporal and Thread-Level Analysis

Temporal analysis examined how sentiment distribution evolved during the initial response window. Posts were grouped by time intervals to identify early narrative formation and shifts in dominant sentiments. Previous research suggests that early discourse patterns can strongly influence long-term public perception of crises (Coombs, 2007). Thread-level analysis examined how sentiments propagated within discussions. This included identifying whether certain sentiments were more likely to appear in original posts versus replies, and whether humor or sarcasm functioned as responses to expressions of concern or anger.

Reliability and Ethical Considerations

As discussed earlier, to ensure analytical consistency, a subset of posts were manually reviewed to validate sentiment classifications. Tagging procedures and aggregation methods were documented to support reproducibility. Results were compared using both raw and engagement-weighted counts to compare and contrast stability of findings. All data analyzed were publicly available Reddit posts. Usernames were not recorded or reported, and findings are presented in aggregate form. No interaction with users occurred, and the study adheres to standard ethical practices for research using publicly accessible social media data (Proferes et al., 2021).

```
BEGIN PROCEDURE
  Load structured CSV containing Reddit posts and comments

  FOR EACH post/comment in dataset:
    Build thread context (chain parent comments for replies)

    Analyze relevance using OpenAI gpt-5.2-2025-12-11
      // Outputs: Relevant_to_attack, Relevant_to_thread + explanations

    Analyze sentiment tags using OpenAI gpt-5.2-2025-12-11
      // Outputs: 20 boolean tags + explanation
      // Tags: Denial, Anger, Bargaining, Acceptance, Anxiety, etc.

    Record results (relevance + tags + explanations) to dataset
  END FOR

  Save enhanced CSV with analysis results
END PROCEDURE
```

Figure 2: Pseudo code developed to conduct multi-dimensional sentiment tagging and analysis

Results and Discussion

This section presents the quantitative findings specifically, sentiment distribution and engagement patterns summarized in first, while thread-level dynamics and interpretation in the second. Lastly, implications and limitations connecting the results to broader cybersecurity and resilience considerations are detailed.

Sentiment Distribution and Engagement Patterns

The final filtered most relevant dataset comprised 300 posts across 15 subreddits within the early-response window following the cyber incident, including 20 OPs (6.7%), 66 comments (i.e., response made to an OP) (22%), and 214 replies (i.e., response made to a comment) (71.3%). The initial extraction yielded a larger candidate set, including 30 OPs after automated relevance tagging and manual contextual verification, this was refined to 20 original posts retained for analysis. A similar two-stage filtering process was applied to comments and replies, where non-relevant or context-detached items were removed. This hybrid human-in-the-loop screening ensured that only posts directly tied to the incident and its interpretation were included in the quantitative analysis.

Across the dataset, as can be seen in Figure 3, security concern appeared in almost half (50%) of all posts, emerging as the dominant sentiment category and reflecting widespread public attention to infrastructure vulnerability, system reliability, and accountability. Expressions of Anger (~20–25%) and Denial (<5%) were also present, particularly in original posts where users reacted to initial reports or questioned the severity and legitimacy of the incident. At the same time, Humor (~20%) and Sarcasm (~25–30%) appeared consistently across comments and replies, suggesting the use of coping and critique mechanisms common in online discourse during disruptive events. The remaining Kübler-Ross–

derived categories namely bargaining, depression, and acceptance (<15%) appeared in smaller but notable proportions, indicating a gradual shift in tone as discussions progressed from initial shock and skepticism toward reflection and normalization.

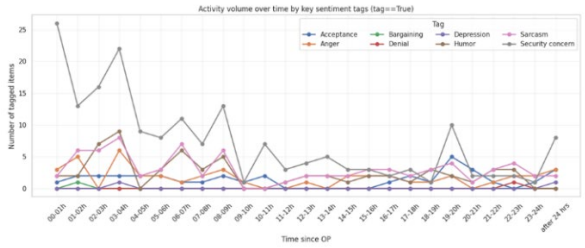


Figure 3: Temporal distribution of activity across different sentiments

Figure 4 reveals from the engagement patterns that a large proportion of total interactions occurred within the earliest hours of discussion. Nearly 50% posts were generated within the first six hours, while approximately 90% of cumulative engagement score was concentrated in this same early period, underscoring the rapid formation of public narratives. High-engagement posts were not always those expressing the most frequent sentiment, rather, posts combining concern with critique or humor often attracted the most interaction. The highest average endorsement per item occurred within the first hour after posting, suggesting that early contributions disproportionately shaped perceived credibility and importance. While original posts frequently introduced concerns or questions about the incident, replies tended to amplify or reinterpret these narratives, often through humor or sarcasm. This distribution highlights how public reaction to cyber disruptions unfolds through layered conversational dynamics rather than isolated expressions.

The analytical workflow used to generate relevance and sentiment tags involved automated LLM-assisted tagging followed by targeted human validation across sampled posts, ensuring alignment between computational classification and contextual interpretation. This process confirmed consistency between automated outputs and manual review, supporting the reliability of the final curated dataset and subsequent quantitative aggregation.

Thread-Level Dynamics and Interpretation

Thread-level analysis as shown in Figures 5 and 6 demonstrates that Reddit discussions function as decentralized spaces for collective sensemaking during infrastructure-related crises. For example, left to right represented from 0 (OP) to 1 (comment) to 2 (replies) and so on. Initial posts typically framed the event in terms of risk and vulnerability while comments and replies elaborated on these themes through personal interpretations, technical speculation, or institutional critique. For instance, security concern score – grey dropped from 73% to 47% while anger – orange increased to 23% and other sentiments peaked in the latter depths of discussions. Notably, denial is the least with about 1.6% which corroborates what was observed through the anger sentiment statistics observed. Activity volume peaked at early reply layers with most interaction occurring within the first two levels of response (layers 0,1 score > 1600 while layer 4 score ~ 400), indicating that collective interpretation emerges quickly once discussions begin. In several threads, expressions of concern were followed by humorous or sarcastic remarks that reframed the incident in more accessible or socially resonant terms (e.g., also evident from color tones in layers 1 and more but nonexistent in 0). Such patterns suggest that humor and sarcasm operate not merely as distractions but as mechanisms through which users negotiate uncertainty and express skepticism or distrust.

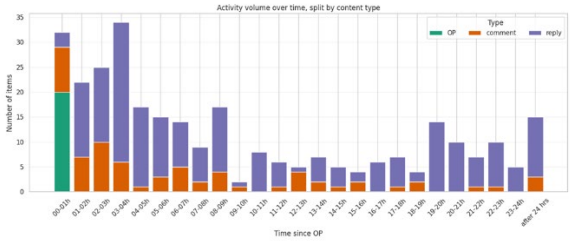


Figure 4: Distribution of activity type and volume over time

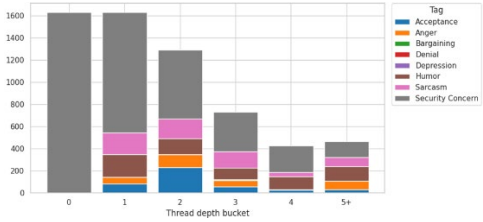


Figure 5: Thread depth weighted scores across sentiments

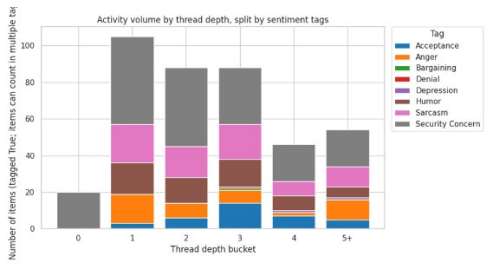


Figure 6: Thread depth activity across sentiments

The presence of Denial (rare at deeper levels) and Bargaining (moderate mid-thread presence) sentiments indicates attempts by some users to minimize perceived risk or reinterpret the incident's implications. Over time, however, these sentiments often gave way to Acceptance-oriented content (blue) in later discussion layers, reflecting normalization of systemic vulnerability discussions. This progression reflects the dynamic, non-linear nature of public response to disruptive events and supports the use of a multi-dimensional sentiment framework that captures both emotional and interpretive reactions.

From an analytical standpoint, these findings address previously identified gaps in understanding how decentralized online communities shape trust and risk narratives following cyber incidents. As guided by Situational Crisis Communication Theory (SCCT), the non-expert nature of this data is not a limitation but the

focal point of the study. The sample reflects public speculation, uninformed opinion, and casual commentary, which provides a direct, measurable lens into how sentiment evolves and which narratives gain traction to help agencies tailor their messaging and manage trust. The results demonstrate that public discourse is not only measurable but also structured by engagement dynamics that influence which narratives gain prominence during a crisis response. The iterative reduction from the initial candidate dataset including the refinement from 30 to 20 original posts and corresponding filtering across comments and replies also highlights the importance of combining automated tagging with human contextual verification when interpreting crisis-related social media discourse. By examining thread-level interactions rather than isolated posts, the analysis highlights how trust, skepticism, and concern are negotiated collectively within online communities. This supports the view of social media platforms as socio-technical environments where public perception and crisis interpretation co-evolve.

Implications and Study Limitations

The findings suggest that early-stage social media discourse can serve as a valuable feedback signal for cybersecurity communication and infrastructure governance, particularly within the architecture, engineering, construction, and facility management (AECFM) industry. The prominence of security concern across posts indicates strong public awareness of infrastructure vulnerabilities, while the prevalence of humor and sarcasm highlights the need for communication strategies that acknowledge and address public skepticism. Monitoring sentiment distributions and engagement patterns may help agencies and infrastructure developers identify emerging trust deficits and adjust messaging accordingly, aligning closely with SCCT principles. For AECFM practitioners, while based on a single-incident design, this approach offers an essential first step for integrating public perception into adaptive communication and lifecycle management strategies, thereby supporting more transparent and trust-oriented responses to cyber disruptions.

Several limitations should be acknowledged. First, the analysis focuses on a single incident within a limited early-response window, which may restrict the generalizability of findings. Second, Reddit users represent a specific subset of the broader population, and their discourse may not reflect wider public opinion. Third, while the multi-dimensional sentiment framework captures a range of emotional and interpretive responses, categorizing complex posts into single dominant sentiments may overlook nuance. Finally, LLM-assisted tagging introduces potential classification variability despite manual verification. These limitations highlight the need for broader comparative analysis across multiple incidents and longer timeframes. Expanding the dataset to include additional cyber events and cross-sector infrastructure contexts will enable more robust identification of recurring patterns and improve the generalizability of findings.

Conclusions and Ongoing Work

This study contributes to the existing body of knowledge by providing empirical, quantitative evidence on how decentralized online communities respond to CI cyber disruptions. By analyzing 300 Reddit posts across 15 subreddits, the study identifies dominant sentiment patterns, engagement dynamics, and thread-level interactions that shape public interpretations of cyber incidents. The findings show that security concern emerges as a primary narrative driver, while humor and sarcasm play significant roles in mediating public reaction and critique. These observations demonstrate how early-stage online discourse forms rapidly, concentrates engagement, and amplifies particular narratives that influence public perception. Such insights are relevant for cybersecurity agencies, infrastructure operators, policymakers, and specifically AEC professionals seeking to embed human-centered resilience into future infrastructure projects. By addressing AEC infrastructure vulnerabilities directly, these findings provide actionable suggestions for the industry to proactively incorporate public feedback into lean-inspired cybersecurity frameworks and facility management crisis response strategies.

However, the study is subject to several limitations. It focuses on a single incident and a short early-response window, which may not capture longer-term discourse evolution or variations across sectors and regions. The dataset is limited to Reddit users, whose demographic and behavioral characteristics may differ from the broader population. Therefore, it is worth noting that presented analysis and discussion serves as an initial methodological first step rather than a generalized conclusion. To address these limitations, ongoing work is expanding the dataset to include multiple cyber incidents across different infrastructure sectors and geographic contexts to validate these pattern claims and improve generalizability. This extended analysis will support the development of more robust, evidence-based frameworks for integrating social media feedback into cybersecurity communication and resilience planning, ultimately contributing to more adaptive and human-centred approaches to infrastructure protection.

References

Aggarwal, V. K., and Reddie, A. W. (2018). Comparative industrial policy and cybersecurity: a framework for analysis. Journal of Cyber Policy, 3(3), 291-305.

Arazzi, M., Nicolazzo, S., Nocera, A., and Zippo, M. (2023). The importance of language for the evolution of online communities. ESA, 222, 119847.

Baykara, M., and Das, R. (2018). A novel honeypot based security approach for real-time intrusion detection and prevention systems. JISA, 41, 103-116.

Bossong, R., and Wagner, B. (2017). A typology of cybersecurity and public-private partnerships in the context of the EU. Crime, Law and Social Change, 67, 265-288.

CISA (Cybersecurity and Infrastructure Security Agency) (2021). Compromise of U.S. Water Treatment Facility. Available at https://www.cisa.gov Accessed-2026/2/6.

Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. Corporate reputation review, 10(3), 163-176.

DHS (2020). Critical infrastructure security. U.S. Department of Homeland Security, www.dhs.gov/topic/critical-infrastructure-security

EC (2006). On a European programme for critical infrastructure protection. Communication from the Commission, COM(2006) 786 final, EC.

Filatova, E. (2012). Irony and sarcasm: Corpus generation and analysis using crowdsourcing, LREC.

Fottner A, Okhrin Y, Pfahler J, Wustl J. Reddit financial image post sentiment dataset. Data Brief. 2022 Nov 17;45:108759. doi: 10.1016/j.dib.2022.108759.

Garcia de Soto, B., Georgescu, A., Mantha, B., Turk, Z., Maciel, A., and Sonkor, M. S. (2022). Construction cybersecurity and critical infrastructure protection: new horizons for Construction 4.0. J. Inf. Technol. Constr., 27, 571-594.

George, A. S., Baskar, T., and Srikaanth, P. B. (2024). Cyber threats to critical infrastructure: assessing vulnerabilities across key sectors. Partners Universal International Innovation Journal, 2(1), 51-75, DOI:10.5281/zenodo.10639463.

Goel, R., & Sharma, R. (2020). Understanding the MeToo movement through the lens of Twitter. In S. Aref et al. (Eds.), Social Informatics (Lecture Notes in Computer Science, Vol. 12467, pp. 67–80). Springer.

Heath, R. L., Lee, J., Palenchar, M. J., and Lemon, L. L. (2018). Risk communication emergency response preparedness: Contextual assessment of the protective action decision model. Risk analysis, 38(2), 333-344.

Horawalavithana, S., Bhattacharjee, A., Liu, R., Choudhury, N., Hall, L. O., & Iamnitchi, A. (2019). Mentions of security vulnerabilities on Reddit, Twitter and GitHub. WI ’19 Proceedings.

Kahlawi, A., Masri, F., Ahmed, W., & Vidal-Alaball, J. (2025). Cross-cultural sense-making of global health crises: a text mining study of public opinions on social media related to the COVID-19 pandemic in developed and developing economies. JMIR, 27, e58656.

Kim, K., & Kim, S. (2025). Large language models’ accuracy in emulating human experts’ evaluation of public sentiments about heated tobacco products on social media. JMIR, https://doi.org/10.2196/63631.

Krishna, B., Krishnan, S., and Sebastian, M. P. (2023). Examining the relationship between national cybersecurity commitment, culture, and digital payment usage: an institutional trust theory perspective. ISF, 25(5), 1713-1741.

Kübler-Ross, E. (1969). On death and dying. Macmillan.

Kwon, S., & Park, A. (2023). Examining thematic and emotional differences across Twitter, Reddit, and YouTube: The case of COVID-19 vaccine side effects. Computers in Human Behavior, 144, 107734.

Lehto M. (2013). The cyberspace threats and cyber security objectives in the cyber security strategies. IJCWT, 3(3):1-18.

Lehto, M. (2022). Cyber-attacks against critical infrastructure. In Cyber security: Critical infrastructure protection (pp. 3-42). Cham: Springer.

Maitlis, S., and Christianson, M. (2014). Sensemaking in organizations: Taking stock and moving forward. Academy of management annals, 8(1), 57-125.

Mantha, B. R., Sonkor, M. S., and Garcia de Soto, B. (2024). Investigation of the cyber vulnerabilities of construction networks using an agent-based model. Developments in the Built Environment, 18, 100452.

Martin, Rod A., and Thomas Ford. 2nd Ed. (2018) The psychology of humor: An integrative approach. Academic press, ISBN-13: 978-0128121436

Ninan J, Mantha BRK, Kesavan B (2025), "Human-centred cybersecurity for critical infrastructure: the case of the Florida water plant hack" ECAM, 32(13).

Pacheco, J., Benitez, V. H., and Pan, Z. (2019). Security framework for IoT end nodes with neural networks. IJMLC, 9(4), 381-386.

Palleti, V. R., Adepu, S., Mishra, V. K., and Mathur, A. (2021). Cascading effects of cyber-attacks on interconnected critical infrastructure. Cybersecurity, 4, 1-19.

Proferes, N., Jones, N., Gilbert, S., Fiesler, C., & Zimmer, M. (2021). Studying Reddit. Social Media + Society, 7(2).

Rinaldi, S. M., Peerenboom, J. P., and Kelly, T. K. (2001). Identifying, understanding, and analyzing critical infrastructure interdependencies. IEEE CSM,21(6),11-25.

Romanosky, S. (2016). Examining the costs and causes of cyber incidents. Jrn. of Cybersecurity, 2(2), 121-135.

Sonkor, M. S., and García de Soto, B. (2021). Operational technology on construction sites: A review from the cybersecurity perspective. JCEM, 147(12), 04021172.

Stellios, I., Kotzanikolaou, P., Psarakis, M., Alcaraz, C., and Lopez, J. (2018). A survey of iot-enabled cyberattacks: Assessing attack paths to critical infrastructures and services. IEEE-CST,20(4),3453-95.

Tounsi, W., and Rais, H. (2018). A survey on technical threat intelligence in the age of sophisticated cyber attacks. Computers and security, 72, 212-233.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DEFINING A BLOCKCHAIN-VERIFIED SPECIMEN PASSPORT FOR 3DCP STANDARDISATION

Giancarlo Di Marco[1], Davide Lombardi[1], and Theodoros Dounas[2]
[1]Xi'an Jiaotong-Liverpool University, Suzhou, CN
[2]Heriot-Watt University, Edinburgh, UK

Abstract

3D Concrete Printing (3DCP) adoption is limited by prescriptive standards that cannot accommodate interdependent process parameters (e.g., nozzle speed, layer height, curing) affecting mechanical performance. Current approaches lack a tamper-proof, end-to-end linkage between production parameters and mechanical outcomes. We propose a blockchain-verified 3D-Printed Concrete specimen passport, defining its data schema, smart-contract logic, and integration with performance-based standardisation. The passport records production process, mechanical profile, and a 28-day compressive strength grade. Blockchain's immutable ledger and smart contracts automatically verify performance claims and assign 3DCP grades, enabling flexible, auditable certification. This provides a foundational reference for data-driven 3DCP compliance, fostering industrial adoption.

Introduction

The construction industry faces a digital transformation, driven by the convergence of advanced manufacturing technologies, digital information systems, and distributed ledger technologies (Lavagna et al., 2023; Samarasinghe and Wood, 2021). Among these innovations, 3DCP represents a paradigm shift in construction methodology, offering unprecedented design freedom, material efficiency, and potential for automation (Buswell et al., 2018). Unlike conventional cast concrete, where material properties can be reasonably predicted from standardised mix designs and controlled curing conditions, 3DPC mechanical properties are determined by a complex interplay of process parameters, including extrusion pressure, printing speed, layer deposition timing, and environmental conditions, that directly influence the mechanical performance of the final element (Wolfs et al., 2019). This variability, particularly affecting the interlayer adhesion, represents a fundamental challenge to structural certification and has been identified as a primary obstacle to the integration of 3DCP into regulated construction practice (Di Marco et al., 2025; Diks, 2019; Van Der Putten et al., 2022).

The general approach to traditional construction standardisation is based on the required concrete performance, which depends on design and operational conditions and is obtained through corresponding concrete mix variation; however, traditional quality assurance methods, developed for relatively simple casting processes, lack the granularity, transparency, and immutability required to establish robust traceability and stakeholder trust in digitally fabricated construction elements (Elghaish et al., 2023). Current standardisation efforts (ICC-ES, 2022; ISO/ASTM, 2023) provide foundational quality assurance requirements but do not yet offer a mechanism for immutable, end-to-end traceability of process parameters to performance outcomes. Therefore, the establishment of a unified, verifiable framework for certifying the mechanical performance and production integrity of additively manufactured concrete elements must be framed in a flexible standardisation paradigm, due to the already mentioned extreme variability of the process, and entail a layer of transparency, auditability, and traceability.

Blockchain technology, originally developed as the underlying infrastructure for cryptocurrencies, has emerged as a promising solution for addressing trust, transparency, and traceability challenges across various industries (Li et al., 2019; Turk and Klinc, 2017). In the construction sector, blockchain applications have been explored for supply chain management, contract execution, and information flow coordination (Darabseh and Poças Martins, 2023; Ye et al., 2023). The technology's core characteristics, decentralisation, immutability, transparency, and cryptographic security, align closely with the requirements for robust quality assurance in advanced manufacturing contexts.

The concept of digital passports for construction materials has recently gained traction as a mechanism for documenting the lifecycle history of construction materials and components, particularly in the context of circular economy initiatives (Senarathne et al., 2025; Trubina et al., 2024). Material passports serve as comprehensive digital records that capture information about material composition, origin, processing history, and performance characteristics, enabling informed decision-making regarding reuse, recycling, and end-of-life management (Kc et al., 2025). The EU's Digital Product Passport (DPP) (Damvakeraki and Vlachos, 2024), mandated by the Ecodesign for Sustainable Products Regulation (ESPR), is redefining product information management for construction products by requiring standardised, machine-readable digital records that accompany products throughout their lifecycle.

Stakeholder studies have identified improved traceability and resource efficiency as key benefits, alongside integration with Building Information Modelling (BIM) (Ruismäki et al., 2025). Blockchain-based material passport systems have been proposed to address the unreliability of centralised databases in construction's dispersed supply chain, with researchers arguing that blockchain enables decentralised, transparent, and immutable information management (Kc et al., 2026). Furthermore, blockchain has been explored for additive manufacturing traceability, including cloud 4D printing service platforms and digital twin creation for trusted data provenance (Liu et al., 2025).

Building on these foundations but focusing specifically on 3DCP specimen certification within a performance-based standardisation framework—a context not yet addressed in the literature—this paper introduces a blockchain-verified material passport specifically designed for 3DPC specimens. The proposed system cryptographically anchors a comprehensive data record that spans from digital design through material sourcing to real-time fabrication monitoring and final performance validation. The 3DPC specimen passport constitutes the cornerstone of a performance-based standardisation framework. By creating an immutable and auditable chain of custody, the system addresses the core challenges of 3DCP standardisation while mitigating information asymmetry among stakeholders.

Without a digital passport system for 3DCP, the construction industry faces several foreseeable consequences: each printed element would require case-by-case regulatory approval, significantly delaying project timelines and increasing costs; the lack of tamper-proof traceability would enable data manipulation, undermining trust among clients, insurers, and regulators; performance data from one print batch could not be reliably linked to subsequent batches, preventing continuous quality improvement and statistical process control; the construction sector would miss the opportunity to align with emerging EU Digital Product Passport regulations, which are rapidly redefining supply chain transparency for construction products.

Related work

Standardisation paradigms for 3DCP

The effective management of industrial production necessitates a strong approach to standardisation, which can fundamentally be characterised by two distinct philosophical paradigms: prescriptive and flexible.

The prescriptive approach is grounded in the principles of scientific management and statistical control (Shewhart, 1930). This paradigm seeks to achieve uniformity, interchangeability, and repeatability by specifying precise dimensions, values, tolerances, and processes. It operates on the principle that variation is a primary source of quality defects. The primary objective of this approach is to eliminate deviations and ensure perfect adherence to predefined specifications. This is illustrated by models such as Technical Specifications, which offer formal, detailed descriptions of requirements (ISO/IEC, 2021), and Geometrical Product Specifications (GPS), an international standards system that defines geometrical features—including form, position, and size—on technical drawings with clarity (Henzold, 2021; ISO 8015, 2011).

The second paradigm adopts a flexible approach, guided by systems theory and the necessity for strategic adaptability (Baldwin and Clark, 2000; Sanchez and Mahoney, 1996). This approach establishes clear goals, performance criteria, and interfaces while allowing considerable autonomy in the methods used to achieve these objectives. Its purpose is to ensure interoperability and alignment of goals across diverse and dynamic production environments without hindering innovation. A prominent model in this regard is Performance-Based Standards, which define the required functional outcomes without dictating specific technical solutions, thus fostering innovation in design and materials. (Foliente, 2000).

Prescriptive standards are primarily applied to critical, high-volume, or safety-sensitive components to ensure a fundamental level of quality and interchangeability. In contrast, flexible standards operate at the system and architectural levels, fostering product diversity, process innovation, and adaptability to change. This hybrid approach, integrating the rigidity of prescriptive models with the flexibility of adaptive models, serves as a hallmark of strong and competitive industrial operations (Sanchez and Mahoney, 1996).

Blockchain technology in construction supply chains

Blockchain technology has attracted significant attention from the construction sector as a potential solution to longstanding challenges related to information management, trust, and coordination among fragmented stakeholders (Wilson et al., 2024). Li et al. conducted a systematic review of blockchain applications in the built environment and construction industry, identifying key use cases including supply chain management, contract execution, and asset lifecycle management (Li et al., 2019). Their conceptual models highlight blockchain's potential to address information asymmetry, reduce transaction costs, and enhance transparency across construction project networks.

Turk and Klinc explored the potential of blockchain technology for construction management, emphasising its ability to create immutable transaction records and enable automated contract execution via smart contracts (Turk and Klinc, 2017). Their work identified several promising application areas, including payment automation, supply chain traceability, and quality documentation. However, they also noted significant challenges related to scalability, interoperability, and the need for industry-wide standards.

More recent research has focused on practical implementations and integration frameworks. Darabseh et al. presented a novel framework for blockchain orchestration and implementation in construction, developing a blueprint for blockchain applications using

Hyperledger Firefly (Darabseh and Poças Martins, 2023). Their work provides practical perspectives on infrastructure requirements and deployment strategies, building upon earlier Ethereum-based implementations. Wilson et al. proposed a multilayer blockchain framework for provenance tracking in construction supply chains, addressing critical challenges in verifiability, scalability, privacy, and interoperability through a privacy-centric solution that combines decentralised and centralised storage (Wilson et al., 2024).

Material passports and circular economy

Material passports have emerged as a critical tool for enabling circular-economy practices in the construction sector, providing comprehensive documentation of material characteristics, origins, and lifecycle histories. Trubina et al. conducted an in-depth analysis of digital technologies and material passports for circularity in buildings, highlighting the significant potential of integrating material passports with BIM, artificial intelligence, scanning technologies, blockchain, and the Internet of Things (Trubina et al., 2024). Senarathne et al. developed a standardised materials passport framework to unlock the full circular potential of the construction industry, incorporating technologies such as BIM, blockchain, QR codes, and radio-frequency identification (Senarathne et al., 2025). Dounas et al. introduced an innovative approach using Non-Fungible Tokens (NFTs) on the blockchain to build components, creating a digital infrastructure layer that can integrate with AEC supply chains (Dounas et al., 2021). Their work on non-fungible building components demonstrates how smart contracts can enable material passports and circular economies, providing an infrastructure for whole-life-cycle BIM and the digital encapsulation of value in architectural design.

3D printed concrete–quality control and standardisation challenges

The field of 3D printed concrete has experienced rapid growth over the past decade, yet significant challenges remain in quality control and standardisation. Buswell et al. provided a comprehensive roadmap for 3DPC research, identifying key technical challenges including material development, process control, structural performance prediction, and the development of testing protocols and standards (Buswell et al., 2018). A critical challenge in 3DPC quality assurance is the influence of process parameters on interlayer adhesion, which directly affects the mechanical performance and structural integrity of printed elements. Wolfs et al. investigated the hardened properties of 3D printed concrete, specifically examining how process parameters influence interlayer adhesion (Wolfs et al., 2019). Their research demonstrated that factors such as time gap between layer deposition, printing speed, and material rheology significantly affect bond strength between layers, introducing variability that must be carefully controlled and documented for structural applications. Di Marco et al. advanced the field by developing stress-informed non-planar slicing strategies for 3D concrete printing, demonstrating how computational design approaches can optimise print paths to enhance structural performance (Di Marco et al., 2025). Their work highlights the complex relationship between digital design, process execution, and final mechanical properties, underscoring the need for comprehensive documentation systems that can capture this entire chain of causality. In another work, Di Marco et al. outlined a list of 3DCP process parameters that contribute to defining the mechanical performance and grade of 3DPC (Di Marco and Cheung, 2025). The 3DPC standardisation challenge extends beyond technical testing protocols to encompass broader issues of certification, liability, and regulatory acceptance. Traditional concrete standards, developed for cast-in-place and precast applications, do not adequately address the unique characteristics of layer-by-layer additive manufacturing. The absence of established standards creates uncertainty for designers, contractors, and regulatory authorities, hindering the technology's adoption in structural applications.

Research gaps, aims, and objectives

The lack of a shared, validated standardisation framework undermines the industry's adoption of 3DCP, constituting the principal research gap. Working towards a hybrid standardisation model, our broader research endeavour proposes adopting a blockchain-enabled, decentralised framework to collect production reports, detect production-to-grade patterns, and derive, validate, and distribute production standards. This paper focuses exclusively on the foundational first objective: to define a 3DPC specimen passport.

Methodology

A hybrid approach at the intersection of design science methods and decentralised science is adopted, i.e. the recording of artefacts and governance onto blockchains for the production of new knowledge. A new blockchain-verified 3DPC specimen passport is built on a multi-layered architecture that integrates a digital twin for performance prediction and optimisation, blockchain storage, smart contract execution, and a user interface. Its architecture follows principles established in recent blockchain for construction applications (Darabseh and Poças Martins, 2023; Wilson et al., 2024), while incorporating specific adaptations for the unique requirements of 3DPC quality assurance.

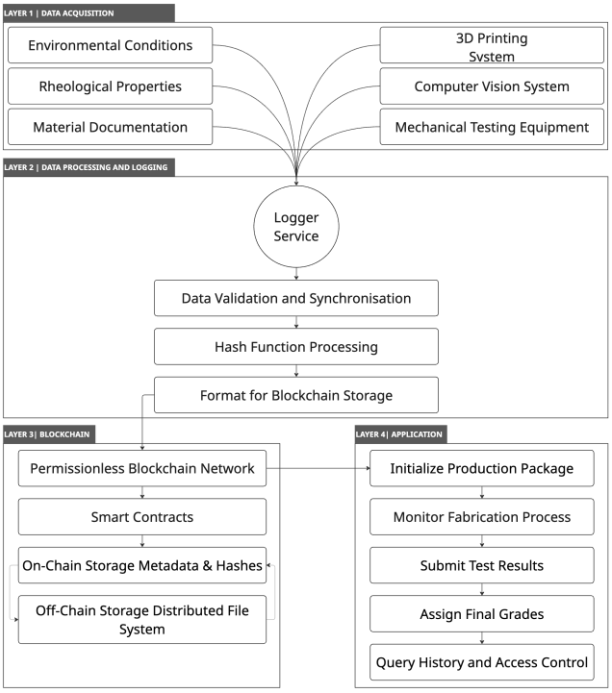


Figure 1: Framework layered structure

The system comprises four primary layers (Figure 1):

- Data Acquisition Layer: This layer encompasses all data sources that capture information about the 3DPC specimen fabrication process. It includes material batch documentation, in-line rheology monitoring, environmental conditions monitoring during printing and curing, machine control data from the robotic printer, computer vision imaging for geometric quality assessment, and mechanical testing conditions and results for performance validation.
- Data Processing and Logging Layer: A dedicated Logger Service processes raw sensor data, synchronises timestamps, performs data validation, and formats information for blockchain storage. This layer implements hash functions for large data objects (such as G-code files and vision data) to enable efficient on-chain storage while maintaining data integrity verification capabilities.
- Blockchain Layer: The core of the system utilises a permissionless blockchain network to store specimen passport data and execute smart contracts. Following best practices for construction applications (Ye et al., 2023), the system employs a hybrid on-chain/off-chain storage strategy: critical metadata and cryptographic hashes are stored on-chain, while large data objects are stored in distributed file systems, with their hashes recorded on the blockchain for verification.
- Application Layer: User-facing interfaces enable stakeholders to initialise production packages, monitor fabrication progress, submit test results, assign final grades, and query the complete history of any 3DPC element. Access control mechanisms ensure that different stakeholders have the appropriate permissions to view and modify data.

Study

An extensive material characterisation campaign with 60 cases of specimen tracking (an excerpt of the results is shown in Table 1) conducted in our robotic workshop, allowed to obtain a precise understanding of the connection between process parameters and the final 3DPC grade.

Table 1: Specimen Tracking and Grade

Specimen ID	250612-004	250612-001
Temperature (°C)	23.9	23.9
Humidity (%RH)	82.8	82.8
Concrete (kg)	82.3	99.5
Water (l)	12.35	14.9
Added Water (l)	0	0
w/i	0.15	0.15
Load orientation	Parallel	Perpendicular
Strength (MPa)	16.4	37.0
Grade	M15	M35

With reference to Figure 1, this was equivalent to implementing the data acquisition layer of the framework.

Framework overview

Beyond the basic material passport, our system manages the whole lifecycle of the 3DPC specimen, from production process to mechanical tests. The framework is purpose-built for construction 3DCP, addressing the issues of data manipulation, production logs can be altered after a failure; lack of independent verification – a single party often certifies its own prints; fragmented traceability, material provenance, process parameters, and test results are stored in disconnected systems, making root-cause analysis nearly impossible.

The material passport for 3DPC specimens is structured as a comprehensive Production Package containing 13 distinct information elements (Table 2), each serving a specific role in establishing traceability and enabling quality verification.

Table 2: Specimen Passport

Information	What it indicates
Production Package ID & Metadata	A unique identifier for the specific print job and its basic context.
Raw Material IDs & Certificates	The specific batch and origin of constituent materials (cement, aggregates, admixtures).
Target Mix Design	The intended material composition (e.g., water-to-cement ratio, admixture doses).

Target Rheological Properties	The desired workability and printability of the concrete (yield stress, viscosity).
Print G-code (Digital File Hash)	The exact instructions for the robotic printer's toolpath.
Environmental Targets	The intended ambient conditions (Temperature, Relative Humidity) for the print.
In-Process Sensor Data	The real-time state of the material and process (pressure, flow rate, actual rheology).
In-Process Vision & Scan Data	The visual and geometric quality of the print during fabrication (bead shape, layer alignment).
Mechanical Performance Data	The achieved structural properties of the printed element (compressive/flexural strength).
Final Assigned Grade	A holistic quality score (e.g., Grade M20) for the finished element.
Final Certification Hash	A digital seal that locks the package and confirms its integrity and completeness.

This structure draws upon material passport frameworks developed for circular construction (Senarathne et al., 2025; Trubina et al., 2024), but extends them with additive-manufacturing-specific fields required by emerging standards (ICC-ES, 2022; ISO/ASTM, 2023).

The experimental implementation of the decentralised material passport system for 3DPC samples is designed using blockchain technology to ensure immutable data provenance and transparent validation processes. The system architecture comprises three core smart contracts deployed on an Ethereum Virtual Machine (EVM)-compatible blockchain, integrated with the InterPlanetary File System (IPFS) for distributed data storage. This hybrid approach balances the need for permanent, tamper-proof record-keeping on the blockchain with the practical requirements for storing large volumes of scientific data off-chain.

The **ProductionPassportNFT** contract serves as the central registry, implementing the ERC-721 non-fungible token standard to create unique digital identities for each concrete production run. The **ValidationRegistry** contract implements a decentralised validation mechanism that requires consensus among multiple independent validators, addressing the critical need for peer review in scientific data. The **DataProvenance** contract maintains a comprehensive chain of custody, tracking materials from sourcing through final testing to ensure complete traceability of the production process.

The architecture follows a modular design principle, with each smart contract addressing a specific aspect of the material passport lifecycle (Figure 2). This modular architecture provides several advantages for scientific research collaboration. First, it enables different laboratories to participate in the ecosystem without requiring a central coordinating authority. Second, it allows each contract component to evolve independently as standards and requirements evolve. Third, it facilitates regulatory compliance by clearly separating concerns for data certification, validation, and provenance tracking. Fourth, it enables integration with existing Laboratory Information Management Systems (LIMS) through well-defined interfaces.

ProductionPassportNFT smart contract

The ProductionPassportNFT contract uses the OpenZeppelin ERC-721 implementation, providing the foundational framework for creating and managing production passports as non-fungible tokens. Each NFT represents a single production run and contains metadata linking on-chain records to off-chain scientific data stored on IPFS. The use of the ERC721 NFT standard provides inherent transferability, enabling production passports to be licensed or transferred between institutions while maintaining complete provenance of ownership. The contract addresses a fundamental challenge in scientific data management: how to create permanent, tamper-proof records of experimental work while accommodating the iterative nature of scientific workflows. Unlike traditional database systems, where records can be modified or deleted, blockchain-based records are immutable once finalised. However, the contract recognises that scientific data often requires revision before final publication—preliminary results may need correction, additional tests may provide new insights, or methodological improvements may necessitate data updates. The contract resolves this tension through a versioning system that allows updates before finalisation while ensuring absolute immutability after certification. The NFT standard ensures global uniqueness and verifiable ownership history. In construction, this enables a digital chain of custody – an auditor can see who minted, validated, and currently holds the passport.

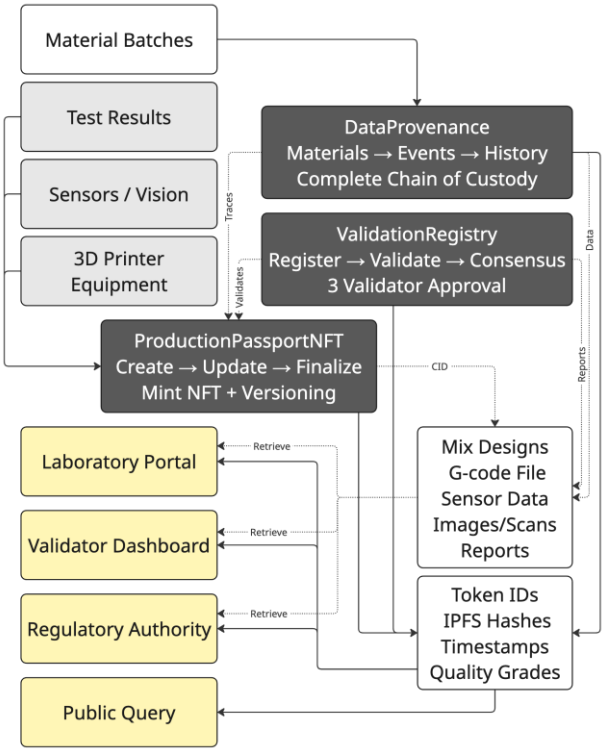


Figure 2: Blockchain-based System Architecture

Three mappings maintain contract state: passports maps token IDs to passport data structures, packageIdToToken enables lookup of token IDs by human-readable package identifiers, and _tokenIdCounter maintains the next available token ID for minting. Events are declared for all state-changing operations, enabling off-chain applications to efficiently monitor contract activity without polling every transaction.

ValidationRegistry smart contract

The ValidationRegistry contract implements a multi-validator consensus mechanism that replicates peer-review processes in scientific research (Lombardi et al., 2020). Unlike centralised validation systems, where a single authority determines data validity, this contract requires consensus from multiple independent validators, reducing bias and increasing trust in validated results. Requiring consensus among three independent validators (e.g., two accredited labs + one owner's engineer) replicates academic peer review. Collusion becomes harder, and the system provides a publicly auditable record of who approved what and when.

The architecture addresses several challenges inherent in decentralised validation. First, it must prevent collusion where validators coordinate to approve invalid data. The three-validator requirement increases the cost of such collusion and makes it more likely that at least one validator will act honestly. Second, it must maintain validator accountability while preserving the ability for legitimate criticism. The reputation scoring system provides a quantitative measure of validator reliability without creating barriers to participation. Third, it must balance transparency (all validations are publicly recorded) with practicality (validators are not penalized for disagreement with the eventual consensus).

DataProvenance Smart Contract

The DataProvenance contract implements comprehensive chain-of-custody tracking for materials and processes, addressing the critical need for traceability in construction materials. Unlike validation (which assesses quality) or certification (which attests to final results), provenance tracking creates a minute-by-minute history of everything that happened during production. This granular tracking serves multiple purposes: it enables root-cause analysis when quality issues arise, provides evidence of regulatory compliance, supports continuous improvement through statistical analysis, and increases transparency for end users who want to understand exactly how their concrete was manufactured.

The described blockchain-based architecture recognises two distinct types of provenance data that require different tracking mechanisms. Material provenance tracks the sources and certification of raw materials. These two provenance types are complementary: material provenance ensures quality inputs, while process provenance ensures proper handling of those inputs.

Summary of scientific rationalisation.

Each design decision is driven by a documented weakness in current 3DCP practice:

- NFT identity → solves counterfeit or duplicate batch records.
- Multi-validator consensus → solves the single-point-of-trust problem.
- Granular provenance → solves the traceability gap in layer-wise failures.
- Hybrid IPFS/blockchain → solves the data volume vs. immutability trade-off.
- Versioning → solves the tension between iterative research and final compliance.

Discussion and conclusions

This work is a fundamental part of the broader general aim to establish a flexible standardisation paradigm for 3DCP, encompassing diverse material choices and hardware configurations, and advocating the adoption of a decentralised framework underpinned by blockchain technology to enable verifiable compliance and traceability whilst fostering innovation. In this context, the 3DPC specimen passport enables traceability and auditability of different material choices and hardware configurations.

The introduction of a 3DPC specimen passport, as proposed in this study, represents a pragmatic step towards reconciling innovation and accountability in digital fabrication. This digital identity, encompassing the complete lifecycle of a printed element, from specific ink recipe and admixture ratios to hardware configuration and environmental conditions, shifts the paradigm of quality assurance to verifiable parameters and required performance. It functions as a cornerstone for enhanced traceability and auditability, creating an immutable record

that moves beyond batch-level documentation to unit-specific provenance, enabling a flexible, performance-based standardisation model. Rather than enforcing uniformity, the passport system accommodates heterogeneity in materials and hardware by recording these variables as intrinsic data points within a compliance framework. Rather than mandating identical processes (prescriptive standardisation), this flexible standardisation defines permissible performance envelopes and requisite data integrity; notably, this approach allows connecting 3DCP to existing standards (e.g., conventional concrete grade). A printer using a novel geopolymer and another using a traditional cementitious mix can both be deemed compliant if their respective passport data demonstrate that the final printed specimen meets the same structural and durability benchmarks. This fosters an ecosystem where technological diversification and experimentation are not hindered by standardisation but are instead its documented foundation.

In conclusion, the specimen passport is posited as an enabler of a responsive, evidence-based regulatory environment. By providing a mechanism to audit performance against specific production choices, it allows standardisation to evolve from a restrictive checklist into a dynamic framework that actively supports the tailored and innovative application of 3DCP technology.

Limitations and future work

The current implementation of the 3DPC specimen passport relies on manual input of process data; however, the reliable, auditable collection of process data should be ensured by adopting digital technologies and IoT equipment. Furthermore, the integrity of the data chain, from sensor to passport must be safeguarded, potentially through blockchain or other distributed ledger technologies, to prevent tampering and establish trust. Widespread adoption will also require industry-wide agreement on the minimal dataset for a viable passport, the definition of a 3DCP ontology, and its integration into existing BIM and quality management workflows, highlighting the current interoperability gaps include missing LIMS integration, material-type standardisation, and BIM connectivity. In the current stage, the preliminary identification of process parameters relevant to the determination of mechanical performance was finalised, based on material characterisation campaigns, and the design of the system architecture and smart contracts has been completed. The next stage will consist of embedding computer vision data in the implementing the blockchain; therefore, no EVM platform has been chosen nor tokenisation mechanism was presented in this paper. Future work will prioritise Layer 2 (Figure 1) deployment for cost reduction, governance token implementation enabling decentralised control and validator rewards, automated reputation systems, cross-chain accessibility, data marketplace integration, API development for existing systems, and regulatory pilot programs demonstrating compliance benefits and practical viability.

Acknowledgments

This work was supported by Xi'an Jiaotong-Liverpool University Research Development Fund [RDF-23-01-081] and Xi'an Jiaotong-Liverpool University Design School Funded Project [DES-SFP-20221127001].

References

Baldwin, C.Y., Clark, K.B., 2000. Design Rules: The Power of Modularity. The MIT Press. https://doi.org/10.7551/mitpress/2366.001.0001

Buswell, R.A., Leal de Silva, W.R., Jones, S.Z., Dirrenberger, J., 2018. 3D printing using concrete extrusion: A roadmap for research. Cement and Concrete Research, SI : Digital concrete 2018 112, 37–49. https://doi.org/10.1016/j.cemconres.2018.05.006

Damvakeraki, T., Vlachos, I., 2024. Digital Product Passports: A Blockchain-based Perspective. https://doi.org/10.13140/RG.2.2.36103.41125

Darabseh, M., Poças Martins, J., 2023. Blockchain Orchestration and Transformation for Construction. Smart Cities 6, 652–675. https://doi.org/10.3390/smartcities6010031

Di Marco, G., Cheung, L.H., 2025. Real-Time Defect Detection, Analysis and Suggestion in Adaptive 3D Concrete Printing via Multimodal LLM Integration, in: Proceedings of the IASS 2025 Symposium. Presented at the IASS2025 The Living Past as a Source of Innovation, Ingenta Connect, Mexico.

Di Marco, G., Yau, H.M., Lombardi, D., 2025. Stress-Informed Non-Planar Slicing for 3D Concrete Printing, in: Proceedings of the IASS 2025 Symposium. Presented at the IASS2025 The Living Past as a Source of Innovation, Ingenta Connect, Mexico.

Diks, T., 2019. The roadmap to standards for 3D concrete printing. Research on the interplay between technological and legislative developments (info:eu-repo/semantics/bachelorThesis). University of Twente, Netherland.

Dounas, T., Jabi, W., Lombardi, D., 2021. Non-Fungible Building Components: Using Smart Contracts for a Circular Economy in the Built Environment, in: Blucher Design Proceedings. Presented at the XXV International Conference of the Iberoamerican Society of Digital Graphics, Editora Blucher, online, pp. 1189–1198. https://doi.org/10.5151/sigradi2021-20

Elghaish, F., Hosseini, M.R., Kocaturk, T., Arashpour, M., Bararzadeh Ledari, M., 2023. Digitalised circular construction supply chain: An integrated BIM-Blockchain solution. Automation in Construction 148, 104746. https://doi.org/10.1016/j.autcon.2023.104746

Foliente, G., 2000. Developments in performance-based building codes and standards. Forest Products Journal 50, 12–21.

Henzold, G., 2021. Geometrical Dimensioning and Tolerancing for Design, Manufacturing and Inspection. Elsevier. https://doi.org/10.1016/C2020-0-01081-9

ICC-ES, 2022. AC509: 3D Automated Construction Technology for 3D Concrete Walls.

ISO 8015, 2011. ISO 8015:2011(en) Geometrical product specifications (GPS) — Fundamentals — Concepts, principles and rules.

ISO/ASTM, 2023. ISO/ASTM 52939:2023 - Additive manufacturing for construction — Qualification principles — Structural and infrastructure elements.

ISO/IEC, 2021. ISO/IEC Directives, Part 2 - Principles and rules for the structure and drafting of ISO and IEC documents.

Kc, A., Senaratne, S., Perera, S., Nanayakkara, S., 2026. Blockchain-Based Material Passports: A Review of Managing Built Asset Information for Material Circularity. Buildings 16, 658. https://doi.org/10.3390/buildings16030658

Kc, A., Senaratne, S., Perera, S., Nanayakkara, S., 2025. Review of current digital technologies for material passports to enhance circularity towards net zero. Built Environment Project and Asset Management 15, 1055–1072. https://doi.org/10.1108/BEPAM-03-2024-0085

Lavagna, M., Giorgi, S., Pimponi, D., Porcari, A., 2023. Enabling technologies for circular economy in the construction sector. Techne 214–224. https://doi.org/10.36253/techne-13130

Li, J., Greenwood, D., Kassem, M., 2019. Blockchain in the built environment and construction industry: A systematic review, conceptual models and practical use cases. Automation in Construction 102, 288–307. https://doi.org/10.1016/j.autcon.2019.02.005

Liu, J., Zainudin, E.S., As'arry, A.B., Ismai, M.I.S.B., Zhang, C., 2025. Modeling of the blockchain-empowered cloud 4D printing services collaboration digital twin platform oriented on supply–demand. Soft Comput 29, 977–1004. https://doi.org/10.1007/s00500-025-10461-x

Lombardi, D., Dounas, T., Cheung, L.H., Jabi, W., 2020. Blockchain Grammars for Validating the Design Process. https://doi.org/10.5151/sigradi2020-56

Ruismäki, W., zu Castell-Rüdenhausen, M., Pohjalainen, E., Wahlström, M., 2025. Stakeholder perspectives on digital product passports for construction products. Cleaner Logistics and Supply Chain 17, 100275. https://doi.org/10.1016/j.clscn.2025.100275

Samarasinghe, D.A.S., Wood, E., 2021. Innovative Digital Technologies:, in: Underwood, J., Shelbourn, M. (Eds.), Advances in Civil and Industrial Engineering. IGI Global, pp. 142–163. https://doi.org/10.4018/978-1-7998-6600-8.ch006

Sanchez, R., Mahoney, J.T., 1996. Modularity, flexibility, and knowledge management in product and organization design. Strategic Management Journal 17, 63–76. https://doi.org/10.1002/smj.4250171107

Senarathne, H.N.Y., Weerasinghe, N.P., Zhang, G., 2025. Developing a Standardized Materials Passport Framework to Unlock the Full Circular Potential in the Construction Industry. Sustainability 17, 6337. https://doi.org/10.3390/su17146337

Shewhart, W.A., 1930. Economic Quality Control of Manufactured Product1. Bell System Technical Journal 9, 364–389. https://doi.org/10.1002/j.1538-7305.1930.tb00373.x

Trubina, N., Leindecker, G., Askar, R., Karanafti, A., Gómez-Gil, M., Blázquez, T., Güngör, B., Bragança, L., 2024. Digital Technologies and Material Passports for Circularity in Buildings: An In-Depth Analysis of Current Practices and Emerging Trends, in: Ungureanu, V., Bragança, L., Baniotopoulos, C., Abdalla, K.M. (Eds.), 4th International Conference "Coordinating Engineering for Sustainability and Resilience" & Midterm Conference of CircularB "Implementation of Circular Economy in the Built Environment," Lecture Notes in Civil Engineering. Springer Nature Switzerland, Cham, pp. 690–699. https://doi.org/10.1007/978-3-031-57800-7_64

Turk, Ž., Klinc, R., 2017. Potentials of Blockchain Technology for Construction Management. Procedia Engineering 196, 638–645. https://doi.org/10.1016/j.proeng.2017.08.052

Van Der Putten, J., Hoogeveen, M.J., Bruurs, M.J.A.M., Laagland, H.L.M., 2022. 3DCP Structures: The Roadmap to Standardization, in: Buswell, R., Blanco, A., Cavalaro, S., Kinnell, P. (Eds.), Third RILEM International Conference on Concrete and Digital Fabrication. Springer International Publishing, Cham, pp. 50–55. https://doi.org/10.1007/978-3-031-06116-5_8

Wilson, S., Adu-Duodu, K., Li, Y., Sham, R., Almubarak, M., Wang, Y., Solaiman, E., Perera, C., Ranjan, R., Rana, O., 2024. Blockchain-Enabled Provenance Tracking for Sustainable Material Reuse in Construction Supply Chains. Future Internet 16, 135. https://doi.org/10.3390/fi16040135

Wolfs, R.J.M., Bos, F.P., Salet, T.A.M., 2019. Hardened properties of 3D printed concrete: The influence of process parameters on interlayer adhesion. Cement and Concrete Research 119, 132–140. https://doi.org/10.1016/j.cemconres.2019.02.017

Ye, X., Tao, X., Cheng, J.C.P., König, M., 2023. Blockchain-BPMN Integrated Framework for Construction Management, in: Capone, P., Getuli, V., Pour Rahimian, F., Dawood, N., Bruttini, A., Sorbi, T. (Eds.), Proceedings e Report. Firenze University Press, Florence, pp. 309–317. https://doi.org/10.36253/979-12-215-0289-3.29

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

GOVERNING CONSTRUCTION MATERIAL REUSE AS A BLOCKCHAIN COMMONS: OSTROM'S DESIGN PRINCIPLES AND EU DIGITAL PRODUCT PASSPORTS

Theodoros Dounas[1,3], Michele Victoria[2], Magdalena Blasiuzak, Zhe Jiang[1], and Hico McDonald[3]
[1]Heriot Watt University, Galashiels, UK
[2]Robert Gordon University, Aberdeen, UK
[3]Adventurous systems, London, UK

Abstract

Scotland's construction industry operates at a circularity rate of just 1.3%, exemplifying a wider EU governance failure: the Construction Products Regulation and Ecodesign for Sustainable Products Regulation mandate Digital Product Passports (DPPs) for construction materials, assuming linear supply chains, with no governance for reclaimed material re-entry. This paper applies Ostrom's principles for commons governance to a blockchain-based marketplace for construction material reuse. Using hybrid methods, we present a deployed prototype architecture of seven smart contracts, built on VeChainThor and implemented as the TRACE platform (trace.adventurous.systems), designed for EU DPP compliance while implementing commons governance for material reuse's institutional context.

Introduction

The construction industry is responsible for approximately 37% of global CO^2 emissions and consumes roughly half of all extracted materials worldwide (UNEP, 2022). In Scotland, despite ambitious circular economy targets, the construction sector achieves a circularity rate of only 1.3% (Zero Waste Scotland, 2024)—a figure that reflects a broader EU-wide challenge, as the governance infrastructure for material reuse does not yet exist in any member state. The vast majority of demolition materials are downcycled or landfilled rather than reused in their original form. Beyond an environmental failure this is a governance failure: the institutional infrastructure required to coordinate material reuse across fragmented stakeholders does not exist.

Two converging policy developments create both urgency and opportunity. First, the EU's revised Construction Products Regulation (CPR), which entered into force on 7 January 2025, mandates Digital Product Passports for construction products, requiring structured digital records of material composition, environmental performance, and compliance data (European Commission, 2024a). Second, the Ecodesign for Sustainable Products Regulation (ESPR, EU 2024/1781) establishes the broader DPP framework, specifying JSON-LD data formats, GS1 Digital Link identifiers, and tiered access controls for consumers, professionals, and regulators (European Commission, 2024b). Construction product DPPs are expected to become mandatory between 2029 and 2032 through phased harmonised standards.

However, these regulations are designed for linear supply chains: a manufacturer produces a product, declares its performance, and places it on the market with a CE marking and accompanying DPP. The regulatory framework offers no model for what happens when a construction material is deconstructed from a building, assessed for condition, and re-enters circulation through a community-run reuse hub. Who issues the DPP for a reclaimed steel beam? Under what authority? With what quality assurance mechanisms? How are disputes resolved when the 'manufacturer' is a volunteer-run charity?

We argue that this governance vacuum is precisely the institutional challenge that Elinor Ostrom's (1990) design principles for managing common-pool resources can address. Construction materials available for reuse constitute a shared resource that is rivalrous (one builder's use of a reclaimed beam excludes another's) and partially excludable (access requires knowledge of availability and logistics coordination). The community of reuse hubs, suppliers, and buyers face classic commons dilemmas: how to maintain quality without central authority, how to share information without exploitation, and how to distribute costs and benefits fairly across heterogeneous participants.

This paper presents findings from the TRACE (Transforming Resources and Advancing Circular Economy) project, which develops a blockchain-based digital marketplace for construction material reuse hubs. We address the following research question: *How can Ostrom's design principles for commons governance inform the design of EU DPP-compliant blockchain infrastructure for construction material reuse?*

Related work

Ostrom's design principles and digital commons

Ostrom's (1990) eight design principles for successful commons governance—clearly defined boundaries, congruence between appropriation and provision rules, collective-choice arrangements, monitoring, graduated sanctions, conflict-resolution mechanisms, minimal recognition of rights to organise, and nested enterprises—were originally derived from analysis of natural resource commons. Subsequent work has extended these principles

to digital contexts. Schweik and English (2012) applied the framework to open-source software communities. Fuster Morell (2014) adapted it to online creation communities. Most relevant to this research, Rozas et al. (2021) systematically mapped Ostrom's principles to six blockchain affordances: tokenisation, self-enforcement and autonomy of rules, decentralisation of authority, transparency, and codification of trust through cryptographic mechanisms. The Mozilla Foundation's practical framework (Ruhaak et al., 2021) further operationalised Ostrom for data commons governance.

EU digital product passports for construction

The EU's DPP framework represents the most significant regulatory development in construction product information management in decades. The revised CPR (EU 2024/1305) folds DPPs into the CE marking workflow, requiring Declarations of Performance and Conformity to be digitally accessible alongside environmental data including Global Warming Potential, which becomes mandatory from 2026 (European Commission, 2024a). The ESPR (EU 2024/1781) establishes horizontal DPP requirements: machine-readable structured data (JSON-LD aligned with Schema.org), persistent unique product identifiers (GS1 GTIN with serial-level granularity via GS1 Digital Link), QR code or NFC data carriers, tiered access controls, and mandatory third-party backup storage (European Commission, 2024b). Environmental product data must align with EN 15804+A2 standards for lifecycle assessment.

Critically, these regulations assume a manufacturer-to-market linear chain. The ESPR designates the 'economic operator'—typically manufacturer or importer—as responsible for creating and maintaining the DPP (European Commission, 2024b). For reclaimed materials, this responsibility chain breaks: the original manufacturer may no longer exist, the material's properties may have changed through use and ageing, and the entity placing the reclaimed material on the market is typically a small reuse hub or social enterprise with limited resources for conformity assessment. This institutional gap is the specific problem our research addresses.

Blockchain for construction material traceability

Blockchain applications in construction have focused primarily on material traceability and supply chain transparency. Dounas et al (2021) and Lombardi (2022) provided the first comprehensive treatment of non fungible tokens on blockchain based smart contracts for building components for construction, covering smart contracts, non-fungible tokens for building components. Circularise (Amsterdam) and ReCheck (Netherlands) have implemented blockchain-based material identity systems, while the EU BAMB (Buildings as Material Banks) project explored material passport concepts. However, these implementations share a common limitation: they treat blockchain as a technical ledger for recording provenance, without addressing the governance question of who has authority to assert material properties, how quality disputes are resolved, and how collective decisions about standards and parameters are made. Circularise, for example, records material data but relies on manufacturer self-declaration, leaving the governance vacuum unchanged when no manufacturer exists. BAMB developed material passport schemas without addressing the institutional infrastructure required to issue, verify, and dispute those passports in a commons setting. This gap—between technical traceability and institutional governance—is precisely what this research addresses through the Ostrom framework.

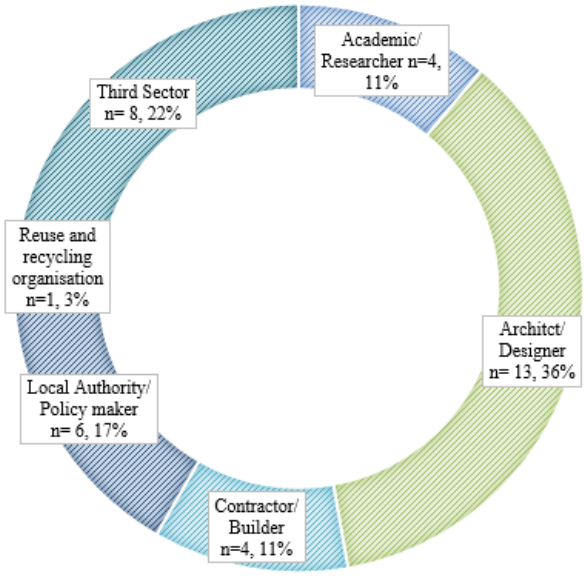


Figure 1: distribution of stakeholders' professional backgrounds

Methodology

Oversized This research follows a Design Science Research (DSR) methodology (Hevner et al., 2004) structured through a Design Thinking process (Stanford d.school, 2010). The current paper reports on the Empathise, Define, and Ideate phases, with the prototype architecture now implemented and deployed as the TRACE platform on VeChainThor.

Data collection

Four participatory workshops were conducted with 36 stakeholders across six categories: architects (n=6), third-sector organisations (n=7), local authorities (n=6), academics (n=5), contractors (n=6), and reuse organisations (n=6). Workshops were organised around four thematic areas: (1) materials and practical use, (2) trust and quality, (3) information and decision-making, and (4) process and experience. Each workshop used Microsoft Teams with Miro whiteboards for collaborative data capture. Discussions were recorded and transcribed with participant consent.

Analytical framework

Data was analysed using a hybrid deductive-inductive framework. The primary method followed Gioia et al.'s

(2013) approach: first-order codes captured participant language, second-order themes represented researcher-interpreted patterns, and aggregate dimensions mapped to theoretical constructs. An Ostrom Institutional Analysis and Development (IAD) overlay was applied to code for action arenas, rules-in-use, biophysical conditions, community attributes, and outcomes. This dual-layer analysis enabled systematic translation from stakeholder insights to governance requirements to prototype specifications.

Findings

Thematic analysis revealed three critical governance gaps that consistently surfaced across all stakeholder categories and all four thematic areas: (1) assurance mechanisms—the absence of verification pathways for reclaimed material quality; (2) logistics infrastructure—the lack of coordination systems for material availability and movement; and (3) information systems—the absence of structured digital records comparable to those available for new construction products. These gaps map systematically onto Ostrom's design principles, as presented below.

Boundaries and material viability (P1)

Stakeholders identified clear material viability hierarchies. Mineral-based materials (stone, slate, granite), brick and masonry, and structural timber were considered most viable for reuse due to durability and professional familiarity. Concrete, moisture-sensitive timber, and specification-specific steel were cited as most challenging. Participants consistently demanded clear boundary definitions: what materials qualify for reuse, under what conditions, and with what documentation. In Ostrom's terms, the resource system requires clearly defined boundaries before appropriation rules can function.

Monitoring and trust infrastructure (P4)

Trust emerged as the dominant concern across all stakeholder groups, particularly intensified by the post-Grenfell regulatory environment. Participants emphasised that reclaimed materials must be supported by documentation comparable to new-material technical data sheets. Third-party verification, recognised certification schemes, and professional expertise were identified as essential. Current checking practices rely on a combination of document-led assurance (certificates, warranties), expert-led assessment (structural engineer reports, visual grading), and operational boundary-setting (focusing on lower-risk applications). The absence of systematic monitoring mechanisms was identified as the primary barrier to scaling reuse beyond informal, relationship-dependent exchanges, in some cases with public procurement as a stimulus for circular practices.

Information asymmetry and collective choice (P3)

Decision-critical information identified by stakeholders included precise specifications (quantity, dimensions, cost, performance data), testing and certification evidence, and traceability data establishing provenance. A significant finding was the distinction between essential information (compliance and risk data) and desirable information (embodied carbon data, material history). Stakeholders from all categories expressed that current information sourcing is informal and relationship-dependent, particularly for smaller projects, while regulated buildings require substantially higher documentation standards. This asymmetry maps to Ostrom's collective-choice arrangements: stakeholders need voice in determining what information standards are required at what levels.

Congruence and cost distribution (P2)

A persistent temporal mismatch emerged: the costs of material documentation and quality assurance are borne at the point of reclamation, while the benefits accrue at the point of reuse—potentially months or years later. Third-sector organisations expressed concern that documentation costs could create barriers for smaller hubs operating on minimal budgets. Participants identified that a viable system must distribute costs proportionally to benefits derived, aligning with Ostrom's congruence principle.

Prototype architecture: implementation and design

Based on the governance requirements derived from stakeholder analysis, Ostrom mapping, and EU DPP technical specifications, we have implemented a prototype architecture comprising seven smart contracts integrated through a commons governance layer. The prototype is deployed as the TRACE platform (trace.adventurous.systems), built on VeChainThor—an EVM-compatible enterprise blockchain with built-in fee delegation (enabling users to interact without holding cryptocurrency), multi-clause batch transactions, and an energy-efficient proof-of-authority consensus model aligned with sustainability goals. The architecture follows a 'thin on-chain, rich off-chain' philosophy: the full material passport data resides in a conventional database (PostgreSQL) as structured JSON-LD, while the blockchain stores content hashes for integrity verification, ownership records, marketplace escrow logic, and governance votes. The frontend is implemented as a Progressive Web App (Next.js 14), accessible at trace.adventurous.systems, with a QR code scanner enabling public passport verification without any blockchain knowledge.

EU DPP-compliant material passport layer

The material passport data model is designed from the ground up for ESPR and CPR compliance. Each passport is structured as JSON-LD aligned with Schema.org and GS1 Web Vocabulary, carrying: a globally unique GS1 GTIN identifier with serial-level granularity; a GS1 Digital Link URI resolvable via QR code scanning; environmental data structured per EN 15804+A2 (including Global Warming Potential); and tiered access controls separating public, professional, and regulatory data views as mandated by the ESPR.

For reclaimed materials entering circular reuse, the core DPP schema is extended with fields capturing: the source

building and deconstruction context, condition assessment scores (structural, aesthetic, environmental), photographic evidence, estimated remaining useful life,

MaterialRegistry (P1: Boundaries) anchors material passport hashes on-chain, providing tamper-evidence for off-chain JSON-LD data. Each registered material

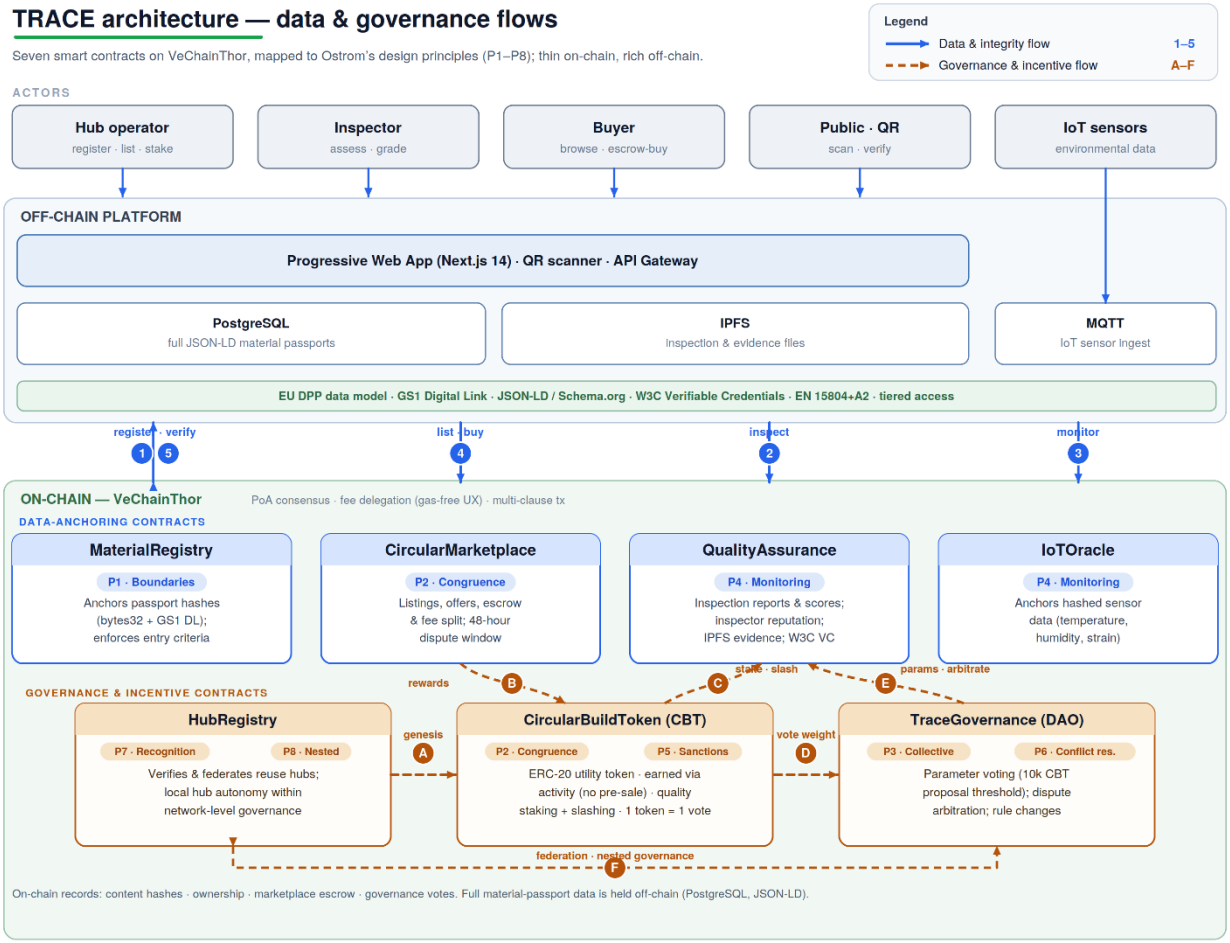


Figure 2: Implemented software architecture (deployed at trace.adventurous.systems)

and critically, the carbon savings achieved by reusing versus manufacturing new—data not captured by the standard DPP framework but essential for circular economy accounting. The extended schema also explicitly captures material-specific technical performance data referenced to applicable standards (EN 1995 for timber, EN 1992 for concrete, EN 1993 for steel), measured residual properties, and recommended use-class constraints, since reclaimed materials must continue to satisfy performance requirements in their new application context. Quality certifications are issued as W3C Verifiable Credentials (VCDM 2.0), enabling cryptographic verification of inspector qualifications and assessment results without requiring trust in a central authority. This directly addresses the ESPR's requirement for conformity assessment while providing a decentralised alternative to the manufacturer-issued Declaration of Performance that does not exist for reclaimed materials.

Smart contract modules

Seven smart contracts implement the on-chain governance layer, each mapped to specific Ostrom principles and DPP requirements (Table 1):

receives a unique bytes32 identifier linked to its GS1 Digital Link URI. The contract enforces entry criteria—minimum documentation requirements that must be met before a material can be registered—implementing the boundary definitions demanded by stakeholders.

CircularMarketplace (P2: Congruence) manages listing creation, offer handling, escrow, and fund release with configurable fee distribution. A 48-hour dispute window after delivery confirmation provides buyer protection. Fee structures are adjustable through governance to ensure costs remain proportional to benefits, addressing the temporal mismatch identified in workshops.

CircularBuildToken (CBT) is an ERC-20 utility token serving as the governance and incentive mechanism. CBT is not a payment currency; material purchases settle in fiat via conventional payment rails. CBT provides governance weight (one token, one vote), quality staking (hubs stake CBT against material quality claims, with slashing for disputes), and activity rewards (minted for registering complete passports, completing transactions, and providing quality reviews). The initial CBT supply is bootstrapped through a genesis allocation to verified founding hubs (Stirling Reuse Hub and network partners),

with subsequent supply minted exclusively through verified on-platform activity—there is no pre-sale or speculative allocation. This design ensures governance weight is earned through participation rather than purchase, directly implementing Ostrom's congruence principle at the token level. The staking mechanism implements Ostrom's graduated sanctions (P5): first disputes result in partial CBT slashing; repeated violations escalate to hub suspension.

QualityAssurance (P4: Monitoring) records inspection reports on-chain with inspector reputation tracking. Each report stores material ID, multi-dimensional quality scores, and an IPFS reference to the full inspection documentation. Inspector reputation accumulates from successful (undisputed) inspections, creating an emergent quality assurance system that does not depend on a single certifying authority—essential for reclaimed materials where no manufacturer exists to issue conformity declarations.

IoTOracle (P4: Monitoring, automated) receives hashed sensor data from environmental monitoring systems (temperature, humidity, structural strain), providing on-chain evidence that storage conditions were maintained. For the prototype, QR code scanning serves as the primary physical-digital bridge, with IoT sensor integration as a secondary capability.

TraceGovernance (P3: Collective Choice, P6: Conflict Resolution) implements a Decentralised Autonomous Organisation for platform parameter voting, dispute arbitration, and community rule changes. A proposal threshold of 10,000 CBT ensures only committed participants can propose changes, while all token holders can vote. This provides the structured collective-choice and conflict-resolution mechanisms that stakeholders identified as absent from current informal reuse practices.

HubRegistry (P7: Recognition of Rights, P8: Nested Enterprises) manages the federation of reuse hubs. Each hub registers on-chain with its operator address and metadata hash, receives platform verification, and operates semi-autonomously within the network. This addresses the regulatory gap where reuse hubs are not recognised as 'economic operators' under the CPR: the HubRegistry creates a parallel institutional recognition system governed by the commons rather than by the state.

Bridging the regulatory gap

The architecture specifically addresses the institutional gap between EU DPP regulations and circular economy practice through three mechanisms. First, *DPP issuance for reclaimed materials*: where the ESPR assigns DPP responsibility to the manufacturer or importer, the TRACE system assigns it to the verified reuse hub, with quality assurance distributed across the inspector network rather than concentrated in a single entity. Second, *conformity assessment through commons governance*: where the CPR requires manufacturer-issued Declarations of Performance, the TRACE system substitutes blockchain-anchored quality assessments issued as W3C Verifiable Credentials by reputation-scored inspectors, governed by DAO-set standards. Third, *federated authority*: where EU regulations assume national market surveillance authorities, the TRACE system implements Ostrom's nested enterprises principle, with local hub autonomy operating within network-level governance—a structure that could complement rather than replace regulatory oversight.

Discussion

The mapping presented in Table 1 reveals a systematic correspondence between Ostrom's governance requirements, stakeholder-identified gaps, EU DPP technical requirements, and blockchain smart contract capabilities. This is not coincidental: the EU DPP framework and Ostrom's principles both address the fundamental challenge of governing shared information resources, but from different institutional perspectives. The EU approach is top-down and regulatory, designed for hierarchical supply chains with identifiable responsible entities. Ostrom's approach is bottom-up and institutional, designed for polycentric governance where authority is distributed across self-organising participants.

The circular economy's material reuse context sits awkwardly between these two models. Materials that have been deconstructed and re-enter circulation require the structured data and consumer protection that EU DPP regulations provide, but they lack the institutional infrastructure (manufacturer, importer, notified body) through which those regulations are implemented. Our findings suggest that Ostrom-governed blockchain infrastructure can fill this gap—not as a replacement for regulatory oversight, but as a complementary institutional layer that enables reuse hubs to issue EU-compliant material passports through a collectively governed, reputation-based system. It is important to be clear on regulatory positioning: the TRACE system does not claim to issue legally binding DPPs under current EU law, where formal authority remains with recognised economic operators. Rather, it demonstrates the institutional and technical architecture that a future regulatory adaptation—such as recognition of 'verified reuse hub' as an economic operator category under CPR delegated acts—would need to build upon. This is consistent with the EU's own approach to DPP implementation, which explicitly anticipates stakeholder-driven standard development through delegated acts. Construction companies managing material reuse independently of hubs would interface with the same schema and smart contract layer, with the HubRegistry federation providing the interoperability layer between actors.

Several limitations must be acknowledged. First, while the prototype is now deployed and operational at trace.adventurous.systems, empirical evaluation at scale remains ongoing; the Stirling Reuse Hub pilot currently underway will provide the first rigorous assessment of adoption, usability, and governance effectiveness in practice.

Table 1: Mapping of Ostrom's design principles to stakeholder evidence, EU DPP requirements, and implemented blockchain features (TRACE platform).

Ostrom Principle	Workshop Evidence	Governance Gap	EU DPP Requirement	TRACE Smart Contract
P1: Boundaries	Material viability hierarchies; demand for clear grading standards	No defined criteria for reuse eligibility	Unique product identifier (GTIN); product classification	MaterialRegistry: entry criteria, passport hash anchoring
P2: Congruence	Cost-benefit temporal mismatch; small hub burden	Documentation costs borne at reclamation, not reuse	Proportionate compliance based on product risk	CBT token: rewards proportional to contribution; fee distribution
P3: Collective choice	All groups want voice in quality standards	No forum for stakeholder input on reuse standards	Delegated acts developed via stakeholder consultation	GovernanceDAO: parameter voting, proposal system
P4: Monitoring	Trust as dominant barrier; post-Grenfell anxiety	No systematic quality verification for reclaimed materials	Conformity assessment; Declaration of Performance	QualityAssurance: inspector reports, reputation; IoTOracle
P5: Graduated sanctions	Need for accountability without excluding marginal suppliers	No penalty mechanism for quality misrepresentation	Market surveillance; penalties for non-compliance	CBT staking/slashing: graduated penalties for disputes
P6: Conflict resolution	Current disputes resolved informally; need structure	No arbitration mechanism for reuse transactions	National market surveillance authorities	GovernanceDAO: arbitration workflow, evidence-based resolution
P7: Recognition of rights	Regulatory uncertainty; building control unfamiliarity	Reuse hubs not recognised as 'economic operators' under CPR	Economic operator registration; EOID	HubRegistry: on-chain hub verification and federation
P8: Nested enterprises	Local vs regional coordination needs; independence desired	Siloed hub operations; no network coordination	Interoperability between DPP systems across sectors	HubRegistry: federated architecture; cross-hub protocols

Second, the regulatory compatibility we describe is architectural (schema compliance, identifier standards) rather than legal—formal recognition of blockchain-issued DPPs under EU law remains an open question. Third, the digital literacy requirements of blockchain interaction present barriers for some reuse hub staff, particularly older volunteers. The prototype addresses this through a 'blockchain-invisible' design where the user interface abstracts away wallet management and on-chain transactions. Fourth, the physical-digital verification gap persists: a blockchain record cannot independently verify that a reclaimed beam has the structural properties claimed. The quality assurance system mitigates but does not eliminate this fundamental limitation.

The broader implication extends beyond construction. As EU DPP requirements expand across product categories through 2030, every sector with significant secondary material flows—textiles, electronics, furniture—will face the same institutional gap: regulation designed for manufacturing supply chains applied to contexts where products re-enter circulation through informal or community-based channels. The commons governance approach demonstrated here may offer a transferable institutional model.

Conclusion

This paper has presented a theoretically grounded approach to governing construction material reuse as a digital commons, bridging Ostrom's design principles with the EU's emerging Digital Product Passport regulatory framework through blockchain smart contracts. Our contribution is twofold. First, we provide empirical evidence from 36 stakeholder workshop participants that the governance challenges of construction material reuse map systematically onto Ostrom's eight design principles, identifying assurance, logistics, and information systems as the three critical gaps. Second, we present a deployed seven-contract prototype architecture, implemented on VeChainThor and accessible at trace.adventurous.systems, that is simultaneously EU DPP-compliant (JSON-LD, GS1 Digital Link, W3C Verifiable Credentials, EN 15804+A2) and commons-governed (DAO voting, reputation-based quality assurance, graduated sanctions, federated hub architecture).

The key finding is that EU DPP regulations and Ostrom's governance principles are not competing frameworks but complementary layers: the regulations define *what* data a material passport must contain, while Ostrom's principles define *how* the institutional infrastructure for issuing and

governing those passports should be organised in contexts—such as material reuse—where the regulatory assumption of a responsible manufacturer does not hold.

The TRACE prototype is now deployed and operational at trace.adventurous.systems, with ongoing pilot evaluation at Stirling Reuse Hub providing empirical assessment of whether blockchain-enabled commons governance can deliver the trust, traceability, and coordination that stakeholders identified as prerequisites for scaling construction material reuse in Scotland and beyond.

Figure 3 presents the public landing page of the running platform. The deployed system operationalises the architecture described above through a deliberately blockchain-invisible interface: a visitor can browse the materials marketplace, register or list reclaimed materials, or scan a passport QR code to verify a material's integrity against its on-chain hash—none of which requires a cryptocurrency wallet or any blockchain knowledge. The landing page foregrounds the three commitments that distinguish the platform—EU Digital Product Passport compliance, blockchain-anchored integrity on VeChainThor, and circular-economy traceability of deconstruction origin, condition grade, carbon savings and reuse history—while the seven-contract governance layer of Figure 2 operates entirely beneath this surface. To our knowledge, this is among the first publicly accessible material-passport platforms to pair EU DPP-aligned data structures with an Ostrom-grounded commons-governance layer, and it constitutes the live system on which the Stirling Reuse Hub pilot is now being evaluated.

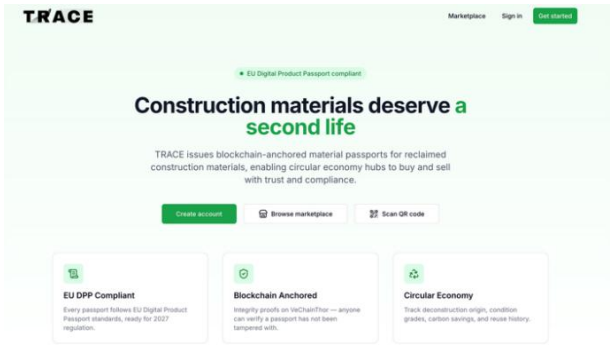


Figure 3: The deployed TRACE platform (trace.adventurous.systems): public landing page presenting EU DPP-compliant, blockchain-anchored material passports for reclaimed construction materials.

Acknowledgements

This research was funded through Scotland Beyond Net Zero. The authors thank Stirling Reuse Hub and the workshop participants for their collaboration. TD and MV conceptualised the research, MB and ZJ conducted the workshops and drafted the first findings, while TD developed the classification per Ostrom and EU regulations. TD and HMcD conceptualised the contract architecture and developed the platform.

References

Dounas, T, Jabi, W and Lombardi, D (2021) Topology generated non-fungible tokens: Blockchain as infrastructurefor a circular economy in architectural design In: CAADRIA 2021: Projections, 2021-3-29 - 2021-4-1.

European Commission (2024a). Regulation (EU) 2024/1305 of the European Parliament and of the Council laying down harmonised conditions for the marketing of construction products (revised CPR). Official Journal of the European Union.

European Commission (2024b). Regulation (EU) 2024/1781 of the European Parliament and of the Council establishing a framework for setting ecodesign requirements for sustainable products (ESPR). Official Journal of the European Union.

Fuster Morell, M. (2014). Governance of online creation communities for the building of digital commons. In Frischmann, B., Madison, M. & Strandburg, K. (eds.) Governing Knowledge Commons. Oxford University Press.

Gioia, D.A., Corley, K.G. & Hamilton, A.L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. Organizational Research Methods, 16(1), 15–31.

GS1 in Europe (2024). GS1 Standards Enabling the EU Digital Product Passport. Brussels: GS1 in Europe.

Hevner, A.R., March, S.T., Park, J. & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.

Ostrom, E. (1990). Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press.

Rozas, D., Tenorio-Fornés, A., Díaz-Molina, S. & Hassan, S. (2021). When Ostrom meets blockchain: Exploring the potentials of blockchain for commons governance. SAGE Open, 11(1).

Ruhaak, A., Bloom, G., Motz, G. et al. (2021). A Practical Framework for Applying Ostrom's Principles to Data Commons Governance. Mozilla Foundation.

Schweik, C.M. & English, R.C. (2012). Internet Success: A Study of Open-Source Software Commons. MIT Press.

Stanford d.school (2010). An Introduction to Design Thinking Process Guide. Hasso Plattner Institute of Design at Stanford.

UNEP (2022). 2022 Global Status Report for Buildings and Construction. United Nations Environment Programme.

Zero Waste Scotland (2024). Circular Economy in the Built Environment: Material Flows Analysis. Edinburgh: Zero Waste Scotland.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

BLOCKCHAIN-ENABLED REAL ESTATE OWNERSHIP TRANSFER IN BELGIUM: A HYBRID INSTITUTIONAL FRAMEWORK

Aishuak Kerimzhanova[1], Theodoros Dounas[2]
[1]University of Antwerp, Antwerp, Belgium
[2]Heriot-Watt University, Edinburgh, Scotland, UK

Abstract

This paper addresses the tension between blockchain's technical potential and the need for legal compliance in highly regulated property markets. Drawing on Institutional Cryptoeconomics, we propose a hybrid governance architecture for real estate transactions in Belgium that complements, rather than replaces, existing legal frameworks. The system combines smart contracts for automated escrow and settlement, an AI-based assistant to reduce ex ante coordination costs, and civil notaries as institutional oracles for legal validation. By embedding algorithms within mandated structures, this "trustworthy" governance model aims to reduce transaction costs while ensuring regulatory alignment, offering a scalable template for adoption in real estate.

Introduction

Real estate transactions are among the most institutionally complex forms of economic exchange, characterised by high asset values, long settlement periods, illiquidity, and substantial information asymmetry (Garmaise & Moskowitz, 2004). As a result, buyers and sellers rely on third parties who perform authentication functions and act as trust intermediaries. Although these intermediaries generate institutional trust, they also impose high transaction costs and cause significant time delays (Nooteboom, 2005).

Blockchain technology has been widely proposed as a technical solution to real estate inefficiencies through its immutable records, transparent distributed ledger, and automated transactions enabled by smart contracts (Konashevych, 2021). Early academic and industry discussions framed blockchain as a "trustless" technology capable of disintermediating traditional institutions (The Economist, 2015).

However, empirical evidence from pilot projects and case studies increasingly demonstrates that fully decentralised models are difficult to integrate with the legal and regulatory frameworks governing real estate markets. Notably, a famous example of the Swedish blockchain and land registry initiative, although technologically successful, has been abandoned in favour of a non-blockchain solution, showcasing that technical feasibility alone is insufficient for institutional adoption (Proskurovska & Dörry, 2022).

Although prior studies have explored the use of blockchain technology in land administration and property transactions, they primarily focus on technical feasibility or propose universal solutions. As a result, limited attention has been given to how such systems can be institutionally embedded to complement, rather than replace, mandatory legal frameworks. This limitation is particularly important, as blockchain-based solutions cannot yet be uniformly applied across jurisdictions; real estate regulations vary significantly even among EU member states. Furthermore, the growing importance of circular real estate and circular construction introduces additional requirements for data continuity and lifecycle management that blockchain can support (Shojaei & Naderi, 2024).

To address this gap, this paper adopts the perspective of Institutional Cryptoeconomics (IC), which conceptualises blockchain not merely as a technical tool but as an institutional technology capable of reducing transaction and coordination costs within an economy (Allen et al., 2020). Drawing on the Transaction Cost Economics (TCE) theory of Williamson (1985), blockchain is viewed as a mechanism that reduces ex post enforcement and verification costs through smart contracts, while some ex ante costs, such as drafting, information processing, and negotiation, remain unresolved.

Building on this theoretical foundation and a gap in the literature, this study develops a hybrid "trustworthy" governance framework for blockchain-enabled real estate transactions in Belgium. The framework aims to provide security, transparency, and automation through blockchain immutability and smart contract execution, while assigning legal certainty and trust to Belgian civil law notaries as institutional oracles. Additionally, the framework introduces an AI-based assistant to reduce ex ante transaction costs associated with bounded rationality and procedural complexity.

Point of departure

This study builds on three interconnected literature topics that together motivate the need for a hybrid, institutionally grounded approach to blockchain-enabled real estate transactions. First, prior research on blockchain in real estate and land administration reveals a consistent gap between technical feasibility and real-world adoption. Second, the literature on trust demonstrates that trust is

not merely a social construct or a technological outcome but an institutionally produced condition. Third, IC conceptualises blockchain as an institutional technology that can reduce coordination and transaction costs within an economy. Together, these three topics feed into the central proposition of this study that blockchain-based real estate systems require institutional embedding rather than purely technical design.

Blockchain technology in real estate and land administration

The application of blockchain technology in real estate and land administration has been a growing topic both in academic research and industry practice. Existing studies suggest that blockchain can address many persistent challenges in real estate, such as inefficiency, fraud, corruption, and high transaction costs (Allessie et al., 2019; Kempe, 2017; Konashevych, 2021). These challenges are commonly associated with unreliable or missing land registrations, lengthy administrative processes, and the structural illiquidity of real estate assets (Shuaib et al., 2020; Wong et al., 2012). Due to the manual, paper-based, and fragmented nature of records, they are prone to errors, fraud, and corruption (Ameyaw & de Vries, 2021).

Distributed Ledger Technology (DLT) has been proposed as a technical solution to reduce these inefficiencies by enabling immutable records and automated transaction execution via smart contracts (Konashevych, 2021). Empirical studies suggest that DLTs can reduce information asymmetries between market participants and improve transparency (Chen & Yu, 2024). Several frameworks have been proposed to operationalise these claims. For example, Shuaib et al. (2020) present an 11-step ownership transfer model, while Verma et al. (2024) and Mohanty et al. (2025) develop blockchain frameworks for property registration and operations, reporting improvements in security, efficiency, and transparency. Bărbuţă & Alexandrescu (2024) further extend the design space by proposing dNFTs as updatable on-chain property representations.

However, recent academic research increasingly highlights that blockchain adoption faces greater constraints from legal and regulatory compliance and the complexity of system integration than from technical feasibility alone (Ajayi et al., 2026; Akhmetbek & Špaček, 2021). Podshivalov (2022) similarly argues that blockchain implementation in real estate registration should be introduced gradually alongside the development of appropriate legal and regulatory frameworks. Case studies of pilot projects show that fully decentralised systems rarely align with the legal frameworks governing real estate markets (Proskurovska, 2023). A notable example is the Swedish Lantmäteriet initiative, which was the first to test blockchain technology for property conveyance but did not advance beyond the pilot phase. Instead, the system was replaced by a non-blockchain platform, Tambur (Proskurovska & Dörry, 2022). Proskurovska (2023) explains that this results from a mix of resistance from existing institutions and incompatibilities with current strategies aimed at financial stability.

These findings thus support a shift from technological optimism toward institutionally embedded adoption strategies. Rather than advocating full-scale disruption, several authors argue for a hybrid approach rather than a full-scale process (Allessie et al., 2019; Bennett et al., 2021; Saari et al., 2022).

Trust

Across the literature, trust was found to influence market confidence and participation directly (Gurun & Booth, 2024), which in turn affects price formation (CFA Institute, 2022). While Nakamoto (2008) introduced blockchain as a technology that does not require trust, in which users rely on cryptographic protocols, consensus mechanisms, and immutable ledgers, subsequent research has challenged this assumption. Blockchain can facilitate a form of institution-based trust by increasing transparency and enabling automated execution (Chen & Yu, 2024; Lumineau et al., 2021). However, it does not eliminate the need for trust altogether. Instead, trust remains embedded in social, legal, and organisational contexts (Szabo, 2017). Furthermore, trust was found to be a crucial prerequisite for blockchain adoption rather than its outcome (Dupuis et al., 2021). Adoption depends on socio-cultural factors (Mutambik et al., 2024), governance structures (Li et al., 2023), and the overall usability of blockchain-based applications (Smits & Hulstijn, 2020).

In land administration, legal and regulatory compliance has been identified as a key factor in stakeholder trust (Ajayi et al., 2026). This suggests that algorithms alone cannot produce trust in blockchain-enabled real estate transactions; they must be anchored in institutional validation and legal enforceability.

Institutional cryptoeconomics and hybrid governance

IC conceptualises blockchain not only as a database technology but also as a new institutional technology for governing the “knowledge commons” (Berg et al., 2019). IC applies the principles of New Institutional Economics (Coase, 1937; Williamson, 2000) to explain how blockchain technology reduces the “costs of trust” and coordination within an economy. The central principle of IC is the reduction of transaction costs (Davidson & Potts, 2022). Transaction costs arise from three primary constraints: bounded rationality, opportunism, and asset specificity (Williamson, 1985). Berg et al. (2019) demonstrate that blockchain technology can be analysed through Williamson’s (1985) framework. While smart contracts can reduce ex post costs, such as enforcement and monitoring, through automated execution and transparency, they do not eliminate ex ante costs related to drafting, negotiation, and information processing. Moreover, a smart contract designer might still engage in fraudulent activities during the design process (Davidson & Potts, 2022), giving rise to the so-called "oracle problem” (Caldarelli, 2020). Recent research, therefore, highlights the emergence of hybrid governance models, in

which algorithmic mechanisms coexist with social institutions (Bennett et al., 2021; Lumineau et al., 2021; Saari et al., 2022). Such models combine the efficiency of automation with the legitimacy of traditional institutions. In real estate, this hybridisation is not optional but structurally required due to the legal, high-value, and high-risk nature of ownership rights.

Institutional context

EU regulatory framework

The European Commission has adopted several legislative measures to regulate crypto-assets and protect consumers and investors (EU, 2026). Crypto-assets classified as financial instruments are regulated under MiFID II and must comply with the same rules as traditional financial instruments (ESMA, 2025). Other crypto-assets are governed by the Markets in Crypto-Assets Regulation (MiCA). Data protection remains a major concern, especially regarding compliance with the General Data Protection Regulation (GDPR) and the "right to be forgotten" (Zafar, 2025). The European Data Protection Board (EDPB) has addressed this by issuing guidelines on processing personal data through blockchain technologies (EDPB, 2025): storing personal data directly on the blockchain should be avoided, except for identifiers or proof of identity. Personal data should be stored off-chain, with cryptographic proof on the blockchain. This approach allows the "right to be forgotten" to be enforced by deleting data off-chain.

Belgian real estate law

The main regulations governing the purchase of real estate are set out in the Belgian Civil Code, particularly Book 3, entitled "Assets," as of 1 September 2021. Additionally, the transaction process follows the general rules of contract law in Book 5, entitled "Obligations". The real estate ownership and registry system in Belgium was described in detail by van de Werve (2025) in Deloitte's Legal Handbook:

The process of purchasing property in Belgium generally involves two main steps: signing a private sales agreement and executing a deed before a Belgian civil-law notary. Within four months of signing the private sales contract, it must be registered with the mortgage registry office and the land registry (Kadaster). The registered titles in Kadaster are for informational purposes only and do not serve as confirmation of ownership. Civil law notaries play a crucial, legally mandated role in Belgian real estate transactions. Their duties include validating the transfer deed, verifying the identities of the parties, collecting registration fees, ensuring there are no encumbrances, and confirming compliance with urban planning and land use laws. The Royal Decree strictly regulates Belgian notarial fees, which are consistent across regions and are set on a proportionally progressive scale based on the property's value.

The process of real estate purchase and registration in Belgium differs from, for example, Sweden, where real estate transactions do not require notarization (Fjellström, 2025). Instead, their system is heavily dependent on real estate brokers that mediate over 90% of all housing sales in Sweden (Proskurovska & Dörry, 2022). Notably, during the transaction, they are typically granted the power of attorney, which allows them to access and handle personal data of the involved parties (Proskurovska, 2023). These differences underline the importance of national legal systems in the development of blockchain-enabled real estate transaction solutions.

Methodology

This study employs a Design Science Research (DSR) methodology to develop a hybrid governance framework for blockchain-enabled real estate transactions in Belgium. Following Hevner et al. (2004) and Peffers et al. (2007), DSR is appropriate as it focuses on creating innovative artefacts that address identified problems in real-world contexts. This paper covers Phases 1-3: problem identification, objectives definition, and artefact design, leaving development, demonstration, and evaluation for future work.

DSR phases 1-2: problem identification and objective definition

Prior blockchain implementations in land administration, notably Sweden's Lantmäteriet pilot, demonstrate that technical feasibility alone is insufficient for institutional adoption (Proskurovska & Dörry, 2022). The problem is further complicated by three factors: high transaction costs resulting from information asymmetry and intermediary fees, regulatory requirements mandating notarial involvement in Belgian property transfers, and the absence of frameworks that meaningfully integrate blockchain with existing legal structures rather than attempting to replace them.

This phase involved a literature review across three domains: blockchain applications in real estate and land administration, trust mechanisms in property transactions, and hybrid governance models. The review reveals a critical gap between fully decentralised "trustless" blockchain implementations and the regulatory realities of property markets, motivating the development of a hybrid institutional architecture.

DSR phase 3: artefact design

The artefact is a hybrid governance architecture comprising technical components, institutional interfaces, and procedural workflows, organised into four interconnected activities: Institutional Actor Modelling, Transaction Workflow Architecture, Technical Component Specification, and Regulatory Compliance.

Proposed hybrid transaction architecture

The proposed framework integrates three governance components:

(1) blockchain and smart contracts for automated execution and record-keeping, (2) an AI-assistant for information processing and coordination, and (3) civil law notaries as institutional oracles responsible for legal validation and enforcement. The platform is governed as a permissioned consortium, with node validation rights

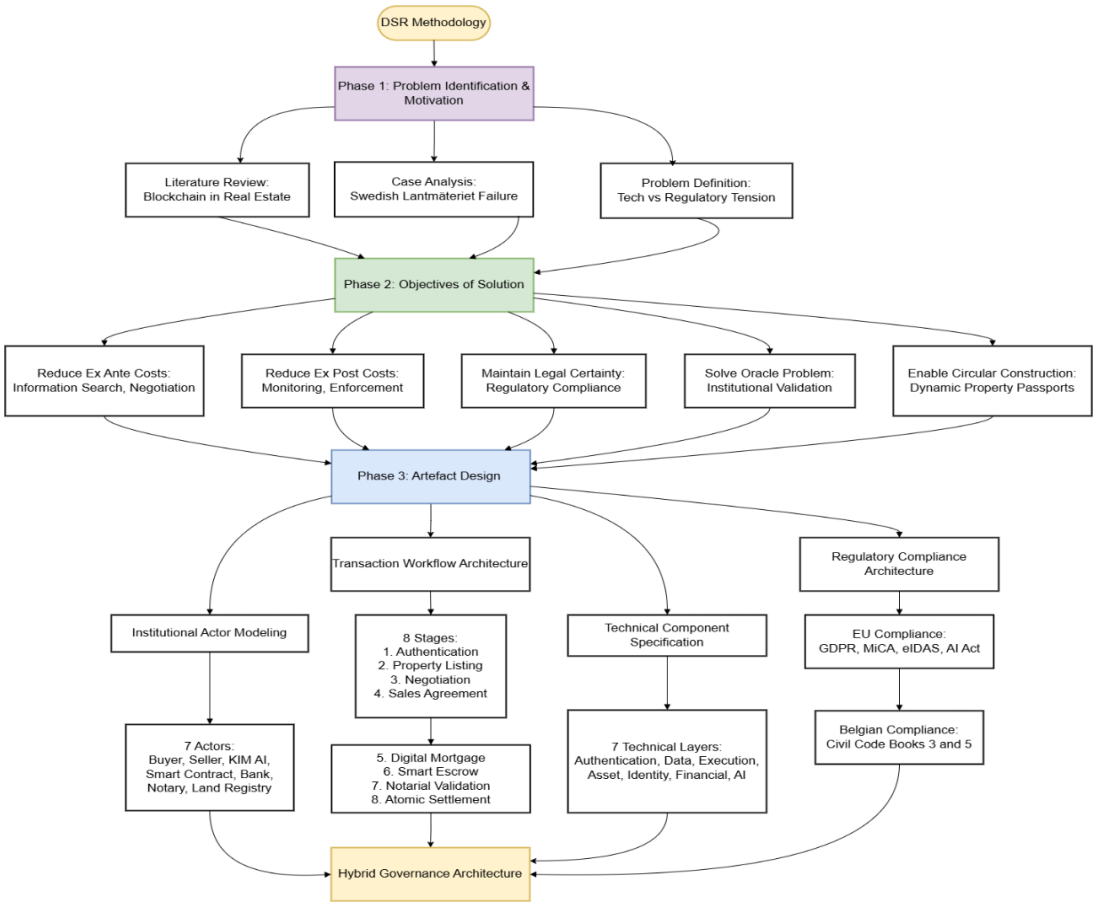


Figure 1: Phases 1-3 of the DSR methodology employed for this paper.

restricted to a defined set of institutional participants. In the Belgian context, the proposed consortium comprises civil law notaries represented through the Federation of Notaries (Fednot), the land registry, and participating mortgage banks. Platform operation and KIM's infrastructure would be managed by a neutral technical operator under contractual obligations to the consortium.

Figure 2 illustrates the sequence diagram of the proposed eight-stage transaction workflow. The framework involves the following actors:

- **Buyer (Alice) and Seller (Bob)**: the primary transacting parties.
- **AI-based assistant (KIM - Kunstmatige Intelligentie Makelaar):** KIM operates through a perceive-reason-act cycle without requiring explicit human instruction at each stage. Its agentic capabilities include: (1) independent calls to government APIs, including the Kadaster, soil databases, regional EPC registries, and municipal urban planning systems; (2) multi-source synthesis of retrieved data; (3) document generation, such as standardised property listings and the Private Sales Agreement (PSA). By automating property searches, data retrieval, document preparation, and eligibility coordination, which are currently performed by real estate agents, KIM could reduce their role and, in turn, lower ex ante costs associated with bounded rationality and information asymmetry (Williamson, 1985). Importantly, KIM's outputs are decision support, not autonomous determinations – all binding actions require a notary's authorisation. To ensure compliance with the EU AI Act (Regulation (EU) 2024/1689), KIM is architecturally restricted to a "limited-risk" or "transparency-obligated" AI system.
- **Smart contract (DLT)**: the execution layer that governs the escrow, suspensive conditions, and atomic settlement, deployed on a permissioned consortium blockchain.
- **Mortgage bank**: responsible for credit risk assessment and issuance of MiCA-compliant E-Money Tokens (EMTs).
- **Notary**: acts as an institutional oracle and legal authority.
- **Land registry and mortgage Office**: maintains off-chain records of ownership and encumbrances.

Stage 1: authentication

To ensure non-repudiation and compliance with Belgian AML/KYC requirements, the framework integrates the Belgian digital identity solution, Itsme, as the primary authentication mechanism. Under eIDAS (Regulation (EU) No 910/2014), a Qualified Electronic Signature (QES) generated via Itsme has the same legal validity as a handwritten signature. Both Alice and Bob authenticate

via Itsme, each receiving cryptographic credentials derived from their respective Itsme-based verifications. These credentials enable participants to generate zero-knowledge proofs (ZKPs) demonstrating that they are legally authorised to participate in property transactions without revealing their personal data to the platform or the blockchain.

Stage 2: property listing and data verification

Bob submits property information to the platform. KIM verifies ownership status through the land registry's API and retrieves certified property-related data from governmental databases, including cadastral records, energy performance certificates (EPCs), soil reports, and urban planning documentation. Based on these inputs, KIM helps Bob to generate a standardised property listing. For Alice, KIM provides decision support by filtering properties according to her preferences and explains the necessary steps and the legally complex documentation.

Stage 3: viewing and price negotiation

KIM schedules a property viewing for Alice using Bob's preferences, ensuring that both Alice and Bob are authorised participants before the address is shared to reduce security risks. If Alice submits a bid, KIM initiates a privacy-preserving eligibility verification workflow in which Alice interacts directly with the mortgage bank's secure system via Itsme. The Mortgage Bank, after an internal off-chain evaluation, sends the signed via Itsme attestation to the platform, reducing opportunism and lowering risk.

Stage 4: private sales agreement

Once Bob accepts Alice's bid via the platform, KIM generates a Private Sales Agreement (PSA) with the agreed price and conditions. Both Bob and Alice sign the PSA via Itsme. A cryptographic hash of the signed PSA is anchored on the blockchain to prevent both Bob from taking a higher offer after agreeing to Alice's or Alice from lowering her offer at the last minute.

Stage 5: digital mortgage

Alice formally applies for a mortgage. Once approved, the Mortgage Bank issues a conditional payment commitment to the smart contract. Funds are represented as EMTs, ensuring regulatory alignment with EU law.

Stage 6: smart escrow

After the PSA is signed and mortgage approval is granted, Alice sends the deposit to the smart contract escrow in EMTs. The Mortgage Bank transfers the remaining amount to the same smart contract escrow in EMTs. The funds are released only when the notary signs the deed off-chain and authorises execution of the smart contract. This mechanism reduces ex post enforcement costs by automating escrow management while preserving legal safeguards.

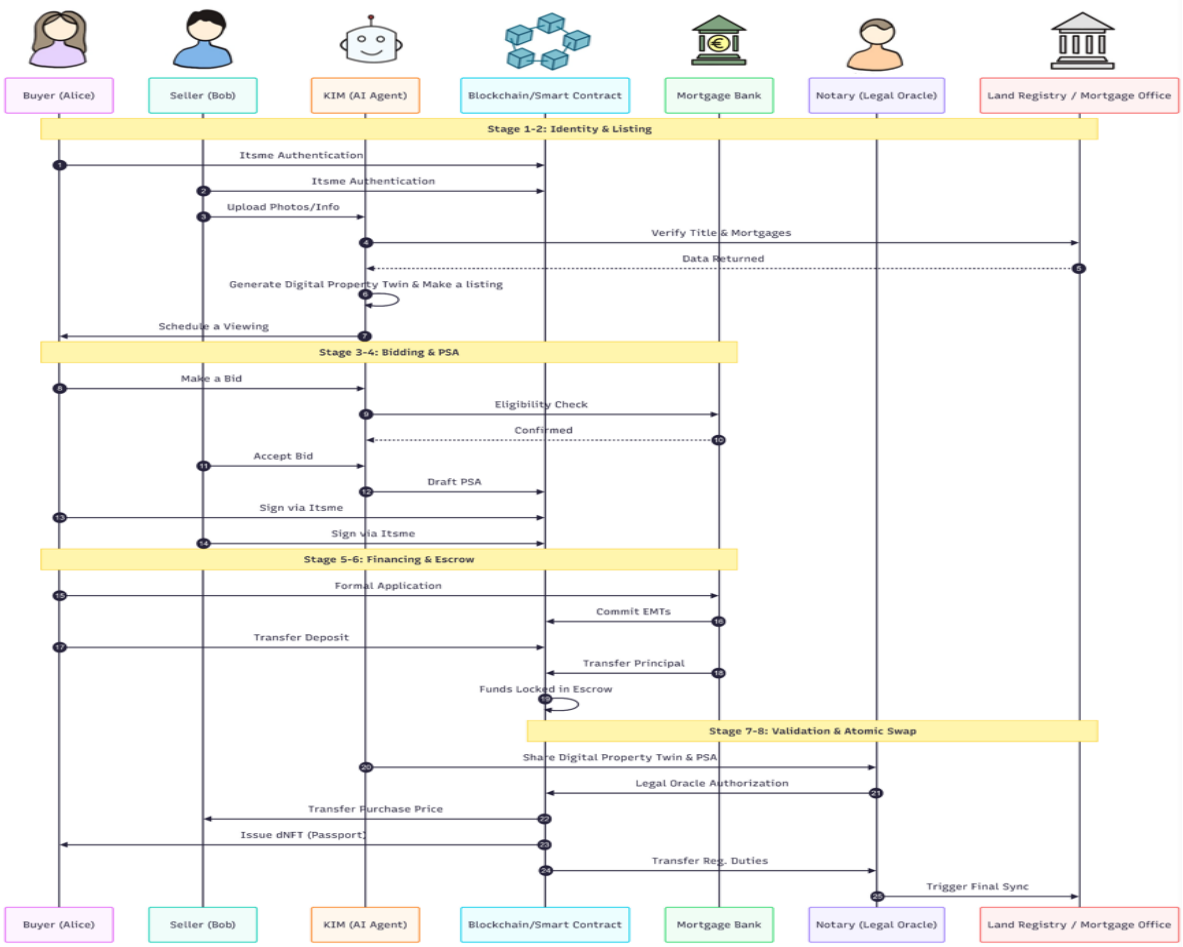


Figure 2: Sequence diagram of proposed property transfer architecture.

Stage 7: notarial validation

The civil law notary receives the "Digital Property Twin" from KIM, stored off-chain via InterPlanetary File System (IPFS). However, the notary remains responsible for verifying the data provided by KIM and performing other mandated checks that require human judgment and institutional authority. Upon completion, the notary issues an on-chain authorisation signal to the smart contract. The oracle problem is addressed through institutional legitimacy.

Stage 8: atomic settlement

Upon notarial authorisation, the smart contract executes an atomic settlement: Bob receives payment, the notary receives the registration fee, and Alice receives a dynamic Non-Fungible Token (dNFT) and a physically signed deed representing ownership. Unlike static hashing, which fails to capture the dynamic nature of a real estate asset, dNFT can be updated and thus functions as a Property Passport. Off-chain data can be erased in compliance with GDPR requirements, while the DLT retains cryptographic proofs. The idea of using dNFTs for real estate assets was mentioned by Bărbuță & Alexandrescu (2024). Simultaneously, the system notifies the official land registry and mortgage office to ensure synchronisation between the blockchain ledger and the off-chain records.

Discussion and conclusion

This study proposed a hybrid framework for blockchain-enabled real estate transactions in Belgium that embeds blockchain technology within existing legal structures rather than disintermediating legally mandated institutions. Existing research shows that blockchain achieves meaningful impact when integrated into legal and governance frameworks rather than disrupting them (De Filippi & Wright, 2018). By assigning operational and informational tasks to algorithms and validation functions to notaries, the framework addresses specific inefficiencies while maintaining legal certainty. In line with Saari et al. (2022), who emphasised that blockchain implementation in land administration will likely take a hybrid form. The framework employs a permissioned consortium blockchain, restricting validation to civil law notaries represented through Fednot, the land registry and mortgage office, and participating banks. This structure ensures that no single private entity controls the ledger while enabling faster transaction processing and controlled governance (Peters & Panayi, 2016).

Drawing on Williamson's (1985) TCE, the framework addresses both ex ante and ex post transaction costs. In Belgium, real estate agent commissions typically range from 2% to 5% of the property's value, and the average settlement period is three to four months. Ex ante, KIM reduces these costs by automating real estate agent functions, including property searches, data retrieval, document preparation, and eligibility coordination. Importantly, KIM does not require the elimination of real estate agents; buyers and sellers who prefer professional mediation can still choose that option. Instead, the platform lowers the barrier for those who cannot afford or decide not to use an agent, thereby reducing transaction costs more significantly. KIM's outputs are strictly decision support: all outputs are subject to notarial verification, as there is a risk of errors and inaccuracies inherent to LLM-based systems. Ex post, smart contracts and atomic settlement eliminate enforcement costs associated with escrow and manual fund transfers (Davidson & Potts, 2022), while consortium consensus reduces fraud risk and balances efficiency with institutional trust (Kowalski et al., 2021).

The framework proposes a trustworthy environment where algorithmic trust is anchored in institutional trust through notarial oversight. Moreover, the architecture uses dNFTs that, in combination with Digital Twins, could enable longitudinal tracking of construction materials, maintenance logs, and energy performance certificates, contributing to circular real estate and circular construction (Naderi & Shojaei, 2024). Crucially, dNFTs represent a data layer rather than actual ownership and do not require legal equivalence to a notarial deed.

The framework has several limitations. First, it has not yet been tested in real-world deployments, and its performance with high-volume transactions remains unverified. Second, the long-term legal interpretation of the use of dNFTs in property law continues to evolve. Third, relying on notarial oversight might limit scalability in other countries with different legal systems. Fourth, KIM introduces risks specific to LLM-based agent systems, such as outputs that depend on the availability and accuracy of government APIs, residual hallucination risks that cannot be fully eliminated, and user over-reliance on AI-generated documents, which could weaken informed consent. Mandatory notarial verification partly addresses these risks but also creates a dependency on notaries' digital literacy, a skills gap that needs targeted professional development as part of the deployment plan.

Future work will validate the framework through prototype development and multi-dimensional evaluation: (1) technical performance, assessing system reliability, latency, and interoperability; (2) transaction cost analysis, comparing ex ante and ex post costs against traditional Belgian workflows; (3) stakeholder evaluation, conducted through semi-structured interviews with key actors in the Belgian real estate ecosystem, using a protocol grounded in the Technology Acceptance Model (TAM), trust assessment constructs, and perceived adoption barriers (Davis, 1989); and (4) circular construction extensibility, exploring how the dNFT model can support multi-stakeholder material provenance tracking and lifecycle management.

In conclusion, this paper offers a legally grounded template for blockchain-enabled real estate transactions that does not compromise institutional integrity. By embedding algorithms with mandated structures rather than replacing them, this framework provides actionable insights for public authorities seeking to combine technological innovation with regulatory compliance in property markets.

Acknowledgments

This project has received funding from the European Union's Horizon Europe 2021-2027 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 101169048 (QuiVal). Views and opinions expressed are, however, those of the author only and do not necessarily reflect those of the European Union or REA. Neither the European Union nor the granting authority can be held responsible for them.

References

Ajayi, O., Kirsikka, R., Adekola, O. (2026). User-centric dimensions of trust and acceptance in blockchain adoption for land administration systems: A systematic literature review. Land Use Policy, 164, 107944.

Akhmetbek, Y., Špaček, D. (2021). Opportunities and Barriers of Using Blockchain in Public Administration: The Case of Real Estate Registration in Kazakhstan. The NISPAcee Journal of Public Administration and Policy, 14(2).

Allen, D.W.E., Berg, C., Markey-Towler, B., Novak, M., Potts, J. (2020). Blockchain and the evolution of institutional technologies: Implications for innovation policy. Research Policy, 49(1), 103865.

Allessie, D., Sobolewski, M., Vaccari, L., Pignatelli, F. (2019). Blockchain for digital government: An assessment of pioneering implementations in public services. Publications Office of the European Union.

Ameyaw, P.D., de Vries, W.T. (2021). Toward Smart Land Management: Land Acquisition and the Associated Challenges in Ghana. A Look into a Blockchain Digital Land Registry for Prospects. Land, 10 (3), 239.

Bărbuţă, D.-E., Alexandrescu, A. (2024). A Secure Real Estate Transaction Framework Based on Blockchain Technology and Dynamic Non-Fungible Tokens. International Conference on System Theory, Control and Computing.

Bennett, R.M., Pickering, M., Kara, A.-K. (2021). Hybrid approaches for smart contracts in land administration: Lessons from three blockchain proofs-of-concept. Land, 10 (2), 220.

Berg, C., Davidson, S., Potts, J. (2019). Understanding the Blockchain Economy: An Introduction to Institutional Cryptoeconomics. Edward Elgar Publishing.

Braun, V., Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3 (2), pp.77–101.

Caldarelli, G. (2020). Understanding the Blockchain Oracle Problem: A Call for Action. Information, 11 (11), 509.

CFA Institute. (2022). Guide to 2022 CFA Program Curriculum Changes. CFA Institute.

Chen, Y., Yu, Z. (2024). Digitalization, trust, and sustainability transitions: Insights from two blockchain-based green experiments in China's electricity sector. Environmental Innovation and Societal Transitions, 50, 100801.

Coase, R.H. (1937). The Nature of the Firm. Economica, 4 (16), pp.386–405.

Davidson, S., Potts, J. (2022). Institutional Cryptoeconomics. In: The Economics of Blockchain and Cryptocurrency, pp.1–12.

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13 (3), pp. 319–340.

De Filippi, P., Wright, A. (2018). Blockchain and the Law: The Rule of Code. Harvard University Press.

Dounas, T., Hunhevicz, J., Byers, B. (2023). Design dimensions for blockchain oracles in the AEC industry. EC3 Conference Proceedings. DOI: 10.35490/EC3.2023.297.

Dupuis, I., Toohey, L., Grimstad, S., Follong, B., & Bucher, T. (2021). Blockchain: The Paradox of Consumer Trust in a Trustless System - a Systematic Review. 2021 IEEE International Conference on Blockchain, pp. 505–512.

European Data Protection Board. (2025). Guidelines 02/2025 on processing of personal data through blockchain technologies.

European Securities and Markets Authority. (2025). MiFID II. Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments and amending Directive 2002/92/EC and Directive 2011/61/EU.

European Union. (2014). Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC.

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

European Union. (2026). Crypto-assets: how the EU is regulating markets. Available at: https://www.consilium.europa.eu/en/policies/crypto-assets-how-the-eu-is-regulating-markets.

Fjellström, J. (2025). Real Estate Law in Sweden. In: Legal Handbook for Real Estate Transactions. Deloitte Legal.

Garmaise M.J., Moskowitz T.J. (2004). Confronting Information Asymmetries: Evidence from Real Estate

Markets. Review of Financial Studies, 17 (2), pp. 405–437.

Gurun, A., Booth, G.G. (2024). The Importance of Trust to Financial Markets and Institutions: Exploring the Current Literature. Theoretical Economics Letters, 14, pp. 27–40.

Hevner, A.R., March, S.T., Park, J., Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28 (1), pp. 75–105.

Kempe, M. (2017). The land registry in the blockchain – testbed. A development project with Lantmäteriet, Landshypotek Bank, SBAB, Telia company, ChromaWay and Kairos Future. Stockholm: Kairos Future.

Konashevych, O. (2021). "GoLand Registry" case study: Blockchain/DLT adoption in land administration in Afghanistan. DGO2021: The 22nd Annual International Conference on Digital Government Research. ACM, pp. 489-494.

Kowalski, M., Lee, Z. W. Y., & Chan, T. K. H. (2021). Blockchain technology and trust relationships in trade finance. Technological Forecasting and Social Change, 166, 120641.

Li, Y., Liao, A., Li, L., Zhang, M., Zhao, X., & Ye, F. (2023). Reinforcing or weakening? The role of blockchain technology in the link between consumer trust and organic food adoption. Journal of Business Research, 164, 113999.

Lumineau, F., Wang, W., Schilke, O., & Huang, L. (2021). How Blockchain Can Simplify Partnerships. Harvard Business Review Digital Articles.

Mohanty, A.K., Rao, D.C., Rath, A.K. (2025). A Comprehensive Framework for Facilitating Real Estate Operations Employing Smart Contracts on the Blockchain. Journal of the Institution of Engineers (India): Series B, 106, pp. 1033–1049.

Mutambik, I., Almuqrin, A., Zhang, J. Z., & Alharbi, Z. H. (2024). Trust in Cryptocurrency Payments: Journal of Organizational and End User Computing, 36(1).

Naderi, H., Shojaei, A. (2024). Digital twin non-fungible token (DT-NFT): Enabling data ownership in the AEC industry. Automation in Construction, 168 (A), 105777.

Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. Satoshi Nakamoto Institute.

Nooteboom, B. (2005). Innovation, Learning and Cluster Dynamics. Clusters and Regional Development, pp. 155–181.

Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24 (3), pp. 45–77.

Podshivalov, T.P. (2022). Improving implementation of the Blockchain technology in real estate registration. Journal of High Technology Management Research, 33, 100440.

Proskurovska, A. (2023). Re-inventing housing finance with blockchain. The case of Sweden. Geoforum, 147, 103884.

Proskurovska, A., Dörry, S. (2022). The blockchain challenge for Sweden's housing and mortgage markets. Environment and Planning A: Economy and Space, 54 (8), pp. 1569-1585.

Saari, A., Vimpari, J., Junnila, S. (2022). Blockchain in real estate: Recent developments and empirical applications. Land Use Policy, 121, 106334.

Shojaei, A., Naderi, H. (2024). Blockchain Technology for a Circular Built Environment. A Circular Built Environment in the Digital Age.

Shuaib, M., Alam, S., Daud, S. (2020). Improving the Authenticity of Real Estate Land Transaction Data Using Blockchain-Based Security Scheme. International Conference on Advances in Cybersecurity 2020.

Shuaib, M., Daud, S.M., Alam, S., Khan, W.Z. (2020). Blockchain-based framework for secure and reliable land registry system. Telekomnika, 18(5), pp. 2560–2571.

Smits, M., & Hulstijn, J. (2020). Blockchain Applications and Institutional Trust. Frontiers in Blockchain, 3.

Szabo, N. (2017). Money, Blockchains, and Social Scalability. Satoshi Nakamoto Institute.

The Economist. (2015). The trust machine: The technology behind bitcoin could transform how the economy works.

Van de Werve, M. (2025). Real Estate Law in Belgium. In: Legal Handbook for Real Estate Transactions. Deloitte Legal.

Verma, S., Pathak, S.K., Chaudhary, P., Yadav, P., Yousuf, S. (2024). Land Registry System Using Blockchain Technology. International Conference on Emerging Innovations and Advanced Computing (INNOCOMP). Sonipat, India, 2024, pp. 170–177.

Williamson, O.E. (1985). The Economic Institutions of Capitalism. Simon and Schuster, 1985.

Williamson, O.E. (2000). The New Institutional Economics: Taking Stock, Looking Ahead. Journal of Economic Literature, 38 (3), pp.595–613.

Wong, S.K., Yiu, C.Y., Chau, K.W. (2012). Liquidity and Information Asymmetry in the Real Estate Market, 45, pp.49–62.

Zafar, A. (2025). Reconciling blockchain technology and data protection laws: regulatory challenges, technical solutions, and practical pathways. Journal of Cybersecurity, 11 (1)

Sustainability & Climate Action

Sub-Tracks:

Decarbonization & Carbon Footprint Analytics

Circular Economy & Digital Material Management

Climate Resilient Design & Adaptation

Energy Efficiency & Smart Building Operations

Integrated Delivery for Sustainability Goals

Performance-Based Sustainability Feedback Loops

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

AUTOMATED GENERATION OF DISASSEMBLY GRAPHS BASED ON DETAILED 2D SECTION DRAWINGS

Kasimir Forth[1], Mikhael Johanes[1], Elvire de Clermont Tonnerre[1], and Catherine De Wolf[1]
[1]ETH Zurich, Zurich, Switzerland

Abstract

Disassembly and circularity assessments are limited by the lack of structured, machine-readable data on material layers and connections, which are typically represented in expert-readable 2D section drawings. This paper explores how disassembly-relevant information can be automatically derived from such drawings represented as graphs using computer vision for segmentation, vision language models for material labeling, and language-based reasoning for semantic enrichment. The prototype extracts components and their connectivity and material information and generates graph representations capturing assembly hierarchies, enabling automated disassembly reasoning. This work marks a first step toward integrating heterogeneous data sources into unified graph models for circularity assessment.

Introduction

The built environment is one of the largest consumers of raw materials and a major contributor to global waste streams. Within Europe, construction and demolition activities generate close to 40% of all waste (García et al., 2024). Growing urbanization and rising living standards are expected to intensify this demand, particularly for mineral-based construction materials (OECD, 2019). Furthermore, as resource extraction and processing continue to drive greenhouse gas emissions, which are projected to rise to 4.6 Gt CO_2-equivalent annually by 2060 (UN Environment Programme, 2022), the construction sector faces increasing pressure to reduce its environmental footprint.

To address these challenges, the concept of a circular economy has gained prominence in the construction sector. It promotes keeping materials and components in the use stage for as long as possible by designing buildings that can be adapted, dismantled, dis- and reassembled rather than demolished. Achieving such reusability, however, depends on reliable knowledge of how building parts are connected and composed. This information of materials and components in existing contexts is often missing or inconsistently represented in current digital workflows, limiting the ability to evaluate a building's potential for disassembly.

Motivation and research gap

Existing circularity metrics, such as the material circularity indicator (MCI) (Goddin et al., 2019) and disassembly potential (Durmisevic, 2006), require detailed information on material composition, connection types, and assembly hierarchies. In previous studies (Forth and De Wolf, 2025), we showed that representing this disassembly information using knowledge graphs, hereinafter referred to as disassembly graphs, supports circular decision-making. However, the extraction of all relevant data for existing assemblies remains time consuming and dependent on expert knowledge.

Current digital methods for assessing disassembly potential rely on incomplete as-built data or manual modeling, while detailed two-dimensional (2D) section drawings contain rich latent information on assemblies and materials that remains underutilized. Automating the interpretation of such drawings can bridge the gap between design documentation and disassembly-oriented building information, enabling large-scale circularity assessments and disassembly planning and thus the reuse of components and materials.

This study investigates how disassembly graphs can be automatically derived from detailed 2D as-designed section drawings to represent the hierarchical and material relationships of existing building assemblies. We hypothesize that by combining vision-language reasoning techniques, it is possible to infer building component connectivity and semantic relationships sufficient to construct a graph representation that supports circularity assessments that focus on disassembly potential.

Background and related work

This section introduces the two main concepts of representing disassembly-related information using graphs and reasoning techniques for automatically extracting information from 2D architectural drawings.

Graph representations of disassembly information

The representation of disassembly information has evolved from manual heuristic assessments to sophisticated graph-based models that capture the complex interdependencies between building components. Fundamental to this domain is the work of Durmisevic, who established the theoretical groundwork for quantifying disassembly potential (Durmisevic, 2006). She proposed modeling building configurations as relational diagrams of components and their connections to represent physical dependencies. By analyzing these relational structures, Durmisevic demonstrated that the disassembly potential is inversely propor-

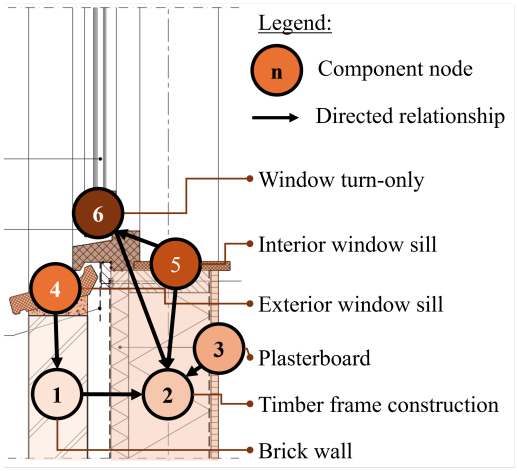


Figure 1: Example of information for disassembly potential of a wall and window according to van Vliet (2018)

tional to the complexity of the dependencies of components (Durmisevic, 2006). Specifically, hierarchical and parallel relational patterns facilitate higher recoverability compared to integrated, sequential dependencies.

Building on these topological foundations, Smith et al. introduced the disassembly sequence structure graph (DSSG), a formalism designed to optimize the disassembly process itself (Smith et al., 2012). Unlike static dependency diagrams, the DSSG models the process by generating a graph in which nodes represent specific parts and edges represent feasible removal sequences. This approach addresses the ”selective disassembly” problem by allowing for the identification of optimal paths to remove specific target components for maintenance or reuse while minimizing the total number of operations.

In more recent applications focusing on circularity assessment, van Vliet operationalized Durmisevic’s relational patterns to evaluate the circular potential of existing whole buildings (van Vliet, 2018). He simplified Durmisevic’s approach by focusing on the most important properties, such as ”type of connection”, ”assembly sequence”, ”accessibility to connection, as well as ”assembly shape”, ”independency”, ”type of relational pattern” or ”method of fabrication” to calculate a disassembly potential score. This method highlights the practical utility of graph representations in translating abstract circularity principles into quantifiable metrics for building elements.

Most recently, the focus has shifted toward using semantic knowledge graphs to manage the data required for these assessments. Bosma investigated the use of linked data to create digital product passports (DPPs) (Bosma, 2024). Instead of purely geometric or process graphs, Bosma utilizes an ontology-based graph structure, where products, materials, and their properties (such as location and warranty) are interlinked nodes. This semantic graph approach ensures that the disassembly information, such as material composition and manufacturer data, remains accessible and able to be queried.

Reasoning of 2D architectural drawings

In the domain of pixel-based analysis, convolutional neural networks (CNNs) and vision transformers (ViTs) have become the standard for tasks requiring visual pattern recognition, such as material detection and drawing classification. For instance, Pizarro et al. reviewed widespread methodologies involving CNNs for extracting walls, doors, and rooms from raster images. They note that these methods often discard essential topological metadata (Pizarro et al., 2022). More specialized applications include the work of Kairlapova et al., who successfully deployed CNN architectures like MobileNetV2 and ResNet50 to detect and classify building materials in 2D drawings to facilitate material passport generation (Kairlapova et al., 2025). Similarly, Carrara et al. compared ViTs against graph-based methods for classifying construction drawings into categories like discipline and project phase. They demonstrated that while pixel-based transformers are effective, they may fall short in capturing complex structural relations (Carrara et al., 2025a).

To address the limitations of raster approaches in capturing geometric and topological relationships, recent research has increasingly focused on vector and graph-based representations. Graph neural networks (GNNs) are gaining prominence for their ability to model drawings as connected entities rather than mere collections of pixels. Ko and Lee proposed a framework that converts architectural detail drawings into graph structures. Their framework utilizes GNNs to improve error detection and drawing classification by learning the interrelationships between drawing elements Ko and Lee (2025). Further advancing this domain, Carrara et al. introduced VectorGraphNet, a method that parses PDF drawings into scalable vector graphics graphs (Carrara et al., 2025b). By applying graph attention transformers, they achieved state-of-the-art results in line-level semantic segmentation, thus validating that structure-aware models can offer superior scalability and accuracy for complex technical drawings compared to conventional computer-vision-based techniques. Complementing these extraction methods, Schönfelder and König addressed the downstream challenge of data interoperability (Schönfelder and König, 2025). They introduced the drawing analysis ontology (DAnO), a framework designed to standardize the output of various computer vision tools into knowledge graphs, thereby enabling automated validation and semantic reasoning to refine the reconstructed building models.

Most approaches focus on segmentation and object detection on the floor plan level. To date, only Ko and Lee have focused on detailed architectural sections (Ko and Lee, 2025). However, they did not take textual input of legend and descriptions into account to further reason quantities, material, and connection information. Consequently, the work presented here extends the field toward multi-modal reasoning of detailed section drawings. It leverages segmentation, legend parsing, and graph-based modeling to infer assembly hierarchies and material relationships from

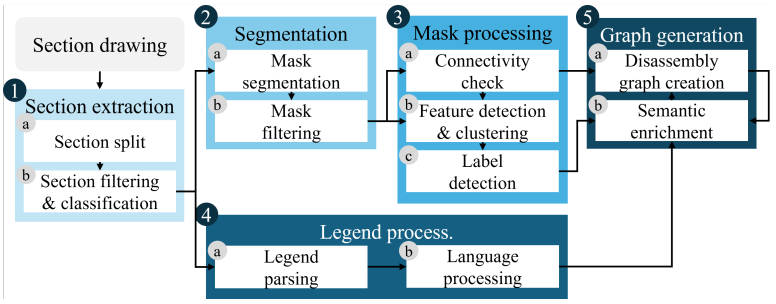


Figure 2: Overview of the proposed framework of deriving disassembly graphs based on 2D section drawings

as-designed documentation.

Method of generating disassembly graphs based on 2D section drawings

This study introduces a novel workflow for the automated generation of disassembly graphs based on 2D section drawings using computer vision, vision language models (VLM), and large language models (LLM) as well as graph-based reasoning techniques. Following Peffers' Design Science Research methodology Peffers et al. (2007), we identified the problem and defined the objective. Next, we designed and developed the overall framework, the steps of which are explained here. Finally, we demonstrated the framework with a case study and evaluated the results.

Overall framework

The overall framework consists of five main steps with several substeps, as shown in Figure 2. First, we split the whole drawing into its individual sections (1.a) and filter or classify them (1.b). Those section images that do not contain color gradients or legend text are further processed for segmentation (2). In the second step, each section is segmented into masks (2.a), and the masks are filtered (2.b): meaningful masks are kept, and almost blank masks are discarded. In the next step (3), the filtered masks from the second step are further processed and combined with the processed legend information. In the first substep (3.a), the connectivity between neighboring masks is identified. Next, features are detected (3.b), and finally, every mask is labeled either according to the detected features or the identified legend association (3.c). The section that was classified as legend is further processed by parsing and correcting the text (4.a) and then restructuring the text according to geometric and semantic information (4.b). In the final step (5), the disassembly graph is generated by, first, creating nodes for each mask and edges for each connection (5.a) and then enriching the results of the labeling and legend information (5.b). Depending on the drawing style, the order of these steps may change, but the general steps can be robustly applied to several drawing styles.

Section extraction

The section extraction-and-filtering pipeline automates the decomposition of sections across the entire drawing. Initial regions of interest are isolated by detecting structural contours within the image, which separates potential layout blocks from the background. Text-only areas are excluded using CLIP-based semantic filtering (Radford et al., 2021). These candidates are then validated using a visual classifier that checks for low color saturation and high edge density. This process effectively distinguishes valid technical sections, which are characterized by dense, monochromatic line work, from non-geometric artifacts, such as logos or photorealistic renderings.

Segmentation

We apply an automatic mask generator based on a Segment Anything Model (SAM) (Kirillov et al., 2023) that densely samples point prompts across the drawing to produce a large set of overlapping candidate segments. These proposals are subsequently post-processed to resolve overlaps and extract a set of disjoint graphical regions corresponding to distinct drawing components. To ensure reproducibility and stabilize color-encoded outputs, each detected mask is assigned a fixed, distinct color generated via a deterministic color scheme. Each cropped mask is evaluated using grayscale statistics, including brightness distribution, proportion of dark pixels, edge variance, and intensity histogram entropy. Segments are retained only if they exhibit (i) sufficient dark content relative to the background, (ii) meaningful edge structure indicative of technical line work or texture, and (iii) adequate intensity diversity throughout the region. Near white or homogeneous fragments that do not meet these criteria are discarded.

Mask processing

Different processing steps are necessary to bridge the gap between the segmented masks and the processed legend in order to generate semantic-rich disassembly graphs. We propose three main processing steps: connectivity check (4.a), feature detection and clustering (4.b), and label detection (4.c).

The first step checks the connectivity between masks representing physical adjacency between them. In technical drawings, lines of masks that should theoretically touch are often separated by tiny gaps of 1–2 pixels due to rasterization (converting vector lines to pixels) or anti-aliasing (smoothing edges). To take this into account, we use a 3x3 kernel dilation, which expands every segment by 3 pixels

in all directions. Now the artificially expanded segment bridges tiny gaps, ensuring that segments which are almost touching are recorded as adjacent. This robust adjacency detection ensures that compound structures (e.g., a wall composed of multiple material layers) are correctly linked as a coherent assembly.

To enable classification of similar materials within each mask, the system extracts a high-dimensional feature vector for each segment. While models like CLIP are powerful for semantic image-text alignment (Radford et al., 2021), we used DINO (self-supervised Distillation with No labels) in our implementation (Siméoni et al., 2025). The decision to use DINO stems from the nature of technical drawings. Architectural component materials are defined by local geometric textures (hatching patterns) rather than the global semantic contexts. The extracted features are subsequently clustered using K-Means algorithm, which groups segments with visually similar hatching patterns into common classes. This visual clustering serves as a fallback mechanism. If a specific segment lacks a direct text label from the next step, then its material properties can be inferred from its cluster neighbors that have been successfully identified.

The final step bridges the gap between the component segment connectivity graph and the parsed legend data. While the previous steps characterize the shape of the segments where they touch, this step identifies what label it contains by detecting numerical callouts. The system employs a VLM, specifically configured with a structured prompting strategy, to scan the original section images and the segments. The model is provided with the cropped section and the bounding box coordinates of the segments. To further facilitate spatial reasoning, the input image is augmented with a visible grid aligned with the image's pixel resolution. This grid provides an explicit, shared spatial reference frame that simplifies the localization and description of leader lines and numerical callouts by the language model. It is then tasked with identifying leader lines and numerical indicators pointing to specific geometries. The output is a structured mapping linking a segment ID to a legend number. This direct assignment allows the system to propagate the rich material metadata extracted from the legend in the next step to the corresponding nodes in the graph, thereby completing the semantic enrichment of the graph in the final step.

Legend parsing

To bridge the gap between abstract geometric segments and their physical properties, we detect and interpret the drawing's legend. Initial attempts to digitize this information relied on conventional optical character recognition (OCR) engines, such as Tesseract. However, these methods proved insufficient for the provided dataset. The low resolution and rasterization artifacts inherent in the technical drawings led to significant character misrecognition, resulting in incoherent text outputs that lacked semantic continuity.

To overcome these limitations, we use a VLM, in our case OpenAI's "GPT-4o-mini" (OpenAI et al., 2024). Unlike standard OCR, which processes characters individually, the VLM analyzes the visual context of the legend table, allowing it to infer meaning even from pixelated text. The model is prompted to extract the legend text and restructure it into a machine-readable dictionary, as shown for one layer in 4.b in Figure 3. This prompting strategy explicitly instructs the model to decompose the raw text descriptions into granular attributes for each layer. The resulting data structure differentiates for every component between material composition and layer thickness and, if available, connection type. We have, to date, implemented this workflow for only one style of legend and need to adapt it to further styles in future.

Graph generation

The final stage of the framework synthesizes the geometric, topological, and semantic data into a unified graph database. This step transforms the isolated segmentation masks into a disassembly graph that can be queried, where the building assembly is represented as a network of interconnected components. We use labeled property graphs since component properties can directly be stored in nodes and connection properties in edges, and implement the disassembly graph using Neo4j.

For every segmentation mask, a distinct component node is instantiated within the graph database. To ensure full traceability and visual consistency between the disassembly graph and the original drawing, each node is populated with a comprehensive set of properties, including a unique segment identifier, the source case study and section reference, and the specific RGB color values of the mask, which allow for manual traceability. Additionally, classification data such as the detected label and the cluster ID from feature analysis are stored as node properties to serve as keys for subsequent semantic mapping.

Following node creation, the physical relationships between components are encoded as edges to model the structural topology of the assembly. Using the adjacency list generated during the connectivity analysis, the system establishes connection relationships between interacting nodes, effectively distinguishing between isolated elements and composite structures.

Finally, the graph is enriched with the detailed metadata parsed from the legend. By querying the label property of each node against the structured legend dictionary, the system infers all available information, such as material composition and layer thickness, into the geometric entities.

Case study and results

To validate the proposed framework, we prototypically implemented the workflow and evaluated the segmentation, connectivity, and label detection step in two case studies to show the accuracy and stability performance at each step.

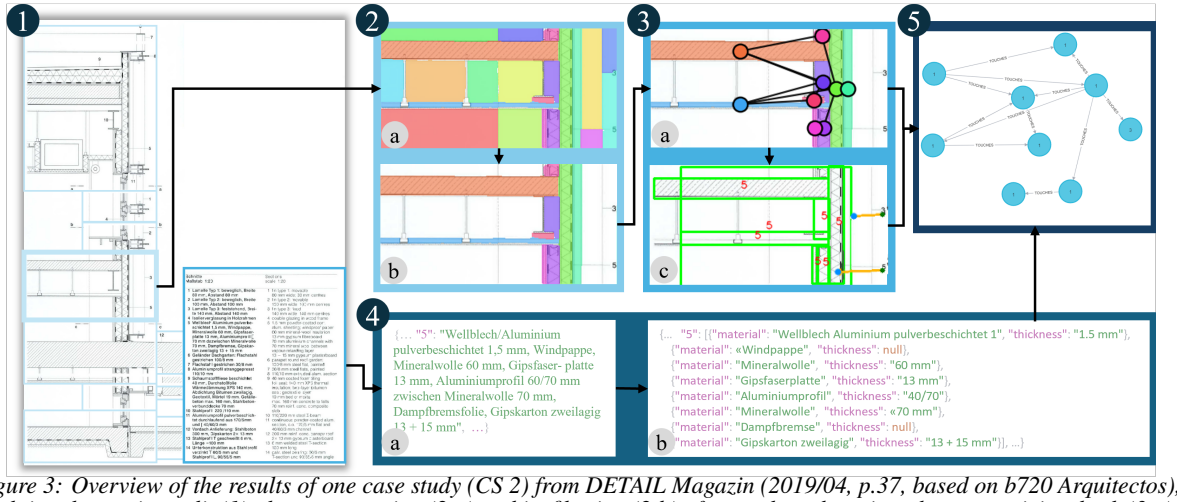


Figure 3: Overview of the results of one case study (CS 2) from DETAIL Magazin (2019/04, p.37, based on b720 Arquitectos), highlighting the section split (1), the segmentation (2.a) and its filtering (2.b) of one selected section, the connectivity check (3.a) and detected labels (3.c), the related legend excerpt (4), and the resulting disassembly graph of this section (5).

Case studies

We selected two exemplary detailed section pages from the architecture trade magazine *DETAIL*, one containing six sections and the second four sections. These reflect a standardized style of detailed section drawings and contain a legend in German. Figure 3 shows one exemplary section with the results of each step. The case studies show multi-layered facades, challenging the segmentation, connectivity, and labeling algorithm to distinguish repetitive geometries, continuous layers, and air gaps.

Results and discussion

In the following subsections, we evaluate three main steps of the overall workflow, focusing on the segmentation, the resulting connectivities, and the identification of labels for the relevant segments. We ran the implementation five times for both case studies to quantitatively validate accuracy and stability for the three selected steps, as well as analyzing error patterns qualitatively selecting one example section.

Segmentation

The segmentation performance was evaluated by comparing predicted instance masks against manually annotated ground truth across all ten section drawings from both case studies. Instance-level evaluation was performed using greedy matching at an IoU threshold of 0.5, resulting in a precision of 52.4%, a recall of 61.5%, and an F1-score of 56.6%. These results indicate limited agreement at the instance level.

Further analysis shows that the instance-level scores are primarily affected by systematic over- and under-segmentation rather than poor spatial localization. When instance boundaries are ignored and evaluation is performed on pixel-wise overlap between the unions of ground truth and predicted masks, the measured agreement is substantially higher, with a precision of 73.7%, a recall of 93.2%, and an F1-score of 82.3%. This discrepancy indicates that most relevant regions are spatially detected but not consistently partitioned into the correct instances.

Qualitative inspection, shown in Figure 4, confirms this behavior. The overlay view (top) shows good geometric alignment between predictions (colored fill) and ground truth (gray outlines), indicating that errors are primarily due to incorrect mask extent rather than mislocalization. The middle view highlights under-segmentation, where multiple adjacent components are merged into a single predicted instance, producing overly large masks that introduce false-positive regions. The bottom view illustrates over-segmentation, where elements with large hatched regions are split into multiple predicted instances, particularly across internal boundaries introduced by hatch patterns.

These results highlight the need for more domain-specific instance segmentation approaches for architectural and construction drawings. Although the use of SAM-based segmentation is promising, the model is not well adapted to the abstraction level and graphical conventions of section drawings, leading to systematic under- and over-segmentation. In particular, thin, contiguous, and hierarchically composed elements are not consistently grouped into coherent instances despite being geometrically well localized. This limitation suggests that generic image-driven segmentation models require additional architectural priors, drawing-aware constraints, or task-specific fine-tuning to reliably support instance-level reasoning in this domain.

Connectivity results

The performance of the connectivity algorithm was quantitatively assessed by comparing the automatically generated connectivity graphs against a manually annotated ground truth across ten detailed 2D architectural sections. Missing connections lead to false negatives (red), and wrong connections lead to false positives (orange), as

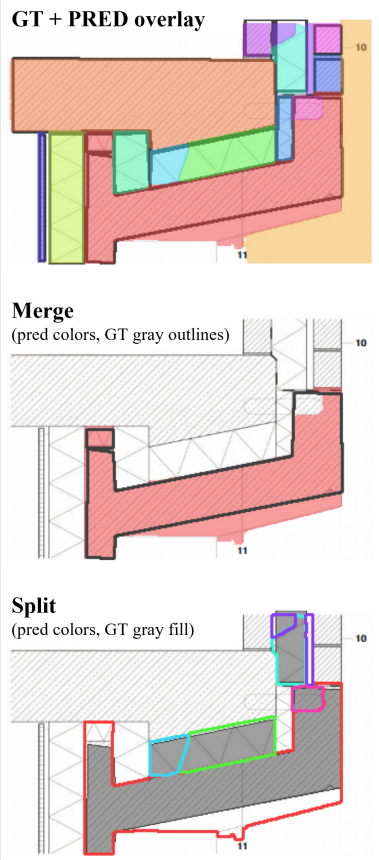


Figure 4: A qualitative comparison of ground truth and predicted segmentation for a representative section drawing shows generally sufficient alignment but inconsistent instance boundaries due to over- and under-segmentation.

shown exemplarily in Figure 5. The system achieved balanced and stable performance in all five iterations, yielding a precision of 82.52%, a recall of 87.41%, and a resulting F1 score of 84.89%. These metrics indicate that the framework is highly reliable for identifying standard physical adjacencies, effectively translating segmented mask data into a functional connectivity graph. The near parity between precision and recall suggests that the model does not suffer from a systemic bias toward over-prediction or under-prediction, providing a stable baseline for automated disassembly reasoning.

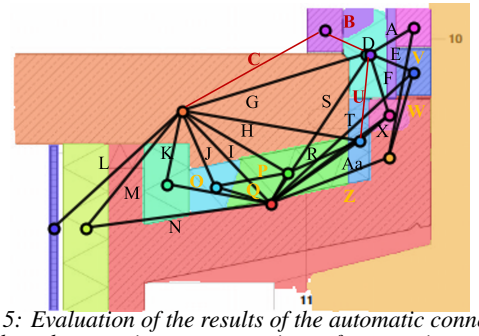


Figure 5: Evaluation of the results of the automatic connectivity check based on previous segmentations of one section example

A critical challenge identified during the evaluation is missed connections due to insufficient proximity between masks. In several instances, as illustrated in Figure 5, the human expert identified essential structural connections that the algorithm failed to detect because the masks were separated by small gaps. These gaps, often only a few pixels wide, frequently result from anti-aliasing artifacts during segmentation or the presence of thin architectural membranes that do not register as overlapping pixels. Because the current logic relies on strict pixel-to-pixel contact (extending by 3 pixels), these semantically connected components are treated as isolated entities. This limitation is particularly evident in complex joints where multiple materials converge but do not perfectly intersect in the digital representation.

Furthermore, the accuracy of the connectivity graph is heavily dependent on the upstream success of the computer vision segmentation phase. As highlighted in Figure 5, the absence of specific material masks directly results in a cascade of missing connections. These missing segments or over-segmented masks, such as unidentified insulation layers (light blue and green) in Figure 5, represent most of the recorded false positive connections. When the segmentation model fails to identify a mask for a specific material layer, the graph generator cannot establish any adjacency for that component, leading to incomplete assembly hierarchies. This underscores that the robustness of the final disassembly graph is intrinsically linked to the recall of the initial segmentation and labeling processes.

Segment label identification

Table 1 shows the average results of precision, recall, and F1 score as well as their standard deviations (σ) of all five iterations for each case study and overall (mean).

Table 1: Performance metrics for label identification across five iterations and the overall average

Average	**Precision**	**Recall**	**F1 Score**
CS 1	50.47%	86.00%	61.04%
CS 2	27.83%	100.00%	43.52%
Mean	37.68%	90.75%	52.91%
Std. Dev.	**Precision**	**Recall**	**F1 Score**
CS 1	9.58%	21.91%	2.39%
CS 2	1.82%	0.00%	2.23%
Mean	1.41%	14.58%	2.37%

The results show that case study 1 achieved a balanced performance, with an average F1 score of 61.04%, characterized by a recall of 86.00% and precision of 50.47%. However, it exhibited significant instability in detection rates. This was indicated by a high standard deviation in recall ($\sigma = 21.91\%$). Conversely, case study 2 demonstrated

perfect retrieval stability (recall 100.00%, $\sigma = 0.00\%$) but significantly lower precision (27.83%), resulting in a lower F1 score of 43.52%. This comparison highlights that while the system reliably detects all labels in complex configurations (CS 2), it is prone to assign labels too broadly compared to the more balanced but variable results of standard details (CS 1).

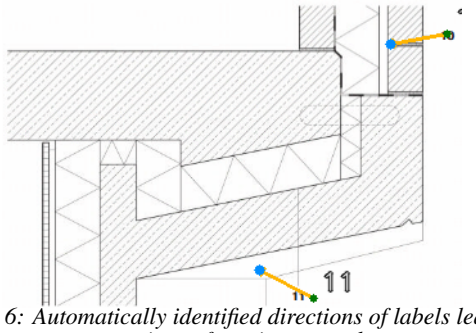


Figure 6: Automatically identified directions of labels leading to sections of one ion example.

The sources of inaccuracy were further investigated by visually inspecting and describing one exemplary section using the intermediate and final outputs. Figure 6 illustrates the system's interpretation of the leading lines. The green start point identifies the numeric label text, indicated by a green start point, and the blue end point the identified end point at the target segment, connected by a yellow leader line. While the leading line of label "10" accurately follows the initial direction, the leading line of label "11" deviates significantly, causing wrong labeling.

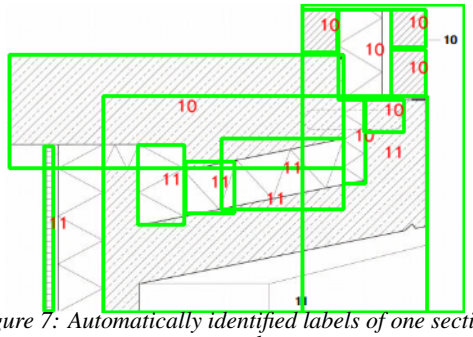


Figure 7: Automatically identified labels of one section example.

Figure 7 presents the resulting label assignments, revealing distinct success and failure modes. The system demonstrates robust performance in standard conditions, as evidenced by the correct labeling of the four top right segments. However, errors primarily arise from geometric inconsistencies in vector interpretation. A notable example occurs in the bottom-left region, where a segment is incorrectly labeled as "11" because the interpreted leading line vector deviates from its intended termination point and intersects a neighboring mask, as shown in Figure 6 . Additionally, the system exhibits a tendency to assign false positives in ambiguous zones. Two empty segments are incorrectly assigned label "10" despite the absence of explicit vector connectivity, suggesting an over-reliance on spatial proximity when graphical signals are weak.

In summary, the results indicate that while the VLM-based extraction is highly effective at retrieving semantic data (high recall), the geometric mapping mechanism requires refinement. The prevalence of false positives highlights the need for stricter vector constraints and collision detection algorithms to resolve ambiguities in dense or complex drawing regions. Thereby, we are aligning the module's precision with its high retrieval capability. In practice, better performance is observed in sections with larger components, while smaller or finer elements tend to generate more ambiguous mappings.

Limitations

The primary limitation pertains to the validated case studies. The current workflow is optimized using high-quality, standardized architectural detail drawings from a DETAIL magazine. Consequently, the system's robustness against the heterogeneity of broader architectural drawings, such as archival drawings, varying hatching standards, or hand-drawn sketches, remains unverified. The pipeline's adaptability to unstructured annotations or non-standardized material indications has not yet been evaluated.

Methodologically, quantitative benchmarking was restricted to the geometric processing modules. More specifically, it confirm only the accuracy of segmentation, connectivity detection, and label identification. Performance of legend processing and visual feature extraction have not yet been validated. Finally, the resulting disassembly potential calculations for every section have yet to be validated against a manual ground truth.

Technically, the prototype utilizes a specific configuration of models (SAM, DINOv3, GPT-4o) without comparative analysis. Ablation studies benchmarking these choices against alternative architectures, e.g., CLIP for feature extraction or open-source LLMs for parsing, were not conducted. Therefore, the current implementation may not represent the optimal balance between computational cost and inference accuracy.

Conclusion and outlook

This study proposes a novel computational framework for the automated generation of knowledge graphs interpreting disassembly-related data from 2D architectural section drawings. By integrating geometric segmentation (SAM), visual feature encoding (DINOv3), and semantic parsing (VLM), the pipeline successfully converts detailed section drawings with dense domain knowledge into disassembly graphs that can be queried. This transformation bridges a critical gap in the built environment: It enables the systematic assessment of reusability of components and materials and assembly hierarchies from legacy data sources that were previously inaccessible to digital circularity tools.

Future research will focus on validating the end-to-end fidelity of the generated disassembly graphs by comparing disassembly potential calculation results against manually curated ground truth. Next, the pipeline's robustness

will be tested on heterogeneous datasets that extend beyond standardized details to include diverse archival and hand-drawn documentation. Additionally, systematic ablation studies will benchmark alternative architectures. We will compare different models for computer vision, VLM, and LLM to optimize the trade-off between computational cost and inference accuracy, and focus on connection type properties. Finally, we aim to research how to link and fuse data with other data sources such as point clouds and BIM models (Forth and De Wolf, 2025) to enable a data-driven workflow for holistic circularity assessments.

Acknowledgments

This research has been funded by the SWIRCULAR project supported by Innosuisse (107.235 FS-EE). A collaborating author was supported through a Design++ fellowship program at ETH Zurich.

References

Bosma, J. A. M. (2024). From Material Passports to Digital Product Passports Creating and validating Linked Data-Based Digital Product Passports for the AEC industry.

Carrara, A., Nousias, S., and Borrmann, A. (2025a). Content-based classification of construction drawings : a comparative study of vision transformers and graph attention networks. In EG-ICE 2025: International Workshop on Intelligent Computing in Engineering. University of Strathclyde Publishing, Edinburgh, UK.

Carrara, A., Nousias, S., and Borrmann, A. (2025b). VectorGraphNET: Graph Attention Networks for Accurate Segmentation of Complex Technical Drawings. Journal of Computing in Civil Engineering, 39(6):04025085.

Durmisevic, E. (2006). Transformable building structures. Design for dissassembly as a way to introduce sustainable engineering to building design & construction. [s.n.], [S.l.].

Forth, K. and De Wolf, C. (2025). Formalizing information for disassembly potenttial of buildings using bim and labeled property graphs. In Proceedings of the 2025 European Conference on Computing in Construction, volume 6, Porto, Portugal. EC3.

García, C., Caro, J., Gallo, G., and Tonini, F. (2024). Techno-economic and environmental assessment of construction and demolition waste management in the European Union - Status quo and prospective potential. Technical report, European Commission.

Goddin, J., Marshall, K., Pereira, A., Tuppen, C., Herrmann, S., Jones, S., Krieger, T., and Lenges, C. e. (2019). Circularity Indicators: An Approach to Measuring Circularity.

Kairlapova, A., Forth, K., Carrara, A., and Borrmann, A. (2025). Material detection and classification in 2d architectural drawings using computer vision for material passports. In Proceedings of the 2025 European Conference on Computing in Construction, volume 6, Porto, Portugal. EC3.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., and Girshick, R. (2023). Segment Anything. arXiv:2304.02643.

Ko, J. and Lee, D. (2025). Graph neural networks for classification and error detection in 2D architectural detail drawings. Automation in Construction, 170:105936.

OECD (2019). Global Material Resources Outlook to 2060: Economic Drivers and Environmental Consequences. Technical report, OECD Publishing, Paris. ISBN: 9789264307445.

OpenAI, Hurst, A., Lerer, A., and Goucher, A. P. e. (2024). GPT-4o System Card. arXiv:2410.21276 [cs].

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3):45–77.

Pizarro, P. N., Hitschfeld, N., Sipiran, I., and Saavedra, J. M. (2022). Automatic floor plan analysis and recognition. Automation in Construction, 140:104348.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning transferable visual models from natural language supervision.

Schönfelder, P. and König, M. (2025). Ontology-based reasoning in automatic floor plan analysis. Advanced Engineering Informatics, 68:103761.

Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., and et al. (2025). DINOv3. arXiv:2508.10104 [cs].

Smith, S., Smith, G., and Chen, W. H. (2012). Disassembly sequence structure graphs: An optimal approach for multiple-target selective disassembly sequence planning. Advanced Engineering Informatics, 26(2):306–316.

UN Environment Programme (2022). 2022 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector.

van Vliet, M. (2018). Disassembling the steps towards Building Circularity Redeveloping the Building Disassembly assessment method in the Building Circularity Indicator. Master's thesis, Eindhoven University of Technology.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

TOWARDS STANDARDISED ENVIRONMENTAL PRODUCT DATA INTEGRATION IN IFC

Martina Jakubowska[1], and Kristoffer Negendahl[1]
[1]Technical University of Denmark, Denmark

Abstract

Whole Building Life Cycle Assessments are currently hindered by manual, labour-intensive workflows and fragmented BIM and LCA data, making iterative use impractical. Existing automated workflows often lack adherence to standards, undermine data quality, or introduce data migration burdens. This study investigates integrating WBLCA data within IFC, focusing on the Danish context, by formalising Danish LCA requirements on bSDD. We propose a data model for implementing EPDs as PDTs, according to ISO 22057, in IFC using the `IfcProjectLibrary` exchange format defined in ISO 17549-2 and ISO 16757-5. A prototype implementation of PDT-to-IFC conversion tool demonstrates and evaluates the data model's feasibility.

Introduction

Life Cycle Assessment

Life Cycle Assessment (LCA), standardised through the ISO 14000 series, is the established methodology for quantifying the environmental impacts. It evaluates multiple environmental indicators across the life cycle, from raw material extraction to end of life, and is essential for avoiding burden shifting from one environmental problem to another. LCA applied to buildings is referred to as Whole Building Life Cycle Assessment (WBLCA). A WBLCA aggregates the impacts of construction products and combines them with impacts from the operational phase.

Product-level results in building LCAs are typically communicated through Environmental Product Declarations (EPDs), defined by ISO 14025 and harmonised in the construction sector through the Product Category Rules (PCRs) standard EN 15804. Product LCA data may be product-specific or generic, e.g. the ÖKOBAUDAT, which is based on representative assumptions rather than individual products. At building level, EN 15978 provides the assessment framework.

While EPDs are still mainly published as PDFs, machine-readable formats are gaining importance. In Europe, ILCD+EPD is widely used, alongside newer approaches like ISO 22057, which defines EPDs as product data templates (PDTs) for direct integration into BIM workflows.

Current WBLCA Workflow

Today, performing a WBLCA is complex and labour-intensive, relying heavily on manual processes such as extracting quantities from models, mapping materials to appropriate EPDs, and coordinating inputs across disciplines. Interdisciplinary communication relies on disconnected tools, while required environmental data is frequently incomplete or unavailable in digital formats. These steps introduce additional workloads, are error-prone, and limit the feasibility of iterative assessments.

Long-term data storage is a further challenge. Although buildings last for decades, LCA results and their underlying assumptions are often difficult to access or reuse after only a few years, reducing transparency and leading to information loss.

Existing automated workflows offer some improvements, but often lack alignment with standards, compromise data quality, or depend on data transfers between systems that require additional mapping and validation. There is clear potential to improve WBLCA by reducing manual effort, enabling earlier and more iterative assessments. Such workflows should align with data standards, maintain scientific rigour, and ensure that models remain transparent and self-documenting over time.

Background

Strategies for Storing and Sharing EPD Data

Addressing these issues requires tighter integration of LCA data into BIM-based workflows (Klumbyte et al. (2023), Bari et al. (2019), Safari and AzariJafari (2021)), with a focus on open standards, such as IFC, and vendor-neutral, collaborative processes like openBIM, to ensure interoperability and long-term data accessibility.

Initial attempts to integrate LCA data into IFC rely on the property sets `PSet_EnvironmentalImpactIndicators` and `PSet_EnvironmentalImpactValues`. These property sets store environmental indicators per functional unit and calculate scaled results based on element quantities. However, their practical use is currently limited. The main limitation is the inability to represent multiple lifecycle modules for a single element. The `LifeCyclePhase` attribute allows only one phase to be assigned, which prevents full "cradle to grave" representation. In addition, the supported indicators are not fully aligned with EN 15804, agreement on

indicator definitions is incomplete, and LCA data cannot be consistently defined at material level or aggregated at building level.

ILCD+EPD and ISO 22057

The ILCD+EPD format, a subset of European Commission's ILCD framework, was an early and significant step in digitising EPDs, using XML to make environmental data machine-readable. Initially developed to meet German governmental requirements, it has since 2022 become the mandatory format for all new European EPDs.
Despite this progress, ILCD+EPD remains limited in its applicability within BIM workflows, particularly for representing construction objects. This limitation motivated the development of ISO 22057, which aims to better integrate EPDs into BIM workflows and support fully machine-interpretable EPDs in the future.
A significant benefit of ISO 22057 lies in its integration within systems already familiar to manufacturers, who will predominantly be responsible for maintaining and publishing EPDs moving forward. The task of digitising EPDs should not fall on architects or LCA practitioners. Additionally, the implementation of EPDs as PDTs presents opportunities for integration with existing systems such as SMART CE, ETIM, and GTIN/GS1. A JSON-based representation of ISO 22057, called DT JSON (Erlandsson et al. (2024)), is also mentioned in the EU Taxonomy and as the potential base-format for Digital Product Passports.

ISO 17549-2 and CODview2

The standard ISO 17549-2 defines an MVD of the IFC schema, named Construction Object Data View 2 - CODview2 for short. The purpose of CODview2 is to store and exchange configurable construction objects and their requirements using the `IfcProjectLibrary` mechanism of IFC. The MVD can be used to exchange both empty PDTs and filled product data sheets (PDSs). CODview2 does not support geometrical representations, but allows for exchanging geometric properties, focusing on complex, dynamic requirement constraints, and product specifications using embedded JavaScript functions. ISO 17549-2 does not define any construction-specific semantic entities or IFC element classes. Instead, it recommends using `IfcBuildingElementProxy` to represent all object types and defines classification through external data dictionaries based on ISO 12006-3:2022, such as the buildingSMART Data Dictionary (bSDD), as the sole sources of construction semantics.

ISO 16757-5 and IFC Catalogues

ISO 16757-5:2025 - *"Data structures for electronic product catalogues for building services - Product catalogue exchange format"* builds on the structure defined in ISO 17549-2 by, among others, introducing means of representing parametric geometry and generation of IFC geometries for selected product variants. The result is an IFC-based product catalogue, dubbed IFC catalogue, bridging the gap between manufacturer's catalogues and BIM project models. ISO 16757-5 keeps the strong connection to data dictionaries, relying on them for definitions of product classes and related properties.

Table 1: Central literature on LCA-IFC integration

Ref	Focus	Link to ÖKOBAU	Custom Pset	UUID Link	Spreadsheet with ERs	MVD	Custom schema
Lambertz, M. et al. (2019)	Analysis of multiple methods; IDM; IFC model architectures; UUID	✓	✓	✓	✓	✓	
Horn et al. (2020)	ifcXML; custom XML schema; DGNB; LCA MVD	(✓)	(✓)		✓	✓	✓
Schmid et al. (2022)	EPDs in IFC using IfcTask & IfcResource; bSDD; ISO 22057		✓		(✓)		
Theißen et al. (2020)	Building services; Revit; LCA MVD (custom Psets)	✓	✓	✓	✓		
Llatas et al. (2022)	SBA (LCA, LCC, S-LCA); custom Dynamo scripts		✓				

Existing Research on LCA and BIM Integrations

Research on BIM–LCA integration is extensive (Bolognesi et al. (2025), Parekh and Trabucco (2024), Potrč Obrecht et al. (2020)). Wastiels and Decuypere (2019) identified five levels of integration. These range from exporting Bills of Quantities (BoQs) and processing them in spreadsheets or dedicated LCA tools, to importing BIM models directly into LCA software, and performing real-time estimations using LCA plug-ins within native BIM software. A further level, relatively new in the LCA context, involves using manufacturer-specific BIM objects enriched with environmental data.
Most research and software implementations do not place IFC or openBIM at the core of the workflow. Studies that do focus on IFC or openBIM, are comparatively few. The main contributions in this area are summarised in Table 1. All of the approaches rely on creating custom IFC property sets. Some avoid embedding EPD data in the model, instead referencing external datasets via unique identifiers (UUIDs). While this reduces model size, it introduces risks related to data persistence, responsibility, and broken links.
The buildingSMART Sustainability Group's "Industry Report on Environmental Impact Indicators" (Schmid et al. (2022)) is the only approach identified that integrates ISO 22057 data into IFC and links to a DD. It defines LCA and EPD modules in the bSDD and maps them into IFC using existing entities such as `IfcTask` and `IfcResource`. However, the resulting structure is highly complex, relying on deep nesting and repurposing IFC concepts beyond their original intent to represent even simple EPD data. This limits software support and may reduce clarity for users. The approach also lacks scenario handling, diverges from ISO 22057 in several respects, is incompatible with the CODview2 MVD, and does not support easy reuse or exchange of EPD data across projects or BIM objects.
Notably, Kremer et al. (2025) recently demonstrated an implementation of IFC product catalogues with dynamic requirements, defined in ISO 16757-5, using a heat pump as

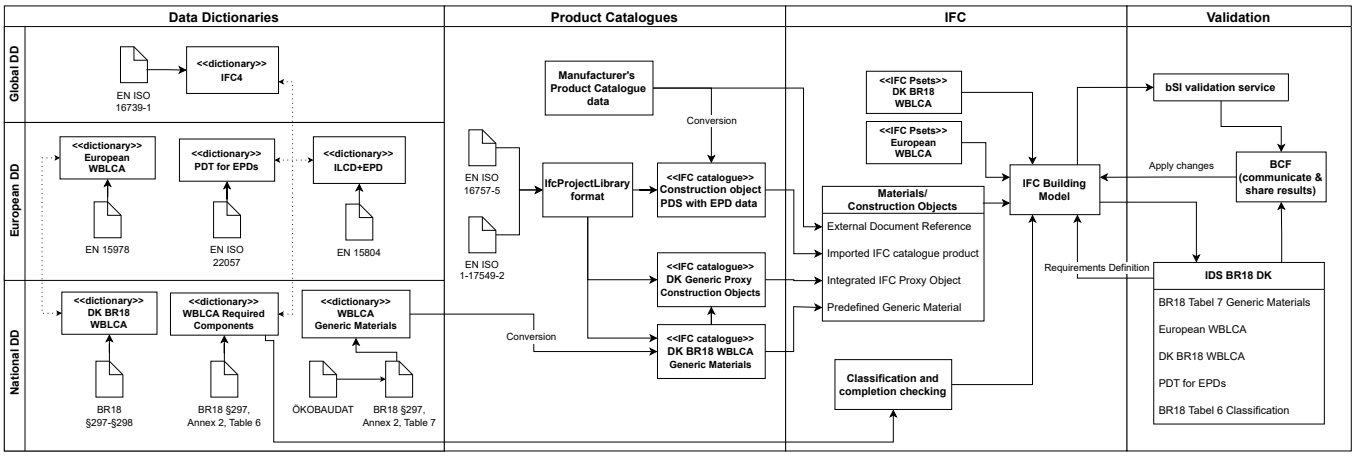


Figure 1: Data flow diagram of storing LCA information in an IFC context.

a case study.

None of the identified research investigates the implementation and enrichment of IFC catalogues, based on ISO 17549-2 and ISO 16757-5, with EPD data aligned with ISO 22057 as a means to enable transparent environmental data exchange and support automated WBLCA. Addressing this gap constitutes the primary focus and contribution of this study.

Danish Regulatory Context

In 2023, Denmark introduced new energy and climate regulations (§297–298), requiring WBLCAs. All new buildings, except unheated constructions under 50 m², need to comply with Global Warming Potential (GWP) limits ranging from 4 to 8 kg CO_2-eq/m²/year depending on function and size.

To define a consistent and practicable LCA scope, Annex 2, Table 6 of BR18 §297 specifies which building elements must be included and which may be excluded. Rather than requiring full material coverage, Table 6 focuses on elements that account for most of a building's climate impact, balancing analytical accuracy with feasibility and ensuring comparability across projects.

According to BR18, the environmental impact of included elements must be calculated using either product-specific EPDs or generic data defined in Annex 2, Table 7. Table 7 provides standardised generic values based on average data from the ÖKOBAUDAT database. No other generic datasets may be used, and any more detailed assessments must rely on third-party verified EPDs.

Methodology

This study proposes a method for integrating data needed to support a WBLCA process in a BIM context, with a specific focus on the Danish regulatory framework. The main objective is to enable practical integration of product-level EPD data into IFC models, with minimal manual input, clear data traceability, and long-term usability. Building-level LCA integration is addressed at a conceptual level.

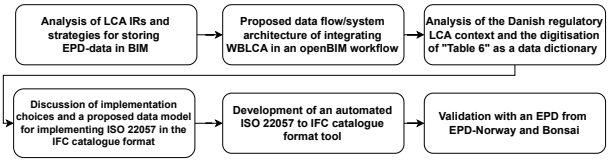


Figure 2: Summary of the methodological approach.

Figure 2 summarises the methodological approach. Firstly, existing approaches for storing, sharing, and reusing product data in IFC-based BIM workflows are then reviewed, alongside relevant sources and schemas for machine-readable, product-level LCA information.

Based on this analysis, an openBIM data model and workflow for WBLCA integration is proposed, linking IFC models with products stored in IFC catalogues, data dictionaries (DD), and national information requirements specified using Information Delivery Specification (IDS). Potential DD architectures are explored to support building-level, product-level, and Danish national requirements. As part of this work, BR18 §297, Annex 2, Table 6, is formalised, mapped to IFC classes, and digitised as a DD in the bSDD.

Implementation choices are then discussed and a data model for implementing ISO 22057 in the IFC catalogue format is proposed. This includes the handling of information modules, alternative scenarios and sub-scenarios, and the representation of EPDs at different modelling levels. Finally, a proof-of-concept automated conversion tool from ISO 22057 to the IFC catalogue format is developed using the IfcOpenShell library and validated with a real EPD from EPD-Norway using the Bonsai software.

Proposed Data Flow of LCA-IFC Integration

Figure 1 presents the proposed data model and its associated data flow. It illustrates the DDs required to support WBLCAs, structured across three levels: global, European, and national. In practice, an additional layer may be introduced for company- or project-specific data. Each dictionary is linked to the relevant standards that define its content. An implementation of ISO 22057 as a DD (shown as "PDT for EPDs") can build on existing initiatives, such as the buildingSMART Sustainability Group's "LCA In-

dicators and Modules" dictionary (Schmid et al. (2022)) discussed earlier.

Manufacturer's product catalogues provide input to construction objects shared through an IFC catalogue in the `IfcProjectLibrary` format, as defined in ISO 16757-5 and ISO 17549-2. The Danish context is represented through national data dictionaries and potential generic construction objects and materials in accordance with BR18 §297, Annex 2, Table 7.

An IDS tailored to Danish LCA requirements is used both to define information requirements and to validate the completed model. Final verification is carried out using the buildingSMART validation service, with results communicated via the BIM Collaboration Format (BCF), either externally or returned to the model for correction.

Implementation Analysis and Discussion

BR18 §298, Annex2, Table 6 as a Data Dictionary

A key part of the presented WBLCA workflow in the Danish context, is ensuring that it includes the required building components (part of inventory scope) defined in Table 6, Annex 2 of BR18 §298. Table 6 is provided by BR18 as an Excel sheet with nested categories of building elements, boolean specifying inclusion or omission, comments, and examples. To minimise oversights, support the integrated modelling process and be able to use the BIM model as a single source of truth (SSoT) in reporting and long-term archiving, Table 6 was formalised as a classification system to be used in IFC.

Using properties or classification systems defined in the bSDD has become one of the main recommended ways of extending IFC by buildingSMART. With the provision to avoid duplication of existing definitions, this approach presents a semi-standardised alternative to the many national and project-specific property sets that otherwise characterise the BIM landscape. Many BIM software today also provide direct integration with the bSDD API.

In line with this philosophy, we implemented the required elements from Table 6 as classification system on the bSDD[1]. Firstly, we made an English translation and numerical categories to organise the hierarchy. All required elements were then mapped to a corresponding IFC class. While this mapping should not be seen as a mandatory assignment to specific IFC classes, it serves as guidance for BIM practitioners during modelling and could be used in BIM software to recommend a classification assignment to the user. Figure 3 shows how an `IfcWallType` would be classified as a masonry wall ("Opmurede vægge") through nested `IfcClassificationReference` entities when using Table 6 as a DD.

Implementation of ISO 22057 in IFC

In this paper we present an integration of the ISO 22057 standard and its Annex A Excel workbook into the IFC schema using the IFC catalogue format defined in ISO

[1] A preview version of the Table 6 DD can be found here.

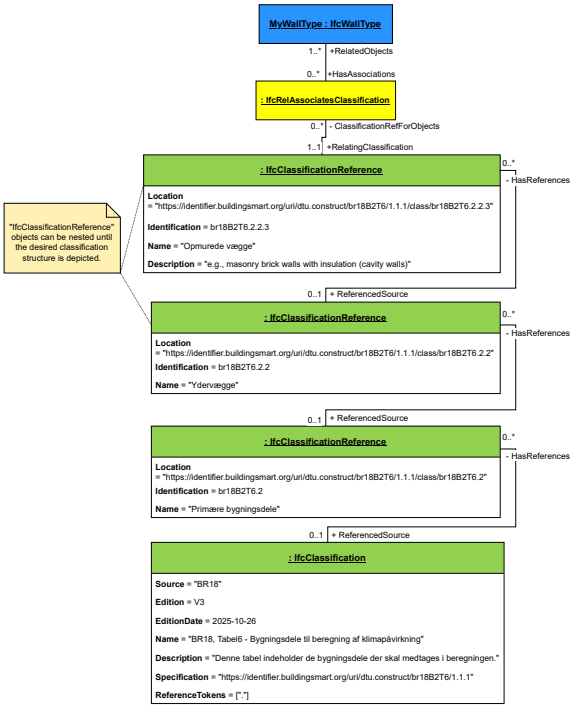


Figure 3: Example object diagram of association of BR18 Table 6 — "Masonry Walls" on the bSDD as a classification in IFC.

16757-5 and ISO 17549-2. In general, the IFC implementation adheres as closely to the standard as possible, preserving property names, descriptions, units, and other salient details unchanged.

Minor, IFC-specific implementation choices include; converting property names into PascalCase to align with the IFC naming conventions, prefacing property sets with `LCAPset_` to differentiate custom property sets from official `IfcPropertySets` beginning with "`Pset_`", mapping "Data type" fields to the appropriate `IfcValue` type. In case of enumerations, the property type was sat to `IfcPropertyEnumeration` and individual enumeration values were assigned to appropriate data types as indicated in the standard. More substantial modifications, or implementation choices, are described in the following sections.

Omitting Enumerations with Boolean Values

ISO 22057 defines several properties with "yes" and "no" enumerations, seen defined in Tables 9 and 10 in the standard. Representing these binary states as enumerations is redundant where a boolean data type such as `IfcBoolean` is available. Using `IfcBoolean` directly simplifies the data model and reduces unnecessary complexity.

Flattening the Nested Structure

ISO 22057 defines a deeply nested data template, organised into property groups with three levels of nesting in some cases (Figure 4). Annex D of the standard ("EPD in Smart CE Marking Declarations") describes how this structure is mapped to the smart CE XML format, which supports only two levels of nesting; "Property Level 1"

groupings and “Property Level 2” individual properties. ISO 22057 introduces additional tags (e.g. <EPDGeneralInformation>) to represent higher-level categories (shown in blue on Figure 4).

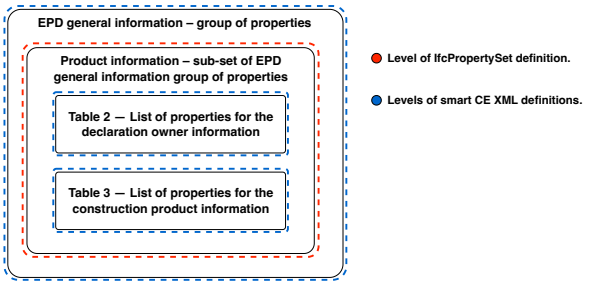


Figure 4: EN ISO 22057 nested structure adjusted for IFC and smart CE XML.

As IFC supports only a single level of property grouping, the intermediate groupings are implemented as `IfcPropertySets`. For example, the grouping highlighted in red in Figure 4 is represented as the property set `"LCAPset_ProductInformation"`. This balances the trade-off between restricting the number of properties in a single property set and minimising the overall number of property sets.

Implementing Information Modules and Scenarios

Representing EPD results for a single indicator, such as GWP limited to modules A1–A3, as commonly found in EPDs aligned with EN 15804+A1, is relatively simple and can be achieved with a single property (e.g. “`GWP_A1toA3`”). However, comprehensive assessment requires support for multiple indicators and life cycle modules, which substantially increases data model complexity. In ISO 22057, life cycle phases are defined through the “information module” property in the EPD methodological framework (Table 9, ISO 22057). This property links environmental indicators, waste categories, and output flows to specific modules. Complexity is further increased by the use of scenarios, which allow multiple declared values for the same indicator under different assumptions.

Several alternative approaches were considered for representing life cycle modules and scenarios in IFC, including dependent properties, repeated property sets following the ISO 22057 Smart CE XML implementation from Annex D, `IfcPropertyTableValue` properties, and the use of `IfcTask` and `IfcResource` entities (Schmid et al. (2022)). These approaches were rejected due to excessive property duplication, reliance on naming conventions, poor alignment with CODview2, increased model complexity, or limited software support.

The chosen approach represents environmental indicator results using `IfcTable`, referenced from a property via `IfcPropertyReferenceValue`. This enables a compact and structured representation closely aligned with the tabular format commonly used in EPDs, with indicators listed as rows and life cycle modules as columns. Unlike `IfcPropertyTableValue`, `IfcTable` allows an arbitrary number of columns and richer column definitions, supporting clearer semantics and better scalability as the number of indicators and modules increases. Units can be assigned at column level, improving machine interpretability while avoiding the repetition of properties for each module or scenario.

The `IfcTable` is composed of `IfcTableRow` and `IfcTableColumn` objects, with the former containing a list of values in their respective data types. `IfcTableColumn` enables the specification of an “identifier” attribute, allowing it to be referenced across multiple tables. It also supports the definition of “description”, “unit”, and a “reference path” that links to the object and attribute targeted by the column’s data via the `IfcReference` entity. The nesting capability of `IfcReference` entities further allows for direct linkage of an `IfcTableColumn` to a particular property within a property set of a specific object.

Given that `IfcTableColumns` can define units while `IfcTableRows` cannot, an alternative approach involves transposing the table, positioning indicators as columns and modules as rows. Unlike modules, indicators require unit definitions, and assigning them proper IFC units enhances computational interpretability, compared to listing them as IfcLabels in an additional column. However, this approach would complicate the direct referencing of modules to their defining property sets, and it may be more user-friendly to adhere to the conventional table format familiar from PDF EPDs.

A potential drawback of this approach is that the `IfcExternalReferenceRelationship`, utilised for linking properties to external definitions such as those in the bSDD, is not applicable to `IfcTableRows`. This constraint implies that while individual indicators would be able to be directly linked to their definitions in a DD, the overarching property could reference a compiled list of indicators.

While current support for visualising `IfcTable` is limited in some software, the data structure itself is robust, extensible, and well suited to automated processing.

Implementation of Alternative Scenarios

ISO 22057 allows two interpretations for handling alternative scenarios; either by defining separate data sheets for each scenario or by consolidating multiple scenarios within a single data sheet (as presented in Annex D).

Both approaches were examined in an IFC context, in the first of which, each scenario is modelled as a separate object definition (for example, alternative `IfcWallTypes`), each containing a complete set of environmental indicator results. Emission values are organised in a table linked to the `EN15804+A2CoreLCIAIndicators` property, under a new property set, `LCAPset_EnvironmentalIndicators`, housing all relevant indicator values (Table 2). Since `IfcTableColumns` can be referenced by multiple tables, the `ReferencePath` attribute, alongside an `IfcReference` object, points to the specific

Table 2: Implementation of modules using IfcPropertyReferenceValue and IfcTable

Pset.Name:	LCAPset_EnvironmentalIndicators					
Property Name	PropertyReference					
EN15804+A2CoreIndicators	Table.Name: EN15804+A2CoreLCIIndicatorsResults					
	Indicator	Unit	A1	A2	A3	...
	GWP_Total	kg CO2-eq				
	GWP_FossilFuels	kg CO2-eq				
	GWP_Biogenic	kg CO2-eq				
	...	...				

property set that corresponds to the module scenario, as outlined for modules A4 to D in the standard. This approach defines only a single scenario per module, suggesting that for alternative scenarios, the entire structure would be duplicated, rather than duplicating individual properties or property sets.

An alternative is to encode multiple scenarios within the same object definition by introducing additional properties and property sets for alternative scenarios and linking them to a base scenario. While this can be advantageous in early design phases, where rapid comparison of alternative assumptions is valuable, it relies on project-specific naming conventions and increases internal object complexity. It also departs from the main-text guidance of ISO 22057 and complicates standardisation through data dictionaries, as only the base scenario can be predefined.

Overall, modelling alternative scenarios as separate construction objects provides a straightforward and schematically clearer method. This approach is particularly suitable for the later stages of design, where the building LCA's primary goals is documentation and long-term storage of product information. However, if LCAs were reimagined as not merely calculations yielding a single numeric results but as comprehensive uncertainty analyses in and of itself resulting in a range of potential outcomes, the second approach could become relevant even in final LCAs. Figure 6 shows the implementation of alternative scenarios as separate construction objects.

Implementation of Sub-scenarios

Sub-scenarios, such as waste handling or transport, introduced in section 8.4.1 of ISO 22057, consist of an intricate data model with nested property dependencies. This complexity, as shown in Annex D for the smart CE XML, could be simplified by repeating all properties relevant to a module scenario. The IFC framework does, however, offer a mechanism for defining nested property sets through `IfcComplexProperty`, which allows grouping several properties under a single entry within an `IfcPropertySet`. Figure 6 details how sub-scenarios can be actualised within the IFC model using `IfcComplexProperty` and `IfcPropertyDependencyRelationship` to establish property dependencies. However, due to inconsistent implementation in BIM software, this mechanism is considered optional, with direct repetition of sub-scenario properties per scenario representing a valid solution.

Representing EPDs at Different Modelling Levels

Construction objects carrying EPD information may exist at different levels of detail and should therefore be definable at multiple levels within a BIM model. Where an EPD represents a complete building element, such as a prefabricated wall or a door, it can be assigned at the element type level. The element type defines LCIA indicator values per declared unit, while the corresponding element instance, with quantities defined through an `IfcQuantitySet`, represents the scaled results. These scaled values may be explicitly stored, for example in `IfcTables` for communication or visualisation, or derived implicitly by storing only a scaling factor and a reference to the type. This approach is exemplified in Figure 5 where an `IfcWallType` specifies $1m^2$ of a prefabricated wall, and the `IfcWall` instance represents the scaled results for $320m^2$ of the wall.

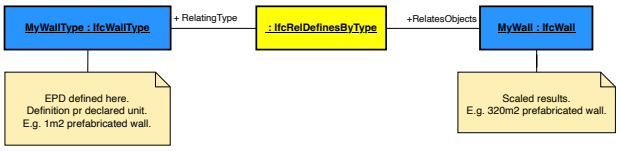


Figure 5: Construction object EPD definition on element level.

However in some cases, EPDs may describe smaller components or raw materials rather than complete building elements. Here, EPDs can be assigned at the material level, such as to an `IfcMaterialLayer` within a larger assembly, as exemplified on Figure 7. Defining EPDs at the type or material level avoids redundant property sets on individual instances, reducing file size and improving model quality. For materials requiring scaling, typically declared per mass or volume, ISO 22057 introduces reference quantities and scaling factors. In IFC, these can be implemented using attributes such as `LayerThickness` together with user-defined scaling properties, allowing EPDs to be defined once at the `IfcMaterial` level and consistently applied across different material set definitions.

PDT-to-IFC Implementation

The data model presented in the previous sections was evaluated by an implementation in the IFC STEP Physical File format (SPF-IFC), using a custom conversion tool developed with the IfcOpenShell library. The implemen-

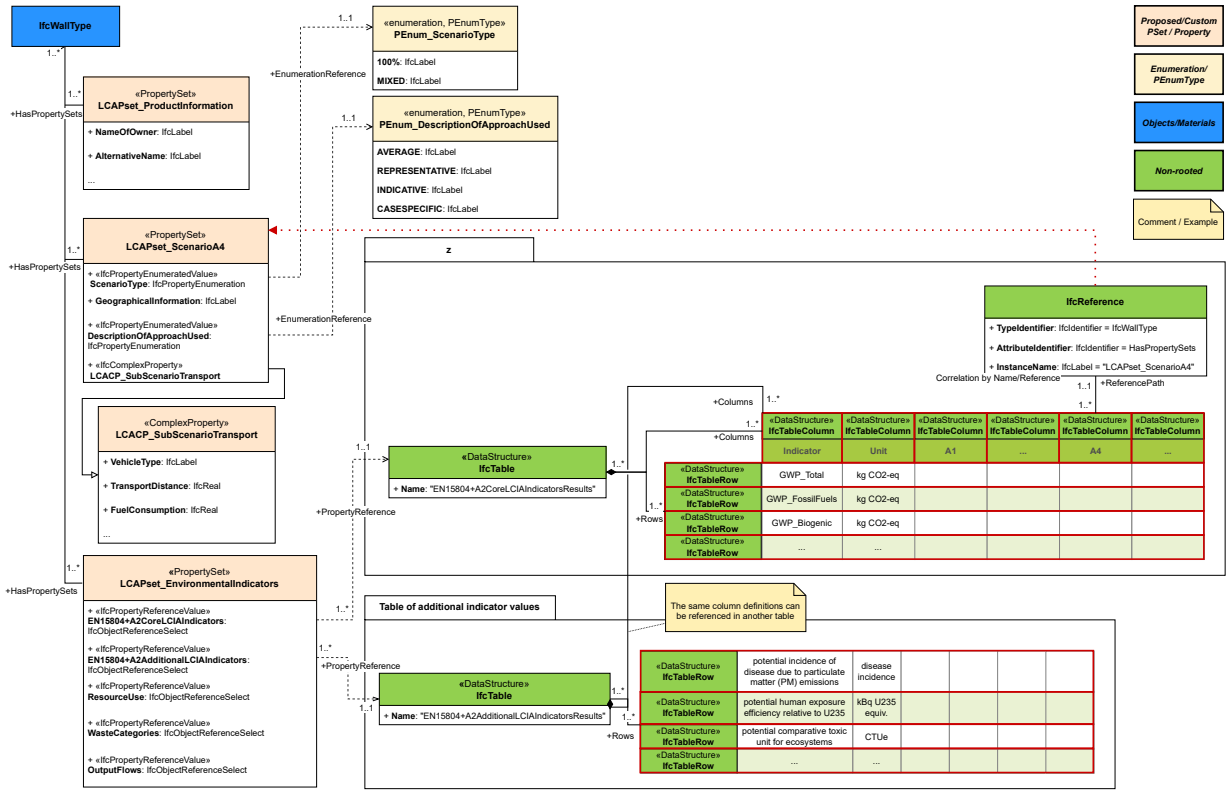


Figure 6: Alternative scenarios defined in different construction objects - with sub-scenarios.

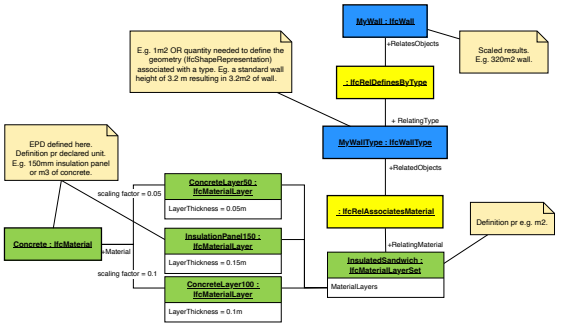


Figure 7: Construction object EPD definition on material level.

tation is based on the CODview2 MVD and the accompanying example IFC files.

The workflow begins with the creation of a `IfcProjectLibrary` model, including reference libraries. A user-defined element type is then introduced as the anchor for all subsequent LCA data. Property sets with simple or complex properties are created, including the dedicated `LCAPset_EnvironmentalIndicators`, which stores LCIA indicator results using `IfcTable`. All properties are linked to bSDD definitions via the `Specification` attribute of `IfcProperty`.

The ISO 22057 product catalogue implementation was demonstrated using a real EPD from EPD-Norway. The pilot project "EPD in ISO 22057 format", which provides a JSON-based representation of the standard, was used as input for the conversion. As an example, the EPD for "Massiv betongelement - Vegg, lavkarbonbetong" was selected, with Listing 1 showing a subset of properties from the `LCAPset_ProductInformation` property set.

As the available EPDs did not include scenario data, an extended version with proxy data was created to demonstrate the full implementation, including `IfcPropertyEnumeration`, `IfcPropertyEnumeratedValue`, and `IfcComplexProperty`. The resulting IFC files were validated using the Bonsai add-on. The full source code, generated IFC files, and supporting spreadsheets are available online.

```
#13=IFCPROPERTYSET('3PqJDfAD140QLz7fZhO18q',$,'
    LCAPset_ProductInformation','Document reference:
    ISO 22057 8.2.2 - Table 2 and Table 3',(#14,#15,#
    16,#17,#18,#19,#20,#21));
#14=IFCPROPERTYSINGLEVALUE('NameOfOwner','https://
    identifier.buildingsmart.org/uri/LCA/LCA/3.0/prop/
    nameofowner',IFCLABEL('Jaro AS'),$);
#15=IFCPROPERTYSINGLEVALUE('UniqueIdentifier','https
    ://identifier.buildingsmart.org/uri/LCA/LCA/3.0/
    prop/uniqueidentifier',IFCIDENTIFIER('976 241 053'
    ),$);
#16=IFCPROPERTYSINGLEVALUE('WebDomain','https://
    identifier.buildingsmart.org/uri/LCA/LCA/3.0/prop/
    webdomain',IFCURIREFERENCE('https://jaro.no/'),$);
#17=IFCPROPERTYSINGLEVALUE('ProductName','https://
    identifier.buildingsmart.org/uri/LCA/LCA/3.0/prop/
    productname',IFCLABEL('Massiv betongelement -
    Vegg, lavkarbonbetong'),$);
```

Listing 1: Fragment of the "Massiv betongelement - Vegg lavkarbonbetong" EPD from EPD-Norge in IFC

Conclusions

This study contributes to the integration of product-level LCA data into IFC-based openBIM workflows. It reviews existing methods for storing, sharing, and reusing product data in BIM, together with available schemas for machine-readable LCA data. ISO 22057 is identified as the most suitable standard for product-level LCA

information, while the IFC catalogue format based on `IfcProjectLibrary`, as defined in ISO 17549-2 and ISO 16757-5, is selected as the main mechanism for integrating this data into IFC.

We propose an openBIM workflow for WBLCA that improves data longevity, traceability, documentation, and regulatory compliance, while reducing manual work for LCA practitioners. The approach relies on coordinated data dictionaries at national, European, and global levels. As a concrete implementation, BR18 §297, Annex 2, Table 6 is formalised, mapped to IFC classes, and digitised as a data dictionary in the bSDD.

A detailed IFC data model for integrating ISO 22057 into the IFC catalogue format is developed. Key modelling choices are discussed, including the handling of information modules, alternative scenarios, sub-scenarios, and different levels of EPD representation. These concepts are tested through a proof-of-concept tool that converts EPDs from the ISO 22057 JSON format to IFC-SPF. The approach is demonstrated using a real EPD from EPD-Norway and validated with Bonsai.

The proposed solution also has some limitations. Although it can be implemented in IFC-SPF today, many of the more advanced IFC features used, such as complex property structures beyond `IfcPropertySingleValue`, are not yet supported by most BIM software. This is a tooling limitation rather than a conceptual one and should be addressable if demand increases.

In addition, while ISO 22057 represents a significant step towards machine-interpretable EPDs, recent work has highlighted the remaining challenges that require further refinement of the standard Aragón et al. (2025).

This study focuses on product-level LCA data. Future research should extend the approach to building-level information, including a systematic analysis of information requirements and their implementation within IFC-based workflows.

References

Aragón, A., Spudys, P., Pupeikis, D., Nieto, O., and Alberti, M. (2025). Bridging Interoperability Gaps Between LCA and BIM: Analysis of Limitations for the Integration of EPD Data in IFC. Buildings, 15(15):2760.

Bari, R. d., Jorgji, O., Horn, R., Gantner, J., and Ebertshäuser, S. (2019). Step-by-step implementation of BIM-LCA.

Bolognesi, C., Bassorizzi, D., Balin, S., and Manfredi, V. (2025). An Overview on LCA Integration in BIM: Tools, Applications, and Future Trends. Digital, 5(3):31.

Erlandsson, M., Görman, F., Thrysin, A., Häkkinen, T., Eckerberg, K., Pesu, J., and Dalborg, M. (2024). Nordic view on data needs and scenario settings for full life cycle building environmental assessment.

Horn, R., Ebertshäuser, S., Di Bari, R., Jorgji, O., Traunspurger, R., and Both, P. v. (2020). The BIM2LCA approach: An industry foundation classes (IFC)-based interface to integrate life cycle assessment in integral planning.

Klumbyte, E., Georgali, P.-Z., Spudys, P., Giama, E., Morkunaite, L., Pupeikis, D., Jurelionis, A., and Fokaides, P. (2023). Enhancing whole building life cycle assessment through building information modelling: Principles and best practices. Energy and Buildings, 296:113401.

Kremer, N. C., Junge, M., Domme, E., and Beetz, J. (2025). Integration of Parametric Product Catalogue Requirements in IFC. volume 6 of Computing in Construction. European Council on Computing in Construction.

Lambertz, M., Theißen, S., Höper, J., Wimmer, R., Meins-Becker, A., and Zibell, M. (2019). LCA and BIM in sustainable construction. Technical report, Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) im Bundesamt für Bauwesen und Raumwesen (BBR).

Llatas, C., Soust-Verdaguer, B., Hollberg, A., Palumbo, E., and Quiñones, R. (2022). BIM-based LCSA application in early design stages using IFC. Automation in Construction, 138:104259.

Parekh, R. and Trabucco, D. (2024). Recent progress in integrating BIM and LCA for sustainable construction: A Review. International Journal of Science and Research Archive, 13(1):907–932.

Potrč Obrecht, T., Röck, M., Hoxha, E., and Passer, A. (2020). BIM and LCA Integration: A Systematic Literature Review. Sustainability, 12(14):5534.

Safari, K. and AzariJafari, H. (2021). Challenges and opportunities for integrating BIM and LCA: Methodological choices and framework development. Sustainable Cities and Society, 67:102728.

Schmid, H., Frischknecht, R., Werner, D. F., and Chipman, T. (2022). Industry Foundation Classes (IFC) for data sharing in the Construction and Facility Management Industries — Integration of Life-Cycle Phases and Environmental Impact Indicators in bSDD.

Theißen, S., Höper, J., Drzymalla, J., Wimmer, R., Markova, S., Meins-Becker, A., and Lambertz, M. (2020). Using Open BIM and IFC to Enable a Comprehensive Consideration of Building Services within a Whole-Building LCA. Sustainability, 12(14):5644.

Wastiels, L. and Decuypere, R. (2019). Identification and comparison of LCA-BIM integration strategies. IOP Conference Series: Earth and Environmental Science, 323(1):012101.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

GRAPH-BASED INSTANCE TRACEABILITY FOR BUILDING PRODUCT REUSE

Silvia Vangelova[1], Clara P. Blum[1], Kasimir Forth[1], Jens J. Hunhevicz[1,2], Pritha Jacob[1] and Catherine De Wolf[1]
[1]ETH Zurich, Zurich, Switzerland
[2]Empa, Dübendorf, Switzerland

Abstract

Tracing information related to the life cycle of building products supports decision-making for circular construction. However, current digital tools and standards in construction primarily address product identification at the type rather than instance level, resulting in information loss that hinders reuse. To trace products across projects at the instance level, we propose a graph-based workflow that maps globally unique identifiers (GUIDs) of element instances. While information continuity can be established manually, preliminary workflow evaluation in a case study reveals that manual GUID verification is impractical. We conclude that realising a digital reuse supply chain requires software automation for instance-level identification.

Introduction

Reuse in construction, one of the most resource-intensive industries, plays a key role in achieving a sustainable built environment (Lima et al., 2021; Eurostat, 2024). Reuse, in general, can be understood as a process that enables multiple use cycles of products (Alcayaga et al., 2019). Capturing and updating product life-cycle information at each use cycle and storing it throughout the product's lifetime is an important capability for enabling product reuse (Wan and Jiang, 2025). However, for construction products, this capability remains a challenge due to the long, multi-phased life cycle of buildings. Novel information management solutions are required to ensure long-term data availability and maintainability (Hunhevicz and De Wolf, 2023).

Current element-based information management solutions for building product reuse build on existing digital tools but remain only prototypical. To achieve instance-level traceability for reuse, potential solutions extend these tools by integrating globally recognised supply chain standards (Hradil et al., 2023). Specifically, the GS1 framework enables product instance identification using Serialised Global Trade Item Numbers (SGTIN) (GS1, 2024). By assigning a unique SGTIN to a physical product, this system distinguishes individual instances from otherwise identical products of the same type. Yet, maintaining instance-level traceability over a product's lifecycle remains challenging; critical data disconnects occur between phases even within a single project (Borrmann et al., 2018), let alone across multiple projects separated by years or decades.

In reuse scenarios, preserving instance-level traceability can be crucial, as secondary products often require individual structural validation or condition assessment. Type-level identification can lead to the loss of instance-specific information, such as material condition or load history, that often differs across reused elements of the same type (Blum et al., 2026). Although identification at the instance level is technically more challenging and often abandoned in favour of simpler type-level tracking, it remains valuable for reuse decisions (Byers and De Wolf, 2023).

While linking type- and instance-based identifiers via semantic models can help mitigate this information loss (Kebede et al., 2022), traditional building information modelling (BIM) software lacks the built-in functionality to implement such semantic models (Pauwels et al., 2017). This disconnect is compounded because BIM tools are primarily designed for construction using new products (Pocobelli et al., 2018), abstracting elements into replicable object-oriented types rather than tracing unique physical instances. By referencing a single parent definition to instantiate multiple elements of the same type (Costin et al., 2022), these applications prioritise computational efficiency at the expense of historical data continuity. Consequently, incorporating reused products into a design results in the instantiation of new elements with new application-generated identifiers, which can effectively overwrite their previous instance-level identity. Consequently, traditional software offers limited support for instance-level identification for product reuse.

Despite limited cross-project instance traceability, BIM authoring tools handle type-to-instance mapping within single projects very well by using relational data schemas internally, which can be translated into Industry Foundation Classes (IFC) for interoperable data exchange. The problem is that these data structures are enclosed within single project files, and there is no established mechanism to query highly granular, instance-level information across separate models. To overcome these native software limitations and explicitly define semantic relationships, BIM models can be linked to a graph in which elements are represented as nodes and their spatial relationships as edges. This approach decouples information from BIM models by enabling storage in graph databases (Cao et al., 2022) and can optimise life-cycle data management by linking project specifications to external product databases (Forth and De Wolf, 2025), facilitating disassembly sequencing (Sanchez et al., 2021) and tracing incremental changes

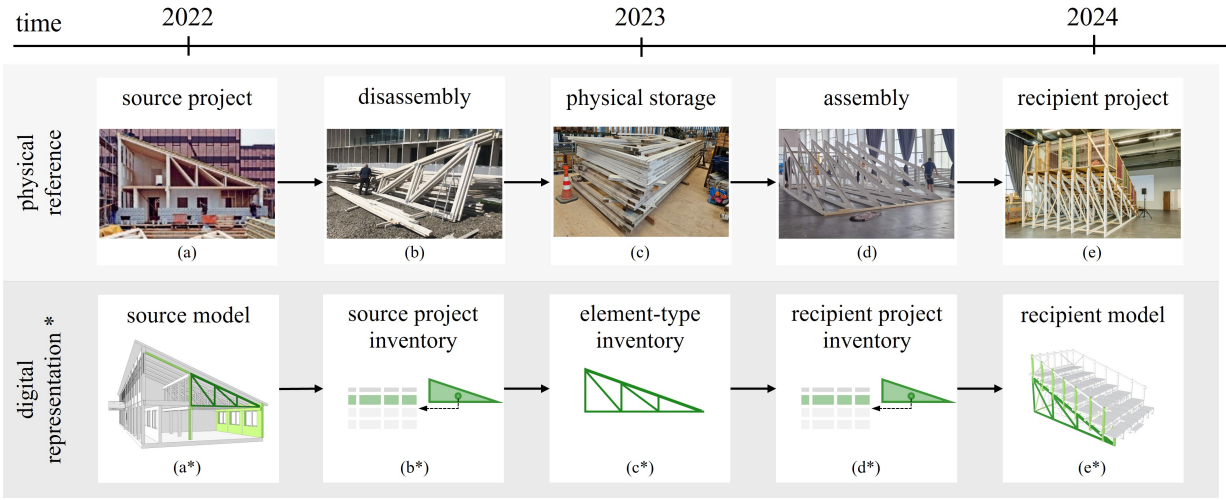


Figure 1: Physical and digital mapping of element instances across projects exemplified by the reuse of the Huber Pavilions. (*) Alignment in time is only approximate. Photo and model credit is provided in the Acknowledgements section

across project phases (Esser et al., 2023).

To address the gap in instance-level element identification, this study proposes a graph-based workflow for tracing product instances across different use cycles. The proposed workflow uses element-based graphs to map globally unique identifiers (GUIDs) from a source project to a recipient project, thereby preserving instance-level identification that is often lost during software exchanges. Following preliminary validation through a case study and a custom visualisation tool, we argue that establishing a digital reuse supply chain requires automated instance mapping that extends the capabilities of current design tools.

Point of Departure

Bridging physical and digital traceability

Reusing elements across multiple projects is a key strategy for enabling a circular economy in construction. However, the information needed to reuse an element is often lost at project handover or becomes outdated during operation and must be reacquired at the end of a building's life cycle (Honic et al., 2021). This acquisition process typically involves a building audit, during which a digital inventory of its constituent elements is created (Honic et al., 2019). To manage this inventory, identifiers such as quick response (QR) codes or near-field communication (NFC) chips can serve as a ''phygital'' link connecting a physical building element to its digital representation (Byers et al., 2025). Unlike the consumer goods industry, where traceability is optimised through linear distribution networks and statistically predictable behaviour, typically without the need for a digital model, construction operates on a project-by-project basis at a higher level of complexity. It involves longer time spans and various stakeholder groups that transfer multidisciplinary data between project phases. Consequently, construction requires coordinating products across two parallel timelines: the physical supply chain and the digital project-related information supply chain (Byers et al., 2025). This interdependency between the physical and digital sides of the construction supply chain makes it challenging to maintain consistent identifiers across both dimensions during reuse.

Type vs. instance traceability

Creating a building model for a project requires distinguishing between element types, which are defined by the shared properties of a product, and element instances, which identify each unique physical occurrence of that product and its specific positioning within the building. This distinction is well established in BIM, where buildings and products are represented in an object-oriented structure with classes and properties organised in a nested hierarchy. Specifically, the IFC schema models type-to-instance relationships in an object-oriented manner by separating universal specifications of a given product (IfcTypeObject) from the unique physical occurrences of this product (IfcObject) within a project (buildingSMART International, 2016). The IfcRelDefinesByType relationship maps each instance to its type, allowing element instances to inherit type-level properties from a single element type definition.

The IFC schema provides standardised property set definitions, such as the Tag property of IfcElement (buildingSMART, 2020), to distinguish specific element instances using serial numbers or application identifiers, such as GUIDs. The IFC Tag was originally intended for standardised global identifiers, such as those defined by GS1. This tag is often used to store element GUIDs, unique codes generated automatically by the authoring software to identify element instances. While these GUIDs are useful for tracking product changes, they are often unreliable for tracing at the instance level across different project phases because BIM authoring workflows often lead to generating inconsistent element identifiers (Esser et al., 2022).

The challenge of instance-level traceability

While there have been developments towards integrating GS1 identifiers into the buildingSMART Data Dictionary (bSDD) for IFC (GS1, 2025), current implementations for tracing across projects are more feasible for product types than for individual instances. One solution for tracing element information at the instance level is provided by Speckle. Speckle uses application IDs, the equivalent of element GUIDs, to maintain a link between an element and its authoring software, and hierarchical graphs to store project information. In these tree-like structures, elements contain other elements and references to elements stored outside the authoring application ecosystem. A suitable way to record these relationships is to store GUID key value pairs in a graph database. In the case of reusing elements from a source project in a recipient project, this requires mapping the original instance GUID from the source project to a product type ID (instance-to-type), and subsequently linking that type ID to the new instance-level GUIDs generated in the recipient project (type-to-instance). However, this process of instance and type mapping is where instance-level traceability breaks down due to information loss when models are transferred across different authoring software (Blum et al., 2026).

Scope and methodology of this study

Given the need to consistently trace product instances of the same type, along with their digital representations, across multiple use cycles in different projects, this study aims to develop a workflow for tracing the history of life-cycle-related information of reused building products across projects.

To answer the research question of how reused building elements of the same type can be digitally represented and identified at the instance level across projects, we analysed a reuse project at ETH Zurich to trace the flows of type- and instance-level information. Based on this analysis, we identified a cycle of instance-to-type mapping that repeats throughout the reuse process. This study therefore suggests that the digital side of the reuse supply chain should follow a cyclic pattern of recurring steps traversed using a graph-based workflow. Following this workflow, we propose that element GUIDs be mapped from source to recipient projects using a graph-based structure that enables the tracing of instance-level changes across product use cycles in line with the recurring steps.

Analysing Instance-to-Type Information Flow

Case study: the Huber Pavilions and Mehrwerk

As a case study, we trace the digital representation of several timber frames of the same type sourced from the deconstructed Huber Pavilions at the ETH Zurich Hönggerberg campus and their subsequent reintegration in Mehrwerk, an installation at the Kunsthalle Zürich (De Wolf, 2025). Specifically, we trace this single element type from its original function in supporting the pavilions' roof structure to its new function in supporting the seating platform of Mehrwerk, demonstrating how instance-level identification can be maintained throughout this transition.

We analyse the case study in five steps, tracing the physical reuse supply chain alongside its parallel digital representation (see Figure 1). Tracing begins with the source project (a), one of the Huber Pavilions (Knecht, 2022), for which a BIM model (a*) was created based on an audit conducted prior to disassembly. During the disassembly (b) of the building, a digital inventory of its constituent elements (b*) was created. The disassembled elements were then relocated to a physical storage location (c), where elements of the same type were physically grouped and represented digitally as a consolidated element-type inventory (c*). During the assembly (d) of the recipient project, Mehrwerk, a new inventory (d*) was created that specified the required elements. After completion of the recipient project (e), its final as-built model (e*) was created.

Identified instance-to-type mapping cycle for reuse

We observed a cycle of mapping instance and type relationships of building elements. We use the term ''cycle'' to encapsulate the pattern of recurring events in product use cycles, distinguishing these events from the physical ''circular'' flow of resources. This terminology emphasises that in reuse scenarios, a building project has one life cycle, whereas building products have multiple use cycles across different projects.

On the sourcing side, buildings at their end-of-life phase typically lack digital documentation. To facilitate reuse and environmental assessments, a digital model is created through a building audit. In this model, the geometry of each physical element is represented as a unique instance with a GUID automatically assigned by the software used to create the model. As the building is deconstructed, these elements lose their original context. When the elements are relocated to a storage facility, individual instance data is condensed into a digital inventory of elements grouped by type. In the recipient project, the steps for element GUID association are repeated by reusing the same elements in a new context. When a designer selects an element type from the inventory, they instantiate it within a new project model. This action generates new instance identifiers, effectively mirroring the initial flow of information by transitioning from the consolidated type-based inventory to element instances in the recipient project.

While current digital tools are built for project life cycles, e.g., managing a single building across specific project phases, they are not built for product use cycles. One example is managing a beam that is sourced from the end-of-life phase of one project to the initial design phase of another. Currently, there is no functionality that allows product information to be retained after the project's end-of-life phase for future use cycles. This is especially important during the early design phases, where implementing circular solutions has the greatest impact (Heisel and McGranahan, 2024).

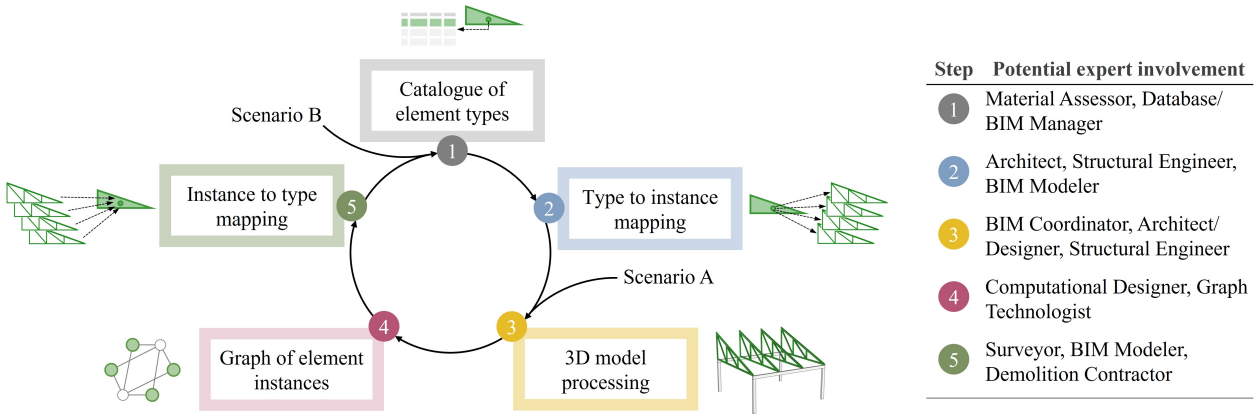


Figure 2: Graph-based workflow for product life-cycle information tracing. Potential expert involvement for each step is proposed. Scenario A starts with an inventory of reused products. Scenario B starts with the reuse of a building constructed with new products.

Workflow for Life-Cycle Information Tracing

In this section, we describe our proposed workflow for tracing information about multiple instances of a single element type in five steps, as illustrated in Figure 2.

In this workflow, we use four types of element instance and type representations (see Figure 3). Type a is a three-dimensional geometric representation, which can be a BIM or CAD model within the authoring software or an IFC export of it. Type b is the graph representation of the model, capturing the connectivities between different elements, where each node represents an element instance, and each edge is a spatial connection between two elements. Type c is an instance-to-type mapping, a key value-pair dictionary, where the keys are the element type IDs and the values are the instance GUIDs. Type d is a tabular dataset, which we refer to as a ‘‘catalogue’’, and can be either a digital inventory of physically stored elements or a specification table of the products to be used in a new design.

To better illustrate the graph-based workflow, we present two scenarios for initiating the reuse cycle. The first scenario (A) is when a building has no catalogue of the products used, it is built with new products, and is inventoried through disassembly for reuse. The second scenario (B) is designing using a given set of element types from an inventory of reused building products.

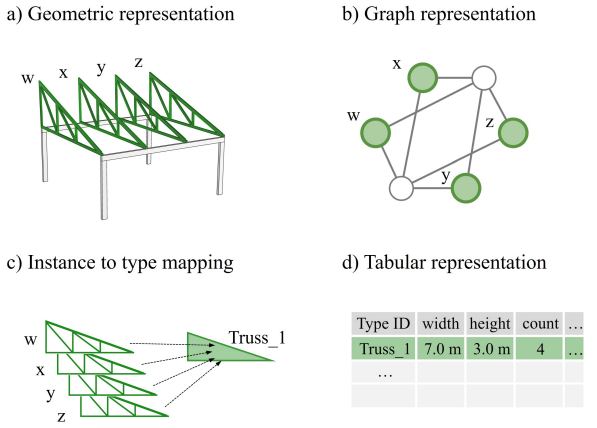


Figure 3: Element instance and type representations used in the graph-based workflow

In the following subsections, we describe our workflow based on scenario B, which consecutively loops through all steps in the circular loop (see Figure 2). The workflow begins with a catalogue of element types (1), which may include both reused and new products. During the design phase, these types are instantiated, establishing a type-to-instance mapping (2) where unique identifiers (GUIDs) are generated for the geometry of each element instance.

Once the design is finalised, 3D model processing (3) standardises the data via IFC export. This geometric data is then used to compute a graph of element instances (4), storing their geometric connectivity, with elements represented as nodes in the graph and linked via their GUIDs. Finally, the workflow concludes during the disassembly or end-of-use phase with instance-to-type mapping (5). In this step, specific element instances are remapped to new element types to repopulate an inventory for future projects, effectively starting a new use cycle.

Alternatively, for scenario A, the cycle can start with the model-processing step (3), as shown in Figure 2. This is how we applied the workflow to the case study since we do not have the original project specifications that catalogue all products used in the building. Additionally, the source building was constructed with new products rather than an inventory of reused ones.

Step 1: Catalogue of element types

Whether creating a model for new construction or reuse projects, the workflow starts with a catalogue of element types. In projects that reuse building elements, this catalogue is usually derived from an inventory, a finite set of reusable elements stored in a physical location, most of which come from a previous building. In new construction projects, this catalogue functions similarly to a bill of materials, specifying the product types that will be assigned to the element instances in the model. In this study, we define the catalogue of element types as a unified tabular

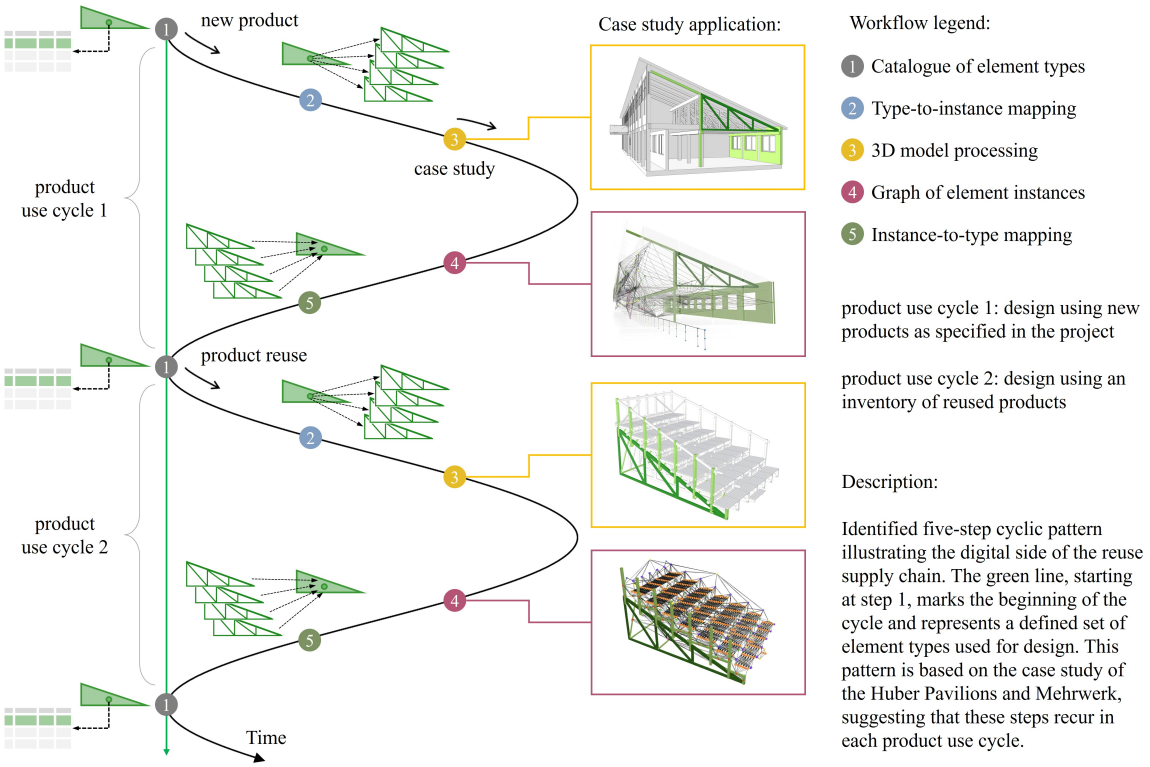


Figure 4: Identified instance-to-type mapping cycle for reuse and its application in the case study

dataset, regardless of whether they are reusable elements or new products. Defined by attributes such as type ID, dimensions, and available quantity, this catalogue serves as a predefined set of element types and is used in the next step of the workflow, where elements are instantiated in the authoring software.

Step 2: Type-to-instance mapping

Type-to-instance mapping occurs implicitly as geometry is created within the authoring software during the design process. The designer assigns an element type ID, which is stored as both a column in the inventory data table and a specific property of the element instance in the model. This linkage is critical for the subsequent model processing in Step 3 because it maintains traceability between the element type and the instance GUID generated by the authoring software in later steps.

Step 3: 3D model processing

To extract and retain relevant life-cycle-related information, we use two distinct processes, depending on the authoring software type of the source file, which are applicable for both scenario A and B. Both processes are based on the IFC schema because it is an open data exchange format and supports product identification through geometric representation and unique identifiers, namely, the tag property of IFC building elements. We differentiate between the two processing workflows:

1. BIM-based processing (e.g., Revit, Archicad): We export these models from native authoring software directly into IFC4, ensuring base quantities and geometry are included in the export configuration.
2. CAD-based processing (e.g., Rhino, AutoCAD): We first ensure the geometry of element instances is assigned to corresponding element-type layers. The model is exported in STEP format, attaching layer information to the geometric representation of each element. We then use IfcOpenShell to convert the STEP file to IFC4, and pythonOCC, a Python wrapper of OpenCASCADE, to generate the geometric representation and compute base quantities.

Step 4: Graph of element instances

Using the IFC file generated in the previous step, we compute a geometric connectivity graph to store how the elements are connected. For this step, we further developed an existing IFC viewer (Forth and Emenegger, 2025) to compute the graph in the back end and edit IFC properties in the front end (see Figure 5). The graph computation is performed using an additionally developed Python back end for the IFC-graph editor, where geometry is extracted using IfcOpenShell and OpenCascade. To optimise performance, we estimate the number of connected pairs based on the distance between oriented bounding boxes. For each candidate pair, we then analyse the actual geometry; if the distance between elements is less than 1 cm, then they are considered to be connected. We justify this assumption because a tolerance of 1 cm to 2 cm falls within standard allowable deviations for building construction (DIN, 2019). Using the final list of pairs, we con-

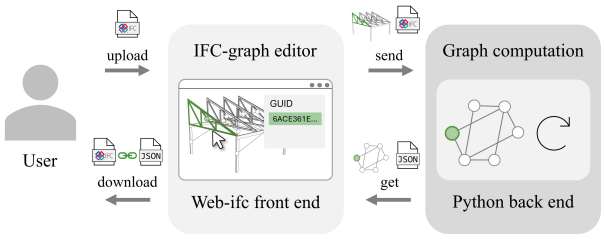


Figure 5: User interaction for assigning a GUID to an element and linking the geometric connectivity graph to the IFC file

struct a graph with NetworkX, attaching the corresponding GUID to each node and convert it to JSON.

Through this graph, for each node of an element, a list of connected elements can be retrieved. The resulting graph is stored as a NetworkX-based JSON. This approach was chosen because JSON is supported by the application programming interfaces (APIs) of commonly used BIM software, allowing the graph and GUIDs to be linked to their respective elements in the authoring software. For example, by using plugins to serialise and deserialise the JSON of the graph. To preserve the link between the graph and the IFC, we assign a unique identifier to the graph itself and store it within a custom IfcPropertySet as an ifcText.

Step 5: Instance-to-type mapping

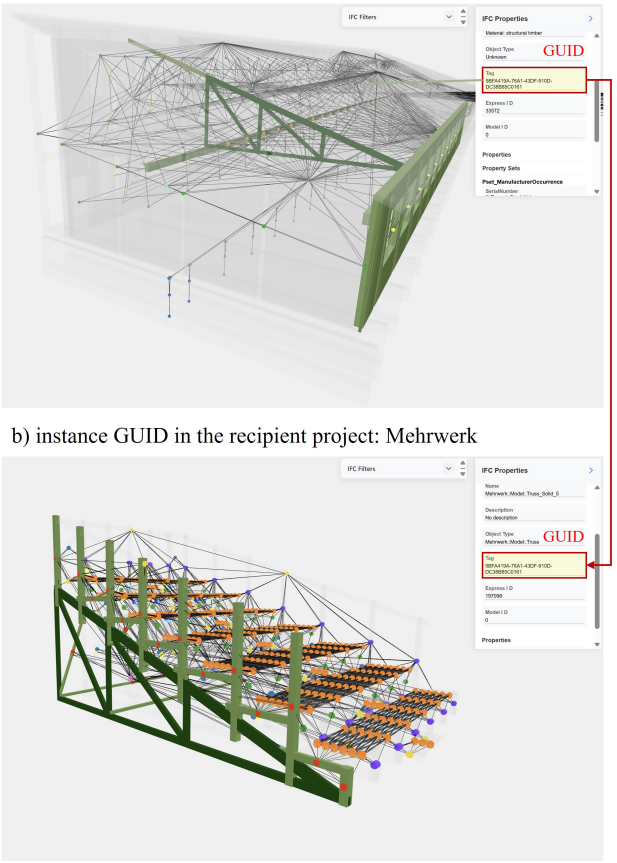
Unlike in Step 2, where mapping occurs in the authoring application, in Step 5, we use the IFC-graph editor to update the mapping between specific instance GUIDs and type IDs. This step is necessary because the as-built model often diverges from the initial design model yet requires the original type IDs to be relinked with the final element GUIDs.

The front end of the IFC-graph editor is used to visualise the IFC files and the graph computed in Step 4. The viewer comprises a TypeScript front end based on web-ifc from That Open Company and the previously mentioned Python back end (see Figure 5). Through this interface, we manually verify and map GUIDs to type IDs for the selected elements to be reused. The application then exports an updated IFC with the correct GUIDs and type IDs and a linked JSON graph, where key value pairs of element GUIDs preserve the semantic link between the elements and their connected neighbours for the next use cycle.

Preliminary Evaluation and Future Work

For initial validation, we applied the graph-based workflow to the case study of the Huber Pavilions and Mehrwerk (see Figure 4) by tracing the reuse of timber frames from the source project (Huber Pavilions) to the recipient project (Mehrwerk). Originally, while the frames were geometrically identical, the transfer between software applications during the design of Mehrwerk resulted in a loss of instance-level traceability. Furthermore, the columns connected to the frames in the source project were modified in the recipient project, creating a distinct set of connections that required a new inventory version. Using our workflow,

a) instance GUID in the source project: Huber Pavilions



b) instance GUID in the recipient project: Mehrwerk

Figure 6: Manual GUID assignment of selected elements using the IFC-graph editor. This action modifies the original IFC model and links it to a graph through the assigned GUIDs.

we re-link the timber frames in both models.

The workflow begins with the source project. We exported the Huber Pavilion model from Archicad as IFC4 (Step 3). Then, using the IFC-graph editor, we computed the element connectivity graph (Step 4). We then mapped the type-to-instance relationships by assigning a specific Type ID property to the elements (Step 5). To link the frames from the source to the recipient project, we initiated the workflow by storing their element type IDs in a table (Step 1). We then assigned these types to the corresponding geometry in the Mehrwerk model (Step 2) before exporting it from Rhino and converting it to IFC4 (Step 3). Using the IFC-graph editor, we computed the graph for this new context (Step 4). Finally, we assigned the original GUIDs from the Huber model to the new frame instances (Step 5). This way, we were able to re-link the element instances across both models (see Figure 6).

Challenges and limitations

A limitation in the case study is the inability to verify the instance-level GUID assignment. While we successfully mapped element GUIDs across the source and recipient models, we retrospectively lacked ground-truth data to confirm which timber frame from the source project cor-

responded to which instance in the recipient project. Furthermore, in the case study, no instance-level differentiation has been implemented. This disconnect highlights a broader issue: collecting and maintaining instance-level data across the reuse supply chain is currently not realised. Even though the instance-to-type mapping solution in the IFC schema allows for the preservation of instance-level variables, manually verifying that every element has a correct GUID corresponding to a physical product requires significant effort, particularly during early design iterations, when geometry is rapidly changing. The labour of manual verification is also due to long project life cycles, during which information is lost across different software.

Although GS1 standards specify the general requirements for instance-level identification through SGTINs, current construction software lacks the capability to integrate these standards efficiently. In particular, tracing building products composed of multiple sub-components or layers is challenging to integrate, as these can be disassembled and reassembled into different elements. To enable instance traceability without the overburden of manual tagging, algorithms need to be developed to automatically detect when a digital element instance in a new design contextually represents a physical element in the inventory, and maintain this link throughout further project phases. On the physical side, a potential solution to these efforts is to use extended reality for the identification of as-built elements. On the digital side, algorithms for measuring geometric similarity between elements can help identify matching pairs (Collins et al., 2024).

Regarding the IFC-graph editor developed for this study, user feasibility testing is needed to validate better graph visualisation approaches and determine whether the front-end graph visualisation actually supports instance-level identification or should be kept in the back end. Future research will focus on formalising the graph-based representation and connecting the viewer to a graph database, thereby enabling the integration of dynamic digital passports that evolve throughout product use cycles.

Conclusion

In this study, we developed a graph-based workflow for tracing product use cycles across projects. We demonstrated the workflow using an IFC-graph editor to map element GUIDs across two projects, using the Huber Pavilions and Mehrwerk as a case study. The workflow shows that preserving element identification outside the original authoring software is a viable strategy to reduce vendor lock-in and facilitate product data transfer across use cycles. However, data provenance remains to be addressed, as individual product data can be isolated from project databases with varying levels of access.

Making instance mapping more efficient by linking digital passports to graph databases of building projects will facilitate reuse. If realised, this functionality would serve two purposes: it would allow building owners and reuse partners to trace the life-cycle history of building products effectively while simultaneously supporting designers to integrate reused products into new designs. Applying the learnings from the case study of the Huber Pavilions and Mehrwerk to active projects will be the next step in advancing the graph-based reuse workflow.

Acknowledgments

This research has been funded by the Flagship projects SWIRCULAR (107.235 FS-EE) and ThinkEarth (107.163 FS-ENG), both supported by Innosuisse. The authors would like to thank Vanessa Schwarzkopf, Elias Knecht and Heidi Silvennoinen for providing the models and on-site photos of Mehrwerk and the Huber Pavilions. In addition, the authors gratefully acknowledge Dr. Roxanne Goldberg for valuable editorial support.

Credits in Figure 1:
(a) Photo by gta Archiv / ETH Zurich (Estate of Benedikt Huber);
(b) Photo by Elias Knecht, Master student, ETH Zurich;
(d) Photo by Vanessa Schwarzkopf, CEA, ETH Zurich;
(e) Photo by Anna Buser, CEA, ETH Zurich;
(a*) 3D model by Heidi Silvennoinen, CEA, ETH Zurich;
(e*) 3D model by Vanessa Schwarzkoph, Vincent Jackow, Tim Cousin, CEA, ETH Zurich;

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this paper, the authors used Google Gemini 3 Pro to enhance the clarity of the text. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

Alcayaga, A., Wiener, M., and Hansen, E. G. (2019). Towards a framework of smart-circular systems: An integrative literature review. Journal of Cleaner Production, 221:622–634.

Blum, C., Vangelova, S., Hunhevicz, J., Forth, K., and De Wolf, C. (2026). Breakpoints in data continuity across construction reuse projects.

Borrmann, A., König, M., Koch, C., and Beetz, J. (2018). Building Information Modeling: Why? What? How? In Borrmann, A., König, M., Koch, C., and Beetz, J., editors, Building Information Modeling: Technology Foundations and Industry Practice, pages 1–24. Springer International Publishing, Cham.

buildingSMART (2020). IfcElement attribute definitions. https://standards.buildingsmart.org/IFC/RELEASE/IFC4/ADD2/HTML/link/ifcelement.htm.

buildingSMART International (2016). IfcRelDefinesByType. https://standards.buildingsmart.org/IFC/RE-

LEASE/IFC4/ADD2/HTML/link/ifcreldefinesbytype.htm.

Byers, B. S. and De Wolf, C. (2023). QR Code-Based Material Passports for Component Reuse Across Life Cycle Stages in Small-Scale Construction. Circular Economy, 1(2):1–16.

Byers, B. S., Stengele, K., and De Wolf, C. (2025). Data carriers for circular construction supply chains: An exploratory quantitative analysis. Journal of Cleaner Production, 494:145053.

Cao, J., Bucher, D. F., Hall, D. M., and Eggers, M. (2022). A graph-based approach for module library development in industrialized construction. Computers in Industry, 139:103659.

Collins, F. C., Braun, A., and Borrmann, A. (2024). Finding Geometric and Topological Similarities in Building Elements for Large-Scale Pose Updates in Scan-vs-BIM. In Skatulla, S. and Beushausen, H., editors, Advances in Information Technology in Civil and Building Engineering, pages 517–530, Cham. Springer International Publishing.

Costin, A., Ouellette, J. W., and Beetz, J. (2022). Building product models, terminologies, and object type libraries. In Pauwels, P. and McGlinn, K., editors, Buildings and Semantics: Data Models and Web Technologies for the Built Environment, pages 3–24. CRC Press, London.

De Wolf, C. (2025). The Art of Connecting: Reuse of the Huber Pavilions. Edition Detail, Munich.

DIN (2019). Tolerances in building construction - Buildings. Standard DIN 18202, Deutsches Institut für Normung.

Esser, S., Vilgertshofer, S., and Borrmann, A. (2022). Graph-based version control for asynchronous BIM collaboration. Advanced Engineering Informatics, 53:101664.

Esser, S., Vilgertshofer, S., and Borrmann, A. (2023). Version control for asynchronous BIM collaboration: Model merging through graph analysis and transformation. Automation in Construction, 155:105063.

Eurostat (2024). Material footprints - details by final use of products. https://doi.org/10.2908/ENV_AC_RMEFD.

Forth, K. and De Wolf, C. (2025). Formalizing Information for Disassembly Potenttial of Buildings Using BIM and Labeled Property Graphs. volume 6. European Council on Computing in Construction.

Forth, K. and Emenegger, P. (2025). IFC and Neo4j Disassembly Graph Viewer. https://github.com/cea-ethz/disassembly-graph-viewer.

GS1 (2024). GS1 Standards enabling the Digital Product Passport. Whitepaper, GS1. https://gs1.eu/wp-content/uploads/2024/12/GS1-Standards-Enabling-DPP.pdf.

GS1 (2025). bSDD GS1 Identifiers. https://identifier.buildingsmart.org/uri/gs1/gs1/0.1.0.

Heisel, F. and McGranahan, J. (2024). Enabling Design for Circularity with Computational Tools. In De Wolf, C., Çetin, S., and Bocken, N. M. P., editors, A Circular Built Environment in the Digital Age, pages 97–110. Springer International Publishing, Cham.

Honic, M., Kovacic, I., Aschenbrenner, P., and Ragossnig, A. (2021). Material Passports for the end-of-life stage of buildings: Challenges and potentials. Journal of Cleaner Production, 319:128702.

Honic, M., Kovacic, I., and Rechberger, H. (2019). Concept for a BIM-based Material Passport for buildings. IOP Conference Series: Earth and Environmental Science, 225(1):012073.

Hradil, P., Jaakkola, K., and Tuominen, K. (2023). RFID-based traceability system for constructional steel reuse. In Life-Cycle of Structures and Infrastructure Systems. CRC Press. Num Pages: 8.

Hunhevicz, J. and De Wolf, C. (2023). Blockchain for a Circular Digital Built Environment. In IoT, Blockchain, and Smart Environments. Design Computation.

Kebede, R., Moscati, A., Tan, H., and Johansson, P. (2022). Integration of manufacturers' product data in BIM platforms using semantic web technologies. Automation in Construction, 144:104630.

Knecht, E. (2022). Future Learning Spaces - Re-Used.

Lima, L., Trindade, E., Alencar, L., Alencar, M., and Silva, L. (2021). Sustainability in the construction industry: A systematic review of the literature. Journal of Cleaner Production, 289:125730.

Pauwels, P., Zhang, S., and Lee, Y.-C. (2017). Semantic web technologies in AEC industry: A literature overview. Automation in Construction, 73:145–165.

Pocobelli, D. P., Boehm, J., Bryan, P., Still, J., and Grau-Bové, J. (2018). BIM for heritage science: a review. Heritage Science, 6(1):30.

Sanchez, B., Rausch, C., Haas, C., and Hartmann, T. (2021). A framework for BIM-based disassembly models to support reuse of building components. Resources, Conservation and Recycling, 175:105825.

Wan, P. K. F. and Jiang, S. (2025). Enabling a dynamic information flow in digital product passports during product use phase: A literature review and proposed framework. Sustainable Production and Consumption, 54:362–374.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATED QUANTITY TAKEOFF AND LIFE-CYCLE COSTING IN CONSTRUCTION PROJECTS USING AN OPENBIM–IFC APPROACH

Farzad Jalaei[1], Arash Hosseini Gourabpasi[1], Abdulwahed Fazeli[2], Ahmad Jrade[3]
[1]National Research Council Canada, Ottawa, Canada
[2]Toronto Metropolitan University, Toronto, Canada
[3]University of Ottawa, Ottawa, Canada

Abstract

BIM-based cost estimation often depends on proprietary tools and ad-hoc mappings to external cost databases, limiting interoperability and auditability. This paper presents an openBIM workflow that enriches material definitions with standardized Canadian-specific cost items and automatically generates element-level quantity takeoff and capital cost breakdowns. A case study demonstrates automated extraction of material, labor, and equipment costs, as well as structured inputs for maintenance-and-replacement-oriented life-cycle cost analysis. In the current implementation, automated computations focus on construction-phase capital costs, while LCC support is limited to structured inputs for external RSMeans-based estimators.

Introduction

Accurate cost estimation remains a continuous challenge in construction projects, where early decisions are made under uncertainty and information is fragmented across stakeholders and tools (Lee et al., 2012). Building Information Modeling (BIM) has enabled automated quantity takeoff (QTO) and facilitated 5D BIM workflows, yet many BIM-based cost estimation pipelines remain tightly coupled to specific authoring platforms or proprietary estimating systems (Fazeli et al., 2021). These constraints impede reproducibility, limit integration with alternative cost databases, and complicate longitudinal use of cost information across the asset life cycle (Ramaji et al., 2020).

OpenBIM addresses these limitations by promoting vendor-neutral data exchange through the Industry Foundation Classes (IFC) schema (Hosseini Gourabpasi et al., 2025; Laakso & Kiviniemi, 2012). IFC is standardized as ISO 16739 and is widely used for interoperable model exchange. However, the representation of cost information in open standard formats is still uneven in practice: cost data are frequently exported as spreadsheets, maintained in separate systems, or stored as project-specific properties without consistent semantics (Khudhair et al., 2023, Hosseini Gourabpasi et al., 2025).

This study develops and evaluates an openBIM-based framework for automated QTO and costing in which cost items are explicitly linked to IFC material definitions and then propagated to building elements for computation. In addition, the framework extracts a defined set of building attributes to support LCC inputs focused on preventive maintenance and replacement planning (ISO 15686-5). The paper also addresses whether embedding detailed cost parameters within the BIM/IFC model is preferable to keeping them in linked external containers. Accordingly, a balanced assessment and approach relative to BIM-Cost-Container frameworks (e.g., DIN 18290-3) is proposed that packages cost determinations externally while maintaining links to model objects. According to ISO 15686-5, life-cycle costing covers acquisition, operation, maintenance, and end-of-life costs over the asset's service life. In this study, the term LCC is used in a narrower, implementation-specific sense: the proposed workflow automates construction-phase capital cost estimation based on IFC quantities and RSMeans unit rates and extracts structured model attributes (e.g., building area, number of floors, perimeter) needed by RSMeans' online LCC estimators for maintenance and replacement planning. Operational energy costs, water and utility charges, and explicit end-of-life scenarios are outside the current prototype's scope and are identified as targets for future work.

Literature Review

BIM-driven cost estimation has been studied for the past two decades and remains an active area of research, including ontology-based approaches for measurement rules and standardized cost semantics (Abanda et al., 2017; Lee et al., 2014, Fazeli et al., 2024), and automated QTO methods for compound and interior elements (Khosakitchalert et al., 2019; Kim et al., 2019, Fazeli et al., 2022). Several studies integrate BIM with cost databases to automate estimation or to support 5D BIM workflows (Elbeltagi et al., 2014; Banihashemi et al., 2022). Previous studies have also explored Industry Foundation Classes (IFC)-based cost estimation pipelines and extensions for construction-related data exchange (Ma et al., 2013; Cassandro et al., 2024; Zhang et al., 2024).

Despite this progress, two gaps are highlighted. First, many implementations rely on exports from proprietary BIM tools and do not preserve a consistent, reusable linkage between model objects and cost items in an open schema. Second, the term life-cycle costing (LCC) is used inconsistently: some studies address full whole-life costing across construction, operation, and end-of-life,

while others focus primarily on construction-phase capital costs or limited maintenance scenarios (Rad et al., 2021). This paper explicitly distinguishes capital cost estimation based on quantities and unit rates and a maintenance-and-replacement-oriented LCC view enabled through structured data extraction and mapping to an external LCC estimator that is precisely used in Canadian context. In parallel, European standardization efforts have advanced the concept of linked containers for multimodal exchange. DIN 18290-2 (2023) defines a BIM-LV container for linking models with bills of quantities, and DIN 18290-3 (2023) defines a BIM cost container for structured exchange of model-based cost determinations linked to IFC reference objects. These container approaches externalize cost documents while maintaining traceable links to the BIM model, offering an alternative to embedding detailed cost parameters directly within the model.

Methodology

This study provides an openBIM-based approach to cost estimation and preparing the model for a high-level LCC. Figure 1 shows the overall design of the current study, which began with quantitative inputs to establish an OpenBIM-based cost-estimation framework. Quantitative inputs include both IFC External database and IFC Internal database. In the IFC External database, cost specification database (RSMeans) will be connected to the IFC Materials of the model. This database includes all Canadian-specific cost information for materials, equipment, labour, and total cost, along with their corresponding Overhead & profit costs. In addition, each item taken from the RSMeans dataset includes ID, Name, Unit, and Crew, which specify the type of material, unit, and workers used in the cost estimation process. This information provides full cost estimates for every element in the IFC model. IFC Internal database also includes all information about quantities of elements, including Area, Volume, and Length and Layer Thickness. The entire process of cost estimation, including the connection between cost information and model quantity, as well as its corresponding calculations, is in the simulation stages, where it evaluates the causes and impacts on various variables and potential reactions in various scenarios (Privé and Errico, 2016).

The proposed methodology comprises four steps. The first step is to develop an IFC model. The next step is to connect cost database into IFC materials. The authors developed a new framework to automatically import all cost data into IFC materials. Following this, the user must enrich elements inside the IFC Model with RSMeans database and new developed IFCMaterials. Finally, the cost estimation process has been conducted using the IFC geometric information and IFC materials inside the model. The implementation uses IfcOpenShell to read and write IFC, create and attach property sets, and compute quantities from IFC physical quantity definitions.

1) Develop an IFC model: 3D modelling capability is regarded as one of the most important benefits of OpenBIM application. 3D modelling can help visualize both elements and data within elements, enabling project practitioners to manage and control time and cost effectively and other data within the project. Accordingly, the whole list of cost items can be stored in a material list of elements in the 3D modelling as inputs for cost estimation.
   i) Required IFC model and Level of Information Need: The proposed workflow presumes an IFC model with necessary data to support quantity takeoff and material-based costing. Concretely, the following minimum Level of Information Need (LOIN) is required for reliable operation:
      a. Geometry: Explicit solid or layered representations for main elements (IfcWall, IfcSlab, IfcCovering, IfcWindow, IfcDoor) with consistent units, enabling derivation of length, area, and volume from IFC physical quantities and material layer sets.
      b. Materials: Each cost-relevant assembly must be associated with one or more IfcMaterials or IfcMaterialLayers that can be enriched with RSMeans-linked properties (cost item ID, unit, formula, unit rates).
      c. Quantities: Either IfcElementQuantity sets or material layer quantities must be present or derivable (e.g., base quantities for walls, slabs, ceilings), as the prototype relies on these definitions for automated QTO
      d. Identification: Stable global identifiers for elements and materials (GlobalId, consistent naming conventions) to support repeatable mapping between IFC objects and cost items.

   In cases where these conditions are not met—for example, models lacking material assignments or relying solely on symbolic representations the workflow may still operate but will require additional pre-processing, manual material mapping, or default quantity assumptions.
2) Develop new IFC Materials with cost specifications: In this step of the process, a list of the most important cost items is retrieved from the RSMeans database. The ID, name, and unit shows the identity of material and units. Crew specify the type of workers needed for the cost estimation process. Then, all other information are the cost information in RSMEANS database. The formula also specifies which geometric information of elements is needed for the cost estimation process. A new framework is developed to import all RSmeans database inside IFC Materials of the project using IFCOpenShell.
3) Equip IFC Elements with new IFC Materials: In this step, project managers can either replace Materials inside elements with IFCMaterials or create new elements with new IFC Materials.
4) OpenBIM-based Cost estimation: In this step, Quantities are retrieved from IFC physical quantities,

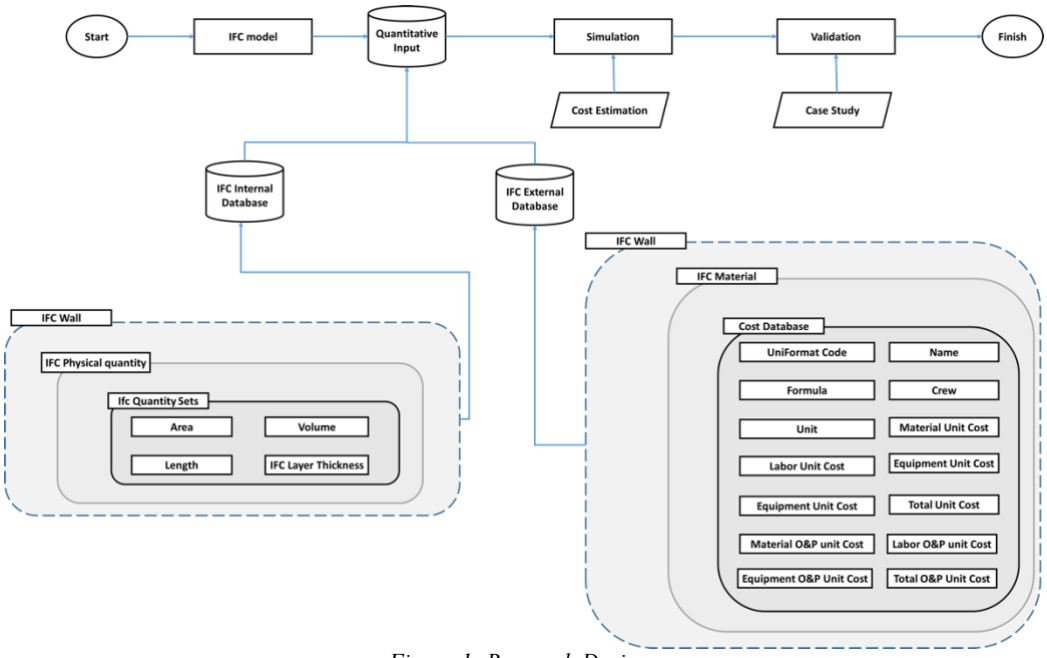


Figure 1: Research Design

evaluated against the cost formula, and multiplied by unit rates to produce element-level cost components. Results are stored back into the IFC as element properties to enable downstream queries, auditing, and visualization. The formula is used to change all units to SI units using a coefficient. Figure 2 shows the overview of the calculation made for cost estimation.

This workflow is most beneficial from late schematic design through detailed design and pre-construction, when key assemblies and material systems have been defined but design iterations and cost feedback are still frequent. At these stages, the IFC model typically contains sufficient geometry and material assignments to support reliable quantity takeoff while remaining flexible enough for design changes, enabling rapid re-estimation as materials, assemblies, or quantities evolve. During early conceptual design, the approach may be used with simplified template assemblies and approximate quantities, whereas in later construction phases it can support auditing and change impact analysis provided that as-built models and updated cost data are available.

Figure 3 shows how IFC structure for wall elements was used to implement the cost estimation of current study. As shown in the figure, the prototype first finds the elements inside the IFC Model. The IFC has standard structure to reach elements of the IFC model. In order to reach elements, firstly, IFCbuidingproduct will be found from IFCRoot. Inside IFC product, there are many elements including IFCWall, IFCSlab, IFCCovering, IFCDoor, and IFCWindows. The next step is to find IFCMaterials of elements inside the element. IFCRelAssociatesMaterials will be used to reach IFCMaterials of elements. Then, all RSMEANS information will be extracted from IFCMaterials through IFCMaterialProperties. The quantities of elements will be extracted through both IFCQuantitySets and IFCMaterialLayerSets. Finally, both quantities and RSMEANS data will be used to

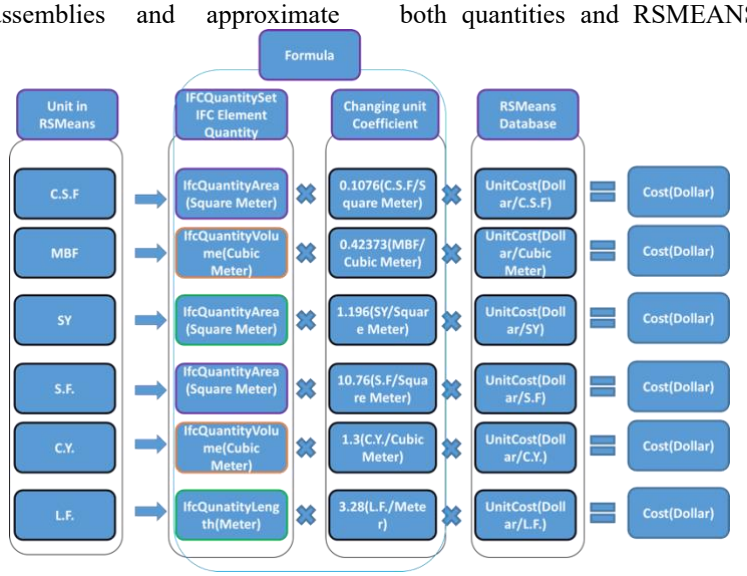


Figure 1: Cost estimation process

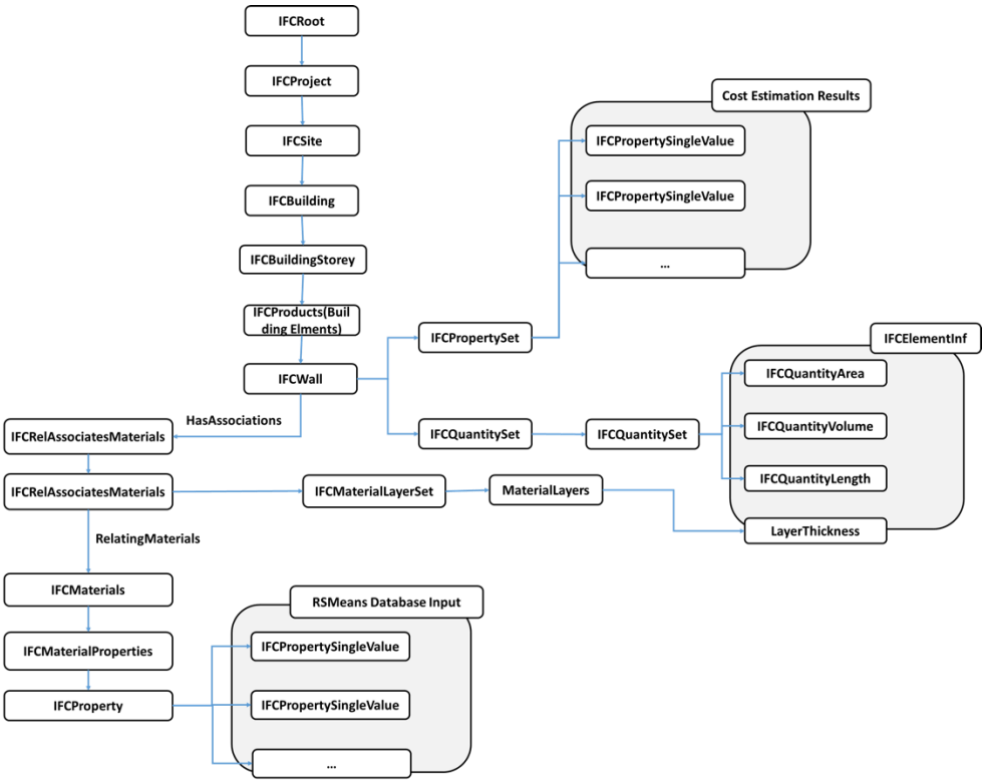


Figure 2: IFC-linked cost data architecture used for estimation

estimate the cost of project. The results of cost estimation will be imported through new developed property inside IFCPropertySet of elements.

The RSMeans website can estimate LCC costs of IFC project using some specific information from IFC project. The process of getting this information from IFC project is very challenging and time consuming. A new framework is needed to retrieve this information from IFC project. The LCC component of the workflow does not compute a full whole-life cost model inside IFC. Instead, it automates extraction of RSMeans-required input parameters (gross area, floor count, floor heights, perimeter) from the IFC model and delegates detailed maintenance and replacement costing to RSMeans' external LCC engine. As a result, the current prototype effectively covers construction-phase capital costs internally, while mid-life maintenance and replacement are handled externally; operational energy performance and end-of-life stages will require coupling with BIM-to-BEM methods and deconstruction/disposal datasets in future extensions. Therefore, In this step of the study, a framework has been developed to retrieve LCC information from IFC project to estimate LCC cost using RSMeans database. This framework can retrieve information, including the number of floors, the total floor area, the floor heights, and the total perimeter of the project. This information can be the primary input on the RSMeans website for LCC estimation. After entering this information on the RSMeans website, the total lifecycle cost is estimated.

Figure 4 shows the overview of the process. The first step is to retrieve LCC information from IFC model. A new prototype is developed to automatically extract all LCC information from IFC model. Then, the information should be imported in the next step and finally, LCC would be estimated from RSMeans. Using the current prototype saves a lot of time and prevents errors.

In this study, a platform has been developed to estimate the cost of construction project. Figure 5 shows the overall view of new developed framework. The first part is element part which has three types of elements including Wall, Floor, and ceiling. The main purpose of element part is to implement quantity take off from all parameters inside IFC model. The second part is edit part. The first button in edit part is the import RSMeans data. This button aims to import all RSMeans database inside

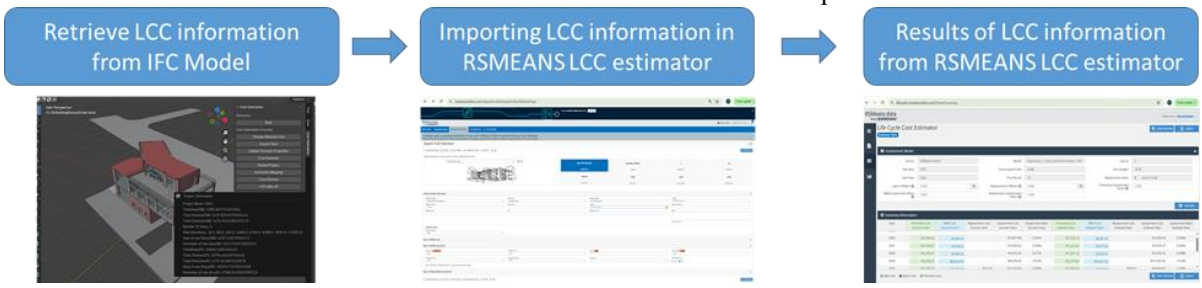


Figure 3: LCC estimation process

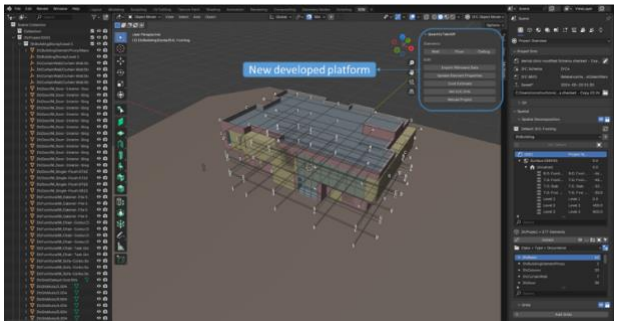


Figure 5 . Overview of the proposed OpenBIM costing workflow

materials of elements in the IFC project. The main purpose of second button (Update Element Properties) is to develop cost parameters in the elements of the project. The third button implement cost estimation from IFC project and update cost parameters in the IFC project.

The proposed framework is designed to align cost properties with shared semantics rather than ad-hoc custom fields. The buildingSMART Data Dictionary (bSDD) can host definitions for cost-related properties and classifications, enabling software-agnostic referencing of property meaning. Thus, Information Delivery Specifications (IDS) can be authored to formally specify which cost properties must be present for a given exchange (e.g., required cost code, unit, currency, and date), enabling automated compliance checking of IFC submissions before estimation. While full operationalization of bSDD- and IDS-driven governance is left as future work, these standards provide a practical path to scale the approach to other cost databases, regional cost indices, and organizational taxonomies without changing the underlying IFC computation logic.

Case Study and Results

The proposed framework was tested on a two-story office building in Toronto as a representative case. This building has a total gross area of 1005.35 m^2 and a net area of 781.91 m^2. The two-story office building model used in this study corresponds to a design-development level of detail, which aligns with the intended primary use-phase of the proposed workflow.

The model includes layered wall, floor, and ceiling assemblies with explicit material assignments. RSMeans unit rates were mapped to the enriched IFC material library, and quantities were extracted from the IFC physical quantities. Table 1 shows the results of cost estimation by the developed framework. The current methodology can estimate three main costs within the project: Material Cost, Equipment Cost, and Labour Cost, for both construction costs and Overhead and Profit costs.

The next step is to estimate the LCC cost of the office building from IFC format using RSMeans website. Figure 6 shows the LCC information required for LCC cost estimation, based on RSMeans requirements.

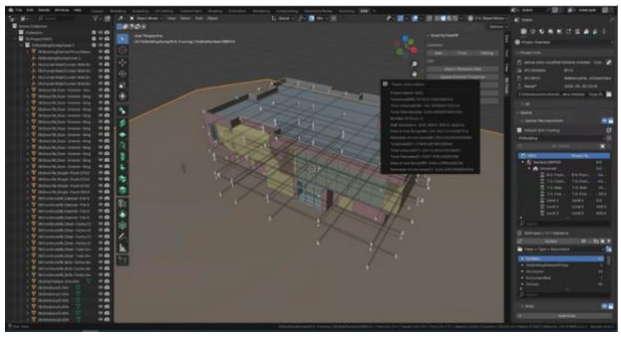


Figure 6: LCC information collection from new developed tool

Figure 7 shows an overview of maintenance cost results from RSMeans, including maintenance costs for the next 10 years, estimated for the case project.

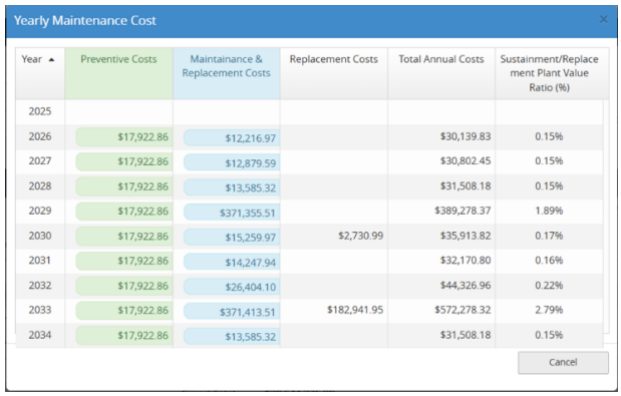


Figure 7. LCC information results from RSMeans

Discussion

The main contribution in this study is to propose an explicit and reusable linkage and between openBIM (in IFC data exchange format for objects like materials and assemblies) and standardized cost items, implemented without dependence on a specific BIM authoring tool or estimating platform. By storing both the mapping (cost code, unit, and formula) and the computed results in the IFC, the workflow improves auditability and supports downstream uses such as element-level cost visualization, change impact analysis, and exchange to other systems via non-proprietary solutions.

Table 1: Cost estimation results

	Material Cost	Equipment Cost	Labor Cost	Total Cost	Material Cost O&P	Equipment Cost o&P	Labor Cost O&P	Total Cost O&P
Wall	42262.75	116528.25	20268.5	179059.5	52407.58832	144276.34	25277.82167	221961.75
Floor	8494.8	17802	5580.94	31877.74	9650.463307	20395.04301	6358.093684	36403.6
Ceiling	1765.8	3801	1071	6637.8	2021.340613	4351.067884	1225.991503	7598.4

As specified in the ISO 15686-5, LCC can cover acquisition through operation and disposal. In the current implementation, automated computation focuses on construction-phase capital costs, and the LCC results are restricted to generating structured inputs for maintenance and replacement planning using RSMeans life-cycle cost estimators based on the Canadian-specific datasets. Operational energy costs and performance-driven OPEX (e.g., energy, water, and carbon-related costs) are not calculated in the current workflow; integrating BIM-to-BEM pipelines and measured operational data is in progress and will be covered in future work.

Cost becomes a time-dependent attribute when it is linked to design revisions, procurement decisions, and construction progress or asset condition. The presented openBIM cost layer can act as a foundational dataset for a cost-focused digital twin by providing a traceable mapping between physical components and cost parameters, enabling systematic updates when quantities, rates, or asset states change.

Embedding cost parameters directly in IFC provides a self-contained, portable package that supports offline review, element-level traceability, and automated computation. However, it introduces governance and maintenance burdens. For instance, cost data can become outdated, licensed cost datasets may not be distributable inside models, and large models may not operate properly with detailed pricing information. Container-based approaches such as DIN 18290-3 (BIM cost container) and DIN 18290-2 (BIM-LV container) mitigate these issues by externalizing cost determinations and bills of quantities while preserving explicit links to IFC reference objects. For projects that require strict control over cost document versions, procurement workflows, or licensing constraints, a linked container may be preferable. For early-stage design, alternative analysis, and internal collaboration, in-model cost embedding can be effective when paired with clear update rules and access controls.

To scale across databases and contexts, the openBIM approach should rely on stable semantics for properties and exchange requirements. bSDD offered by buildingSMART can support standardized property definitions, and IDS can specify information requirements and validation rules for exchanges, reducing ambiguity in cost mappings across tools and stakeholders in various life cycle stages. In practice, bSDD and IDS can be used to formalize the Level of Information Need for model-based estimation by specifying which cost-related properties, material assignments, and quantity sets must be present at each project stage. In fact, it enables automated compliance checking of IFC submissions before running the estimation workflow.

Conclusions

Traditional cost estimation methods were often reliant on manual processes and licensed tools. This creates many inefficiencies and inaccuracies in the process. The integration of BIM has revolutionized this process by enabling digitalized, automated cost estimation. However, there are still challenges in the process due to the use of different BIM tools and data formats, leading to interoperability issues and data loss during data transformation between them. This study aimed to develop an OpenBIM-based approach to cost estimation. IFC facilitates seamless data exchange between different BIM authoring tools, ensuring consistency and accuracy across platforms. Furthermore, incorporating standardized classification systems (e.g. UniFormat) and reliable cost databases, such as RSMeans, enhances the precision and reliability of cost estimates.

Despite the advantages of the current study, it has some limitations. There is a lack of robust software to handle IFC for large projects. The Blender software used in the current study cannot open large IFC projects, so the authors are limited to medium-sized projects in the validation phase. On the other hand, this study focuses on only three categories of elements. To enhance the precision and accuracy of cost estimation, it is necessary to estimate the costs of all other project elements. A notable limitation of the current implementation is its reliance on explicitly modelled elements for quantity takeoff and costing. Thin finishing layers (e.g., paint, screeds), multi-component façade systems, lightweight partitions (e.g., timber or gypsumboard walls), and temporary works (e.g., scaffolding, cranes, site logistics elements) are often omitted or represented only implicitly in design-stage models, which means their costs are not directly captured by the automated workflow. In practice, practitioners could address these categories using: i)Parametric allowances (e.g., applying finish factors or percentage-based mark-ups tied to gross area or perimeter). ii) Assembly templates that aggregate typical finishing and sub-components into a single modelled proxy with an associated composite cost item. iii) Linked cost containers or external BoQs that include non-modelled items but remain traceable to the model context. In fact, Systematic integration of such non-modelled or partially modelled cost drivers especially temporary works and construction logistics remains an open research direction and is not fully supported by the current prototype. Furthermore, life-cycle cost data were sourced exclusively from the RSMeans. This reliance may limit the model's adaptability across different regional contexts or project types. Future study is in progress to independently estimate the life cycle cost of the project and consider all other factors in the life cycle cost process. Finally, While the present case study and implementation rely on Canadian-specific RSMeans datasets, the core workflow is not tied to a specific vendor or region. Because cost logic is encoded as generic IFC material properties (cost code, unit, formula, unit rates) and the linkage semantics can be governed via bSDD and IDS, the same pattern can be applied to other regional or international cost databases (e.g., national price books, company-specific cost libraries) by changing the external dataset and property definitions without modifying the underlying IFC computation engine. Future work will focus on validating the approach with alternative cost sources and regional

indices and on defining reusable IDS templates for cross-context deployment.

Acknowledgement

This study has been supported by the Construction Sector Digitalization and Productivity (CSDP) program under the project "OpenBIM integrated framework to standardize life cycle environmental and economical benchmarking of building projects" at National Research Council Canada. The authors would like to extend their sincere gratitude to all the stakeholders who participated in this research and contributed to the findings presented in this paper.

References

Abanda, F.H., Kamsu-Foguem, B. & Tah, J.H.M. (2017) BIM–new rules of measurement ontology for construction cost estimation. Engineering Science and Technology, an International Journal, 20, p.pp.443–459.

Banihashemi, S., Khalili, S., Sheikhkhoshkar, M. & Fazeli, A. (2022) Machine learning-integrated 5D BIM informatics: building materials costs data classification and prototype development. Innovative Infrastructure Solutions, 7, p.215.

buildingSMART (2024) buildingSMART Data Dictionary (bSDD). Available at: https://www.buildingsmart.org/users/services/buildingsmart-data-dictionary/ (accessed 2026-01-17).

buildingSMART (2024) Information Delivery Specification (IDS). Available at: https://www.buildingsmart.org/standards/bsi-standards/information-delivery-specification-ids/ (accessed 2026-01-17).

Cassandro, J., Mirarchi, C. & Pavan, A. (2024) IFC-based cost estimation: application to a structural model. Computing in Construction, 5, p.pp.868–875.

Deutsches Institut für Normung (DIN). (2023) DIN 18290-2:2023-03 (Draft): Linked BIM data exchange of building models with other specialist models—Part 2: Linked BIM data exchange of building models and specified bill of quantities (BIM-LV container). Berlin: DIN.

Deutsches Institut für Normung (DIN). (2023) DIN 18290-3:2023-11: Linked BIM data exchange of building information models with further specialist models—Part 3: Linked BIM data exchange of building information models and cost determinations (BIM cost container). Berlin: DIN.

Elbeltagi, E., Hosny, O., Dawood, M. & Elhakeem, A. (2014) BIM-based cost estimation/monitoring for building construction. International Journal of Engineering Research and Applications, 4, p.pp.56–66.

Fazeli, A., Banihashemi, S., Hajirasouli, A. & Mohandes, S.R. (2024) Automated 4D BIM development: the resource specification and optimization approach. Engineering, Construction and Architectural Management, 31, p.pp.1896–1922.

Fazeli, A., Jalaei, F., Khanzadi, M. & Banihashemi, S. (2022) BIM-integrated TOPSIS-fuzzy framework to optimize selection of sustainable building components. International Journal of Construction Management, 22, p.pp.1240–1259.

Fazeli, A., Dashti, M.S., Jalaei, F. & Khanzadi, M. (2021) An integrated BIM-based approach for cost estimation in construction projects. Engineering, Construction and Architectural Management, 28, p.pp.2828–2854.

Gholamzadehmir, M., Cassandro, J., Mirarchi, C. & Pavan, A. (2025) Advancing cost estimation through BIM development: focus on energy-related data associated with IFC elements. Applied Sciences, 15, p.7814.

Hosseini Gourabpasi, A., Jalaei, F. & Ghobadi, M. (2025) Developing an openBIM information delivery specifications framework for operational carbon impact assessment of building projects. Sustainability, 17, p.673.

IfcOpenShell (2025) IfcOpenShell: the open source IFC toolkit. Available at: https://ifcopenshell.org/ (accessed 2026-01-17).

International Organization for Standardization (ISO). (2024) ISO 16739-1: Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries—Part 1: Data schema. Geneva: ISO.

International Organization for Standardization (ISO). (2017) ISO 15686-5: Buildings and constructed assets—Service life planning—Part 5: Life-cycle costing. Geneva: ISO.

Khosakitchalert, C., Yabuki, N. & Fukuda, T. (2019) Improving the accuracy of BIM-based quantity takeoff for compound elements. Automation in Construction, 106, p.102891.

Khosakitchalert, C., Yabuki, N. & Fukuda, T. (2020) Automated modification of compound elements for accurate BIM-based quantity takeoff. Automation in Construction, 113, p.103142.

Khudhair, A., Li, H. & Ren, G. (2023) Knowledge-based openBIM data exchange for building design. Automation in Construction, 156, p.105144.

Kim, S., Chin, S. & Kwon, S. (2019) A discrepancy analysis of BIM-based quantity take-off for building interior components. Journal of Management in Engineering, 35, p.05019001.

Laakso, M. & Kiviniemi, A. (2012) The IFC standard: a review of history, development, and standardization. ITcon, 17.

Lee, S.-K., Kim, K.-R. & Yu, J.-H. (2014) BIM and ontology-based approach for building cost estimation. Automation in Construction, 41, p.pp.96–105.

Lee, G., Park, H.K. & Won, J. (2012) D3 City project—Economic impact of BIM-assisted design validation. Automation in Construction, 22, p.pp.577–586.

Ma, Z., Wei, Z. & Zhang, X. (2013) Semi-automatic and specification-compliant cost estimation for tendering of building projects based on IFC data of design model. Automation in Construction, 30, p.pp.126–135.

Naderi, M., Nazari, A., Shafaat, A. & Abrishami, S. (2025) Enhancing accuracy in construction overhead cost estimation: a novel integration of activity-based costing and building information modelling. Smart and Sustainable Built Environment, 14, p.pp.758–775.

Privé, N.C. & Errico, R.M. (2016) Observing system simulation experiments: an overview. Weather and Chaos.

Rad, M.A.H., Jalaei, F., Golpour, A., Varzande, S.S.H. & Guest, G. (2021) BIM-based approach to conduct life cycle cost analysis of resilient buildings at the conceptual stage. Automation in Construction, 123, p.103480.

Ramaji, I.J., Messner, J.I. & Mostavi, E. (2020) IFC-based BIM-to-BEM model transformation. Journal of Computing in Civil Engineering, 34, p.04020005.

RSMeans (2025) Using the life cycle cost estimator in RSMeans Data Online. Available at: https://www.rsmeans.com/resources/using-lifecycle-cost-estimator-rsmeans-online (accessed 2026-01-17).

Wahab, A. & Wang, J. (2022) Factors-driven comparison between BIM-based and traditional 2D quantity takeoff in construction cost estimation. Engineering, Construction and Architectural Management, 29, p.pp.702–715.

Yang, B., Zhang, B., Wu, J., Liu, B. & Wang, Z. (2019) A BIM-based quantity calculation framework for frame-shear wall structure. Structural Engineering International, 29, p.pp.282–291.

Zhang, S., Zhang, S., Wang, C., Zhu, G., Liu, H. & Wang, X. (2024) Extended IFC-based information exchange for construction management of roller-compacted concrete dam. Automation in Construction, 163, p.105427.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

CITYRETROFIT: DEVELOPMENT OF A DISTRICT-SCALE DIGITAL TWIN

J. J. McArthur[1], Saeed Ranjbar[2], Dana Al Jadaa[1], Masood Shamsaiee[2], Oriol Gavaldà-Torrellas[2], Hossein Bagherzadeh[1], Sanam Dabirian[2], Jackson Adebisi[1], Ursula Eicker[2]
[1]Toronto Metropolitan University, Toronto, Canada
[2]Concordia University, Montreal, Canada

Abstract

CityRetrofit integrates publicly available datasets with building archetypes to create urban building energy models (UBEMs). The tool integrates geospatial data, occupancy schedules, construction assemblies, and HVAC/energy system characteristics for both baseline and potential retrofit scenarios to support retrofit strategy development at the urban scale. This paper summarizes the development of this workflow in a mandatory building energy disclosure context (Montreal) and its extension to a jurisdiction without such disclosure (Toronto). The resultant case study demonstrates the workflow. The results show CityRetrofit's adaptability across jurisdictions to inform urban decarbonization policy.

Introduction

Global cities face mounting pressure to reduce CO_2 emissions but the sheer scale of retrofits is overwhelming, particularly given the paucity of building energy disclosure in many jurisdictions, existing building heterogeneity, and the lack of appropriate simulation tools.

Previous work has explored various approaches to urban building energy modeling (UBEM) , most notably single- and multi-zone bottom up simulation and top-down energy mapping. While the latter was enabled with mandatory building energy disclosure (MBED) laws, the ability to simulate building system performance was necessary to model and evaluate potential retrofit pathways. In this context, bottom-up models play a critical role.

Urban Building Energy Modeling (UBEM) tools have been developed to analyze city-scale energy flows, with a strong focus on building stocks and their interactions within the urban context, such as shading, thermal exchanges, urban heat island effects, and vegetation (Allegrini et al. 2015).Over time, a wide range of tools have emerged, using different modeling approaches—from reduced-order and resistance–capacitance models to detailed dynamic simulations—and varying in scope, input formats, and integrated features like transportation, life-cycle analysis, or cost–benefit assessment .

Despite these advances, the literature shows that existing Urban Building Energy Modeling (UBEM) tools have progressively improved the representation of urban energy flows—particularly building stocks and their interactions—using a wide range of modeling approaches and simulation engines (Monien et al. 2017). However, most tools remain strongly conditioned by the specific use cases for which they were originally developed, leading to limited modularity, scalability, data flexibility, cross-sector integration, and automation, as well as difficulties in clearly communicating results.

To address these gaps, recent research argues for moving away from tool-specific, ad hoc solutions toward an open-source, flexible framework built around a central urban data model capable of handling heterogeneous data sources and interfacing with multiple simulation tools at different levels of detail. This approach aims to support diverse urban sustainability and building retrofit use cases by integrating different processes and software within a unified, adaptable structure rather than relying on a single modeling engine, following (Gavaldà-Torrellas et al. 2025).

To populate such workflows, archetype development has been a topic of significant research globally. In the North American context, two sets of archetype libraries were available: 'reference models' for older (pre-2001) buildings, and 'prototypical' code-minimum models aligning with each version of the ASHRAE 90.1 energy efficiency standard, which was released codified on a three-year basis. This has been adapted regionally, for example the Canadian Building Technology Assessment Platform (BTAP) (CanMet Energy, 2023) that adapt these to local codes. The latter was a reasonable approximation for newer building stock, since this standard has been codified into regional codes across North America.

The former reference models (US DOE, 2012), however, categorize older buildings into two broad vintages: pre-1980 and 1980-2004, regardless of building type or form. This simplification fails to capture critical differences in construction practices, materials, and code requirements within each period. As a result, these archetypes often misrepresent the performance of older buildings, leading to inaccurate energy predictions (Li & Feng, 2025). The development and integration of more granular characterizations was thus critical to capture these differences.

Within Europe, there has been substantial archetype development, using both bottom-up and top-down aggregation methods; see (Mata, Kalagasidis, & Johnsson, 2014) for a comprehensive review. Across

multi-country studies, the use of code minimum performance has been shown to consistently estimate building stock performance with low (<5%) error (Mata et al., 2014).

This paper presents the development of the CityRetrofit tool for Montreal, its supporting workflows, and its extension to incorporate bottom-up approaches allowing its mapping to new jurisdictions with significantly limited public data. A case study for Toronto, a cold climate (ASHRAE Climate Zone 5 (CZ5)) dense urban context was presented to demonstrate and validate this process. The resultant tool supports its adaptation and replication, providing a scalable digital twin suitable for decarbonization policy and district energy planning across global urban contexts.

The paper was structured as-follows. First, the methodology guiding the development of CityRetrofit was presented, including geometric data processing, energy model input mapping, and the detailed development of refined inputs (to address the above oversimplification limitations) to provide more accurate baseline and retrofit pathway building-level models for UBEM integration. Next, the resultant inputs were summarized, and the platform integration was presented. Finally, the extension to other jurisdictions was discussed and the future research trajectory to address current limitations was presented.

Methodology

The CityRetrofit platform was developed through four overlapping phases: (1) workflow development; (2) ingestion of building geometry; (3) building archetype refinement; (4) retrofit design and analysis; and (5) integration and testing using a second urban area (Toronto) as a use case.

Workflow development

Existing UBEMs rely on use-case–specific architectures and back-end simulation engines, which constrains modularity and limits their computational scalability when extended to large building stocks, multiple cities, or cross-sectoral applications, as evidenced by their typically small-scale deployments in prior studies. In line with previous benchmarking research, the limited automation of data ingestion and preprocessing remains a key bottleneck, increasing human effort, reducing reproducibility, and hindering systematic comparison across tools and case studies. Moreover, the trade-off between reduced-order and detailed dynamic simulation approaches restricts their suitability for different decision contexts, ranging from rapid policy screening to high-resolution retrofit analysis. The proposed CityRetrofit platform addresses these gaps through an open-source, centralized data model that automates data handling, enables transparent benchmarking across simulation engines, and supports scalable, decision-relevant outputs tailored to policymakers, urban planners, and researchers. From a detailed perspective, the CityRetrofit platform integrates empirical data, archetype-based modeling, simulation, and comparative assessment to support scalable, portfolio-level retrofit decision-making through five main stages: (1) Building energy demand calculations, including geometric data acquisition and preprocessing, construction and usage archetype assignment, and weather file selection; (2) HVAC system assignment and energy disaggregation; (3) Retrofit identification and characterization; (4) Retrofit assessment; and (5) Visualization (Fig. 1). This modular structure enables scalability and adaptation to different building stocks and jurisdictions.

Building-level data availability determines the modeling approach adopted. For bottom-up analyses, geometric information is required and can be extracted from CityGML, 3DS, or GeoJSON files at varying levels of detail (LOD1/LOD2), all of which were processed through a unified geometry factory. When detailed geometry is unavailable, predefined geometrical archetypes enriched with building attributes were employed instead. These archetypes were developed by clustering extensive geometric and monthly energy datasets from data-rich cities, resulting in a set of transferable, high-resolution building representations integrated into the tool.

Building geometry ingestion

The first step for building modeling was the creation of a 3D city model. This model can be generated from multiple

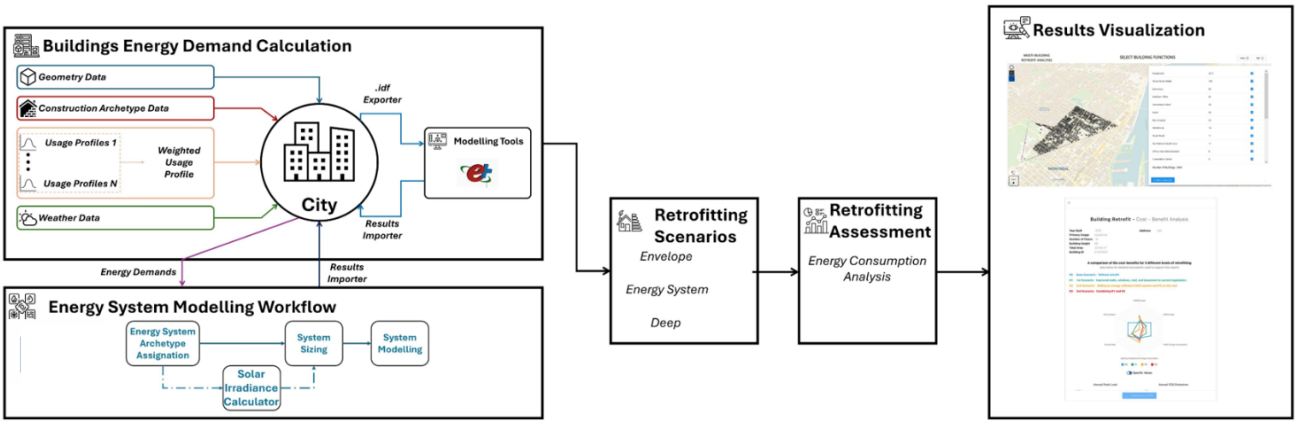


Figure 1: CityRetrofit workflow

formats, including GeoJSON, obj, 3ds and CityGML. Automated processes to validate and integrate different data sources was developed. For Québec cities, the input data was created by combining several open-source datasets including LiDAR aerial data (Montréal 2026a), property assessments (Montreal 2026a), Building Footprints (Canada 2026), and Layer 2 cadastral data for disaggregated building use types (Université Laval 2026). A validation and cleaning process, comparing the results with sources of ground truth (utility data for 2019-2021 divided by building function and Forward Sortation Area) was developed and helps validate the process of merging initial low-quality datasets, which was a usual situation in large scale city data. The comparison between simulated and measured energy use showed good agreement, with deviations generally within 15–20%, consistent with previous validation studies of the proposed UBEM workflow (Gavaldà-Torrellas, Monsalvete, Ranjbar, & Eicker, 2025).

Required building inputs include year of construction, building function, and height. The process begins by creating an instance of the *City* class, which was then populated with city object information extracted from the input file through a dedicated importer. This importer enriches the city model by integrating both geometric and attribute data. Once the city objects were extruded to form a 3D representation, the internal zoning of buildings was defined. At present, each building is modeled as a single thermal zone, although several multi-zoning approaches are currently under development. An important aspect of the approach taken was the capacity of ingesting different levels of detail of data (presently LOD1 and LOD2) depending on the quality of the original data sources.

Archetype development

Non-geometric building attributes were captured using archetypal constructions, systems, and scheduled, to map to the geometric archetypes. Existing reference and prototypical building models – e.g. DOE/ASHRAE and BTAP reference and prototype models – were used to provide the starting point for archetype development, providing reference geometry, construction assemblies, HVAC system definitions, and building schedules by period and type. A review of these archetypes against the existing building stock indicated a need for more granular construction and HVAC system characterization for the pre-2001 commercial (office) and multi-unit residential (MURB) building typologies while smaller buildings were more consistent with respect to both construction and HVAC systems. In all cases, occupancy schedules follow ASHRAE 90.1 as these were typical across the building stock.

Archetype development in a mandatory building energy disclosure context

Because Montreal had publicly available building energy use, that data was used to validate and refine inputs from the literature and allowed building clustering similar to what was achieved for the geometry.

Construction archetypes

Construction assemblies were re-evaluated through recalculation of U-values using explicit material properties and consistent assumptions regarding layer composition and air gaps. This process led to improved representation of post-2000 lightweight cladding systems, correction of legacy U-values where discrepancies were identified, and harmonized treatment of air gaps in accordance with building physics best practices. In parallel, the material library was expanded to include underrepresented but commonly used materials, such as stucco finishes and explicit air-gap layers, improving transparency and physical realism. Terminology was standardized across archetypes to ensure consistency with BTAP, NECB, and ASHRAE conventions, facilitating interoperability and collaboration.

To enable a robust and comparable assessment of building energy performance, this study applies a set of statistical diagnostic methods to both measured utility data and simulated baseline consumption. The primary analytical framework was the PRISM (PRInceton Scorekeeping Method), which characterizes the relationship between energy use and outdoor air temperature through segmented linear regression, yielding key indicators such as heating sensitivity (Fig 2).

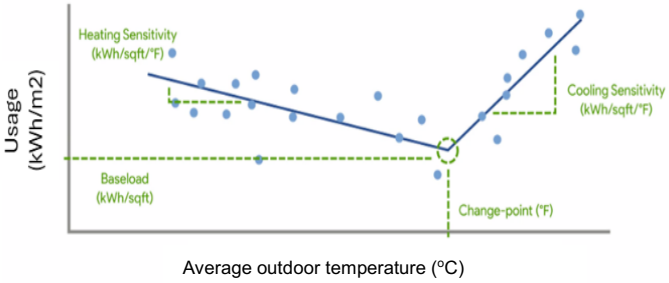


Figure 2: PRISM methodology

At the portfolio scale, kernel density estimation (KDE) was used to identify a reference centroid representing typical thermal performance within the building stock, allowing individual buildings to be benchmarked relative to representative baselines rather than absolute thresholds.

HVAC/Energy system archetypes and data model

In addition to construction and usage archetypes, each building must have the HVAC system assigned, which permits energy disaggregation. This process was done automatically where the assignation criteria can be defined based on: (1) Regional statistics; (2) Current building codes (for retrofit performance requirements); and (3) Scientific literature.

Process automation was enabled through the development of a generic object-oriented energy system data model where all the energy system-related data needed in simulation and optimization were structured in a parametric, generic, and extensible way. This provides a structured, object-oriented representation of building energy systems that explicitly separates data representation from simulation and optimization logic.

All system components – namely generation, storage, distribution, and emitters – were described through standardized classes with well-defined attributes, relationships, performance curves, and cost functions.

This separation enables the same data structures to be reused consistently across feasibility screening, hourly simulation, life-cycle costing, and multi-objective optimization, without rewriting model logic for each application. The object-oriented design supports modularity and extensibility: new technologies or manufacturer-specific components can be introduced by adding data entries or subclasses, without modifying the computational workflow. As a result, the data model serves as a robust computational backbone that enables scalable automation, cross-climate comparability, and relevant system analysis, while remaining compatible with semantic annotation or ontological layers if broader interoperability was required in future extensions. The appropriate energy system archetype was assigned. The system component parameters were retrieved from the energy system data and passed on to the simulation model. System retrofit analysis is implemented by changing the energy system archetype from that of the base case.

Extension to the low data availability context

Because of the lack of publicly available building energy data in Toronto, a more empirical approach was necessary to refine archetypes. Buildings types for CZ5 were characterized and clustered according to construction assemblies and HVAC systems to enrich existing building archetypes to reflect the full range of variations in form, use, and system configurations. This was done in three key phases: (1) historical analysis; (2) relevant code review; and (3) validation.

The construction historical analysis was based on an architectural study, which characterized the various forms, window-wall-ratios (WWRs), and constructions associated with each of the 20th century architectural movements. These movements were broadly used to define building styles and tend to be international in scope, providing a broadly-applicable method for building classification. Each architectural movement period was mapped to identify periods where multiple distinct styles were present within a single period. This approach allowed the archetypes to capture the diversity of common building types across typologies and identify the defining characteristics for each. The HVAC historical analysis similarly looked at the invention and period of adoption for various equipment and system types, grounded in industry documentation and previous historical studies (Carrier 2012, US DOE 2015).

While the historical analysis provided insight on typical equipment and construction types, these did not provide sufficient detail to model the performance of the HVAC /energy systems and construction assemblies. For this, a detailed code review was undertaken, where the various iterations of relevant building and energy codes were reviewed and prescribed performance was documented at each point. Once this was complete, periods with different code mandated-performance were used to subdivide the building sets and required RSI/U-values for construction and efficiency/COP for HVAC were documented for each. For HVAC systems, the resultant characterization sets were then cross-referenced with the typical systems and efficiencies identified in the reference models (US DOE 2012), building code requirements (NRCan 1997, NRCan 2011a) and guidance documents (NRCan 2011b).

To validate the characterizations, the Delphi technique, a procedural research method involving repeated rounds of input by experts until consensus was reached (Linstone 1985, Barrett and Heale 2020) was used. This approach was widely used to develop best practice guidance in the absence of available data (Barkhordari Ahmadi, et al. 2023) and is increasingly adopted in the energy modeling context (Phichetkunbodee, et al., 2023). Two leading building science firms, each with substantial (200+) energy model databases of past project in both pre- and post-retrofit conditions supported this process, referring to their libraries of energy models to confirm or clarify the proposed characterizations. The code minimum construction thermal performance and construction assumptions were documented in a set of tables circulated to both consultant teams at this stage. Because of the limited service lives of HVAC equipment, likely replacement periods for each building set and the replacement equipment – whether like-for-like or an anticipated new type (e.g. oil boilers replaced with natural gas boilers) were also included in the HVAC system review table.

Two rounds of review were used, along with an in-person meeting after consensus was reached to present the final archetype inputs and provide a venue to identify any further issues warranting consideration or inclusion in the model. Finally, cluster analysis was undertaken to identify constructions and HVAC/energy systems with similar characteristics. For construction assemblies, clustering used effective conductivity ($\mathrm{Cond}_{\mathrm{eff}}$; Eq. 1) vs effective solar heat gain coefficient ($SHGC'_{\mathrm{eff}}$; Eq. 2) and both Euclidean and Mahanobilis distances were used with 2 to 15 tested.

$$\mathrm{Cond}_{\mathrm{eff}} = \mathrm{U}_{\mathrm{wall}} * (1 - WWR) + \mathrm{U}_{\mathrm{window}} * WWR \quad (1)$$

$$SHGC_{\mathrm{eff}} = SHGC_{\mathrm{window}} * WWR \quad (2)$$

The resultant construction and HVAC system inputs by period were then translated into EnergyPlus input files.

Retrofit development and prioritization

Retrofit scenarios were generated for each building archetype through a three-stage process for each of construction and HVAC system retrofits. In the first stage, the research team reviewed each of the base building combinations and identified the full spectrum of possible retrofits based on hygrothermal, thermodynamic, or practical considerations. For example, for building construction upgrades, insulation strategies that risked moving the freeze point into a porous cladding such as brick were avoided. Similarly, local/terminal equipment requiring high temperatures limited fuel switching to those capable of providing the appropriate temperature

range. For practical reasons, retrofits requiring a complete replacement of all HVAC devices in a building or requiring the installation of significantly larger equipment were not considered as they were cost prohibitive.

In the second stage, the building consultants were again engaged through a Delphi technique where the researchers identified potential upgrades feasible for each a) construction and b) HVAC system type and created tables for each base condition to summarize these options. Notes regarding excluded retrofit options were also shared. These tables were circulated to the consultants for their review against their project databases and comments were received regarding any unconsidered technical challenges preventing listed retrofit measures, the frequency of such retrofits, and performance characteristics (i.e. revised U-value or COP). After another two rounds of review, consensus was reached, informing a range of retrofits representing different levels of ambition (Level 1 Most Typical (Short Payback) vs Level 2 Deep Retrofit) and for HVAC, retrofits enabled by the presence of a thermal energy network or ground-source heat pump system (Level 3). Again, an in-person meeting with all parties was held after retrofit finalization to present the results and provide an additional opportunity for clarification or refinement. The finalized retrofit inputs for both construction and HVAC systems were then characterized and translated into EnergyPlus input files. The final workflow is shown in Figure 3.

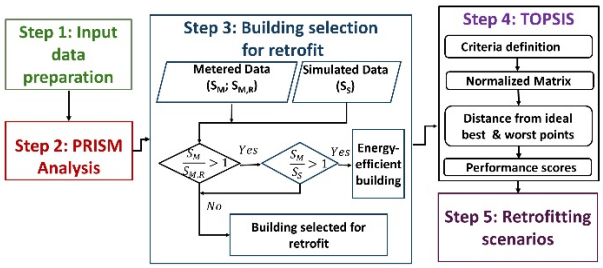


Figure 3. Building portfolio retrofit prioritization workflow

To further support transparent and data-driven retrofit prioritization, the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) was employed as a multi-criterion ranking method. Buildings were evaluated using a combination of measured and simulated indicators, including heating sensitivity slope, energy use intensity, total annual energy consumption, and volume-to-area ratio. After normalization, each building's distance to ideal and negative-ideal solutions was computed to derive a composite performance score. By integrating operational performance metrics with physical characteristics, this approach enables a balanced and justifiable ranking of retrofit candidates, ensuring that investment decisions reflect multiple dimensions of energy impact and align with broader energy and policy objectives.

Integration

Once the structural components of the model were defined, their composition was formalized through workflows, whose primary function was to manage the flow of data from external sources to one or more models, execute those models, and collect and distribute the resulting outputs. Regardless of the specific model being managed, each workflow follows a common sequence of fundamental steps: importing heterogeneous input data from proprietary formats, filling data gaps using archetypal data for construction, usage and energy systems and converting them into a shared central data model; optionally pre-processing these data through cleaning, calibration, pruning, or compression; optionally constructing or adapting model structures based on input characteristics such as building geometry; running the selected models one or multiple times using appropriate data exporters and model-specific input formats; retrieving and standardizing model outputs back into the central data model; and finally delivering the results through files, visual interfaces, or availability for subsequent workflows and external applications. This standardized workflow structure ensures interoperability, extensibility, and consistent data handling across diverse models, use cases, and urban-scale analyses. The re-running of the cases was crucial to obtain retrofit results once the model was validated.

District energy network optimization

To evaluate the impact of building retrofits on a district system, a 5th generation (heat recovery) thermal energy network (TEN) assessment/feasibility tool was developed in MATLAB to identify thermally balanced micro-TENs within urban blocks. The tool reads hourly building heating/cooling time series from building-specific EnergyPlus simulations results in Excel format and assembles a matrix in which each column represents one building's 8,760-hour profiles. After the building with the highest net heat rejection was selected as an anchor, the remaining buildings within the block were considered in n={1,2,3} sets to support adding building pairs with complementary net heating and cooling loads. Building profiles were aggregated along with the thermal storage state (charge/ discharge) to compute the outstanding mismatch, which then informs inclusion/exclusion decisions for the remaining candidate buildings. An objective function (Eq. 3) was used with forward selection to minimize the total energy imbalance, $Q_{net.}$

$$minimize\ \mathrm{Q_{net}}(t) = \Delta Q(t) + \mathrm{Q_s}(t) \qquad (3)$$

where: ΔQ was the instantaneous imbalance of connected buildings, and Q_s was the energy stored or used from storage. stored energy available. This method was applicable for both baseline and retrofit cases, along with future development scenarios.

Building energy simulations were performed using the EnergyPlus engine, with all archetypes translated into IDF input files and simulated at hourly resolution. Model calibration and validation were conducted using measured utility data where available. In the Montreal case, statistical calibration was performed using the PRISM method, enabling the derivation of heating sensitivity and base load parameters, which were compared against simulated outputs. At the portfolio level, kernel density

estimation (KDE) was used to define representative performance benchmarks. The resulting simulated energy use intensity (EUI) values were found to be consistent with measured datasets, with deviations generally within typical UBEM uncertainty ranges reported in the literature (<10–20%). In data-scarce contexts such as Toronto, archetypes were validated through expert-based Delphi processes informed by large databases of calibrated energy models (>400 cases).

Results

Construction characterization

Figure 4 shows the evolution of building codes (left) along with representative architectural styles (right) across periods. The overlapping of styles across multiple instances of the building code, along with differing residential vs non-residential code requirements led to an initial 23 archetypes, validated using Delphi technique.

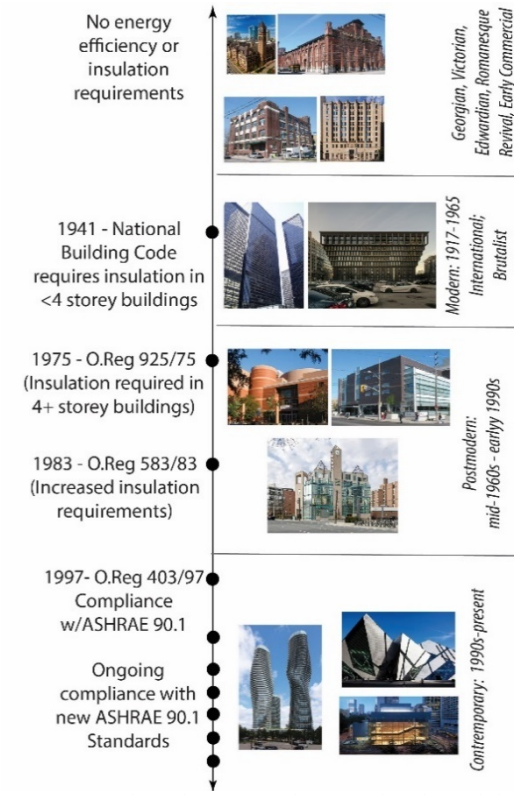


Figure 4: Timeline showing Architectural Styles Subdivided by Evolving Building Code Thermal Performance Requirements

Cluster analysis (Figure 6) was then used to identify buildings with highly similar thermal performance – both with respect to conductivity (area-weighted U-value/RSI) and radiation (area-weighted solar heat gain coefficient (set to 0 for opaque walls) – reducing the set to nine distinct construction characteristics.

The clustering exercise identified eight clusters and while the resultant areas and cluster identification orders were different, the cluster assignments themselves were identical for both approaches. The resultant construction assembly clusters are: {1975-2004 low-rise (brick), mid-high rise (concrete), and high-rise (steel) structures; 1975 -2004 Mid/High rise office (spandrel panel Postmodernist); 1975-2004 MURBS (all); 1941 1975 Mid/High rise office (curtain wall international); 1941-1975 mid/high rise offices (brick); All pre-1940 buildings + 1941-1975 low-rise MURBs; 1941-1975 low-rise office (brick) and mid-high rise MURBs (brick); 1941-1975 mid/high rise office & MURBs (concrete)}. These were uploaded into the CityRetrofit tool and mapped by address to each of the buildings based on year of construction and type. The input data for all clusters is published online at: https://github.com/DanaAlJadaa/Pre-2004-Canadian-CZ5-Envelope-focused-IDF-models.

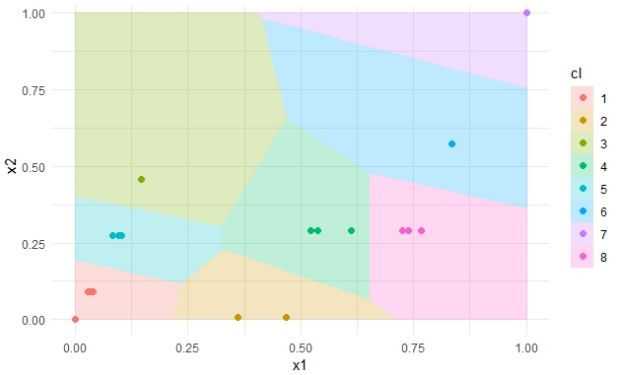


Figure 5: Construction Clustering using Euclidean Distances

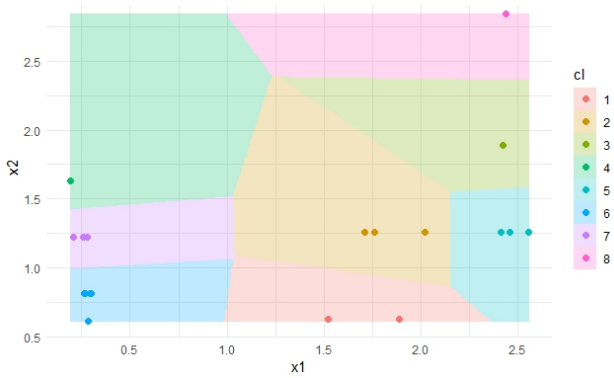


Figure 6: Construction Clustering using Mahanobilis Distances

HVAC system characterization

The HVAC system characterization identified nine representative HVAC systems across commercial offices and MURBs of various sizes, increasing the number of systems from DOE by two and providing more insight on shifts in fuel type (oil vs natural gas) and distribution (constant vs variable air volume) than the latter. These, with both baseline and the three levels of retrofit, were provided at https://github.com/jakjak42069/Baseline-E-Models-Canada-CZ5A. All other typologies were found to be consistent with DOE reference models, thus requiring no modification. Again, these were uploaded into the CityRetrofit library and mapped to buildings by type and period of construction.

District energy network optimization

The "Golden Mile" was an area within Toronto with significant planned densification (28,000 new residential

units to be added (Fig 7, blue buildings) was used as a demonstration case, consisting of 35 buildings including a large grocery store used as the anchor. Results were shown after each iteration in Fig. 8.



Figure 7: Demonstration Area

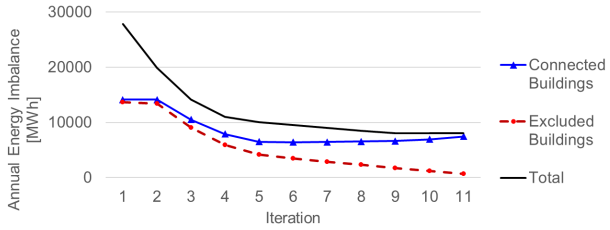


Figure 8: District interaction results

With this existing building set, 26 out of the total 35 buildings were identified for connection to the optimal TEN. This allowed 77% of the total heating and cooling loads to be met using heat recovery and seasonal storage, reducing the total heating and cooling energy required from by 77%, from 32,065 MWh/year to 7,449 MWh/year. The remaining load must be met from other sources, whether central (for the TEN) or at the building-level. Compared with individual building heat recovery using ground-source heat pumps with seasonal storage, the optimal TEN reduces net energy demand by 70%.

Platform integration and performance

In addition to the Golden Mile used for district interaction, the complete Toronto 2030 district (a 16km^2 area with over 30 million m^2 of built space across 7000+ buildings) has been integrated CityRetrofit tool (Fig. 10). The tool simulate up to 1,000 buildings in less than 1 hour using a single Intel® Xeon® E-2400 processor, with up to 8 cores at 95W, allowing the full district to be modeled in just over 7 hours, at which point the values were saved for future recall.

To use the tool, the building set of interest is identified, either through the GUI or by uploading a new set of buildings with archetype mappings as a .csv. Sample visualizations selected for implementation of the tool ae shown in Figure 10 alongside baseline energy consumption data.

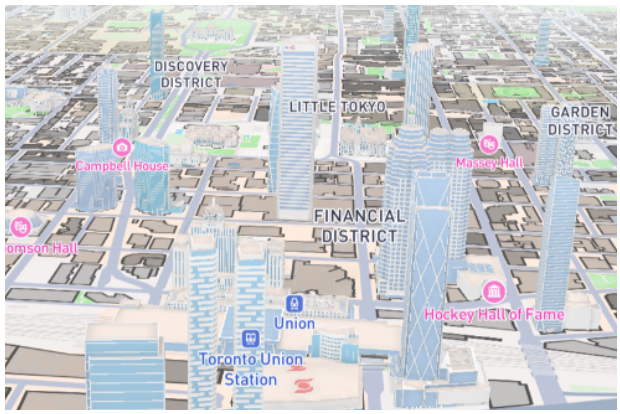


Figure 9: Toronto 2030 district (cropped)

Discussion and conclusions

The CityRetrofit platform presents a cost-effective way to create UBEMs suitable for retrofit pathway analysis at the portfolio, district, or city scale. Extending the tool to a new jurisdiction requires any new envelope constructions and HVAC/energy systems not already in the library to be added, which can be done by appending csv files.

Compared to existing UBEM platforms such as CitySim, UMI, or UrbanOPT, which are typically built around specific simulation engines or predefined workflows, CityRetrofit adopts a modular, data-centric architecture that decouples data models from simulation engines. This enables interoperability across tools, automated data ingestion from heterogeneous sources, and scalable application across jurisdictions with varying data availability. Additionally, the integration of empirical calibration methods (PRISM) and expert-informed

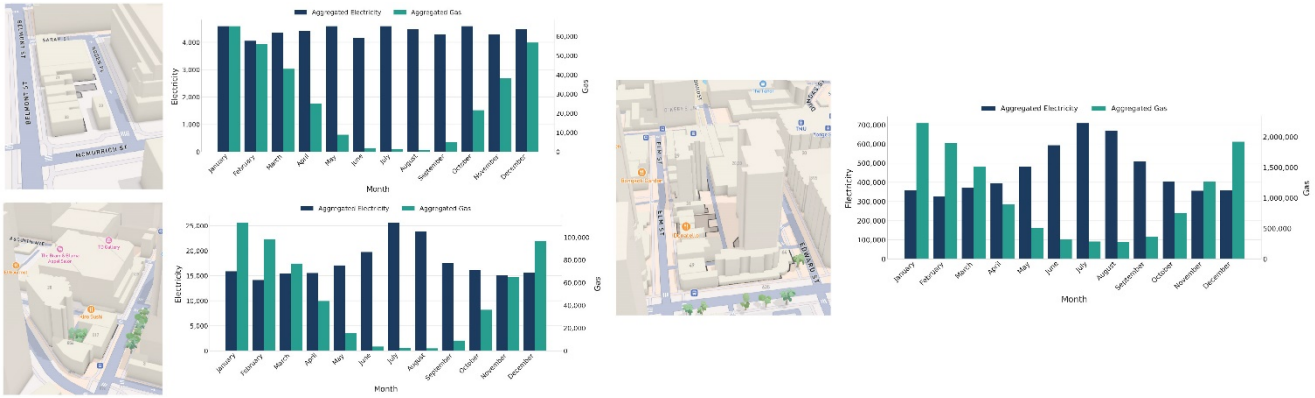


Figure 10: CityRetrofit sample visualizations (baseline condition)

archetype refinement provides improved adaptability in both data-rich and data-scarce contexts.

There were some limitations of this work. First, the lack of publicly available building energy data has made it impossible to validate the archetypes using utility so the Delphi study, informed by over 400 calibrated energy models, was used to confirm archetype inputs. Negotiations are ongoing to obtain access utility data for archetype calibration purposes and thus confirm the accuracy of these inputs. Second, the current modeling is single-zone only; multi-zone (by storey) is currently being tested for integration into the tool. Third, only forward-selection has been integrated for TEN optimization; future research will incorporate Backward Elimination and a binary Genetic Algorithm using tournament selection and uniform crossover, enforcing anchor buildings to remain included in all chromosomes. Finally, while the archetypes developed have been fully integrated into the back end of the platform and the district energy models are running, the full Toronto 3D model has not been loaded into a web-hosted platform and thus the Montreal GUI has not yet been implemented for the Toronto case. This is anticipated to be completed by August 2026.

Data Availability Statement

The full set of archetypal construction and HVAC system baseline and construction retrofit inputs were available in EnergyPlus format on GitHub as noted in the manuscript. The repository for the tool presented is available at https://ngci.encs.concordia.ca/gitea/CERC/hub.

Acknowledgments

This research was funded by Volt-Age (CFREF) through the CityRetrofit project.

References

Allegrini, J., Orehounig, K., Mavromatidis, G., Ruesch, F., Dorer, V., Evins, R. (2015). A Review of Modelling Approaches and Tools for the Simulation of District-Scale Energy Systems. Renewable and Sustainable Energy Reviews 52 (C), pp.1391-404.

Barkhordari Ahmadi, F., Esmaeeli, B., Shariat Moghani, S., Rahimi, R., & Shariati, K. (2023). A Literature Review of Group Decision-Making: The Case Study of Delphi Method. Medical Education Bulletin, 4(4), pp.824-833.

Barrett, D., Heale, R. (2020, May 19). What were Delphi studies? Evidence Based Nursing, 23(3).

Canada (2026) Automatically Extracted Buildings - Open Government Portal. Natural Resources Canada. https://open.canada.ca/data/en/dataset/7a5cda52-c7df-427f-9ced-26f19a8a64d6.

CanMet Energy. (2023). BTAP. https://github.com/canmet-energy/btap_batch

Carrier. (2012). Carrier. Available at https://www.carrier.com/commercial/en/us/about/history/

Montréal. 2026a. LiDAR aérien 2015 - Site web des données ouvertes de la Ville de Montréal. https://donnees.montreal.ca/fr/dataset/lidar-aerien-2015.

Montréal. 2026b. Unités d'évaluation foncière - Site web des données ouvertes de la Ville de Montréal. https://donnees.montreal.ca/dataset/unites-evaluation-fonciere.

Gavaldà-Torrellas, O., Monsalvete, P., Ranjbar, S., Eicker, U. (2025) The Urban Building Energy Retrofitting Tool: An Open-Source Framework to Help Foster Building Retrofitting Using a Life Cycle Costing Perspective - First Results for Montréal. Smart Cities 8 (1), p. 17.

Jayawardena, M., Heijndermans, E., Meiers, M. W., Watkins, R., Butsscher, J., Shenghui, F., & Berrah, N. (2022). Delphi Technique: How Expert Panels Predict Emerging Opportunities and Challenges in Renewable Energy. IEG Methods and Evaluation Capacity Development Working Paper.

Li, Y., & Feng, H. (2025). Pathways to urban net zero energy buildings in Canada: A comprehensive GIS-based framework using open data. Sustainable Cities and Society.

Linstone, H. A. (1985). The Delphi Technique. Environmental Impact Assessment, Technology Assessment, and Risk Analysis. Berlin: Springer.

Mata, É., Kalagasidis, A., & Johnsson, F. (2014). Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK. Building and Environment, 81, pp.270-282.

Monien, Dirk, Aneta Strzalka, Athanasios Koukofikis, Volker Coors, i Ursula Eicker. 2017. Comparison of Building Modelling Assumptions and Methods for Urban Scale Heat Demand Forecasting. Future Cities and Environment 3 (1): 2.

NRCan. (1997). Model National Energy Code of Canada for Buildings 1997. Ottawa: Government of Canada.

NRCan. (2011a). Model National Energy Code of Canada for Buildings 2011. Ottawa: Government of Canada.

NRCan. (2011b, October 25). Guide to Canada's Energy Efficiency Regulations. Ottawa: Government of Canada.

Université Laval. 2026. Geoindex - Public. https://geoapp.bibl.ulaval.ca/

US DOE. (2012). Commercial Reference Buildings. US Department of Energy.

US DOE. (2015, July 20). History of Air Conditioning. US Department of Energy.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

BREAKPOINTS IN DATA CONTINUITY ACROSS CONSTRUCTION REUSE PROJECTS

Clara P. Blum[1], Silvia Vangelova[1], Jens J. Hunhevicz[1,2], Kasimir Forth[1] and Catherine De Wolf[1]
[1]ETH Zurich, Zurich, Switzerland
[2]Empa, Dübendorf, Switzerland

Abstract

This study investigates where and why material identification and digital representation fail to persist across reuse life cycles. Digital material tracking in circular construction is commonly organized through separate data pipelines for logistics and design, resulting in discontinuities as building elements move between projects. A unified analytical framework is proposed and used to define six recurring classes of breakpoints in data continuity. These breakpoints are empirically identified and analyzed across eight construction reuse case studies. The findings establish a systematic breakpoint typology and provide recommendations for continuity-oriented data infrastructures that support traceable material reuse in construction.

Introduction

Material reuse is increasingly promoted as a key strategy for advancing circular construction (De Wolf et al., 2024), yet its practical implementation remains limited. While reclaimed materials are often physically available, their integration into new projects is constrained by insufficient or unreliable data (Wijewickrama et al., 2021). In many reuse scenarios, design decisions are hindered not by material scarcity but by uncertainty regarding material identity, condition, and provenance.

Across multiple reuse projects, data continuity was repeatedly observed to break as building elements move through typical reuse stages, including deconstruction, transport, planning and material storage, and construction. These process stages are illustrated in Figure 1 through the example of the Huber Pavilions, former student workshop pavilions at ETH Zurich, which were disassembled in 2022 and served as a source project for several recipient projects on and off the university campus (Knecht, 2022; De Wolf, 2025).

Each construction project establishes its own physical and digital infrastructures, including modeling environments, documentation practices, physical identification and labeling methods on building elements (Byers et al., 2024c). These infrastructures are typically optimized for a single project and its immediate life cycle (Whyte and Hartmann, 2017; Papadonikolaki et al., 2020). When elements are transferred between projects, their associated physical and digital systems are rarely maintained or preserved. Each of these infrastructural components, shown in Figure 2, namely the physical storage, physical element, physical data carrier, user interface, ID, digital model, and digital storage, exist at each process step depicted in Figure 1 but may change or be lost between these steps and along the reuse process.

Based on these recurring observations, this study argues that material data discontinuities in reuse projects emerge in part because information is handled through two established yet poorly connected data pipelines: supply chain logistics and design. Logistics-oriented approaches in practice commonly focus on tracking, aggregation, and inventory management and rely on standardized identification and data exchange systems, such as GS1, which define globally unique identifiers, hierarchical packaging structures, and event-based tracking across supply chains (GS1, 2025). In research, emerging track-and-trace frameworks extend these logistical principles to construction materials, and even reuse (Byers et al., 2024a). Design-oriented workflows generally rely on digital representations as the primary means of organizing and managing project information. In contemporary construction practice, this role is most commonly fulfilled by Building Information Modeling (BIM) environments, which use internally defined element identifiers, object hierarchies, and project-specific data structures to support design, coordination, and fabrication (ISO, 2018).

As illustrated in Figure 2, both logistics- and design-oriented pipelines address identification and representation, but operate according to different data structures: logistics emphasize persistent identifiers across inventories and transactions, while design environments prioritize context-dependent model elements and project-bound semantics. When these two structures are not aligned, material identities are frequently reduced, duplicated, or detached from physical elements altogether.

Despite growing interest in digital integration and reuse, there remains limited empirical documentation of how and where these logistic and design data pipelines break down in practice as materials move between projects. Therefore, this study investigates where and why material data continuity breaks down between logistics and design in reuse-oriented construction life cycles.

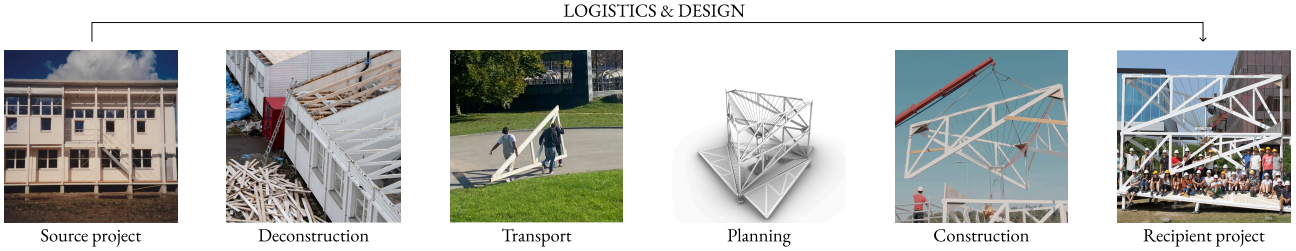


Figure 1: Process stages of construction reuse, illustrated through the Huber Pavilions'deconstruction and material reuse. Photo and model credit is provided in the Acknowledgments section.

Research design

Through direct involvement in multiple reuse projects, recurring overlap and misalignment between logistic and design-related data flows were observed, particularly for identification and digital representations of elements. Based on these observations, a combined analytical framework was formulated that integrates logistics and design perspectives across digital infrastructures corresponding to physical components, namely data carriers, user interfaces, IDs, digital models, and digital storage (Figure 2). Upon establishing this framework, potential breakpoints in persistence of data across reuse cycles were identified.

In line with the framework, a set of past reuse projects were analyzed to determine where and why these breakpoints occur. The analysis focused on how material data is structured, transferred, and transformed, with emphasis on its implications for material tracking and tracing. The analysis follows a comparative multiple-case study design (Yin, 2018) to identify recurring patterns and systemic gaps and to synthesize identified breakpoints in current digital infrastructures.

Data collection combines document analysis, including digital models, databases, and material documentation, with process reconstruction of data handovers between life cycle stages. Where available, expert insights are used to contextualize observed data practices.

The goal is to improve understanding of how existing digital systems support, constrain, or disrupt persistence of data in reuse-oriented construction processes.

Proposed framework for data continuity

The integration of infrastructures present in both logistics and design results in the combined infrastructure shown in Figure 2. Rather than treating logistics and design as separate data domains, material identity should be established, accessed, and maintained across both perspectives. The **physical element** is commonly considered only on the logistical side, being transported and stored in a **physical storage** or installed within a building, whilst in design it is abstracted as a geometric and semantic entity and exists only in the form of the digital model. The framework requires the physical element to remain linked to all digital representations, which is facilitated through a physical data carrier, the user interface, and an identifier (ID). **Physical data carriers** (e.g. direct product marking (DPM), Quick Response (QR) codes, barcodes, and near-field communication (NFC)) provide the primary interface between the physical element and digital systems. By scanning the physical data carriers, **user interfaces** mediate access to element data in logistics. In design, the user interface functions as an access point to the digital model. While these two forms of user interfaces are often implemented separately, the framework highlights their shared function: resolving an identifier into actionable information. The **ID** acts as the key link across all layers. In logistical systems, IDs are often assigned based on type or batch, whilst in design systems, IDs are typically instance-based and embedded in digital models. The framework positions the ID as a shared anchor that must persist across systems and representations to enable continuity, rather than being redefined at each stage. **Digital models** primarily emerge from the design perspective, where elements are represented geometrically and semantically. In the combined framework, the digital model is linked via the shared IDs, which allows material data originating in logistics to be associated with specific modeled elements. The digital infrastructures require **digital storage**. While these are often separated in practice, the framework considers linked or interoperable storage structures to prevent duplication and manual synchronization. Digital storage is thus understood as a shared backend that supports both inventory tracking and design integration. In line with the functionality of the identifiers, this considers both the storage of the digital model and the IDs and how these correspond.

Identified breakpoints

By aligning the combined stack in a single framework, overlaps, gaps and misalignment become apparent. Beyond illustrating the system as a whole, six breakpoints (A–F) were identified and are shown in Figure 2.

Breakpoint A addresses the interface between the physical construction element and the physical data carrier, which is commonly attached, engraved, or imprinted. Since construction and reuse processes typically involve transformations of the element, such as cutting it to the dimensions required for a subsequent project, the connection between the data carrier and the physical element may be disrupted. This can lead to loss of access to the element's identity and the associated stored data.

Breakpoint B highlights discrepancies that arise when

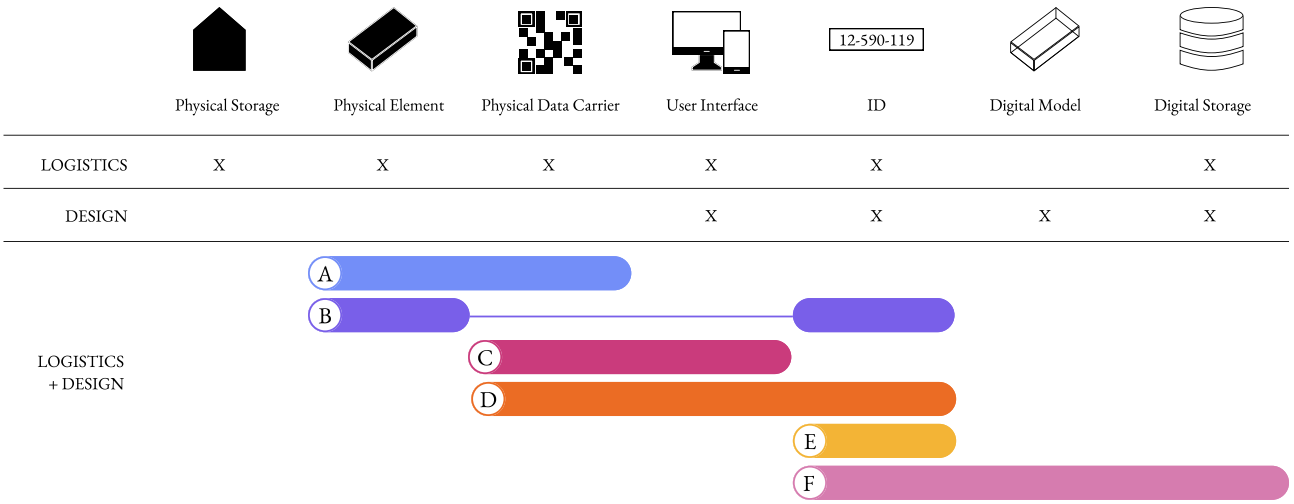


Figure 2: Infrastructure integrating logistics and design and tabular presence of digital infrastructures for identification and representation of construction elements separated between logistics and design. Breakpoints in the separated data pipeline are highlighted and in order: Physical–Digital Interface Loss (A), ID Granularity Mismatch (B), Data Carrier Degradation (C), Static Carrier Limitation (D), Cross-System ID Inconsistency (E), Unlinked Data Repositories (F)(Bottom)

groups of elements are recorded not by instance-specific IDs but by generalized identifiers, such as type-based classifications.

Breakpoint C refers to the physical data carrier and its accessibility via the user interface. In particular, the readability of the data carrier is influenced by its durability, which directly affects long-term data access.

Breakpoint D points to the link between the physical data carrier, the user interface, and the ID. It specifically addresses whether the data carrier is static or dynamic, and thus its capacity for adaptability over time.

Breakpoint E emphasizes inconsistencies in the definition and use of IDs across logistics and design systems, where the ID assigned to a physical element in logistics may not correspond to the ID used in the digital design model.

Breakpoint F addresses data storage, particularly the separation of ID storage and model storage. When these are not linked, duplicate information arises, requiring manual updates on both sides whenever modifications occur.

Breakpoints across case studies

Based on the identified breakpoints, eight case studies conducted at the Chair of Circular Engineering for Architecture at ETH Zurich between 2020 and 2025 were analyzed. Figure 3 synthesizes these case studies by illustrating the flow of digital infrastructures across reuse processes. Source projects are labeled with Roman numerals and recipient projects with Arabic numerals, with the project legend shown in the top-right corner. Instances of identified breakpoints are annotated by breakpoint class (A–F) and project number (1–8). Where multiple materials were used within a single project (e.g., case study 8), an additional index is appended. Figure 3 also includes keywords indicating the specific infrastructural components involved (e.g., QR codes or DPM for physical data carriers and Rhino for digital models). Detailed descriptions are provided in the following text, which should be read together with the figure for full interpretation. Breakpoint instances are bolded in the text to facilitate efficient cross-referencing between the descriptive analysis and the synthesized overview in Figure 3.

Timber from an automobile warehouse in Geneva, Switzerland (I), was used for the construction of Dome 1 (1), a geodesic dome that was later disassembled and reconstructed as Dome 1.2 (2) and Dome 1.3. For the construction of Dome 1, QR codes were engraved on the timber elements (Byers et al., 2022; Byers and De Wolf, 2023), and assembly logic was recorded directly on the elements using DPM. During the reconstruction of Dome 1 as Dome 1.2, the DPM information was no longer present (**A.2**), while the engraved QR codes remained readable. During reconstruction, elements were temporarily misplaced due to out-of-sequence staging and inaccessible labels. These elements were subsequently identified using the remaining QR codes (Byers and De Wolf, 2023). In case studies 1 and 2, all elements originated from the same source and were reconstructed using identical procedures. Within this context, no functional difference was observed between instance-specific IDs and categorical IDs (**E.2**). QR code content was limited to material properties and nomenclature, while higher-level project information such as assembly and disassembly instructions was not accessible through the system (**F.2**). Following the construction of Dome 1.2, the online material passport was no longer actively maintained (**D.2**).

Dome 2 (3) was built with a similar design as Dome 1, and Dome 1.2, with timber reclaimed from the deconstruction of a music pavilion in Oetwil am See, Switzerland (II). In Dome 2, the technical database did not include mechanisms for privacy or controlled accessibility, allowing unrestricted editing and deletion of material data (**D.3**) (Byers and De Wolf, 2023).

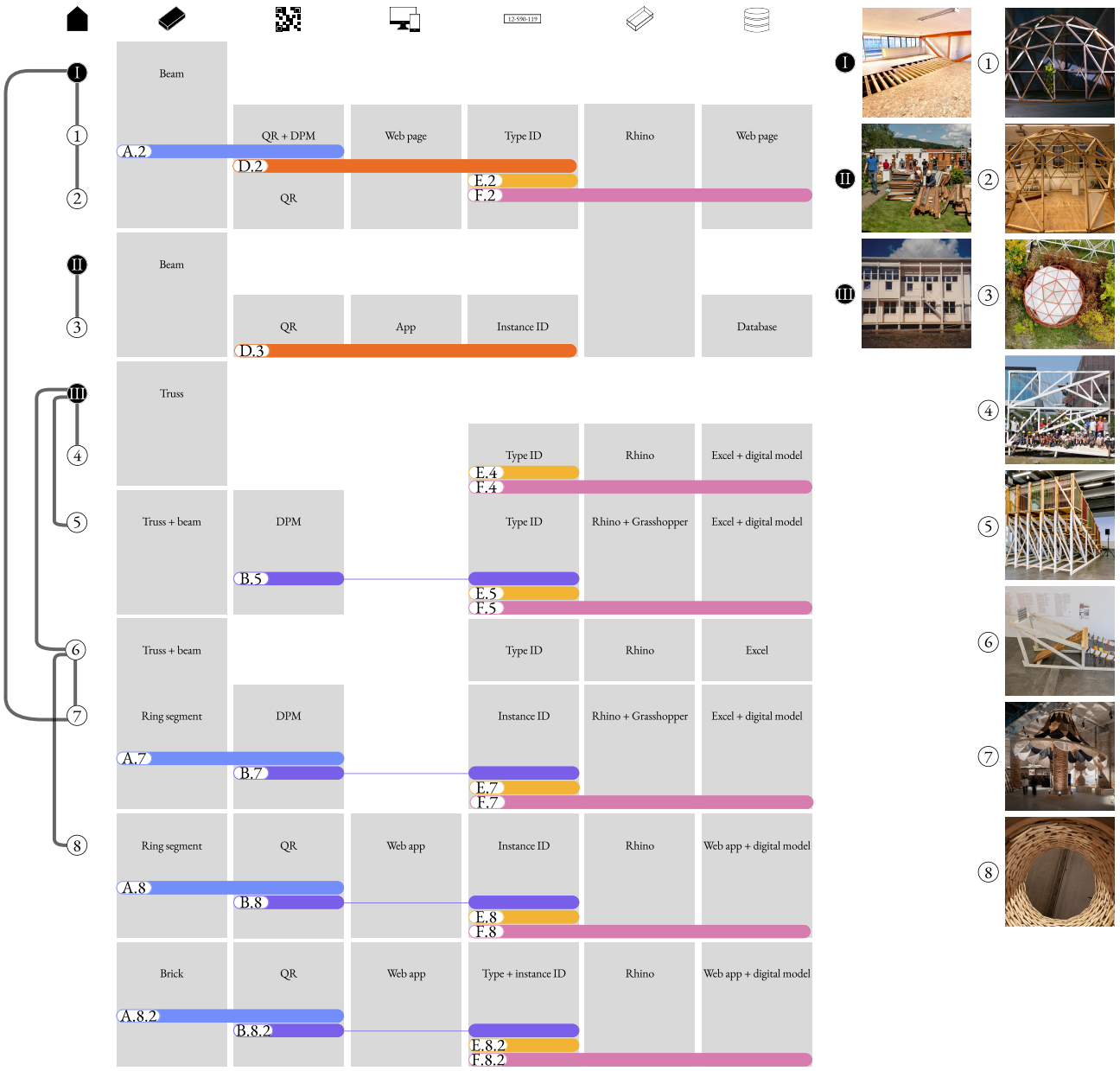


Figure 3: Flow of digital infrastructures across reuse case studies. Breakpoint instances are annotated by class (A–F) and project number (1–8), with additional indices indicating multiple material instances per project. Source projects are labeled with Roman numerals and recipient projects with Arabic numerals (see project legend on the right) in order: automobile warehouse in Geneva, Switzerland (I); music pavilion in Oetwil am See, Switzerland (II); Huber Pavilions (III); Dome 1 (1); Dome 1.2 (2); Dome 2 (3); CircÛbi (4); Mehrwerk (5); playground structures (6); VAMO (7); Insieme (8). Photo and model credit is provided in the Acknowledgments section.

Subsequently, the Huber Pavilions (III) were deconstructed, providing material for further projects, including CircÛbi (4), Mehrwerk (5), and a set of playground structures (6). In case study 4, IDs were not documented for the wooden trusses used, resulting in missing element identification for subsequent reuse (**E.4**). In addition, discrepancies between Excel-based documentation and the digital model resulted in unlinked ID storage locations (**F.4**).

Case study 5 showed the same patterns of missing or inconsistent element IDs (**E.5**) and unlinked data storage between documentation and digital models (**F.5**). Beam IDs were documented only by type, quantity, and dimensions without instance-specific identifiers (**B.5**). Only type-based DPMs were present on the physical elements, while beam condition varied.

Following the deconstruction of the playground structures (6), part of the reclaimed material was reused in two additional projects: VAMO (7) and Insieme (8). Insieme also included reclaimed bricks from several sources. Case study 7 sourced material from both source project I and case study 5. Wooden beams from case study 6 were transformed into cold-bent laminated ring segments. The original DPM markings were not transferred or documented for the new elements, resulting in the loss of the phys-

ical element–data carrier connection (**A.7**). Floorboards sourced from project I were used in a similar transformation process but did not carry documentation of their prior life cycle. Because they had no DPM, they are disregarded here. In this context, beam IDs were documented only by type, quantity, and dimensions and not by instance (**B.7**). Instance-specific IDs were also not assigned during recycling (**E.7**), and beams were treated as generically available material, while projects 7 and 8 ran in parallel using the same inventory. Beam IDs were documented in an Excel file by type but were neither referenced in the digital model nor assigned to model elements intended for the ring segments (**F.7**).

In case study 8, similar transformation-related losses occurred: beams were cut and cold-bent laminated without transferring or documenting DPM information for the new elements and their QR codes (**A.8**). Prior to this project, beam IDs had been documented only by type, quantity, and dimensions (**B.8**). Instance-specific IDs were again disregarded during recycling (**E.8**). A resulting cold-bent lintel was assigned an instance ID stored as a JSON file in a web application; this ID was not linked to the corresponding element in the digital model (**F.8**). Case study 8 also incorporated reclaimed bricks from multiple sources. The data and identification infrastructure for the bricks followed the same structure as that for the wooden elements, corresponding to observations (**A.8.2**), (**B.8.2**), (**E.8.2**), and (**F.8.2**). In contrast to the timber elements, some bricks were assigned instance-specific IDs due to documented damage and repair histories.

Results

Physical–digital interface loss (Breakpoint A)

Repeated loss of physical data carriers (**A.2, A.7, A.8, A.8.2**), namely DPM, during disassembly and element transformation indicates that data carrier strategies tied to original geometries are vulnerable in fabrication-intensive reuse workflows. To maintain the connection between physical elements and their data, and thus their identity, data carriers must persist through material transformation. Other approaches may be to use different data carriers during certain process stages to ensure the presence of data carriers (Dervishaj et al., 2023) or to enforce protocols for reattaching data carriers and recording provenance and transformation events in respective models (Guillem et al., 2025). Such protocols should include unique element identifiers, timestamps and actors involved in transformation, descriptions of geometric and material changes, and links to updated digital representations (e.g., BIM elements or graph nodes), enabling traceability across transformations within a shared data infrastructure.

ID granularity mismatch (Breakpoint B)

The reliance on type-based IDs (**B.5, B.7, B.8, B.8.2**) obscures instance-specific differences between elements, which become relevant when material condition or history varies within the same type. As each project is different and neither the future use of materials nor the respective requirements of data granularity can be predicted, one approach might be to use a multi-level ID schema. This would mean defining both a type ID and an instance ID with structured metadata, enabling differentiation by condition and role through semantic data models and otologies (Kebede, 2024; Forth et al., 2025). The brick reuse in case study 8 demonstrates that combining type-based and instance-specific IDs can mitigate several of the challenges observed in breakpoints **B**, **E**, and **F**, when element-specific histories are relevant. Nevertheless, the added value of maintaining instance-specific identifiers depends on their relevance for the recipient project: while tracking large numbers of interchangeable elements of the same type (e.g., bricks) may provide limited benefit, instance-level identification becomes essential when element-specific properties or regulatory requirements, such as fire safety labels on doors, must be individually verified. Complementary work builds on these observations by proposing instance-level tracing methods that address limitations of type-based identification in reuse scenarios (Vangelova et al., 2026). In practice, such multi-level ID systems could be implemented using graph-based data structures, where type nodes define shared properties and instance nodes capture condition, provenance, and transformation history, enabling flexible queries depending on project requirements.

Data carrier degradation (Breakpoint C)

Breakpoint C was not directly represented and observed in the analyzed used cases. Beyond the analyzed cases, Byers and De Wolf (2023) highlight that insufficient durability and readability of physical data carriers can undermine long-term data accessibility. Future implementations may therefore require redundancy strategies, to ensure continued accessibility even if individual carriers degrade.

Static carrier limitation (Breakpoint D)

The absence of long-term maintenance and access-control mechanisms highlights governance-related challenges for material passport systems when reuse cycles extend beyond a single project (**D.2, D.3**). Versioned records and stable persistent links (e.g., URLs encoded in carriers) that update as element information evolves, supported by access control in the backend, may respond to this challenge (Hunhevicz, 2023; Byers and De Wolf, 2023; Byers et al., 2024b). Such mechanisms align with emerging data space approaches, where persistent identifiers are resolved through federated services and access rights are managed across stakeholders, ensuring that data remains accessible and trustworthy beyond individual project life cycles.

Cross-system ID inconsistency (Breakpoint E)

While instance-specific IDs did not provide measurable benefits in homogeneous reuse scenarios (case studies 1 and 2), later case studies suggest their relevance increases in multi-source and parallel reuse contexts (**E.4, E.5, E.7, E.8, E.8.2**). Unifying the ID logic across logistics and

design databases encourages the use of BIM as a shared digital representation to link external identifiers to model data. Digital product passport research also emphasizes interoperable identification across supply chain and design systems (Zhang and Seuring, 2024; Vangelova et al., 2025) and the need for consistent identifiers in BIM exchange. This could be supported through semantic or knowledge graph models, where persistent identifiers act as nodes linking BIM elements, logistics records, and external data sources, enabling consistent referencing across systems and lifecycle stages.

Unlinked data repositories (Breakpoint F)

The recurring separation between ID storage (e.g., Excel, Rhino model, web apps) and digital design models (e.g., BIM and Rhino) reveals a structural limitation in current infrastructures, where material allocation and tracking remain manual (**F.4, F.5, F.7, F.8, F.8.2**). Integration of ID storage and digital model syncing could employ cloud-based data spaces that reduce redundancy and support linked data for the design phase (Bucher, 2019; van Zandwijk et al., 2025a). Beyond technical fragmentation, the separation of ID storage and model storage also exposes limitations in data longevity and governance, as material information must persist beyond individual software tools and project-specific repositories, which may motivate future exploration of distributed and decentralized data infrastructures (Bucher et al., 2025). Building on distributed data storage, data spaces could automate currently manual synchronization processes by enabling real-time data exchange between stakeholders, reducing coordination effort and improving data consistency across tools (van Zandwijk et al., 2025b; Hagedorn et al., 2025).

Discussion

This study proposes that integrating logistical and design data pipelines has the potential to improve data continuity across stakeholders in reuse-oriented construction. The contribution is threefold. First, a framework is defined that integrates logistics and design perspectives within a shared data infrastructure. Second, breakpoints within this framework, namely where data continuity is disrupted, are identified and defined. Third, instances of these breakpoints are examined and analyzed across multiple case studies. Based on these findings, recommendations for addressing material data–related challenges in reuse projects are derived.

The results show that material identities are frequently detached from physical elements during transformation, reduced to type-based classifications, or stored in isolated systems that are not linked to design models. As a result, logistical inventories and design representations diverge, forcing manual coordination and limiting digitally integrated workflows. Data resolution typically drops after initial construction or first reuse cycles, when material passports and identifiers are no longer maintained or integrated into subsequent planning phases, which aligns with findings by Whyte and Hartmann (2017) and Papadonikolaki et al. (2020). This also indicates that one primary barrier to linking logistics and design lies in the absence of consistent cross-system IDs. Without robust interfaces between physical elements and digital identities across life cycle phases, material data cannot reliably support reuse-oriented design decisions. The results of this study help address these structural discontinuities by making data breakpoints between logistics and design explicit, thereby providing a basis for developing continuity-oriented infrastructures that enable traceable and effective material reuse in circular construction.

Research on digital product passports and material passports in both product and construction sustainability states persistent, unique identifiers as a prerequisite for lifecycle traceability across design, use, deconstruction, and reuse (Byers et al., 2024c, 2025). To extend traceability beyond static documentation, semantic and ontology-based data structures are increasingly proposed to represent enriched metadata, provenance, and relationships resulting from material transformation (Vangelova et al., 2026). At the same time, interoperability research in BIM emphasizes aligning identifier-based data structures with cross-system standards to enable consistent integration across logistics databases, digital design models, and fabrication workflows (Forth et al., 2025). Despite these advances, material identity is still frequently detached from physical elements, reassigned new identifiers, or reduced to coarse classifications, disregarding materiality and product decomposition completely. This raises the question of whether increased data granularity and detailed representation in design and construction actually improve data continuity into subsequent reuse phases. In this context, emerging construction data space initiatives and decentralized data infrastructures provide a promising direction, addressing persistent identification, provenance tracking, and cross-organizational data exchange at scale, complementing the breakpoints presented in this study.

Limitations and outlook

This study is based on a limited number of case studies, most of which were analyzed retrospectively and conducted at the same institution, the Chair of Circular Engineering for Architecture at ETH Zurich. All case studies are short-term, student-led, pavilion-scale projects, intentionally designed as exploratory testbeds to assess data carrier feasibility, with limited incentives for maintaining track-and-trace beyond the course duration. While this allows for in-depth analysis of reuse processes, it limits the generalizability of the findings. In addition, data collection relied partly on informal expert insights in cases where partial data was unavailable. Although these insights provide critical contextual understanding, they introduce a degree of subjectivity that cannot be fully eliminated.

The study does not propose or validate a specific technical implementation, data standard, or software system. Instead, it adopts an analytical and diagnostic perspective to

identify recurring infrastructural challenges across reuse scenarios. As such, the findings do not constitute a prescriptive solution but rather highlight structural conditions that future systems must address. Validation of the identified breakpoints will require applied system development and longitudinal testing in further reuse projects. To support analytical rigor despite these limitations, all case studies were examined using the same framework and evaluated against the same set of breakpoints (A–F). Multiple case studies allowed isolated incidents to be distinguished from recurring patterns, strengthening internal validity. While the results remain context-dependent, the systematic application of the framework and transparent mapping of observations to breakpoints provides a robust basis for further empirical and technical investigation. Future research should focus on translating the identified breakpoints into implementable system architectures, for example by combining persistent identifier systems, semantic graph models, and construction data spaces to enable FAIR (Findable, Accessible, Interoperable, Reusable) (Wilkinson et al., 2016) material data across reuse cycles. In addition, questions of long-term governance and incentive alignment remain critical: responsibilities for maintaining data quality, storage, and access must be shared across stakeholders, potentially supported by decentralized governance infrastructures (Hunhevicz et al., 2024) that reduce individual burden while preserving collective value.

Acknowledgments

This research has been funded by the Flagship projects ThinkEarth (107.163 FS-ENG) and SWIRCULAR (107.235 FS-EE) supported by Innosuisse. The authors would like to thank Vanessa Costalonga, Vanessa Schwarzkopf, Brandon Byers, and Elias Knecht for providing required data and additional information in the form of informal interviews. In addition, the authors would like to thank Dr. Roxanne Goldberg for her invaluable contributions to this work in editing and content refinement.

Credits in Figure 1: (Source project) Photo by gta Archiv / ETH Zurich (Estate of Benedikt Huber); (Deconstruction) Photo by Elias Knecht, Master student, ETH Zurich; (Transport) Photo by Elias Knecht, Master student, ETH Zurich; (Planning) 3D Model by Elias Knecht, CEA, ETH Zurich; (Construction) Photo by Ivy Filmstudio, CEA, ETH Zurich; (Recipient project) Photo by Elias Knecht, CEA, ETH Zurich.

Credits in Figure 3: (I) Photo by Catherine De Wolf, CEA, ETH Zurich; (II) Photo by Anna Buser, CEA, ETH Zurich; (III) Photo by gta Archiv / ETH Zurich (Estate of Benedikt Huber); (1) Photo by Alexandre Attias, CEA, ETH Zurich; (2) Photo by Catherine De Wolf, CEA, ETH Zurich; (3) Photo by Artan Hajrullahu, CEA, ETH Zurich; (4) Photo by Catherine De Wolf, CEA, ETH Zurich; (5) Photo by Anna Buser, CEA, ETH Zurich; (6) Photo by Anna Buser, CEA, ETH Zurich; (7) Photo by Lloyd S. Lee; (8) Photo by Lloyd S. Lee;

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 5.2 in order to increase clarity. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Bucher, D. (2019). Development of a cloud-based Common Data Environment for purposes of a construction company. PhD thesis.

Bucher, D. F., Lab, J. J. H. D. U. E. S., Byers, B. S., Honic, M., Wolf, C. D., and Hall, D. M. (2025). Decentralized Data Networks for Lifecycle Management in the Built Environment. Journal of Information Technology in Construction (ITcon), 30(34):826–851.

Byers, B., Cheriyamulla, S., Jephta, E., Hall, D., and De Wolf, C. (2022). Using engraved QR codes to connect building components to materials passports for circular construction.

Byers, B., Emmenegger, P., Hunhevicz, J., Schumm, D., Heisel, F., Hall, D., and De Wolf, C. (2024a). Decentralized Phygital Identifiers for Circular Construction.

Byers, B., Hunhevicz, J. J., Honic-Eser, M., and De Wolf, C. (2024b). Survey Responses on Tokenized Product Passports: Advancing Circular Construction and Supply Chains. Medium: text/csv,47.45 kB.

Byers, B., Stengele, K., and De Wolf, C. (2024c). Data Carriers for Product Data in Circular Construction Supply Chains: A Quantitative Analysis.

Byers, B. S. and De Wolf, C. (2023). QR Code-Based Material Passports for Component Reuse Across Life Cycle Stages in Small-Scale Construction. Circular Economy, 1(2):1–16.

Byers, B. S., Emmenegger, P., Hunhevicz, J. J., Schumm, D., Heisel, F., Hall, D. M., and De Wolf, C. (2025). Decentralized phygital identifier systems for digital passports in circular construction: a design science evaluation. Construction Innovation, pages 1–35.

De Wolf, C., editor (2025). The Art of Connecting. Reuse of the Huber Pavilions. Detail Special. DETAIL, München.

De Wolf, C., Çetin, S., and Bocken, N. M. P., editors (2024). A Circular Built Environment in the Digital Age. Circular Economy and Sustainability. Springer International Publishing, Cham.

Dervishaj, A., Hernández Vargas, J., and Gudmundsson, K. (2023). Enabling reuse of prefabricated concrete components through multiple tracking technologies and digital twins.

Forth, K., Kaltenegger, J., Petrova, E., and De Wolf, C. (2025). BIM-based material property enrichment using text-based and ontology-based semantic similarity matching. IOP Conference Series: Earth and Environmental Science, 1554(1):012043.

GS1 (2025). bSDD GS1 Identifiers.

Guillem, A., Abergel, V., Roussel, R., Comte, F., Pamart, A., Luca, L. D., Guillem, A., Abergel, V., Roussel, R., Comte, F., Pamart, A., and Luca, L. D. (2025). Bridging the Provenance Knowledge Gap Between 3D Digitization and Semantic Interpretation. Heritage, 8(11).

Hagedorn, P., Donkers, A., Petrova, E., and König, M. (2025). Decentralised Data Exchange in Construction Data Spaces Using Information Containers. volume 6 of Computing in Construction, pages 0–0. European Council on Computing in Construction.

Hunhevicz, J. (2023). Blockchain for a Circular Digital Built Environment. In IoT, Blockchain, and Smart Environments. Design Computation.

Hunhevicz, J. J., Hall, D. M., Brasey, P.-A., Bonanomi, M. M. M., and Fischer, M. (2024). Decentralized project delivery on the crypto commons: Conceptualization, governance mechanisms, and future research directions. Project Leadership and Society, 5:100132.

ISO (2018). ISO 19650-1:2018(en), Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) — Information management using building information modelling — Part 1: Concepts and principles.

Kebede, R. Z. (2024). Ontology and knowledge graph-based product data representation for sustainable buildings : A case of data integration and digital product passport.

Knecht, E. (2022). Future Learning Spaces - Re-Used.

Papadonikolaki, E., Krystallis, I., and Morgan, B. (2020). Digital transformation in construction-Systematic literature review of evolving concepts.

van Zandwijk, K., Donkers, A., Petrova, E., and Pauwels, P. (2025a). Transitioning to AECO Data Spaces: Challenges, Opportunities and User Requirements.

van Zandwijk, K. v., Donkers, A., Alkemade, F., Petrova, E., and Pauwels, P. (2025b). Enabling Reuse of Construction and Demolition Waste with Data Spaces: Potentials and Limitations in the Transition to Circularity.

Vangelova, S., Blum, C. P., Forth, K., Jacob, P., Hunhevicz, J. J., and Wolf, C. D. (2026). A Graph-Based Circular Workflow for Advancing Reuse in Construction by Tracing Building Products Across Projects. Computing in Construction. European Council on Computing in Construction.

Vangelova, S., Forth, K., and De Wolf, C. (2025). Towards a Compatible Digital Product Passport Data Template in Construction. IOP Conference Series: Earth and Environmental Science, 1554(1):012053.

Whyte, J. K. and Hartmann, T. (2017). How digitizing building information transforms the built environment. Building Research & Information, 45(6):591–595. _eprint: https://doi.org/10.1080/09613218.2017.1324726.

Wijewickrama, M., Rameezdeen, R., and Chileshe, N. (2021). Information brokerage for circular economy in the construction industry: A systematic literature review. Journal of Cleaner Production, 313:127938.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3(1):160018.

Yin, R. K. (2018). Case study research and applications: design and methods. SAGE, Los Angeles London New Delhi Singapore Washington DC Melbourne, sixth edition edition.

Zhang, A. and Seuring, S. (2024). Digital product passport for sustainable and circular supply chain management: a structured review of use cases. International Journal of Logistics Research and Applications, 27(12):2513–2540. _eprint: https://doi.org/10.1080/13675567.2024.2374256.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

SCENARIO-BASED MATERIAL-FLOW AND CIRCULARITY ASSESSMENT OF DEEP RENOVATION PROJECTS

Mohamed T. Elnabwy[1], Omar Doukari[1], and Pablo Martinez[1]
[1]School of Architecture and Built Environment, Faculty of Science and Environment, Northumbria University, Newcastle upon Tyne, NE1 8ST, United Kingdom

Abstract

Waste generation in deep renovation is driven by onsite activities and their temporal sequencing and scheduling, yet most existing circularity assessment approaches remain scenario and execution agnostic. This paper proposes a schedule-aware BIM-based approach that integrates activity-zone modelling with precedence-constrained scheduling to enable time-based estimation of material flows and waste generation. Implemented through the Circular Intelligent-based Waste Engine (CIWE), the approach supports activity-level waste estimation and onsite circularity assessment. A multi-residential renovation case in Lille, France is used to compare three renovation scenarios. Results show that renovation sequencing significantly influences waste intensity, recovery potential and circularity outcomes.

Introduction

The construction sector is a major contributor to global resource consumption and waste generation, making it a central target for circular economy transitions (dos Santos Gonçalves et al., 2025). In Europe, deep renovation of existing buildings is increasingly promoted to reduce operational energy demand while limiting embodied environmental impacts associated with new construction (Hertwich et al., 2019). However, deep renovation causes substantial material inflows and demolition waste that must be effectively managed to support circular economy goals such as waste reduction, reuse, and recycling (Hossain et al., 2017).

Building Information Modelling (BIM) has significantly advanced construction and demolition waste estimation by enabling element and activity-level quantification and supporting evaluation of renovation and deconstruction strategies (Cheng and Ma, 2013). Recent studies have extended BIM-based approaches toward circularity principles by incorporating material recyclability, reuse potential and whole-life performance indicators (Akanbi et al., 2018). Nevertheless, the literature shows that most BIM-based circularity assessments remain largely static, relying on aggregated material quantities and overlooking the temporal and operational dynamics of renovation execution (Zhao et al., 2025).

Renovation sequencing, spatial dependencies and access constraints influence waste intensity, recovery feasibility and onsite logistics, particularly in multi-storey residential renovation context (Mrema et al., 2023). However, many existing BIM-based frameworks treat renovation activities as temporally independent, failing to integrate scheduling logic with material-flow and circularity assessment (Ashrafi et al., 2025). As summarized in Table 1, prior approaches generally lack zone-based modelling, precedence-constrained scheduling and time-resolved waste profiling. To address these gaps, this paper proposes a scenario-based, schedule-aware BIM-enabled framework integrating activity-zone modelling with precedence-constrained scheduling, implemented through CIWE, and demonstrated on a deep renovation case study in Lille, France.

Table 1: BIM-based waste and circularity frameworks.

Framework	Cheng and Ma (2013)	Akanbi et al. (2018)	Jiang et al. (2025)	This Study
BIM-based material data	✓	✓	✓	✓
Activity-level waste estimation	✓	✓	✓	✓
Zone-based modelling	×	×	×	✓
Execution-aware scheduling (sequencing & precedence)	×	×	×	✓
Execution-level time-resolved waste profiling	×	×	×	✓
Circularity & material-flow assessment	×	✓	✓	✓

Material and Methods

Framework overview

The proposed framework enables schedule-aware circularity assessment by structuring renovation projects around activity-zone execution logic and precedence-constrained scheduling. By integrating BIM-derived quantities with execution sequencing, it supports time-resolved quantification of material inflows, waste outflows and circularity indicators, which captures the influence of renovation sequencing on waste generation and recovery outcomes. The methodological workflow consists of five sequential stages:

1. Definition of renovation activities and scenarios
2. BIM-based activity-zone mapping
3. Precedence-constrained renovation scheduling

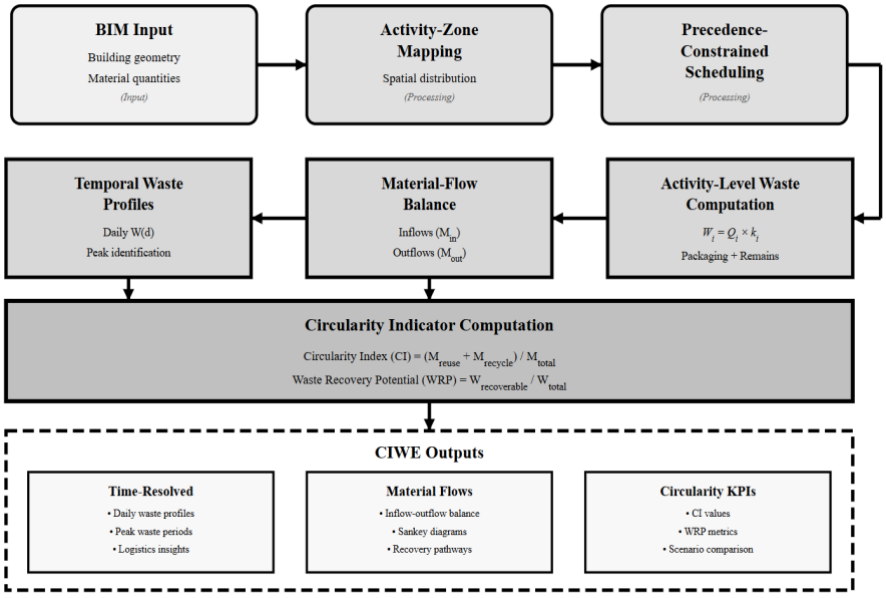


Figure 1: Overview of the CIWE framework for schedule-aware material-flow and circularity assessment

4. Activity-level waste and material-flow estimation
5. Computation of circularity indicators

Figure 1 illustrates the overall workflow and highlights the interaction between BIM data, activity sequencing, and temporal material-flow modelling implemented through CIWE. The CIWE, developed in this study, is implemented as a plug-in tool within Autodesk Revit, enabling automated extraction and activity-level quantification of material flows and waste generation during renovation projectsParts of the paper.

Case study

The proposed framework is demonstrated using a multi-residential building undergoing deep renovation in Lille, France, representative of a typical European multi-storey residential typology targeted by current renovation and circular economy policies (Figure 2). The building includes five floors, envelope elements, circulation areas, and external zones, enabling analysis of both vertical and horizontal renovation sequencing. A BIM model was developed as the primary source of geometric, material, and zoning data, supporting activity-zone mapping and consistent definition of renovation activities within CIWE. The model enables extraction of activity-level quantities required for material-flow and waste estimation. Three renovation scenarios with increasing intervention depth are analysed. All scenarios share identical geometry and zoning to ensure comparability, with differences arising only from renovation scope and system upgrades. Scenarios vary in envelope insulation levels, window specifications, HVAC configurations, ventilation systems and renewable energy integration, including photovoltaic and solar-thermal installations (see Table 2). It should be noted that the three scenarios are defined as alternative renovation strategies rather than sequential extensions of one another. Differences in waste generation arise primarily from variations in activity composition and execution logic, particularly the balance between demolition-intensive and installation-based interventions.

Activity-zone execution model

Renovation works are structured using an activity-zone execution model that explicitly links each renovation activity to the building zones where it is performed. The defined activities include envelope insulation, window and door replacement, HVAC and ventilation installations, internal insulation measures, pipe insulation, and renewable energy systems. This structure enables consistent allocation of material inflows and waste outputs at activity level while preserving spatial differentiation across the building.



Figure 2: Case study building in Lille, France, corresponding BIM model and CIWE GUI

Table 2: Summary of renovation scenarios applied to the Lille demonstration case

Element / System	External-wall insulation	Roof insulation	Basement-ceiling insulation	Windows	Heating /Cooling	Ventilation	RES electricity	RES thermal	Lighting
Scenario 1	EPS 12 cm (U= 0.25 W/m²K)	EPS 20 cm (U = 0.20 W/m²K)	--	Double-glazed Low-E	Decentralised ASHP	Natural	--	--	--
Scenario 2	EPS 12 cm (U= 0.25 W/m²K)	EPS 20 cm (U = 0.20 W/m²K)	--	Triple-glazed Low-E	Decentralised ASHP	Decentralised MVHR	Roof PV 15 kW	--	LED partial
Scenario 3	EPS 12 cm (U= 0.25 W/m²K)	EPS 20 cm (U = 0.20 W/m²K)	EPS 3 cm (U = 0.83 W/m²K)	Triple-glazed Low-E	Decentralised ASHP	Decentralised MVHR	Roof PV 20 kW	Solar-thermal 12 m²	Full LED replacement

Legend: ● = Activity executed in this zone | -- = Activity not applicable to this zone

Activity Group	Roof	Façades	Floor 4	Floor 3	Floor 2	Floor 1	Ground	Stairs	Entrance	External
ENVELOPE SYSTEMS										
Roof insulation replacement	●	--	--	--	--	--	--	--	--	--
Façade insulation + render	--	●	--	--	--	--	--	--	--	--
Windows and doors replacement	--	●	●	●	●	●	--	--	●	--
INTERNAL INTERVENTIONS										
Internal insulation (basement ceiling)	--	--	--	--	--	--	●	--	--	--
Pipe insulation	--	--	●	●	●	●	--	--	--	--
LED lighting replacement	--	--	●	●	●	●	--	●	--	--
BUILDING SERVICES										
HVAC system installation	--	--	●	●	●	●	--	--	--	--
Decentralised MVHR units	--	--	●	●	●	●	--	--	--	--
Ventilation ductwork	--	--	●	●	●	●	--	●	--	--
RENEWABLE ENERGY SYSTEMS										
Roof-mounted PV panels	●	--	--	--	--	--	--	--	--	--
Façade-integrated BIPV	--	●	--	--	--	--	--	--	--	--
Solar thermal collectors	●	--	--	--	--	--	--	--	--	--
Storage tanks and connections	--	--	--	--	--	--	●	--	--	--

Figure 3: Activity-zone execution matrix defining the spatial distribution and repetition of renovation activities across building

Repetitive activities, such as internal insulation, HVAC installations and pipe insulation are replicated across the building floors, reflecting vertical execution patterns typical of multi-storey renovation projects. In contrast, envelope-related and renewable energy activities are assigned to specific zones, including roofs, external façades, and external areas, capturing horizontal differentiation in renovation execution. Figure 3 presents the activity-zone execution matrix, which links spatial activity allocation with BIM-derived quantities to support consistent material-flow and waste estimation.

Precedence-constrained renovation scheduling

Renovation execution is governed by a zone-level precedence logic that defines spatial and temporal dependencies between activities, reflecting access, safety, and coordination constraints typical of multi-storey deep renovation projects. This ensures feasible sequencing while maintaining operational continuity.

Three categories of precedence constraints are applied. Vertical precedence enforces a top-down execution order for internal works, requiring upper-floor activities to be completed before lower-floor interventions. Horizontal precedence coordinates envelope and internal activities, ensuring façade insulation and window replacement are completed before internal installations commence, while preserving functional access routes.

Controlled parallelisation allows selected activities to proceed concurrently across zones once prerequisite conditions are satisfied. All precedence relationships are formalised as a directed scheduling network integrated with BIM-derived activity-zone mappings and activity durations. This enables CIWE to generate feasible renovation schedules that underpin time-resolved allocation of material inflows, waste generation, and circularity assessment. Figure 4 illustrates the zone-level precedence network applied in this study.

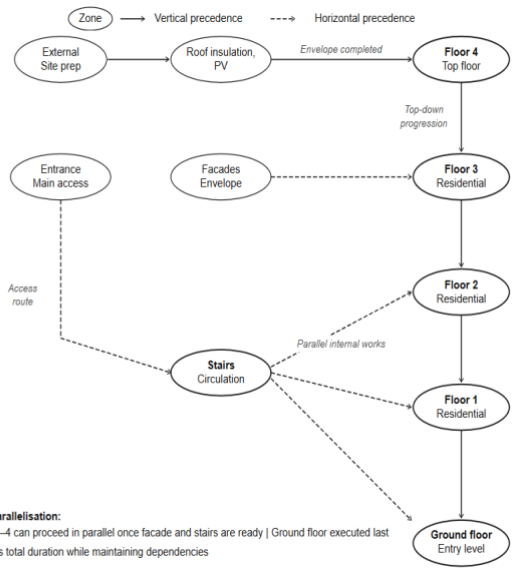


Figure 4: Zone-level precedence network governing spatial and temporal renovation sequencing in CIWE.

Activity-level waste and material-flow estimation

Waste generation is estimated at activity level using predefined waste and demolition coefficients applied to BIM-derived quantities. For each activity ($\boldsymbol{i}$), total waste generation is calculated as:

$$W_i = Q_i \cdot k_i \quad (1)$$

where (W_i) is the total waste generated by activity ($\boldsymbol{i}$), (Q_i) is the BIM-derived quantity of the activity ($\boldsymbol{i}$), and (k_i) is the corresponding waste coefficient of activity ($\boldsymbol{i}$).

The activity-level waste coefficients, packaging shares, reuse, recycling and recovery assumptions applied in this study are derived from literature (Llatas and Osmani, 2016; Quiñones et al., 2022) and expert-informed calibration (Table 3). To account for the temporal distribution of waste, activity-level waste is evenly allocated over the scheduled duration of each activity:

$$W_i(d) = \frac{W_i}{D_i} \text{for } d \in [S_i, E_i] \quad (2)$$

Where ($W_i(d)$) is the waste generated on day (d), (D_i) is the activity duration, and (S_i) and (E_i) are the scheduled start and end days respectively for activity **($\boldsymbol{i}$).**

Daily project-level waste generation is then computed as the sum of waste contributions from all active activities:

$$W(d) = \sum_{i=1}^{n} W_i(d) \quad (3)$$

Waste outputs are divided into packaging waste and remains waste. Packaging waste corresponds to installation and transport losses, while remains waste represents demolition debris and residual materials.

For each activity ($\boldsymbol{i}$), packaging and remains waste are calculated respectively as follows:

$$W_{i,pack} = W_i * p_i \quad (4)$$

$$W_{i,rem} = W_i * (1 - p_i) \quad (5)$$

where (p_i) is the packaging share in Table 3. In parallel, material inflows associated with new installations are quantified, allowing the construction of material inflow-outflow balances:

$$M_{\text{net}} = \sum M_{\text{in}} - \sum M_{\text{out}} \quad (6)$$

Where (M_{in}) represents materials introduced through new installations and (M_{out}) represents materials removed as waste.

Circularity indicator computation

Circularity performance is evaluated using two indicators derived from CIWE outputs. The Circularity Index (CI) quantifies the share of material flows directed towards reuse and recycling pathways relative to total material throughput:

$$CI = \frac{\text{Recovered Material}}{\text{M}_{throughput}} = \frac{\sum_{i=1}^{n} \text{W}_i\left(r_i^{reuse} + r_i^{recycle}\right)}{\sum_{i=1}^{n} M_i^{in} + \sum_{i=1}^{n} \text{W}_i} \quad (7)$$

where $W_i$is the waste generated by activity i, $M_i^{in}$is the material inflow associated with activity i, $r_i^{reuse}$and $r_i^{recycle}$are the reuse and recycling coefficients assigned to activity i, respectively, and $M_{throughput}$represents the total material throughput of the renovation system, defined as the sum of all material inflows and waste outflows across renovation activities.The Waste Recovery Potential (WRP) represents the proportion of total generated waste that can be directed to recovery pathways. In this study, material-specific recovery routes refer to the predefined post-generation pathways assigned to each activity-related waste stream, namely reuse, recycling, and landfill disposal, based on the recovery assumptions reported in Table 3. Accordingly, WRP captures the share of generated waste that is recoverable through reuse and recycling pathways:

$$WRP = \frac{\sum_{i=1}^{n} \text{W}_i \times r_i^{recover}}{\sum_{i=1}^{n} \text{W}_i} \quad (8)$$

Where, $r_i^{recover} = r_i^{reuse} + r_i^{recycle}$represents the total recoverable fraction of waste for activity i. Both indicators are computed at scenario level by aggregating activity-level results using the recovery coefficients reported in Table 3.

Table 3: Activity-level waste coefficients, waste fractions, and recovery assumptions

Activity Category	Activity	Unit	Waste Coefficient (k)	Packaging Share (%)	Reuse Potential (r_i^{reuse}) (%)	Recycling Potential (r_i^{recycle}) (%)	Total Recovery Potential (r_i^{recover}) (%)
Envelope	Roof insulation replacement	dm³/m²	23	35	20	50	70
	Façade insulation + render	dm³/m²	28	25	15	50	65
	Window replacement	dm³/unit	30	40	30	45	75
Internal works	Internal insulation	dm³/m²	18	30	20	48	68
	Pipe insulation	dm³/m	8	45	25	55	80
	LED lighting replacement	dm³/unit	5	65	20	65	85
Building services	HVAC system installation	dm³/unit	30	70	18	60	78
	MVHR unit installation	dm³/unit	25	75	17	65	82
	Ventilation ductwork	dm³/m	12	55	20	55	75
Renewable energy	Roof PV panels	dm³/kW	15	80	15	70	85
	Solar thermal collectors	dm³/m²	10	75	15	65	80
	Storage tanks	dm³/unit	20	60	25	50	75

$W_i$represents the total waste generated from activity i, and the recovery coefficients $r_i^{reuse}$and $r_i^{recycle}$are applied to the total waste, not limited to the packaging fraction. Adapted from Llatas and Osmani (2016); Quiñones et al. (2022), with expert-informed calibration.

Recovery coefficients are applied to the total waste generated from each activity, not limited to the packaging fraction, and are defined at the aggregated activity level to ensure modelling consistency.

Results and Discussion

This section presents a schedule-aware assessment of renovation waste on the Lille case study. It examines how material inflows, waste generation, spatial execution, and temporal sequencing interact across the three renovation scenarios. Waste is analysed as a time-dependent and activity-driven process rather than a static cumulative outcome, to enable explicit evaluation of waste intensity, recovery potential, and circularity trade-offs.

Scenario-level material inflows and waste generation

Table 4 summarises scenario-level material inflows and waste outputs for the Lille renovation case study. As renovation depth increases, total new materials installed rise from 18,500 kg in Scenario 1 to 24,100 kg in Scenario 3, primarily due to the introduction of renewable energy systems and expanded building services. In contrast, material inflows associated with insulation and window replacement remain relatively stable across scenarios, indicating that envelope upgrades constitute a baseline intervention independent of renovation depth.

Table 4: Scenario-level material inflows and waste outputs for the Lille renovation case.

Indicator	Unit	Sc. 1	Sc. 2	Sc. 3
Material Inflows				
Total new materials installed	kg	18,500	21,200	24,100
Insulation materials (EPS)	kg	4,200	4,200	4,850
Windows and doors	kg	3,100	3,600	3,600
HVAC systems	kg	2,800	3,400	3,400
Renewable energy systems	kg	--	1,850	3,200
Other building services	kg	8,400	8,150	9,050
Material Outflows (Waste)				
Total waste generated	dm^3	50,664	42,127	33,736
Packaging waste	dm^3	13,173	11,796	10,121
Remains waste	dm^3	37,491	30,331	23,615
Packaging fraction	%	26	28	30
Remains fraction	%	74	72	70

Across the scenarios, higher material inflows do not translate into higher waste quantities, indicating a clear decoupling between material throughput and waste generation. This outcome reflects differences in renovation activities rather than overall material volumes, as scenarios with greater emphasis on system installation generate proportionally less demolition-related waste. The results highlight that waste generation in deep renovation is primarily driven by activity composition and execution logic, rather than by the magnitude of material inflows alone. At the same time, waste composition shifts toward packaging-dominated streams as renovation depth increases. This qualitative change reflects the growing contribution of packaged building services and renewable energy systems and has direct implications for on-site handling and recovery strategies. These findings demonstrate that scenario-level waste performance cannot be evaluated solely through cumulative waste quantities, motivating analysis at activity, spatial and temporal levels.

While Table 4 provides a scenario-level summary of material inflows and total waste outputs, it does not explicitly reveal the origin of waste generation across renovation systems. To improve interpretability, Table 5 presents a disaggregated breakdown of waste by renovation system and activity for each scenario.

Table 5: Waste generation by renovation system and activity across scenarios (dm^3)

Renovation System	Activity	Sc. 1 (dm^3)	Sc. 2 (dm^3)	Sc. 3 (dm^3)
Envelope		**48396**	**18363**	**1883**
	Flat roof insulation	**21302**	**0**	**0**
	Façade insulation	**27094**	**0**	**1883**
	Windows & doors replacement	**0**	**18363**	**0**
Internal works		**0**	**3794**	**0**
	Internal insulation (all floors)	**0**	**3794**	**0**
Building services		**2268**	**19975**	**31448**
	Pipe insulation (all floors)	**0**	**19975**	**0**
	Wall-mounted HVAC / storage	**2268**	**0**	**0**
	Radiant floor installation	**0**	**0**	**24511**
	Mechanical ventilation	**0**	**0**	**2320**
	Mini split installation	**0**	**0**	**2368**
	Condensing boiler installation	**0**	**0**	**2249**
Renewable energy		**0**	**0**	**405**
	Photovoltaics (roof)	**0**	**0**	**405**
Total waste generated		**50664**	**42127**	**33736**

Table 5 clarifies the origin of waste generation across scenarios. Scenario 1 is dominated by envelope-related activities, particularly façade and roof insulation, which are associated with demolition-intensive processes and therefore generate high waste volumes. In Scenario 2, part of this contribution shifts toward windows and internal works, resulting in a more distributed waste profile.

In contrast, Scenario 3 shows a significant reduction in envelope-related waste and a shift toward building

services and renewable energy systems. These activities are primarily installation-based and generate lower waste per unit compared to demolition-driven processes. Additionally, the integrated execution of multiple systems within the same intervention cycle reduces redundant removal and rework operations. This explains why Scenario 3 generates lower total waste despite involving a broader scope of interventions.

Activity-level and zone-level waste generation

Figure 5 indicates that waste generation across all scenarios is dominated by envelope-related and building services activities, with broadly consistent contribution patterns as renovation depth increases. Envelope insulation and window replacement remain major contributors, although their relative importance decreases as renovation strategies shift toward more system-based interventions. Internal insulation and HVAC-related activities contribute a stable share of total waste across scenarios, reflecting the influence of spatial repetition across residential floors rather than renovation depth. In contrast, renewable energy systems introduce a distinct pattern, increasing the share of installation- and packaging-related waste as renovation depth increases.

From a spatial perspective, activities repeated across floors generate substantial cumulative waste, while envelope and roof-related works produce more localised concentrations. Overall, these results confirm that waste generation is driven primarily by activity composition and spatial execution logic rather than by renovation scope alone, underscoring the value of activity-zone modelling for effective waste management planning.

Temporal waste profiles and schedule-induced intensity effects

Figure 6 and Table 6 illustrate the temporal waste intensity patterns across the three renovation scenarios, revealing substantial differences driven by activity sequencing and scheduling logic rather than total waste quantities. Although cumulative waste generation decreases from Scenario 1 to Scenario 3, the timing and concentration of waste generation vary markedly across scenarios. Scenario 1 exhibits the highest waste intensity, with an average daily waste generation of 160 dm³/day and a peak of 377.7 dm³/day, resulting in a peak-to-average ratio of 2.36. These pronounced peaks coincide with extensive activity overlap, generating elevated operational pressure on on-site waste handling, storage capacity, and collection logistics. Scenario 2 presents a more balanced temporal profile. Reduced parallelisation and more sequential execution of major activities lower the average daily waste generation to 113 dm³/day and the peak-to-average ratio to 1.38. This smoother distribution indicates improved alignment between execution sequencing and on-site waste management capacity.

Table 6: Daily waste generation intensity indicators across renovation scenarios.

Indicator	Sc. 1	Sc. 2	Sc. 3
Project duration (working days)	321	376	317
Total waste generated (dm³)	50,664	42,127	33,736
Average daily waste (dm³/day)	160	113	108
Peak daily waste (dm³/day)	377.7	156	136.2
Peak-to-average ratio	2.36	1.38	1.26

Despite the highest material inflows, Scenario 3 achieves the lowest waste intensity, with an average daily waste generation of 108 dm³/day and a peak-to-average ratio of 1.26. This demonstrates that extended execution durations and controlled activity overlap can substantially mitigate waste intensity, even in deeper renovation scenarios. Overall, these results confirm that waste intensity is primarily a scheduling-driven phenomenon rather than a function of cumulative waste volume. Peak waste formation, driven by execution overlap, represents a critical determinant of operational feasibility and directly influences the effectiveness of on-site logistics, labour coordination and waste segregation strategies.

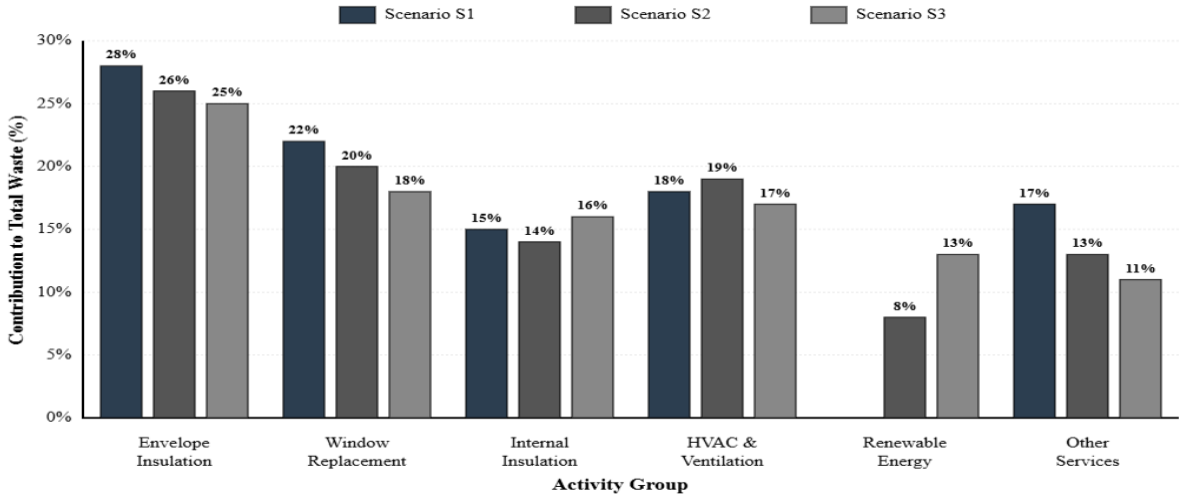


Figure 5: Waste contribution by activity group across the three renovation scenarios

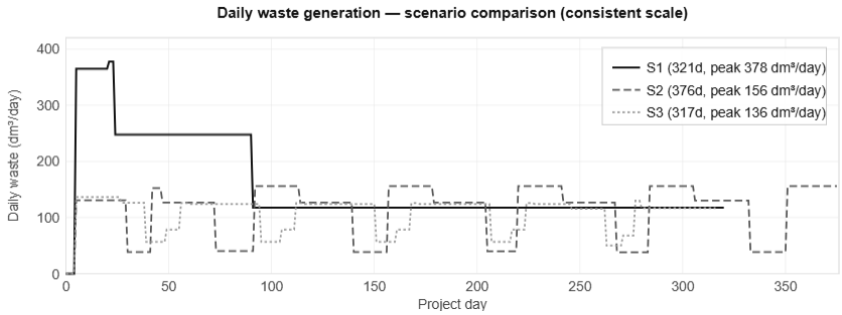


Figure 6: Daily waste generation profiles across the three renovation scenarios plotted on a consistent scale (0-400 dm³/day).

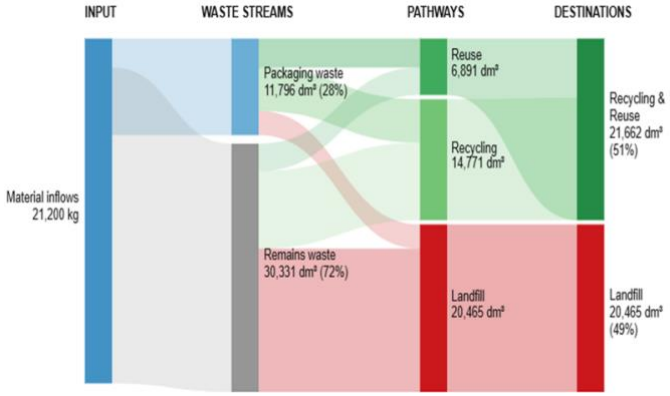


Figure 7: Material-flow pathways and recovery routes (Scenario 2)

Material-flow pathways and recovery routes

Figure 7 illustrates the material-flow pathways and reuse and recycling routes for Scenario 2 by linking material inflows, waste streams, and final destinations. From total material inflows of 21,200 kg, the renovation process generates 42,127 dm³ of waste, consisting of 11,796 dm³ of packaging waste (28%) and 30,331 dm³ of remains waste (72%). The Sankey diagram indicates that a substantial share of both waste streams is directed toward reuse and recycling pathways. In total, 29,910 dm³ (71%) of generated waste is directed toward recycling and reuse, while 12,217 dm³ (29%) is sent to landfill. Packaging waste exhibits high reuse and recycling potential due to its relatively homogeneous and uncontaminated nature, making it well suited to established reuse and recycling processes. In contrast, remains waste constitutes the primary contributor to landfill disposal, reflecting the presence of mixed, damaged, or contaminated materials generated during removal and demolition activities

It should be noted that recovery coefficients for remains waste are literature-derived estimates representing theoretical potential under optimized sorting conditions, and actual rates may be lower in practice due to material contamination and heterogeneity. Although remains waste dominates total volumes, its relatively lower recoverability drives landfill dependency, underscoring the need to distinguish waste streams at activity and scheduling levels. Scenario 2 illustrates how structured material flows aligned with reuse and recycling pathways can enhance overall recovery performance.

Circularity performance and scenario trade-offs

Circularity performance across the three renovation scenarios is evaluated using the Circularity Index and the Waste Recovery Potential, which together capture the balance between material recovery and overall waste generation. These indicators provide an integrated view linking material inflows, waste composition, and recovery outcomes. Circularity performance improves consistently with increasing renovation depth (Table 7).

Table 7: Scenario-level circularity indicators for the Lille renovation case.

Indicator	Sc. 1	Sc. 2	Sc. 3
Circularity Index (CI)	0.58	0.62	0.65
Waste Recovery Potential (WRP)	0.67	0.71	0.74
Total material throughput (kg)	28,400	31,800	35,200
Recoverable waste fraction (%)	67	71	74
Landfill fraction (%)	33	29	26

CI improves from 0.58 to 0.65 and WRP from 0.67 to 0.74 across scenarios, alongside a rise in total material throughput from 28,400 to 35,200 kg. Recoverable waste rises from 67% to 74% and landfill fraction decreases from 33% to 26%, confirming that deeper renovation strategies enhance circularity when recovery pathways are effectively integrated. However, as relative indicators can mask absolute flow increases, results should be interpreted alongside the absolute waste quantities in *Tables 4 and 5.*

Implications, limitations, and outlook

The results demonstrate that schedule-aware waste modelling provides actionable decision support beyond cumulative assessments by capturing activity-level, spatial, and temporal waste dynamics. This enables more informed planning of onsite logistics, waste segregation, and recovery strategies, particularly through identifying waste-intensity peaks and their scheduling drivers during renovation execution. Limitations include fixed waste coefficients, uniform daily waste distribution assumptions, and fixed recovery coefficients that reflect theoretical pathways rather than dynamic site conditions. In practice, packaging waste may occur in short bursts, and sorting efficiency may vary under peak generation. Another limitation concerns the use of relative circularity indicators. While CI and WRP improve consistently across scenarios, relative indicators can mask increases in absolute material flows. As total material throughput grows from 28,400 kg in Scenario 1 to 35,200 kg in Scenario 3, absolute landfill quantities may not decrease proportionally despite improved relative indicators. This rebound-effect risk suggests that the results should be interpreted alongside the absolute waste quantities reported in Tables 4 and 5, not solely through relative indicators. These findings should therefore be considered in light of the underlying modelling assumptions. Future work will focus on calibrating the framework using empirical site data and extending it to capture dynamic execution conditions through real-time sensing and digital twin integration, thereby strengthening its robustness and applicability for execution-aware circular renovation planning.

Conclusions

This paper introduced a schedule-aware, BIM-based approach linking activity-zone modelling with precedence-constrained scheduling to enable time-resolved assessment of material flows, waste generation and circularity performance in deep renovation. Application to a multi-residential case in Lille demonstrated that scenarios with comparable material throughput can exhibit markedly different waste intensity profiles and recovery outcomes, confirming that activity sequencing, spatial repetition and controlled parallelisation are decisive drivers of circularity performance. Deeper renovation strategies can enhance circularity despite increased material inputs when execution logic is appropriately managed. These findings position scheduling as a central lever for circular renovation performance, underscoring the need to integrate execution planning into material-flow and circularity assessment.

Acknowledgements

This project has received funding from the European Union's Horizon Europe research and innovation programme (HORIZON) under the Marie Skłodowska-Curie Actions Postdoctoral Fellowship, Grant Agreement No. 101205009.

References

Akanbi, L.A., Oyedele, L.O., Akinade, O.O., Ajayi, A.O., Delgado, M.D., Bilal, M., Bello, S.A., 2018. Salvaging building materials in a circular economy: A BIM-based whole-life performance estimator. Resour. Conserv. Recycl. 129, 175–186. https://doi.org/10.1016/j.resconrec.2017.10.026

Ashrafi, S., Vrijhoef, R., Wamelink, H., 2025. Circular renovation in construction at the meso scale: a systematic literature review and framework development. Front. Built Environ. 11, 1649637. https://doi.org/10.3389/FBUIL.2025.1649637/FULL

Cheng, J.C.P., Ma, L.Y.H., 2013. A BIM-based system for demolition and renovation waste estimation and planning. Waste Management 33, 1539–1551. https://doi.org/10.1016/J.WASMAN.2013.01.001

dos Santos Gonçalves, J., Claes, S., Ritzen, M., 2025. Measuring Circularity of Buildings: A Systematic Literature Review. Buildings 2025, Vol. 15, Page 548 15, 548. https://doi.org/10.3390/BUILDINGS15040548

Hertwich, E.G., Ali, S., Ciacci, L., Fishman, T., Heeren, N., Masanet, E., Asghari, F.N., Olivetti, E., Pauliuk, S., Tu, Q., Wolfram, P., 2019. Material efficiency strategies to reducing greenhouse gas emissions associated with buildings, vehicles, and electronics—a review. Environmental Research Letters 14, 043004. https://doi.org/10.1088/1748-9326/AB0FE3

Hossain, M.U., Wu, Z., Poon, C.S., 2017. Comparative environmental evaluation of construction waste management through different waste sorting systems in Hong Kong. Waste Management 69, 325–335. https://doi.org/10.1016/J.WASMAN.2017.07.043

Llatas, C., Osmani, M., 2016. Development and validation of a building design waste reduction model. Waste Management 56, 318–336. https://doi.org/10.1016/J.WASMAN.2016.05.026

Mrema, S.A., Lau, H.H., Liew, S.C., Ekambaram, P., Alam, M., Lee, V.C.C., 2023. Building Information Modeling (BIM) for Construction and Demolition Waste (CDW) Management: Scientometric and State-of-the-Art Review. Lecture Notes in Electrical Engineering 1072 LNEE, 191–211. https://doi.org/10.1007/978-981-99-5547-3_15

Quiñones, R., Llatas, C., Montes, M.V., Cortés, I., 2022. Quantification of Construction Waste in Early Design Stages Using Bim-Based Tool. Recycling 2022, Vol. 7, Page 63 7, 63. https://doi.org/10.3390/RECYCLING7050063

Zhao, W., Hao, J.L., Gong, G., Fischer, T., Liu, Y., 2025. Applying digital technologies in construction waste management for facilitating sustainability. J. Environ. Manage. 373, 123560. https://doi.org/10.1016/J.JENVMAN.2024.123560

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

TOWARDS HYBRID DIGITAL TWIN FOR PERFORMANCE-BASED CONTROL EVALUATION IN HYDROGEN FACILITIES

José L. Hernández[1], Ismael Lozano[1], Víctor I. Serna[1], Luis V. Jimeno[1], Jesús Samaniego[1] and Francisco J. Miguel[1]
[1] CARTIF technology center, Boecillo, Spain

Abstract

Current digital twins for hydrogen energy systems mainly rely on physics-based simulations. Still, data-driven approaches are not fully deployed, limiting the ability to support systematic evaluation of control strategies under diverse operating scenarios. This paper addresses this gap by proposing a hybrid digital twin framework for hydrogen-based testing and experimentation facilities, integrating physics-based and data-driven models, which combines time-series and contextual data. The framework enables scenario-based experimentation and performance-oriented control evaluation. While no experimental results are yet presented, the proposed approach establishes a scalable foundation for performance-based feedback loops, supporting future development of adaptive, energy-efficient, and low-carbon control strategies.

Introduction

Hydrogen energy systems are increasingly recognized as a cornerstone technology for decarbonizing energy systems and enabling high penetration of renewable energy sources. The major potential of hydrogen (H_2) lies in its versatility in production, storage, and utilization, as well as its ability to mitigate grid imbalance through long-term energy storage (Taghavi, 2026). As a versatile energy carrier, hydrogen can support sector coupling between electricity, heating, and mobility, while providing long-term storage and flexibility to energy systems with high shares of variable renewable generation.

Recent research highlights the expanding role of hydrogen in integrated energy systems. The objective is to enhance system flexibility across electricity, heat and transport sectors (Bhandari, 2024). A significant growth in both pilot projects and commercial demonstrations, including industrial facilities, is exemplified by large-scale deployments such as the Fukushima Hydrogen Energy Research Field, which produces green hydrogen from solar energy (JapanGov, 2021).

Despite these advances, hydrogen energy systems present significant technical and operational challenges due to their multi-domain nature. First of all, there is a need for efficient storage technologies and dynamic interactions between production, conversion and end-use subsystems (Kamran, 2024). Second, its design and operation are complex due to their multi-domain nature, non-linear dynamics and strong dependency on external conditions such as renewable generation, demand profiles, and operational constraints. These challenges drive to solutions to ensure safe, efficient and low-carbon operation through the application of advanced analysis and control approaches capable of capturing system interactions across multiple time scales and operating scenarios. Conventional modelling and management approaches frequently lack the capability to capture complex system dynamics under diverse operating conditions, constraining systematic performance assessment and optimization.

Under this perspective, digital twins (DTs) arise as a research initiative for hydrogen systems and this solution has expanded rapidly. In this work, a DT is defined as a dynamic digital representation of a physical system that combines data-driven and model-based components to replicate system behavior, support simulation under varying conditions, and enable decision-making through virtual experimentation.

Its main motivation is the need for a safer operation, higher reliability and scalable deployment across the hydrogen value chain (Naanani, 2025). Recent works show how hydrogen DTs are increasingly used for monitoring, anomaly detection, and risk-informed operation, but also stress the ongoing challenges of model fidelity, data availability, and real-time integration across heterogeneous subsystems (Naanani, 2025). Some works, for instance, some DT methods have been proposed to support operational decision-making for electrolyzers. They combine structured modelling workflows with plant data to maintain alignment between the virtual and physical assets (Monopoli, 2024). Beyond single assets modelling, architecture-level DT concepts have been introduced for fleets of electrolyzers (Alsharif, 2024), enabling unified data management and scalable analytics across distributed hydrogen production units.

Scaling up the asset perspective, some studies focus on facility level. These works demonstrate DT modelling for integrated hydrogen energy systems, coupling electrolysers, fuel cells, hydrogen storage and batteries. This research shows the direction towards holistic, system-level twins suitable for experimentation and control evaluation (Weerakoon, 2025). In the same line, (Alsharif, 2023) proposes software and reference

architectures for electrolyser digital twins intended to support scale-up and industrial digitalization.

Complementary to the DTs, experimental platforms and testing and experimentation facilities (TEFs) are gaining importance as environments to validate AI (Artificial Intelligence)-enabled control and operational strategies in realistic, safety-critical conditions. Some EU initiatives explicitly include hydrogen testing platforms within large-scale TEF ecosystems, reflecting the strategic role of experimentation infrastructures for accelerating innovation and deployment (CORDIS, 2025).

However, existing research rarely integrates physics-based and data-driven approaches within a unified framework. In this context, a hybrid digital twin is defined as a digital representation in which physics-based and data-driven models are combined to jointly represent system behavior. This approach leverages the complementary strengths of both paradigms. On the one hand, physics-based models rely on first-principles knowledge, ensuring physical consistency. On the other hand, data-driven models capture complex, non-linear patterns directly from operational or sensor data through adaptive learning mechanisms. By combining both approaches, hybrid digital twins can improve predictive accuracy, computational efficiency, and robustness, particularly in scenarios where neither modelling paradigm alone is sufficient (Hunde, 2025). Within this work, hybridization is implemented at component level, where different modelling paradigms are assigned to system components depending on data availability and physical knowledge.

According to (Langlotz, 2022), these hybrid methodologies follow a sequential integration. Data-driven modules perform adjustments or corrections thanks to the physics models. By coupling both frameworks, the active interaction between models is enabled (Langlotz, 2022). Moreover, hybrid modeling facilitates real-time monitoring, predictive analytics and performance optimization by continually updating models with real-world data while constrained by governing physical laws, thus enhancing reliability in complex systems (ANSYS, 2024). Hybrid analysis further show that embedding physics knowledge can reduce uncertainty and improve generalizability of data-driven predictions (Pawar, 2021).

Despite these advances, existing digital twin approaches for hydrogen systems and TEFs remain limited in their ability to support systematic, performance-based evaluation of control strategies within testing and experimentation facilities. Although efforts are put in place, current implementations typically focus on either physics-based simulation or data-driven analytics, with limited integration between both paradigms. Moreover, many DT solutions prioritize monitoring or fault detection rather than scenario-based experimentation, and they often rely predominantly on time-series data while underutilizing contextual information such as system configurations, operational modes, and boundary conditions. As a result, the potential of digital twins as virtual experimentation environments for control strategy assessment and iterative performance improvement remains largely underexplored. Additionally, existing research on hydrogen digital twins addresses different levels of granularity, ranging from detailed component-level models, such as electrolyzers, to system-level representations of integrated hydrogen energy facilities. While component-focused approaches enable high-fidelity modeling of specific assets, system-level digital twins aim to capture interactions across multiple subsystems, albeit often at the expense of detailed component dynamics. Nevertheless, no works are represented in the coupling of multi-level granularity.

Keep in mind these gaps, this paper addresses these challenges by proposing a conceptual hybrid digital twin framework for hydrogen-based testing and experimentation facilities. The proposed approach integrates physics-based and data-driven models and explicitly combines time-series and contextual data to enable scenario-based experimentation and performance-oriented control evaluation. Rather than presenting experimental results, this work focuses on defining a scalable and modular digital twin architecture that supports performance-based feedback loops and systematic comparison of control strategies. Moreover, this work combines multiple levels of granularity, while modelling single components, optimization algorithms combined them to provide a whole system recommendation for optimal operation. The lasting contribution of this paper lies in establishing a foundation for future implementation and validation of adaptive, energy-efficient, and low-carbon control solutions within hydrogen testing and experimentation environments.

The rest of the paper is organized as follows. First of all, the hydrogen-based testing facility is described. Next, the digital twin concept is explained, with the detailed approach that is the contribution in this paper. Following, the combined data sources, time-series and contextual information, are explained as the foundation of the digital twin. Finally, the discussion and conclusions section is stated.

Hydrogen-based testing facility

Hydrogen-based testing and experimentation facilities are designed to provide controlled, flexible and safe environments for the development, validation, and assessment of hydrogen technologies and operational strategies. By the integration of multiple energy systems (both generation and consumption assets and storage subsystems) with AI-driven management strategies, it is enabled the experimentation under realistic operating conditions, taking into consideration real-time data and operation conditions. At the same, safety and regulatory constraints are applied. Their modular and configurable nature makes them particularly suitable for evaluating advanced control approaches, system interactions, and scenario-based operational strategies.

The hydrogen-based testing facility considered in this work couples electrical, hydrogen and thermal subsystems in order to emulate the operation of a small-

scale, multi-domain hydrogen energy system. The facility is designed to support experimentation under different operating configurations while respecting safety, power, and communication constraints inherent to laboratory-scale infrastructures.

The hydrogen testing facility is located in the CARTIF technology center building, placed in Valladolid (Spain). Figure 1 illustrates the facilities.

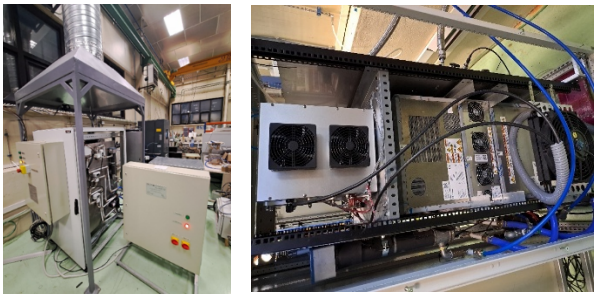


Figure 1: Hydrogen testing system in CARTIF building

The facility is centered on an AEM (Anion Exchange Membrane) electrolyzer with a hydrogen production capacity of up to 500 Nl/h (1.08 kg/day) and nominal power consumption between 2 and 3 kW, operating at pressures up to 35 bar. Hydrogen conversion is supported by a 2 kW PEM fuel cell, enabling cogeneration with fast start-up and flexible load response. The electrolyzer operates within a defined power range, introducing minimum power constraints that are representative of real-world hydrogen production systems. The installation includes metal hydride hydrogen storage, providing a storage capacity of approximately 1.6 kg.

The electrical subsystem provides power conversion bidirectional DC/DC converters. These converters couple the fuel cell output to the electrical bus, enabling regulated power delivery from hydrogen-based electricity generation. This subsystem also includes electrical energy storage in form of lithium-based batteries with a nominal capacity of approximately 2.56 kWh and a nominal voltage of 25.6 V. These batteries support high charge and discharge currents, enabling short-term buffering and fast dynamic response during experimental scenarios. Finally, electronic loads allow controlled consumption of electrical power to reproduce variable demand profiles.

The thermal subsystem provides additional flexibility for future experimentation involving thermal energy management. It includes a heat pump capable of delivering both heating and cooling, as well as an electrically heated water storage tank with a capacity of 300 liters. The thermal assets allow partial coupling between electrical, hydrogen, and thermal domains, supporting future extensions of the digital twin towards integrated energy management scenarios.

The facility is also equipped with an advanced data acquisition and control system based on PLCs (Programmable Logic Controller) and Node-RED communicating with all main components (electrolyzer, fuel cell, hydride tank, batteries, and heat pump). It allows real-time monitoring of parameters such as temperature, pressure, flow, current, and voltage, as well as the implementation of AI-based predictive and adaptive control strategies.

This hydrogen-based system is designed to operate under multiple configurations and operating modes, allowing flexible interaction between subsystems. Hydrogen can be produced stored, or converted back to electricity depending on system conditions, control objectives, and external inputs such as renewable generation and demand profiles. Electrical and thermal loads can be varied to represent different usage patterns, while storage systems provide buffering and flexibility across time scales. This multi-domain interaction introduces non-linear dynamics and strong coupling effects, which are characteristic of real-world hydrogen energy systems. In this sense, the system can reproduce variable generation and demand profiles, enabling the study of renewable integration, load shifting, and storage optimization at building and microgrid scales. For that reasons, the facility supports scenario-based analysis by enabling controlled variation of boundary conditions, system configurations, and control strategies. Forecasts of renewable generation and energy demand can be incorporated to emulate prospective operational scenarios.

Last but not least, the constraints should be taken into account. Minimum and maximum power limits for the electrolyzer, DC/DC converters, and power supply, a maximum system power rating of 2 kW, limited configurability of battery electrical connections, and the absence of direct control over fuel cell internal operating parameters. Additionally, communication limitations between certain components introduce non-ideal behaviors, which further motivate the use of digital twin-based experimentation to evaluate control strategies under realistic conditions.

In short, this hydrogen-based testing facility provides the physical foundation for the digital twin framework proposed in this paper. Its modular structure, multi-domain nature, and configurable operation make it well suited for investigating performance-based control evaluation, system-level interactions, and feedback-driven optimization through virtual experimentation.

Hybrid digital twin for performance-based control evaluation

The complexity and multi-domain interactions of hydrogen-based testing facilities require advanced digital representations capable of supporting systematic control analysis beyond conventional simulation approaches. In this context, hybrid digital twins emerge as an effective solution by combining physics-based models with data-driven components to capture both system dynamics and operational variability. This section introduces the proposed hybrid digital twin framework, designed to enable performance-based evaluation of control strategies through scenario-based experimentation. By integrating time-series and contextual data, the digital twin provides a virtual environment for analyzing control performance, exploring alternative operational configurations, and

supporting feedback-driven decision-making without interfering with real facility operation.

Digital twin conceptual approach

Figure 2 illustrates the conceptual approach of the proposed hybrid digital twin. The digital twin is fed with two data inputs: trained models and real time data. The models are trained using both time-series and contextual data to provide component-level models to be operated with the AI algorithms. The outputs of this process include control strategies and performance indicators used for evaluating operational configurations to compare the original strategy and the AI-based one, as indicated in Figure 4.

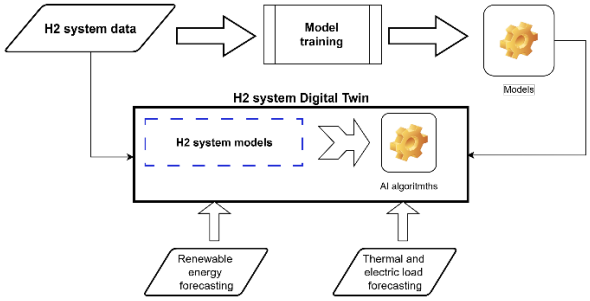


Figure 2: Digital twin conceptual approach

Entering into details, at the top level, historical, operational and contextual H_2 system data are collected from the testing facility and used within the model training process. This stage is responsible for preparing and calibrating the digital representations of the physical system. These digital representations are implemented at component level (see next section for more details). These models include both physics-based and data-driven details. The outcome of this process is a set of validated models that capture the dynamic behavior of the hydrogen system under different operating conditions.

The lower layer represents the operational digital twin, where the trained models are instantiated as H_2 system. Combined with AI-based algorithms, optimal configuration and parametrization can be obtained. Scenario-based experimentation is thus enabled through the integration of renewable energy forecasting and thermal and electrical load forecasting to determine future conditions of the operational stage. These external inputs provide prospective boundary conditions to the digital twin, allowing the evaluation of control strategies under varying demand and generation profiles.

By decoupling model training from operational experimentation and embedding forecast-driven scenarios, the proposed digital twin supports systematic and repeatable evaluation of control strategies. This structure enables performance-based comparison of alternative control approaches and establishes the foundation for feedback-driven optimization and future adaptive control deployment in hydrogen testing and experimentation facilities.

Detailed digital twin components

As explained previously, the proposed hybrid digital twin is composed of a set of modular, component-level models that collectively represent the behavior of the hydrogen-

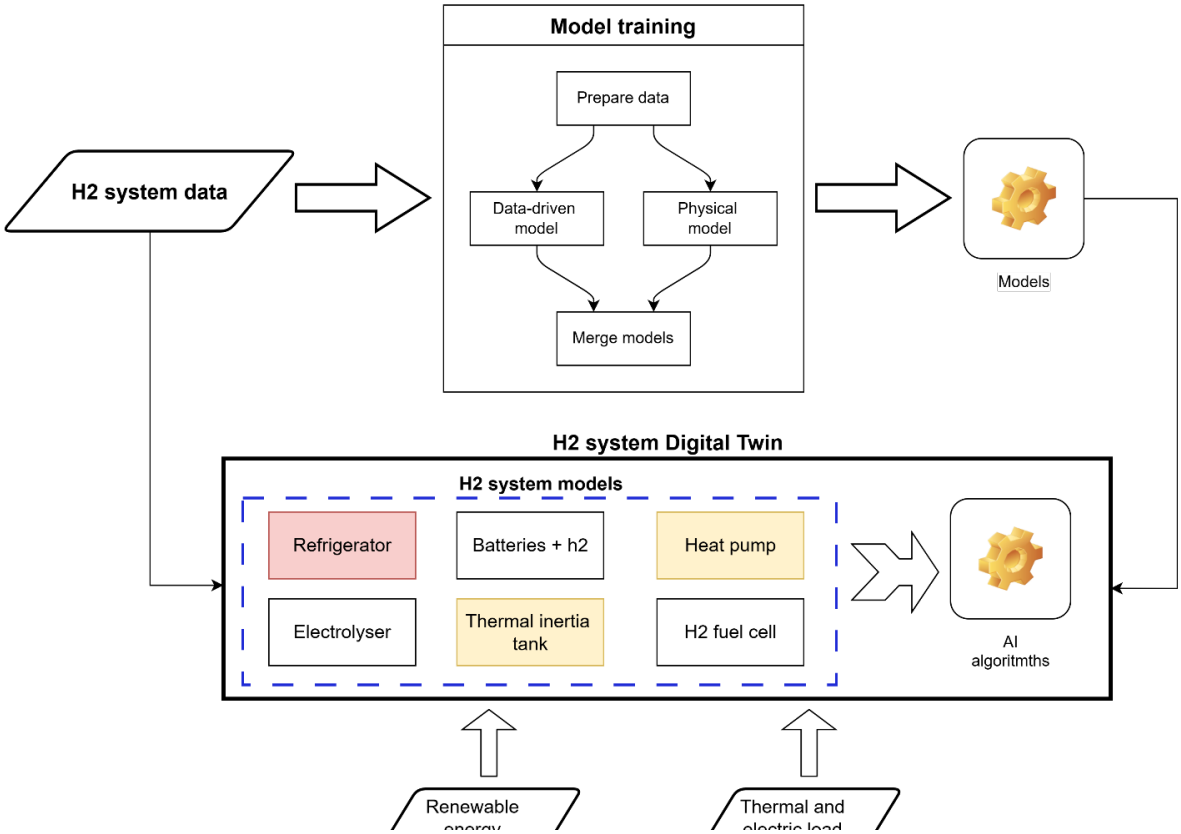


Figure 3: Detailed configuration of the digital twin for the H_2 facilities

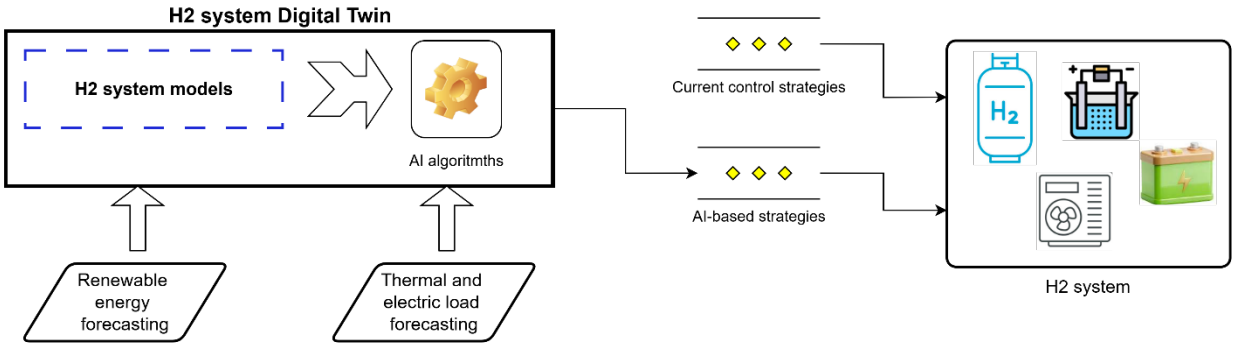


Figure 4: Comparative operational strategies for performance-based control scenarios

based testing facility. These models correspond to the main subsystems of the facility, including hydrogen production, storage, conversion, and associated electrical and thermal components. Within this layer, the digital twin operates as a virtual experimentation environment, enabling the execution of multiple scenarios without interfering with real system operation.

Figure 3 illustrates how the different component-level models are grouped and how they interact with AI-based algorithms to support scenario-based control evaluation. Each digital twin component captures the dynamic behavior and operational constraints of a specific physical asset, while their coordinated integration enables system-level analysis across electrical, hydrogen, and thermal domains for a system-level analysis.

In Figure 3, the color coding is used to explicitly represent the hybrid modelling approach adopted in the proposed digital twin.

Red components correspond to mathematical or estimation-based models, which is used for the refrigerator. The reason for this approach relates to the lack of real physical assets, while the estimation of the amount of energy used for refrigeration is estimated via indirect measurements. This allows to approximate system behavior by estimations.

Yellow components denote physics-based models derived from first-principles formulations. In the proposed architecture, this approach is adopted for the heat pump and the thermal inertia tank, where thermodynamic relationships are used to represent heat generation, storage, and dissipation processes. The use of physics-based models in these assets is motivated by the limited availability of real-time operational data.

Finally, white components represent data-driven models trained directly from historical and operational data. This modelling approach is applied to the electrolyzer, batteries, hydrogen storage system, and fuel cell, for which sufficient experimental data are available from real operation. These models capture complex, non-linear behaviors and component-specific dynamics that are difficult to describe analytically.

By combining mathematical estimations, physics-based representations, and data-driven models within a unified framework, the proposed digital twin achieves a balanced trade-off between model fidelity, adaptability, and computational efficiency. This hybrid structure enables systematic, performance-based evaluation and comparison of control strategies under forecast-driven operating conditions.

Performance-based strategies comparison

Figure 4 illustrates how the proposed hybrid digital twin is used to support performance-based comparison of alternative control strategies. In this stage, the integrated H_2 system models are coupled with AI-based algorithms. These AI algorithms interact with the system models to generate control actions, assess system responses, and compute performance indicators related to efficiency, energy balance, and operational feasibility. These algorithms include model predictive control (MPC) or reinforcement learning (RL).

To further formalize the control evaluation process, the problem addressed by the digital twin can be formulated as a constrained optimization problem. The objective is to minimize energy consumption from the grid, while maximizing renewable energy use and system efficiency. This can be expressed as:

$$\min\left(\alpha_1 \cdot E_{grid} + \alpha_2 \cdot C_{oper} + \alpha_3 \cdot H_{2_loss}\right) \quad (2)$$

where, E_{grid} is the energy supplied by the grid, C_{oper} is the cost of the operation and H_{2_loss} is hydrogen losses.

Scenario definition is driven by renewable energy forecasting and thermal and electrical load forecasting, which provide prospective boundary conditions for system operation. These forecasts allow the digital twin to emulate future operating conditions, such as variations in renewable generation availability and demand profiles. For each scenario, AI algorithms determine control actions related to hydrogen production, storage usage, energy conversion, and subsystem coordination.

Each candidate control strategy is executed within the digital twin, where the system response is simulated at component level and, then, aggregated at system level. Performance indicators are computed based on predefined metrics, split into three groups: model calibration, AI-based predictive control and hybrid energy integration and flexibility.

Table *1*,

Table 2 and Table 3 summarizes these Key Performance Indicators (KPIs) with the baseline (monitored before the implementation of the DT) and target values to be

achieved by the application of the digital twin. By repeating this process across multiple scenarios and strategies, the digital twin enables objective, quantitative comparison of alternative control approaches.

Table 1: KPIs for the model calibration

KPI	KPI description	Baseline	Target value
Model accuracy index	Quantifies deviations between simulated and experimental variables.	±10% average deviation	< ±5% deviation across all monitored variables
System response time	Time to reach steady state after a setpoint change.	> 60s average	< 30s for 90% of test cases
Operational data completeness	Proportion of data successfully integrated into the AI data lake.	85% completeness	≥ 95% completeness

Table 2: KPIs for the AI-based predictive control

KPI	KPI description	Baseline*	Target value
Prediction accuracy of AI models	Measures forecasting error for energy demand and hydrogen production	MAE = 0.15	MAE ≤ 0.05
Energy cost reduction through AI control	Quantifies reduction in energy consumption.	0%	≥ 10% reduction
Renewable energy utilization rate	Ratio of renewable input energy used in hydrogen production.	70%	≥ 90%
Ratio H2 stored energy vs renewable	Ratio of energy stored in the H2 system versus the renewable input energy.	n.a.	> 65%
Ratio battery stored energy vs renewable	Ratio of energy stored in the batteries versus the renewable input energy.	n.a.	> 30%
Ratio H2 cell energy vs energy demand	Ratio of energy generated by the H2 cell versus the energy that is demanded.	n.a.	< 30%
Net consumed energy	Energy that is used from the grid.	n.a.	
Hydrogen storage cycle efficiency	Ratio between energy recovered and energy input per cycle.	80%	≥ 90%

** n.a. refers no baseline is available, just a target value to be achieved with the implementation of the DT.*

Table 3: KPIs for the hybrid energy integration and flexibility

KPI	KPI description	Baseline	Target value
System flexibility index	Ability of the integrated system to adapt to variable generation and demand.	Flexibility ratio = 0.6	≥ 0.85
Thermal–electrical coupling efficiency	Performance of heat recovery and utilization in cogeneration mode.	70%	≥ 85%

The outcomes of this comparison process support the identification of control strategies that best satisfy performance objectives and operational constraints. Moreover, the results provide structured feedback for refining control logic and parameterization, establishing the basis for performance-driven optimization and future adaptive control deployment within hydrogen testing and experimentation facilities.

Time-series and contextual data integration

The effectiveness of the proposed hybrid digital twin relies on the coherent, explicit and structured integration of time-series and contextual data originating from the hydrogen-based testing facility. This approach ensures ensure consistent model instantiation, execution, and evaluation. Unlike conventional approaches that treat operational data in isolation, this work defines a unified data representation in which dynamic system behavior and static system description are jointly exploited within the digital twin lifecycle. The proposed architecture is drawn in Figure 5, where both data types are managed through a data lake-oriented approach, ensuring continuous synchronization between them.

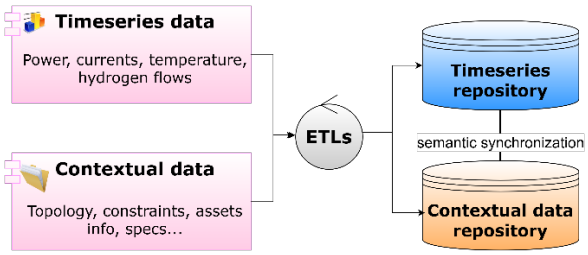


Figure 5: Integration of time-series and contextual data for hybrid digital twin execution

Data typology and roles

Time-series data represent the dynamic evolution of the hydrogen-based testing facility and consist of high-frequency measurements acquired from the physical system. These include electrical variables (power, current, voltage), thermodynamic variables (temperature, pressure), hydrogen flows, and state variables such as battery state of charge or hydrogen storage level. Additionally, forecasted inputs, including renewable generation and thermal and electrical demand profiles, are incorporated as time-dependent boundary conditions for scenario-based experimentation.

In contrast, contextual data provide the structural and semantic description of the system. This includes asset definitions (e.g., electrolyzer, fuel cell, storage units), system topology, operational constraints (e.g., power limits, pressure thresholds), nominal capacities, and configuration parameters. Contextual data also define the relationships between variables and assets, enabling the correct interpretation of measurements and ensuring consistency across models and scenarios.

This integration enables the digital twin to instantiate component-level models with the correct configuration and constraints while being continuously driven by dynamic operational data and forecasted inputs. During scenario-based experimentation, contextual data define the system configuration and control boundaries, while time-series and forecasted data provide evolving boundary conditions. As a result, the digital twin can execute reproducible simulations, evaluate control strategies under varying conditions, and compute performance indicators in a consistent and traceable way.

By combining time-series and contextual data within a unified and governed data infrastructure, the proposed approach ensures that digital twin simulations remain physically meaningful, adaptable to changing conditions, and suitable for systematic performance-based control evaluation in hydrogen TEFs.

Data model and semantic mapping

To enable integration, both data types are represented through a shared data model based on three core entities: assets, variables and observations. Each asset (e.g., electrolyzer) is associated with a set of variables (e.g., input power, hydrogen production rate), which are linked to time-series observations through unique identifiers. Contextual data define the attributes and constraints of each asset, while time-series data provide their temporal evolution. Formally, a time-series observation can be represented as the equation (2):

$$o = (a_i, v_j, t_k, x_{ijk}) \quad (2)$$

where, a_i is the asset, v_j is the associated variable, t_k the timestamp and x_{ijk} the measured or forecasted value. Contextual data define the admissible domain of the measurement, as well as the relationships between variables across assets.

This mapping enables the digital twin to interpret each data point within its physical and operational context, avoiding ambiguity and ensuring that model inputs are aligned with system constraints.

Integration and synchronization mechanism

While time-series data are stored in a repository optimized for temporal queries, contextual data are maintained in a structured repository capturing system configuration and semantics. Therefore, a synchronization mechanism is required to both repositories through:

- shared asset identifiers,
- timestamp alignment for dynamic data,
- semantic tagging of variables and units.

The data model is designed to be compatible with semantic standards such as SAREF and SAREF4ENER, enabling formalized representation of assets, variables, and their relationships, and supporting scalable and interoperable digital twin implementations.

This process ensures that each time-series input used by the digital twin is consistently mapped to its corresponding asset and variable definition. Furthermore, updates in contextual data (e.g., configuration changes or operational constraints) are propagated to the digital twin to ensure that simulations reflect the current system state.

This integrated approach enables the digital twin to operate in two complementary modes: model instantiation and scenario execution. During model instantiation, contextual data are used to configure component-level models, defining parameters such as capacities, limits, and connectivity between subsystems. During scenario execution, time-series data and forecasted inputs drive the dynamic behavior of the models, providing evolving boundary conditions.

Discussion and conclusions

This paper has presented a conceptual hybrid digital twin framework for hydrogen-based testing and experimentation facilities, aimed at enabling performance-based evaluation of control strategies under realistic and forecast-driven operating scenarios. The proposed approach addresses key limitations identified in the current state of the art, namely the fragmented use of physics-based and data-driven models, the limited exploitation of contextual data and the lack of systematic support for scenario-based control comparison within testing environments.

By combining component-level digital twins with multiple modelling paradigms, such as mathematical estimations, physics-based representations, and data-driven models, the proposed framework achieves a balanced trade-off between physical fidelity, adaptability, and computational efficiency. This hybrid structure allows the digital twin to capture both the dynamic behavior of hydrogen systems and the operational constraints imposed by real testing facilities. The explicit integration of time-series and contextual data further ensures consistency, traceability, and reproducibility across scenarios, which is essential for meaningful performance-based control evaluation.

The framework has been designed with a strong focus on modularity and scalability, enabling its application across different levels of granularity, from individual components to integrated multi-energy systems. In this sense, the digital twin acts as a virtual experimentation environment, supporting the systematic comparison of alternative control strategies based on clearly defined performance indicators. Although no experimental validation results are presented in this work, the proposed architecture establishes the necessary foundations for structured experimentation, optimization and decision support in hydrogen TEFs.

Future work will focus on the implementation and experimental validation of the proposed hybrid digital twin within the CARTIF hydrogen testing facility. This includes the deployment of the component-level models, the integration of real-time data streams, and the execution of representative operational scenarios to assess the effectiveness of different AI-based control strategies. Further developments will address the extension of the control layer towards closed-loop adaptive operation, the calculation of performance indicators, and the inclusion of additional assets and use cases. In the longer term, the proposed framework will be evaluated for scalability and transferability to other hydrogen-based facilities and multi-energy systems within large-scale testing and experimentation infrastructures.

Acknowledgments

This work has been done under the EU funded project EnergyGuard (Large-Scale Testing and Experimentation Facility (TEF) for Assessing, Validating, and Enhancing AI-Powered Next-Generation Energy Solutions), under Grant Agreement 101172705. The authors would like to thank the rest of the contributing partners of the project for the support and help in the developing of the work.

During the preparation of this work the authors used DeepL and ChatGPT in order to improve the grammar and readability of some sentences. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Alsharif, S., Huxoll, N., Raeiszadeh, A, Brandt, T., Brand, M. and Lehnhoff, S. (2023). Digital Twins for Scaling up Hydrogen Electrolysis. ETG Congress 2023, Kassel, Germany, 2023, pp. 1-7.

Alsharif, S., Huxoll, N., Jelke Wibbeke, Grimm, T., Brand, M. and Lehnhoff, S. (2024). Digital Twin concept and architecture for fleets of hydrogen electrolysers. Frontiers in Energy Efficiency, 2. doi: 0.3389/fenef.2024.1437214.

Ansys.com. (2024). Hybrid Digital Twin - Ansys. [online] Available at: https://www.ansys.com/applications/hybrid-analytics [Accessed 28 Jan. 2026].

Bhandari, R. and Adhikari, N. (2024). A comprehensive review on the role of hydrogen in renewable energy systems. International Journal of Hydrogen Energy, 82, pp.923–951. doi: 10.1016/j.ijhydene.2024.08.004.

CORDIS, cordis.europa.eu (2025). Large-Scale Testing and Experimentation Facility (TEF) for Assessing, Validating, and Enhancing AI-Powered Next-Generation Energy Solutions. [online] CORDIS | European Commission. Available at: https://cordis.europa.eu/project/id/101172705 [Accessed 28 Jan. 2026].

Hunde, J.M., Ochono, T.S., Senevirathne, D., Eneyew, D.D., Bitsuamlak, G.T., Capretz, M.A.M. and Grolinger, K. (2025). Data-driven and physics-based modeling approaches and their integration in building digital twins: A systematic review. Journal of Building Engineering, 114, p.114214. doi: 0.1016/j.jobe.2025.114214.

Kamran, M. and Marek Turzyński (2024). Exploring hydrogen energy systems: A comprehensive review of technologies, applications, prevailing trends, and associated challenges. Journal of energy storage, 96, pp.112601–112601. doi:10.1016/j.est.2024.112601.

Langlotz, P., Klar, M., Yi, L., Hussong, M., Sousa, F.J.P. and Aurich, J.C. (2022). Concept of hybrid modeled digital twins and its application for an energy management of manufacturing systems. Procedia CIRP, 112, pp.549–554. doi:10.1016/j.procir.2022.09.098.

Monopoli, D., Semeraro, C., Abdelkareem, M.A., Alami, A.H., Olabi, A.G. and Dassisti, M. (2024). How to build a Digital Twin for operating PEM-Electrolyser system – A reference approach. Annual Reviews in Control, 57, pp.100943–100943. doi: 0.1016/j.arcontrol.2024.100943.

Naanani, H., Nachtane, M. and Faik, A. (2025). Advancing hydrogen safety and reliability through digital twins: Applications, models, and future prospects. International Journal of Hydrogen Energy, 115, pp.344–360. doi: 10.1016/j.ijhydene.2025.02.440.

Pawar, S., Ahmed, S.E., San, O. and Rasheed, A. (2021). Hybrid analysis and modeling for next generation of digital twins. Journal of Physics: Conference Series, 2018(1), p.012031. doi: 10.1088/1742-6596/2018/1/012031.

Taghavi, A., Niknam, T., Shojaeiyan, S. and Rodriguez, J. (2026). The future of hydrogen as a strategic enabler in integrated energy systems: Technological developments, barriers, and policy implications. Energy Strategy Reviews, 63, p.101999. doi: 0.1016/j.esr.2025.101999.

The Government of Japan - JapanGov. (2021). The World's Largest Hydrogen-Production Facility on the Path to Zero Emissions. [online] Available at: https://www.japan.go.jp/kizuna/2021/03/hydrogen-production_facility.html [Accessed 28 Jan. 2026].

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

A MULTI-MODEL FRAMEWORK INTEGRATING CLIMATE PROJECTIONS AND SYSTEM DEGRADATION FOR CLIMATE-AWARE RETROFIT PLANNING

Avinash Myakal[1], Ranjith K. Soman[1], and Igor Nikolic[1]
[1]Delft University of Technology, Delft, the Netherlands

Abstract

Current building energy simulations often rely on static assumptions, neglecting dynamic factors such as climate change and component degradation, leading to performance gaps and suboptimal retrofit decisions. This research applies a multi-model framework that integrates climate projections with building envelope and HVAC degradation models within an automated workflow. Using an office building case study, the results demonstrate the efficacy of the multi-model framework while also improving understanding of maintenance–retrofit trade-offs under climate uncertainty. The analysis reveals key sensitivities and limitations of dynamic energy modelling, advancing insight into how the multi-model framework can support climate-aware, lifecycle-oriented retrofit decision-making.

Introduction

The global building sector represents a primary driver of energy demand and environmental impact, currently accounting for approximately 40% of total energy use and 36% of greenhouse gas emissions (Farahani et al., 2019). Within the European context, a significant majority of the building stock was constructed before the implementation of stringent energy efficiency standards, with many structures remaining in operation for over 50 years (Feng et al., 2022). Consequently, the pressure on primary energy resources is increasing, necessitating a rapid increase in the rate and effectiveness of energy retrofitting (Roetzel and Tsangrassoulis, 2012). Energy retrofitting, defined as the upgrade of existing building technologies, materials, and features, is essential for reducing operational costs, improving occupant comfort, and achieving long-term sustainability goals (Wang et al., 2014; Asadi et al., 2014).

Despite the importance of these interventions, the planning and scheduling of retrofits are often hampered by inaccurate energy consumption projections. Building energy consumption is a highly dynamic and complex system shaped by an intricate interplay of thermophysical properties, construction quality, local climatic conditions, and the aging of mechanical systems (Delzendeh et al., 2017; Li et al., 2021). Conventional simulation methods frequently rely on static inputs that ignore the degradation of the building envelope or the shifting baseline of future weather patterns driven by climate change (Thomas et al., 2016b). Research indicates that ignoring these factors can lead to discrepancies between predicted and actual energy use of up to 100%, resulting in misdirected investments and policy failures (Thomas et al., 2016b; Delzendeh et al., 2017).

In the domain of computing in construction, the emergence of digital twins and integrated modeling offers a pathway to address these complexities. Digital twins provide virtual abstractions of physical assets, enabling continuous monitoring and scenario-based analysis throughout the building's lifecycle (Lu et al., 2025). However, the integration of heterogeneous models (such as climate projections, material science degradation laws, and mechanical engineering maintenance factors) remains a significant technical challenge (Nikolic et al., 2019). Most existing workflows are either too simplified to capture physics-based volatility or too labor-intensive to scale across large scenario grids (Li et al., 2021; Waddicor et al., 2016).

This paper presents a novel Multi-Model (MM) framework, the Multi-Model Building Energy Consumption Simulator (MM-BECS), designed to automate and integrate the simulation of building energy systems under varying future scenarios. By leveraging the multi-modelling infrastructure (MMI) principles established in the MMviB project, the framework ensures loose coupling between participating models, facilitating the reuse of validated components (Nikolic et al., 2019; Matthijssen et al., 2023). The research question addresses how an integrated multi-model framework can provide high-fidelity annual energy trajectories to inform effective retrofitting and maintenance decisions over a building's entire 60-year lifespan. The primary contribution of this work is the development and demonstration of a containerized, tool-agnostic workflow that bridges the gap between specialised and validated physics-based models and lifecycle investment planning. The novelty of this work lies not in individual sub-models, but in their integration into a fully automated, containerised, and tool-agnostic workflow capable of generating annual energy trajectories over a full 60-year lifecycle.

Background

The scientific literature surrounding building energy simulation has increasingly recognized the limitations

of single-model approaches when applied to long-term projections. Traditional decision-support tools often frame building retrofit planning as a life cycle cost optimization (LCO) problem, utilizing Net Present Value (NPV) to identify cost-effective interventions (Wang et al., 2014; Ascione et al., 2015).

While studies by Wang et al. (2014) and Luo and Oyedele (2022) provide valuable economic insights, they frequently treat energy savings as fixed, stable inputs. This assumption is a significant deficiency, as it ignores the reality that insulation performance and HVAC efficiency decline as components age, while climate warming shifts the underlying demand baseline (Eleftheriadis and Hamdy, 2018). Consequently, many LCO and Life Cycle Assessment (LCA) models rely on externally generated, time-invariant energy trajectories that do not endogenously account for combined, time-varying effects. Such simplifications reduce the realism of the guidance provided to facility managers and policy makers.

Prior research has attempted to integrate multiple influencing factors through various coupling methods, yet each approach presents distinct trade-offs. Scripted multi-run approaches, as exemplified by de Wilde et al. (2011), are transparent and tool-agnostic but become computationally and manually burdensome as the scenario space expands. Waddicor et al. (2016) utilized decadal snapshots to evaluate the joint impact of climate warming and system aging on a historic library in Turin, finding that rising temperatures and declining chiller efficiency can double cooling demand. However, the decadal resolution used in such studies is a notable limitation, as it smooths out the inter-annual volatility driven by extreme weather events, which are critical for sizing mechanical systems and assessing resilience. In contrast, runtime co-simulation allows for bi-directional feedback between models, but introduces significant technical challenges in time synchronization and interface validation (Thomas et al., 2016a).

Existing methods also suffer from limitations in temporal horizon and spatial applicability. While Li et al. (2021) produced annual trajectories, their analysis was restricted to a 25-year horizon, which is insufficient for representing a building's full 60-year service life. Furthermore, many current climate-forcing methods rely on statistical extrapolation or morphing techniques that fail to capture the rising frequency of climate uncertainties and extremities such as heatwaves (Rodrigues et al., 2023). Finally, many approaches are geographically constrained due to reliance on local probabilistic datasets, limiting their applicability across regions in the building sector.

The critical research gap identified is the absence of an automated end-to-end methodology capable of generating annual energy consumption series for the entire operational life of a building while accounting for joint trajectories of climate change and envelope degradation. Despite the successful application of multi-modelling infrastructures in other complex energy domains, such as national infrastructure planning (Huang and Nikolic, 2024), this approach has not yet been realized for high-fidelity building energy simulation. There remains a lack of a containerized, tool-agnostic workflow that uses standardized interfaces to execute large scenario ensembles without manual oversight, thereby producing the consistent annual trajectories required for robust lifecycle financial and environmental appraisal.

Method

The methodology adopted in this study follows a structured multi-model framework, focusing on the development of an integrated simulation system that can represent the building energy consumption system as a whole (Matthijssen et al., 2023). In this paper, the term 'multi-model framework' is used consistently to describe the integration of independently developed simulation components. The framework is conceptualized using the XLRM structure, which organizes Exogenous Uncertainties (X), Policy Levers (L), Relationships (R), and Performance Metrics (M) to enable systematic decision analysis (Lempert et al., 2003). The XLRM framework for building energy systems decomposes the complex dynamics of building energy use into manageable components.

The Exogenous Uncertainties (X) include climate change pathways (RCP 2.6, 4.5, and 8.5) alongside the degradation rates of insulation materials and HVAC components. The Policy Levers (L) represent the decision space, encompassing maintenance regimes, the choice of energy-saving measures, and the timing of retrofit interventions. The Relationships (R) are governed by the physics-based energy simulation engine, while the Performance Metrics (M) are the annual values of site energy consumption, heating/cooling loads, and cumulative energy savings.

Case



Figure 1: Office building at Green Village, Delft.

The developed system will be demonstrated using the

case of the Office building at Green Village, Delft. This building features XPS polystyrene as insulation and a four-pipe fan coil unit as an HVAC system. Additionally, passive ventilation measures have been integrated into the building's design to enhance energy efficiency. This building was selected as a case study because it represents the implementation of traditional Dutch construction designs and practices. The size of the building was suitable for demonstration purposes.

Model Identification and Formulation

The research identifies validated models for three distinct domains: future weather data generation, building component degradation, and whole-building energy simulation.

Weather Model

Dynamic Downscaling and Bias Correction.
To provide future climate forcing, the study utilizes dynamically downscaled weather data from the Coordinated Regional Downscaling Experiment (CORDEX) archive. For the office building case study in Rotterdam, the MPI-ESM-LR global model and KNMI-RACMO22E regional model were selected. A hybrid bias-correction scheme was implemented using Multivariate Bias Correction (MBCn) for inter-dependent variables (temperature, humidity, wind) and univariate Quantile Delta Mapping (QDM) for shortwave irradiance (Machard et al., 2024).

Ageing Model

Degradation Laws and Maintenance Scenarios.
The degradation of the building envelope is modeled via the thermal conductivity (λ) of insulation materials. For extruded polystyrene (XPS), a polynomial degradation law was applied (Eleftheriadis and Hamdy, 2018):

$$\lambda_{\text{age}} = \Big(-0.000097502(\text{age}-1)^5 + 0.0054(\text{age}-1)^4 - 0.1126(\text{age}-1)^3 + 1.0895(\text{age}-1)^2 - 4.9677(\text{age}-1) + 28.7371 \Big)^{-1} \ \text{W m}^{-1}\,\text{K}^{-1} \quad (1)$$

For HVAC components, the study utilizes linear and exponential degradation laws for the Coefficient of Performance (COP) as a function of time (t) and maintenance factor (r) (Griffith et al., 2008; Hendron, 2006):

$$\text{COP}(t) = \text{COP}(0)\,(1 - r\,t) \quad (2)$$

$$\text{COP}(t) = \text{COP}(0)\,(1 - r)^t \quad (3)$$

Simulation Model

EnergyPlus Configuration.
EnergyPlus was selected as the simulation engine due to its feature-rich capabilities and widespread scientific validation (Crawley et al., 2008; Barbosa et al., 2015). The

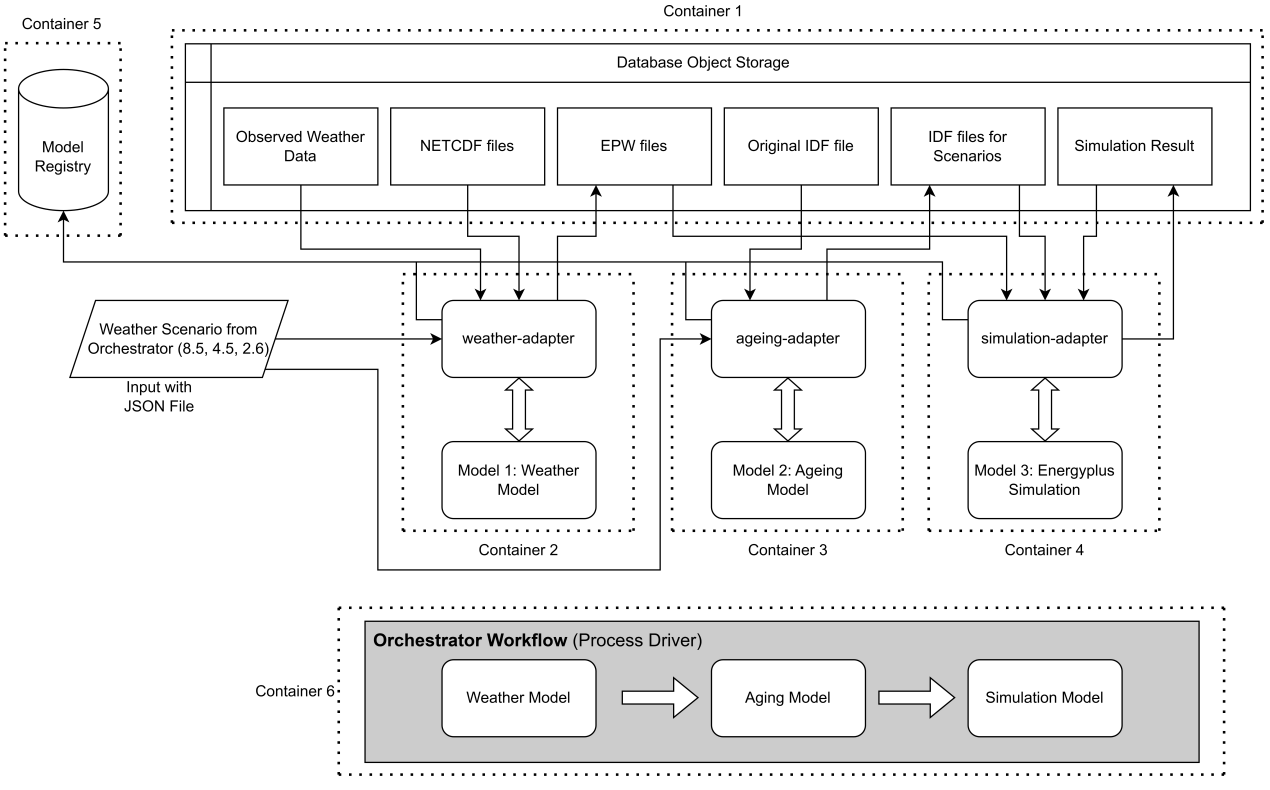


Figure 2: Multi-model Building Energy Consumption Simulator (MM-BECS) system.

engine was configured to run year-by-year simulations by pairing scenario-specific building description files (IDF) with annual weather files (EPW).

Implementation

The technical realization of the MM-BECS system utilizes a containerized architecture to achieve loose coupling and high portability, following the architectural guidelines of the MMviB project (Matthijssen et al., 2023; Nikolic et al., 2019).

Architecture and Service Adapters

Each participating model (Weather, Ageing, and Simulation) is packaged as a Docker container, providing isolation of libraries and dependencies (Matthijssen et al., 2023). Communication is mediated by adapter services, which are lightweight web applications exposing standardized RESTful APIs. These adapters handle the interplay between model-specific inputs and the central object storage service, Minio, which serves as the data repository for the multi-model workflow.

Orchestration and Workflow Automation

A model registry service maintains a catalog of available models and their health status. The orchestrator, implemented as a Directed Acyclic Graph (DAG) in Apache Airflow, automates the sequence of events: (1) preparing future weather inputs, (2) generating scenario-specific building models, and (3) executing the EnergyPlus simulations. This automated execution allows for large scenario grids to run unattended, ensuring computational tractability as the scenario space expands (Huang and Nikolic, 2024).

Results

The MM-BECS system generated annual site energy consumption, cooling load, and heating load for 28 distinct scenarios over a 60-year lifespan (2025–2084) based on the designed simulation configuration. Scenarios E1–E28 represent combinations of maintenance regimes, HVAC replacement timing, and envelope retrofit interventions.

Baseline Performance and Volatility

The baseline scenario (E1) under the RCP 8.5 climate pathway indicates a steady rise in total annual energy consumption, increasing by approximately 43% over the horizon. This trend is driven by the compounding effects of climate warming and the gradual degradation of insulation and mechanical systems. Notably, annual cooling loads nearly double, reflecting rising sensitivity to extreme summer temperatures as chiller performance decays.

Evaluation of Retrofitting and Maintenance Scenarios

A single HVAC replacement in 2055 (Scenario E2) reduces cumulative lifecycle energy consumption by 8.0%. However, implementing regular maintenance without any

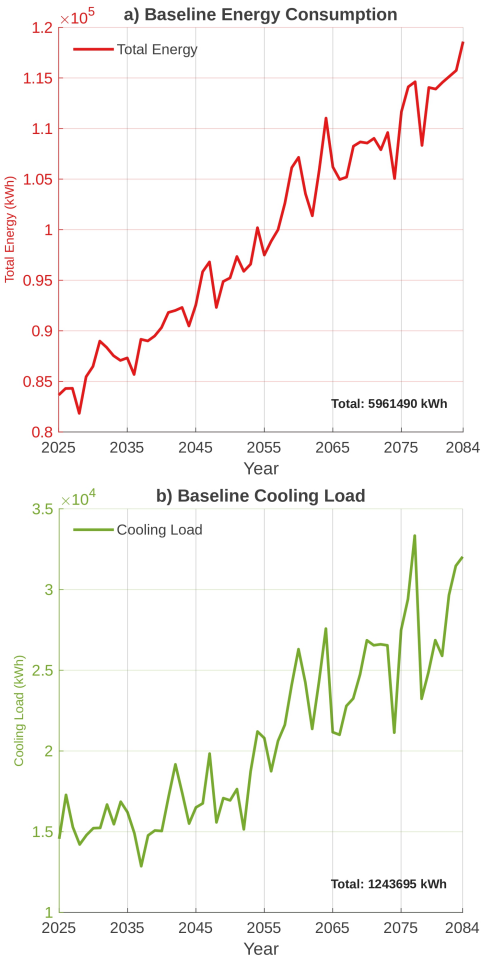


Figure 3: Performance metrics in baseline scenario under RCP 8.5.

HVAC replacement (Scenario E8) achieves lifecycle savings of 10.4%—identical to a dual-replacement scenario (E4) without maintenance. This suggests that a continuous maintenance policy can be as effective as capital-intensive retrofits if initiated early.

Integrated Policy Pathways

Scenario E28, combining dual HVAC replacements, an envelope retrofit in 2055, and regular maintenance, reduces total energy use by 13.6%. While this represents the maximum savings, Scenario E3 demonstrates that a single HVAC retrofit in 2055 followed by regular maintenance still captures 75% of these maximum savings, offering a balanced approach for resource-constrained asset management.

Discussion

The results generated by the MM-BECS system underscore the critical need for integrating dynamic factors into building energy assessments. Traditional methods relying on static models fail to capture the long-term trends and intra-decadal volatility revealed by this multi-model framework (Waddicor et al., 2016; Li et al., 2021).

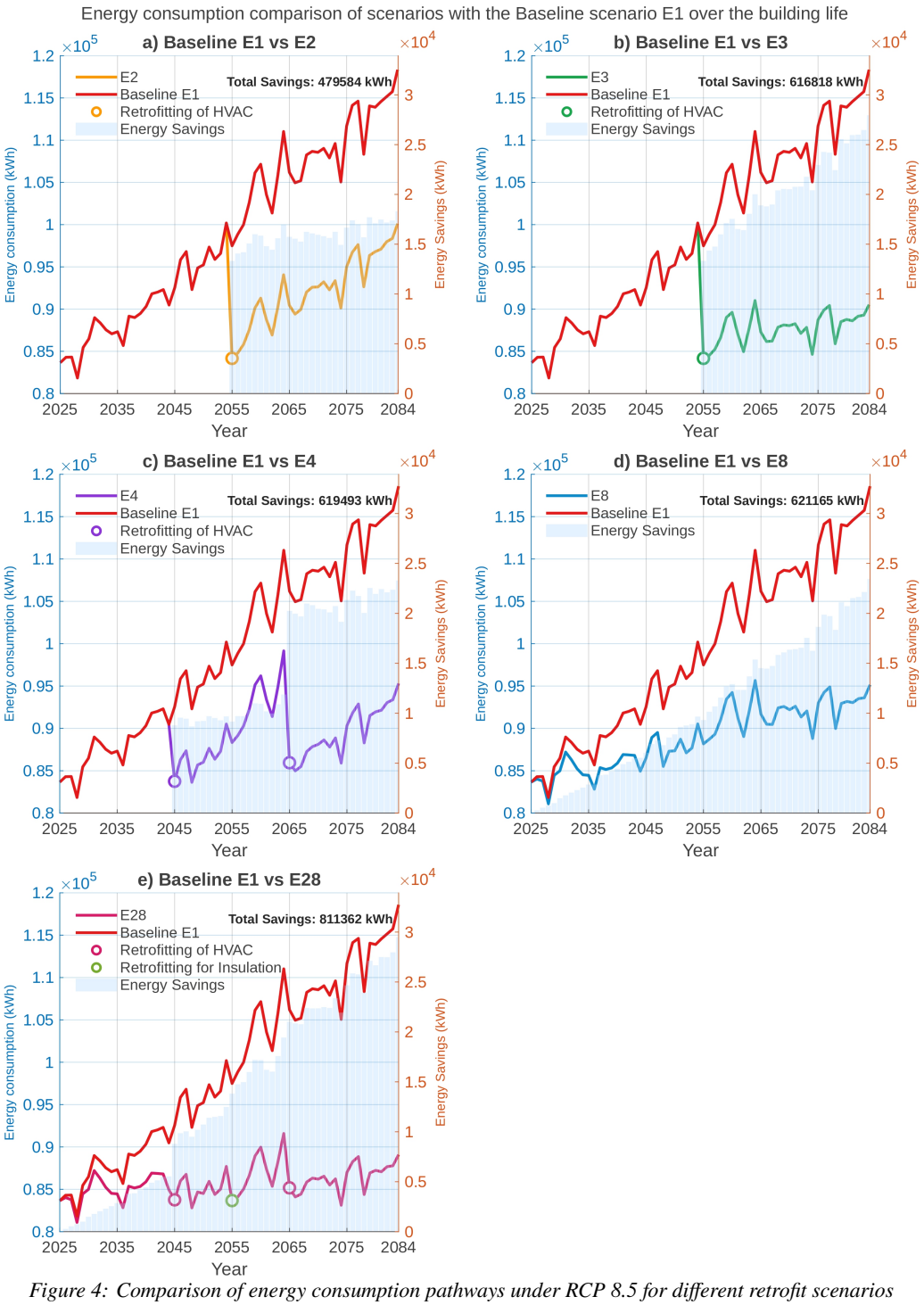


Figure 4: Comparison of energy consumption pathways under RCP 8.5 for different retrofit scenarios

The comparative analysis demonstrates that the value of an intervention is highly dependent on its timing relative to climate forcing and equipment aging (Farahani et al., 2019). Maintenance acts not only to reduce efficiency loss but also to dampen the volatility of cooling loads. For facility managers, this suggests that investment in regular maintenance can defer the need for expensive capital retrofits while ensuring performance stability.

Methodologically, the containerized architecture addresses interoperability challenges that plague bespoke simulation scripts (Huang and Nikolic, 2024). The use of RESTful adapters and centralized object storage allows for a ‘‘plug-and-play’’ environment where sub-models can be updated as new data becomes available. Furthermore, the automation of these workflows significantly reduces human labor; while traditional practice for building energy modeling can take 8 to 32 hours per case, automated pipelines like MM-BECS (and comparative frameworks like Data2BEM) can execute large-scale analyses in minutes (Lu et al., 2025). The MM-BECS framework is

transferable to other building types and regions, provided appropriate climate datasets, degradation parameters, and building models are available. The modular architecture facilitates substitution of sub-models without altering the overall workflow.

Despite its advantages, the current system treats occupant behavior as a static schedule. Future iterations should integrate agent-based occupancy models to capture socio-technical feedbacks (Delzendeh et al., 2017; Thomas et al., 2016a). Additionally, moving toward stochastic ensembles would allow for the generation of probabilistic energy trajectories, providing risk-aware guidance for investment (de Wilde et al., 2011). Furthermore, each sub-model (climate forcing, degradation, and energy simulation) is based on previously validated approaches in the literature. However, the integrated system has not been empirically validated against long-term measured datasets, which remains an area for future work. Also, this study is limited to a single office building archetype and a specific climatic context (Delft, NL). While the framework is generalisable, quantitative results are case-specific. Additionally, occupant behaviour is represented using static schedules, and uncertainty is explored through scenario variation rather than probabilistic ensembles.

Conclusions

This research has demonstrated a multi-model framework for building energy simulation that successfully integrates the dynamic, interdependent influences of climate change, envelope degradation, and HVAC aging. By operationalizing the Multi-Model Building Energy Consumption Simulator, the study has addressed the critical performance gaps inherent in conventional static energy assessments relying on fixed assumptions, neglecting dynamic factors such as climate change and component degradation. The research successfully identified and validated modular sub-models for climate forcing, envelope degradation, and system aging (linear/exponential efficiency loss). These were integrated into a containerized architecture that leverages loose coupling, enabling tool-agnostic execution across large scenario grids. Through the demonstration of a 60-year analysis horizon for an office building, the framework proved capable of generating annual site energy consumption series that capture both long-term trends and intra-decadal volatility. The systematic comparison of 28 scenarios highlighted the efficacy of strategic maintenance as a policy lever, showing that early and consistent maintenance can deliver energy savings comparable to capital-intensive retrofits while damping demand fluctuations caused by extreme weather events.

Methodologically, the study advances the state-of-the-art by moving from decadal snapshots to continuous annual trajectories, providing the high-fidelity data required for sophisticated lifecycle assessments and financial appraisals. The use of standardized interfaces and orchestration logic (Apache Airflow) ensures that the workflow is scalable, reusable, and reproducible—a significant improvement over the bespoke, labor-intensive scripts common in current professional practice. Practically, the findings empower facility managers to optimize the timing of interventions, showing that clustering or splitting retrofits based on divergent heating and cooling load trajectories can significantly improve the cost-effectiveness of energy transition strategies.

Currently, the annual trajectories produced by MM-BECS require manual export for financial evaluation. Future development should focus on the direct integration of lifecycle cost and lifecycle assessment models into the automated pipeline to provide real-time economic and environmental performance indicators. Additionally, the incorporation of a monitoring and calibration layer—utilizing real-time sub-metering and local weather observations—would transform the framework into a climate-responsive digital twin. Such a cognitive system could iteratively update degradation parameters and occupant behavior patterns, allowing for adaptive management that ensures long-term building resilience and sustainability in an increasingly volatile climatic context. This work provides a foundational template for the next generation of data-driven, portfolio-scale decision-support tools in the built environment.

Acknowledgments

The Authors acknowledge the TU Delft- IIT Delhi seed fund for enabling exchange of ideas for the development of this research. Authors also acknowledge contributions from Dr Ann Francis and Dr Daniel Hall during the execution of this research work.

References

Asadi, E., Silva, M. G. D., Antunes, C. H., Dias, L., and Glicksman, L. (2014). Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application. Energy and Buildings, 81:444–456.

Ascione, F., Bianco, N., Stasio, C. D., Mauro, G. M., and Vanoli, G. P. (2015). A new methodology for cost-optimal analysis by means of the multi-objective optimization of building energy performance. Energy and Buildings, 88:78–90.

Barbosa, R., Vicente, R., and Santos, R. (2015). Climate change and thermal comfort in southern europe housing: A case study from lisbon. Building and Environment, 92:440–451.

Crawley, D. B., Hand, J. W., Kummert, M., and Griffith, B. T. (2008). Contrasting the capabilities of building energy performance simulation programs. Building and Environment, 43:661–673.

de Wilde, P., Tian, W., and Augenbroe, G. (2011).

Longitudinal prediction of the operational energy use of buildings. Building and Environment, 46:1670–1680.

Delzendeh, E., Wu, S., Lee, A., and Zhou, Y. (2017). The impact of occupants' behaviours on building energy analysis: A research review.

Eleftheriadis, G. and Hamdy, M. (2018). The impact of insulation and hvac degradation on overall building energy performance: A case study. Buildings, 8.

Farahani, A., Wallbaum, H., and Dalenbäck, J. O. (2019). The importance of life-cycle based planning in maintenance and energy renovation of multifamily buildings. Sustainable Cities and Society, 44:715–725.

Feng, K., Lu, W., Wang, Y., and Man, Q. (2022). Energy-efficient retrofitting under incomplete information: A data-driven approach and empirical study of sweden. Buildings, 12.

Griffith, B., Long, N., Torcellini, P., Judkoff, R., Crawley, D., and Ryan, J. (2008). Methodology for modeling building energy performance across the commercial sector. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO.

Hendron, R. (2006). Building america performance analysis procedures for existing homes. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO.

Huang, Y. and Nikolic, I. (2024). Towards a multi-model infrastructure for integrated decision-making in energy transition. In Proceedings of the 23rd International Multidisciplinary Modeling & Simulation Multiconference (I3M 2024), Tenerife, Spain. Delft University of Technology. Open Access under CC BY-NC-ND 4.0.

Lempert, R. J., Popper, S. W., and Bankes, S. C. (2003). Shaping the next one hundred years : new methods for quantitative, long-term policy analysis and bibliography. RAND.

Li, L., Wang, Y., Wang, M., Hu, W., and Sun, Y. (2021). Impacts of multiple factors on energy consumption of aging residential buildings based on a system dynamics model–taking northwest china as an example. Journal of Building Engineering, 44.

Lu, J., Zheng, Z., Langtry, M., Jackson, M., Zhao, Y., Feng, C., Zhang, R., Zhang, C., Zhang, J., and Choudhary, R. (2025). Automated building energy modeling for energy retrofits using a large language model-based multi-agent framework. iScience, 28(11).

Luo, X. J. and Oyedele, L. O. (2022). Life cycle optimisation of building retrofitting considering climate change effects. Energy and Buildings, 258.

Machard, A., Salvati, A., Tootkaboni, M. P., Gaur, A., Zou, J., Wang, L. L., Baba, F., Ge, H., Bre, F., Bozonnet, E., Corrado, V., Luo, X., Levinson, R., Lee, S. H., Hong, T., Olinger, M. S., e.Silva Machado, R. M., da Guarda, E. L. A., Veiga, R. K., Lamberts, R., Afshari, A., Ramon, D., Ngo, H. N. D., Sengupta, A., Breesch, H., Heijmans, N., Deltour, J., Kuborn, X., Sayadi, S., Qian, B., Zhang, C., Rahif, R., Attia, S., Stern, P., and Holzer, P. (2024). Typical and extreme weather datasets for studying the resilience of buildings to climate change and heatwaves. Scientific Data, 11.

Matthijssen, E., Werkman, E., and Keskin, S. (2023). Multimodelling documentation. Revision `b9f89f16`.

Nikolic, I., Warnier, M., Kwakkel, J., Chappin, E., Lukszo, Z., Brazier, F., Verbraeck, A., Cvetkovic, M., and Palensky, P. (2019). Principles, challenges and guidelines for a multi-model ecology. White paper, Delft University of Technology, Faculteit Techniek, Bestuur en Management, Delft, Netherlands.

Rodrigues, E., Fernandes, M. S., and Carvalho, D. (2023). Future weather generator for building performance research: An open-source morphing tool and an application. Building and Environment, 233.

Roetzel, A. and Tsangrassoulis, A. (2012). Impact of climate change on comfort and energy performance in offices. Building and Environment, 57:349–361.

Thomas, A., Menassa, C. C., and Kamat, V. R. (2016a). An lcm framework to couple spatially distributed energy simulation and occupancy models for optimizing building energy consumption. In Construction Research Congress 2016, pages 1071–1080.

Thomas, A., Menassa, C. C., and Kamat, V. R. (2016b). System dynamics framework to study the effect of material performance on a building's lifecycle energy requirements. Journal of Computing in Civil Engineering, 30.

Waddicor, D. A., Fuentes, E., Sisó, L., Salom, J., Favre, B., Jiménez, C., and Azar, M. (2016). Climate change and building ageing impact on building energy performance and mitigation measures application: A case study in turin, northern italy. Building and Environment, 102:13–25.

Wang, B., Xia, X., and Zhang, J. (2014). A multi-objective optimization model for the life-cycle cost analysis and retrofitting planning of buildings. Energy and Buildings, 77:227–235.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

HEATING DEMAND AS AN INDICATOR OF RENOVATION DEPTH IN SOVIET-ERA APARTMENT BUILDINGS

Fariha Harun, Ergo Pikas, Jaanus Hallik
Tallinn University of Technology, Tallinn, Estonia

Abstract

Incremental renovation creates uncertain inputs for archetype-based UBEM, especially in Soviet-era apartment blocks with incomplete renovation records. This study proposes a dynamic archetyping methodology using heating demand as a diagnostic indicator for renovation depth (none/partial/deep). Simulating 44 buildings in Tallinn using the ReST (Renovation Strategy Tool), this study derives performance ranges rather than deterministic values, which account for uncertainties in user behavior and construction quality. Results show a proportional decrease in heating demand with respect to renovation depth. Deep renovation forms a distinct low-demand tier, while overlap among partial states limits identification of specific measures from energy data alone.

Introduction

In the European Union, the built environment accounts for roughly 40% of energy consumption and 36% of CO_2 emissions (Commission, 2020). Significantly reducing these numbers requires robust methods to estimate energy use across scales, from individual units to nationwide systems. While Urban Building Energy Models (UBEM) are essential for large-scale planning, their accuracy is frequently undermined by a lack of reliable, building-level input data. To address this, many UBEM studies utilize an archetype approach, grouping buildings by typology, construction period, and use (Ferrando et al., 2020; Verellen and Allacker, 2019). While archetype-based UBEM can be suitable for aggregated city-level analyses, it can be unreliable at the level of a district or individual buildings, where renovation history, system condition, and occupant behavior strongly influence energy use (Ferrando et al., 2020).

Renovation often occurs incrementally and unevenly over several decades, meaning that buildings of the same archetype often differ substantially in current envelope condition, system upgrades (such as air-to-air heat pump installation or ventilation system modification), and operational characteristics. Furthermore, national registries like the Estonian Building Register (EBR) frequently lack up-to-date records (Hamburg and Kalamees, 2019; Iliste et al., 2023). Although geometric information is often available, missing data on building service systems and envelope upgrades introduce substantial uncertainty into UBEM. Consequently, policies are often implemented based on limited information, despite significant variations in actual building conditions, rendering static, archetype-based solutions unreliable. To address this challenge, systematic data collection is essential to minimize modelling errors and uncertainty (Iliste et al., 2023).

This study addresses this gap by developing the methodological foundation for a dynamic archetyping framework that uses heating-demand distributions to infer renovation depth. To establish this framework, we adapted the European Commission's renovation depth classifications traditionally defined by primary energy savings as light (<30%), medium (30–60), and deep (>60) (European Commission, 2019) into operational renovation classes: baseline (no savings), partial (covering light and medium renovations, from (1%-60% savings), and deep (>60% savings).

This study analyzes how simulated heating demand varies across incremental renovation scenarios and evaluates whether these energy patterns can be translated into these operational renovation-depth tiers. The core research question is: Can heating demand be used to classify renovation depth (none/partial/deep) in Soviet-era apartment buildings? As a simulation-based proof-of-concept, this analysis establishes the classification logic and methodological structure required for future tool development.

Literature Review

Building envelope renovation is an effective way to reduce energy consumption and improve indoor comfort (Cai et al., 2025). Evidence from Northern European contexts: Ruokamo et al. (2025) analysed 416 Finnish housing companies' data and found that partial, maintenance-driven actions, such as window or door replacements, are significantly more common than deep renovations. These low-cost measures typically serve as the initial step in a renovation sequence, while capital-intensive upgrades like façade insulation and heating system modernization occur at later stages (Ruokamo et al., 2025).

Similar patterns are reported in Eastern Europe. Nowak-Dzieszko and Rojewska-Warchał (2015) found that many

large-panel buildings in Poland have undergone basic modernization, like balcony glazing, which can reduce heating demand in adjacent rooms by up to 22% due to increased solar gains. Large-scale renovation studies confirm that multifamily renovation typically follows a stepwise, component-by-component sequence (Abdul Hamid et al., 2018). According to Abdul Hamid et al. (2018), renovation projects commonly start with envelope measures such as façade insulation, roof improvements, and window upgrades because they directly influence energy performance, durability, and indoor comfort. Next in the sequence are actions targeting ventilation and heating systems, particularly upgrades from existing natural ventilation to mechanical ventilation with heat recovery (MVHR) or installing heat pumps. Plumbing system upgrades appear further down the renovation order and are less frequently implemented (Abdul Hamid et al., 2018). The reviewed literature indicates that technical needs, such as façade degradation and energy-performance priorities, consistently establish envelope measures as the starting point in multifamily renovation pathways.

In the Estonian context, heating energy use is strongly influenced by envelope condition, service-system operation, renovation quality, and occupant behavior. Iliste et al. (2023) show that actual wall thermal performance often deviates from calculated values due to air leakage, insulation gaps, and workmanship defects. Operational factors such as indoor temperature settings, ventilation airflow, and air leakage can further affect heating demand (Hamburg et al., 2020; Hamburg and Kalamees, 2019). Even buildings with similar typology can therefore display substantial variation in energy use.

Based on the reviewed studies, renovation in multifamily apartment buildings typically follows a sequential and incremental pathway. Moving from window replacement and minor envelope repairs, followed by roof and façade insulation, while more complex upgrades such as ventilation and heating-system improvements are generally implemented at later stages or within coordinated deep-renovation projects. This gradual evolution produces uncertain intermediate renovation states. Because heating demand may not uniquely distinguish between specific renovation combinations, this study assesses whether these patterns can instead be used to reliably infer overall renovation depth in Soviet-era buildings.

Methodology

The primary objective is to evaluate whether heating demand can be used to classify renovation depth (none/partial/deep). To demonstrate and validate this approach, a pilot study was conducted in a residential neighborhood in Mustamäe, Tallinn, Estonia. The dataset includes 44 multifamily apartment buildings (Figure 1), of which 26 belong to the 464A series and 18 to the 317A series. In Estonia, buildings are grouped by series, reflecting the industrialized, standardized designs used to address post-WWII housing shortages. (Iliste et al., 2023). Series-level classification is more reliably available and widely used in practice. While each series contains subtypes (Type) varying by story count or minor details, this study adopts the series level as the primary analytical unit. This choice ensures the tiering approach remains practical and applicable even when building registries lack granular subtype data. To verify that the proposed heating-demand tiers reflect physical renovation depth rather than modeling artifacts, a subset of 12 buildings belonging to subtype 1-464A-1kE is analyzed separately as a validation case.



Figure 1: Study area in Mustamäe, Tallinn, apartment buildings

Modelling approach

All buildings were modelled in ReST tool (Renovation Strategy Tool) (Hallik et al., 2024). Each model was based on building-specific as-built data obtained from the Estonian Building Register (EHR). Input parameters include building geometry, net heated floor area, envelope properties, window areas, and façade orientation.

Although buildings are grouped by series, they are not treated as identical archetypes; each building retains its specific physical characteristics. Therefore, baseline heating demand may still vary within the same series due to subtype-driven differences in geometry, heated area, and façade orientation. To isolate the effect of envelope modifications, all non-envelope assumptions (including user behavior, ventilation settings, air leakage parameters, and heating-system settings) were held constant across simulations. At the subtype (type) level, buildings sharing identical geometric and envelope inputs produce single baseline values. However, unless geometric and orientation-specific inputs are identical, variation in baseline heating demand can still occur between buildings.

Renovation scenarios and input parameters

Renovation scenarios were generated by systematically modifying selected envelope thermal properties within each building model. Specifically, thermal transmittance values for windows (U_{window}), walls (U_{wall}) and roof (U_{roof}) were applied according to the combinations described in Table 1. For renovation-depth classification *Table 1*, scenarios were grouped into three categories: (i)

non-renovated, (ii) Partial (one or two envelope measures), and (iii) Deep (full envelope upgrade: windows, walls, and roof). No ventilation system upgrades (e.g., heat-recovery ventilation) or HVAC system modifications were modelled in this study. In this study, 'Wall insulated' refers exclusively to full envelope insulation to ensure a consistent comparison of renovation depth. To represent typical incremental renovation pathways in Soviet-era apartment buildings, a set of envelope scenarios was defined by varying insulation thickness and the corresponding thermal transmittance (U-values) of walls and roofs.

Table 1: Renovation scenarios and renovation-depth labels used in the ReST simulations

Scenario	Envelope measures applied	Renovation depth
None	Baseline	None
Single measure	Windows replaced	Partial
Single measure	Roof insulated	Partial
Single measure	Wall insulated (full)	Partial
Double measure	Window and roof	Partial
Double measure	Window and wall	Partial
Double measure	Wall and roof	Partial
All Combined	Window, roof, and wall	Deep

For both building series, insulation thicknesses of 150 mm and 250 mm were applied to both walls and roofs. This corresponds to a range between $U_{wall,max}$= 0.21 W/m²K and $U_{wall,min}$ = 0.17 W/m²K, and for roof $U_{roof,max}$= 0.21 W/m²K and $U_{roof,min}$ = 0.17 W/m²K, respectively, to represent a practical minimum-maximum range for heating demand. Window performance was modelled using $U_{window,max} = 1.8$ W/m²K (improved condition) and $U_{window,min} = 0.9$ W/m²K (best-case performance). In this study, all existing windows were assumed to be replaced in window upgrades involved in renovation scenarios. To complement the absolute heating-demand ranges, the impact of each renovation scenario is expressed as a percentage reduction relative to the non-renovated baseline $HD_{baseline}$. Percentage reduction is calculated as:

$$\%\Delta HD = \frac{(HD_{baseline} - HD_{scenario})}{HD_{baseline}} \times 100 \qquad (1)$$

Where HD represents annual heating demand. This proportional view supports comparison across buildings with different baseline levels and clarifies the practical effect of each renovation package.

Results

Across both building series (317A and 464A) (Table 2), window-only and roof-only measures deliver modest heating-demand reductions. Full window replacement results in reductions of 12–14%, while roof insulation yields 13–28%. In contrast, wall insulation produces significantly larger reductions, ranging from 45% to 55%. Combinations involving wall insulation result in even higher performance gains. For both series, double measure (e.g., window and wall) achieves reductions of 57–64%, while full-envelope renovation reaches the highest overall savings at 68–75%.

Table 2: Heating demand reduction (%) relative to non-renovated baseline min/max range across buildings per series.

Scenario	464A (%)	317A (%)
Windows replaced	12–14	12–13
Roof insulated	14	13–28
Wall insulated	47–50	45–55
Window and roof	24–26	23–24
Window and wall	59–61	57–64
Wall and roof	58–62	56–66
Window, roof, and wall	68–73	69–75

The intervals in Table 2 show the minimum and maximum heating reductions across the entire building series (n=44), accounting for variations in geometry, orientation, and the 150-250 mm insulation thickness used in the simulations. The pattern observed in Table 2 is consistent with the incremental renovation sequence identified in the literature review. Window-only measures result in relatively modest reductions in heating demand and are typically associated with early-stage, maintenance-driven interventions.

In contrast, wall insulation and full envelope renovation lead to substantially larger reductions, reflecting their more comprehensive and capital-intensive nature. This gradual increase in reduction levels across measures supports the interpretation of heating-demand performance as an indicator of renovation depth.

Analysis of renovation impacts on heating demand (464A series)

Figure 2 summarizes the heating-demand distributions (kWh/m²·yr) for the 464A building series.

The reported intervals and violin densities represent the overall results obtained across the entire dataset of 26 buildings, capturing the full range of performance variations caused by building-specific geometry and orientation. Heating demand decreases consistently as renovation measures accumulate, indicating a consistent correlation between renovation depth and energy performance. The violin plots provide deeper insight into the data. The "width" of each violin illustrates the probability density of the data, while the median values identify the typical performance for a building in this series under each scenario.

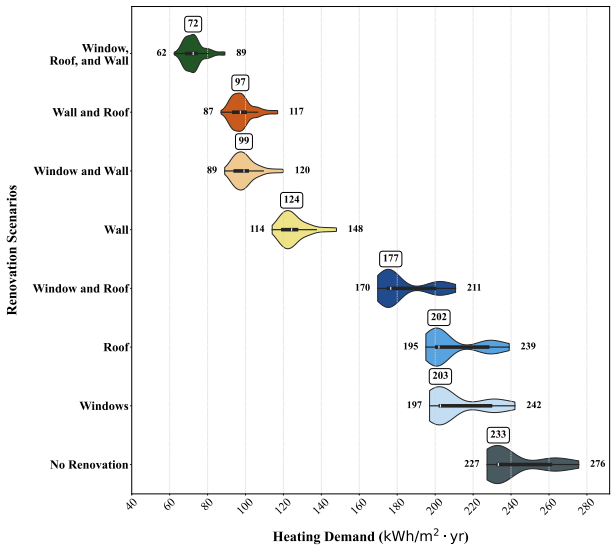


Figure 2: Heating-demand distributions (kWh/m²·yr) for renovation scenarios (464A series). The intervals and violin densities represent the overall results obtained across the entire building dataset.

However, despite the clear decrease in heating demand, most partial renovation states are not uniquely identifiable because their heating-demand ranges and densities overlap significantly. The yellow-orange shade range has been assigned to visualize wall measures better.

Influence of individual renovation measures

Wall (façade) insulation produces the most significant reduction in heating demand, forming a clearly lower performance band (114–148 kWh/m²·yr) compared to the non-renovated baseline (227–276 kWh/m²·yr). As shown in Table 2, this separation corresponds to a substantial 47–50% reduction. This shift is visually captured by the yellow-orange-shaded violins in Figure 2, where the sharp drop in the median to 124 kWh/m²·yr from 202-203(roof and window) highlights wall insulation as the key driver for lower heating demand.

In contrast, window replacement and roof insulation deliver more modest improvements. Full window replacement results in reductions of 12–14% (spanning 197–242 kWh/m²·yr with a median of 203 kWh/m²·yr), while roof insulation yields a 14% reduction (195–239 kWh/m²·yr with a median of 202 kWh/m²·yr). The blue-shaded scenarios exhibit substantial distribution overlap with each other and the baseline. The wide, irregular violins in Figure 2 show that single-measure upgrades affect buildings differently. Because of variations in size and orientation across the dataset, the impact of an upgrade can vary. This high variability makes these partial states (except wall insulation) difficult to distinguish from unrenovated buildings using heating-demand data alone.

Effects of combined renovation measures

Combining renovation measures further reduces heating demand, though several partial combinations remain difficult to distinguish. In particular, the window and wall (89–120 kWh/m²·yr) and wall and roof (87–117 kWh/m²·yr) scenarios fall within a nearly identical band. The medians for these two scenarios, 99 kWh/m²·yr and 97 kWh/m²·yr, respectively, indicate that once wall insulation is included, adding a second measure provides limited additional separation in heating-demand ranges. However, the distinct clustering of these two-measure scenarios provides strong evidence that wall insulation is present, making heating demand a reliable indicator for identifying wall-involved renovations. Although window replacement is common in practice, in this analysis, window upgrades are treated as a distinct scenario and are not assumed to be universally present across the building stock.

In contrast, full envelope renovation (window, wall, and roof) forms a clearly isolated, low-demand tier (62–89 kWh/m²·yr). With a median of 72 kWh/m²·yr, this scenario shows no overlap with any single-measure case. This clear separation indicates that deep renovation can be robustly identified from heating-demand data alone. These results align with the percentage reductions in Full window replacement results in reductions of 12–14%, while roof insulation yields 13–28%. In contrast, wall insulation produces significantly larger reductions, ranging from 45% to 55%. Combinations involving wall insulation result in even higher performance gains. For both series, double measure (e.g., window and wall) achieves reductions of 57–64%, while full-envelope renovation reaches the highest overall savings at 68–75%.

Table 2, where full envelope upgrades consistently deliver the highest performance gains across the entire building series.

Analysis of renovation impacts on heating demand (317A series)

Figure 3 shows broader ranges and stronger overlap across the partial renovation scenarios compared to 464A. Despite this increased variability, the data follow a consistent trend: heating demand decrease is proportional to the increase in renovation measures. The "width" of the violins in Figure 3 highlights higher densities, while the median values identify the typical performance for a building in this series.

Effects of Individual and Combined Measures (317A)

Wall (façade) insulation remains the primary driver of heating demand, shifting the median demand from 242 kWh/m²·yr(baseline) down to 137kWh/m²·yr, causing a reduction of approximately 45–55%. While combined measures yield larger absolute reductions, they remain partially non-unique (Figure 3). For instance, the window-and-roof scenario displays a wide range (163–240 kWh/m²·yr) that overlaps heavily with both the window-only state and the non-renovated baseline. This reinforces the finding that envelope upgrades excluding wall insulation are difficult to identify reliably from heating demand data alone. In contrast, wall-involved combinations reduce demand more substantially and consistently. Scenarios such as window-and-wall (76–136 kWh/m²·yr) and wall-and-roof (73–140 kWh/m²·yr) occupy nearly identical low-demand tiers. This suggests

that once wall insulation is implemented, adding a secondary measure provides limited additional separation in the demand range. However, full envelope renovation (window, wall, and roof) forms the lowest and most distinct band (54–104 kWh/m²·yr).

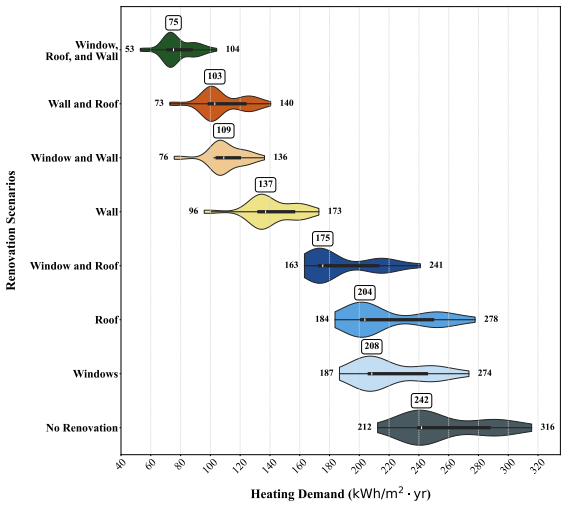


Figure 3: Heating-demand distributions (kWh/m²·yr) for renovation scenarios (317A series). The intervals and violin densities represent the overall results obtained across the entire building dataset.

Energy-Based Tiering and Classification Logic

The heating-demand distributions across both series (Figure 2 and Figure 3) indicate that while renovation depth generally correlates with heating demand reduction, specific partial renovations often overlap. To translate these simulated distributions into a practical screening tool, a consistent classification logic was applied to both the 464A and 317A series. While building-specific geometry and orientation create overlapping ranges, the following two rules were established to ensure robust categorization:

Prioritization of Common Practice: Since window replacement is the most frequent initial renovation measure in Estonia, the overlapping intervals between the non-renovated baseline and window-only ranges are intentionally assigned to Tier 2. By setting a high threshold for Tier 1, we minimize the risk of misclassifying partially renovated buildings as entirely unrenovated.

Median-Based Classification: For deeper renovations (Tiers 3 and 4), where ranges often intersect, boundaries are defined using median values rather than absolute ranges. Medians provide high-confidence statistical centers that allow for distinct classification despite the overlapping "tails" of the distributions. For the 464A series, the Tier 3 wall-involved median (99 kWh/m²·yr) is significantly higher than the Tier 4 deep-renovation median (72 kWh/m²·yr) (Figure 2). Similarly, for the 317A series, the wall-involved median (103 kWh/m²·yr) is clearly distinguished from the full-envelope median (72 kWh/m²·yr). Using this logic, typology-specific thresholds were established to categorize buildings into four distinct operational tiers (Table 3).

Table 3: Heating-demand tiering ranges (kWh/m²·yr) for renovation depth classification

Renovation Tier	464	317
Tier 1	≥242	≥275
Tier 2	170-240	163-274
Tier 3	87-147	104-162
Tier 4	<87	<103

From Table *3*, the tiers are as follows:

- **Tier 1** = None: non-renovated baseline
- **Tier 2**= Partial: Window and/or roof only
- **Tier 3** = Partial: Wall-involved renovation
- **Tier 4** = Deep: Full envelope renovation

Robustness of Heating-Demand Tiers

To ensure that the identified heating-demand tiers and their overlapping boundaries reflect the physical relationship between envelope measures and renovation depth rather than modeling artifacts, we performed two targeted checks on the 464A building series. First, we tested the stability of the boundaries against highly optimistic envelope-performance assumptions. Second, we isolated the effects of geometric variability by analyzing a subtype with identical inputs.

Impact of Optimistic U-Values

To verify the stability of the identified heating-demand tiers, we performed a robustness check on the 464A building series. We evaluated the theoretical best-case scenario for the same buildings from the 464 series using optimistic envelope values (U_{window} = 0.9; $U_{wall,min}$= 0.17 and U_{roof} = 0.08 W/m²K). Figure 4 shows the largest absolute reductions occurring in combined renovation scenarios.

Figure 4 demonstrates that every renovation scenario involving wall insulation (green) results in a non-overlapping heating-demand tier. This isolation is more pronounced under optimistic assumptions than in the realistic ranges (grey). While overlaps continue in the orange categories, the actual area of overlap is significantly reduced compared to the grey. Despite these absolute shifts, both cases follow a consistent logic: as envelope renovations intensify, heating demand decreases. This proves that the relationship between renovation depth and demand reduction is a stable physical property of the building series, rather than an artifact of specific U-value inputs.

Impact on Type-level (1-464A-1kE)

To isolate the effect of geometric variability, the same renovation scenarios were applied to subtype 1-464A-1kE. Because buildings within this subtype share identical geometric and envelope parameters, the baseline heating demand remains constant at 233 kWh/m²·yr (Table *4*).

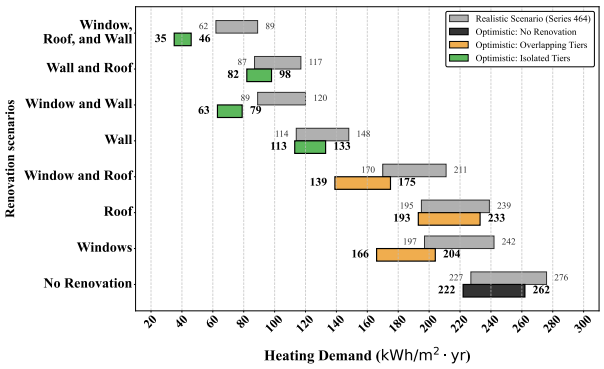


Figure 4: Robustness check (464A series) heating-demand tiers. Grey bars (realistic) vs. colored bars (optimistic U-values) show a consistent lower demand proportional to renovation scenarios. Orange signifies overlapping tiers, and green bars are non-overlapping tiers. Values represent min/max heating demand (kWh/m²·yr).

This deterministic result confirms that the wide variability observed at the broader series level is fundamentally driven by building-to-building geometric differences rather than modeling artifacts. For the renovated states, the remaining "Highest-Lowest" variation shown in Table *4* reflects envelope-performance levels (U-values). The "Highest" demand values corresponds to standard performance inputs ($U_{window,max}$ = 1.8; $U_{wall,max}$= 0.21 and $U_{roof,max}$ = 0.21 W/m²K), while the "Lowest" demand scenario corresponds to higher-performance specifications (U_{window} = 0.9; $U_{wall,min}$= 0.17 and $U_{roof,min}$ = 0.17 W/m²K).

Table 4: Heating demand (kWh/m²·yr) for subtype 1-464A-1kE, with identical geometric inputs and varying envelope U-value.

Variables	Highest	Lowest
No Renovation Applied	233	233
Windows Replaced	203	182
Wall insulation	125	120
Roof insulation	202	198
Window Replaced and Wall Insulated	99	74
Window Replaced and Roof Insulated	177	152
Wall Insulated and Roof Insulated	98	90
Window, Roof, and Wall	73	45

While real-world buildings certainly show performance variation due to partial window replacements, workmanship quality, and operational differences, those exact mechanisms are not explicitly modelled here. Instead, the high-low U-value intervals are used to represent those plausible real-world performance variations and construct realistic heating-demand ranges for each envelope upgrade.

Compared to the broader series level, the type-level analysis demonstrates significantly fewer overlapping ranges (green colors for non-overlap scenarios in Table 4), allowing for cleaner renovation tiering. However, even at this more granular type level, some overlaps remain. As marked in Table 4, these fall into two groups: single-measure overlaps (orange: windows [182–203 kWh/m²·yr] vs. roof [198–202kWh/m²·yr]) and combined-measure overlaps (blue: window and wall [74–99kWh/m²·y] vs. wall and roof [90–98kWh/m²·y]). This persistent overlap confirms that heating demand alone is insufficient to uniquely identify specific renovation states, suggesting that supplementary information (e.g., renovation records or imagery) may be required to distinguish between overlapping states.

Small Vignette:

The proposed workflow (Figure 5) is the first step of a decision-support pipeline using a rule-based classification. The tool takes building data (ID, series, and heating demand) as inputs, maps typology-specific rules, and assigns each building to a renovation-depth tier (none/partial/deep). To assist in decision-making, the output includes a confidence proxy based on the building's proximity to the nearest typology threshold. The workflow can classify a national (or large-scale) building stock into heating-demand tiers, supporting identification of deeply renovated buildings and prioritization of renovation targeting across large scales.

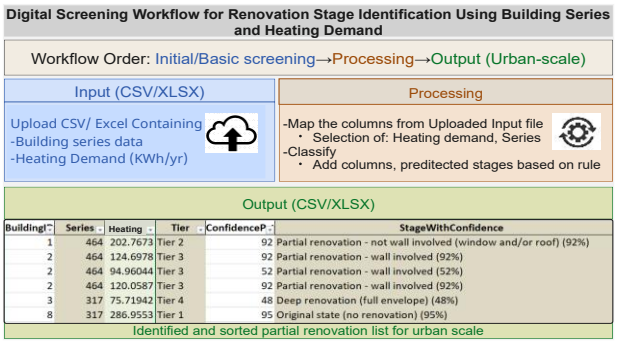


Figure 5: Digital screening workflow to classify renovation stage from building series and energy use (CSV/XLSX input and output)

By automatically post-processing heating-demand patterns as output, this workflow fills gaps in missing renovation records and establishes the operational logic for a dynamic archetyping framework.

Discussion

The findings demonstrate how simulated heating-demand patterns can be translated into renovation-depth indicators at the stock level. By combining series-level modelling, subtype validation, and robustness testing, the study establishes the analytical basis of a dynamic energy-profiling framework.

Across both building series, heating demand decreases proportionally with renovation depth, supporting a coarse renovation-depth (none/partial/deep) classification. However, substantial overlap among partial renovation scenarios shows that heating demand alone cannot uniquely identify specific intermediate measures (e.g., window-only vs roof-only). This happens because an average 1 °C increase in indoor temperature can raise

heating demand by 15% (Hamburg et al., 2020), thereby masking the 12-14% savings gained from window replacement. Consequently, when demand falls within overlapping ranges, use the logic from Table 3, or use the interactive evidence-based extension (Figure 6) to refine the classification.

A key finding is that wall insulation represents the clearest threshold intervention. In both 464A and 317A, wall-involved scenarios consistently shift buildings into a lower heating-demand tier than window-only or roof-only upgrades. While heating demand cannot fully resolve intermediate measures, it provides strong evidence for identifying wall-involved renovation states.

The robustness analysis further demonstrates that while optimistic envelope assumptions shift absolute heating demand downward, the qualitative patterns remain stable: wall insulation remains the dominant threshold intervention, deep renovation remains identifiable, and the overlap among shallow partial renovation states persists. This indicates that uncertainty among partial renovation states is not an artifact of a specific parameter choice, but rather a structural limitation of energy-only classification.

The type-level analysis is included as a controlled validation case and demonstrates how the approach could be refined when reliable subtype data are available. Because the subtype buildings share identical geometric and envelope input parameters, the baseline heating demand remains constant and does not form a range. Under these conditions, renovation scenarios also produce single deterministic values, as non-envelope assumptions are held constant (Table 4). The only remaining variation reflects differences in envelope U-value levels, resulting in a narrower and clearly defined performance range. This confirms that the broader ranges observed at the series level arise from building-to-building geometric and input heterogeneity rather than modelling instability. When reliable subtype data are available, the tiering approach can be applied at that level to improve classification accuracy. While absolute tiering ranges are series-specific, the methodology is transferable to other building stocks by using specific building geometry as the primary analytical unit to derive new, series-specific medians.

Limitations

This study is a simulation-based proof-of-concept using ReST-derived heating demand under predefined envelope scenarios. Ventilation system type can significantly affect heating demand; incorporating alternative ventilation strategies (e.g., exhaust vs balanced with HRV) may change absolute demand values. Here, ventilation was held constant to focus on envelope measures. In this study, operational and behavioural assumptions were held constant across buildings, which can affect heating consumption. Furthermore, in real buildings, uncertainty arises from (i) EBR data inaccuracies (Iliste et al., 2023), (ii) operational and occupant behavior (e.g., thermostat setpoints, ventilation operation, window airing) (Hamburg et al., 2020; Hamburg and Kalamees, 2019), and (iii) physical degradation and workmanship-related performance gaps (Hamburg et al., 2020; Hamburg and Kalamees, 2019; Iliste et al., 2023) (e.g., air leakage, insulation defects). These factors can shift heating demand from one renovation tier to another. Finally, as tier thresholds are series and heating-demand-dependent, transferability requires typology data and is limited in cooling-dominated regions, such as Southern Europe, without first adjusting the models.

Future Work

Future research will focus on calibrating this classification logic against real-world performance gaps by validating simulated results against measured heating data. This will include refining typology definitions through automated façade feature recognition and testing

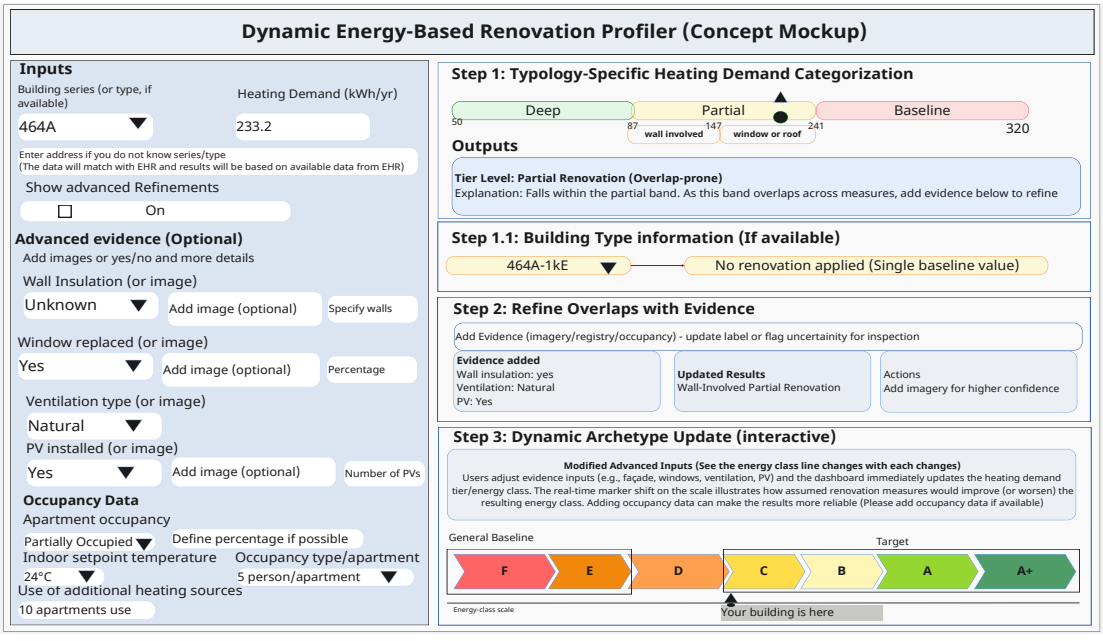


Figure 6: Concept mockup of an interactive Energy-Based Renovation Profiler

a wider variety of ventilation strategies (e.g., heat recovery variants). Furthermore, integrating heterogeneous data such as occupancy data and imagery-based assessments, through Bayesian calibration, offers a promising pathway to reduce UBEM performance gaps further (Wang et al., 2020). These steps will transition the current proof-of-concept into a reliable, multi-source decision-support tool for large-scale renovation.

Interactive heating-demand based tiering (mock-up)

The proposed heating-demand tiering provides a practical first step for archetype updating when renovation records are missing: buildings can be assigned a typology-specific renovation-depth tier from heating demand, with deep renovation being the most reliably detectable. However, overlap among partial renovation states means that energy-only classification should not be used to confidently infer specific measures (e.g., window-only versus roof-only). To address this limitation, we outline a conceptual interactive heating-demand-based tiering dashboard (Figure 6). The interface translates typology-specific thresholds into a decision-support workflow: users enter a building series and baseline heating demand to obtain an initial renovation-depth tier (none/partial/deep). Optional evidence inputs (e.g., ventilation type, registry attributes, occupancy behavior, or image-based façade) can then be incorporated to refine the classification. Users may test alternative renovation assumptions (façade, windows, roof, ventilation, PV) and observe how the assigned tier changes. This interactive layer allows rapid comparison of scenarios and highlights how additional evidence or renovation choices influence the inferred renovation status and energy class. This illustrates how typology-specific thresholds and supplementary evidence could be combined within a practical decision-support setting.

Conclusions

This study demonstrates that heating demand can be used to classify renovation depth (none/partial/deep) in Soviet-era multifamily buildings. Series-level analysis captures realistic stock heterogeneity, while subtype-level validation confirms deterministic model behaviour under identical inputs. The findings are a practical step toward establishing the methodological foundation for a dynamic archetyping framework, defining the analytical logic required for future integration with measured energy data and auxiliary building information within a decision-support environment.

Acknowledgments

This work has been supported by the Estonian Research Council grant (PSG963) and by the Estonian Centre of Excellence in Energy Efficiency, ENER (grant TK230), funded by the Estonian Ministry of Education and Research.

References

Abdul Hamid, A., Farsäter, K., Wahlström, Å., Wallentén, P., 2018. Literature review on renovation of multifamily buildings in temperate climate conditions. Energy Build. 172, 414–431.

Cai, Y., Rayegan, R., Aryal, A., 2025. An automated framework for tailored building envelope retrofit planning. J. Build. Eng. 108, 112838.

Commission, E., 2020. A Renovation Wave for Europe – Greening Our Buildings, Creating Jobs, Improving Lives (No. COM(2020) 662 final). European Commission, Brussels, Belgium.

European Commission, 2019. Commission Recommendation (EU) 2019/786 of 8 May 2019 on building renovation Authority.

Ferrando, M., Causone, F., Hong, T., Chen, Y., 2020. Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches. Sustain. Cities Soc. 62, 102408.https://doi.org/10.1016/j.scs.2020.102408

Hallik, J., Arumägi, E., Pikas, E., Kalamees, T., 2024. Comparative assessment of simple and detailed energy performance models for urban energy modelling based on digital twin and statistical typology database for the renovation of existing building stock. Energy Build. 323, 114775.

Hamburg, A., Kalamees, T., 2019. How well are energy performance objectives being achieved in renovated apartment buildings in Estonia? Energy Build. 199, 332–341.

Hamburg, A., Kuusk, K., Mikola, A., Kalamees, T., 2020. Realisation of energy performance targets of an old apartment building renovated to nZEB. Energy 194, 116874. https://doi.org/10.1016/j.energy.2019.116874

Iliste, E., Lomp, S., Pikas, E., Arumägi, E., Hallik, J., Õiger, K., Kisel, Ei., Liiv, I., Kalamees, T., 2023. Heat loss characteristics of typology-based apartment building external walls for a digital twin-based renovation strategy tool. J. Phys. Conf. Ser. 2654, 012125.

Nowak-Dzieszko, K., Rojewska-Warchał, M., 2015. Influence of the Balcony Glazing Construction on Thermal Comfort of Apartments in Retrofitted Large Panel Buildings. Procedia Eng. 108, 481–487.

Ruokamo, E., Tikkanen, T., Lukkarinen, J., Karhinen, S., 2025. Energy retrofits in shared ownership: exploring motivations in Finnish housing companies. Build. Res. Inf. 0, 1–17.

Verellen, E., Allacker, K., 2019. Inventory of the existing residential building stock for the purpose of environmental benchmarking. IOP Conf. Ser. Earth Environ. Sci. 323, 012029.

Wang, C.-K., Tindemans, S., Miller, C., Agugiaro, G., Stoter, J., 2020. Bayesian calibration at the urban scale: a case study on a large residential heating demand application in Amsterdam. J. Build. Perform. Simul. 13, 347–361.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

INTEGRATING INDICATOR REVIEW AND SOCIAL NETWORK ANALYSIS FOR SUSTAINABLE PUBLIC INFRASTRUCTURE

Anagha Divakaran, Ann Francis
Department of Civil and Environmental Engineering,
Indian Institute of Technology Delhi, New Delhi, 110016, India

Abstract

The construction sector's current sustainability frameworks are centered around private buildings, with little consideration for public infrastructure. This study reviews global rating frameworks to address these gaps and examines 629 indicators grouped into quantitative and qualitative categories. Social Network Analysis (SNA) is applied to examine the relationships among indicators, identifying 113 distinct indicator groups with a high modularity of 0.975, suggesting weak cohesion and fragmentation among the frameworks. Based on network structure centrality analysis, refined 56 cross-cutting indicators are identified. A three-level hierarchical framework is developed to select and make decisions about indicators for sustainable public infrastructure in India.

Introduction

Infrastructure development plays a significant role in the nation's overall development. Several studies confirm that the infrastructure has an important and beneficial impact on productivity, output, and long-term growth rates (Boix-Cots, Pardo-Bosch and Pujadas, 2023). Beyond economic performance, public infrastructure systems enable the delivery of critical services essential for social justice, public health, economic productivity, and overall quality of life. However, the long-term viability of these services is increasingly threatened by climate change, making sustainability and climate resilience central concerns for infrastructure planning and operation. The operational and structural integrity of the infrastructure is at risk due to the growing frequency and intensity of extreme weather events, including heatwaves, hurricanes, droughts, and floods. (Das, 2025) found that these vulnerabilities lead to service disruptions, economic losses, and increased social inequality in impoverished areas. Sustainability encompasses an assessment of environmental conservation, economic feasibility, and social equity throughout the life cycle of an asset (Sánchez-Silva and Baker, 2024). On the other hand, climate resilience involves planning for, responding to, and recovering from climate-based disruptions while maintaining services (Ferrari, 2020). Incorporating both sustainability and climate resilience into decision-making is critical for developing a sustainable and climate-resilient infrastructure (Cardin, Mijic and Whyte, 2023). This paper presents a conceptual framework that connects short-term development and performance metrics with long-term sustainability and climate resilience goals. The findings aim to assist policymakers, planners, infrastructure agencies, and asset owners by providing a cohesive method for indicator selection, benchmarking, and decision-making in public infrastructure planning and policy formulation. The following section discusses the study's literature background.

Research Background

The rating framework is not homogeneous in assessing environmental, social, and economic aspects across project lifecycles (Khan and Henschel, 2024). Widely used frameworks such as LEED, GRIHA, and IGBC primarily support environmental performance in buildings, focusing on aspects such as energy efficiency, water conservation, and indoor environmental quality in both commercial and residential buildings (Leal *et al.*, 2024), while GRIHA and IGBC focus on integrated habitat assessment and green townships according to the Indian climatic and regulatory context (Kumar, Santara and Banerji, 2020).

However, these frameworks have a drawback in that they primarily focus on private-sector buildings and individual assets with relatively short ownership cycles (Liu *et al.*, 2018). Also, these frameworks do not account for the unique characteristics of public infrastructure, including networked structures (e.g., highways, water grids), multi-decadal lifecycles, multi-level governance coordination, climate risk exposure, and long-term resilience (Liu *et al.*, 2018; Sanad *et al.*, 2025).

There are also infrastructure-centric rating frameworks globally, such as ENVISION, CEEQUAL, Greenroads, SuRe, and the IS Rating tool (Mehraban *et al.*, 2025). However, upon arrival in India, no overarching system exists for rating public infrastructure across environmental, social, economic, and resilience dimensions. Existing rating frameworks, such as GRIHA and IGBC, are for buildings or newly developed highway adaptations and do not provide a comprehensive, integrative, and harmonising system for the entire infrastructure (Kumar, Santara and Banerji, 2020; Jain and Ananda, 2024). Consequently, the existing global rating frameworks, such as ENVISION, CEEQUAL, and IS rating tools, show more bias towards the environmental dimension, and the least concentration on other

dimensions of sustainability, such as social and economic (Diaz-Sarachaga *et al.*, 2016).

The significant research gap in developing countries like India is the lack of an integrated, coherent, and consistent framework that integrates environmental, social, economic, and resilience aspects across the lifecycle of public infrastructure. Moreover, developed frameworks in advanced economies fail to include location-specific vulnerabilities, climate risks, or multi-level governance (Diaz-Sarachaga *et al.*, 2016; Kumar, Santara and Banerji, 2020) . Finally, limited use of network analysis to map interdependencies among indicators, stakeholders, and vulnerabilities hinders the transition from isolated assessments to holistic, adaptive decision-making (Khan and Henschel, 2024).

As a result, they fall short in addressing long-term performance, resilience, and adaptability under climate uncertainty, which are central to public infrastructure systems (Sanad *et al.*, 2025). Therefore, there is a lack of harmonisation of sustainability indicators across different global rating systems, such as LEED, BREEAM, and Envision. Although these frameworks include overlapping environmental, social, and economic indicators, they remain fragmented and non-standardised due to their origins in building-scale and private-sector contexts (Mansell *et al.*, 2019). This fragmentation in the industry has negative implications for sustainability assessment in public infrastructure, due to the lack of robust benchmarking, cross-framework comparisons, and comprehensive integration into decision-making (Liu *et al.*, 2018; Mehraban *et al.*, 2025).

To address these challenges, this study combines a systematic review of indicators with Social Network Analysis (SNA) to better understand how sustainability and resilience indicators relate to one another. SNA uses network-based methods to examine patterns, connections, and groupings within complex systems (Rousseau, 2002). In research concerning infrastructure sustainability, SNA elucidates intricate interconnections among indicators, stakeholders, and vulnerabilities, evolving from discrete assessments to comprehensive, adaptive frameworks that identify critical linkages and systemic issues (Khan and Henschel, 2024). While SNA has been applied in construction management research, including supply chain coordination, site operations, and health and safety (Wang, Zhang and Lu, 2018). Its use in structuring sustainability and resilience indicators for public infrastructure remains limited. Therefore, this study adopts a system-oriented perspective to examine public infrastructure as an interconnected network. Studies applying SNA in similar contexts include those that integrate vulnerability indicators with stakeholder objectives in sustainable infrastructure decision-making (Khan and Henschel, 2024). By combining a comprehensive review of sustainability and resilience indicators with Social Network Analysis (SNA), the research aims to reveal how indicators relate to one another, identify gaps and overlaps, and identify which themes dominate current assessment practices. The study further seeks to reorganise these indicators into meaningful clusters that reflect environmental, social, economic, and resilience indicators within the Indian context.

Methodology and Experiment

A four-phase methodology to develop a 3-level hierarchical conceptual framework is proposed, as shown in *Figure 1*.

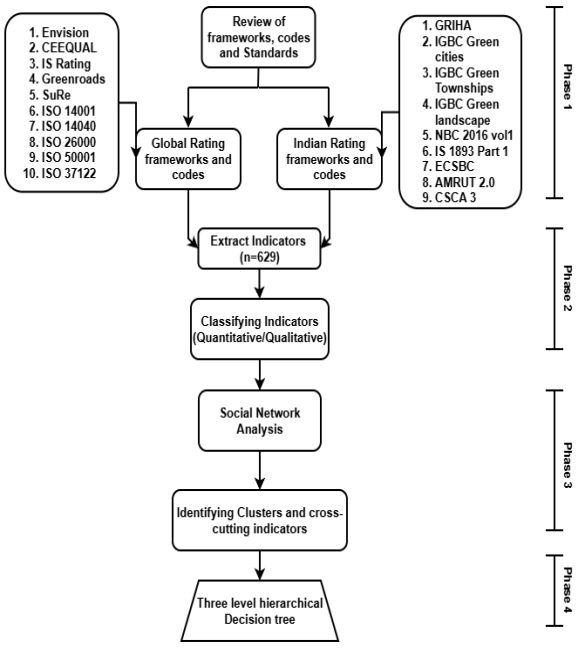


Figure 1: Framework for the development of a 3-level hierarchical decision tree.

Phase 1: Review of Existing Frameworks, Codes, and Standards

A structured review was conducted to identify and analyse sustainability and resilience assessment systems relevant to public infrastructure. This review targets formal rating systems, international standards, national codes, and policy frameworks that define indicator-based assessment criteria. The review covered five major categories:

i. Global sustainability rating systems such as CEEQUAL, SuRe, IS rating, and Greenroads (ICE, 2007; University of Washington, 2011; GIB, 2018; ISCA, 2018; ISI, 2018),
ii. Associated international standards such as ISO 14001:2015, BS EN ISO 14040:2006, ISO 26000; ISO 50001, ISO/CD 37122 (BIS, 2006; ISO, 2011; ISO/TC, 2012),
iii. Indian sustainability rating frameworks such as GRIHA, IGBC (Green township rating, Green cities, and Green landscape rating) (IGBC, 2010, 2013, 2015; TERI, 2010),
iv. Indian building and infrastructure codes such as NBC 2016, ECSBC 2024, & IS 1893 Part 1:2016 (BIS, 2016; IS, 2016; BEE, 2024),
v. Indian policy frameworks, including AMRUT 2.0 and the Climate Smart Cities Assessment Framework 3.0 (MOHUA, 2021, 2022).

Table 1 represents the classification of the Existing Global and Indian frameworks, codes, standards, and policies.

Table 1: Global and Indian sustainability rating systems, codes, and policy instruments

Global rating	Global codes	Indian rating	Indian codes	Indian policies
CEEQUAL	ISO 14001:2015	GRIHA	NBC 2016 VOL 2	AMRUT 2.0
Greenroads	BS EN ISO 14040:2006	IGBC Green Cities	IS 1893 Part 1:2016	Climate Smart Cities Assessment Framework
Envision	ISO 26000	IGBC Green Landscape	ECSBC 2024	
IS Rating	ISO 50001	IGBC Green Townships		
SuRe	ISO/CD 37122			

Frameworks were selected based on their relevance to public infrastructure, including sustainability and climate resilience dimensions, and on the availability of clearly defined indicators.

Phase 2: Indicator Extraction and Classification

In this Phase, sustainability and climate-resilience indicators identified in Phase 1 were extracted and compiled in an Excel database for further analysis. The extracted indicators were then classified into quantitative and qualitative categories to facilitate systematic analysis and standardisation. Indicators were harmonised to remove duplication while retaining their conceptual intent. Quantitative indicators represent measurable variables such as percentages of emission reductions, energy efficiency metrics, and lifecycle costs, while qualitative indicators capture aspects such as stakeholder engagement, community-influenced impact evaluations, and adaptability planning judgments. The quantitative indicators were subsequently used for Social Network Analysis.

Phase 3: Network Construction and Analysis

In this phase, Social Network Analysis (SNA) was used to analyse the relationships among the quantitative sustainability and climate resilience indicators identified in Phase 2. The process involved the following steps.

i. Network construction where each indicator is represented as a node and edges are established based on co-occurrence across multiple rating and assessment frameworks.
ii. Network visualisation and analysis using Gephi, applying the ForceAtlas2 layout to reveal structural patterns.
iii. Community detection or modularity analysis to identify clusters of closely related indicators.
iv. Centrality analysis to identify influential and cross-cutting indicators within the network.

SNA was particularly useful because it enables quantifying inter-indicator relationships, identifying tightly connected clusters, and influential indicators through centrality-based measures. These properties are critical for prioritising indicators that play a bridging or coordinating role in integrated sustainability and resilience decision-making.

Key SNA metrics, such as centrality, density, and modularity, were used to interpret the network structure and identify key structural patterns.

Phase 4: Framework Development

In this phase, the results obtained from SNA were synthesised to develop the conceptual framework. Indicator clusters identified through modularity analysis were interpreted as thematic groupings representing different dimensions of sustainability and climate resilience. The process involved the following steps:

i. Identification of thematic clusters based on community detection results.
ii. Selection of key indicators using centrality measures to identify cross-cutting and influential indicators.
iii. Refinement of indicators based on relevance to the Indian public infrastructure context and removal of redundancy.
iv. Development of a hierarchical structure organising indicators into three levels: indicators, criteria, and requirements.

Based on this process, a conceptual decision framework was established to arrange indicators into coherent thematic hierarchies, thereby facilitating integrated assessment and decision-making for India's sustainable, climate-resilient public infrastructure.

Results and Discussion

Indicator Dataset and Classification

A total of 629 sustainability and climate-resilient indicators were identified from the structured review of global and Indian frameworks, codes, standards, and policy documents relevant to public infrastructure. This comprehensive set of indicators encompasses a broad spectrum of sustainability considerations, including environmental, social, and economic dimensions relevant to infrastructure development.

A total of 387 quantitative and 242 qualitative indicators were identified. To examine the evolution of these frameworks and standards, the temporal distribution of their evolution was studied.

Figure 2 presents a timeline figure illustrating the temporal distribution and contextual diversity of the reviewed global and Indian frameworks, codes, and policy initiatives. The X-axis represents the years from 2000 to 2025, while the Y-axis is classified into five categories: global frameworks, Global codes, Indian

rating systems, Indian codes, and Indian policy documents. The graph depicts the evolution of sustainable governance at the global and national levels. Global framework development was dominated in the early to mid-2000s, reflecting the initial international push for sustainability, climate action, and infrastructure evaluation. Global codes became more prominent between 2008 and 2020.

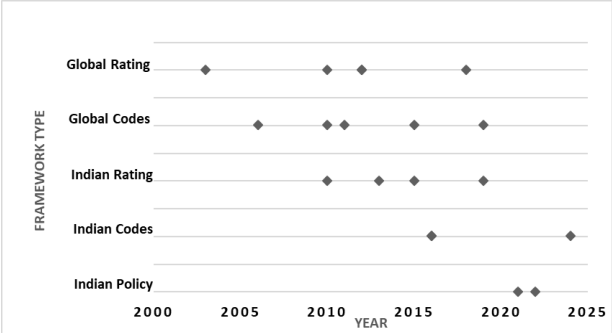


Figure 2: Temporal distribution of global and Indian frameworks, codes, and policy initiatives across different categories.

In the Indian context, rating systems emerged around 2010. This trend indicates the localisation and adaptation of global sustainability concepts into performance-based assessment tools with a specific focus on the building sector, rather than infrastructure systems. Indian codes appear later, primarily after 2015, reflecting a gradual institutionalisation of sustainability principles into formal building regulations and standards. After 2020, the Indian policy, which emphasises the growing role of buildings as key levers for sustainability, decarbonisation, and climate resilience, while broader infrastructure-scale integration remains comparatively limited.

Structural Properties of the Indicator Network

The indicator network was developed by representing each sustainability indicator as a node, with edges indicating co-occurrence across the reviewed frameworks. This strategy enabled systematic mapping of indicator interdependence and redundancy, revealing underlying thematic groupings and highly influential individual indicators. The network was constructed using 387 quantitative indicators identified in phase 2. The indicator network comprises 387 nodes, representing all identified sustainability and climate-resilient indicators. The network exhibits limited connectivity, with an average degree of 1.622, indicating that most indicators are connected to only one or two others. The graph density is very low (0.3%), indicating that only a small fraction of possible connections exists. The diameter of 8 and the average path length of 2.554 indicate efficient local connectivity within the clusters. Multiple connected components exhibited siloed groupings, indicating incomplete integration across the frameworks.

Community Detection and Framework Fragmentation

To understand the internal structure of the indicator network and identify key indicator clusters, community detection was performed using a modularity-based algorithm as discussed in Phase 3. This method enables the identification of clusters of indicators that are more densely connected to each other than to the rest of the network.

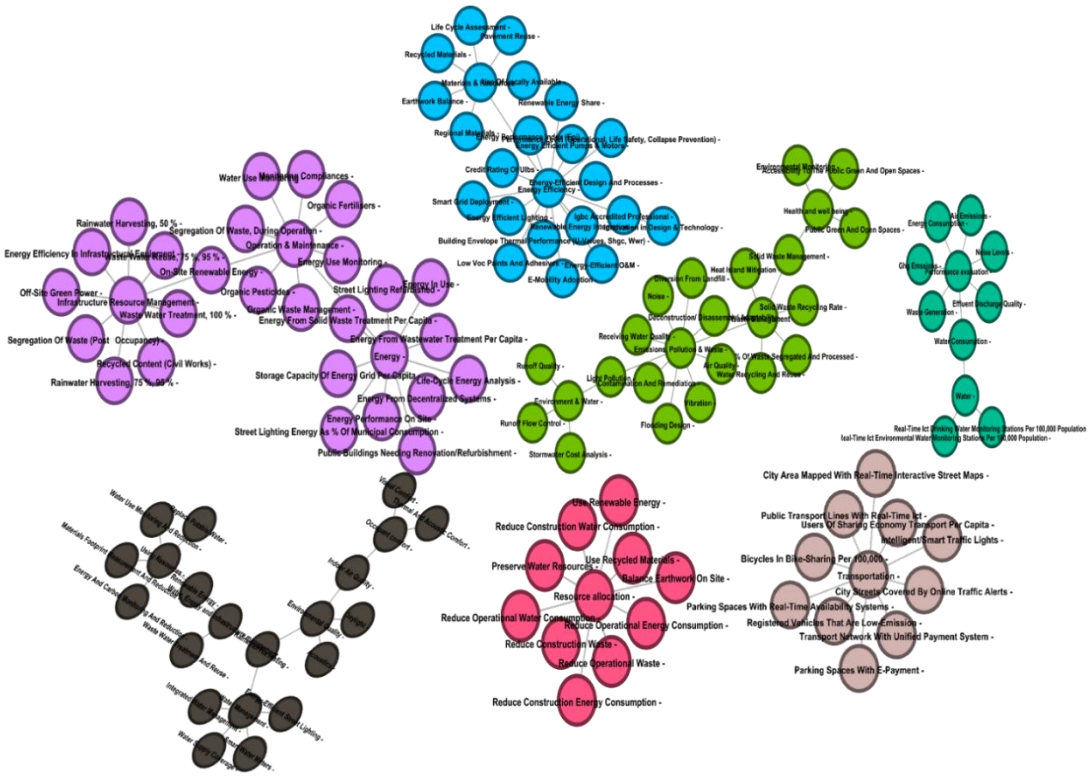


Figure 3: Community structure of the sustainability indicator network showing modularity-based clusters.

The analysis resulted in a high modularity value of 0.974, indicating a strong community structure within the network. A total of 113 distinct communities were identified, suggesting that the indicator network is highly fragmented into multiple, relatively independent groups. This high modularity value indicates strong clustering and limited interaction between groups, confirming a fragmented indicator structure. This highlights the nature of the existing sustainability assessment framework, given the large number of detected communities. Rather than forming an integrated, cohesive assessment system, the indicators tend to cluster by specific thematic domains, sectoral focuses, or framework origins. This implies that current frameworks do not adequately address interdependence among environmental, social, and economic dimensions.

Figure 3 illustrates the community structure of the indicator network, where distinct clusters are visible as separate groupings of indicators. Different groups of indicators were identified by modularity-based community detection, with each color representing a distinct cluster. Indicators in the same color category are more closely related than indicators in other groups.

Identification of Cross-Cutting Indicators

To address the fragmented community structure identified in the indicator network, cross-cutting indicators were identified using betweenness centrality as the primary metric. Based on this analysis, the first 15 indicators were selected to illustrate cross-cutting aspects from the total set of 387 quantitative indicators. These indicators exhibit the highest betweenness centrality values and span multiple modularity classes, confirming their role in bridging distinct indicator communities. Closeness centrality was examined as a supporting measure to contextualise their overall connectivity within the network.

Table 2 presents the identified cross-cutting indicators and their associated network metrics. These indicators provide an empirical basis for integrating fragmented sustainability frameworks and for developing a structured, hierarchical conceptual framework for public infrastructure. These indicators served as key inputs to the development of the conceptual framework presented in the subsequent section.

Table 2: Cross-cutting sustainability indicators identified using network analysis.

ID	Modularity Class	Closeness centrality	Betweenness centrality
Operation & Maintenance	64	0.5090909	278
Energy	64	0.4516129	225
Energy Efficiency	18	0.7586207	209
Emissions, Pollution & Waste	13	0.5	198
Infrastructure Resource Management	64	0.3733333	188
On-Site Renewable Energy	64	0.4375	171
Waste Management	13	0.46	152
Rainwater Harvesting	12	0.3921569	131
Materials & Resources	18	0.5789474	111
Water, Energy, and Infrastructure Management	12	0.3571429	97
Environmental Quality	12	0.3448276	93
Light Pollution	13	0.3833333	76
Solid Waste Management	13	0.359375	76
Renewable Energy	12	0.3076923	75
Using Resources	12	0.2631579	70

Conceptual Framework for Sustainability and Climate-Resilient Infrastructure Assessment

Based on the results of the network analysis, a conceptual framework was developed to support the assessment of sustainability and resilience in public infrastructure in the Indian context.

From the Modularity-based analysis, distinct thematic clusters were identified, and indicators were grouped based on their thematic similarity. These were further examined to identify key indicators using centrality measures. Based on their structural importance, relevance to the Indian context, and to avoid redundancy. A refined subset of 56 indicators was identified from the clustered network based on centrality analysis, thematic relevance, and contextual suitability. The framework is structured into three levels: indicators, criteria, and requirements, to enable systematic aggregation, comparison, and decision support. The framework is designed to support users in selecting indicators based on project-specific priorities by focusing on the refined subset of 56 indicators, which represent the most structurally influential and cross-cutting indicators identified through centrality analysis. Users can further tailor this subset by selecting relevant thematic clusters aligned with project goals and contextual requirements. *Figure 4* represents the conceptual framework.

By combining fragmented indicator groupings and prioritising system-level, cross-cutting indicators, the framework creates a consistent structure that can be applied to various public infrastructure sectors and

decision-making contexts. At the highest level, criteria are consolidated into five broad requirements: (R1) Management and Governance, (R2) Social Well-being and Inclusivity, (R3) Environment and Resource Efficiency, (R4) Energy and Infrastructure, and (R5) Climate Resilience. These requirements serve as modular assessment dimensions that represent strategic sustainability and resilience goals.

At the lowest level, indicators represent measurable variables that capture the institutional, social, environmental, infrastructural, and resilience-related attributes of infrastructure systems.

These indicators are organised into criteria (C), which serve as intermediary aggregation units that represent distinct functional categories. Criterion-level aggregation allows normalisation and consolidation of heterogeneous indicators while preserving thematic coherence. The matrix-based layout of the framework facilitates cross-dimensional evaluation, allowing indicators under a single requirement to contribute to multiple criteria and enabling interlinkages between sustainability and resilience dimensions. Overall, the framework provides a transparent medium for evaluating, summarising, and benchmarking infrastructure projects. Unlike conventional frameworks that present static and often independent indicator lists, the proposed approach explicitly captures interdependencies among indicators, enabling the identification of cross-cutting indicators that enhance integration, reduce redundancy, and improve the coherence of sustainability assessment. It is intended to support evidence-based decision-making, policy formulation, and the prioritisation of interventions to enhance sustainability and climate resilience in public infrastructure systems. The framework enables users to systematically select and prioritise indicators based on their structural importance, thematic relevance, and alignment with project-specific requirements.

Implications for Practice and Research

Existing sustainability assessment frameworks are often fragmented, making decision-making complex and inconsistent across projects. This study provides a structured approach to identify and prioritise indicators using network-based metrics such as centrality, enabling practitioners, including project managers, planners, and policymakers, to focus on the most influential and cross-cutting indicators. The proposed three-level hierarchical framework facilitates systematic selection, aggregation, and comparison of indicators aligned with project-specific requirements, thereby supporting more transparent, consistent, and evidence-based decision-making in public infrastructure planning and evaluation.

From a research perspective, this study demonstrates the applicability of Social Network Analysis as a replicable, data-driven approach for examining interdependencies among sustainability indicators. The framework establishes a foundation for future work on indicator weighting, validation, and integration with multi-criteria decision-making methods, while also opening avenues for extending the approach to different infrastructure contexts and geographic settings. Overall, the study bridges the gap between indicator identification and practical application by offering a structured, scalable approach to sustainability assessment in infrastructure projects.

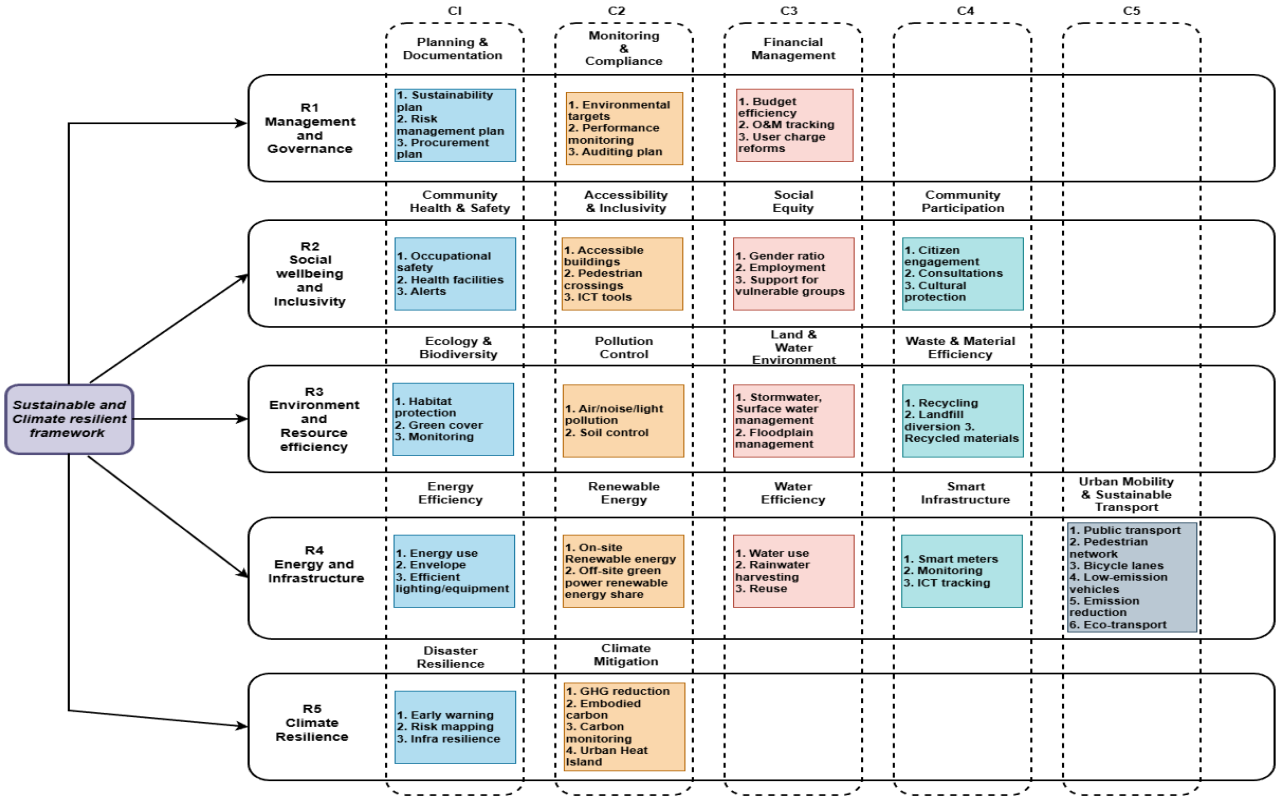


Figure 4: Conceptual sustainability and climate-resilience assessment framework illustrating the hierarchical organisation of indicator-level inputs into criteria and requirement-based modules for integrated evaluation of public infrastructure systems

Limitations and Future Research

On the one hand, this work proposes a conceptual framework for classifying sustainability indicators for public infrastructure. However, the framework is not validated through stakeholder engagement or real-world case applications, and no indicator weighting is performed in this study.

Additional research is required to validate and improve the suggested framework through expert consultation and stakeholder participation. Multi-criteria decision-making methods, such as the Analytic Hierarchy Process, can be employed to assign indicator weights and operationalise the framework. Furthermore, the framework could be extended to a representative set of case studies on public infrastructure projects in India to assess its practical usability and robustness. Periodic amendments to the indicator network should also be made to account for changes in sustainability priorities and climate resilience considerations.

Conclusions

This research develops a conceptual decision framework for sustainable and resilient public infrastructure in India by examining the structure and integration of existing sustainability indicators. A systematic review of global and Indian frameworks, codes, standards, and policy documents identified 629 indicators, which were classified as quantitative and qualitative. The relationships among the quantitative indicators were analysed using Social Network Analysis. The results reveal a highly sparse and fragmented indicator network with very high modularity of 0.974, indicating strong internal categorisation but weak cross-dimensional integration. This structural fragmentation points to limitations in prevailing sustainability frameworks, particularly in their support for integrated assessment of public infrastructure.

To address this limitation, the study identifies a focused set of cross-cutting indicators using centrality-based measures. These indicators capture key interconnections across multiple sustainability dimensions and provide a basis for improving coherence and integration. Based on the modularity analysis, indicators were grouped into different thematic clusters, and a refined subset of 56 indicators was identified from these clusters using centrality measures. Building on this insight, a conceptual decision framework is proposed that reorganises disjointed indicator sets into a context-sensitive structure suited to Indian public infrastructure. The framework supports transparent multi-criteria decision-making and provides a methodological and practical foundation for developing integrated, consistent, and climate-resilient infrastructure assessment and rating systems.

This is an ongoing study, and future work will focus on empirical testing, stakeholder engagement, and operational refinement of the proposed framework. The performance of the proposed framework can be benchmarked against existing frameworks through comparative case studies, evaluating criteria such as reduction in indicator redundancy, consistency in decision outcomes, and the ability to capture cross-dimensional linkages in sustainability assessment. The proposed framework benefits policymakers, infrastructure agencies, planners, asset owners, and researchers by providing a transparent, integrated basis for sustainability- and resilience-informed decision-making and assessment of public infrastructure.

Acknowledgment

The authors gratefully acknowledge the financial support provided by NBCC (India) Limited for this research.

References

BEE (2024) *Energy Conservation and Sustainable Building Code 2024*. Available at: https://www.beeindia.gov.in.

BIS (2006) *IS/ISO 14044 Environmental Management-Life Cycle Assessment-Requirements and Guidelines*. Available at: https://www.iso.org/standard/38498.html (Accessed: 28 January 2026).

BIS (2016) *National building code of India, Volume 2*. Available at: https://archive.org/details/nationalbuilding02/in.gov.nbc.2016.vol2.digital/ (Accessed: 28 January 2026).

Boix-Cots, D., Pardo-Bosch, F. and Pujadas, P. (2023) 'Analysis and Comparison of the Infrastructure Report Cards as a Decision Making Tool for Sustainable Development', *Buildings*, 13(9). doi: 10.3390/buildings13092166.

Cardin, M. A., Mijic, A. and Whyte, J. (2023) 'Data-driven infrastructure systems design for uncertainty, sustainability, and resilience', in *Life-Cycle of Structures and Infrastructure Systems - Proceedings of the 8th International Symposium on Life-Cycle Civil Engineering, IALCCE 2023*. CRC Press/Balkema, pp. 2565–2572. doi: 10.1201/9781003323020-312.

Das, D. K. (2025) 'Resilient and Sustainable Urban Infrastructure in the Global South: Strategies to Address Climate Change-Linked Vulnerabilities', *Urban Planning*, 10. doi: 10.17645/up.10037.

Diaz-Sarachaga, J. M. *et al.* (2016) 'Evaluation of existing sustainable infrastructure rating systems for their application in developing countries', *Ecological Indicators*, 71, pp. 491–502. doi: 10.1016/j.ecolind.2016.07.033.

Ferrari, M. (2020) 'Reflexive governance for infrastructure resilience and sustainability', *Sustainability (Switzerland)*, 12(23), pp. 1–8. doi: 10.3390/su122310224.

GIB (2018) *The Standard for Sustainable and Resilient Infrastructure*. Available at: https://sure-standard.org/wp-content/uploads/2019/10/ST01_Normative_Standard_v1.1_clean.pdf (Accessed: 28 January 2026).

ICE (2007) *CEEQUAL Manual V3.1*. Available at: www.ceequal.com (Accessed: 28 January 2026).

IGBC (2010) *IGBC Green Townships*. Available at: https://igbc.in/frontend-assets/html_pdfs/abridged/IGBC%20Green%20Townships%20-%20Abridged%20Reference%20Guide%20(Pilot%20Version).pdf (Accessed: 28 January 2026).

IGBC (2013) 'IGBC Green Landscape'. Available at: https://igbc.in/frontend-assets/html_pdfs/abridged/IGBC%20Green%20Landscape%20-%20Abridged%20Reference%20Guide%20(Pilot%20Version).pdf (Accessed: 28 January 2026).

IGBC (2015) *IGBC Green Cities*. Available at: https://www.inidesignstudio.com/wp-content/uploads/2020/11/IGBC_Green_Cities.pdf (Accessed: 28 January 2026).

IS (2016) *Criteria for Earthquake Resistant Design of Structures Part 1*. Available at: https://law.resource.org/pub/in/bis/S03/is.1893.1.2002.pdf (Accessed: 28 January 2026).

ISCA (2018) *IS Technical Manual, V-1.2*. Available at: https://d3n8a8pro7vhmx.cloudfront.net/themes/5a72941f5ee54d4c43000000/attachments/original/1543788661/ISTechnicalManual_v1.2_Nov18.pdf?1543788661 (Accessed: 28 January 2026).

ISI (2018) *ENVISION Sustainable infrastructure framework V3*. Available at: www.sustainableinfrastructure.org.

ISO (2011) *ISO 50001 Energy management systems*. Available at: https://www.iso.org/files/live/sites/isoorg/files/store/en/PUB100400.pdf (Accessed: 28 January 2026).

ISO/TC (2012) *ISO/CD 37122 Sustainable cities and communities*. Available at: https://www.iso.org/standard/69050.html (Accessed: 28 January 2026).

Jain, A. and Ananda, B. K. (2024) *Examination of a traditional residential development in a Tier-II urban area in India according to GRIHA standards to advance sustainable infrastructure*, *International Journal of Environmental Pollution and Environmental Modelling*.

Khan, K. and Henschel, T. (2024) 'LCT-Based Framework for the Assessment of Sustainability: From the Perspective of Literature Review', *Social Indicators Research*, 175(3), pp. 1–20. doi: 10.1007/s11205-024-03333-8.

Kumar, S., Santara, A. and Banerji, H. (2020) 'Pentagram sustainability model framing of a holistic sustainability dimension model for assessment of Neighbourhood-Level Urban Communities (NLUCs) using a semi-automated systematic literature review', *International Review for Spatial Planning and Sustainable Development*, 8(2), pp. 100–117. doi: 10.14246/irspsda.8.2_100.

Leal, O. J. U. *et al.* (2024) 'A Systematic Review of Integrated Frameworks for Resilience and Sustainability Assessments for Critical Infrastructures', *Structural Engineering International*, 34(2), pp. 266–280. doi: 10.1080/10168664.2023.2265965.

Liu, Y. *et al.* (2018) 'Sustainable Infrastructure Design Framework through Integration of Rating Systems and Building Information Modeling', *Advances in Civil Engineering*, 2018. doi: 10.1155/2018/8183536.

Mansell, P. *et al.* (2019) 'Assessing the impact of infrastructure projects on global sustainable development goals', *Proceedings of the Institution of Civil Engineers: Engineering Sustainability*, 173(4), pp. 196–212. doi: 10.1680/jensu.19.00044.

Mehraban, R. A. *et al.* (2025) 'Comprehensive Analysis of Sustainability Rating Systems for Road Infrastructure', *Infrastructures*. Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/infrastructures10010017.

MOHUA (2021) *AMRUT 2.0 Making Cities Water Secure Operational Guidelines*. Available at: https://amrut.mohua.gov.in/ (Accessed: 28 January 2026).

MOHUA (2022) *Climate Smart Cities Assessment Framework 3.0*. Available at: https://niua.in/c-cube/sites/all/themes/zap/assets/pdf/CSCAF_3_0_Technical_document.pdf (Accessed: 28 January 2026).

Rousseau, R. (2002) *Social network analysis: a powerful strategy, also for the information sciences*. doi: https://doi.org/10.1177/016555150202800601.

Sanad, H. *et al.* (2025) 'Optimizing sustainability in infrastructure projects: a framework integrating time, economic, social and environmental perspectives', *Construction Innovation*. doi: 10.1108/CI-12-2024-0398.

Sánchez-Silva, M. and Baker, J. W. (2024) 'Dynamic Infrastructure Systems: advancing sustainable urbanization and climate change', *Environment Systems and Decisions*, 44(3), pp. 489–499. doi: 10.1007/s10669-024-09969-8.

TERI (2010) *GRIHA- Introduction to National Rating System, V1*. Available at: https://www.grihaindia.org/ (Accessed: 28 January 2026).

University of Washington (2011) *Greenroads, manual, V1.5*. Available at: https://www.uni-groupusa.org/ (Accessed: 28 January 2026).

Wang, H., Zhang, X. and Lu, W. (2018) 'Improving Social Sustainability in Construction: Conceptual Framework Based on Social Network Analysis', *Journal of Management in Engineering*, 34(6). doi: 10.1061/(asce)me.1943-5479.0000607.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

FROM INSPECTION TO DIGITAL PASSPORT: TRACING THE PROVENANCE OF TESTING DATA OF CONCRETE COMPONENTS FOR CIRCULAR REUSE

Lukas Guntermann[1], Nina Krautgartner[1], Hendrik Zielonka[2], Philipp Hagedorn[1], Peter Mark[2], and Markus König[1]
[1]Chair of Computing in Engineering, Ruhr University Bochum, Bochum, Germany
[2]Institute of Concrete Structures, Ruhr University Bochum, Bochum, Germany

Abstract

Circular reuse of construction components, prioritized by the EU over recycling, is limited by the lack of legally reliable, element-specific information in BIM environments. This work proposes a literature-based parameter set for the reuse of reinforced concrete elements as part of a product passport. Data from legally signed inspection certificates are integrated with digital BIM representations using established ontologies (BOT, Metadata4ing, ISOProps). Information is extracted from signed PDF certificates and linked to IFC models. Feasibility is demonstrated through a case study, enabling an interoperable, auditable workflow supporting End-of-Life-to-reuse information continuity for reinforced concrete elements.

Introduction

To reduce the negative environmental impact of the construction industry, the reuse of structural components is increasingly important regarding waste reduction (Sakthibala et al., 2025) and is assigned a higher priority in the waste hierarchy of the European Commission (2008). Reinforced concrete is particularly relevant in this context, as its high mass, significant greenhouse gas emissions, and long service life offer considerable potential to reduce environmental impacts through targeted reuse. Within the Circular Economy (CE), when planning with components that are reused at their End-of-Life (EoL), it is necessary to capture and store component- and reuse-specific information, e. g. from Pre-Demolitions Audits (PDA) according to DIN SPEC 91484. Information requirements for circular components have already been documented in general terms (cf. Akbarieh et al. (2023); Al-Qazzaz (2024)). However, for solid construction and structural components, additional information on the condition of components from detailed inspections and reuse-oriented component testing is necessary, since these typically need to be verified prior to reinstallation in order to determine structural properties such as load-bearing capacity.

For a Building Information Modeling (BIM)-supported reuse workflow, inspection data must be linked with Industry Foundation Classes (IFC) in a machine-readable form (Hagedorn et al., 2025), while the signed inspection certificates for the components must remain available for legal reasons so that it can be determined with certainty which properties can be assumed in subsequent structural planning. These inspection certificates ensure liability for the structural properties of the components for reuse, and are already being used for the reuse of steel components, as defined in EN 10204:2004 and EN 10168:2004. However, the reuse of reinforced concrete (RC) involves other product tests for which no reuse-specific material parameters or inspections have yet been established. This necessary flow of information leads to the following Research Question (RQ).

RQ: *How can information be extracted from PDF-based inspection certificates for RC elements and reliably transferred to the BIM model, ensuring a continuous flow of information in the components lifecycle?*

To answer the RQ, the remainder of the paper is structured as follows. Related work on reuse, product passports, ontologies, and information extraction is gathered. A methodology and workflow is proposed and implemented. Eventually, the workflow is demonstrated and validated before discussing the results and drawing conclusions.

Related work

Reuse of reinforced concrete (RC)

Küpfer and Fivet (2024) examine the structural reuse of concrete pieces, a relatively new practice evidenced by scarce yet diverse built projects. They trace the slow evolution of precast reuse since the 1980s and the recent emergence of cast-in-place reuse, driven by sustainability and design exploration, emphasize reliance on local knowledge and better case documentation, and advocate viewing soon-to-be-demolished structures as material stock. Mark et al. (2025) describe modular reuse of existing structures and focus on component testing in order to overcome usage restrictions and building regulations. Rose et al. (2025) investigate efficient reuse of RC using optimization algorithms and highlight the relevance of load-bearing capacity. Consequently, M–N (Moment-Normal force) charts may be pertinent for future structural assessment of reused components. Krautgartner et al. (2025) propose a network-based approach that integrates GIS (Geographic Information Systems) to manage building deconstruction and match demand for reuse elements. They derive requirements for system components, parameters, and process flows from the literature and formalize them.

Product passports in construction

Aftab and Agliata (2026) review the state of Material Pass-

ports (MPs) and discuss them generically, without addressing the complexity of composite, load-bearing materials such as RC. Their study characterizes BIM primarily as a dynamic repository, supporting the premise that BIM models serve as the starting point for Linked Data to consolidate additional information for reuse.

MPs address material-specific documentation, often in recycling, whereas Digital Product Passports (DPP) cover whole construction products; both rely on similar BIM- and Linked Data–based methods and can cover different scales from building materials to overall assessments. Ruismäki et al. (2025) examine stakeholder perspectives on DPPs for construction products and identify traceability, BIM integration, and resource efficiency as key benefits. Interviews show broad agreement on including EoL data and usage data (e.g., recorded damages). Stakeholders prioritized steel structures in a phased implementation owing to ease of handling, reuse, and major environmental benefits; concrete products were also considered suitable, but onsite modifications complicate information management. Ruismäki et al. (2025) further emphasize the complexity of concrete, which has contributed to the prevailing focus on steel. Senarathne et al. (2025) propose a concrete-specific DPP framework that includes physical properties and parameters across logistics, chemical, biological, cost, safety and health, among others. As it has not yet been validated in real-world settings, this paper adopts a smaller example focused on reuse-specific data instead. Vahidi et al. (2024) focus on recycled concrete and Radio Frequency Identification (RFID)-based MPs. As their work addresses crushed recycled concrete rather than component reuse, its relevance here is limited; however, RFID remains a potential option for tracking reused components. Bernier and Danash (2024) propose DPP based on ontologies. Blomqvist et al. (2023) propose a network of core ontologies for the CE and include process-, actor-, material- and product-information.

Ontologies for circular construction components

Ontologies and Knowledge Graphs (KG) enable machine-interpretable information for the built environment, typically represented in the Resource Description Framework (RDF) and the Web Ontology Language (OWL). The definition of terminology using ontologies is referred to as the TBox, while the instantiation and assertion of facts according to these ontologies is referred to as the ABox. (Pauwels et al., 2022, 2018)

Moreover, ontologies can be categorized as top-level ontologies for providing abstract generic concepts and domain ontologies for representing domain knowledge. Beyond domain-specific efforts, Li et al. (2023) investigate cross-industry general ontologies for the circular economy to create an ontology network that links core ontologies across multiple industry domains. To facilitate ontology discovery and reuse in AECO (Architecture, Engineering, Construction, and Operations), Akbarieh et al. (2025) developed the Built Environment Ontology Lookup Service (BE-OLS) (EC3 M&S Committee, 2025). Within the built environment, examples of domain ontologies include the Decommissioning and Reuse Ontology (DOR); for EoL decision-making, the Construction Circular Product Ontology (CCPO) provides a rule-based EoL product ontology that considers Product Health State, recyclability, and reference service life (Adu-Duodu et al., 2025). Additional ontologies relevant to DPP workflows include the Material Passport Ontology (MPO), the Steel Element Reuse Ontology (SERO), the Building Element Ontology (BEO) for classification of building elements, the Building Product Ontology (BPO), and a Concrete Damage Ontology (CDO). For data structuring and property description in the BIM-based information management, Zentgraf et al. (2026) developed the ISO 23386 Property Ontology (ISOProps), providing modular and reusable descriptions for properties and construction objects. Liu et al. (2021) introduced the Building Concrete Monitoring Ontology (BCOM) for maintaining testing data of cast-in-place fresh concrete properties on the construction site. However, their ontology is not a generic, extendable property dictionary but a strict, encoded TBox ontology with a fixed schema. In the context of managing arbitrary test data from the broader domain of engineering sciences, the Metadata4Ing ontology (Lanza et al., 2025) is relevant. It provides a framework for the semantic description of research data and the entire data-generation process, covering the object of investigation, sample and data manipulation methods and tools, the data files, and the roles of persons and institutions. Its structure and application follow modularity and inheritance, with the processing step as a central element. Ultimately, ontologies also serve information management by structuring extracted data from reports (Zentgraf and Zimmermann, 2025). In summary, existing ontologies already model circular processes and data structuring in AECO but do not address testing-level details for concrete.

Methodology

To answer the RQ proposed, the methodology is defined as shown in Fig. 1. First, relevant background and literature is gathered and analyzed regarding relevant component parameters for the reuse. Next, a methodology and concept are developed on how these parameters can be extracted from signed PDF certificates and be integrated with relevant metadata of the testing procedures for documenting the provenance. As a result, an ABox dictionary of properties assignable to concrete elements based on inspection tasks is developed, reflecting the parameters defined in the literature. This ABox dictionary is based on terminology from multiple ontologies, such as the Building Topology Ontology (BOT) for IFC model alignment (Rasmussen et al., 2021), Metadata4ing for test procedure metadata and material properties (Lanza et al., 2025), and ISOProps for maintaining the tested properties in accordance with ISO 23386 (Zentgraf et al., 2026). The proposed concept is implemented and demonstrated on a case study of a rein-

forced concrete element; results are discussed and conclusions drawn.

Component reuse data flow

Figure 2 outlines the workflow and data concept for component reuse. The process starts with deconstruction; components are tested in the laboratory, and a certificate is issued that must be handwritten-signed for legal accountability. This signature confirms that testing was performed, liability is assumed, and the described properties can be used for future planning in case of reuse. Information from the certificate is extracted and stored in a BIM model. The PDF inspection certificate is linked to the component's IFC model, ensuring that, in subsequent planning, information is available in two forms: within the IFC model for BIM-based design and as a linked PDF report that can be accessed at any time to provide planners with assurance for component reuse. The data linking the component and the inspection results can thus be transferred to the next life cycle of the component.

Inspection certificate for reuse of RC

Properties for a product passport for reinforced concrete are proposed, derived from the literature: Senarathne et al. (2025) present a DPP for concrete; Mecka et al. (2025) outline a framework for the reuse of reinforced concrete components; Ottosen et al. (2024) review non-destructive testing to document properties for reuse; Suchorzewski et al. (2023) propose a classification system for disassembled concrete elements (DCE). Relevant properties were compiled from these sources. The proposed inspection certificate can form part of a DPP and focuses on properties determined through laboratory testing and inspection of the elements. No claim to completeness is made; the proposal serves to demonstrate and model implementation and data flow in this paper. Table 1 lists and explains required parameters, and specifies the investigation methods by which they can be determined. These investigation methods are central to laboratory test reports. Section 1 of the table provides general information; Sections 2 and 3 present mechanical properties, geometry, damage, and durability.

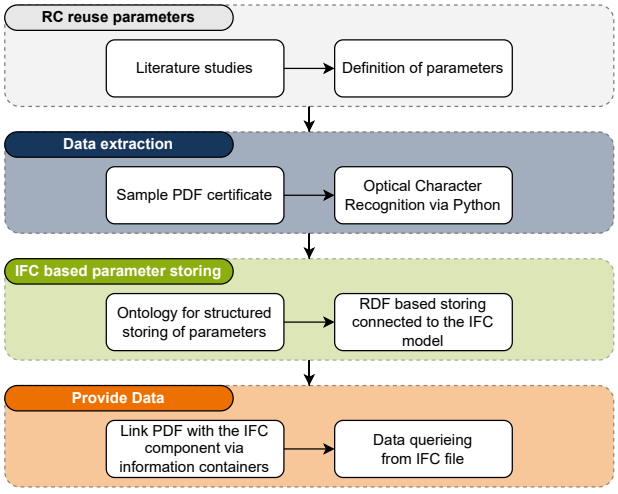


Figure 1: Methodology

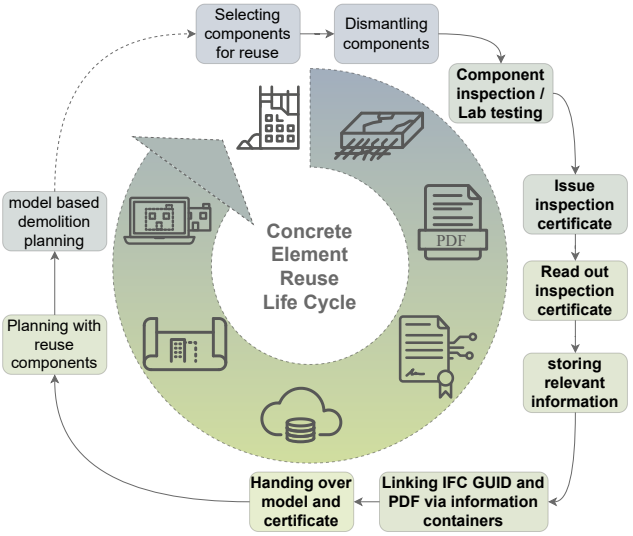


Figure 2: Data Concept for reuse components

Implementation

After having analyzed the literature and provided a template and parameter set for inspection reports according to the methodology in Figure 1, in this section, data extraction is performed. For the first testing of the procedure, a Python library for Optical Character Recognition (OCR) was used. The extracted parameters are ingested into a KG using the Concrete Element Reuse Ontology (CERO) published at `https://w3id.org/cero#` in previous work (Krautgartner et al., 2026). CERO provides standardized ABox properties as `owl:DatatypeProperty` instances annotated with standardized metadata according to the Metadata4Ing ontology and maintenance workflows for the property dictionary according to the ISOProps ontology and ISO 23386. Property values are assigned to concrete components represented in RDF graphs after being converted from the IFC model. Eventually, the extracted data is provisioned into Information Containers for Linked Document Delivery (ICDD) (ISO 21597-1, 2020) and linked to both the PDF and the IFC model to provide the provenance of the properties. Using the ICDD platform as it has been presented in Hagedorn et al. (2023), this interlinked data can be queried from the model and the PDF file of the inspection certificate.

Ontology for structuring the parameters

To embed Table 1 into the Linked Data concept, it must be converted into processable entities following a meaningful structure. Because BE-OLS does not contain an ontology for RC component inspection certificates (as of January 2026, (EC3 M&S Committee, 2025)) yet, a custom implementation of the parameter set into the CERO ontology was necessary. The namespaces used for the implementation are presented in Table 2.

This dataset inside CERO maps the entirety of the proposed information of the test certificates according

to Table 1. The parameters from Section 1 are directly instantiated and persisted into the CERO ontology. The parameters from Sections 2 and 3 are also created as parameters in the CERO ontology, but are recorded via the Metadata4Ing ontology. This means that each test method (see Table 1) has been created there as `metadata4ing:Tool`. In addition, the possible test configurations (`metadata4ing:Configuration`) have been stored, which contain, which parameters can be recorded with which test methods. According to the Metadata4Ing Ontology, for each row, `metadata4ing:Configurations` were initially created to link instances of `metadata4ing:Methods` (e.g., `certificate:CompressiveStrengtTest`) with the testing tools (`metadata4ing:Tools`) (e.g., `certificate:Reboundhammer`) as shown in Figure 3. Numerical parameter Values are stored in the `metadata4ing:ProcessingSteps` in `metadata4ing:NumericalAssignments` in a `metadata4ing:NumericalVariable`. The `metadata4ing` Ontology uses the `qudt` ontology for storing units (as here MPa for the compressive strength).

Table 1: Proposed testing parameters for inspection certificates of reused reinforced concrete components

Section 1: General component and project information		
Parameter	***Description***	
Identification	Identification number/ID/serial number	
Origin	Original building, year of construction, pre-use	
Element type	Type of element (beam, column, slab, etc.))	
Overall dimensions	External geometry of the element, span after processing	
Load capacity	Reduced cross-sectional load-bearing capacity after taking into account all mitigating factors, e.g. via M-N interaction diagram	
Environmental actions	Relevant exposure classes	
Environmental impacts	Environmental data on the original element, on processing the elements after use, and on the cost of dismantling and transport	
Section 2: Condition assessment		
2.1 Mechanical properties		
Parameters	***Description and units, if applicable***	***Investigation method***
Concrete compressive strength	Characteristic compressive strength (fck) [MPa; N/mm²]	Core extraction; non-destructive testing (e.g. rebound hammer); as-built documentation
Concrete stiffness	Modulus of elasticity [MPa; N/mm²]	Core extraction; non-destructive testing (e.g. ultrasonic testing); as-built documentation
Reinforcing steel tensile strength	Yield strength (fyk) [MPa; N/mm²]; modulus of elasticity [MPa; N/mm²]	Reinforcement sampling with tensile test; as-built documentation
2.2 Geometry and damage		
Concrete cover	Minimum and mean concrete cover [mm]	Ground penetrating radar (GPR); as-built documentation
Reinforcement layout	Location, diameter and number of reinforcement bars; effective depth	Reinforcement scanner; covermeter; ground penetrating radar (GPR); as-built documentation; partial exposure of reinforcement
Cracks	Localisation and maximum crack width [mm]	Visual inspection; laser scanning; digital imaging
Internal defects	Presence of delamination, voids or flaws from loading history	CT scans; X-ray tomography
Section 3: Durability and remaining service life		
Carbonation depth	Measured depth compared to existing concrete cover	Phenolphthalein Indicator, Drill dust analysis
Chloride content	Concentration at reinforcement level [% by mass of cement]	Staggered drilling dust analysis
Remaining service life	[a]	Two-phase model according to Tuutti
Classification	According to DCE classification (Suchorzewski et al., 2023)	

Table 2: List of namespaces used in this study

Prefix	**Namespace IRI**
Existing namespaces used:	
rdf	`http://www.w3.org/1999/02/22-rdf-syntax-ns#`
rdfs	`http://www.w3.org/2000/01/rdf-schema#`
xsd	`http://www.w3.org/2001/XMLSchema#`
sh	`http://www.w3.org/ns/shacl#`
bot	`https://w3id.org/bot#`
props	`https://w3id.org/props#`
isoprops	`https://w3id.org/isoprops`
metadata4ing	`http://w3id.org/nfdi4ing/metadata4ing/1.4.0/`
cero	`https://w3id.org/cero#`
Defined namespaces:	
component	`https://example.org/component#`
certificate	`https://example.org/certificate#`

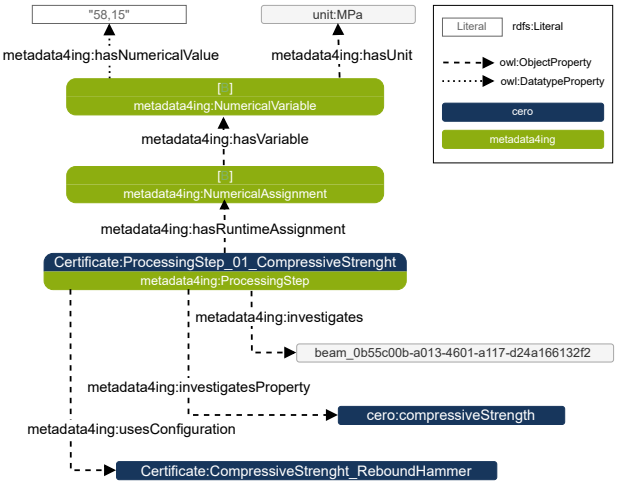


Figure 3: Property for concrete compressive strength

Information extraction of inspection certificates

The information extraction and ontology population for this workflow are inspired by existing research approaches. Sithisint et al. (2025) present engineering drawing interpretation, employing bounding box detection via YOLOv11n and text extraction via OCR with EasyOCR. Schönfelder and König (2025) provide an ontology based approach to structure extracted information from floor plans. Vaatz et al. (2023) show how to link semantic information to BIM using ontologies and ICDD.

In this article, the information extraction from the sample inspection certificates is done by a Python script and the PyTorch library EasyOCR[1]. To do this, the parameters were converted into different spellings in regex pattern, which are then used to search the imported certificates. The following values are then transferred into a template file according to the CERO parameter set.

[1] `https://github.com/JaidedAI/EasyOCR`

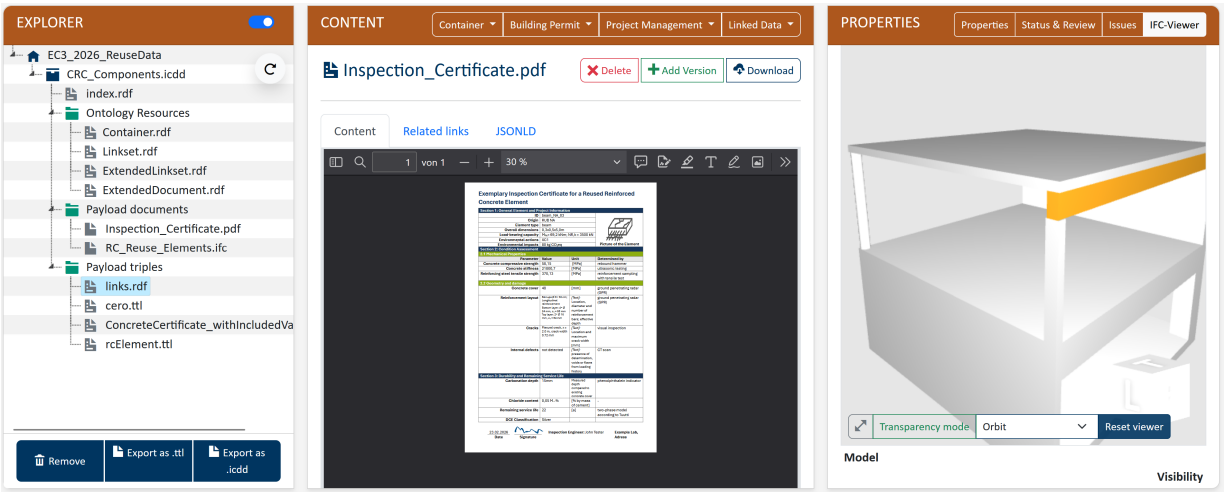


Figure 4: Link of PDF inspection certificate (middle) and RC reuse component (right)

Link of inspection certificate and IFC entity via ICDD

ICDD containers organize and link heterogeneous project data on a dedicated platform. Through Linked Data, documents and files can be associated with a project and with specific IFC elements in the central model. RDF/OWL semantics enable machine-readable linkage and integration with standardized BIM data (e.g., IFC). The ICDD platform has been used for a ReSource Marketplace, where individual components are offered and presented together with their IFC-based data (Hagedorn et al., 2025).

Consequently, the platform's Linked Data concept is suitable for establishing a connection between the PDF inspection certificate and the RC Reuse Element in the BIM model via the IFC GUID, as shown in Figure 4.

Validation and discussion

To validate the concept, an IFC4 model of a solid construction support structure consisting of reinforced concrete components was used (see Figure 4). For this purpose, a sample inspection certificate with realistic values for an RC beam was created, containing all parameters proposed in Table 1 (see Figure 5). In addition, the test engineer's data is provided as an example. The information extraction of this sample inspection certificate was carried out as described, whereby it became apparent that the OCR algorithm has shortcomings with multi-line table entries. The extracted data is stored in the `certificate` KG template using Python by manipulating the text in the Turtle file of the KG. The ontology, the IFC Model and its RDF representation is then uploaded to the ICDD platform (see Figure 4).

To ensure the fundamental functioning of the ontology, Competency Questions (CQs) are formulated according to Bezerra et al. (2013) and checked using SPARQL (SPARQL Protocol and RDF Query Language) queries within the ICDD platform. The aim is to be able to retrieve all data contained in the PDF inspection certificate in the IFC model as well. Due to the different use cases of the data sections, two CQs are checked and the corresponding SPARQL queries are formulated (see Listings 1 and 2).

Exemplary Inspection Certificate for a Reused Reinforced Concrete Element

Section 1: General Element and Project Information			
ID	beam_NA_03		Picture of the Element
Origin	RUB NA		
Element type	beam		
Overall dimensions	0,3x0,5x5,0m		
Load-bearing capacity	M_{Rk}= 69,2 kNm; NR,k = 3500 kN		
Environmental actions	XC1		
Environmental impacts	80 kg CO_2eq		
Section 2: Condition Assessment			
2.1 Mechanical Properties			
Parameter	**Value**	**Unit**	**Determined by**
Concrete compressive strength	58,15	[MPa]	rebound hammer
Concrete stiffness	21800,7	[MPa]	ultrasonic testing
Reinforcing steel tensile strength	370,13	[MPa]	reinforcement sampling with tensile test
2.2 Geometry and damage			
Concrete cover	48	[mm]	ground penetrating radar (GPR)
Reinforcement layout	Stirrups Ø 8 / 30 cm; Longitudinal reinforcement Bottom layer: 4× Ø 24 mm, c_v = 68 mm Top layer: 2× Ø 16 mm, c_v = 64 mm	*[Text]:* Location, diameter and number of reinforcement bars; effective depth	ground penetrating radar (GPR)
Cracks	Flexural crack, x = 2.0 m, crack width 0.72 mm	*[Text]:* Location and maximum crack width [mm]	visual inspection
Internal defects	not detected	*[Text]:* presence of delamination, voids or flaws from loading history	CT scan
Section 3: Durability and Remaining Service Life			
Carbonation depth	15mm	Measured depth compared to existing concrete cover	phenolphthalein indicator
Chloride content	0,05 M.-%	[% by mass of cement]	-
Remaining service life	22	[a]	two-phase model according to Tuutti
DCE Classification	Silver		

23.02.2026 **Date** — **Signature** — **Inspection Engineer:** John Tester — **Example Lab, Adress**

Figure 5: Example of a printed reuse certificate based on the proposed parameters

CQ1: *Can all ''General Element and Project Information'' (Certificate Section 1) be queried in the model?*
CQ2: *Can all ''condition assessment'' and ''durability and remaining service life'' information be queried in the model? (Certificate Sections 2 and 3)?*
Both SPARQL queries were executed on the ICDD platform, returning all values from Figure 5 in tabular form

```
PREFIX inst: <https://icdd.vm.rub.de/converter-ws/model/ae874c60-a8ec-4915-a1a3-bc6d9b0df9ac#>
SELECT ?property ?value
WHERE {
?property isoprops:hasGroupOfProperties cero:Group_1_GeneralInformation .
  inst:beam_0b55c00b-a013-4601-a117-d24a166132f2 ?property ?value
}
```

Listing 1: SPARQL query for CQ1

```
PREFIX inst: <https://icdd.vm.rub.de/converter-ws/model/ae874c60-a8ec-4915-a1a3-bc6d9b0df9ac#>
SELECT ?procedure ?val
WHERE {
  ?procedure metadata4ing:investigates inst:beam_0b55c00b-a013-4601-a117-d24a166132f2> .
  ?procedure (metadata4ing:hasRuntimeAssignment/metadata4ing:hasVariable) ?var .
  { ?var metadata4ing:hasValue ?val .}
  UNION
  { ?var metadata4ing:hasNumericalValue ?val . }
}
```

Listing 2: SPARQL query for CQ2

and demonstrating that the proposed workflow enables information storage and querying in the model.

This research approach is limited by the availability of real inspection certificates, so a sample certificate was provided and validated. Moreover, the implementation of the OCR pipeline is limited because the focus of the study is not the employment of an optimized OCR but on the ontology implementation for inspection certificates with traceable test data provenance and the modeling of the information flow for a reuse RC Element. The OCR algorithm can be supplemented with segmentation to enable the correct reading of other layouts or the extraction of correct information from diagrams (e.g., M–N charts). Another avenue to improve information extraction could be to employ a visual–language model (VLM). However, since text recognition using OCR and VLMs involves uncertainties in data extraction, it should be noted that printed component certificates will be replaced by digitally signed inspection certificates in the long term. Currently, PDF-based certificates are used in the field of steel component reuse, as revealed in an expert interview conducted as part of the ‘‘CLUE’’ (Circular Building Design) research project.
The information flow from an inspection certificate to a digital passport, regardless of product or material passport, relies on the review of literature and has not yet been extensively tested in practice. Employing the proposed workflow is intended as a demonstration case on a reference reuse construction site in the next step. Moreover, an integration of the proposed inspection certificate parameters in addition with the the linking of the signed PDF Version into reuse Marketplaces such as this presented by Hagedorn et al. (2025) would be a meaningful next step in establishing the market for reused components.
Ontology alignment with other CE-specific ontologies remains to be addressed; particularly with the superior approach for CE ontologies across different industries by Blomqvist et al. (2023). The KG for inspection certificates requires further maintenance; missing test methods may need to be added, and additional parameters can be included if necessary. If CERO is further developed, adjustments must be made to the `certificates` KG.

Conclusion and outlook

This paper models the data flow for reused components. The proposed structure of inspection certificates was successfully used to (1) extract data using OCR and transfer it to the corresponding IFC model, and (2) maintain a persistent linkage to the reuse component using ICDD so that the signed PDF remains available at any time. The following research question was addressed as part of the prototype implementation of the workflow: *How can information be extracted from PDF-based inspection certificates for RC elements and reliably transferred to the BIM model, ensuring a continuous flow of information in the components lifecycle?*
To create the inspection certificate, a literature-based analysis of the required information was conducted; see Table 1. In addition to the necessary parameters, measurement methods are central, which is why possible measurement methods for parameter determination are listed. The compilation was implemented as an ontology using existing ontologies, including BOT, Metadata4ing, and ISOProps, in compliance with ISO 23386.
The information can be inserted into the ontology by text recognition (in this work with a simplified OCR algorithm) from PDF-based inspection certificates. To provide the information to users, the ICDD platform was chosen. This platform uses Linked Data to directly link different files to the corresponding IFC GUIDs, and the information can be queried directly on the model via SPARQL queries, which offers potential for further automation through the ICDD platform’s API (Application Programming Interface).
The ontology was validated using two CQs. Although

CERO is still under development, the example presented here shows that CERO supports CE processes for RC reuse elements. The OCR algorithm employed to demonstrate the complete information flow (see Figure 2) is error-prone and cannot be recommended for future work; it would need to be improved or replaced, or it may become obsolete if fully digital product passports become established.
The concept is suitable and expandable to other reuse products, e.g., steel and timber. Future work should investigate approaches without upstream paper documents, i.e., fully digital product passports. There may be new regulations for product passports through the Circular Economy Act. The developed ontologies can also be aligned with the parameters of DIN SPEC 91525 (''Post-Use Concept (PUC) for construction products from existing buildings''), which was published in February 2026, after the implementation of this work was finalized.

Acknowledgments

The authors acknowledge financial support for this research as part of the ''CRC 1683'' funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 523205694 and financial support as part of the project ''CLUE'' (project number EFRE-20800305) funded by the European Union and the state of North Rhine-Westphalia as part of the EFRE/JTF program NRW 2021-2027. Further information can be found at www.efre.nrw.

References

Adu-Duodu, K., Wilson, S., Li, Y., Oladimeji, A., Huraysi, T., Barati, M., Perera, C., Solaiman, E., Rana, O., Ranjan, R., and Shah, T. (2025). A Circular Construction Product Ontology for End-of-Life Decision-Making. In Hong, J., Battiato, S., Esposito, C., Park, J. W., and Przybyłek, A., editors, Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing, pages 1943–1952, New York, NY, USA. ACM.

Aftab, U. and Agliata, R. (2026). Material passports for digital and circular construction: a bibliometric and thematic analysis. Journal of Cleaner Production, 543:147651.

Akbarieh, A., Osadcha, I., Farghaly, K., Bosché, F., and Puust, R. (2025). Development of a Built Environment Ontology Lookup Service (BE-OLS) with a New Ontology Evaluation. In Proceedings of the 2025 European Conference on Computing in Construction. European Council for Computing in Construction.

Akbarieh, A., Teferle, F. N., and O'Donnell, J. (2023). Semantic material bank: A web-based linked data approach for building decommissioning and material reuse. In Scherer, R. J., Sujan, S. F., and Hjelseth, E., editors, ECPPM 2022 - eWork and eBusiness in Architecture, Engineering and Construction 2022, pages 69–76. CRC Press, London.

Al-Qazzaz, I. (19.12.2024). Circularity bSDD: Version 0.0.1 (preview). Published in the bSDD by University of Nottingham.

Bernier, C. and Danash, F. (2024). Digital Product Passport (DPP) Prototypes.

Bezerra, C., Freitas, F., and Santana, F. (2013). Evaluating Ontologies with Competency Questions. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), pages 284–285. IEEE.

Blomqvist, E., Li, H., Keskisärkkä, R., Lindecrantz, M., Pour, M. A. N., Li, Y., and Lambrix, P. (2023). Cross-domain Modelling – A Network of Core Ontologies for the Circular Economy. In 14th Workshop on Ontology Design and Patterns (WOP 2023) - Colocated with the 22nd International SemanticWeb Conference (ISWC 2023).

EC3 M&S Committee (2025). Built Environment Ontology Lookup Service.

European Commission (2008). Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain Directives.

Hagedorn, P., Donkers, A., Petrova, E., and König, M. (2025). Decentralised data exchange in construction data spaces using information containers. In 2025 European Conference on Computing in Construction & 42nd CIB W78 Conference on IT in Construction.

Hagedorn, P., Pauwels, P., and König, M. (2023). Semantic Rule Checking of Cross-Domain Building Data in Information Containers for Linked Document Delivery using the Shapes Constraint Language. Automation in Construction, 156.

ISO 21597-1 (2020). Information container for linked document delivery: Exchange specification - Part 1: Container. International Organization for Standardization, Geneva, Switzerland, 1 edition.

Krautgartner, N., Günther, F., Hagedorn, P., Hafner, A., Thewes, M., and König, M. (2026). A Semantic Digital Twin for the Sustainability Assessment of Reusable Structural Concrete Elements. Proceedings of the International Conference on Computing in Civil Engineering (i3CE 2026).

Krautgartner, N., Hagedorn, P., Thewes, M., and König, M. (2025). GIS-Integrated Simulation of Logistics Processes for the Reuse of Structural Concrete Elements. In Proceedings of ACCCBE 2025 & CONVR 2025, Lecture Notes in Civil Engineering. Springer, Cham.

Küpfer, C. and Fivet, C. (2024). Structural component reuse of precast and cast-in-place reinforced concrete in architecture since the late 1960s in europe. In Proceedings of the 8th International Congress on Construction History (Construction Matters), pages 1077–1084.

Lanza, G., Iglezakis, D., Fuhrmans, M., Jordan, M., Farnbacher, B., Sosa Rodriguez, A. A., Leimer, S., Hachinger, S., Arndt, S., Terzijska, D., Wellmann, A., Theissen-Lipp, J., and Munke, J. (2025). Metadata4Ing: An ontology for describing the generation of research data within a scientific activity. v1.4.0.

Li, H., Abd Nikooie Pour, M., Li, Y., Lindecrantz, M., Blomqvist, E., and Lambrix, P. (2023). A Survey of General Ontologies for the Cross-Industry Domain of Circular Economy. In Companion Proceedings of the ACM Web Conference 2023, pages 731–741, New York, NY, USA.

Liu, L., Hagedorn, P., and König, M. (2021). An ontology integrating as-built information for infrastructure asset management using BIM and Semantic Web. In Proceedings of the 2021 European Conference on Computing in Construction, pages 99–106.

Mark, P., Sanio, D., Balzani, D., Curoșu, I., Hafner, A., König, M., Kuhlenkötter, B., Thewes, M., Baille, W., Hackl, K., Hagedorn, P., Meschke, G., Nestorović, T., Neu, G. E., Sauer, R., Tafili, M., Vogel, A., Weber, S., Wichern, M., Wichtmann, T., Niederleithinger, E., Freitag, S., Menges, A., and Simon, S. (2025). Modulares Wiederverwenden von Bestandstragwerken. Bautechnik, 102(8):424–435.

Mecka, R., Geng, J., Schubert, A., Bos, F. P., Nübel, K., and Fischer, O. (2025). Ein Handlungsrahmen für die Wiederverwendung von Stahlbetonbauteilen. Beton- und Stahlbetonbau, 120(10):785–797.

Ottosen, L. M., Kunther, W., Ingeman-Nielsen, T., and Karatosun, S. (2024). Non-Destructive Testing for Documenting Properties of Structural Concrete for Reuse in New Buildings: A Review. Materials (Basel, Switzerland), 17(15).

Pauwels, P., Costin, A., and Rasmussen, M. (2022). Knowledge Graphs and Linked Data for the Built Environment. In Industry 4.0 for the Built Environment, pages 157–183. Springer International and Springer International Publishing, Cham.

Pauwels, P., McGlinn, K., Törmä, S., and Beetz, J. (2018). Linked data. In Building information modeling: Technology foundations and industry practice, pages 181–197. Springer.

Rasmussen, M. H., Lefrançois, M., Schneider, G. F., and Pauwels, P. (2021). BOT: The building topology ontology of the W3C linked building data group. Semantic Web, 12(1):143–161.

Rose, J., Forman, P., Stieler, D., Menges, A., and Mark, P. (2025). Combinatorial optimization approach for the efficient reuse of RC components. Structural Concrete, 26(6):7180–7203.

Ruismäki, W., zu Castell-Rüdenhausen, M., Pohjalainen, E., and Wahlström, M. (2025). Stakeholder perspectives on digital product passports for construction products. Cleaner Logistics and Supply Chain, 17:100275.

Sakthibala, R. K., Vasanthi, P., Hariharasudhan, C., and Partheeban, P. (2025). A critical review on recycling and reuse of construction and demolition waste materials. Cleaner Waste Systems, 12:100375.

Schönfelder, P. and König, M. (2025). Ontology-based reasoning in automatic floor plan analysis. Advanced Engineering Informatics, 68:103761.

Senarathne, H. N. Y., Weerasinghe, N. P., and Zhang, G. (2025). Developing a Standardized Materials Passport Framework to Unlock the Full Circular Potential in the Construction Industry. Sustainability, 17(14):6337.

Sithisint, V., Laoprom, S., Wattanapornprom, W., Chotibuth, N., Nupakorn, M., Satittham, N., and Bumrungvongsiri, T. (2025). Automated Bounding Dimension Retrieval: A Machine Learning-Driven Framework for Engineering Drawing Interpretation. In 2025 IEEE International Conference on Cybernetics and Innovations (ICCI), pages 1–6. IEEE.

Suchorzewski, J., Santandrea, F., and Malaga, K. (2023). Reusing of concrete building elements – Assessment and quality assurance for service-life. Materials Today: Proceedings.

Vaatz, A., Hamdan, A.-H., Al-Sadoon, N., Wogan, M., and Menzel, K. (2023). Integration of semantic temporal information in BIM using ontologies. In Proceedings of the 2023 European Conference on Computing in Construction and 40th International CIB W78 Conference, Computing in Construction, Heraklion, Crete, Greece. European Council for Computing in Construction.

Vahidi, A., Gebremariam, A. T., Di Maio, F., Meister, K., Koulaeian, T., and Rem, P. (2024). RFID-based material passport system in a recycled concrete circular chain. Journal of Cleaner Production, 442:140973.

Zentgraf, S., Hagedorn, P., and König, M. (2026). A BIM-based framework for automated building code extraction and compliance checking: under review. Advanced Engineering Informatics.

Zentgraf, S. and Zimmermann, L. (2025). Automated Extraction of Information Requirements from Building Regulations using LLMs and Ontologies. In Proceedings of the 36th Forum Bauinformatik 2025, pages 155–162, Aachen. RWTH Aachen.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

INTEROPERABILITY FOR A CIRCULAR BUILT ENVIRONMENT: TOWARDS A ROADMAP

Mustafa Selçuk Çıdık, Irene Josa, Kell Jones, Nicola Moretti and Karim Farghaly
The Bartlett School of Sustainable Construction, University College London, London, UK

Abstract

Circular built environment relies on reliable, reusable information across life-cycle stages; however, current initiatives remain fragmented. Identifying interoperability as a primary barrier, this paper draws on insights from a multi-stakeholder workshop to present a systemic roadmap. The research consolidates expert perspectives into sequenced, mutually dependent interventions across five domains: shared strategic vision; aligned standards and requirements; organisational and human capability; incentives and value realisation; and integrated system architecture. By clarifying cross-domain dependencies, this framework provides an actionable basis for prioritising pilots, policy mechanisms, and digital infrastructure to scale circular practices beyond isolated demonstrators.

Introduction

The Circular Economy (CE) has emerged as a transformative model to address environmental pressures and resource inefficiencies by replacing linear "take-make-dispose" patterns with regenerative systems that preserve material value and reduce waste (Ellen MacArthur Foundation, 2013; Kirchherr, Reike & Hekkert, 2023). The construction sector is a critical arena for this transition due to its high material throughput, complex supply chains, and significant waste generation (Ghisellini, Cialani & Ulgiati, 2016; Peng & Cao, 2025). In practice, CE in construction relies on strategies such as reuse, refurbishment, recycling, and design for disassembly, which are approaches that reduce life-cycle burdens while supporting longer-term sustainability (Konietzko, Bocken & Hultink, 2020). Yet, despite strong policy and industry interest, implementation remains uneven and difficult to scale beyond isolated pilots (Köhler et al., 2025).

Scaling CE depends on life cycle and systems thinking, as circular strategies require actors to coordinate decisions across stages (design, procurement, construction, operation, end-of-life) and across organisational boundaries (Voulvoulis et al., 2022). This coordination is information-intensive: to enable reuse and high-value recovery, stakeholders must trace materials, access reliable product data, assess reuse potential, and align compliance requirements over time (Giorgi et al., 2022). However, the built environment is characterised by fragmented processes and heterogeneous stakeholder needs, creating persistent challenges in communication and accountability (Jones et al., 2022). Consequently, circular initiatives often stall because information sharing is inconsistent, unverifiable, or untrusted across the value chain (Jäger-Roschko & Petersen, 2022; Mosgaard, Kristensen & Bocken, 2025).

Digitalisation is frequently presented as an enabler: through Building Information Modelling (BIM) tools, Digital Product Passports (DPP), Digital Building Logbooks (DBL), Lifecycle Assessment (LCA) tooling, and emerging "data spaces" approaches. However, digital adoption does not automatically translate into circularity (Dervishaj & Gudmundsson, 2024). A central barrier is interoperability: even when data is exchanged, it is often inconsistent in meaning, granularity, and context (Farghaly & Jones, 2023; Parte et al., 2025). This challenge is amplified because circular construction operates as a system-of-systems, where independently governed subsystems must coordinate without losing autonomy (Sadeghi et al., 2024; Farghaly et al., 2025). While technical advances in ontologies and knowledge graphs are progressing, they are often pursued in isolation from governance, incentives, and capability development; this leads to fragmented solutions that do not scale (Kebede et al., 2024; Shaw et al., 2025). What is missing is an integrated change model that clarifies how standards, systems, and data initiatives should be sequenced and reinforced through complementary institutional and market interventions.

This paper addresses this gap by proposing a systemic roadmap for interoperability in support of circular construction. Synthesising emerging insights from a multi-stakeholder workshop, the research organises priority actions into five interdependent domains (i.e. shared strategic vision; aligned standards and requirements; organisational and human capability; incentives and value realisation; and integrated system architecture). The framework is intended as a basis for further development.

Background

Interoperability has been widely discussed in the literature, yet it does not have a single unified definition (Lalli, 2025). It is generally understood as the ability of different systems or components to exchange information and use it effectively for a given purpose. However,

studies show that interoperability has evolved from a purely technical concern to a multidimensional socio technical concept, leading to the emergence of numerous definitions, models, and classifications. A recent tertiary study identified thirty-six different interoperability types and highlighted that ambiguity and overlap across definitions still exist, which can hinder communication and implementation across domains (Maciel et al., 2024).

Within the policy context, the European Interoperability Framework for Smart Cities and Communities identifies five key types of interoperability that should be addressed together, namely technical, semantic, organisational, legal, and culture interoperability. Within this broader landscape, particular attention in both research and practice has increasingly centred on semantic and technical interoperability.

Semantic interoperability refers to the ability of heterogeneous systems to exchange information such that the meaning of the information is preserved and can be interpreted consistently by different tools and stakeholders. In the built environment, this requirement spans multiple data types and scales (e.g., design models, product information, operational records, and assessment outputs) and must hold across long asset lifecycles and organisational boundaries. This is particularly challenging in construction, where fragmentation and specialisation create persistent discontinuities in how information is produced, owned, and maintained (Jones et al., 2022), and where interoperability increasingly resembles a system of systems problem with independently governed subsystems that must coordinate while retaining autonomy (Farghaly et al., 2025). This challenge is often approached through digitalisation and data exchange, yet exchange alone does not guarantee shared meaning. Even when information can be transferred between tools, inconsistencies in terminology, classification, granularity, and context can prevent reliable reuse and automation. This is evident in sustainability-oriented workflows where requirements and evidence must be mapped across heterogeneous sources and interpretations, making consistent meaning a prerequisite for trustworthy decision-making (Farghaly & Jones, 2023). Similar challenges appear in broader digital ecosystems (e.g., IoT-driven environments), where semantic interoperability is explicitly needed to ensure that multiple systems "act on" shared information consistently (Parte et al., 2025).

For circular construction in particular, circularity strategies depend on the availability of reliable lifecycle information to support traceability, assessment, and end-of-life decisions. Digital tools such as lifecycle assessment platforms and emerging passport/logbook concepts are promising, but their ability to deliver circular outcomes depends on the interoperability of underlying data and concepts rather than on tool adoption alone (Dervishaj & Gudmundsson, 2024). Recent work on digital product passports in the built environment demonstrates how ontology-based modelling can improve reuse and interpretability of product information by structuring lifecycle content and aligning concepts with standards, thereby improving cross-platform and cross-stakeholder consistency (Kebede et al., 2024). In parallel, knowledge-graph approaches demonstrate the potential to connect requirements and reporting across policy and practice contexts, supporting more robust and machine-interpretable lifecycle analysis and disclosure (Shaw et al., 2025). However, the literature also highlights persistent constraints that prevent semantic approaches from scaling beyond demonstrators. Information sharing across circular value chains remains difficult due to organisational boundaries, uneven incentives, and limited trust in shared data, all of which reduce the willingness and ability of actors to invest in producing and maintaining high-quality information (Jones, Farghaly & Toth, 2025; Jäger-Roschko & Petersen, 2022).

These socio-technical barriers do not operate independently but interact directly with technical constraints. In the absence of shared requirements, consistent semantics, and distributed implementation capacity across organisations, even technically robust solutions struggle to achieve effective adoption. In this context, "well-designed" refers to artefacts that are internally coherent, formally specified, and semantically expressive, often providing comprehensive coverage of domain concepts and relationships. However, these qualities can also introduce practical challenges and remain difficult to generalise and have seen limited in-practice uptake (Shaw et al., 2025).

Digital solutions in the built environment, such a DPPs and DBLs, are explicitly designed to improve data accessibility, lifecycle transparency, and cross-stakeholder coordination. Yet, despite operating in a data-rich environment, the sector remains highly fragmented, and much of this information is underutilised due to structural discontinuities across the value chain. Crucially, recent empirical evidence shows that barriers to integration are not primarily technical but procedural and organisational, including unstructured data, unclear responsibilities, and weak incentives (Çetin et al., 2026). Even where technical infrastructures are considered advanced, interoperability is constrained by the lack of shared definitions and semantic alignment, making consistent data exchange difficult to operationalise (Jones, Farghaly & Toth, 2025). At the same time, reliance on unstructured formats (e.g., PDFs and fragmented documentation) further limits the usability and automation of data within these systems (Çetin et al., 2026). These constraints are compounded by organisational and human factors. Responsibilities for data provision are often ambiguous, and stakeholders lack both incentives and capacity to contribute to shared systems, resulting in low engagement and inconsistent implementation. Moreover, the proliferation of overlapping digital tools and frameworks can create additional complexity and "digital fatigue," reducing willingness to adopt new systems (Supriyadi et al., 2025).

Consequently, solutions that are "well-designed" in a technical sense may still reproduce fragmentation in practice. Rather than eliminating heterogeneity, they shift the burden of integration onto individual organisations,

each of which must align its internal data structures, classifications, and workflows with shared models. Interoperability is therefore achieved at the level of architectural intent, but not consistently realised in operational settings. This helps explain why, despite their conceptual robustness and policy support, many such systems struggle to scale beyond limited implementations and fail to achieve widespread, routine use across the sector.

Taken together, the above motivates the need for a systemic approach that integrates technical measures (standards, semantics, architecture) with complementary governance and adoption conditions – providing the rationale for the roadmap proposed in this paper.

Methodology

This research is grounded in a structured stakeholder workshop designed to elicit, consolidate, and sequence practical actions for enabling semantic interoperability in circular construction.

In particular, a full day workshop was convened involving 45 participants representing key segments of the value chain and enabling ecosystem (including asset owners, designers, contractors, manufacturers, software providers, policymakers, and standardisation bodies). The participant sample comprises 45 individuals across four stakeholder groups. Solution implementers form the largest group (n=16), with 10 involved in design and construction and 6 in operation and in-use. This is followed by academics (n=13), governance stakeholders (n=9), and solution providers (n=7). Academic participants range from professors to research associates, covering digital technologies, sustainability, project management and supply chains. Governance stakeholders are involved in international and national standards and regulations, while solution providers are primarily focused on data integration and digitalisation. Participants were selected through purposive sampling to ensure coverage of different perspectives (technical, regulatory, delivery).

The workshop followed a three-phase structure:

1. **Shared framing:** Collaborative definition of the interoperability problem, establishing boundaries and identifying critical pain points.
2. **Action elicitation:** Facilitated breakout discussions where participants proposed concrete technical, organisational, and policy interventions required to achieve data interoperability.
3. **Convergence and sequencing:** Consolidation of actions and dependency mapping (identifying which actions are prerequisites for others).

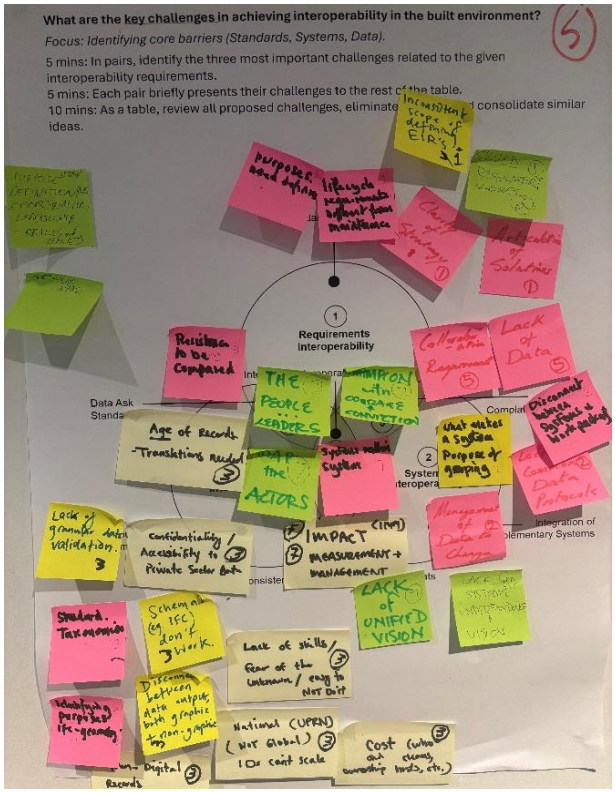


Figure 1: Example of one of the activities carried out in the workshop

Evidence included facilitator notes and group artefacts and worksheets (Figure 1). After the workshop session, the artifacts were digitised using spreadsheets and dashboards[1] for thematic analysis and clustering of the results. Subsequently, research team conducted an iterative qualitative synthesis to identify recurrent themes with the consolidation of overlapping proposals into coherent action statements.

The workshop findings played a central role in shaping the development of the Centre for Built Environment Data Sharing (C-BEDS) direction. A strong consensus emerged during the workshop that many initiatives are already addressing interoperability challenges. However, these efforts remain fragmented and largely disconnected. This creates a situation where solutions exist but their collective value is not realised. Based on this insight, the development process followed two linked directions. First, the activities of initiatives are aligned through a collective impact approach to support coordination, shared goals, and mutually reinforcing actions. Second, CBEDSync platform has been developed which has an open and queryable structure that captures existing initiatives, their outputs, and the relationships between them. This allows the connections between efforts to become visible and usable. In this way, the workshop findings directly informed both the conceptual design and the functional structure of the C-BEDS initiative.

[1] https://app.powerbi.com/view?r=eyJrIjoiNjUyOWM4MWQtYTQ0Zi00MjY4LTkxMzMtMzRhN2UwMDJmOGZkIiwidCI6ImMyMTM2Y2QxLWFkOGEtNGQ3MC05Yjc5LTBiNjM0NTNlNmJjYiJ9

Results and discussion

The synthesised actions were organised into five interdependent domains: (1) shared strategic vision; (2) aligned standards and requirements; (3) organisational and human capability; (4) incentives and value realisation; and (5) integrated system architecture. These are described in what follows and summarised in Table 1.

The actions elicited from the workshop reflect both practical concerns and an interconnected set of considerations. While these interventions are grouped into five thematic domains, their success depends not on isolated implementation but on how they reinforce, enable, or depend upon one another. This section interprets these emerging findings as part of a holistic roadmap, synthesising the identified challenges and enablers with the strategic role each domain plays. The goal is to clarify their function within a coherent change model, highlighting interdependencies and complementarity.

Domain 1: shared strategic vision

A shared strategic vision acts as the anchor of the proposed roadmap; it defines what interoperability is for, whom it benefits, and how value is distributed. Without such a vision, efforts to standardise, invest, or reform institutional practices remain fragmented. Participants widely agreed that strategic misalignment across governance levels and organisations perpetuates inefficiencies and undermines the circular economy agenda. Consequently, the actions in this domain form the foundation upon which all others depend.

These actions emphasise the need to convene key stakeholders early to articulate a shared purpose, formalise commitment across sectors, and embed this vision within regulation. Critically, participants viewed this not merely as a leadership issue but as an operational necessity: without a shared "why," technical or behavioural changes lack legitimacy. Enablers such as cross-sector collaboration, political leadership, and regulatory clarity support this alignment, while the absence of clear attribution remains a systemic risk.

This domain requires institutional continuity. Embedding the vision in procurement, funding frameworks, and leadership roles ensures adaptation to evolving technological and regulatory contexts. It is not merely the starting point of a linear plan, but a persistent condition for success across the roadmap.

Domain 2: aligned standards and requirements

Standards and requirements form the semantic backbone of interoperability. They define how information is structured, understood, and exchanged, underlying every attempt to automate, audit, or regulate the flow of data.

Workshop participants identified that the primary challenge is not the absence of standards, but their incoherence and incompatibility. Fragmented schemas, redundant classification systems, and diverging national regulations impede scalability and render meaningful data reuse nearly impossible.

As such, an interesting and recurring finding across the workshop was that standards are both indispensable and problematic: interoperability depends on shared standards, yet the proliferation of overlapping standards, schemas, and jurisdiction-specific requirements is itself a major source of fragmentation.

The actions proposed for this domain reflect a pragmatic, phased approach: first mapping and analysing existing standards, then consolidating and simplifying them into shared, machine-readable formats. Participants emphasised the need for open definitions and digital tools that support validation, traceability, and compliance without imposing excessive complexity. Importantly, these standards must be dynamic, namely able to evolve with emerging needs and remain adaptable across sectors.

Crucially, standards are not self-sufficient. Their relevance and uptake depend on how they are embedded within procurement and policy frameworks and whether they are understood by the workforce. The system-level implication is clear: aligned standards cannot succeed without prior strategic clarity and simultaneous investment in organisational capability. In this sense, standardisation is both a technical and a political act, defining not just what is possible, but what is prioritised.

Domain 3: organisational and human capability

Where shared vision gives direction and standards provide structure, organisational and human capability enables action. This domain addresses persistent capacity deficits: low data literacy, inconsistent leadership, and siloed technical expertise. These issues are not resolved by training alone; they reflect a broader cultural landscape in which data is undervalued, ownership is ambiguous, and resistance to change remains high.

The actions outlined for this domain propose a multi-layered response: building awareness of data value, implementing training for key roles, and embedding data responsibilities into procurement and professional standards. These are long-term investments in behavioural and institutional transformation rather than quick fixes. Enablers identified by participants (such as cross-sector taskforces, targeted leadership programmes, and revised contracting norms) support this shift. Thus, this domain bridges strategic intention and operational implementation. Without capability, actions in other domains remain aspirational. Moreover, strong leadership and a skilled workforce create feedback loops that accelerate adoption and continuous improvement, making capability the operational foundation of systemic change.

Domain 4: incentives and value realisation

While capability enables action, incentives determine whether actors choose to act, and how consistently. Discussions highlighted that, in the current system, data is often perceived as a cost centre rather than a strategic asset. Ownership is unclear, accountability is weak, and value is poorly articulated. Participants repeatedly stressed that unless data-related behaviours are tied to clear benefits (financial, reputational, or operational), interoperability will not scale.

Table 1: Roadmap domains and aspects for interoperability

Roadmap domains	Core focus	Key challenges	Key enablers	Emerging actions	Strategic role
Domain 1 Shared strategic vision	Establishing purpose, coherence, and direction	Fragmented goals, unclear ownership, siloed strategies	Governance coordination, regulatory alignment, leadership buy-in	Convening stakeholders, formalising commitments, embedding vision in regulation	Vision is the strategic anchor. It defines direction, legitimises action, and enables cross-sector coherence.
Domain 2 Aligned standards & requirements	Creating a shared, usable foundation for data exchange	Incompatible formats, misaligned regulations, excessive complexity	Open ontologies, bsDD, policy alignment, machine-readability	Mapping standards, simplifying formats, developing validation tools	Standards are the semantic backbone. Their adoption depends on clarity, usability, and integration into compliance tools.
Domain 3 Organisational & human capability	Enabling people and institutions to adopt and sustain change	Low data literacy, weak leadership, fragmented skills	Targeted training, communities of practice, procurement reform	Building awareness, role-specific training, revising professional standards	Capability is the human infrastructure. It translates standards into practice and enables culture change across roles.
Domain 4 Incentives & value realisation	Making interoperability worthwhile and accountable	Unclear data ownership, low perceived value, lack of ROI	Outcome-based models, new business cases, financial levers	Demonstrating ROI, embedding contract accountability, risk mitigation	Incentives are the behavioural engine. They must reward outcomes, not just compliance, and reinforce data as a shared asset.
Domain 5 Integrated system architecture	Building a flexible but connected digital ecosystem	Legacy systems, poor integration, document-centric workflows	Open-source tools, federated infrastructure, aligned APIs	Federated architecture, aligned ontologies, prototyping real-time flows	Architecture is the operational backbone. Modular, open systems enable the technical realisation of all other domains.

To address this gap, the roadmap suggests diverse mechanisms: from demonstrating ROI and linking data quality to outcomes, to embedding accountability in contracts and utilising financial instruments to mitigate risk. These actions aim to align economic logic with system goals, transforming data interoperability from a compliance burden into an investment opportunity. The strategic implication is twofold: first, incentives depend on visibility – without clear standards and trustworthy infrastructure, value is difficult to verify; second, incentives are transformative – when tied to outcomes like reuse rates, they shift market behaviour. This domain therefore acts as the behavioural engine for the system, reinforcing positive norms and accelerating diffusion.

Domain 5: integrated system architecture

Finally, the technical realisation of interoperability depends on an integrated but decentralised system architecture. The workshop revealed persistent fragmentation, with data locked in static documents, proprietary systems, or disconnected platforms. The enablers identified – open ontologies, modular APIs, federated platforms – point to a future where interoperability is achieved through structured compatibility rather than centralisation.

Actions in this domain articulate this vision, including developing a federated architecture, aligning ontologies, reducing API complexity, and creating usable tools for real-time data flows. Importantly, these actions position prototyping and demonstration not as side activities, but as essential components for building confidence. From a system perspective, architecture is the operational backbone supporting all other domains. If designed in isolation, it risks reproducing existing silos; but if aligned

with the broader roadmap, it becomes the delivery mechanism for semantic interoperability.

Synthesis of the framework

Together, these five domains form a cohesive implementation framework. Strategic Vision provides the guiding purpose; Standards define the rules of engagement; Capabilities ensure actors can act; Incentives align behaviour with value; and Architecture operationalises the system through technology. The actions synthesised in this report are not isolated fixes, but interdependent interventions requiring coordination across governance levels, sectors, and systems.

Way forward: operationalising the roadmap through C-BEDS and C-BEDSync

The roadmap presented in this paper emphasises that semantic interoperability for a circular built environment cannot be achieved through isolated technical deliverables. Implementation requires sustained coordination across actors and initiatives to reduce duplication, build alignment, and maintain continuity of purpose and measurement.

As a practical response, the authors of this article have established C-BEDS (Centre for Built Environment Data Sharing) as a collaborative "network of networks" designed to connect, support, and amplify existing work rather than introducing an additional competing initiative.

Governance

C-BEDS is framed as a sector-wide collaboration committed to making data-sharing initiatives accessible, connected, and valued, recognising that new activities must build on substantial work already underway. It is explicitly intended to support alignment and collective progress through shared infrastructure, communication, and coordination, while maintaining distributed leadership and ownership across the ecosystem.

The governance framework is informed by a Collective Impact approach and organises alignment through five areas: (i) common agenda, (ii) shared measurement systems, (iii) mutually reinforcing activities, (iv) continuous communication, and (v) backbone support (Kania et al., 2022). Backbone support is described as being coordinated by an organisation while being collectively shaped and governed by the wider community, to sustain momentum, trust, and fairness across participating organisations.

C-BEDSync and the Memorandum of Understanding as implementation instruments

To operationalise "shared measurement" and make ecosystem progress visible, C-BEDS defines two primary outputs. The first is C-BEDSync, a platform that collects, categorises, and visualises data-sharing initiatives to support awareness, learning, and the avoidance of duplication. The governance framework proposes to assess C-BEDSync through indicators such as (a) the number of initiatives captured, (b) the quality of categorisation, and (c) the strength of links established between initiatives (Figure 2). To maintain the live mapping of standards, schemas, and related initiatives as regulations and market practices evolve, CBEDS envisages a backbone stewardship function supported by a cross-sector working group including standards bodies, regulators, software providers, and industry

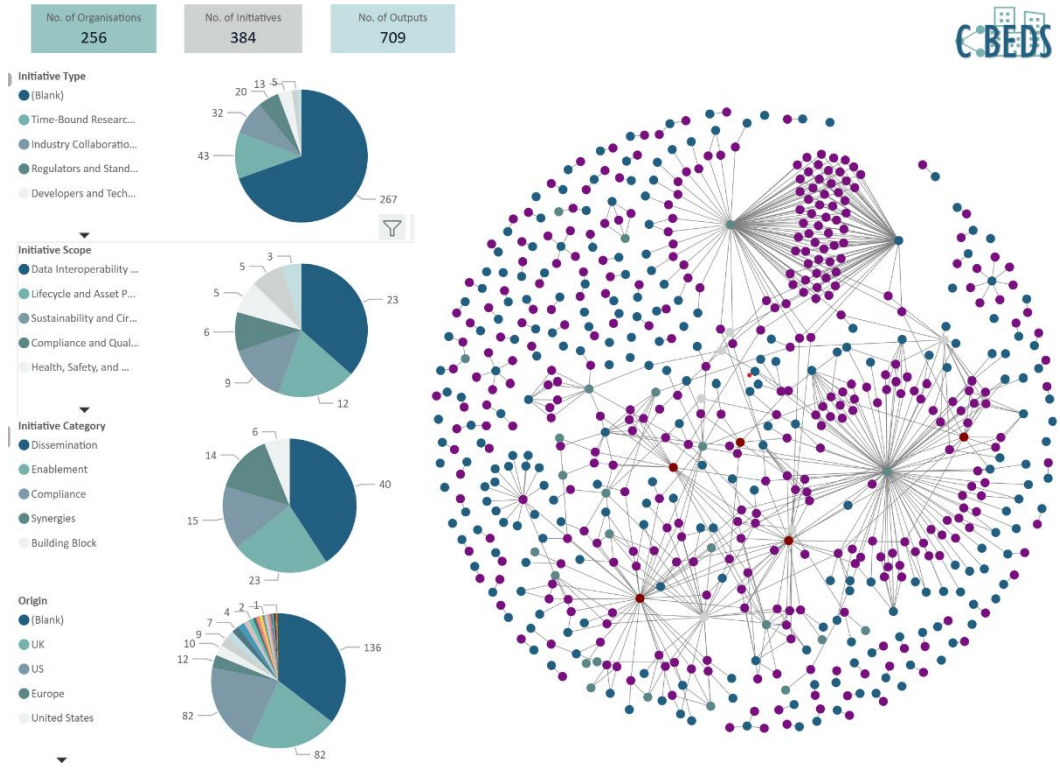


Figure 2: Screenshot from C-BEDSync

Table 2: Implementation pathway across the five domains

Roadmap domains	C-BEDS approach
Domain 1 Shared strategic vision	"Common agenda" mechanisms support alignment on outcomes and priorities across initiatives and governance levels.
Domain 2 Aligned standards & requirements	C-BEDSync's mapping function can improve discoverability of relevant standards, ontologies, and templates and identify overlaps, gaps, and opportunities for consolidation.
Domain 3 Organisational & human capability	Recurring workshops, webinars, and community engagement create a channel for shared learning and diffusion of practices, supporting capability development at ecosystem scale.
Domain 4 Incentives & value realisation	Shared measurement and the MoU provide mechanisms to make collaboration visible and to translate intent into accountable joint action.
Domain 5 Integrated system architecture	By linking initiatives and their outputs, C-BEDSync can help prioritise interoperability pilots and reduce duplicated platform building, increasing the likelihood that architectural work converges on usable, shared semantics.

representatives. This group would be responsible for periodic updates, validation of mappings, and publication of agreed changes.

The second instrument is a Memorandum of Understanding (MoU) intended to formalise collaboration among organisations and enable deeper alignment of objectives. Progress is to be tracked not only through the number and diversity of signatories, but also through tangible collaborative outcomes (e.g., formation of consortia or joint research and innovation proposals).

In addition, C-BEDS outlines a portfolio of mutually reinforcing activities such as workshops and webinars, white papers and academic outputs, community engagement, C-BEDSync platform development, and coordination/project management explicitly positioned to complement rather than duplicate existing networks and initiatives.

How this "way forward" connects to the roadmap dimensions

Positioned against the five-domain roadmap proposed in this paper, C-BEDS provides a practical implementation pathway as shown in Table 2. Overall, C-BEDS and C-BEDSync exemplify a concrete "roadmap-to-action" approach: providing governance, measurement, and connectivity infrastructure that can help translate interdependent actions into coordinated delivery, while preserving distributed ownership across the ecosystem.

Conclusions

This paper reframed semantic interoperability for a circular built environment as a systemic challenge rather than a purely technical integration task. Based on multi-stakeholder workshop evidence, we synthesised a roadmap that organises mutually dependent interventions across five domains – shared strategic vision, aligned standards and requirements, organisational and human capability, incentives and value realisation, and integrated system architecture – and highlighted why progress depends on coordinated sequencing rather than isolated actions. The "Way forward" section illustrates how this logic can be operationalised through cross-initiative coordination mechanisms such as C-BEDS and the C-BEDSync mapping platform, which aim to reduce duplication and strengthen alignment across the ecosystem. Future work should validate the roadmap through applied pilots and shared measurement, assessing whether cross-domain coordination improves data quality, traceability, and circular decision-making at scale.

References

Çetin, S., Mêda, P., Farghaly, K. and Hwang, S.-A. (2026). 'Linking digital product passports and digital building logbooks: Socio-Technical challenges and a pathways for integration'. *Developments in the Built Environment*, 25, p.100877.

Dervishaj, A. and Gudmundsson, K. (2024). 'From LCA to circular design: A comparative study of digital tools for the built environment'. *Resources, Conservation and Recycling*, 200, p.107291.

Ellen MacArthur Foundation (2013). *Towards the Circular Economy: Economic and business rationale for an accelerated transition*. Ellen MacArthur Foundation. Available [Online] at: https://ellenmacarthurfoundation.org.

Farghaly, K. and Jones, K. (2023). 'Enhancing Requirement-Information Mapping for Sustainable Buildings: Introducing the SFIR Ontology'. In *Advances in Conceptual Modeling*. Cham: Springer Nature Switzerland, pp.242–248.

Farghaly, K., Mêda, P., Shaw, C., O'Donnell, J., Cecconi, F.R. and Moretti, N. (2025). 'Design Wire Engineer Leverage and Transfer (DWELT) Framework for Building-Level Digital Twins'. In *Building Digital Twin International Congress*. Springer, pp.1–12.

Ghisellini, P., Cialani, C. and Ulgiati, S. (2016). 'A review on circular economy: the expected transition to a

balanced interplay of environmental and economic systems'. *Journal of Cleaner Production*, 114, pp.11–32.

Giorgi, S., Lavagna, M., Wang, K., Osmani, M., Liu, G. and Campioli, A. (2022). 'Drivers and barriers towards circular economy in the building sector: Stakeholder interviews and analysis of five European countries policies and practices'. *Journal of Cleaner Production*, 336, p.130395.

Jäger-Roschko, M. and Petersen, M. (2022). 'Advancing the circular economy through information sharing: A systematic literature review'. *Journal of Cleaner Production*, 369, p.133210.

Jones, K., Farghaly, K. and Toth, Z. (2025). 'A vision for data in the built environment: transforming sustainability reporting'. Available [Online] at: https://www.bpie.eu/wp-content/uploads/2025/03/DBL-Vision-Report-final.pdf.

Jones, K., Mosca, L., Whyte, J., Davies, A. and Glass, J. (2022). 'Addressing specialization and fragmentation: product platform development in construction consultancy firms'. *Construction Management and Economics*, 40 (11–12), pp.918–933.

Kania, J., Kramer, M., Juster, J. and Ide, S. (2022). 'Collective impact'. In *International Encyclopedia of Civil Society*. Springer, pp.1–8.

Kebede, R., Moscati, A., Tan, H. and Johansson, P. (2024). 'A modular ontology modeling approach to developing digital product passports to promote circular economy in the built environment'. *Sustainable Production and Consumption*, 48, pp.248–268.

Kirchherr, J., Reike, D. and Hekkert, M. (2023). 'Conceptualizing the circular economy: An analysis of 114 definitions'. *Resources, Conservation and Recycling*, 127, pp.221–232.

Köhler, J., Egerter, A.M., Hall, D.M. and Thuesen, C. (2025). 'Product platforms as enablers for the circular economy in construction: an integrative review'. *Computers & Industrial Engineering*, p.111277.

Konietzko, J., Bocken, N. and Hultink, E.J. (2020). 'Circular ecosystem innovation: An initial set of principles'. *Journal of Cleaner Production*, 253, p.119942.

Lalli, G. (2025). 'Defining Interoperability: A Universal Standard'. *Journal of ICT Standardization*, 13 (2), pp.139–156.

Maciel, R.S.P., Valle, P.H.D., Santos, K.S. and Nakagawa, E.Y. (2024). 'Systems Interoperability Types: A Tertiary Study'. *ACM Comput. Surv.*, 56 (10), p.254:1-254:37.

Mosgaard, M.A., Kristensen, H.S. and Bocken, N. (2025). 'Competition or coopetition? Collaboration strategies for the circular economy'. *Journal of Cleaner Production*, 513, p.145686.

Parte, M.S.E.D.L., Martínez-Ortega, J.-F., Martínez, N.L. and Díaz, V.H. (2025). 'SISS: Semantic Interoperability Support System for the Internet of Things'. *IEEE Internet of Things Journal*, pp.1–1.

Peng, L. and Cao, W.-J. (2025). 'Measuring circularity in the construction industry: A systematic literature review from the stakeholder perspective'. *Journal of Cleaner Production*, 496, p.144805.

Sadeghi, M., Carenini, A., Corcho, O., Rossi, M., Santoro, R. and Vogelsang, A. (2024). 'Interoperability of heterogeneous Systems of Systems: from requirements to a reference architecture'. *The Journal of Supercomputing*, 80 (7), pp.8954–8987.

Shaw, C., de Andrade Pereira, F., de Riet, M., Hoare, C., Farghaly, K. and O'Donnell, J. (2025). 'Knowledge graph for policy- and practice-aligned life cycle analysis and reporting'. *Automation in Construction*, 176, p.106282.

Supriyadi, T., Sulistiasih, S., Rahmi, K.H., Pramono, B. and Fahrudin, A. (2025). 'The impact of digital fatigue on employee productivity and well-being: A scoping literature review'. *Environment and Social Psychology*, 10 (2).

Voulvoulis, N., Giakoumis, T., Hunt, C., Kioupi, V., Petrou, N., Souliotis, I., Vaghela, C. and binti Wan Rosely, WIH. (2022). 'Systems thinking as a paradigm shift for sustainability transformation'. *Global Environmental Change*, 75, p.102544.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DECISION SUPPORT FRAMEWORK FOR INTEGRATING CONSTRUCTION CONSOLIDATION CENTRE INTO SUSTAINABLE WASTE MANAGEMENT

Jintian Xu, Thijs Nooijen, Qi Han and Dujuan Yang
Eindhoven University of Technology, The Netherlands

Abstract

Inner-city construction projects face severe spatial constraints during the finishing stage, often leading to mixed construction and demolition waste (C&DW) handling and reduced circular performance. This paper proposes a project-level decision-support framework to evaluate outbound C&DW logistics strategies, addressing the lack of project-level comparative tools. The framework integrates economic, environmental, and social indicators in a scenario-based comparison of conventional handling and CCC-enabled reverse logistics. Applied to three Dutch projects, results show cost and emission reductions, with slightly higher logistics intensity. The framework supports ex ante decision-making under spatial constraints.

Introduction

Problem Context

Inner-city construction projects operate under conditions that fundamentally constrain material handling and logistics. Dense urban environments are characterised by limited on-site space, restricted access time windows, and regulatory constraints on vehicle movements. These conditions significantly reduce the feasibility of conventional construction logistics and waste management practices, particularly when compared to projects located in less spatially constrained areas (Guerlain et al., 2019).

These spatial constraints become especially critical during the finishing stage of construction—i.e., the late phase after the building shell is weather-tight, when interior fit-out, final installations, and snagging activities dominate until practical completion. Although the overall volume of construction and demolition waste (C&DW) often decreases towards project completion, the diversity and complexity of waste streams typically increase. This stage produces a heterogeneous mix of materials (e.g., wood off-cuts, drywall, insulation, plastics, metals, and packaging), which requires frequent handling and effective on-site separation—tasks that are difficult to perform on inner-city sites with limited storage capacity (Gálvez-Martos et al., 2018). Although the finishing stage does not necessarily generate the highest waste volume, it is particularly relevant for this study because it combines high waste-stream heterogeneity, frequent handling, packaging waste, and intensive subcontractor coordination. Under inner-city space constraints, these conditions make on-site source separation especially difficult and increase the likelihood that recyclable materials are disposed of as mixed waste.

Under such conditions, mixed C&DW handling often becomes the default practice. Instead of separating waste streams on-site, different materials are commonly disposed together in a single container and transported to recycling facilities in an unsegregated state. This practice is not the result of a deliberate preference for low-quality recycling, but rather a pragmatic response to severe spatial and logistical constraints at the project site (Wang et al., 2010).

The downstream consequences of mixed waste handling are significant. At recycling facilities, unsegregated waste requires additional mechanical sorting, which increases operational costs and energy consumption. In many cases, the economic feasibility of high-quality separation is limited, resulting in valuable materials being downcycled into low-grade applications rather than reused or recycled at their original value level (Bayram and Greiff, 2023). In addition to material degradation, mixed waste handling contributes to higher transport-related emissions and increased urban disturbance through inefficient logistics operations and additional freight movements (Guerlain et al., 2019).

Decision-Making Challenge

Project teams operating in inner-city construction contexts must continuously balance economic cost, environmental performance, and urban impacts under tight spatial and temporal constraints. Decisions regarding C&DW handling during the finishing stage are often made under pressure, with limited flexibility to revise logistical arrangements once construction activities are underway.

In practice, waste-handling decisions at the project level tend to be reactive and fragmented. Waste is frequently treated as an operational by-product rather than as a strategic component of construction logistics planning. As a result, decisions are commonly driven by short-term cost considerations and immediate site feasibility, without a structured ex ante comparison of alternative outbound waste logistics strategies (Brusselaers et al., 2024).

This fragmented decision-making environment is further reinforced by the absence of transparent, project-specific decision-support tools. While general sustainability frameworks and guidelines exist, they rarely provide actionable support for evaluating how different waste handling strategies perform under the specific spatial constraints of inner-city construction sites. Consequently, project managers lack the analytical means to systematically assess trade-offs between cost, emissions, and urban impacts during the finishing stage, when waste complexity is at its highest (Yuan, 2013).

Research Question

In response to the challenges outlined above, this study addresses the following research question:

How can a Construction Consolidation Centre (CCC) be integrated into outbound C&DW management during the finishing stage of inner-city construction projects, and how can the resulting economic, environmental, and social impacts be quantified and compared to support project-level decision-making?

Contribution

This study contributes to construction logistics and waste management in three related respects. It develops a project-specific framework for comparing outbound C&DW strategies during the finishing stage of inner-city projects. It also extends the discussion of CCCs beyond inbound supply logistics by examining their role in reverse logistics and source-separated waste collection. Finally, it brings cost, emissions, and urban disturbance into one comparative structure, allowing alternative waste-handling strategies to be assessed under the same project conditions.

Literature Review and Research Gap

C&DW Management Assessment

Assessment of construction and demolition waste (C&DW) management has been widely addressed in the literature, with a strong emphasis on life-cycle-based evaluation methods. Life Cycle Assessment (LCA) is frequently applied to quantify environmental impacts associated with waste treatment, transport, and recycling processes, while Life Cycle Costing (LCC) is commonly used to assess economic performance across collection, sorting, transport, and disposal activities (Finnveden et al., 2009). Together, these approaches provide valuable insights into the environmental and economic implications of alternative waste management strategies.

However, the majority of existing assessments operate at an aggregate, sectoral, or policy-oriented level. Many studies focus on national waste streams, generic building types, or average treatment scenarios, rather than on individual construction projects with specific spatial and logistical constraints (Yuan, 2013). As a result, the outcomes of such assessments are often descriptive in nature and insufficiently sensitive to project-level conditions such as site size, access restrictions, or construction phase.

In addition, social impacts are only marginally addressed in most C&DW assessment studies. Where social considerations are included, they are often treated qualitatively or discussed conceptually, rather than being operationalised through quantifiable indicators (Zhao et al., 2023). Urban disturbance effects related to freight movements, congestion, and nuisance are therefore rarely integrated alongside environmental and economic performance in a systematic manner (Muñuzuri et al., 2005).

Consequently, while existing C&DW assessment literature provides important methodological foundations, it offers limited support for project-level, scenario-based decision-making in complex inner-city construction contexts, particularly during the finishing stage when waste diversity and logistical pressure are the highest.

Construction Consolidation Centre (CCC)

A Construction Consolidation Centre (CCC) is an off-site logistics facility, typically located outside the dense urban core, where construction materials are received, consolidated, and re-dispatched to site in coordinated deliveries. Existing research shows that CCCs can improve delivery reliability, reduce inbound vehicle trips, and lower transport-related impacts in urban construction projects (Muñuzuri et al., 2005; Russo et al., 2021). They have been widely studied as a logistics solution for improving material delivery efficiency, with consistent evidence of reduced vehicle trips and emissions through coordinated last-mile logistics (Russo et al., 2021).

Most CCC studies focus explicitly on inbound material flows, evaluating performance in terms of delivery efficiency, cost savings, and emission reductions associated with supplier-to-site logistics. These analyses typically assess CCC performance using indicators such as vehicle kilometres travelled, delivery frequency, and load factors, often in comparison with conventional direct-to-site delivery models (Muñuzuri et al., 2005).

By contrast, the application of CCC to outbound logistics, and particularly to C&DW flows, remains largely unexplored (Rubio et al., 2019). Existing CCC literature rarely considers reverse logistics or the potential role of CCC in enabling source-separated waste collection under spatially constrained site conditions. Where waste is mentioned, it is typically treated as a secondary or operational issue rather than as a core component of logistics system design (Guerlain et al., 2019).

As a result, there is a lack of analytical tools and empirical studies that evaluate CCC-enabled reverse logistics for C&DW at the project level. In particular, the literature does not provide structured methods for comparing conventional mixed-waste handling with integrated CCC-based waste strategies under real inner-city constraints.

Research Gap

Despite useful insights from both strands of literature, a practical project-level gap remains. On the one hand, C&DW assessment studies predominantly rely on LCA and LCC approaches that are insufficiently tailored to

project-specific, scenario-based decision-making in spatially constrained environments (Finnveden et al., 2009). On the other hand, CCC research has largely focused on inbound logistics, with limited attention to outbound waste flows and reverse logistics applications (Russo et al., 2021).

What is still missing is a project-level method for comparing conventional mixed-waste handling with CCC-supported reverse logistics under the same site conditions. Empirical studies that make this comparison in spatially constrained inner-city settings are still limited. Accordingly, this study develops an ex ante, project-level decision-support framework that explicitly links outbound waste-handling strategies to inner-city logistical constraints and evaluates their economic, environmental, and social implications during the finishing stage of inner-city construction projects. This gap directly informs the framework and method presented in the following section.

Decision-Support Framework and Method

Framework Overview

The framework is used to compare alternative outbound C&DW logistics strategies during the finishing stage of inner-city construction projects. This stage is characterised by high material diversity, frequent logistics operations, and severe spatial constraints, which together complicate conventional waste-handling practices.

The framework adopts a scenario-based comparative logic, enabling project stakeholders to evaluate alternative waste-management strategies under identical project conditions. Two distinct strategies are defined:

- Baseline scenario, representing conventional mixed-waste handling practices commonly observed in spatially constrained inner-city projects (Haag and Jünger, 2023).
- Integrated scenario, representing a coordinated waste-handling strategy enabled through a Construction Consolidation Centre (CCC) with reverse logistics integration (Paddeu, 2018).

The decision context is explicitly project-level and ex ante. The framework is intended to be applied during the procurement and planning phase, when waste-handling strategies can still be adjusted without disrupting ongoing construction activities. The primary stakeholders supported by the framework include contractors, project managers, CCC operators, and, indirectly, municipal authorities responsible for regulating urban freight and construction logistics (Melkonyan et al., 2022; Sobotka and Sagan, 2021).

This setup allows the baseline and integrated scenarios to be compared on the same economic, environmental, and social indicators under the actual spatial conditions of each case (Maxner et al., 2025).

Key Performance Indicator (KPI) Framework

To support multi-criteria decision-making, the framework evaluates outbound C&DW management strategies using a structured set of Key Performance Indicators (KPIs) (Figure 1). The KPIs are limited to indicators that are decision-relevant and operationally estimable during procurement-stage planning and are grouped into three dimensions—economic, environmental, and social—as summarised in Table 1, together with their definitions and units. The framework is intended as a project-level, ex ante comparative assessment of finishing-stage outbound waste logistics rather than a full cradle-to-grave LCA/LCC.

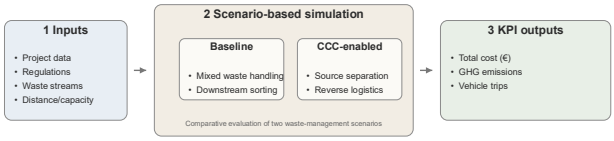


Figure 1: Decision-support framework for outbound C&DW logistics

Table 1: KPI Framework Overview

Dimension	Indicator	Unit	Description
Economic	Transport cost	€	Waste transport costs
Economic	Container handling	€	Container rental and handling
Economic	Sorting & labour	€	On-site sorting labour
Economic	Recovery revenue	€	Income from recyclables
Economic	CO_2 cost	€	Monetised carbon emissions
Environmental	Transport emissions	kg CO_2-eq	Freight-related emissions
Environmental	Sorting emissions	kg CO_2-eq	Waste treatment emissions
Social	Vehicle trips	number	Proxy for urban disturbance

Economic dimension

Economic performance is evaluated as a project-level cost–benefit balance for outbound C&DW logistics, enabling ex ante comparison across scenarios (Brusselaers et al., 2024; Yuan, 2013). Where relevant, carbon pricing is applied to express greenhouse gas emissions as a monetised externality within the economic comparison (Greenstone et al., 2013). Physical emissions are reported separately in the environmental dimension; the CO_2 cost term is included only to support integrated scenario comparison and does not replace the environmental assessment.

Environmental dimension

Environmental performance is assessed using greenhouse gas emissions associated with transport and mixed-waste sorting processes, allowing consistent comparison across logistics configurations (Butera et al., 2015; Hossain et al., 2017; Bayram and Greiff, 2023). Greenhouse gas emissions were selected as the primary environmental indicator because they can be estimated consistently using available project-level inputs, whereas broader impact

categories would require a level of process detail beyond the scope of this ex ante project-level assessment.

Social dimension

Social performance is captured through vehicle trips as an operational proxy for urban disturbance associated with waste logistics movements, providing a transparent basis for project-level comparison under typical data constraints (Çolak et al., 2016). Congestion, noise, and traffic-related disturbance are difficult to quantify directly at project level, whereas trip frequency offers a simple and comparable indicator of logistics intensity in dense urban contexts.

Scenario-Based Simulation Logic

The decision-support framework evaluates performance using two explicitly defined simulation scenarios. Both scenarios are modelled using identical project inputs to ensure comparability.

Baseline scenario

The baseline scenario represents conventional C&DW handling practices commonly applied in inner-city construction projects:

- Waste is collected in a single mixed container.
- No on-site waste separation is performed.
- Waste collection is reactive, triggered when the container is full.
- Waste is transported directly to recycling facilities, where mechanical sorting is required.

This scenario reflects a common fallback practice where limited site space and constrained layout conditions reduce the feasibility of stable on-site source separation (Tsegay et al., 2023).

Integrated scenario

The integrated scenario models a coordinated outbound waste-handling strategy enabled through a CCC with reverse logistics integration:

- Waste is separated on-site using multiple small, material-specific containers.
- Empty containers are supplied (and swapped when needed) via the same coordinated inbound delivery trips from the CCC, i.e., no dedicated trips are assumed solely for container provision.
- Inbound material deliveries and outbound waste collection are coordinated.
- Vehicles delivering materials to the site collect full waste containers on return trips.
- Weekly logistics planning is used to improve the coordination and utilisation of construction transport operations (Brusselaers et al., 2024).

By integrating inbound and outbound logistics, the integrated scenario aims to enable source separation without substantially increasing transport demand.

Decision-Support Tool Implementation

The framework is implemented as a project-level decision-support tool (DST) for the ex ante evaluation of alternative outbound waste-handling strategies. The tool uses project-specific inputs, including project type and size, site constraints, finishing-stage waste-generation parameters, distances to CCCs and recycling facilities, and cost, emission, and carbon-pricing parameters (Greenstone et al., 2013; Çolak et al., 2016; Katsela et al., 2022).

The tool simulates baseline and CCC-enabled scenarios under identical case conditions and converts the outputs into comparable economic, environmental, and social KPIs. Assumptions and parameters remain explicit and adjustable, enabling transparent scenario testing prior to implementation (Melkonyan et al., 2022).

Case Studies and Results

Case Study Description

The decision-support framework was applied to three inner-city construction projects in the Netherlands, selected to represent varying degrees of spatial constraint while involving comparable finishing-stage logistics challenges. The cases include two projects in Utrecht and one in Amsterdam, all involving intensive finishing-stage activities under constrained site conditions (Tsegay et al., 2023).

The Renaissance project in Amsterdam represents a highly constrained inner-city context, characterised by extremely limited on-site storage space, strict access windows, and intensive interaction with surrounding urban functions. These conditions severely restrict conventional waste-handling practices during the finishing stage.

The Tivoli project in Utrecht is a heritage renovation located in a dense historic urban area. While spatial constraints are slightly less severe than in the Renaissance case, the project faces access regulations, heritage protection requirements, and restricted on-site container placement, imposing both spatial and regulatory limitations on waste-handling decisions.

The Beneluxlaan project in Utrecht serves as a comparatively less constrained urban case. Although still embedded within the urban fabric, the project benefits from improved access conditions and marginally greater on-site space, providing a contrast for assessing how strategy performance varies with the degree of spatial constraint.

Across all three cases, the finishing stage constitutes the primary focus of analysis due to its high waste-stream heterogeneity and operational complexity. For each project, two waste-handling strategies were evaluated: a baseline scenario representing conventional mixed-waste handling and an integrated scenario based on CCC-enabled reverse logistics.

Table 2: Characteristics of analysed case studies

Case	City	Project type	Site constraint	Finishing complexity
Renaissance	Amsterdam	Redevelopment	Very high	High
Tivoli	Utrecht	Heritage renovation	High	Medium–high
Beneluxlaan	Utrecht	New construction	Moderate	Medium

Economic Performance Results

Economic performance is evaluated based on project-total finishing-stage waste-handling costs for each case study. Cost components include container rental and handling, transport operations, sorting and labour activities, disposal fees, and revenues from recovered materials, consistent with the framework defined in the previous section. Absolute cost levels differ substantially across the three cases because the reported results reflect project totals rather than normalised values; they therefore support within-case scenario comparison more directly than cross-case benchmarking. Figure 2 illustrates the project-total cost difference between the baseline and integrated scenarios for each case.

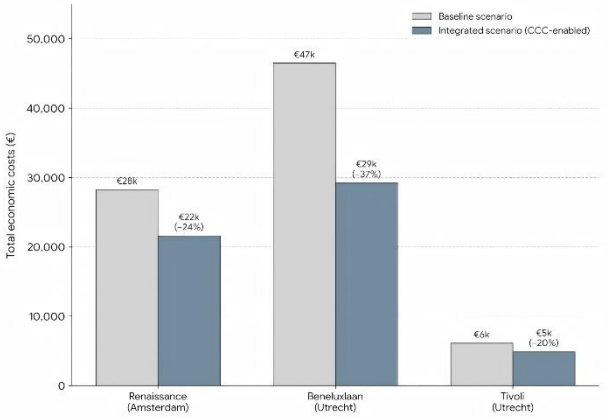


Figure 2: Project-total finishing-stage waste-management costs by case and scenario

In all three cases, the CCC-enabled scenario was less costly than the baseline mixed-waste handling strategy. Total cost reductions range from approximately 20% to 37%, depending on project characteristics and spatial constraints. Differences in absolute cost levels should be interpreted in relation to project-specific scale, finishing-stage scope, and site conditions rather than as like-for-like benchmarks between the three cases. The largest relative savings are observed in the most spatially constrained cases, where baseline logistics are characterised by high transport frequency and handling inefficiencies.

Cost breakdown analysis indicates that savings are primarily driven by reduced transport inefficiencies and increased revenues from material recovery. Although the integrated scenario introduces additional sorting and coordination activities, these costs are outweighed by efficiencies achieved through consolidated logistics operations. In contrast, the baseline scenario exhibits higher downstream disposal and sorting costs due to the treatment of mixed waste streams.

Overall, the economic benefits of CCC-enabled reverse logistics were strongest in cases where limited space and inefficient mixed-waste treatment made the baseline arrangement particularly costly.

Environmental Performance Results

Environmental performance is evaluated based on project-total transport- and mixed-waste-sorting-related CO_2 emissions associated with finishing-stage waste handling (Butera et al., 2015; Bayram and Greiff, 2023). These emissions are calculated consistently across the baseline and integrated scenarios using the parameters defined in the decision-support framework and are intended primarily for within-case comparison rather than cross-case benchmarking. Figure 3 illustrates the difference in project-total transport- and sorting-related CO_2 emissions between the two scenarios in each case.

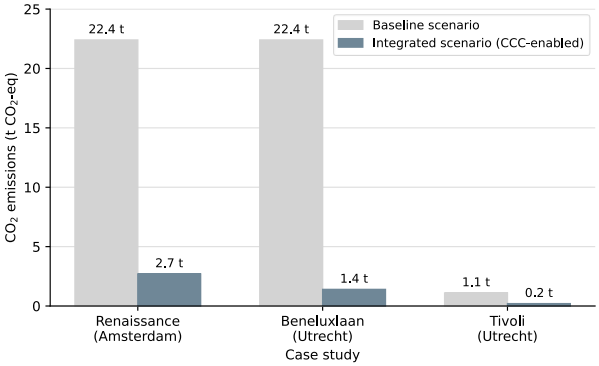


Figure 3: Project-total transport- and sorting-related CO_2 emissions by case and scenario

In all three cases, the integrated scenario produced substantially lower transport- and sorting-related CO_2 emissions than the baseline, with total reductions of roughly 85% to 94%. The dominant contributor to this reduction is the avoidance of energy-intensive downstream sorting processes associated with mixed-waste handling.

Transport-related emissions are also lower in the integrated scenario due to coordinated inbound and outbound logistics enabled by the CCC. Although logistics operations may occur more frequently, improved vehicle utilisation and reduced reliance on dedicated waste transport trips lead to lower overall emissions within the defined project-level assessment boundary (Wu et al., 2023).

Environmental performance gains are particularly pronounced in projects where baseline practices rely heavily on mixed containers and downstream mechanical sorting. In such contexts, source separation enabled through CCC-based reverse logistics offers significant emission avoidance potential.

Social Impact Results

Social impacts are evaluated using the number of vehicle trips as a proxy indicator for urban disturbance, including traffic congestion, noise, and safety risks in dense urban environments (Li et al., 2024).

In contrast to the economic and environmental results, the integrated scenario is associated with a moderate increase in modelled vehicle trips, ranging from approximately 8% to 15% across the three cases. This increase is primarily linked to the more frequent collection of smaller, source-separated waste containers during the finishing stage. Figure 4 shows the modelled number of vehicle trips under the baseline and integrated scenarios for each case.

However, this increase should be interpreted with care. Under the integrated scenario, waste collection is coordinated with inbound material deliveries, meaning that some outbound waste movements may be combined with return legs that would otherwise occur without payload. The vehicle-trip indicator should therefore be understood as an operational proxy reflecting handling rhythm and logistics intensity, rather than as a direct count of net-new urban freight movements. In the baseline scenario, fewer trips may be observed, but these are typically less coordinated and may involve less efficient vehicle utilisation and longer on-site dwell times.

This points to a clear trade-off: the integrated scenario improves cost and emissions performance, but it can also increase the frequency of on-site waste-handling operations. This trade-off highlights the importance of coordination between project-level waste-handling decisions and broader urban logistics planning in highly constrained inner-city environments.

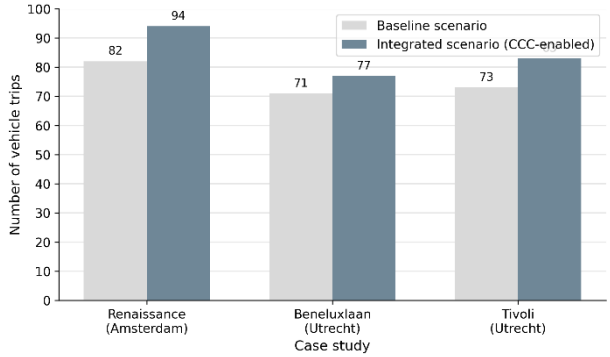


Figure 4: Number of vehicle trips associated with outbound C&DW logistics under baseline and CCC-enabled scenarios

Sensitivity Analysis

To test the robustness of the case-study findings, a threshold-based sensitivity analysis was conducted, focusing on dominant economic and project-constraint parameters. Specifically, we identified (i) the break-even mixed-waste recycling cost at which the baseline strategy becomes economically comparable to the CCC-enabled scenario, and (ii) estimated tipping points for key project constraints, including project size and available on-site storage area, at which the relative performance of the integrated strategy changes.

Across the analysed cases, the CCC-enabled scenario remains environmentally favourable across the tested ranges, while economic performance is more sensitive to market and constraint factors. As summarised in Table 3, the baseline strategy would require a substantial reduction in mixed-waste recycling costs (23.6–30.7%) to become economically comparable to the integrated scenario. In addition, the viability of the integrated strategy is shaped by project scale and available on-site staging capacity, indicating that CCC-enabled reverse logistics should be interpreted as a context-sensitive rather than universally optimal solution.

The thresholds reported in Table 3 are case-specific and should therefore be interpreted as indicative break-even conditions rather than transferable benchmark values.

Table 3: Case-specific sensitivity thresholds: indicative break-even recycling cost and project-constraint tipping points

Case	Break-even recycling cost (€/m³)	Required reduction (%)	Economic scale threshold (m²)	Environmental scale threshold (m²)	On-site storage threshold (m²)
Renaissance	147	25.1	1880	900	100
Tivoli	150	23.6	1900	1850	20
Beneluxlaan	136	30.7	1800	900	25

Discussion

Scalability and Transferability

The framework should not be read as a source of universally applicable threshold values. Its value lies instead in offering a comparison logic that can be adapted to projects with different spatial constraints. In this study, these constraints are operationalised mainly through on-site staging capacity and accessibility conditions, which shape feasible container configurations, handling frequency, and collection feasibility in the baseline and integrated strategies.

Across the three cases, performance differences are driven by how these constraints interact with project-specific scale and duration. Tighter sites restrict material-specific containers and increase swap frequency, whereas less constrained sites allow more stable source separation. The resulting cost and emission outcomes should therefore be interpreted as contingent on local spatial and logistical conditions rather than fixed benchmarks.

Even so, the analytical structure can still be applied to other projects, provided that case-specific inputs are retained. By combining scenario-based comparison with a consistent KPI set, the framework supports ex ante assessment across projects while allowing case-specific inputs to vary.

Sensitivity and Key Enablers

The sensitivity analysis shows that the economic performance of CCC-enabled reverse logistics depends mainly on a small number of market and project-constraint variables. As summarised in Table 3, its economic advantage is particularly sensitive to downstream mixed-waste treatment conditions: the

baseline strategy would require a substantial reduction in mixed-waste recycling costs (approximately 23.6–30.7%) to become economically comparable to the CCC-enabled scenario. This indicates that, under typical market conditions, the economic case for integrated reverse logistics is relatively robust, although it may weaken under unfavourable pricing environments (Wu et al., 2023).

Project scale constitutes a second decisive factor. The tipping-point analysis (Table 3) indicates scale thresholds around 1800–1900 m², below which the fixed coordination requirements of CCC-enabled logistics are less likely to be amortised. Smaller or short-duration projects may therefore struggle to realise economic benefits despite potential emission reductions. The analysis further shows that on-site staging capacity acts as a practical feasibility constraint: where available storage space falls below the identified thresholds, the achievable degree of source separation and the operational rhythm of container swaps may be restricted.

Beyond these quantitative sensitivities, institutional and organisational conditions remain key enablers. Effective implementation depends on coordination among contractors, CCC operators, and municipal authorities, particularly with respect to delivery windows, vehicle access regulations, and on-site handling arrangements. In addition, mechanisms such as carbon pricing may strengthen the relative performance of integrated logistics by internalising transport-related emissions that are otherwise excluded from project-level decision-making (Katsela et al., 2022).

CCC-enabled reverse logistics is therefore better understood as a conditional solution: it performs well in some inner-city settings, but not under all project conditions. The proposed framework provides a structured means to identify when integrated outbound waste logistics constitutes a robust and defensible choice under real inner-city constraints.

Limitations

This study has several limitations. It is based on three Dutch case studies, limiting geographical and regulatory generalisability. It provides a project-level, ex ante comparison of finishing-stage outbound waste logistics rather than a full cradle-to-grave LCA/LCC. Social impacts are approximated through vehicle trips, and mixed-waste sorting emissions and other operational parameters rely partly on secondary assumptions. Results are therefore most robust for within-case comparison rather than direct cross-case benchmarking.

Conclusions

This study developed a framework for comparing outbound C&DW strategies during the finishing stage of inner-city construction projects. Using a baseline mixed-waste scenario and a CCC-enabled reverse-logistics scenario, it examines how cost, emissions, and urban disturbance change under different site conditions.

The case-based comparisons suggest that CCC-enabled reverse logistics may improve cost and emission performance under the analysed conditions, while also increasing logistics intensity in some cases. The main contribution is not a universal prescription, but a practical way to judge when integrated outbound waste logistics is feasible and worthwhile under specific spatial and organisational conditions.

Future research

Future research could extend the framework in several directions. BIM-based waste estimation could improve the automation and accuracy of project-level waste forecasting (Guerra et al., 2019).

Real-time logistics data could support more dynamic decision-making during project execution (Brusselaers et al., 2024).

In addition, coupling the tool with urban transport simulation models may help capture network-level effects, while expanding the social assessment could account for worker-related impacts such as handling burden, safety, and usability. Together, these extensions would improve the framework's relevance for circular decision-making in dense urban construction contexts.

References

Bayram, B., Greiff, K., 2023. Life cycle assessment on construction and demolition waste recycling: a systematic review analyzing three important quality aspects. Int. J. Life Cycle Assess. 28, 967–989. https://doi.org/10.1007/s11367-023-02145-1

Brusselaers, N., Fredriksson, A., Gundlegård, D., Zernis, R., 2024. Decision support for improved construction traffic management and planning. Sustain. Cities Soc. 104, 105305. https://doi.org/10.1016/j.scs.2024.105305

Butera, S., Christensen, T.H., Astrup, T.F., 2015. Life cycle assessment of construction and demolition waste management. Waste Manag. 44, 196–205. https://doi.org/10.1016/j.wasman.2015.07.011

Çolak, S., Lima, A., González, M.C., 2016. Understanding congested travel in urban areas. Nat. Commun. 7, 10793. https://doi.org/10.1038/ncomms10793

Finnveden, G., Hauschild, M.Z., Ekvall, T., Guinée, J., Heijungs, R., Hellweg, S., Koehler, A., Pennington, D., Suh, S., 2009. Recent developments in Life Cycle Assessment. J. Environ. Manage. 91, 1–21. https://doi.org/10.1016/j.jenvman.2009.06.018

Gálvez-Martos, J.-L., Styles, D., Schoenberger, H., Zeschmar-Lahl, B., 2018. Construction and demolition waste best management practice in Europe. Resour. Conserv. Recycl. 136, 166–178. https://doi.org/10.1016/j.resconrec.2018.04.016

Greenstone, M., Kopits, E., Wolverton, A., 2013. Developing a Social Cost of Carbon for US Regulatory Analysis: A Methodology and Interpretation. Rev. Environ. Econ. Policy 7, 23–46. https://doi.org/10.1093/reep/res015

Guerlain, C., Renault, S., Ferrero, F., 2019. Understanding Construction Logistics in Urban Areas and Lowering Its Environmental Impact: A Focus on Construction Consolidation Centres. Sustainability 11, 6118. https://doi.org/10.3390/su11216118

Guerra, B.C., Bakchan, A., Leite, F., Faust, K.M., 2019. BIM-based automated construction waste estimation algorithms: The case of concrete and drywall waste streams. Waste Manag. 87, 825–832. https://doi.org/10.1016/j.wasman.2019.03.010

Haag, P., Jünger, H.C., 2023. Turning a spotlight on construction logistics for a sustainable urban environment—a review of current policy concepts and literature. Front. Built Environ. 9. https://doi.org/10.3389/fbuil.2023.1202091

Hossain, Md.U., Wu, Z., Poon, C.S., 2017. Comparative environmental evaluation of construction waste management through different waste sorting systems in Hong Kong. Waste Manag. 69, 325–335. https://doi.org/10.1016/j.wasman.2017.07.043

Katsela, K., Pålsson, H., Ivernå, J., 2022. Environmental impact and costs of externalities of using urban consolidation centres: a 24-hour observation study with modelling in four scenarios. Int. J. Logist. Res. Appl. 25, 1542–1563. https://doi.org/10.1080/13675567.2021.1915261

Li, C., Wang, D., Chen, H., Liu, E., 2024. Analysis of Urban Congestion Traceability: The Role of the Built Environment. Land 13, 255. https://doi.org/10.3390/land13020255

Maxner, T., Dalla Chiara, G., Goodchild, A., 2025. The State of Sustainable Urban Last-Mile Freight Planning in the United States. J. Am. Plann. Assoc. 91, 88–101. https://doi.org/10.1080/01944363.2024.2324096

Melkonyan, A., Gruchmann, T., Lohmar, F., Bleischwitz, R., 2022. Decision support for sustainable urban mobility: A case study of the Rhine-Ruhr area. Sustain. Cities Soc. 80, 103806. https://doi.org/10.1016/j.scs.2022.103806

Muñuzuri, J., Larrañeta, J., Onieva, L., Cortés, P., 2005. Solutions applicable by local administrations for urban logistics improvement. Cities 22, 15–28. https://doi.org/10.1016/j.cities.2004.10.003

Paddeu, D., 2018. Sustainable Solutions for Urban Freight Transport and Logistics: An Analysis of Urban Consolidation Centers, in: Zeimpekis, V., Aktas, E., Bourlakis, M., Minis, I. (Eds.), Sustainable Freight Transport: Theory, Models, and Case Studies. Springer International Publishing, Cham, pp. 121–137. https://doi.org/10.1007/978-3-319-62917-9_8

Rubio, S., Jiménez-Parra, B., Chamorro-Mera, A., Miranda, F.J., 2019. Reverse Logistics and Urban Logistics: Making a Link. Sustainability 11. https://doi.org/10.3390/su11205684

Russo, S.M., Voegl, J., Hirsch, P., 2021. A multi-method approach to design urban logistics hubs for cooperative use. Sustain. Cities Soc. 69, 102847. https://doi.org/10.1016/j.scs.2021.102847

Sobotka, A., Sagan, J., 2021. Decision support system in management of concrete demolition waste. Autom. Constr. 128, 103734. https://doi.org/10.1016/j.autcon.2021.103734

Tsegay, F.G., Mwanaumo, E., Mwiya, B., 2023. Construction site layout planning practices in inner-city building projects: space requirement variables, classification and relationship. Urban Plan. Transp. Res. 11, 2190793. https://doi.org/10.1080/21650020.2023.2190793

Wang, J., Yuan, H., Kang, X., Lu, W., 2010. Critical success factors for on-site sorting of construction waste: A china study. Resour. Conserv. Recycl. 54, 931–936. https://doi.org/10.1016/j.resconrec.2010.01.012

Wu, F., Mei, S., Xu, H., Hsu, W.-L., 2023. Urban Construction Waste Recycling Path: Robust Optimization. Buildings 13, 2802. https://doi.org/10.3390/buildings13112802

Yuan, H., 2013. Key indicators for assessing the effectiveness of waste management in construction projects. Ecol. Indic. 24, 476–484. https://doi.org/10.1016/j.ecolind.2012.07.022

Zhao, P., Md Ali, Z., Ahmad, Y., 2023. Developing indicators for sustainable urban regeneration in historic urban areas: Delphi method and Analytic Hierarchy Process (AHP). Sustain. Cities Soc. 99, 104990. https://doi.org/10.1016/j.scs.2023.104990

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DATA DRIVEN DECISION SUPPORT FRAMEWORK FOR PRODUCT LEVEL REUSE PRIORITISATION AND BARRIER DIAGNOSIS

Qi Han, Jintian Xu, Dujuan Yang, T. van Tongeren
Eindhoven University of Technology, The Netherlands

Abstract

This paper proposes a three-stage workflow to prioritise reusable building product groups in Dutch demolition outflows, addressing the gap between impact potential and implementation feasibility. Stage 1 screens 139 groups using an impact proxy combining mass outflow and the environmental cost indicator, shortlisting ten candidates. Stage 2 applies a five-criteria multi-criteria analysis with weights from eight experts using the Best–Worst Method, where supply–demand balance dominates (0.296) and sustainability ranks lowest (0.137). Stage 3 links ranking to barrier diagnosis through a masonry case, identifying contamination, certification, and coordination challenges. The workflow supports identifying high priority product groups for early-stage reuse planning.

Introduction

Problem context

The construction sector remains a major contributor to material demand and embodied environmental impacts, which makes improving circular performance a priority in the transition toward a circular economy (Kirchherr et al., 2017). Yet, reuse implementation in practice remains limited, particularly at the demolition stage where large volumes of potentially reusable products leave the building stock and are often downcycled or disposed of (Arora et al., 2020).

This study focuses on demolition outflows from Dutch residential and utility buildings (B&U). In this context, environmental relevance is quantified computationally as mass outflow × MKI, where MKI (Milieu Kosten Indicator) is a Dutch aggregated environmental cost indicator and impact values are operationalised using the A1 indicator set as implemented in the national database (Nationale Milieudatabase (NMD), 2025). This framing enables a consistent comparison of heterogeneous product groups by combining (i) how much material becomes available at the end-of-life and (ii) to what extent environmentally relevant that material is under the same indicator system.

However, identifying environmentally relevant product groups does not automatically translate into reuse uptake. Practical decision-making is often constrained by fragmented information about what leaves the building stock, under what conditions, and whether reuse is feasible, timely, and economically viable within real project constraints (Bellini et al., 2024). This results in a persistent gap between circular ambition and operational reuse uptake in the built environment (Planbureau voor de Leefomgeving (PBL), 2025).

Related work and research gap

This section focuses on decision-support approaches for product-level reuse prioritisation in the built environment, rather than reviewing circular economy literature more broadly.

Research has extensively documented barriers to circular construction and product reuse, including technical constraints, market uncertainty, and institutional and behavioural factors (Kanters, 2020; Wuni, 2022). Two limitations are particularly relevant for this study.

First, high environmental relevance does not imply high reuse feasibility. Product groups with high mass outflow × MKI can still face major implementation barriers such as contamination, certification requirements, liability concerns, and misaligned supply and demand (Giorgi et al., 2022). This implies that computational decision support must separate “impact potential” from “implementability” rather than assuming they overlap (De Wolf et al., 2024).

Second, existing approaches often lack an explicit, reproducible computational pipeline that turns heterogeneous demolition outflow data and expert knowledge into transparent prioritisation outputs. In particular, reuse prioritisation requires combining impact-based signals with feasibility-related dimensions, such as cost barriers, risk barriers, supply–demand balance, and technical feasibility, rather than relying on impact alone (Yu et al., 2023). To address this, the study applies an MCA structure with five criteria—cost, risk, supply–demand balance, technical feasibility, and sustainability impact—and derives criterion weights from eight expert interviews using the Best–Worst Method (BWM) (Rezaei, 2015).

This means that environmental relevance alone is not enough to identify reuse potential. Instead, prioritisation requires explicitly accounting for implementation-oriented criteria such as cost, risk, technical feasibility, and supply–demand balance alongside impact-based indicators.

Computing-oriented circularity research in the built environment increasingly treats the existing building

stock as a quantifiable "secondary resource" that can be measured, forecast, and prioritised through data pipelines rather than narrative assessments (Arora et al., 2021). At the city/building-stock scale, computational approaches typically combine stock modelling with environmental accounting to estimate future availability of recoverable components and associated impact reduction opportunities (Röck et al., 2021). Such pipelines support "where-to-focus" decisions by identifying high-leverage materials/components before stakeholders invest in detailed feasibility studies (Wuyts et al., 2022).

However, much of the stock/potential modelling literature remains oriented toward estimating quantity and theoretical circularity potential, while practical implementation still depends on socio-technical constraints (e.g., dismantling feasibility, market absorption, regulatory permission) (Röck et al., 2021). This calls for a workflow that links data-driven screening to decision criteria such as cost, risk, technical feasibility, and market balance, instead of assuming that high potential also means easy implementation (Van Den Berg et al., 2023).

Multi-Criteria Analysis (MCA) is widely used in construction and demolition contexts because circular decisions are inherently multi-dimensional: environmental benefits must be assessed alongside cost, risk, feasibility, supply chains, and stakeholder acceptance (Banias et al., 2010; Bajwa et al., 2025). Methodologically, MCA offers a transparent aggregation structure—criteria, weights, and scores—that can be audited and updated as assumptions or stakeholder inputs change (Bajwa et al., 2025). Within circular construction research, barrier- and driver-oriented studies often operationalise implementation constraints into evaluative criteria, creating a bridge between qualitative insights and decision models (Kanters, 2020).

In C&D waste and circular practice assessments, MCA has been applied to compare strategies, prioritise interventions, and structure complex trade-offs across stakeholders (Swarnakar and Khalfan, 2024). Yet, a recurring limitation is that MCA is frequently applied after a case or material is selected, rather than being computationally coupled with an upstream, data-driven filtering step that justifies why a given product group deserves priority attention from a system-wide list (Wuyts et al., 2022).

A central methodological challenge in MCA is how to elicit reliable criterion weights, especially when expert time is limited and decision contexts are complex (Peykani et al., 2026). The Best–Worst Method (BWM) has been proposed as a robust multi-criteria weighting approach that reduces the number of pairwise comparisons while improving consistency (Rezaei, 2015). Compared to more comparison-heavy weighting methods, BWM structures expert judgement around two anchors (best and worst criteria), which is easier to communicate and reproduce in a computational workflow (Rezaei, 2016).

For computing-in-construction venues, the relevance of BWM is not only that it yields weights, but that it supports an auditable pipeline: expert inputs can be stored, checked for consistency, and re-run when either the criteria set or the stakeholder group changes, enabling iterative refinement and sensitivity testing (Peykani et al., 2026).

Combine a quantity factor (e.g., mass outflow or material availability) with an impact factor (e.g., environmental burden per unit) to identify candidate materials/components with both sufficient scale and meaningful sustainability leverage (Arora et al., 2020). This logic aligns with urban mining perspectives that frame cities and building stocks as measurable repositories whose recovery potential can be quantified and prioritised (Ghisellini et al., 2022). Building-stock studies further show that "future supply" assessments become more actionable when they are resolved into component/product categories that map onto real recovery processes and market channels (Orenga Panizza and Nik-Bakht, 2024).

In the Dutch context, impact-based filtering can leverage the national environmental cost indicator (MKI) and associated databases/tools to operationalise environmental relevance in a standardised way (Nationale Milieudatabase (NMD), 2025). Sectoral material-flow datasets provide product-group resolution for construction-related inflows/outflows, enabling computational shortlisting across many product groups before deeper feasibility analysis begins (EIB; Metabolic, 2022). The key methodological opportunity is to connect stock flow quantification with decision support prioritisation in one framework, rather than treating them as separate analytical exercises (Wuyts et al., 2022).

Two research gaps are especially relevant here. First, urban mining and stock-flow studies provide useful system-level signals on material availability and environmental impact, but they rarely translate these signals into prioritisation approaches that reflect practical feasibility, such as cost, risk, technical constraints, and supply–demand balance (Röck et al., 2021). Second, barrier and driver studies explain why reuse remains difficult, but they are less often incorporated into an explicit decision workflow. In practice, this means that few studies combine quantitative screening, weighted multi-criteria ranking, and structured barrier diagnosis in a single approach that can support intervention design (Wuni, 2022).

To address these gaps, the study treats reuse prioritisation as a decision-support task that combines impact screening, MCA-based ranking, and barrier diagnosis: product groups are first filtered using impact-based signals, then prioritised using MCA with explicit weighting, and finally connected to a demonstrative barrier diagnosis to show how quantitative ranking can guide targeted implementation strategies (Van Den Berg et al., 2023).

Existing reuse metrics and stock-based assessments are useful for identifying material availability, environmental relevance, or theoretical recovery potential, but they often

stop before implementation conditions are examined. By contrast, this study does not propose another single reuse score. It combines impact-based screening with implementation-oriented multi-criteria ranking and a case-based barrier diagnosis. The key distinction is therefore methodological rather than purely metric-based: the workflow is designed to move from environmental relevance to ranked action candidates and then to an explanation of why high-potential product groups may still face barriers in practice.

Research question

The study addresses the following main research question:

RQ: How can high-potential product groups for reuse be computationally identified from Dutch B&U demolition outflows using mass outflow × MKI, and how can key implementation barriers be diagnosed to inform targeted circular interventions?

Contributions

This study makes three contributions. First, it develops a three-stage workflow for product-level reuse prioritisation that combines impact-based filtering, weighted multi-criteria analysis, and case-based barrier diagnosis. Second, it operationalises public Dutch demolition-flow and environmental-impact data together with structured expert elicitation, making the ranking logic explicit and updateable when new data become available. Third, it shows how ranking outputs can support intervention-oriented diagnosis rather than stopping at score comparison alone, using masonry brickwork as an illustrative case.

Methodology

Research design

This study uses a mixed-method design to identify, rank, and further examine reusable product groups in Dutch demolition outflows. The design combines public product-group flow data, environmental impact data, and two rounds of semi-structured interviews. The first interview round supports Stage 2 by eliciting implementation-oriented judgement for MCA weighting and selected product-level scoring. The second round supports Stage 3 by examining how the highest-ranked product group performs in practice through stakeholder interviews along the brick value chain. This design was chosen because product-level reuse cannot be assessed from environmental relevance alone; it also requires evidence on implementation conditions, actor perspectives, and sector-specific bottlenecks.

Workflow overview

The study develops a three-stage workflow to identify, rank, and further examine reusable product groups in Dutch demolition outflows. The workflow combines impact-based screening, multi-criteria ranking, and barrier diagnosis to support product-level comparison under real implementation constraints (Swarnakar and Khalfan, 2024).

In this paper, ‘computational’ means that product groups are assessed through explicit indicators, weighted aggregation, and reproducible ranking rules (Papamichael et al., 2023).

As illustrated in Figure 1, the framework consists of three sequential stages. First, an impact-based filtering stage is applied to identify product groups with high environmental relevance based on demolition-related material outflows and environmental impact indicators (Atta and Bakhoum, 2024; Nationale Milieudatabase (NMD), 2025). Second, a multi-criteria prioritisation stage evaluates the shortlisted product groups by explicitly accounting for economic, technical, and supply–demand considerations using a structured decision-analysis method (Rezaei, 2015). Third, a case-based barrier diagnosis is conducted for the highest-ranked product group to demonstrate how the framework supports the identification of dominant implementation barriers (Giorgi et al., 2022; Wuni, 2022).

Each stage uses explicit indicators and scoring rules. This makes the workflow easier to reproduce and adapt to other datasets (Yang et al., 2023). The framework is not intended to optimise a single material or project case, but to support early-stage circular construction planning under data constraints (Papamichael et al., 2023).

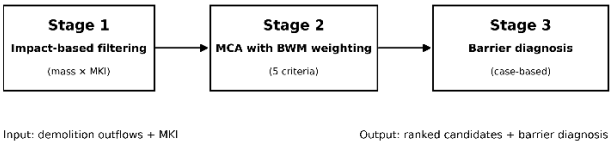


Figure 1: Overview of the computational prioritisation workflow

Stage 1 – impact-based candidate filtering (data-driven)

The first stage aims to reduce the initial set of demolition-derived product groups to a manageable number of candidates that are environmentally relevant for reuse considerations (Wuyts et al., 2022). The analysis is based on a dataset comprising 139 distinct product groups identified within the Dutch residential and utility building stock (EIB; Metabolic, 2022; Nationale Milieudatabase (NMD), 2025).

Each product group is characterised by two quantitative indicators:

(i) the annual mass outflow generated during demolition activities, and

(ii) an aggregated environmental impact indicator, expressed as the Environmental Cost Indicator (Milieu Kosten Indicator, MKI), which reflects life-cycle environmental burdens (Nationale Milieudatabase (NMD), 2025).

An impact score is derived by combining these two indicators, enabling the identification of product groups that exhibit both substantial material quantities and high environmental relevance (EIB; Metabolic, 2022). To

avoid bias towards a single material category, the selection procedure includes both overall high-impact product groups and the highest-ranking representatives within major material classes (EIB; Metabolic, 2022).

As a result, the initial set of 139 product groups is reduced to a shortlist of ten high-impact candidates, which are subsequently evaluated in greater detail in the following stage. The shortlist combines the highest-impact product groups with category representatives to avoid a shortlist dominated by a single material family.

Stage 2 – multi-criteria prioritisation

While the first stage identifies product groups with high environmental relevance, it does not capture constraints related to practical implementation (Byers et al., 2024). Therefore, the second stage applies a multi-criteria decision analysis (MCA) to assess the feasibility of reuse across the shortlisted product groups (EIB; Metabolic, 2022).

Five evaluation criteria are defined to capture key dimensions influencing reuse implementation:

(1) economic barriers,

(2) technical feasibility,

(3) sustainability/impact,

(4) risks related to uncertainty and liability, and

(5) supply–demand balance.

Each product group is scored on a five-point ordinal scale for each criterion using evidence from the literature and expert judgement (Peykani et al., 2026). Criterion weighting and product-level scoring served different purposes. Weighting was used to capture broader implementation priorities across the five reuse criteria, whereas product-level scoring of cost and risk required more specific familiarity with particular product groups. To reflect the relative importance of the criteria, weights were derived using the Best–Worst Method (BWM), which provides a structured form of preference elicitation (Rezaei, 2016).

Criterion weights were obtained from eight semi-structured interviews with relatively independent experts in circular construction and reuse. The interviewees were selected to provide cross-sector perspectives rather than product-specific advocacy. In the BWM step, each expert identified the most important and least important criterion and then compared the remaining criteria to these two anchors on a 1–9 scale. Individual weights were calculated and averaged to obtain the final weight set. Product-level scoring of cost and risk was restricted to interviewees who indicated sufficient familiarity with specific product groups, to avoid speculative ratings at that level. To test robustness, the weighting step was recalculated by excluding one expert at a time. This allowed the ranking to be checked for sensitivity to individual expert input.

The overall prioritisation score for each product group is calculated using a weighted additive model:

$$S_i = \sum_{j=1}^{n} w_j \cdot x_{ij}$$

where S_i denotes the aggregated score of product group i, w_j represents the weight of criterion j, and x_{ij} is the normalised score of product group i under criterion j (Jato-Espino et al., 2014).

This procedure allows different product groups to be compared using the same logic, while keeping environmental relevance separate from implementation feasibility (Byers et al., 2024). Criterion weighting used all eight expert interviews, whereas product-level scoring of cost and risk was restricted to interviewees who indicated sufficient familiarity with specific product groups.

Stage 3 – barrier diagnosis

Stage 3 examines why the product group ranked highest in Stage 2 may still face implementation bottlenecks in practice. Brick masonry is used as the demonstration case. A qualitative analysis is conducted to examine its value-chain structure and identify dominant barriers affecting reuse implementation. The diagnosis draws on five stakeholder interviews across the brick value chain and uses literature-derived barrier categories as a coding frame to identify constraints related to coordination, certification, contamination control, and process organization (Wuni, 2022). The purpose of Stage 3 is analytical explanation rather than statistical validation: it links the ranking output to observed implementation barriers and shows how prioritisation results can support intervention-oriented diagnosis (Rakhshan et al., 2020; Wuyts et al., 2022).

Data sources and reproducibility

The workflow combines public quantitative datasets with interview-based expert judgement. Product-group outflow data were taken from the EIB/SGS-Search mass-flow dataset for Dutch residential and utility buildings, while environmental impact values were retrieved from the Dutch National Environmental Database (NMD) using MKI-based product matches. Data cleaning, product-group merging, thresholding, and impact calculations were implemented in Microsoft Excel using explicit rules and formula-based operations. Public input datasets, decision rules, and scoring procedures are described in the methodology. Interview responses are not reported in full because of ethics and anonymity requirements, but the interview protocol, weighting procedure, and sensitivity checks are summarised in the paper.

Evaluation logic

The framework is evaluated as a demonstration rather than as a full external validation exercise. Its usefulness is assessed in three ways: first, whether Stage 2 meaningfully reorders the Stage 1 shortlist once implementation-oriented criteria are introduced; second, whether the weighting results remain stable under leave-one-out recalculation; and third, whether the masonry

case reveals barriers that are not visible in impact screening alone. Together, these checks show whether the workflow produces distinct, interpretable, and implementation-relevant outputs.

Results

Impact screening results

This section presents the results from the three stages of the workflow. Stage 1 yields a shortlist of high-impact product groups, Stage 2 produces a reuse-oriented prioritisation ranking, and Stage 3 uses masonry brickwork for demonstration only to illustrate the quantitative–qualitative linkage between computational prioritisation and barrier diagnosis (Hopkinson et al., 2019).

After applying the predefined merging rules, the initial product-group list was reduced from 139 to 115 groups. Candidate filtering first applied a mass-flow threshold, retaining groups with relative mass outflow ≥ 0.5%; groups in the 0.5–0.1% band were re-examined for possible merging effects, but no additional merges qualified. The average relative mass outflow of the Top-5 groups was more than 29× that of the 24th group, supporting the assumption that contributions below the threshold can be neglected. For shortlisted groups, environmental impact was computed as mass outflow × MKI, where MKI values were retrieved from the Dutch National Environmental Database (NMD) and multiplied by the corresponding outflow mass to obtain an impact proxy in euros (€). To avoid dominance by stone-like products in the Top-5, the shortlist also includes the highest-impact representatives from steel, wood, glass, mixed and plastics; see Figure 2.

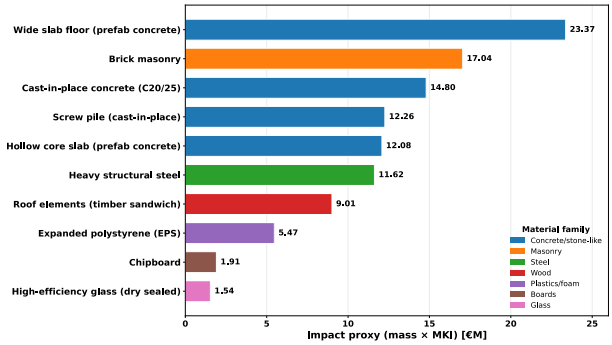


Figure 2: High-impact candidate groups based on impact proxy (mass × MKI)

MCA ranking results

Stage 2 applied MCA to the ten shortlisted product groups. Expert elicitation indicated that supply–demand balance was the most influential criterion (average weight 0.296), more than twice the weight assigned to sustainability/impact (average weight 0.137), underscoring the role of market viability in enabling reuse. This weighting structure provides quantitative evidence that implementation readiness can dominate environmental relevance in computational reuse prioritisation. While all experts contributed to criteria weighting, only a subset provided sufficiently confident and consistent product-level scores for cost and risk; therefore, product scoring draws on those inputs, while the final aggregation uses the combined weights across experts. The resulting MCA ranking shows a clear top tier: brick masonry achieved the highest total score (4.04/5), followed by heavy structural steel (3.29/5), while the third-ranked option (hollow core slabs) scored notably lower (2.78/5). A saturation test indicated that the Top-2 positions remain stable under different expert combinations; see Figure 3.

All eight experts were included in the criterion-weighting step, as they indicated sufficient confidence in judging the relative importance of the five MCA criteria. However, only four experts were retained for product-level scoring of cost and risk. The remaining interviewees reported insufficient familiarity with specific product groups to provide reliable comparative ratings at that level. This separation was intentional: criterion weighting captured broader implementation priorities, whereas product-level cost and risk scoring required more detailed knowledge of particular materials and products.

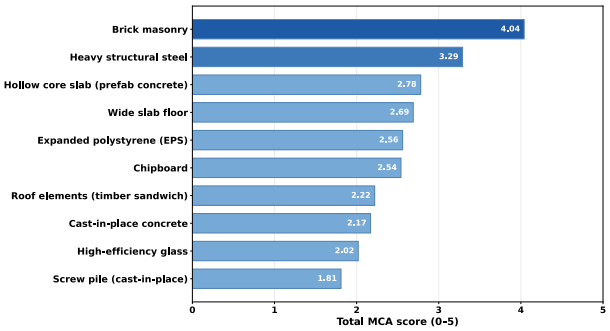


Figure 3: MCA-based reuse prioritisation ranking

Barrier diagnosis results

Although heavy structural steel ranks highly, its end-of-life value retention and established recycling/reuse practices suggest a smaller incremental knowledge gap. By contrast, masonry brickwork is typically downcycled (e.g., crushed for road base), indicating a larger opportunity for value retention if direct reuse can be enabled. Together with its top MCA score and supporting qualitative insights, masonry was selected as the demonstration product group.

To reduce over-interpretation of absolute values, Table 1 is presented as an illustrative comparison. Using estimates reported for an industry tool, direct reuse of brickwork yields an MKI reduction of approximately €24/t, whereas steel reuse yields about €60/t (general reuse) or €80/t (reuse as structural material). However, given masonry's much larger mass flow, the system-level impact of increasing brick reuse is estimated to be 2–3× higher than that of heavy structural steel when scaled by national outflow volumes.

Table 1: Illustrative MKI savings and potential impact reduction (brick vs. steel)

Product	MKI (€/t)	MKI saving (€/t)	Current impact (M€)	Potential impact saving (M€)	Relative savings (%)
Heavy structural steel	120	60	11.62	5.81	50.0
Heavy structural steel (structural reuse)	120	80	11.62	7.74	66.7
Brick masonry	26.3	24	17.04	15.55	91.3

The demonstration barrier diagnosis indicates that the most constraining bottlenecks for brick reuse relate to cleaning and contamination control (including mortar separation), quality consistency and certification, and the time/coordination costs associated with selective deconstruction and matching supply to demand. Corresponding leverage points include selective deconstruction to increase recoverable quality, standardised quality control and certification schemes, and reorganising testing/grading processes across supply-chain nodes to reduce schedule conflicts (Knoth et al., 2022). Based on these constraints, the paper outlines possible intervention directions, such as standardised quality assessment and improved demolition–demand coordination, to show how the ranking results can support practical action.

Discussion

The results indicate that product-group prioritisation for reuse cannot be inferred from environmental relevance alone. While the Stage 1 shortlist is dominated by high-mass, stone-like products, Stage 2 shows that implementation-oriented criteria—especially supply–demand balance—can substantially reorder priorities. This is reflected by masonry brickwork ranking first despite comparatively weaker scores on cost and technical feasibility, suggesting that market fit can outweigh near-term process constraints in determining reuse readiness (Wuni, 2022). This aligns with broader observations that circular implementation is shaped by interacting economic, technical and institutional drivers and barriers rather than single-factor optimisation (Tura et al., 2019). Importantly, the dominance of supply–demand balance over environmental impact in the weighting structure illustrates that reuse prioritisation should not be treated as a single-criterion screening exercise. This suggests that reuse potential should be assessed through both environmental relevance and implementation conditions, rather than environmental relevance alone.

The value of the workflow lies not only in identifying high-impact candidates, but in showing how rankings change once implementation conditions are introduced. In this study, the shift from Stage 1 to Stage 2 demonstrates that environmental relevance alone is not sufficient for prioritisation, while Stage 3 shows why even highly ranked options may still face coordination, certification, or contamination-related barriers. In this sense, the paper offers a proof-of-concept that ranking and diagnosis provide more decision-relevant insight than impact screening alone.

Implications for practice

The barrier configuration observed for brick reuse suggests that scaling requires interventions that simultaneously address (i) quality assurance and certification, (ii) time alignment between demolition and new-build demand, and (iii) process innovations for cleaning and contamination control. Existing reuse toolkits and sector protocols can support standardisation efforts, but their effectiveness depends on embedding them in procurement and early planning routines (SGS Search; Cirkelstad, 2021). In parallel, digital marketplaces and matchmaking platforms may reduce supply–demand uncertainty by enabling earlier visibility of upcoming demolition streams and reusable stock (DigiC, 2024). These implications also matter for computing in construction: reuse is not only a material challenge, but also a coordination challenge involving data availability, timing, and cross-organisational alignment (Knoth et al., 2022).

Limitations and robustness considerations

Several limitations should be noted. First, Stage 1 relies on the availability and consistency of product-group MKI values; while this supports comparability, it also makes results sensitive to database assumptions and boundary definitions (Nationale Milieudatabase (NMD), 2025). Second, the MCA scoring uses ordinal scales and expert judgement, which introduces subjectivity and may vary with expert composition and sector familiarity. Third, the current implementation does not propagate uncertainty from mass-flow estimates, impact factors and scoring into the final ranking; future work could incorporate uncertainty analysis to quantify ranking stability under data and judgement variance. Finally, Stage 3 is deliberately illustrative; it demonstrates linkage to barrier diagnosis but does not validate the effectiveness of specific interventions or business arrangements.

Research and implementation directions

Future development can proceed along three directions. (1) Data enrichment and automation: linking product-group datasets to spatially resolved building-stock models and material-passport data streams to improve timeliness and granularity (Wuyts et al., 2022). (2) Decision-support extensibility: integrating additional objectives (e.g., logistics distance, storage constraints, and risk-adjusted costs) while maintaining the same computational backbone. (3) Governance and standardisation alignment: connecting prioritisation outputs to emerging sector standards and assessment protocols to reduce liability concerns and enable procurement uptake (NEN, 2023).

Together, these directions position the framework as an extensible computational layer that can support circular decision-making across multiple products and contexts.

Conclusions

This paper develops a three-stage workflow for product-level reuse prioritisation in Dutch demolition outflows and shows that reuse potential cannot be inferred from environmental relevance alone. After screening 139 product groups, masonry brickwork emerges as the highest-ranked option, as implementation-oriented factors—especially supply–demand balance—outweigh environmental impact in expert weighting. The masonry case further demonstrates that scaling reuse depends on contamination control, certification, and coordination between demolition and processing.

The main implication is that reuse planning should combine environmental relevance with implementation conditions and use ranking outputs to guide targeted interventions. Although demonstrated on Dutch data and a single case, the contribution lies in the workflow rather than the specific product group, making it transferable to contexts with comparable mass-flow and environmental-impact data. While environmental relevance is operationalised using MKI from the NMD in the Dutch case, alternative indicators—such as monetised LCA metrics or EPD-based impact values—can be adopted in other jurisdictions. The MCA structure is likewise adaptable, provided that criteria and weights are aligned with local market and regulatory conditions (Honic et al., 2021; Giorgi et al., 2022).

References

Arora, M., Raspall, F., Fearnley, L., Silva, A., 2021. Urban mining in buildings for a circular economy: Planning, process and feasibility prospects. Resour. Conserv. Recycl. https://doi.org/10.1016/j.resconrec.2021.105754

Arora, M., West, J., Patterson, M., Daly, M., 2020. Buildings and the circular economy: Estimating urban mining, recovery and reuse potential of building components. Resour. Conserv. Recycl. https://doi.org/10.1016/j.resconrec.2019.104581

Atta, I., Bakhoum, E.S., 2024. Environmental feasibility of recycling construction and demolition waste. Int. J. Environ. Sci. Technol. 21, 2675–2694. https://doi.org/10.1007/s13762-023-05036-y

Bajwa, A.U.R., Siriwardana, C., Shahzad, W., Naeem, M.A., 2025. Material selection in the construction industry: a systematic literature review on multi-criteria decision making. Environ. Syst. Decis. 45, 8. https://doi.org/10.1007/s10669-025-10001-w

Banias, G., Achillas, C., Vlachokostas, C., Moussiopoulos, N., Tarsenis, S., 2010. Assessing multiple criteria for the optimal location of a construction and demolition waste management facility. Build. Environ. 45, 2317–2326. https://doi.org/10.1016/j.buildenv.2010.04.016

Bellini, A., Tadayon, A., Andersen, B., Klungseth, N.J., 2024. The role of data when implementing circular strategies in the built environment: A literature review. Clean. Environ. Syst. 13, 100183. https://doi.org/10.1016/j.cesys.2024.100183

Byers, B.S., Raghu, D., Olumo, A., De Wolf, C., Haas, C., 2024. From research to practice: A review on technologies for addressing the information gap for building material reuse in circular construction. Sustain. Prod. Consum. 45, 177–191. https://doi.org/10.1016/j.spc.2023.12.017

De Wolf, C., Byers, B.S., Raghu, D., Gordon, M., Schwarzkopf, V., Triantafyllidis, E., 2024. D5 digital circular workflow: five digital steps towards matchmaking for material reuse in construction. Npj Mater. Sustain. 2, 36. https://doi.org/10.1038/s44296-024-00034-8

EIB; Metabolic, 2022. Materiaalstromen in de bouw en infra: 2019, 2030, 2050. URL https://www.eib.nl/publicaties/materiaalstromen-in-de-bouw-en-infra/ (accessed 1.21.26).

Ghisellini, P., Ncube, A., Casazza, M., Passaro, R., 2022. Toward circular and socially just urban mining in global societies and cities: Present state and future perspectives. Front. Sustain. Cities 4. https://doi.org/10.3389/frsc.2022.930061

Giorgi, S., Lavagna, M., Wang, K., Osmani, M., Liu, G., Campioli, A., 2022. Drivers and barriers towards circular economy in the building sector: Stakeholder interviews and analysis of five European countries policies and practices. J. Clean. Prod. 336, 130395. https://doi.org/10.1016/j.jclepro.2022.130395

Honic, M., Kovacic, I., Aschenbrenner, P., Ragossnig, A., 2021. Material Passports for the end-of-life stage of buildings: Challenges and potentials. J. Clean. Prod. 319, 128702. https://doi.org/10.1016/j.jclepro.2021.128702

Hopkinson, P., Chen, H.-M., Zhou, K., Wang, Y., Lam, D., 2019. Recovery and reuse of structural products from end-of-life buildings. Proc. Inst. Civ. Eng. - Eng. Sustain. 172, 119–128. https://doi.org/10.1680/jensu.18.00007

Jato-Espino, D., Castillo-Lopez, E., Rodriguez-Hernandez, J., Canteras-Jordana, J.C., 2014. A review of application of multi-criteria decision making methods in construction. Autom. Constr. 45, 151–162. https://doi.org/10.1016/j.autcon.2014.05.013

Kanters, J., 2020. Circular Building Design: An Analysis of Barriers and Drivers for a Circular Building Sector. Buildings 10, 77. https://doi.org/10.3390/buildings10040077

Kirchherr, J., Reike, D., Hekkert, M., 2017. Conceptualizing the circular economy: An analysis of 114 definitions. Resour. Conserv. Recycl. 127, 221–232. https://doi.org/10.1016/j.resconrec.2017.09.005

Knoth, K., Fufa, S.M., Seilskjær, E., 2022. Barriers, success factors, and perspectives for the reuse of construction products in Norway. J. Clean. Prod. 337, 130494. https://doi.org/10.1016/j.jclepro.2022.130494

Nationale Milieudatabase (NMD), 2025. MKI Viewer Tool [WWW Document]. NMD. URL https://milieudatabase.nl/en/latest/news/en-mki-van-de-bouw-blijft-stijgen-bij-business-as-usual/ (accessed 1.21.26).

NEN, 2023. NTA 8713:2023 — Hergebruik van constructiestaal. https://www.nen.nl/nta-8713-2023-nl-307691

Orenga Panizza, R., Nik-Bakht, M., 2024. Building stock as a future supply of second-use material – A review of urban mining methods. Waste Manag. Bull. 2, 19–31. https://doi.org/10.1016/j.wmb.2024.03.001

Papamichael, I., Voukkali, I., Loizia, P., Zorpas, A.A., 2023. Construction and demolition waste framework of circular economy: A mini review. Waste Manag. Res. 41, 1728–1740. https://doi.org/10.1177/0734242X231190804

Peykani, P., Emrouznejad, A., Nouri, M., 2026. Best-worst multi-criteria decision-making method: A review of the literature. Socioecon. Plann. Sci. 104, 102345. https://doi.org/10.1016/j.seps.2025.102345

Planbureau voor de Leefomgeving (PBL), 2025. Integrale Circulaire Economie Rapportage (ICER) 2025. https://www.pbl.nl/publicaties/integrale-circulaire-economie-rapportage-2025

Rakhshan, K., Morel, J.-C., Alaka, H., Charef, R., 2020. Components reuse in the building sector – A systematic review. Waste Manag. Res. 38, 347–370. https://doi.org/10.1177/0734242X20910463

Rezaei, J., 2016. Best-worst multi-criteria decision-making method: Some properties and a linear model. Omega 64, 126–130. https://doi.org/10.1016/j.omega.2015.12.001

Rezaei, J., 2015. Best-worst multi-criteria decision-making method. Omega 53, 49–57. https://doi.org/10.1016/j.omega.2014.11.009

Röck, M., Baldereschi, E., Verellen, E., Passer, A., Sala, S., Allacker, K., 2021. Environmental modelling of building stocks – An integrated review of life cycle-based assessment models to support EU policy making. Renew. Sustain. Energy Rev. 151, 111550. https://doi.org/10.1016/j.rser.2021.111550

SGS Search; Cirkelstad, 2021. Beslisboom hoogwaardig hergebruik bouwproducten [WWW Document]. SGS Search. URL https://www.sgssearch.nl/over-search/search-in-de-media/beslisboom-hoogwaardig-hergebruik-bouwproducten.html (accessed 1.21.26).

Swarnakar, V., Khalfan, M., 2024. Circular economy in construction and demolition waste management: an in-depth review and future perspectives in the construction sector. Smart Sustain. Built Environ. https://doi.org/10.1108/SASBE-02-2024-0056

Tura, N., Hanski, J., Ahola, T., Ståhle, M., Piiparinen, S., Valkokari, P., 2019. Unlocking circular business: A framework of barriers and drivers. J. Clean. Prod. 212, 90–98. https://doi.org/10.1016/j.jclepro.2018.11.202

Van Den Berg, M., Hulsbeek, L., Voordijk, H., 2023. Decision-support for selecting demolition waste management strategies. Build. Cities 4, 883–901. https://doi.org/10.5334/bc.318

Wuni, I.Y., 2022. Mapping the barriers to circular economy adoption in the construction industry: A systematic review, Pareto analysis, and mitigation strategy map. Build. Environ. 223, 109453. https://doi.org/10.1016/j.buildenv.2022.109453

Wuyts, W., Miatto, A., Khumvongsa, K., Guo, J., Aalto, P., Huang, L., 2022. How Can Material Stock Studies Assist the Implementation of the Circular Economy in Cities? Environ. Sci. Technol. 56, 17523–17530. https://doi.org/10.1021/acs.est.2c05275

Yang, X., Hu, M., Shan, W., Zhang, C., Li, T., Pan, Y., 2023. Integrating bottom-up building stock model with logistics networks to support the site selection of circular construction hub. J. Clean. Prod. 430, 139677. https://doi.org/10.1016/j.jclepro.2023.139677

Yu, Y., Yazan, D.M., van den Berg, M., Firdausy, D.R., Junjan, V., Iacob, M.-E., 2023. Circularity information platform for the built environment. Autom. Constr. 152, 104933. https://doi.org/10.1016/j.autcon.2023.104933

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

UTILIZE DISTRICT HEATING SMART METER DATA AND BUILDING INFORMATION TO COMPARE THE HEAT LOAD WITH THE DESIGN GUIDELINES

Martin Frandsen[1,2], Qinjiang Yang[1], Michal Zbigniew Pomianowski[2], Morten Karstoft Rasmussen[3], and Michele Tunzi[1]
[1]Department of Civil and Mechanical Engineering, Technical University of Denmark, Denmark
[2]Department of the Built Environment, Aalborg University, Denmark
[3]Kamstrup A/S, Denmark

Abstract

Recent instability in gas markets has strengthened demand for district heating (DH) solutions. However, design heat loads in DH systems continue to rely on outdated design guidelines that no longer represent the current heat demand. Using high-temporal-resolution data down to 2 seconds from smart heat meters and building information, this study identifies development trends in Danish residential buildings and assesses current design heat loads. The results indicate short-duration peak loads with relatively low heat power, suggesting a potentially over-dimensioned design heat load. These preliminary findings highlight the need for further investigation and may inform future improvements of DH design guidelines.

Introduction

The European Union's REPowerEU roadmap aims to eliminate Russian gas and oil imports by 2027 and to strengthen Europe's energy self-sufficiency (European Commission, 2025). This new initiative to phase out natural gas in building heating systems is driving increased international demand for district heating (DH) technologies. Owing to its high flexibility and capacity to integrate a wide range of energy sources, DH can play a key role in reducing dependence on fossil fuels. The DH sector is currently transitioning toward fourth-generation district heating (4GDH), which is characterized by substantial reductions in average supply and return temperatures, typically around 55/25 °C (Lund et al., 2014). At present, DH supplies 12% of the total building heating demand in the European Union, although significant differences exist among member states (Kranzl et al., 2021). In Denmark, where DH currently accounts for about 50% of the building heating demand, the share is expected to increase to 75% as part of the national strategy for decarbonizing the entire energy system (Mathiesen et al., 2021).

The anticipated growth of Danish DH networks introduces new design challenges because existing guidelines are based on literature and empirical results from the 1980s, when large-scale DH deployment in Denmark was common, and may not reflect the requirements of modern DH systems (Lund et al., 2018). The integration of low-temperature heat sources can impose constraints on the achievable supply temperature levels in current DH networks. In the presence of high return temperatures from end-users, the resulting reduction in temperature difference may create hydraulic bottlenecks in the distribution network. These effects can limit operational flexibility and increase system complexity. Moreover, the future competitiveness of DH compared to individual heating solutions strongly depends on minimizing capital investments in network infrastructure (Holmstedt Hansen et al., 2018) and system components, which are directly influenced by design requirements. Therefore, in-depth knowledge of current heat load demand in the building stock is essential for optimally planning and designing new and renovated DH networks.

In dwellings, the design heat load is typically based on domestic hot water (DHW) production via a heat exchanger (HEX), as this requires the highest instantaneous temperature increase. A Danish study with low temperature DH in terraced houses (Thorsen et al., 2011) showed that the peak heat load for DHW production with a HEX was 20% lower than the design value of 32.3 kW from the Danish Code of Practice for Water Supply Installations (DS 439) (Dansk Standard, 2009) and the Danish reference book Varme Ståbi (Andreasen et al., 2021). Another Danish study (Ramboll, 2013) based on single-family houses (SFHs) connected to existing DH networks showed that the measured peak heat load for DHW with HEX in summer ranges from 12 to 29 kW. However, these studies were constrained by the small number of buildings investigated and the relatively low temporal resolution of the available data. As a result, it was not possible to draw general conclusions or to support a revision of existing design guidelines.

In this context, the Energy Efficiency Directive (European Commission, 2023), launched in 2012, has accelerated the deployment of remotely readable energy meters and submeters at the building level, thereby enabling increased digitalization and opening new opportunities for further investigation using larger datasets and higher temporal resolution data.

The planning and design of new DH networks must consider the different building types and the construction years of the building stock, as this also influences the space heating (SH) demand. In the study by (Kristensen and Petersen, 2021), yearly heating use per heated floor area from different building types in a DH utility was

analyzed by construction year period. The author showed that, for residential buildings, age groups after 1961 had lower annual heating use, down to less than half for the newest age group. Thus, this study only indicates that newer buildings use less heating energy, and not whether the heat load is also reduced.

Novelty

The novelty of this study lies in the use of remotely readable DH smart heat meters (SHMs) to measure the combined heat load for SH and DHW at high temporal resolution in residential buildings. The heat load was recorded at temporal resolutions ranging from 3 minutes down to 2 seconds, allowing the dynamics of heating systems to be analyzed and the actual magnitude of peak heat demand to be revealed. An aspect not always captured in previous studies, many of which are based on hourly measurements. By linking some of these measurements with building registry data, this study establishes an initial methodological structure for assessing building heat demand at scale.

The study's results and conclusions, based on high temporal resolution measurements, are preliminary and based on a small data set. Nevertheless, this creates a basis for revisiting current DH design guidelines and exploring whether prescribed design heat load values can be improved.

Method

The methodology section presents the data sources and analytical workflow applied in this study. It describes the collection and preprocessing of SHM data and explains how building characteristics obtained from the Danish Building and Dwelling Register (BBR; in Danish: Bygnings- og Boligregistret) (Vurderingsstyrelsen, 2025) are used to contextualize the measurements. Design heat load values from existing standards and guidelines are employed as reference benchmarks to evaluate the observed heat demand. The analysis is limited to residential buildings, which constitute the largest share of the Danish building stock.

Design heat load

The design heat load for a building in the Danish design guidelines is calculated as the maximum of the SH or DHW heat load. The SH is designed with an outdoor-indoor temperature set of -12/20 °C and no heat gains. The heat load for DHW preparation can be provided by a HEX or a storage tank. For a single dwelling, the instantaneous design heat load for DHW preparation with HEX is by far the highest. According to the current version of DS 439, it can be as high as 32.3 kW and is based on the maximum heat load required for simultaneous draw-off from a kitchen sink and a shower (Dansk Standard, 2024).

Data collection

The SHM measures the total DH energy use delivered to a building, combining SH and DHW, or separately if two meters are installed. In this study, only SHMs measuring the combined energy use were used, as this is the most common solution. The data were collected hourly, every 3 minutes, and every 2 seconds. Two approaches were used to extract the data, either remotely or on-site.

1) The remote reading is performed using long-distance wireless M-Bus, in which strategically placed concentrators in the utility network communicate with the SHM. The readings are set to hourly by default, but the frequency can be increased to every 3 minutes. Data collection is conducted remotely without access to the building.
2) The on-site reading is performed with an optical eye, Figure 1A, which is mounted on the SHM and communicates via a serial infrared two-way communication. The optical eye is typically used to access the meter's settings. However, with an Arduino and an SD card, a data logger was created, and meter readings are logged every 2 seconds. Access to the building is required to install the data logger. Figure 1B shows an installation of a data logger.

The data collection is conducted in accordance with the General Data Protection Regulation (GDPR). In Denmark, DH utilities and SHM manufacturers have an assigned GDPR agreement with the building owner to collect remote readings. For the on-site readings, the authors have a signed GDPR agreement with the building owner. All data is handled anonymously.

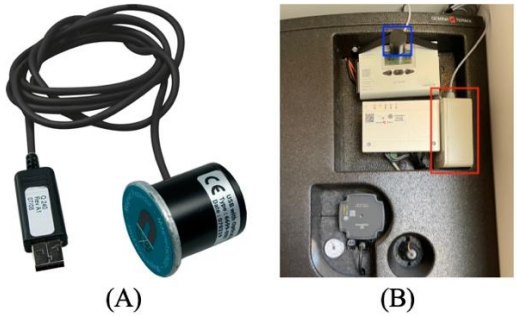


Figure 1: A) Optical eye. B) Optical eye (marked blue) and data logger (marked red) installed on an SHM in campaign 2.

Data preprocessing

Data from remote SHM readings can typically face two problems: first, timestamps may not be aligned to exact clock hours; second, there may be missing data points in the dataset. The reason for both problems is the stability of the remote connection between the SHM and the concentrators located in the utility. A concentrator is a device that collects SHM data and assigns timestamps to the data rows. Depending on the signal strength, location of the installation, and distance from each SHM to the concentrator, there is a deviation between the value's actual timestamp and the assigned timestamp. The signal strength differs between all SHMs in the utility, and changes over time for each individual SHM, depending on interference from electromagnetic signals, and temporally physical blocks e.g. cars and people. If the signal strength is too poor, data can be lost.

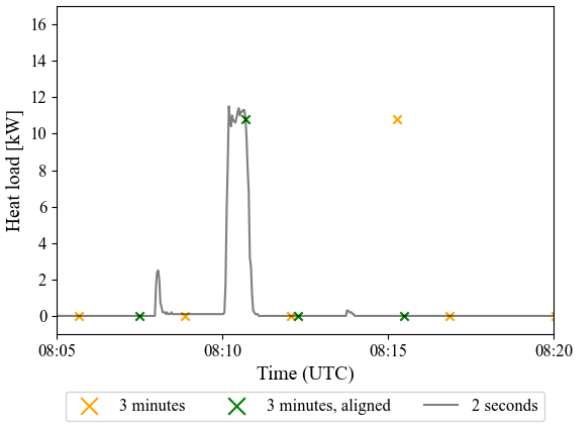


Figure 2: Example of graphical time alignment of 3-minute to 2-second values for one SHM. The 3-minute values in this plot were moved 274 seconds back in time.

The operations to address these problems depend on the use of the data in this paper. For the hourly data, two main operations were performed to address these problems: aligning the original data into regular one-hour data and filling the missing values using linear interpolation. For the 3-minute values, timestamps were manually aligned for each SHM by graphically comparing them with the 2-second data. This is done for the given time of each plot in this paper, as the time deviation can easily change. The observed time difference is a maximum of 5 minutes. In Figure 2, an example of a graphical time alignment for 3-minute remote readings is shown.

Using BBR for building classification

The hourly SHM data in Campaign 1 is combined with the BBR data for each building to compare average and maximum heat loads across residential building categories, age groups, and heating seasons. The BBR includes information about all buildings in Denmark in a publicly available database, e.g., building use category (residential, commercial, industrial, etc.), building area, construction and renovation years (Vurderingsstyrelsen, 2025). In this study, the BBR data set was delivered by the DH utility with anonymized SHM numbers and addresses.

A SHM can measure the total DH use for a building or parts of the building (i.e., one SHM can cover either one apartment in a TH/MFH, or one TH/MFH with several apartments). Since the data was anonymized, it was not possible to group heat meters by building. For simplicity, it was assumed that one SHM supplies one building. For each building, the construction year, heated floor area, and building use code were aggregated into the SHM data. Using the heated floor area, the hourly average heat load was normalized by area. The building's construction year was used to classify buildings into age groups to divide the heat load data in subsequent analyses. The age groups were based on the widely used TABULA project (Wittchen and Kragh, 2012) and the study by (Kristensen and Petersen, 2021), where the age groups were defined by shifts in building construction practices or changes in the Danish building regulations (BR). One additional age group (2018-2020) was added in this paper, as BR18 was implemented in 2018. The building use code defines the building category. From the main category Residential in the BBR, three building groups were created for this paper: single-family house (SFH), terraced house (TH), and multi-family house (MFH). Table 1 provides an overview of the BBR building use code and subcategories included in the defined building groups. The subcategory 130 in the TH group is being phased out and replaced by 131 and 132. However, subcategory 130 was included as there are still buildings listed with the building code in the BBR database.

Table 1: The BBR building subcategories included in the building groups for this paper. (Vurderingsstyrelsen, 2025)

Building group	Subcategories' building use code and description
SFH	120 Detached SFH
	121 Semi-detached SFH
	122 Detached SFH in low-density urban area
TH	130 Row, chain, cluster, or double house (*being phased out*)
	131 Row, chain, and cluster house
	132 Double houses
MFH	140 Multi-story building, multi-family house, or two-family house

Measurement campaigns

The SHM datasets are based on two separate measurement campaigns conducted in two municipalities in Nordjylland, Denmark, and using different data collection methods. The details of the two measurement campaigns are listed in Table 2.

Table 2: Details of the two conducted measurement campaigns.

	Period	No. SHM	Temporal resolution	Variables
Campaign 1	Feb'22 – Mar'22 & Jul'22 – Sep'22	4118	Hourly	Cumulative heat energy (kWh)
Campaign 2	Jul'25 – Dec'25	4	3-minutes, 2-seconds	Heat load (kW)

Campaign 1 includes data from 4118 SHM at hourly resolution, from which the measured cumulative heat energy was used. The data was collected in two separate periods, a heating period (Feb'22 to Mar'22) and a non-heating period (Jul'22 to Sep'22). The stacked plot in Figure 3 shows the distribution of the buildings divided into age and building groups. SFH is the most dominant building type, consistent with the results in (Kristensen and Petersen, 2021). Of the total building stock in the DH utility, the age groups 1961-1972, 1973-1978, and 1979-1998 each account for 19.7%, 15.0%, and 25.4%, respectively. As described in the method, the number of buildings was assumed to be per SHM. However, the

large construction of terraced houses in 1979-1998 may be due to the installation of an SHM per dwelling rather than per building. The age group Before 1851 only includes one SFH and was excluded from further analysis.

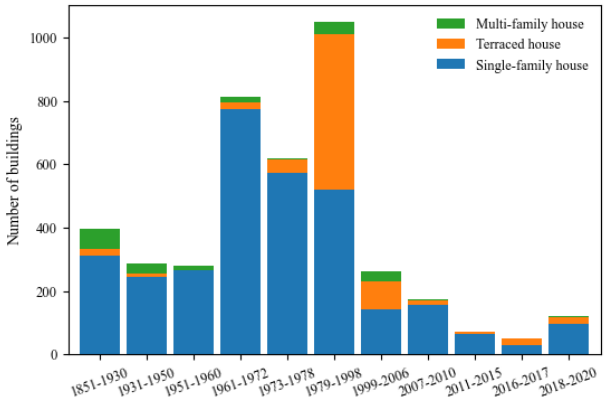


Figure 3: Number of buildings in the three building groups, sorted in age groups. 3168 SFHs, 743 THs, and 206 MFHs.

After analyzing the hourly data in Campaign 1, a need for higher temporal resolution to capture the dynamics of the heat load was evident. Campaign 2 was conducted in another municipality, where the frequency of remote readings was increased to every 3 minutes, and on-site data loggers were installed to measure the actual heat load and capture the short-term peaks. Campaign 2 includes data from four SFHs in the two age groups: 1961-1972 (two buildings) and 1973-1978 (two buildings). The data was collected from 15th July to 31st December 2025 and includes both a heating and a non-heating period. For this analysis, the instantaneous heat load values from the SHM were used.

Results and discussions

The results from the two measurement campaigns are presented and discussed with a focus on heat load characteristics. First, hourly data from Campaign 1 are used to illustrate the distribution of average and maximum heat load at the network level for residential buildings. Second, high-resolution measurements from Campaign 2 are analyzed to assess the magnitude of heat load in four SFH during both heating and non-heating periods. With a temporal granularity down to 2 seconds, this allows the dynamic behavior and true peak values of the heat load to be captured. Finally, a preliminary analysis of the 2-second and 3-minute data in Campaign 2 is conducted for one SFH over one day, during both heating and non-heating periods. 2-second data is focused on, as it shows the actual heat load profile and captures all the peaks from DHW draw-offs, compared to the 3-minute data. However, the 3-minute values are plotted to evaluate the limitations and potential of different temporal resolutions.

Hourly average and maximum heat load on network level – Campaign 1

Using the hourly measurement in Campaign 1, the average heat load for each building over the heating and non-heating periods is determined and normalized by the building's heated floor area. The distribution of the hourly average heat load is shown in Figure 4 as box plots grouped by building group and age group. There is a significant difference in the number of buildings in each building group. For terraced houses and multi-family houses, conclusions should be drawn with care, as some age groups have fewer than 10 buildings. In this case, a few buildings with poorly adjusted DHW or SH systems could easily skew the results, making them outliers. For the SFH group, there is a significant drop in average heat load from age group 1951-1960 to 1961-1972, consistent with the introduction of the first Danish BR in 1961. The BR61 set requirements for the transmission losses in the building envelope. In BR85, a requirement for maximum building energy demand was introduced, which partly can explain the drop in 1979-1998. The average heat load in buildings constructed after 1999 declines with a small gradient across age groups, matching the tightening of the energy demand in the BRs. When comparing the average heat load for SFHs in Figure 4 with the energy use results in (Kristensen and Petersen, 2021), the age groups show the same declining trend and comparable magnitudes. It

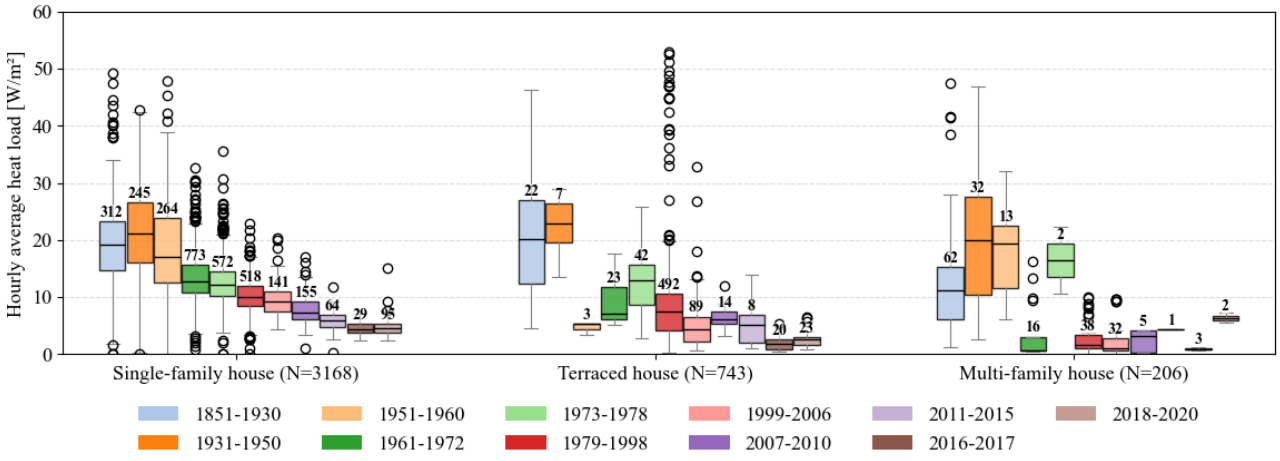


Figure 4: Hourly average heat load normalized per area for all residential buildings in Campaign 1, including the heating and non-heating periods. The boxes show the interquartiles (25th and 75th percentiles) and the median, while the whiskers show the 5th and 95th percentiles. The number above each box plot is the number of buildings in the age group.

should be noted that in the BBR data set, the renovation year for the building is indicated if it has been renovated. This is not considered in this analysis, as renovations can be minimal or radical, and a level of renovation is not indicated in the BBR. Additionally, the BBR data is not error-proof as the long list of inputs is manually entered, which can lead to incorrect entries due to misunderstandings of specific terms. To accommodate this, two outliers in the age group 1931-1950 in the TH building group that skew the general trend are removed from the dataset.

The connection of SFHs is large in Danish DH networks, and the following focuses on this group. For each SFH in Campaign 1, the maximum registered hourly heat load in the non-heating period (July to September) and the maximum registered hourly heat load in the heating period (February to March) are found. In Figure 5, the distribution of maximum hourly heat loads for the heating and non-heating periods is shown as box plots by age group. Obviously, the heating period has a higher maximum load than the non-heating period. An interesting observation for both periods is that, from 1999 onward, the median of the maximum load has increased, and the distribution within each age group is wider. During the heating period, one reason could be the rebound effect (Brøgger et al., 2018; Sorrell et al., 2009), where tenants in low-energy buildings increase the room temperatures to perceive higher thermal comfort. Another reason could be the increased installation of rainshowers in new buildings, which typically have high water flow rates. The use of rainshower affects the heat load throughout the year, but is most evident during the non-heating period, when the heat load for SH is minimal. This should be considered as a hypothesis, as the authors of the paper do not have statistics on rainshowers versus normal showerheads in new buildings.

From Figure 5, it can also be observed that the box plot of the maximum hourly heat load of the five age groups from 1931 to 1998 in the non-heating period is visually very similar. This happens due to two facts. First, the heat energy has a resolution of 1 kWh, making the resolution for heat load 1 kW. Second, box plots summarize the data using the median and quartiles rather than individual values. Since the observed maximum heat load for these age groups is largely concentrated around 4 kW, the resulting medians and quartiles are very close to each other, leading to box plots that appear nearly identical.

Since the data from the first measurement campaign are available only at an hourly resolution, consumption behavior at timescales below one hour cannot be resolved, and a fair comparison with the design heat load specified in the Danish guidelines cannot be made. A higher temporal resolution was therefore required and was adopted in the second measurement campaign.

Actual heat loads in four SFH – Campaign 2

With a relatively small data set in Campaign 2, the focus was on making a preliminary evaluation of the design heat load from DS 439, using the 2-second data. The four SFHs

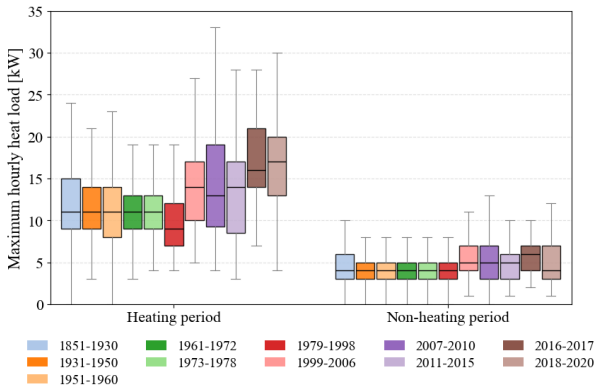


Figure 5: Maximum hourly heat load for single-family houses for both heating and non-heating periods. The boxes show the interquartiles (25th and 75th percentiles) and the median, while the whiskers show the 5th and 95th percentiles.

are connected to the local DH network, and each has a substation with two HEXs for SH and DHW. These houses were selected for the measurement campaign because they have a similar heat load distribution, were constructed according to the BRs, with requirements for transmission losses in the building envelope, and were built before the first requirement for building energy demand. The box plots in Figure 6 show the total measured instantaneous heat load for each SFH, grouped by month. The total heat load is the combined heat load for DWH and SH. The boxes represent the interquartile range (25^{th}-75^{th} percentiles) and the median. The upper whiskers are, as a conservative choice, extended to the 99.9^{th} percentile to include most heat loads for DHW draw-offs while excluding extreme values. The majority of the data in the four SFHs is far below the design heat load of 32.3 kW. The highest 99.9^{th} percentile of the four SFHs is 22.1 kW occuring in SFH2 in December. Of the total data for SFH3, the share of outliers exceeding the design heat load is $2.5x10^{-4}$%. This corresponds to 20 occurrences of a 2-second value exceeding the design value over a 6-month period. These extremely rare and short-term peaks indicate that the design guidelines most certain over-dimensioning the design heat load. The degree of over-dimension cannot be concluded by this preliminary study, and an analysis of more SFHs is needed. The last part of the results section shows an examination of the heat load profile and a discussion about the peaks that exceed the design heat load in SFH3. To understand the outliers, it is necessary to examine the duration of the peaks and the reasons they occur.

Heat load profile and peak analysis in one SFH – Campaign 2

With data at a high temporal resolution of 2 seconds, the actual heat load profile of a building can be visualized and analyzed to better understand the dynamics of DH supplied to a single building. The 3-minute data is also with lower temporal resolution. SFH3 is selected for examination as it has outliers above the design heat load. The SFH is inhabited by a retired couple. The total DH included in the plots to discuss the potential of using data

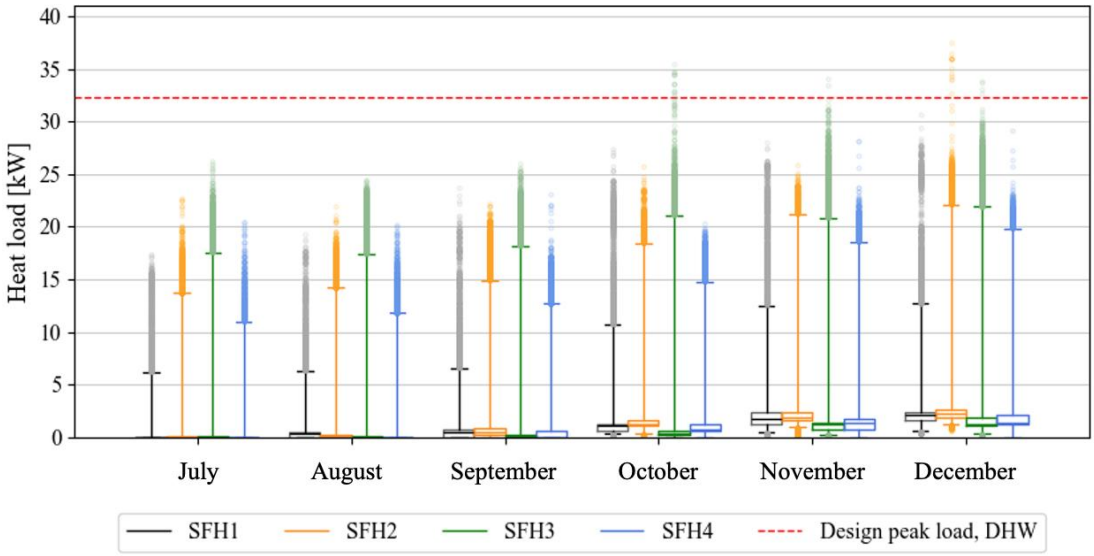


Figure 6: Box plots of the 2-second heat load from the four SFHs in campaign 2, grouped per month. The top-whisker of the box plot is the 99.9th percentile. The design heat load for an SFH is plotted.

load for DHW production and SH is analyzed for one day in the non-heating period (31st of July) and one day in the heating period (16th of November). The 31st of July is chosen as it has the highest peak in the non-heating period, and the 16th of November because it has a high peak, together with a period of low outdoor temperature.

For Thursday, the 31st of July, the heat load profile is plotted in the time range 06:00 to 20:00 in Figure 7. Multiple short peaks occur throughout the day, which aligns well with the occupancy of retired tenants at home using DHW. Two clusters of peaks were located around 7:30 and 17:30, which could be a morning shower and doing the dishes. When there is no peak, the load is 0 W, indicating no need for SH. The plot in Figure 8 is a zoom-in on the time range 09:11 to 09:16 on the same day. It shows the highest peak of 26.3 kW on that day and for the entire month. The interesting outcome is that for this specific peak, the duration is only 12 seconds above 25 kW and 20 seconds above 20 kW.

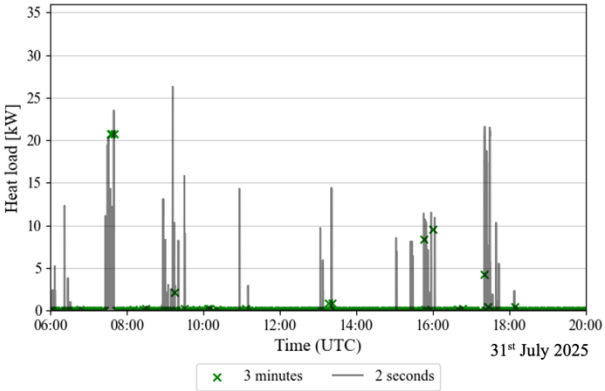


Figure 7: Time series of the heat load over one day in the non-heating period for Campaign 2's SFH3. Remote readings (3 minutes) and on-site readings (2 seconds).

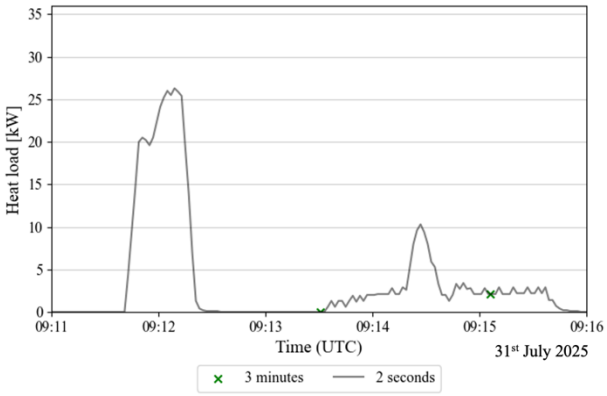


Figure 8: A zoom in on the highest peak in Figure 7.

For Sunday, the 16th of November, the heat load profile is plotted in Figure 9, and a zoom-in on the time range 12:38 to 12:43 is shown in Figure 10. On this day, the peaks also occur throughout the day, but when there is no peak, the base load is between 0.5 and 2.5 kW, covering the heat demand for SH. The highest heat load of the day, see Figure 10, is 33.3 kW. However, the peak duration is significantly short, with only 2 seconds above 32.3 kW. The measurement period for Campaign 2 does not cover the coldest period in Denmark, which typically occurs in January-March, and therefore not the highest SH demand. However, the base load for SH is relatively small compared to peak loads from DHW draw-off and is not anticipated to have a significant impact on the maximum heat load.

Figure 8 and Figure 10 show that the heat load peaks are high for a significantly short time duration. The highest peak in Figure 10 is anticipated to be the start of a DHW draw-off event. The peak has a high spike and drops suddenly to a lower level. This is likely due to the highly reactive thermostatic valve in the HEX, which opens quickly when a DHW draw-off starts and, when the DHW supply temperature is reached, closes more to adjust to the necessary DH flow. These types of control are the main

reasons for the outliers, 0.1% of the measurements, shown in Figure 6. The peaks for DHW draw-off are higher during the heating period than during the non-heating period. The reason for this is, on one hand, due to a higher supply temperature from the DH utility, and on the other hand, due to a colder domestic cold water supply.

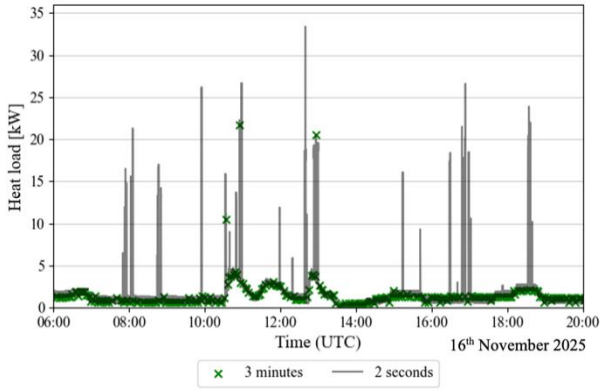


Figure 9: Time series of the heat load over one day in the heating period for Campaign 2's SFH3. Remote readings (3 minutes) and on-site readings (2 seconds).

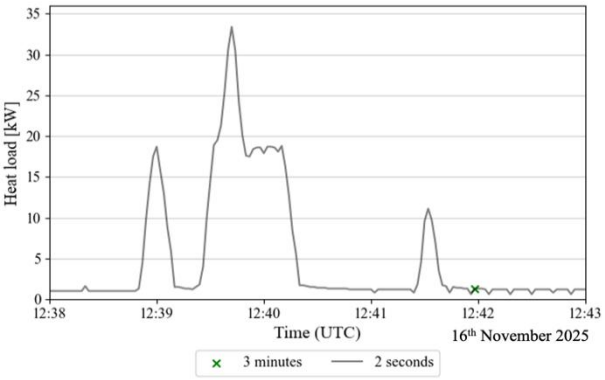


Figure 10: A zoom in on the highest peak in Figure 9.

This suggests that the current design guidelines should not be based on these outliers, which most likely do not affect the tenants' perception of DHW comfort. Nonetheless, there is a need to further investigate, using a larger dataset, the optimal design heat load of a HEX for a single dwelling that ensures tenants' DHW thermal comfort while avoiding the over-dimensioning implied by current design guidelines.

Finally, the 3-minute values are compared to the 2-second values. The 3-minute readings do not capture the maximum peak as seen in Figure 7-10. However, some of the load values in the DHW draw-offs are captured, e.g., in Figure 9 around 11:00 and 13:00. This suggests that, given a sufficiently long observation period, heat loads approaching the maximum peak are likely to be captured. Furthermore, by applying a statistical approach, the maximum peak heat load could be estimated together with an associated level of uncertainty. The use of slightly lower high-resolution data (e.g., 3-minute instead of 2-second values) to identify heat load peaks would allow a larger number of buildings to be analyzed, as data collection can be performed remotely without requiring physical access to the dwellings. This significantly reduces both the time and cost associated with further investigations. The present study reports only preliminary results based on a limited number of single-family houses and therefore does not allow for definitive conclusions. Further in-depth investigations covering a larger sample of SFHs with high temporal resolution are recommended.

Conclusion

Design heat loads in district heating systems continue to rely on design guidelines from the 1980's, which no longer reflect current building heat demand. Many previous studies are based on hourly measurements that shed light on the trend in heating demand, but they do not capture the short-term dynamics of the heat load profile. This work links building registry data with hourly measurements in one measurement campaign and identifies single-family houses for this municipality as the most important building use category for investigating the actual heat load profile. In the second preliminary measurement campaign, the heat load is measured on readable smart heat meters, with a high temporal resolution down to 2 seconds, to capture the actual heat load profile and analyze the magnitude and duration of the peaks. The high temporal resolution allowed a fair comparison to the design guidelines.

This investigation focuses on single-family houses to compare the design heat load of a district heating substation, where domestic hot water is prepared via a heat exchanger, which typically represents the highest instantaneous heat demand. Over a 6-month measurement period across four single-family houses, the majority of measured 2-second heat loads were significantly below the design value. The highest measured 99.9th percentile during a month was 31.6% below the design heat load. The extreme few outliers that exceeded the design heat load were further visually investigated. The outliers occur on 2 seconds when the thermostatic valve for domestic hot water preparation overshoots the opening and immediately reduces it again. This operation is expected to have no impact on tenants' perception of DHW thermal comfort.

The outcome highlights the need for further investigation into the optimal design heat load of a HEX for a single dwelling that aligns with tenants' perceptions of DHW thermal comfort and avoids over-dimensioning in the current guidelines.

Acknowledgements

This work was funded by the Flemish Institute for Technological Research (VITO NV) and the Technical University of Denmark, Department of Civil Engineering under the PhD grant: DTU_PhD_2201_contract; by the national EUDP grants J.nr. 134243-533239 to support the Danish participation in the international IEA project "DHC Annex TS9: Digitalization of District Heating and Cooling"; from the IEA Annex XIV project "FlexVal".

The work was also funded by the Danish EUDP grant no. 640241-521852: "Datadrevet design og drift af lavtemperaturfjernvarme".

References

Andreasen, B., Drivsholm, C., Petersen, F., Poulsen, H., de Wit, J.A., Nielsen, J., Frederiksen, J., Rose, J., Buhl, L., Nielsen, M.B., Winther, N., Schleicher, O., Paulsen, O., Lauritsen, A.B., 2021. Varme Ståbi, 8th ed. Praxis.

Brøgger, M., Bacher, P., Madsen, H., Wittchen, K.B., 2018. Estimating the influence of rebound effects on the energy-saving potential in building stocks. Energy Build. 181, 62–74. https://doi.org/10.1016/j.enbuild.2018.10.006

Dansk Standard, 2024. Code of Practice for domestic water supply installations. København.

Dansk Standard, 2009. Code of Practice for domestic water supply installations. København.

European Commission, 2025. REPowerEU roadmap - European Commission [WWW Document]. European Commission. URL https://energy.ec.europa.eu/strategy/repowereu-roadmap_en?prefLang=da

European Commission, 2023. Energy Efficiency Directive [WWW Document]. European Commission. URL https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficiency-targets-directive-and-rules/energy-efficiency-directive_en (accessed 2.5.26).

Holmstedt Hansen, C., Gudmundsson, O., Kortegaard Støchkel, H., Detlefsen, N., 2018. The competitiveness of district heating compared to individual heating When is district heating the cheapest source of heating?

Kranzl, L., Fallahnejad, M., Büchele, R., 2021. Renewable space heating under the revised Renewable Energy Directive. https://doi.org/https://data.europa.eu/doi/10.2833/525486

Kristensen, M.H., Petersen, S., 2021. District heating energy efficiency of Danish building typologies. Energy Build. 231. https://doi.org/10.1016/j.enbuild.2020.110602

Lund, H., Østergaard, P.A., Chang, M., Werner, S., Svendsen, S., Sorknæs, P., Thorsen, J.E., Hvelplund, F., Mortensen, B.O.G., Mathiesen, B.V., Bojesen, C., Duic, N., Zhang, X., Möller, B., 2018. The status of 4th generation district heating: Research and results. Energy 164, 147–159. https://doi.org/10.1016/j.energy.2018.08.206

Lund, H., Werner, S., Wiltshire, R., Svendsen, S., Thorsen, J.E., Hvelplund, F., Mathiesen, B.V., 2014. 4th Generation District Heating (4GDH). Integrating smart thermal grids into future sustainable energy systems. Energy 68, 1–11. https://doi.org/10.1016/j.energy.2014.02.089

Mathiesen, B.V., Lund, H., Nielsen, S., Sorknæs, P., Moreno, D., Thellufsen, J.Z., 2021. Varmeplan Danmark 2021 - En Klimaneutral Varmeforsyning.

Ramboll, 2013. Effekt- og samtidighedsforhold ved fjernvarmeforsyning af nye boligområder.

Sorrell, S., Dimitropoulos, J., Sommerville, M., 2009. Empirical estimates of the direct rebound effect: A review. Energy Policy 37, 1356–1371. https://doi.org/10.1016/j.enpol.2008.11.026

Thorsen, J.E., Holm Christiansen, C., Brand, M., Olesen, P.K., Larsen, C.T., 2011. Expieriences On Low-Temperature District Heating In Lystrup-Denmark, in: Proceedings of International Conference on District Energy. p. 2025.

Vurderingsstyrelsen, 2025. Bygnings - og Boligregistret [WWW Document]. URL https://bbr.dk/forside (accessed 2.1.26).

Wittchen, K.B., Kragh, J., 2012. Danish building typologies Participation in the TABULA project. Danish Building Research Institute.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

EDGE COMPUTING FOR FEDERATED LEARNING–ENABLED ENERGY MANAGEMENT AND MONITORING SYSTEMS

Mustapha Habib, Youssef Elomari, and Qian Wang
Building Technology and Design, Department of Civil and Architectural Engineering,
KTH Royal Institute of Technology, 11428 Stockholm, Sweden.

Abstract

Despite growing digitalization of energy management systems (EMS) in buildings and local networks, integrating data-driven monitoring and control remains challenging. Edge computing offers enhanced local processing that can be embedded in existing EMS infrastructures. Within the European HYSTORE project, two edge-computing use cases were developed: a network-level platform enabling coordinated EMS operation using building thermal mass estimates, and a building-level controller supporting advanced monitoring and control of thermal energy storage. Both architectures are evaluated for design, integration, and potential for federated learning, enabling secure, efficient, and privacy-preserving multi-agent control.

Introduction

The growing demand for energy-efficient and sustainable building management systems (BMS) has driven significant advancements in technology. Traditional BMSs often rely on centralized computing architectures, which, while effective for basic monitoring and control, fall short in handling the complexities of modern building energy systems. With the rise of distributed energy resources, renewable energy integration, and dynamic grid interactions, there is an increasing need for more responsive, decentralized solutions. Edge computing has emerged as a promising approach, enabling real-time data processing and decision-making closer to the source of data generation. This paradigm not only reduces latency but also enhances system resilience, making it an ideal candidate for addressing the evolving challenges in building management. Combined with artificial intelligence (AI) and the internet of things (IoT), edge computing offers the potential to revolutionize the way energy is monitored, optimized, and controlled in smart buildings.

A few contributions were proposed to employ edge computing-driven advanced services to upgrade building management. The authors in (C. H. Wong et al., 2021) discuss how AI systems can facilitate real-time analysis, optimization, and enhancement of building services. They demonstrate how AI can improve energy efficiency, enhance occupant satisfaction, mitigate risks, and reduce costs in building management and control systems. However, the authors also highlighted some limitations regarding some applications, including restricted access to large datasets, high initial investment costs, and the need for specialized expertise. Authors in (L. A. Husein et al., 2022) review the integration of IoT and AI technologies for improving BMSs, emphasizing the need for greater academic attention to these applications. While the review highlights the potential of IoT-AI methods, it notes that few studies have implemented these technologies in the building sector, and their full potential remains unrealized. The authors call for more research to address these gaps and advance the adoption of IoT and AI in building energy management. G. Carvalho et al. examine the use of AI, particularly machine learning (ML), to enable computation offloading in edge computing environments (G. Carvalho et al., 2020). AI and ML techniques are shown to address the challenges of computation offloading in these systems, offering more accurate, adaptable, and portable decision-making compared to traditional methods. The authors categorize various ML algorithms and provide an in-depth analysis of their application in Vehicular Edge Computing Networks, exploring the benefits and limitations of offloading tasks in edge computing systems. Key challenges discussed include determining when and where to offload tasks, managing decision-making costs, and minimizing the risk of errors that could undermine edge computing's advantages. A. Bourechak et al. explore the intersection of AI and edge computing across various IoT application domains, providing a comprehensive analysis of AI's role in edge-based applications across eight key areas (A. Bourechak et al., 2023). The paper includes a qualitative comparison of the reviewed works, an evaluation of the challenges and limitations of AI model development and implementation in edge environments, and a discussion on how AI can address these obstacles. In (G. Walia et al., 2024), authors present a detailed review of resource management challenges in Fog/Edge computing, categorizing these into areas such as resource provisioning, task offloading, resource scheduling, service placement, and load balancing. The paper examines AI-based and non-AI-based solutions, evaluating their performance metrics, limitations, and challenges. It highlights future research opportunities, including integrating emerging technologies like serverless computing, 5G, blockchain, digital twins, and quantum computing to enhance Fog/Edge frameworks for IoT applications. Challenges discussed include the

complexity of managing heterogeneous resources and workload requirements, as well as the limitations of current non-AI-based solutions. In another study, B. Muniandi at el. discuss AI-driven energy management systems for smart buildings, emphasizing their ability to optimize energy use, enhance operational efficiency, and promote sustainability (B. Muniandi et al., 2024). Key functionalities include predictive analytics, adaptive control of building systems, and integration with renewable energy sources. These systems also enable smart buildings to participate in demand response programs, reducing costs and supporting grid stability. However, limitations such as data privacy and security concerns, as well as scalability and interoperability challenges across diverse building portfolios, are also underlined.

This study investigates two real-world applications in which edge computing offers significant potential to enhance energy management systems (EMSs) and building energy performance monitoring. The work covers the design, assembly, and commissioning of the proposed edge-computing solutions in two operational scenarios: a micro-district heating and cooling (micro-DHC) network and a residential building. Except the federated learning (FL) application for indoor temperature estimation in a micro-DHC–connected building, this study does not delve into the detailed mathematical formulation of each use case. The focus is instead on outlining and discussing the range of application scenarios in which FL-oriented edge-computing solutions can have a practical and impactful role.

Edge computing solutions: concepts and roles

This section presents the technical specifications of the two edge-computing concepts developed within the framework of the European HYSTORE project, along with the engineering details related to their design and integration into the existing systems.

Engineering basis of the proposed edge solutions

Solution 1: micro-district heating and cooling

Problem statement:

Isolated energy networks such as micro-district heating and cooling (micro-DHC) systems often suffer from limited coordination between buildings, which typically operate independently due to the lack of effective inter-building communication and system-level monitoring. These limitations become critical when complex thermal energy storage technologies, such as thermochemical (TCM) and phase-change-material (PCM) systems (see Fig. 1), are integrated, as they require advanced observer-based estimation and control strategies beyond the capabilities of conventional building management systems (BMSs). To address these challenges, the Montserrat micro-DHC demonstration within the HYSTORE project introduces a centralized edge-computing platform that enables network-wide data access and the deployment of advanced algorithms for state-of-charge estimation and model predictive control, supporting coordinated operation and improved energy management at the system level.

Solution 2: residential building substation

Problem statement:

In this situation, the purpose of the edge computing solution is to upgrade the computing capability of the BMS so it can handle the complex dynamics of eventually installed energy systems by incorporating advanced algorithms for control and monitoring.

With the current system, there has been an issue with the central air-handling unit (AHU) that uses a central heat pump system for heating. In extreme weather conditions, the AHU sometimes faces some challenges in ensuring acceptable thermal comfort. One of the potential solutions was to boost the temperature coming from the central HP system when it is overloaded, because it is providing heat to the other two buildings as well. In the HYSTORE project, the solution is partially or fully achieved by integrating an edge-driven sensible/latent energy system to boost the temperature in the AHU heating battery.

Edge-upgraded BMS:

The edge-driven solution for this use case is to develop two industrial panels, low- and high-level control. Each one is:

Low-level control:

- Direct communication of the LTES sensors.
- Multi-protocol enabled by Python programming.
- Handle the translation of the control commands to a set of actions on the hydronic system actuators.

High-level control:

- Long-term data storage.
- Complex AI-based and advanced control.
- Extend the control to be the energy management of the entire microgrid.

Within the framework of the HYSTORE project, the primary objective of the edge-driven solution is to enable high-level control of a hybrid sensible/latent thermal energy storage, PCM-based LTES (see Fig. 2) system through the implementation of observer-based model predictive control (MPC). By leveraging direct access to the system database, the edge platform can also coordinate the operation of the LTES with a heat pump in the Live-In Lab microgrid, thereby exploiting on-site photovoltaic generation and battery energy storage systems (BESS). The functional flowchart of the proposed edge-based solution to upgrade BMS functionality for indoor heating, which incorporates latent thermal energy storage (LTES), is presented in Fig. 3.

On-site installation, commissioning, and algorithms

This section details the installation procedures for each edge-computing solution and discusses the commissioning challenges encountered during deployment.

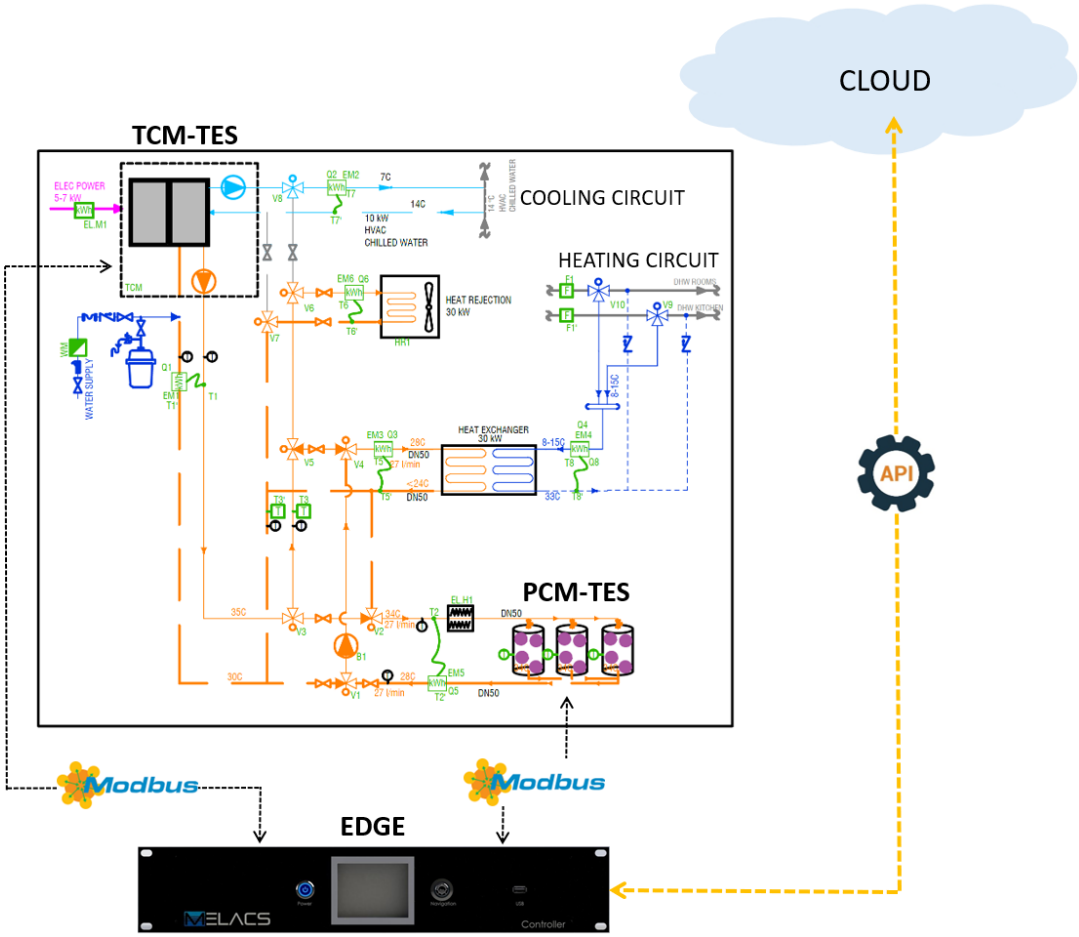


Figure 1: Edge solution as a high-level control for two LTES technologies: PCM and TCM at Montserrat micro-DHC.

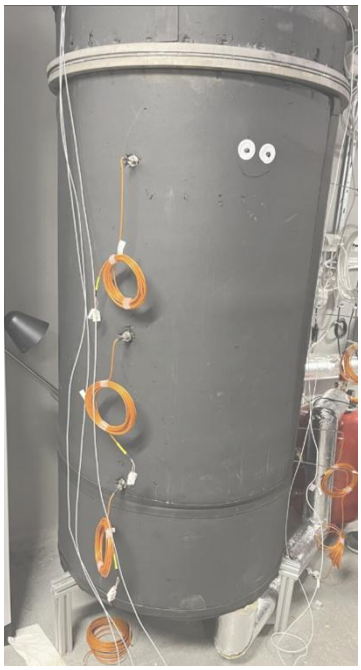


Figure 2: Installation of temperature sensors in the PCM-TES at the KTH live-in lab substation

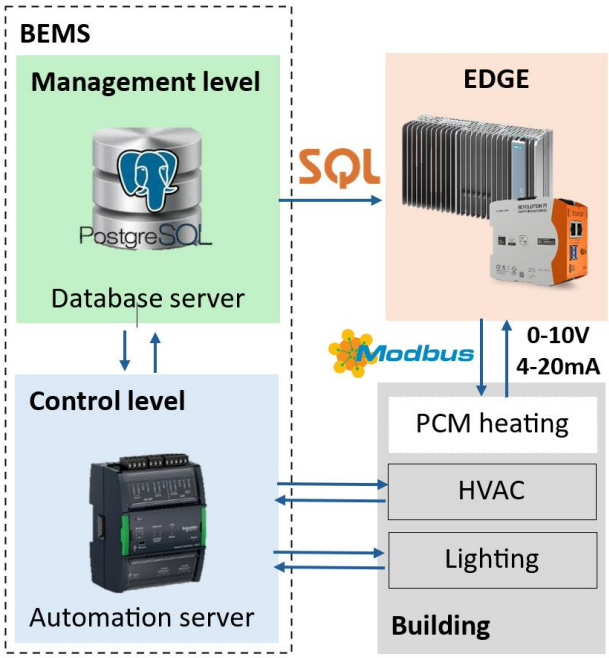


Figure 3: Edge solution as a high-level control for PCM LTES at KTH live-in lab

Solution 1

In Montserrat micro-DHC, a hotel building was selected for the installation of the two thermal energy storage technologies (TCM and PCM), as well as for deploying the second edge-computing solution (see Fig. 1). This choice was driven by the decision to exclude the sensors associated with the two LTES technologies from the central BMS database; as a result, direct physical communication with the devices was adopted. Communication with the various sensors and meters within the micro-DHC, including indoor and HVAC-related measurements from the hotel building, is achieved via the Modbus protocol over the local area network (LAN). Consequently, the edge computer was assigned a static IP address to ensure reliable identification and accessibility by other devices on the network.

Due to time constraints during the on-site commissioning phase of the HYSTORE project, full local communication with all network devices was not possible. To address this limitation, secure shell (SSH) access to the edge computer was configured remotely, allowing continued development and testing of the communication interfaces using Python.



Figure 4: Edge computing solution 1 installed in the hotel substation in Montserrat

Algorithms

In this real-world case, the edge computing platform implemented in Solution 1 serves two primary functions:

- Role 1: Centralized monitoring of the micro-DHC network.
- Role 2: High-level control of complex thermal energy storage (TES) systems.

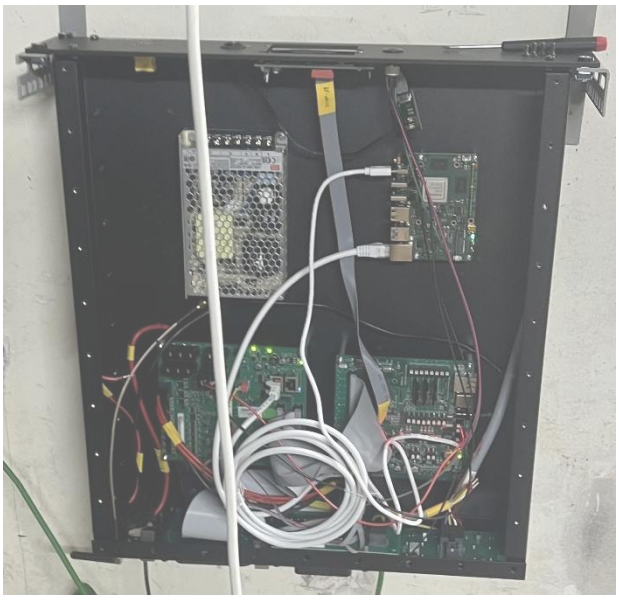


Figure 5: Components of the edge computing solution 1 installed in the hotel substation in Montserrat. It incorporates a GPU-accelerated controller with additional communication cards designed for industrial applications

For Role 1, several functionalities can be envisaged, including anomaly detection systems (ADS), centralized energy management systems (EMS), and coordinated demand-side management (DSM). The presence of a central computing entity communicating with all network components—such as HVAC systems, sensors, meters, and databases—enables a system-level perspective of energy flows and performance. This holistic view supports both network-wide energy management and detailed assessment of building and HVAC performance in heating and cooling generation. Within this scope, the present study focuses on the mathematical formulation of one representative function, which is presented at the end of this section.

For Role 2, a data-driven SoC estimation is implemented using a state observer. The algorithm combines Koopman theory with a linear Luenberger observer (KOOPMAN-LSO) to estimate internal PCM temperatures within the LTES based solely on the heat transfer fluid (water) outlet temperature. This approach has been validated in simulation using real operational data for multiple charging and discharging cycles of compact PCM TES, enabling on-site deployment without the need for intrusive PCM temperature sensors, which are often impractical in real applications.

A similar observer-based strategy will be developed to estimate the SoC of the thermochemical material (TCM) TES, which is co-located with the PCM-TES in the same building substation (see Fig. 1). Accurate SoC estimates allow the edge computer to optimally schedule charging and discharging of both LTES systems, maintaining thermal comfort while minimizing energy consumption. Two candidate control strategies will be tested for this purpose:

- Model predictive control (MPC)
- Reinforcement learning (RL)

Solution 2

In the residential building substation at the KTH Live-In Lab, the installation closely followed the approach used for Solution 1. A static IP address was assigned to the edge panel to enable connection to the LAN and allow real-time data queries from the system database via SQL. This database collects all operational data from the electrical systems, HVAC units, and indoor sensors across the microgrid. Since the edge panel was primarily tasked with managing the charging process of a PCM-TES, it was equipped with a Modbus coupler to allow direct physical wiring of the LTES sensor data to the controller. Additional analog and digital input/output cards were installed to interface with analog sensors and to control actuators, including ON/OFF switches and regulation valves. All remaining data acquisition and control of the heat pump are performed via Modbus TCP/IP.

Algorithms

As stated earlier, the planned role for the edge Solution 2 is to control the charging process of a hybrid latent/sensible using ground-source HP. Two main scenarios will be applied:

- A rule-based control (RBC), where a set of operational modes will be activated if certain conditions are met.
- MPC.

.

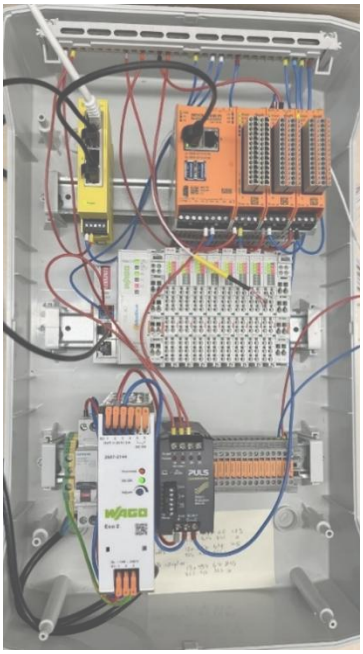


Figure 6: Components of the edge computing solution 2 (low-level control) meant for managing the PCM-TES and acts as an FL client. It incorporates a Python-enabled controller, IO cards, and Modbus coupler with the capacity of integrating 32 4-wire RTD temperature sensors needed to trach the PCM-TES SoC

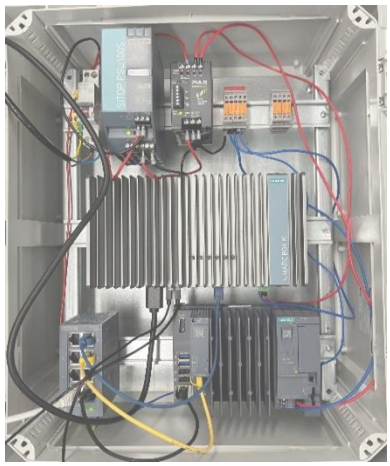


Figure 7: Components of the edge computing solution 2 (high-level control) meant for managing the KTH live-in lab microgrid, along with the different FL clients. It incorporates a high-capability industrial computer for running AI tasks and a PLC to enable communication with the field devices

For the rule-based control (RBC) strategy, a set of predefined operational modes is established, such as heating via the heat pump desuperheater, heating from the PCM-TES, combined operation, PCM-TES charging, and PCM-TES discharging. Each mode corresponds to a specific set of control actions executed by the edge panel through its embedded I/O modules. In this initial stage, the state-of-charge (SoC) of the PCM-TES is estimated using a simple physics-based formulation.

As operational data accumulates during RBC operation and is stored in a time-series database on the high-level control panel, a linear data-driven model of the PCM-TES is subsequently identified. This model enables the implementation of a linear model predictive control (MPC) scheme to maintain the existing heating functionality while achieving cost-effective operation based on dynamic electricity price forecasts and on-site photovoltaic (PV) power generation forecasts.

Federated learning concept

Role 1: microgrid management:

The KTH Live-In Lab can be regarded as a microgrid consisting of a local energy network serving three student residential buildings. The system integrates multiple energy resources, including renewable generation (PV, PVT, and wastewater heat recovery), energy storage systems (battery energy storage systems, sensible thermal energy storage, and geothermal storage), and diverse HVAC technologies, resulting in a sector-coupled energy system that includes heat pumps and air handling units. When operating in island mode, maintaining occupant comfort while optimizing energy utilization requires coordinated control across the three buildings, taking into account heating and domestic hot water demands, the state of charge of BESS and LTES units, renewable energy availability, and dynamic electricity tariffs.

Within this framework, each controllable HVAC system is associated with a federated learning client, implemented at the low-level edge computing layer (an

example is shown in Fig. 5). These clients locally train and update models for load prediction, renewable energy forecasting, and BESS/TES state-of-charge estimation. Only the learned model parameters are transmitted to a central FL server (Fig. 6), which uses them to perform network-level optimization aimed at maximizing renewable energy utilization, preserving thermal comfort, and sustaining islanded operation whenever technically feasible. If required, the server can propagate updated model parameters back to the clients using FedAvg or other aggregation and fitting methods.

Role 2: Indoor temperature / thermal mass prediction:

One of the main advantages of deploying a central monitoring computer in an islanded energy network, such as the Montserrat micro-DHC is its ability to operate as an FL server that coordinates multiple clients distributed across substations and buildings. Each client locally collects indoor temperature and HVAC operational data, trains its own thermal model, and transmits only the learned model parameters to the server. In this study, an RC-model-based observer is developed at each client to estimate indoor temperature dynamics in real time. Only the linear RC model coefficients and the associated observer gains are shared with the FL server, while raw measurements of indoor temperature, heating demand, and HVAC operation remain strictly local, preserving data privacy. Depending on the control objective, the FL server can exploit the received models in two ways. First, it can use the individual observer-corrected models to compute real-time optimal setpoints for each building, enabling coordinated yet decentralized energy management. Second, the server can aggregate the observation-based RC models to construct a global network-level thermal representation, which is then used to derive optimal boiler-side control actions, such as supply temperature and pump operation. This dual use of federated, observation-based models allows the micro-DHC to improve both building-level comfort control and system-level energy efficiency without centralized data collection. The building management system (BMS) collects at each building the following measurements:

- Outdoor temperature T_{out}
- Supply air temperature T_{sup}
- Humidity RH (can be used indirectly)
- Air quality (CO_2 ppm → occupancy proxy)
- Setpoint cooling $T_{sp,c}$
- Setpoint heating $T_{sp,h}$

The dynamic model in this case is based on the state

$$x = T_r$$

Where T_r: room air temperature. Therefore, the continuous-time equation is:

$$C_{eq}\dot{T}_r = \frac{T_{out} - T_r}{\mathrm{R}_{eq}} + Q_{HVAC} + Q_{\mathrm{int}} \qquad (1)$$

Where R_{HVAC} is effective HVAC thermal resistance and Q_{int} internal gains (occupancy + equipment). Thanks to the availability of the air quality sensor data, the heat source emissions related to the occupants (Qin) can be approximated as follows:

$$Q\mathrm{int} = \alpha.(CO_2 - CO_{2baseline}) \qquad (2)$$

Humidity sensor data can be incorporated as a regression term in HVAC heat flow:

$$Q_{HVAC} = \beta_1.(T_{\mathrm{sup}} - T_r) + \beta_2.RH \qquad (3)$$

$$\beta_1 = R_{HVAC}^{-1} = \begin{cases} k_c & T_r > T_{sp,c} \\ k_h & T_r < T_{sp,h} \\ 0 & deadband \end{cases}$$

The input vector is:

$$u = \begin{bmatrix} T_{out} \\ T_{sup} \\ RH \\ \mathrm{CO_2} \end{bmatrix} \qquad (4)$$

Output:

$$y = T_r \qquad (5)$$

Continuous-time state model is:

$$A = -\left(\frac{1}{C_{eq}\,\mathrm{R}_{eq}} + \frac{\beta_1}{C_{eq}}\right) \qquad (6)$$

$$B = \begin{bmatrix} \dfrac{1}{C_{eq}\,\mathrm{R}_{eq}} & \dfrac{\beta_1}{C_{eq}} & \dfrac{\beta_2}{C_{eq}} & \dfrac{\alpha}{C_{eq}} \end{bmatrix} \qquad (7)$$

Discrete state-space representation is:

$$x_{k+1} = A.x_k + B.u_k \qquad (8)$$

Parameters to be identified:

$$\left\{C_{eq},\ \mathrm{R}_{eq},\ R_{HVAC},\ \alpha,\ \beta_1,\ \beta_2\right\}$$

Since no RC model can perfectly capture building energy performance under changing physical conditions, this study proposes incorporating real-time model correction using sensor measurements. A linear Luenberger observer is employed to fuse measured data with model predictions, thereby mitigating the effects of sensor noise, modeling inaccuracies, and external disturbances (e.g., weather variability, occupancy changes, internal heat

gains, and unmodeled HVAC dynamics). The indoor temperature estimate can then be formulated as follows:

$$\hat{x}_{k+1} = A.\hat{x}_k + B.u_k + L(y_k - \hat{y}_k)$$
$$\hat{y}_k = C.\hat{x}_k \quad (9)$$

Local learning objective (client optimization)

Each client independently solves a local prediction-error minimization problem**:**

$$\min_{\theta^i} \sum_{k=1}^{T} \left\| y_k^{(i)} - \hat{y}_k^{(i)}(\theta^{(i)}) \right\| \quad (10)$$

subject to physical constraints on RC parameters (stability, positivity) and local data only.

Federated learning formulation (server-side)

Let: $\theta^{(i)} = \{A^{(i)}, B^{(i)}, L^{(i)}\}$ be the model and observer parameters learned locally. At the communication round r, each client sends $\theta_r^{(i)}$ to the central server

Server aggregation (FedAvg-style)

Federated averaging (FedAvg) computes a weighted average of client parameters:

$$\theta_{global} = \frac{\sum_{i=1}^{N} n_i \theta^{(i)}}{\sum_{i=1}^{N} n_i} \quad (11)$$

n_i is the weight for client $\boldsymbol{i}$, typically number of samples used to train the local model. These aggregated coefficients at the FL server side can be used to represent a global heating demand dynamic of the whole network so it is feasible to optimize the heating energy extracted from the main boilers.

- Let $\theta^{(i)} = \{A^{(i)}, B^{(i)}, L^{(i)}\}$.
- One can compute a global A and B as:

$$A_{global} = \frac{\sum_{i=1}^{N} w_i A^{(i)}}{\sum_{i=1}^{N} w_i}, \quad B_{global} = \frac{\sum_{i=1}^{N} w_i B^{(i)}}{\sum_{i=1}^{N} w_i} \quad (12)$$

- Weights w_i are to be chosen proportional to building thermal mass or typical heat demand.
- No need for L_global, since the observer is only relevant for estimating individual building states.

In the first case, the FL server does not attempt to aggregate building models into a global network representation. Instead, it uses each building's RC model with observer-corrected states to compute real-time optimal setpoints locally. The server sends these setpoints back to the respective clients, allowing each building to regulate its indoor temperature efficiently while respecting occupant comfort and local constraints. Here, the server acts as a coordinator and optimizer, leveraging up-to-date state information to suggest adjustments, but the boiler and network operation are effectively treated as secondary, since the focus is on individual building performance rather than total network-level energy optimization. In the second case, the FL server aggregates the observer-corrected RC models from all connected buildings to construct a global network model. Using this network-level model, the server can simulate how changes in supply temperature, flow rates, or pump operation affect the entire micro-DHC system. With this approach, the server can compute optimal boiler-side control actions that minimize total energy consumption, prevent peak loads, and maintain network stability, while still ensuring buildings receive the heating they need. The observer-based global model allows the server to incorporate real-time deviations in building states, making the control decisions robust to disturbances and heterogeneous building dynamics.

Simulation results

This section details the preliminary simulation results for the edge application of Solution 1. Due to a more streamlined installation and commissioning process, Solution 1 reached the testing phase ahead of Solution 2. Figure 8 illustrates the performance of an RC model in predicting indoor temperatures for a hotel building integrated with a Micro-DHC. Even in its base form – without observer-based corrections – the RC model demonstrates strong predictive accuracy. Future enhancements will involve real-time state observer corrections utilizing micro-DHC heat exchanger data, which will eventually be integrated into an optimal control framework, such as Model Predictive Control (MPC)

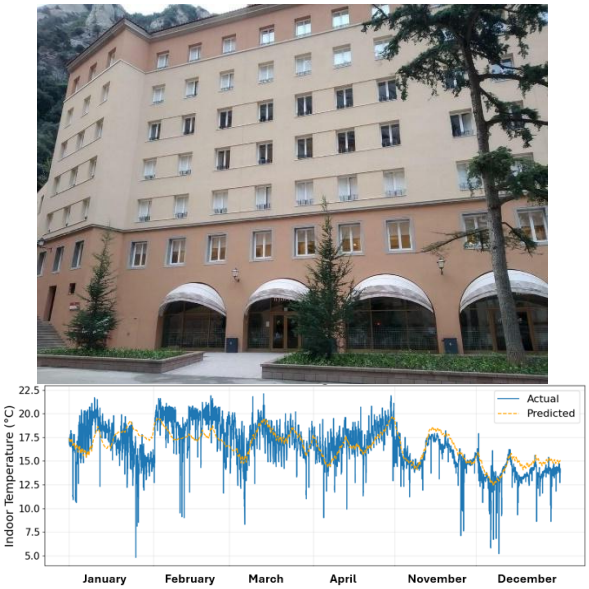


Figure 8: Indoor temperature prediction using trained RC model: (top) the view of the Abat Oliba hotel concerned by the study, (bottom) the seasonal prediction performance for 2023

Fig. 9 illustrates another application featuring an advanced data-driven observer implemented for LTES monitoring. This technique utilizes the KOOPMAN

theory to model the LTES, leveraging collected datasets to construct a linear representation within a lifted observation space. The linearity of this space facilitates the use of a classical Luenberger Observer, which compensates for low-quality measurement data and estimates unmeasured PCM temperature.

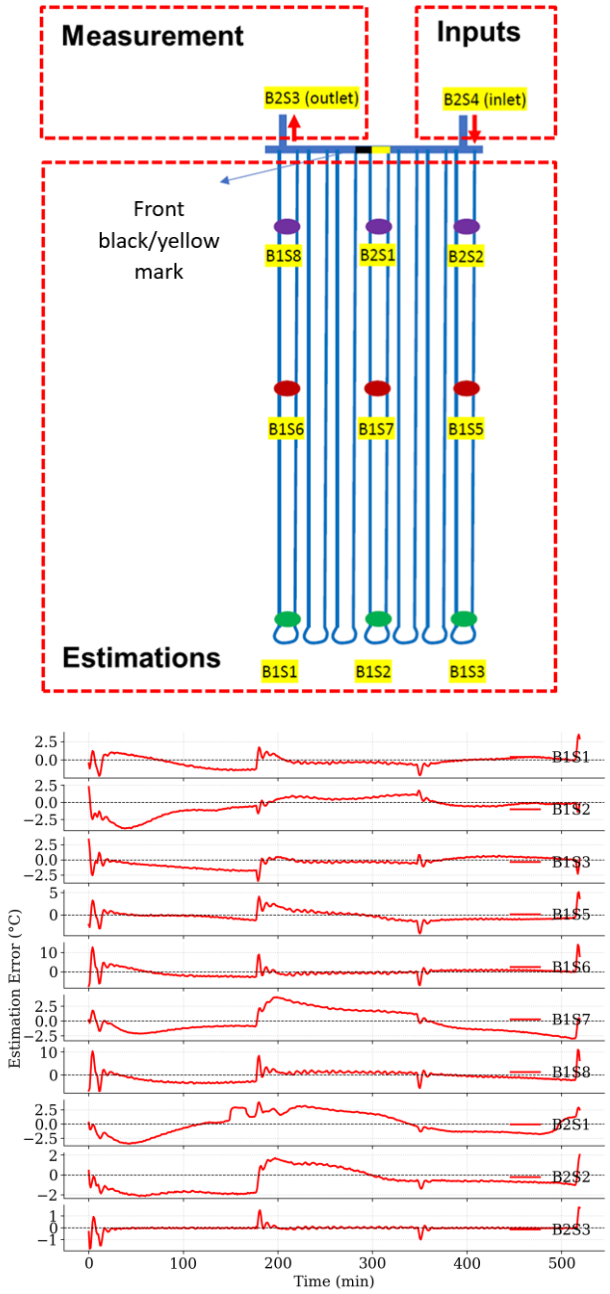


Figure 9: Edge-hosted observer for PCM temperature estimation of LTES (top)sensor placement in the LTES during KOOPMAN model development, (bottom) estimation error for all PCM temperatures

Conclusions

This paper presented two real-world edge-computing implementations developed within the HYSTORE project, targeting coordinated energy management in micro-DHC and residential microgrid environments. In the Montserrat micro-DHC, a centralized edge platform enables network-level monitoring and advanced control of complex thermochemical and PCM thermal energy storage systems using observer-based state estimation and anomaly detection. At the building level, the KTH Live-In Lab implementation demonstrates how similar edge concepts support PCM-TES operation and coordination with heat pumps, photovoltaic generation, and battery storage. Together, these cases show how edge computing can overcome coordination and control limitations of conventional EMSs, enabling observer-driven, privacy-preserving, and scalable control of complex energy systems in isolated energy networks.

Acknowledgments

This work was funded by the European Commission under Horizon Europe (Grant 101096789).

References

C. H. Wong et al, IOP Conf. Ser.: Earth Environ. Sci. 881 012044, 2021, DOI 10.1088/1755-1315/881/1/012044

Lama Alhaj Husein, Imad Alsyouf, Emad Mushtaha and Ahmad Alzghoul, Towards High-Performance Buildings using IoT and AI technologies: A Comprehensive Review, Proceedings of the International Conference on Industrial Engineering and Operations Management Nsukka, Nigeria, 5 - 7 April, 2022.

G. Carvalho, B. Cabral, Vasco Pereira, Jorge Bernardino, Computation offloading in Edge Computing environments using Artificial Intelligence techniques, Engineering applications of artificial intelligence, 2020.

Amira Bourechak, Ouarda Zedadra, Mohamed Nadjib Kouahla, Antonio Guerrieri, Hamid Seridi and Giancarlo Fortino, At the Confluence of Artificial Intelligence and Edge Computing in IoT-Based Applications: A Review and New Perspectives, Sensors 2023, 23, 1639. https://doi.org/10.3390/s23031639

G. Walia, Mohit Kumar, S. Gill, AI-Empowered Fog/Edge Resource Management for IoT Applications: A Comprehensive Review, Research Challenges, and Future Perspectives, EEE Communications Surveys and Tutorials, 2024.

Balakumar Muniandi, Purushottam Kumar Maurya, CH Bhavani, Shailesh Kulkarni, Ramswaroop Reddy Yellu, Dr. Nidhi Chauhan. AI-Driven Energy Management Systems for Smart Buildings, Power System Technology, Volume 48 Issue 1, March 2024.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A DATA-DRIVEN KOOPMAN-MPC FOR MULTI-ACTUATOR AHU CONTROL

Youssef Elomari[1], Mustapha Habib[1], and Qian Wang[1,2]
[1]Division of Building Technology and Design, Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, 114 28, Stockholm, Sweden
[2]Uponor AB, Hackstavägen 1, Västerås, 721 32, Sweden

Abstract

Air-handling units (AHUs) are major energy consumers in HVAC systems due to their coupled actuators and constrained multi-input multi-output operation. This paper proposes a data-driven Koopman-MPC framework that identifies lifted linear dynamics from building management system (BMS) data using EDMDc, enabling constrained MPC design for nonlinear AHU behavior. The approach is validated on a residential AHU in Stockholm, Sweden. Results demonstrate the feasibility of deploying Koopman-based predictive control using operational data, and a BMS-integrated edge implementation workflow is presented for practical building applications.

Introduction

In recent years, data-driven methods have become increasingly important for the operation and optimization of heating, ventilation, and air conditioning (HVAC) systems (Afram and Janabi-Sharifi, 2014). Air-handling units (AHUs) represent a critical component of building HVAC systems, responsible for conditioning and distributing air throughout building zones. Modern AHUs incorporate multiple actuators, including supply and return fans, mixing dampers, heating and cooling coils, and humidifiers, creating a complex multi-input multi-output (MIMO) control problem (Afram and Janabi-Sharifi, 2014; Matetić et al., 2022). Coordinating these actuators to achieve multiple objectives—energy efficiency, thermal comfort, indoor air quality poses significant challenges for conventional control approaches (Yue et al., 2023). Despite this multivariable and constrained nature, AHU control in many buildings still relies on decentralized proportional–integral–derivative (PID) and rule-based controllers implemented in the building management system (BMS) (Zou, Yu and Ergan, 2020). These strategies are robust and easy to commission, but they do not coordinate multiple actuators or explicitly handle constraints. At the same time, the increasing availability of high-resolution BMS data has accelerated progress in data-driven control methods. Most of the literature has focused on fault detection and diagnosis and on supervisory energy management and operational decision-making at scale (Deshmukh, Glicksman and Norford, 2020; Taheri et al., 2021; Matetić et al., 2022).Several studies have attempted to address AHU optimal control using data-driven methods. Optimization-based approaches using simplified models have been demonstrated for AHU regulation in simulation environments, often relying on reduced polynomial structures that cannot represent all internal AHU interactions(Zou, Yu and Ergan, 2020). Learning-based control has also been explored through deep reinforcement learning trained on operational datasets, reporting improvements relative to baseline operation while maintaining predicted discomfort around 10% using an LSTM-based surrogate (Zou, Yu and Ergan, 2020). Neural-network predictive controllers based on dynamic recurrent models have reported up to 69.9% energy reduction on historical data, while also indicating that nonlinear predictors inside the controller (Asvadi-Kermani et al., 2022). (Zheng, Wang and Wang, 2024) conducted a systematic comparison of model-based and data-driven economic MPC using the BOPTEST framework. Their results showed that data-driven methods achieved energy cost reductions within 5–10% Overall, the literature suggests that data-driven control can deliver substantial gains. However, robust AHU optimal control remains challenging due to nonlinear multivariable dynamics.

Koopman operator theory provides a promising route to address these challenges by representing nonlinear dynamics through linear evolution in a lifted space of observables (Wahba et al., 2023). The lifting map captures the nonlinear structure, while prediction evolves linearly in lifted coordinates, enabling the use of standard linear MPC solvers with explicit constraints. A finite-dimensional approximation of the Koopman operator can be learned directly from operational data using extended dynamic mode decomposition with control (EDMDc), which estimates lifted linear dynamics from measured state–input snapshot pairs (Zawacki and Abed, 2023).

In this paper, we develop a Koopman–MPC framework for coordinated multi-input supervisory control of an AHU using operational BMS data from the KTH Live-In Lab. We identify the Koopman predictor via EDMDc, embed the lifted linear dynamics into an MPC formulation, and regulate supply air temperature and exhaust airflow under actuator constraints. Closed-loop performance is evaluated using a high-fidelity model trained on measured data. We also present an edge implementation workflow compatible with BMS-integrated operation

Methodology

This section presents a Koopman–MPC framework for coordinated multi-input control of an air-handling unit (AHU). A finite-dimensional approximation of the Koopman operator is identified from operational building management system (BMS) data using extended dynamic mode decomposition with control (EDMDc). Closed-loop performance is evaluated using a high-fidelity neural state-space (NSS) surrogate model trained on measured data. Figure 1 summarizes the overall architecture and data flow.

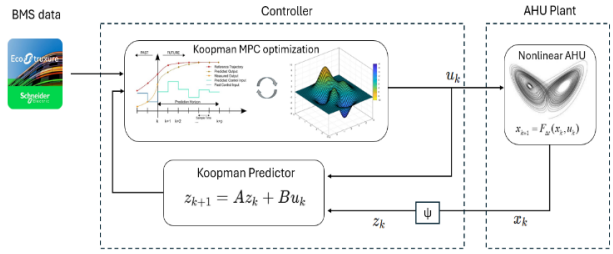


Figure 1: Overview of the proposed Koopman–MPC architecture

System definitions

The AHU is modeled as a nonlinear discrete-time system sampled at interval Δt

$$x_{k+1} = F_{\Delta t}(x_k, u_k)$$

where x_k denotes the system state and u_k the control input at time step k and Δt corresponds to the BMS sampling interval.

Based on available BMS measurements, the state and input vectors are defined as

$$x_k = \begin{bmatrix} T_{\text{sup},k} & F_{\text{sup},k} & T_{\text{rec},k} & T_{\text{exh},k} & F_{\text{exh},k} \end{bmatrix}^{\cdot}$$

$$u_k = \begin{bmatrix} V_{\text{heat},k} & T_{\text{heat},k} & S_{\text{sup},k} & S_{\text{wheel},k} & S_{\text{exh},k} \end{bmatrix}^{\cdot}$$

The control objective focuses on regulating supply air temperature and exhaust airflow, which are defined as the controlled outputs

$$y_k = \begin{bmatrix} T_{\text{sup},k} \\ F_{\text{exh},k} \end{bmatrix}$$

These variables directly affect indoor comfort and energy use and are therefore selected as the primary control targets. This multi-input setup matches the practical control objective. The controller adjusts the heating coil, fans, and heat-recovery wheel in a coordinated way rather than relying on a single actuator.

Koopman-based predictor for MPC

Koopman theory represents nonlinear state evolution through linear evolution of observable functions. For a discrete-time system, the Koopman operator $K_{\Delta t}$ acts on an observable g as

$$(K_{\Delta t} g)(x_k, u_k) = g\big(F_{\Delta t}(x_k, u_k)\big)$$

A practical predictor requires a finite-dimensional approximation. Following the EDMD framework, we select a dictionary of basis functions and project the Koopman operator onto its span.

We define a lifting map ψ that maps the normalized state into a higher-dimensional feature vector

$$z_k = \psi(x_k)$$

where ψ consists of polynomial basis functions selected to capture dominant nonlinearities observed in the data. The lifting map uses normalized measured states and a second-order polynomial dictionary with linear, squared, and pairwise interaction terms. All variables are normalized using the training-data mean and standard deviation. For five measured states, this gives a lifted-state dimension of 21 including the constant term.Using EDMDc, the lifted dynamics are approximated by the linear controlled system

$$z_{k+1} = Az_k + Bu_k$$

$$y_k \approx Cz_k$$

Where matrices A and B describe the evolution of the lifted state under the influence of the control inputs, while C maps the lifted coordinates back to the physical outputs. To identify the Koopman system matrices, snapshot pairs are constructed from the operational dataset

$$Z = \begin{bmatrix} z_1 & \cdots & z_{N-1} \end{bmatrix}, Z^{+} = \begin{bmatrix} z_2 & \cdots & z_N \end{bmatrix}$$

$$U = \begin{bmatrix} u_1 & \cdots & u_{N-1} \end{bmatrix}$$

The lifted system matrices are obtained by solving a regularized least-squares problem

$$\min_{A,B} \left\| Z^{+} - \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} Z \\ U \end{bmatrix} \right\|_F^2 + \lambda \left\| \begin{bmatrix} A & B \end{bmatrix} \right\|_F^2$$

Koopman-MPC formulation

At each control step k, MPC optimizes a sequence of future control inputs over a finite horizon. The Koopman–MPC optimization problem is formulated as

$$\min_{\{u_{k+i|k}\}_{i=1}^{N_p}} \sum_{i=1}^{N_p} \Big[(y_{k+i|k} - y_{r,k+i})^{\cdot} Q (y_{k+i|k} - y_{r,k+i}) + (u_{k+i|k} - u_{k+i-1|k})^{\cdot} R (u_{k+i|k} - u_{k+i-1|k}) \Big]$$

Where $y_{r,k+i}$ denotes the reference trajectory for supply air temperature and exhaust airflow. The first term penalizes tracking error, while the second term penalizes aggressive variations in the control inputs.

The optimization is subject to actuator magnitude and rate constraints

$$u_{\min} \le u_{k+i|k} \le u_{\max}$$

$$\Delta u_{\min} \le u_{k+i|k} - u_{k+i-1|k} \le \Delta u_{\max}$$

The Koopman–MPC uses a 15 min sampling time and a prediction horizon of 4 steps. The optimization is solved using sequential quadratic programming (SQP). Following our earlier AHU MPC setup on the same testbed, the baseline weighting is set to $Q = 1$and $R = 10^{-5}[1,1,1,1,1]$.

High-fidelity AHU model

To evaluate the Koopman-MPC under realistic conditions, a neural state-space (NSS) model is used as a high-fidelity surrogate of the AHU dynamics

$$x_{k+1} = f_\theta(x_k, u_k),\ y_k = h_\theta(x_k)$$

where f_θ and h_θ are nonlinear mappings parameterized by neural network weights θ. The NSS model is trained using one-step-ahead prediction on one year of measured BMS data with a 15-minute sampling resolution.

The NSS model is used solely for closed-loop performance evaluation and benchmarking and is not employed for prediction or control within the Koopman–MPC framework.

Case study

The proposed Koopman–MPC framework is evaluated on an AHU installed at the KTH Live-In Lab. The system serves four student apartments and consists of a rotary heat exchanger, variable-speed supply and exhaust fans, and a water-based heating coil, as illustrated in Figure 2. Historical operational data were collected from the building management system (BMS) at a 15-minute resolution over three years. The dataset includes supply and exhaust air temperatures, airflow rates, heating-coil rate and temperature, fan speeds, and wheel speed. These measurements are used to identify the EDMDc Koopman predictor and to train the NSS surrogate model.

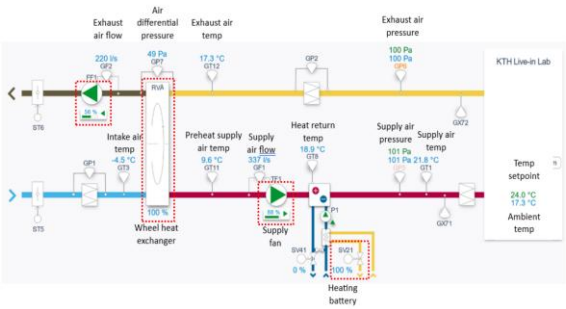


Figure 2: SCADA view of the AHU installed at the KTH Live-In Lab, illustrating the rotary heat exchanger, heating coil, and variable-speed supply and exhaust fans. The highlighted components correspond to the actuators considered in the proposed Koopman–MPC framework

Results

This section presents the results obtained using the proposed Koopman–MPC framework applied to the real AHU system use case described in Section 2. The analysis focuses on validating the high-fidelity neural state-space (NSS) model used as a surrogate plant, evaluating the predictive capability of the EDMDc-based Koopman linear predictor, and assessing the closed-loop performance of the MPC controller under realistic operating constraints.

The predictive accuracy of the NSS model was first assessed to ensure that it provides a reliable reference for closed-loop controller evaluation. The model was trained using one-step-ahead prediction on historical BMS data with a 15-minute sampling interval and validated on unseen data segments covering a wide range of airflow rates, heat-recovery operating points, and heating-coil conditions. Figure 3 compares measured and simulated air-temperature trajectories for the supply, recycled, and exhaust air streams. Minor deviations between measured and simulated signals are primarily observed during periods of rapid airflow variation, where sensor dynamics and short-term turbulence effects introduce additional uncertainty. Despite these effects, the overall agreement remains strong. The mean absolute error for supply-air temperature remains below 0.5 °C, while exhaust airflow prediction errors stay within 5 % over the validation period. These results confirm that the NSS model accurately represents the nonlinear behavior of the full-scale AHU and can be used as a high-fidelity surrogate plant for evaluating the proposed Koopman–MPC strategy.

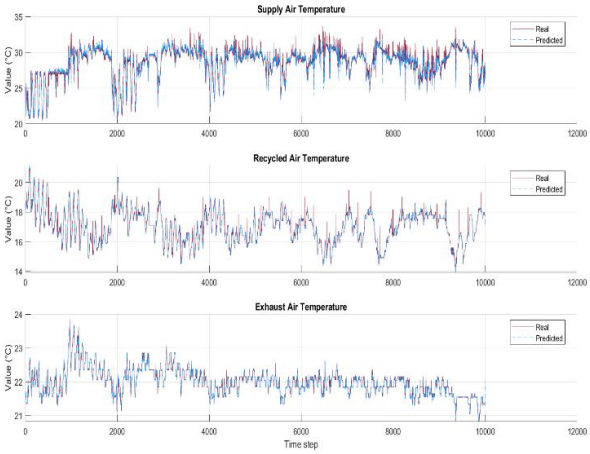


Figure 3: Comparison between measured and NSS-simulated air temperatures (supply, recycled, and exhaust) used for model validation

The multi-step prediction capability of the Koopman-based linear model identified via EDMDc was then evaluated. The lifted predictor is embedded in the MPC to forecast future temperature and airflow trajectories over the prediction horizon. The results show that the Koopman model maintains stable and physically consistent predictions despite the strong nonlinear coupling between airflow, heat recovery, and heating-coil operation. In particular, the predictor accurately captures the influence of airflow modulation on supply-air temperature, which is critical for coordinated multi-input control. As expected, prediction errors increase gradually with horizon length; however, they remain sufficiently small to support reliable MPC optimization.

Figure 4 illustrates the system response during a representative winter operating sequence. The controller successfully maintains both the supply-air temperature and supply airflow close to their respective setpoints while coordinating the heating coil, supply fan, exhaust fan, and rotary heat-exchanger speed. The MPC dynamically adjusts actuator combinations in response to disturbances and load variations, ensuring stable operation without violating actuator magnitude or rate

constraints.

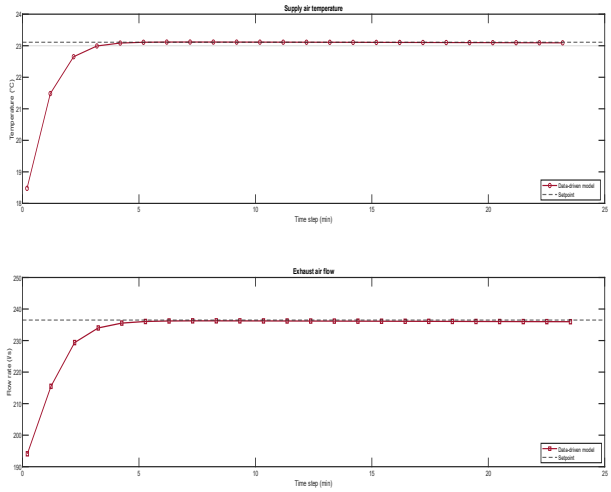


Figure 4:Closed-loop Koopman–MPC response showing supply-air temperature tracking and exhaust-airflow regulation during frost-risk conditions

Technical deployment consideration and edge implementation workflow

This section discusses the technical considerations for deploying the proposed Koopman–MPC framework within a building management system (BMS) using an edge computing architecture. While the control strategy is evaluated in a co-simulation environment in this work, the deployment design is presented to demonstrate practical feasibility and compatibility with existing industrial automation infrastructure.

BMS-integrated edge architectures

The proposed Koopman–MPC framework is designed to operate within a BMS-integrated edge control panel. The hardware architecture, illustrated in Figure 5, consists of a programmable logic controller (PLC), an industrial PC (IPC), and a gateway device.

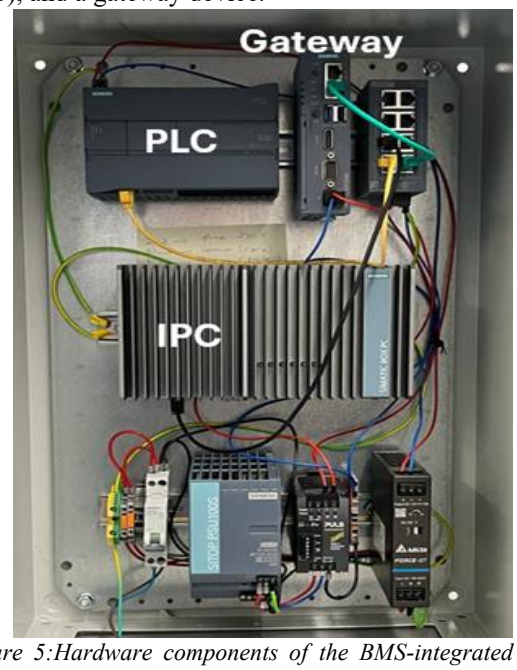


Figure 5:Hardware components of the BMS-integrated edge control architecture considered for Koopman–MPC deployment

The PLC acts as the primary interface with the physical AHU and the BMS. It is responsible for acquiring real-time measurements, enforcing actuator limits and safety logic, and applying control setpoints to the AHU actuators. In the proposed architecture, the PLC retains full authority over final actuation, ensuring safe operation regardless of the control strategy used at higher levels.

The IPC is intended to host the Koopman–MPC algorithm. Its computational resources are sufficient to perform lifted-state prediction and constrained optimization within the control sampling interval considered in this study.

The gateway device enables standardized communication between the PLC, IPC, and supervisory systems. It supports protocol translation between field-level communication (e.g., Modbus) and higher-level interfaces such as OPC UA, facilitating data exchange with the BMS.

Data flow and control integration

The data flow between the AHU, edge devices, and BMS is illustrated in Figure 6. At each sampling interval, the PLC acquires measurements from the AHU and relevant BMS controllers and makes them available to the IPC through the gateway. These measurements constitute the input to the Koopman–MPC algorithm, together with reference setpoints defined at the BMS level.

Based on the received measurements, the IPC executes the Koopman–MPC optimization using the EDMDc-identified linear predictor. The resulting control setpoints are transmitted back to the PLC, where they are evaluated against safety rules, actuator bounds, and operational constraints. If accepted, the setpoints are applied to the AHU actuators; otherwise, the PLC maintains the current operating point or falls back to standard BMS control logic.

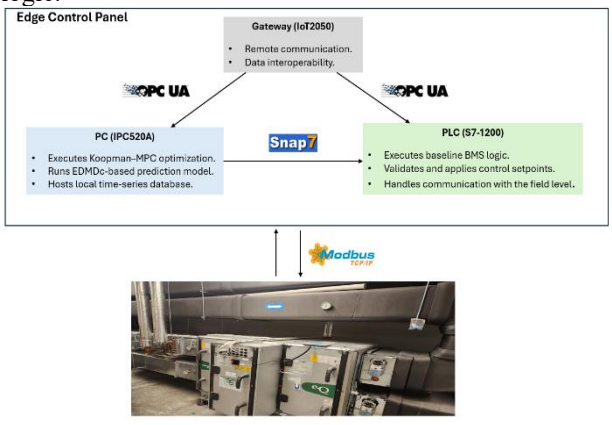


Figure 6:Edge Control Panel architecture used communication protocols for Koopman–MPC integration with the BMS

Conclusions

This paper proposed a data-driven Koopman–MPC framework for coordinated multi-input control of an AHU using operational BMS data. The approach identifies a finite-dimensional linear predictor in lifted space via EDMDc and embeds it in a constrained MPC formulation. The resulting controller targets supply air temperature and

exhaust airflow while coordinating multiple AHU actuators within practical operating limits.

A high-fidelity NSS surrogate trained on measured data was used to evaluate closed-loop performance under realistic nonlinear AHU dynamics. The results indicate that the EDMDc-based Koopman predictor delivers stable short-horizon forecasts of the controlled outputs across representative operating conditions, enabling effective MPC-based regulation.

From a deployment perspective, the workflow is compatible with BMS-integrated edge execution. EDMDc identification and NSS training can be performed offline using historical BMS datasets, whereas online operation requires only state lifting, linear prediction, and constrained optimization at the BMS sampling interval. This separation reduces computational requirements and supports integration with standard automation hierarchies in which PLC-level logic enforces safety constraints and fallback operation.

Future work will focus on experimental validation on the real AHU, with the Koopman–MPC implemented as a supervisory layer in real time and benchmarked against the baseline BMS control. This will enable a quantitative assessment of comfort tracking, actuation behavior, and energy-related performance under comparable disturbance conditions.

Acknowledgements

This work was funded by the European Commission under Horizon Europe (Grant 101096789).

Nomenclature

AHU	Air-Handling Unit
BMS	Building Management System
DRL	Deep Reinforcement Learning
EDMDc	Extended Dynamic Mode Decomposition with Control
HVAC	Heating, Ventilation, and Air Conditioning
IPC	Industrial PC
LSTM	Long Short-Term Memory
MIMO	Multi-Input Multi-Output
MPC	Model Predictive Control
NSS	Neural State-Space
PID	Proportional–Integral–Derivative
PLC	Programmable Logic Controller

References

Afram, A. and Janabi-Sharifi, F. (2014) “Theory and applications of HVAC control systems – A review of model predictive control (MPC),” Building and Environment, 72, pp. 343–355. Available at: https://doi.org/10.1016/j.buildenv.2013.11.016.

Asvadi-Kermani, O. et al. (2022) “Energy Optimization of Air Handling Units Using Constrained Predictive Controllers Based on Dynamic Neural Networks,” IEEE Access, 10, pp. 56578–56590. Available at: https://doi.org/10.1109/ACCESS.2022.3177660.

Deshmukh, S., Glicksman, L. and Norford, L. (2020) “Case study results: Fault detection in air-handling units in buildings,” Adv. Build. Energy Res., 14(3), pp. 305–321. Available at: https://doi.org/10.1080/17512549.2018.1545143.

Matetić, I. et al. (2022) “A Review of Data-Driven Approaches and Techniques for Fault Detection and Diagnosis in HVAC Systems,” Sensors 2023, Vol. 23, 23(1). Available at: https://doi.org/10.3390/s23010001.

Taheri, S. et al. (2021) “Fault detection diagnostic for HVAC systems via deep learning algorithms,” Energy Build., 250, p. 111275. Available at: https://doi.org/10.1016/j.enbuild.2021.111275.

Wahba, N. et al. (2023) “Efficient HVAC system identification using Koopman operator and machine learning for thermal comfort optimisation,” Building and Environment, 242(1), p. 110567. Available at: https://doi.org/10.1016/j.buildenv.2023.110567.

Yue, B. et al. (2023) “Energy-oriented control retrofit for existing HVAC system adopting data-driven MPC – Methodology, implementation and field test,” Energy and Buildings, 295, p. 113286. Available at: https://doi.org/10.1016/j.enbuild.2023.113286.

Zawacki, C.C. and Abed, E.H. (2023) “Dynamic Mode Decomposition for Control Systems with Input Delays,” IFAC-PapersOnLine, 56(3), pp. 97–102. Available at: https://doi.org/10.1016/j.ifacol.2023.12.007.

Zheng, W., Wang, D. and Wang, Z. (2024) “Economic model predictive control for building HVAC system: A comparative analysis of model-based and data-driven approaches using the BOPTEST Framework,” Applied Energy, 374, p. 123969. Available at: https://doi.org/10.1016/j.apenergy.2024.123969.

Zou, Z., Yu, X. and Ergan, S. (2020) “Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network,” Building and Environment, 168, p. 106535. Available at: https://doi.org/10.1016/j.buildenv.2019.106535.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

CAUSALITY-INFORMED TIME-SERIES FOUNDATION MODELS FOR BUILDING PERFORMANCE PREDICTION

Dennis Grauer and Joern Ploennigs
University of Rostock

Abstract

Time-Series Foundation Models, pretrained on large amounts of time-series data, offer strong zero-shot forecasting capabilities. Recent TSFM versions provide improved support for covariate forecasting. This renders them a promising alternative to address current limitations of data-driven building performance prediction approaches: they are highly dependent on data availability, are mostly case-specific, and often favor correlation over causation. In this paper, we compare the predictive performance of two TSFMs with graph neural network approaches and a statistical baseline. We also present an approach for configuring causal covariates for TSFM predictions from graph representations. All models are evaluated using two real-world building datasets.

Introduction

Given the challenges of climate change, it is imperative to improve the energy performance of buildings. Accurate predictions of indoor environmental quality (IEQ) and building energy use can help identify strategies to reduce the environmental impact of buildings (Ali et al., 2024) while maintaining a healthy indoor environment for occupants. With an increasing number of Internet of Things (IoT) sensors deployed in modern buildings, machine learning (ML) approaches that leverage historical data have emerged as an alternative to traditional physical simulations.

Although these approaches perform well on IEQ and building energy management (BEM) tasks, they face two major challenges. First, models often overfit to spurious correlations in the data due to the high redundancy of sensors in buildings (Chen et al., 2025). For example, it is easier for models to learn correlations among temperature sensors than dependencies on heating and cooling systems. Second, models are often limited by the availability and quality of sensor data (Ogundiran et al., 2024; Semasinghe et al., 2025). Consequently, models in the building domain are highly case-specific because they need to be configured for the available sensor data of each building (Ogundiran et al., 2024). Both issues make it difficult to transfer models between buildings, as they do not sufficiently learn generalizable physical causal relationships but instead capture case-specific correlations.

Recently, Time-Series Foundation Models (TSFMs) such as Chronos (Ansari et al., 2024, 2025), TimesFM (Das et al., 2024), Moment (Goswami et al., 2024), and Moirai (Woo et al., 2024) have emerged as an option to address the data availability challenge. They pose to require less training data, as they are pretrained on a variety of time-series datasets from different domains, similar to the transformer architectures of Large Language Models (LLMs) that are trained on large amounts of natural language data. This enables them to make zero-/few-shot forecasts by leveraging a short context window of previous time steps. With increased support for co- and multivariate forecasting in newer model versions, TSFMs promise to incorporate external factors and to predict multiple time series simultaneously.

In this paper, we investigate whether their out-of-the-box performance provides a suitable alternative for short-term forecasting tasks related to Indoor Environmental Quality (IEQ) and Heating, Ventilation, and Air Conditioning (HVAC) loads. We compare the predictive performance of (i) two state-of-the-art TSFMs, (ii) the Autoregressive Integrated Moving Average (ARIMA) model as a strong statistical baseline for time-series forecasting, and (iii) Graph Neural Networks (GNNs) combined with sequence-to-sequence learning methods as a simple way to encode physical causal relationships into machine learning models. By evaluating forecasting accuracy across four tasks, split across two real-world building datasets and varying forecasting horizons, we aim to explore not only the applicability of TSFMs in general but also whether they can be combined with causality-informed approaches such as GNNs. To this end, we *propose a causality-informed TSFM approach* that automatically configures TSFM features from graphs. Our evaluation is conducted on two publicly available datasets: a residential building with sparse HVAC data coverage and no occupancy ground truth, and an office building with diverse HVAC operational data and indoor environment measurements across multiple thermal zones.

This paper is structured as follows. We start with a review of related work on Time-Series Foundation Models and data-driven indoor environment and building energy performance prediction. We then describe our methodology by defining the models and forecasting tasks, followed by an introduction to the scenarios. After presenting and discussing the results, the paper concludes with an outlook.

Background and related works

Indoor environment predictions, HVAC and energy efficiency

Indoor environmental quality is a key factor in ensuring occupants' comfort and health. IEQ is largely governed by heating, ventilation, and air conditioning (HVAC) systems, which also account for a major share of energy use in modern buildings. This relevance makes the resulting complex relationships an important subject of research. Data-driven models that learn from historical data without requiring detailed building information have been employed for various tasks in the IEQ and energy-efficiency domain, including data collection and processing, fault and anomaly detection and analysis, HVAC control management, design optimization, and predictive analysis (Ogundiran et al., 2024). In this paper, we focus on the latter by predicting different IEQ variables and HVAC energy use. HVAC-related predictions mainly involve heating and cooling loads as well as energy consumption, which have been tackled by researchers using deep learning methods (Shi et al., 2024; Liu et al., 2023), neural networks (Borowski and Zwolińska, 2020; Irshad et al., 2024), and graph neural networks (Jia et al., 2024a,b). IEQ prediction is a diverse field, including indoor temperature, CO_2, particulate matter, and thermal comfort. While approaches such as neural networks (Cho and Moon, 2022; Tagliabue et al., 2021; Elmaz et al., 2021) and graph neural networks (Zhang et al., 2023) show convincing results, multiple reviews highlight a key limitation: these approaches remain highly case-specific and are sensitive to the availability and quality of data (Li et al., 2025; Latoń et al., 2025; Ogundiran et al., 2024; Semasinghe et al., 2025). This raises the question of whether zero-/few-shot models could provide an alternative that alleviates this limitation.

Time series foundation models

Time-Series Foundation Models (TSFMs) have recently emerged as a zero-shot alternative for time-series tasks. Built on transformer architectures, they are pretrained on a diverse collection of time-series datasets, enabling them to forecast, detect anomalies, and classify time series. The underlying idea is that TSFMs learn generalizable patterns and correlation structures from these datasets. This can enable zero-shot forecasting, as the model only needs to identify patterns that it has encountered before. In addition, TSFMs support transfer learning by fine-tuning them on new datasets and tasks with fewer data, as the model needs to learn the new task but not the patterns from scratch. Beyond univariate forecasting, models such as Moirai (Woo et al., 2024) and the most recent version of Chronos (Ansari et al., 2025) support integrating past and future covariates into forecasting pipelines.

This suggests that TSFMs could be applicable to different prediction tasks in buildings. However, existing studies report mixed results. Mulayim et al. (2025) evaluate TSFMs for building energy management tasks and find limited zero-shot generalization; TSFMs often fail to outperform classical statistical models. Similarly, Lin et al. (2024) show that TSFMs can handle heterogeneous building data if they are well configured. Extending the analysis to spatio-temporal forecasting, Gupta et al. (2025) report that TSFMs perform competitively at high sampling frequencies but degrade in settings with sparse spatial sensor coverage. Overall, these studies suggest that TSFMs do not necessarily work out of the box but require careful parameterization. This motivates our research question: Can incorporating causal relationships improve TSFM performance in building settings?

Methodology

The aim of this paper is to evaluate the accuracy of TSFMs on different IEQ and energy-consumption tasks and to investigate whether causal information can improve performance. To this end, we perform univariate forecasting with TimesFM and Chronos as two state-of-the-art TSFMs, and covariate forecasting with Chronos using past and future covariates. On the GNN side, we compare two different GNN models as well as our newly proposed approach, which uses GNN graph-representation nodes as covariates for Chronos. We compare these deep learning models with the well-established Autoregressive Integrated Moving Average (ARIMA) model (Newsham and Birt, 2010) as a strong statistical baseline.

Since we want to focus on a comparison of data-driven methods, white-box models are excluded from the scope of this paper. GNN models are favored over other deep learning models, as they allow encoding causal relations in a rather simple manor.

The general task is to predict the next values over a prediction duration P from a set of previous values spanning a context duration D. Since we focus on short-term forecasting in this study, we set $P \in \{1\,\mathrm{h}, 2\,\mathrm{h}, 4\,\mathrm{h}, 8\,\mathrm{h}\}$ to provide an overview of general trends in model forecasting capabilities. The context duration D is defined separately for each model. Depending on the case-specific sampling rate f_s (in minutes; see Scenarios), the context window length C and the horizon length H are defined as $C = (60D)/f_s$ and $H = (60P)/f_s$. In compliance with other work in this domain, we use root mean squared error (RMSE) as the evaluation metric.

Graph neural network models

We employ two types of GNN models: (i) a spatiotemporal GNN (GRUGCN) that uses a Gated Recurrent Unit (GRU) layer to capture temporal regression and a graph convolutional layer, and (ii) a two-layer Graph Convolutional Network with no mechanism to capture temporal dependencies. For both models, we define a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The node set $\mathcal{V}$ denotes a subset $\{v_1, v_2, \ldots, v_i\}$ of available time-series signals from sensor measurements. Each node is associated with a node-feature vector that contains the previous values $\{x_{t-C}, \ldots, x_{t-1}\}$. The edge set $\mathcal{E}$ denotes a set of directed edges $\{e_1, e_2, \ldots, e_i\}$.

Each edge $e = (v_i, v_j)$ represents a directed connection from node v_i to node v_j. We construct these connections based on domain knowledge of physical relationships between the measured signals and dataset exploration, aiming to obtain a graph that captures causal relationships between sensor measurements. Although ontologies like Brick (Balaji et al., 2016) offer schemes for building metadata, they do not cover physical relations and causalities, making automatic graph construction based on causality a complex problem (Langbridge et al., 2023; Grauer and Ploennigs, 2025), not covered in the scope of this paper.

For both models, the context duration is set to $D \in \{4\,\mathrm{h}, 8\,\mathrm{h}\}$. Since we consider different datasets and prediction targets, each requires careful hyperparameter tuning. To this end, we employ the Optuna framework (Akiba et al., 2019) to optimize the learning rate and hidden size of each model layer.

Time series foundation models and ARIMA

We use TimesFM (Das et al., 2024) and Chronos (Ansari et al., 2025) as state-of-the-art TSFMs in our evaluation. We use the most recent versions available at the time of writing, namely TimesFM 2.5-200m and Chronos 2. While TimesFM only supports univariate forecasting, Chronos 2 can incorporate past and future covariates (real-valued or categorical) to support multivariate forecasting. Its in-context learning (ICL) capability shares information within a group of multivariate time series via a group-attention mechanism. We test the effectiveness of this mechanism by adding graph nodes as past covariates to Chronos. Therefore, we define the causal-covariate forecasting task as follows:

$$\hat{y}_{causal} = Chronos(D_{target}, \mathcal{S}) \tag{1}$$

where D_{target} is defined as the previous time-steps of the prediction target and $\mathcal{S}$ is defined as a subset of $\mathcal{V}$:

$$\mathcal{S} = \{v_i \mid \exists e_{i,target} \in \mathcal{E}\} \quad \forall v_i, v_k \in \mathcal{V} \tag{2}$$

where $e_{i,target}$ denotes an edge directed from node v_i to the target node v_{target}. The full pipeline for configuring causal covariates from graph representations is given in Algorithm 1.

Besides the causal approach, we define two additional Chronos covariate settings: one that uses only two past covariates for each prediction target, and one that adds past covariates as well as known future covariates. While past-only covariates add information to the context window D, future covariates are incorporated over the prediction horizon H. To create realistic scenarios, we restrict future covariates to quantities that can be estimated (e. g., future temperature or dew point using weather forecasts), are controlled by building operators (e. g., air supply rates or heating setpoints), or follow time patterns (e. g., indicators for working hours or day of the week).

TimesFM serves as a comparison to Chronos, as it is designed and trained for univariate time series only. We set $D = 2048$ for both models, as this is the maximum context length used in the first pretraining stage of Chronos (Ansari et al., 2025). Although TimesFM 2.5 supports context lengths of up to 16,000 samples, our exploratory experiments suggest that longer context lengths do not necessarily improve performance while increasing computational cost. For ARIMA, we set $D = 1024$. To simplify forecasting, we use AutoARIMA (Smith et al., 17), which automatically fits the model parameters to the data.

Since the data for the GNN models are split into training and test subsets, TSFM and ARIMA predictions are made only on the test subset to ensure a fair comparison across all model types.

Algorithm 1 Causal-covariate TSFM forecasting

Require: Directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, target node v_{target}, historical data matrix $\mathrm{X} \in \mathbb{R}^{|\mathcal{V}| \times T}$, context length C, horizon H

Ensure: Forecast $\hat{y}_{t:t+H}$ for v_{target}

$D_{target} \leftarrow \mathrm{X}[v_{target},\ t{-}C : t{-}1]$ ▷ Target history
$\mathcal{S} \leftarrow \emptyset$ ▷ Causal covariate set
for each $v_i \in \mathcal{V}$ **do**
  if $(v_i, v_{target}) \in \mathcal{E}$ **then**
    $\mathcal{S} \leftarrow \mathcal{S} \cup \{\mathrm{X}[v_i,\ t{-}C : t{-}1]\}$
  end if
end for
$\hat{y}_{t:t+H} \leftarrow \mathrm{Chronos}(D_{target},\ \mathcal{S})$
return $\hat{y}_{t:t+H}$

Scenarios

Residential building

We use the dataset presented by Pandey et al. (2025) to evaluate forecasting performance in residential buildings. The dataset contains sensor data capturing occupant behavior, energy use, and IEQ for eight student dorm units over a period of two years. The data are divided into pre- and post-retrofit periods. During the retrofit, heat pumps were installed alongside a heat recovery ventilation (HRV) system, and the insulation of walls and windows was significantly improved. For our prediction tasks, we use the post-retrofit period, as the HVAC and HRV deployment provides more sensor data for the GNN models. Although energy use for HVAC and HRV is available after the retrofit, other operational signals (e. g., HVAC supply air volume or temperature) are missing. The available sensors include outdoor climate conditions (temperature [°C], solar radiation [W/m²], dew point [°C], humidity [%]), indoor environment variables (relative humidity [%], illuminance, CO_2 [ppm], TVOC [ppb], temperature [lx]), energy use of electrical appliances [kWh], and occupant behavior (door and window state [binary, indicating open or closed]). After data cleaning and preparation, we retain 8.5 months of data sampled at $f_s = 10\,\mathrm{min}$. We use a 70:30 split for training and validation/testing, as recommended for the size of the cleaned dataset (Muraina, 2022). We

aim to predict the following target variables:

- **Indoor relative humidity [%]:** We predict the relative humidity in the living room of the student dorm. Input features for the GNN models and causal-covariate setting include outdoor relative humidity [%], dew point temperature [°C], room temperature [°C], HRV power use [kWh], and window state (0 = closed, 1 = open). Outdoor temperature and dew point temperature are used as both past and future covariates for Chronos.

- **HVAC energy use [Wh]:** We predict the combined energy use of electric baseboard heaters in the dorm and heat pumps. Input features include outdoor temperature [°C], relative humidity [%], dew point temperature [°C], indoor temperature [°C], and indoor relative humidity [%]. Outdoor temperature and humidity are used as past and future covariates.

The graph structures used for the GNNs are shown in Figure 2. Due to sparse operational information on ventilation, the graph remains relatively simple. Moreover, as this is a residential building, humidity and HVAC energy use depend strongly on occupant behavior, daily schedules, habits, and personal preferences. In particular, HVAC energy use is a challenging stochastic prediction task: it exhibits spikes at certain time steps and remains close to zero at others, with no clear seasonality (Figure 1).

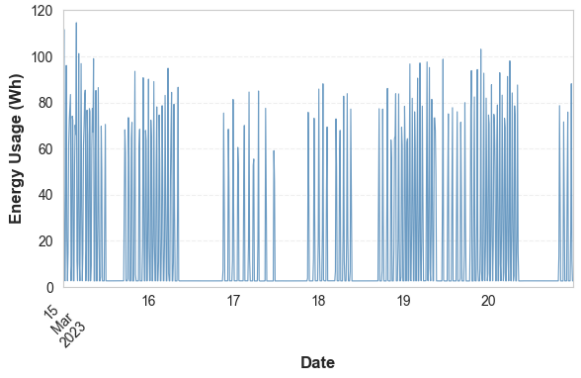


Figure 1: Exemplary HVAC energy usage of student dorm for one week

Office building

We use the Building 59 dataset (Luo et al., 2022) to evaluate forecasting performance in an office building. The records include energy consumption, HVAC operational data, IEQ, and outdoor climate data for a campus building over three years, from January 2018 to December 2020. The IEQ data are separated by thermal zones. For each zone, the zone temperature, cooling and heating temperature setpoints, CO_2 concentration, fan speed for the ventilation air supply, and heating valve position are provided. The building is divided into northern and southern parts, each served by two rooftop units (RTUs). For each RTU, operational data such as supply/return temperature, air flow rates, and setpoints are available. We use data from the southern part of the building, as it provides more detailed measurements. Although the dataset contains a short period of occupancy data, there are no time spans in which both occupancy and indoor-climate data are complete. The following prediction targets are chosen:

- **Zone temperature [°C]:** Zone temperature of an office-space thermal zone. Input features for the graph-based models include temperatures of adjacent zones, outdoor air temperature and humidity, RTU supply-air temperature, and zone fan speeds and heating valve positions. Since office temperatures depend strongly on occupancy, we use a binary indicator for typical working hours (9 am to 5 pm on weekdays) as a past and future covariate for Chronos, alongside outdoor air temperature. The sampling rate is $f_s = 5\,\mathrm{min}$.

- **HVAC energy consumption [kWh]:** HVAC energy consumption is the aggregate of the energy consumption of both RTUs serving the southern part of the building. The graph representation is constructed from each RTU's operational signals (supply/return/mixed-air temperature, supply flow rates), setpoints, fan speeds, heating valve position, temperatures of the thermal zones served by the RTU, and outdoor climate conditions. This yields a complex graph with 115 nodes and 154 edges. For covariate forecasting, we use the working-hours indicator and outdoor air temperature. The sampling rate is $f_s = 15\,\mathrm{min}$, i. e., the original sampling rate of HVAC energy consumption.

Results and discussion

Student dorm - relative humidity

For the task of predicting relative humidity, TSFM and ARIMA significantly outperform the GCN approaches for all horizons (Table 1). With an RMSE below 1 %, these models show superior performance for the shorter prediction horizons of 1 and 2 hours. When the prediction horizon is doubled to 4 and 8 hours, their performance decreases. ARIMA performs at a similar level for the 1-hour prediction but quickly falls behind Chronos and TimesFM as the horizon increases. This is likely due to ARIMA autoregressive stepwise prediction, where errors accumulate for longer horizons. TSFM on the contrary can replicate longer patterns due to the transformer architecture. While adding past-only covariates of outdoor temperature and dew point adds little accuracy, incorporating known future covariates improves performance as the context grows, with particularly strong gains for the 4- and 8-hour horizons. Causal covariates also improve performance but remain behind future covariates.
The GCN model fails to capture the dynamics across horizons and context lengths. Its best performance is achieved

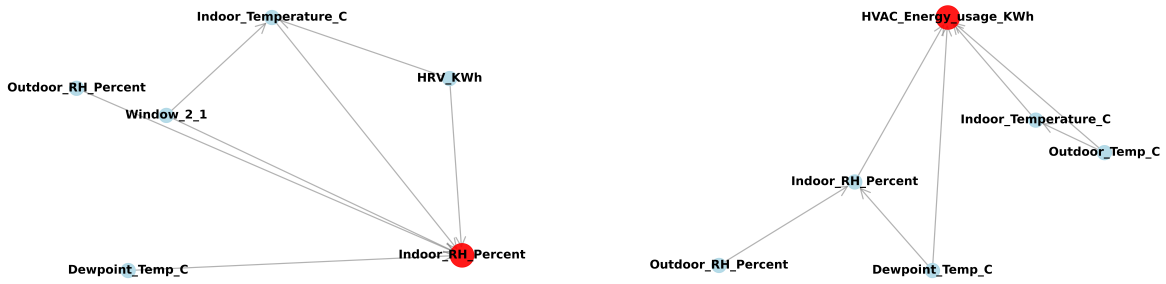


Figure 2: Graph underlying for predictions in the student dorm data set for relative humidity (left) and HVAC energy usage (right).

with the short context length of 4 hours ($C = 24$) for the longest horizon; however, on the 8-hour horizon, performance degrades across all context lengths. The GRUGCN performs best when input and output sequence lengths are equal, yet it still remains behind ARIMA and the TSFMs. These results suggest that TSFMs perform particularly well when predicting variables with a fixed range. Relative humidity, by definition, is bounded between 0 and 100 %; in the test set, values range from 20 % to 57 %, with a mean of 35.1 % and a median of 35.25 %. For machine learning models, the nonlinear relationships between input features can be difficult to learn from scratch, especially given the limited number of samples and meaningful features in the underlying graph.

*Table 1: Results for residential humidity forecasting (RMSE in %), best result for each horizon is marked **bold***

	Horizon P [h]			
Model	1	2	4	8
TimesFM (Univ.)	0.70	1.11	1.80	2.75
Chronos (Univ.)	0.68	1.03	1.65	2.86
Chronos Cov. (past)	**0.66**	1.00	1.60	2.79
Chronos Cov. (p+f)	0.67	**0.93**	**1.24**	**1.79**
Chronos Causal	0.67	0.98	1.52	2.67
ARIMA	0.84	1.59	2.16	3.61
GRUGCN (4 hours)	6.86	7.51	5.01	9.51
GRUGCN (8 hours)	5.24	6.20	6.98	6.25
GCN (4 hours)	9.49	10.81	13.30	8.82
GCN (8 hours)	13.23	8.16	10.55	9.20

Student dorm - HVAC energy usage

Predicting HVAC energy use in the student dorm is a challenging task that none of the models in this study can solve sufficiently well (Table 2). Among all models, Chronos with past covariates achieves the best results for the shortest prediction horizon.

When visualizing 1-hour-ahead predictions over two days in the test set (Figure 3), we find that Chronos can predict peaks in energy use with only a small lag, but it underestimates the magnitude of the energy consumed. The GRUGCN model captures only the baseline, also underestimates peaks, and adds noise around periods with an approximately constant baseline slightly below zero.

For the complex temporal patterns of HVAC energy use, the trend of decreasing accuracy for TSFM and ARIMA forecasts is confirmed, whereas GNN-based models perform similarly across all horizons. For the 4-hour horizon, ARIMA performs best, while the GRUGCN model performs best for the 8-hour horizon.

Because peaks in HVAC energy use are driven by unobserved occupant behavior patterns, all model types struggle to make accurate predictions. Although TSFMs outperform the graph-based models on shorter horizons, the non-seasonal and seemingly random patterns undermine their strengths.

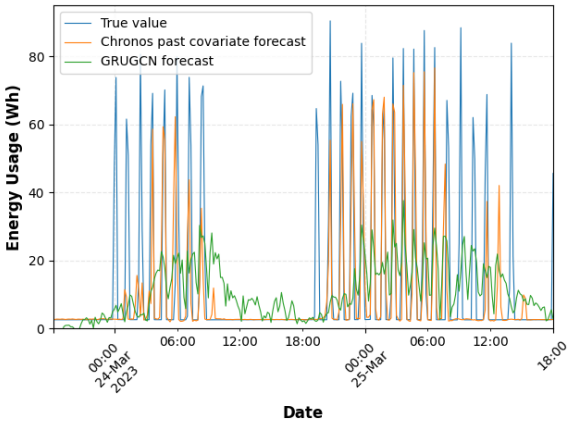


Figure 3: HVAC energy usage of two days in the test set. True value and predictions from GRUGCN with 4 hours context and Chronos past with past covariates for 1 hour horizon are shown.

Office building - zone temperature

Zone temperature in the office building (Table 3) is best predicted by Chronos with causal covariates across all horizons beyond 1 hour. Adding future covariates does not improve longer-horizon predictions at the same scale as in the humidity task: forecasts using only past covariates perform only slightly worse for the 4- and 8-hour horizons, and the RMSE increases with each horizon. The fact that the causal approach performs best for longer horizons

*Table 2: Results for residential HVAC energy forecasting (RMSE in Wh), best result for each horizon is marked **bold***

	Horizon P [h]			
Model	1	2	4	8
TimesFM (Univ.)	24.25	20.25	21.41	22.47
Chronos (Univ.)	15.74	17.09	18.36	19.48
Chronos Cov. (past)	**15.47**	16.92	18.15	19.30
Chronos Cov. (p+f)	15.48	**16.83**	18.12	19.17
Chronos Causal	15.52	16.90	18.15	19.28
ARIMA	16.70	17.30	**17.91**	18.98
GRUGCN (4 h)	17.02	17.93	18.19	18.87
GRUGCN (8 h)	17.53	17.87	18.37	**18.82**
GCN (4 h)	18.85	18.81	19.32	19.87
GCN (8 h)	18.78	18.96	19.41	19.65

(even outperforming predictions that use known future covariates) suggests that Chronos' group-attention mechanism benefits from well-defined causal covariates.

While TimesFM performs best for the 1-hour horizon, it does not appear to scale well for temperature forecasting: it is surpassed by ARIMA for the 2-hour horizon and by both GRUGCN variants for the two longest horizons. Since the sampling rate is doubled compared to the humidity task, TimesFM accuracy decreases substantially for $H > 48$.

As the horizon increases, the GRUGCN model gains ground relative to the TSFMs and ARIMA and surpasses ARIMA for the longest horizon. Compared to humidity forecasting, the gaps between TSFMs and graph-based models shrink. This is likely due to (i) the higher availability of meaningful features for graph learning in the office-building dataset and (ii) the reduced complexity of temperature prediction compared to relative humidity.

*Table 3: Results for office building zone temperature forecasting (RMSE in °C), best result for each horizon is marked **bold***

	Horizon P [h]			
Model	1	2	4	8
TimesFM (Univ.)	**0.160**	0.303	0.611	1.022
Chronos (Univ.)	0.175	0.268	0.365	0.427
Chronos Cov. (past)	0.184	0.302	0.410	0.425
Chronos Cov. (p+f)	0.175	0.284	0.375	0.396
Chronos Causal	0.163	**0.257**	**0.346**	**0.373**
ARIMA	0.183	0.286	0.437	0.696
GRUGCN (4 hours)	0.386	0.418	0.466	0.682
GRUGCN (8 hours)	0.374	0.458	0.496	0.665
GCN (4 hours)	0.600	0.576	0.665	0.766
GCN (8 hours)	0.652	0.734	0.836	0.893

Office building - HVAC energy usage

Among the results for HVAC energy-use forecasting in the office building (Table 4), TimesFM stands out by achieving the best predictions for horizons up to 4 hours. While the GNN-based models and ARIMA perform at a similar level, TimesFM outperforms them but still shows a decrease in accuracy for longer horizons: it falls behind all Chronos variants for the 8-hour horizon. Among the Chronos variants, the past-and-future-covariate setting performs best for the longest horizon. We also find that, in this task adding covariates yields smaller gains than in other cases as the target variable depends on a wide range of factors. In particular, the causal-graph variant, which incorporates a large number of covariates, serves as a scalability test for the in-context learning (ICL) mechanism. The strong drop in performance indicates that this approach may not suitable when many covariates are used. The availability of diverse sensor data highly benefits the graph-based approaches, as they perform similarly to the TSFMs and ARIMA up to the 4-hour horizon. Given the complex graph structure, adding a GRU layer to capture temporal patterns does not improve performance, as results are similar to those of the GCN model.

Although HVAC energy data for the office building are still noisy, the patterns are easier to predict because they are governed by observable behavior represented in the graph.

*Table 4: Results for HVAC energy usage forecasting (RMSE in kWh), best result for each horizon is marked **bold***

	Horizon P [h]			
Model	1	2	4	8
TimesFM (Univ.)	**0.71**	**0.88**	**1.15**	1.50
Chronos (Univ.)	0.94	1.08	1.32	1.44
Chronos Cov. (past)	0.95	1.07	1.30	1.41
Chronos Cov. (p+f)	0.95	1.07	1.29	**1.38**
Chronos Causal	3.58	4.31	4.65	5.62
ARIMA	0.90	1.07	1.32	1.54
GRUGCN (4 hours)	0.96	1.10	1.31	1.56
GRUGCN (8 hours)	0.95	1.14	1.35	1.58
GCN (4 hours)	0.94	1.10	1.31	1.56
GCN (8 hours)	0.96	1.14	1.34	1.58

Conclusion and outlook

In this paper, we compare the predictive performance of two state-of-the-art time-series foundation models with ARIMA and graph neural network (GNN)-based approaches. We also present a new approach for configuring covariates based on causal relationships encoded in graph representations. Using two real-world building case studies of residential and office use, we compare forecasting performance for HVAC energy use, thermal-zone temperature, and relative humidity.

The study finds that TSFMs show strong performance on both examined IEQ tasks, surpassing the GNN-based models. By comparing results across datasets with dense sensor coverage and lower data availability, we conclude that TSFMs are a strong alternative to training machine learning models from scratch, particularly in use cases where high-quality data are scarce. As the number of measurements incorporated for graph learning grows, the performance gap between pretrained TSFMs and the relatively simple GNNs used in our experiments shrinks. In cases with high sensor coverage, we show that Chronos' in-context learning capabilities can benefit from causal covariates and perform better than most Chronos runs with only historical data. However, the HVAC use case also

indicates that identifying a refined set of causal variables is important. This suggests that semantic graph models could help identify more effective covariate sets.

Regarding the two TSFMs we compare, Chronos performs best in three out of four experiments and maintains its accuracy better as the prediction horizon increases. Chronos' support for incorporating known future covariates into the prediction horizon shows potential to help maintain performance for longer horizons in the temperature-forecasting task, but does not consistently improve performance across all experiments. As the office-building HVAC results show, it cannot be assumed that covariate forecasting is superior to univariate forecasting in every case. In addition, although covariate support moves beyond purely historical-context forecasting, current prediction pipelines do not incorporate metadata, semantic context, or natural-language inputs, as proposed, for example, by Mulayim et al. (2025).

Overall, while causality-informed TSFMs show promise for building-related forecasting tasks, their performance varies across use cases, and further research is needed to identify the factors that govern their effectiveness. Beyond predictive accuracy, practical deployment also requires interpretability. Although TSFMs have proven valuable for prediction tasks with sparse data, other aspects remain understudied. Trust in predictions in real-world applications is often undermined by the black-box nature of machine learning models (Chen et al., 2023). While ML and GNN models can be augmented with explainability techniques such as GNNExplainer (Ying et al., 2019), the nature of pretraining and learning solely from time-series data and covariates limits TSFM explainability. For future research, we propose augmenting zero-shot TSFMs with semantic context and exploring synergies that combine the strengths of TSFMs with classical machine learning models and semantically explainable causal information.

References

Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In ACM SIGKDD.

Ali, D. M. T. E., Motuzienė, V., and Džiugaitė-Tumėnienė, R. (2024). Ai-driven innovations in building energy management systems: A review of potential applications and energy savings. Energies, 17(17).

Ansari, A. F., Shchur, O., Küken, J., Auer, A., Han, B., Mercado, P., Rangapuram, S. S., Shen, H., Stella, L., Zhang, X., et al. (2025). Chronos-2: From univariate to universal forecasting. arXiv:2510.15821.

Ansari, A. F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S. S., Pineda Arango, S., Kapoor, S., Zschiegner, J., Maddix, D. C., Mahoney, M. W., Torkkola, K., Gordon Wilson, A., Bohlke-Schneider, M., and Wang, Y. (2024). Chronos: Learning the language of time series. Transactions on Machine Learning Research.

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., et al. (2016). Brick: Towards a unified metadata schema for buildings. In Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments.

Borowski, M. and Zwolińska, K. (2020). Prediction of cooling energy consumption in hotel building using machine learning techniques. Energies, 13(23).

Chen, X., Lv, G., Zhuang, X., Roa, C. D., Schiavon, S., and Geyer, P. (2025). Integrating symbolic neural networks with building physics: A study and proposal. J. of Building Engineering, 111:113033.

Chen, Z., Xiao, F., Guo, F., and Yan, J. (2023). Interpretable machine learning for building energy management: A state-of-the-art review. Advances in Applied Energy, 9.

Cho, J. H. and Moon, J. W. (2022). Integrated artificial neural network prediction model of indoor environmental quality in a school building. J. of Cleaner Production, 344.

Das, A., Kong, W., Sen, R., and Zhou, Y. (2024). A decoder-only foundation model for time-series forecasting. In Forty-first International Conference on Machine Learning.

Elmaz, F., Eyckerman, R., Casteels, W., Latré, S., and Hellinckx, P. (2021). Cnn-lstm architecture for predictive indoor temperature modeling. Building and Environment, 206.

Goswami, M., Szafer, K., Choudhry, A., Cai, Y., Li, S., and Dubrawski, A. (2024). Moment: A family of open time-series foundation models. arXiv:2402.03885.

Grauer, D. and Ploennigs, J. (2025). Does chatgpt know building physics? exploiting foundation models for building performance prediction with gnns. In EC3 Conf.

Gupta, R., Raina, N., Chen, B., Chen, L., Danilov, C., Eckhardt, J., Bernard, K., and Nahrstedt, K. (2025). No one-model-fits-all: Uncovering spatio-temporal forecasting trade-offs with graph neural networks and foundation models. In buildSys.

Irshad, K., Khan, A. I., Zayed, M. E., Algarni, S., and Alqahtani, T. (2024). Performance analysis of air conditioning system integrated with thermal energy storage using enhanced machine learning modelling coupled with fire hawk optimizer. J. of Building Engineering, 98.

Jia, X., Song, H., Nan, X., and Cai, X. (2024a). Prediction of short-term heat load of office buildings based on gnn-lstm modeling. In 2024 6th International Conference

on Frontier Technologies of Information and Computer (ICFTIC). IEEE.

Jia, Y., Wang, J., Reza Hosseini, M., Shou, W., Wu, P., and Mao, C. (2024b). Temporal graph attention network for building thermal load prediction. Energy and Buildings, 321.

Langbridge, A., O'Donncha, F., Ba, A., Lorenzi, F., Lohse, C., and Ploennigs, J. (2023). Causal temporal graph convolutional neural networks (ctgcn). arXiv:2303.09634.

Latoń, D., Grela, J., Ożadowicz, A., and Wisniewski, L. (2025). Artificial intelligence and machine learning approaches for indoor air quality prediction: a comprehensive review of methods and applications. Energies, 18(19).

Li, Y., Arellano-Espitia, F., Aler, R., Igualada, L., and Corchero, C. (2025). Data-driven methods and their applications to building hvac energy consumption prediction: A review. J. of Building Engineering.

Lin, X., Prabowo, A., Razzak, I., Xue, H., Amos, M., Behrens, S., and Salim, F. D. (2024). Exploring capabilities of time series foundation models in building analytics. arXiv:2411.08888.

Liu, H., Liu, Y., Guo, X., Wu, H., Wang, H., and Liu, Y. (2023). An energy consumption prediction method for hvac systems using energy storage based on time series shifting and deep learning. Energy and Buildings, 298.

Luo, N., Wang, Z., Blum, D., Weyandt, C., Bourassa, N., Piette, M. A., and Hong, T. (2022). A three-year dataset supporting research on building energy management and occupancy analytics. Scientific data, 9:156.

Mulayim, O. B., Quan, P., Han, L., Ouyang, X., Hong, D., Bergés, M., and Srivastava, M. (2025). Can time-series foundation models perform building energy management tasks? arXiv:2506.11250.

Muraina, I. (2022). Ideal dataset splitting ratios in machine learning algorithms: general concerns for data scientists and data analysts. In 7th international Mardin Artuklu scientific research conference.

Newsham, G. R. and Birt, B. J. (2010). Building-level occupancy data to improve arima-based electricity use forecasts. In buildSys, pages 13–18.

Ogundiran, J., Asadi, E., and Gameiro da Silva, M. (2024). A systematic review on the use of ai for energy efficiency and indoor environmental quality in buildings. Sustainability, 16(9).

Pandey, P. R., Liu, Y., Wilson, N., and Dong, B. (2025). Dataset on occupant behavior, indoor environment, and energy use before and after dormitory retrofit. Scientific Data, 12(1).

Semasinghe, K., Perera, S., Jin, X., Samaratunga, M., and Nanayakkara, S. (2025). Data-driven approaches to improve indoor environmental quality: A systematic literature review. In Digital Frontiers in Buildings and Infrastructure.

Shi, Z., Zheng, R., Shen, R., Yang, D., Wang, G., Liu, Y., Li, Y., and Zhao, J. (2024). Building heating load forecasting based on the theory of transient heat transfer and deep learning. Energy and Buildings, 313.

Smith, T. G. et al. (2017–). pmdarima: Arima estimators for Python. [Online; accessed <today>].

Tagliabue, L. C., Cecconi, F. R., Rinaldi, S., and Ciribini, A. L. C. (2021). Data driven indoor air quality prediction in educational facilities based on iot network. Energy and Buildings, 236.

Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., and Sahoo, D. (2024). Unified training of universal time series forecasting transformers.

Ying, Z., Bourgeois, D., You, J., Zitnik, M., and Leskovec, J. (2019). Gnnexplainer: Generating explanations for graph neural networks. NeurIPR, 32.

Zhang, J., Xiao, F., Li, A., Ma, T., Xu, K., Zhang, H., Yan, R., Fang, X., Li, Y., and Wang, D. (2023). Graph neural network-based spatio-temporal indoor environment prediction and optimal control for central air-conditioning systems. Building and Environment, 242.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

ADAPTIVE LIFE CYCLE ASSESSMENT AS AN EVOLVING ENVIRONMENTAL KNOWLEDGE STATE FOR DIGITAL BUILDING LOGBOOKS

Kilian Speiser and Jochen Teizer
Technical University of Denmark, Denmark

Abstract

This paper introduces Adaptive Life Cycle Assessment (ALCA) as a time-resolved framework for continuously evaluating building environmental impacts over the lifecycle. Using a conceptual design science approach, the study develops a multidimensional framework integrating prospective and retrospective data and situates it within the Digital Building Logbook (DBL). An illustrative additive manufacturing case demonstrates data flows and feasibility. The results show that ALCA enables continuous updating of accumulated impacts, improves traceability, and supports PDCA-based environmental control. The framework addresses limitations of static and dynamic LCA and provides a governance-oriented approach for lifecycle impact monitoring and regulatory alignment.

Introduction

The building industry is a major contributor to environmental impacts and therefore plays a key role in the transition required to maintain climate stability and planetary habitability (Global Alliance for Buildings and Construction, 2022). Life Cycle Assessments (LCAs) quantify the environmental impacts of buildings and support informed decision-making. This approach helps avoid the unintentional transfer of burdens to other environmental impact categories (Hauschild et al., 2017). In addition, quantification supports the definition of environmental targets, which can guide actions to improve the environmental performance of buildings (International Energy Agency, 2023). Such targets have also translated into quantified thresholds on national levels, where construction projects must hit specific targets, like in Denmark, where only 6.8-8.0 kg CO2 eq are permitted per used m2 and year of a building (Social- og Boligstyrelsen, 2018).

LCA is primarily performed prospectively (ex-ante) but can also be performed retrospectively (ex-post), also referred to as historical LCA (Bruhn et al., 2024; Kristensen et al., 2015). In the remainder of this work, we adopt the terminology for prospective and retrospective LCA proposed by Arvidsson et al. (2024). Aside from the timing of the assessment relative to the product life cycle, time is typically not explicitly resolved in either prospective or retrospective LCAs. Both approaches are generally time-integrated to a reference point (Arvidsson et al., 2024).

LCA is currently performed in the early stages or upon completion and is always time-integrated to a certain point in time (Forth, 2025). A building has a long lifetime with changes through renovations and shifting boundary conditions, meaning that assessments produced early in the project fail to quantify the building's impact once those changes occur. This time-dependency motivates investigation of different LCA approaches.

Dynamic LCA (DLCA) is an approach to LCA that recognizes the changing nature of parameters and processes within assessed systems and incorporates their temporal evolution into the LCA framework. It is therefore time-resolved within the modeling framework, where certain parts can be retrospective and others can consider future events, products, and processes (Arvidsson et al., 2024). In DLCA, temporal evolutions are defined as variations in parameters. Although DLCA explicitly represents time, temporal trajectories are predefined, meaning that the final results are static after LCA completion. Changes in parameters or processes during the product's lifetime are not considered once the LCA is completed. Figure 1 illustrates this limitation: even when time is represented, module contributions remain static once the assessment is produced. In contrast, continuous updating revises module estimates when new evidence becomes available.

Figure 1 suggests that DLCA offers greater opportunities to accommodate time-dependent aspects relevant to the long lifetimes of buildings, but it fails to respond to actual changes over time. The actual impacts after construction, during operation, and renovation are not considered and are based on early predictions, which often significantly deviate from measured energy consumption during operations (De Wilde, 2014). Similar issues occur during the construction phase: as-built LCAs rely on information collected from drawings or Building Information Models (BIM), but significant deviations in the materials used go unnoticed and are hidden from the LCA. This leads to incorrect expectations about environmental impacts.

Ongoing regulatory developments in Europe may provide a framework to address this issue. The Digital Building Logbook (DBL) is defined as a tool to record and assess data, information, and documents that include "maintenance, refurbishment, [...] traceability and characteristics of construction materials, performance

data such as operational energy use [...] and lifecycle emissions" over all phases of a building (Volt et al., 2020). Once such logbooks are available, reported renovation and measured operational energy performance can be used to update prospective LCAs with observed data. In addition, the mandated Digital Product Passport (DPP) can further improve accuracy by enabling instance-specific tracking of building components, so that emissions are linked to a specific installed element rather than to generic impact declarations (European Commission, 2024).

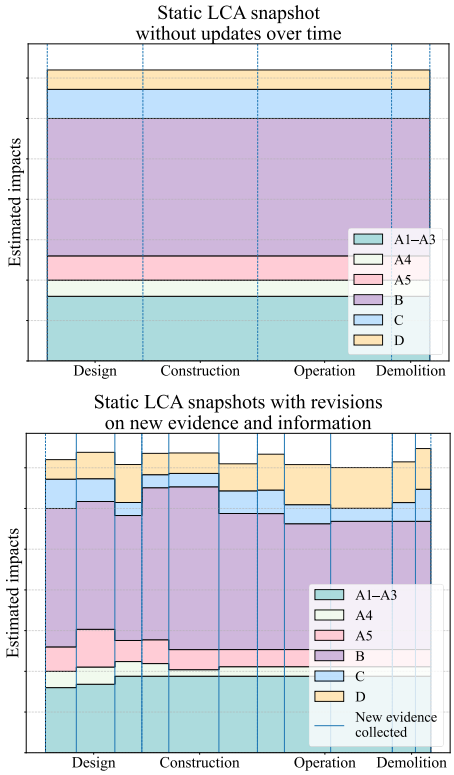


Figure 1: Conceptual comparison of static estimates of LCA and continuous updating based on new evidence. The figure does not show accumulated impacts but the estimates at a given time.

For regulatory environmental impact targets, such as those in Denmark, prospective LCAs are insufficient because they fail to accurately reflect the environmental impacts of buildings as they are actually constructed and operated. The evidence for this gap is well established: as-built deviations systematically alter embodied impacts (Forth, 2025), and the performance gap between predicted and measured operational energy use is widely documented (De Wilde, 2014; IEA, 2023). There is, however, a gap in the conceptual description of how a retrospective LCA can be incrementally developed for buildings over their full lifetime. This would enable more accurate estimates and require stakeholders to act when targets are not met. With the development of DBLs, such updating of impacts could become mandated, but a technical and conceptual description of the information flows and governance structure remains absent.

This work presents the concept of Adaptive LCA (ALCA). It is a time-resolved framework in which LCAs are continuously updated throughout the building's lifetime. Information is updated based on new inputs from either planned changes or retrospective measurements. Previous versions are timestamped and stored, and can become part of the DBL. The paper presents the information dimensions and definitions that align with ALCA and situates it within the information governance context of the Digital Twin Construction (DTC) (Sacks et al., 2020). Additive manufacturing (AM) in construction provides an illustrative context for framework development, as it enables higher traceability of information through sensors and continuous measurements of material use and energy consumption. ALCA is intended as a general framework for buildings; AM is used here as an enabling context that makes information flows more tractable, not as a limiting scope condition.

The paper addresses two research questions:

RQ1: How can a building LCA be conceptualized as a continuously evolving environmental knowledge state that integrates accumulated ex-post impacts and projected ex-ante impacts over the lifecycle?

RQ2: How can ALCA be integrated into a DBL to support traceability, regulatory use, and decision-making?

The paper makes four contributions. First, it introduces ALCA as an evolving environmental knowledge state that integrates accumulated and projected impacts over the building lifecycle. Second, it proposes a multidimensional conceptual framework that formalizes the temporal and epistemic information structure of products and processes within ALCA. Third, it clarifies how ALCA can be integrated into a DBL ecosystem while preserving institutional roles and governance boundaries. Fourth, it uses AM as an illustrative context to show how high-resolution process data, together with DPPs, can reduce uncertainty and improve the instance-specific accuracy of environmental impact declarations.

Research methodology

This paper adopts a conceptual design science research approach (Hevner et al., 2004) to develop the ALCA framework. The framework is derived through three steps. First, limitations in existing LCA approaches, specifically the inability to incorporate observed changes during the building's lifetime, are identified through analysis of established LCA literature and standards (ISO 14040; ISO 14044). Second, the ALCA concept is constructed by synthesizing LCA theory with emerging digital building governance frameworks, including the DBL definition (Volt et al., 2020) and DTC concepts (Sacks et al., 2020). Key framework dimensions (temporal, epistemic, and product-process) are grounded in established LCA terminology and standard practice. Third, the framework is instantiated illustratively in an AM construction context to demonstrate feasibility and to identify key information flows. Evaluation at this stage is argumentative and conceptual: the framework is assessed for internal consistency and alignment with established standards rather than through empirical testing. Empirical validation

in pilot settings is identified as the primary direction for future research.

Conceptual framework for ALCA

ALCA builds on the recognition that retrospective environmental information is increasingly available throughout the building lifecycle. Measurements derived from sensors, construction documentation, operational monitoring systems, and project reports during construction, operation, renovation, and demolition enable systematic observation of realized environmental impacts (Volt et al., 2020; IEA, 2023). Advances in digitization, Internet of Things (IoT) integration, expanded Environmental Product Declaration (EPD) databases, and emerging forms of robotized and AM processes have significantly increased the availability, resolution, and traceability of such data (European Commission, 2024; Schlenger et al., 2025). This development enables the establishment of a closed-loop environmental control logic.

ALCA conceptualizes a building LCA as an incrementally evolving environmental knowledge state. It extends DLCA by not only resolving time within the modelling framework but by updating the assessment incrementally when new information becomes available. ALCA structures the assessment into discrete LCA states. Each state integrates accumulated retrospective (ex-post) impacts and projected prospective (ex-ante) impacts over the remaining lifecycle. At any point in time, one state represents the best available environmental knowledge of the building. When new measured data or planned changes occur, ALCA generates a new timestamped state while preserving the previous version.

ALCA structures environmental knowledge along three explicit dimensions. The temporal dimension distinguishes between retrospective (ex-post) and prospective (ex-ante) impacts across lifecycle phases. The epistemic dimension differentiates primary (measured) data from secondary (estimated or modelled) data. In addition, ALCA separates product-related and process-related impacts to ensure traceability of materials, components, and activities. These dimensions define the information structure required for integration into a DBL, where each LCA state can be timestamped, stored, and audited.

Regarding lifecycle phase characteristics, ALCA's adaptive updating does not apply uniformly across all phases. Module A (product and construction stage) covers activities from raw material extraction through construction completion. Once construction is finished, modules A1-A5 become fully retrospective: no further updating of construction-stage impacts is possible, though their accuracy can be refined with as-built data. By contrast, modules B (use stage) and C (end-of-life) remain partially or fully prospective throughout most of the building's life. The most significant adaptive value of ALCA therefore lies in operational monitoring (module B), renovation cycles (module B4), and end-of-life planning (module C), where continuous updating against realized performance can detect deviations from targets and trigger corrective action. This distinction is important for scoping the practical application and governance responsibilities of ALCA.

LCA methodology steps

According to ISO 14040 and ISO 14044, LCA consists of four phases: goal and scope definition, life cycle inventory (LCI), life cycle impact assessment (LCIA), and interpretation. Information emerges and evolves across all four phases. The goal and scope define the functional unit, system boundaries, assumptions, and impact categories. The LCI quantifies input and output flows across all lifecycle phases of the built asset. In construction projects, these flows correspond to building products and processes. The LCIA translates quantified flows into environmental impacts using selected impact assessment methods and indicators. The interpretation phase evaluates results in relation to regulatory thresholds, project location, and environmental targets.

Although information occurs throughout all phases, the inventory plays a central role. The LCI contains the quantified units and parameters that determine environmental impacts. Changes in quantities, material specifications, process data, or service life assumptions directly affect impact results. Because ALCA incrementally updates environmental knowledge, revisions typically begin at the inventory level and propagate through LCIA and interpretation. The inventory therefore forms the structural backbone of the evolving LCA state.

EPDs provide product-specific LCIA results together with metadata such as reference service life and transport assumptions. However, EPDs cover products only. A building LCA must additionally quantify and assess construction and operational processes, including products that are partially or fully consumed.

Within this methodological structure, data correspond to quantified flows and parameters in the LCI. Information emerges when these flows are characterized through LCIA methods. Knowledge results from interpreting impacts in relation to regulatory requirements and environmental targets. In ALCA, new evidence may require updating inventory data, recalculating impacts, and revising the interpreted knowledge. Each incremental update produces a new LCA state that reflects the best available environmental knowledge at a given time.

Temporal dimension

Prospective and retrospective information both address LCI and LCIA throughout all phases of the building. Prospective information is based on assumptions (design and processes), scenarios (future site conditions), and declared environmental impact data from EPDs. EPD assumptions refer to the declared values and boundary conditions embedded within an EPD dataset, including assumed reference service life, transport distances, and upstream production conditions. As the building progresses through its lifecycle stages, parts of the LCI or LCIA can become historical, meaning that prospective information can be progressively replaced or refined with

retrospective information, such as measured, reported, or verified data, thereby reducing uncertainty.

Retrospective information can address more specific knowledge about units or quantities in the LCI, or more accurate information on impacts in the LCIA (instance-specific EPDs, measured emissions). Additionally, the lifetime of certain elements can be revised based on retrospective evidence.

When updating the LCA using new retrospective information, previous versions are preserved as earlier states. These past projections remain valid representations of former knowledge, enabling learning and retrospective analysis.

Table 1 presents four knowledge states derived from the intersection of two binary dimensions: the temporal dimension (prospective vs. retrospective) and the epistemic dimension (primary vs. secondary). These four states represent analytically distinct combinations relevant to building LCA practice, each characterized by a different uncertainty profile. They are not mutually exclusive within a single ALCA state: a given module may simultaneously draw on both primary and secondary data or combine retrospective and prospective contributions. The four-state structure reflects a minimum necessary set of categories to capture the full range of data quality configurations encountered across the building lifecycle.

Table 1: Temporal-epistemic knowledge states in ALCA

	Secondary (estimated/ modeled)	Primary (measured/ verified)
Prospective	Projected impacts from assumptions • Design-based quantities • Type-level EPDs • Modeled or simulated future scenarios	Empirically grounded projections • Measured baselines • Supplier-specific data • Measured process rates
Retrospective	Reconstructed past impacts • As-built model • Back-calculated activities (fuel/time/waste) • Gap-filled inventories w/o evidence	Observed and measured impacts • Verified quantities (delivery notes, monitored) • Metered process data (energy, telemetry) • Time-stamped records

Epistemic dimension

Information can be measured or estimated. In the context of LCA, these are described as primary (measured) and secondary (estimated). In ALCA, there is no fixed correlation between these and prospective or retrospective information. Instead, primary information reduces uncertainty compared to secondary information. Estimated information refers to assumptions, calculations, simulations, or predictions. Measured information includes raw sensor readings, machine telemetry, verified quantities, and reported data.

Concrete examples of sensor-derived data relevant to ALCA include: electricity consumption of AM robots or construction equipment, directly informing process energy in the LCI; material flow data from weigh cells or volume sensors on concrete pumps, informing material quantities; air quality or particulate sensors on construction sites, informing process emissions; and Building Management System (BMS) data on heating, cooling, and ventilation during operation, replacing assumed operational energy with metered values. Manual data extraction from delivery notes, invoices, and project reports complements sensor data where automated collection is not possible.

The retrospective dimension will still include a significant amount of estimation. For instance, precise on-site fuel consumption tracking does not necessarily linearly translate into specific environmental impact indicators such as NOx. Accurate quantity information in the LCI does not therefore automatically produce accurate impact results; estimation through characterization factors remains necessary. The level of uncertainty, however, should decrease when moving from prospective to retrospective. Yet, uncertainty cannot be eliminated.

Primary and secondary data directly affect the LCI, while LCIA remains largely model-based even when inventory data are measured. LCA interpretation integrates both prospective and retrospective results for decision-making.

Process-product dimension

In addition to temporal and epistemic structuring, ALCA distinguishes between product-related and process-related impacts within the LCI. In standard building LCA practice, although all environmental impacts ultimately originate from physical and chemical processes, they are attributed either to physical building elements (products, with their embodied impacts) or to lifecycle activities such as construction, operation, and demolition. This distinction follows standard LCA practice and enables structured accounting within the inventory.

Product-related impacts refer to impacts embodied in materials, components, and assemblies. These are typically represented through product datasets such as EPDs and quantified as element quantities linked to upstream production processes. Process-related impacts arise from activities such as construction, operation, maintenance, and demolition, and are quantified through activity data, including energy use, transport, equipment operation, and waste handling.

From an ontological perspective, product datasets summarize process-based information into standardized units. While this aggregation enables comparability and practical application, it may obscure instance-specific variations, such as deviations in transport distances, execution conditions, or site-specific energy mixes. Direct process information can therefore reduce uncertainty when available, particularly in adaptive contexts where execution data are progressively observed.

Product impacts are generally object-based and may stabilize once elements are installed, whereas process impacts are activity-based, time-dependent, and execution-sensitive. Within ALCA, both categories coexist in each environmental knowledge state, and updates may affect product quantities, process activities, or both. The distinction serves as a structuring principle within the inventory, enabling clearer allocation, updating, and traceability of impacts across phases.

Conceptual ALCA state

At any given point in the lifetime of a construction project, ALCA is represented by a discrete, timestamped environmental knowledge state. This state reflects the most accurate available environmental knowledge of the building at that moment. It integrates accumulated retrospective impacts and projected prospective impacts across relevant lifecycle modules, preserving traceability of the data and assumptions on which they are based.

Figure 2 represents this multidimensional structure. The three dimensions define the axes of the ALCA knowledge space. The stacked cube representation is scientifically motivated: it visualizes that each lifecycle phase module simultaneously occupies a position along all three dimensions. A given module may involve both prospective and retrospective contributions, draw on both primary and secondary data, and combine product and process impacts. The stacking therefore reflects that no single cell exclusively characterizes a given module; an ALCA state is a composite of multiple co-existing knowledge configurations. When new evidence becomes available, a subsequent ALCA state is generated, preserving prior states for comparison, learning, and regulatory auditability.

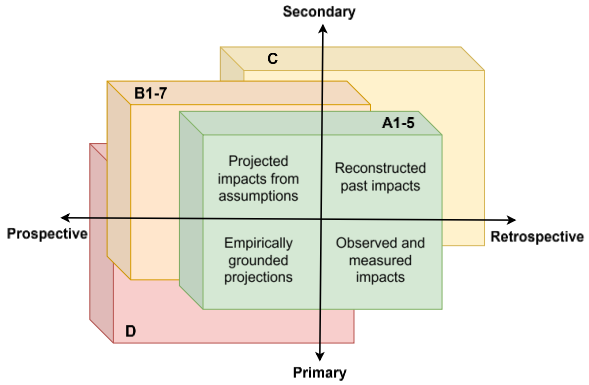


Figure 2: A conceptual ALCA state combining temporal, epistemic dimensions for all building phases.

Continuous PDCA-based control logic

ALCA operationalizes a Plan-Do-Check-Act (PDCA) control logic for environmental performance throughout the building lifecycle. In the planning phase (Plan), prospective LCA defines expected impacts based on design assumptions and scenarios. During execution (Do), construction and operational activities generate retrospective information in the form of measured or reported data, which incrementally replaces earlier assumptions within the ALCA state. Each update produces a revised, timestamped environmental knowledge state, which can be checked.

The checking phase (Check) evaluates the updated ALCA state against environmental targets or regulatory thresholds. If deviations occur, the acting phase (Act) involves mitigation measures such as design adjustments, modified construction methods, or operational interventions. In this way, ALCA transforms LCA from a static reporting instrument into a control mechanism, linking prospective planning with retrospective evidence through structured state transitions.

Regarding operational implementation, several clarifications are relevant. Updates to the ALCA state are triggered by three types of events: phase completion (e.g., end of the fabrication stage), delivery of new monitoring data (e.g., annual metered energy from the BMS), and significant deviations from regulatory thresholds, with specific deviation magnitudes to be defined by regulatory frameworks. Primary responsibility for maintaining the ALCA state lies with the building owner or operator, analogous to existing building maintenance and reporting obligations. Construction teams supply phase-specific process data during active construction stages. Automated data flows from IoT sensors can reduce manual effort significantly, particularly for operational energy monitoring, though manual reporting from delivery notes or invoices remains relevant where automation is unavailable. From a computational standpoint, ALCA does not require substantially new infrastructure: standard LCA calculation engines apply at each state, and the primary technical challenge is data integration and interoperability rather than computational intensity. Building owners and regulatory compliance bodies stand to benefit most from ALCA, as it enables accountable and verifiable impact tracking aligned with regulatory obligations. Design and construction teams also benefit from early deviation detection, which enables timely corrective action before lifecycle thresholds are exceeded.

Additive manufacturing as an enabler for continuous LCA

AM provides a particularly suitable context for illustrating ALCA due to the digital controllability and data richness of AM processes, which directly improve the epistemic quality of LCA inputs. AM processes are typically computer-controlled with integrated logging and are electricity-driven, resulting in accessible and high-resolution energy records. Sensors can additionally track material streams. As a result, material quantities and energy consumption can be recorded at high temporal resolution at the machine level.

Table 2 summarizes generic data characteristics of AM compared to conventional construction. While AM does not eliminate uncertainty it significantly improves the reliability of accumulated impacts during early lifecycle phases. This improvement stabilizes projections of remaining lifecycle impacts and strengthens the basis for continuous evaluation against regulatory thresholds.

Table 2: Generic comparison of data characteristics relevant to Adaptive LCA.

Aspect	Conventional Construction	Additive Manufacturing
Process control	Distributed, manual	Centralized, digital
Quantity tracking	Indirect	Direct or machine-derived
Energy monitoring	Limited	High-resolution, electricity-based
Temporal resolution	Low-medium	High

Impact on LCA uncertainty	High	Reduced (not eliminated)

During the design phase, all information is secondary and prospective. Designers derive quantities and impacts from models and assumptions, covering all lifecycle phases. Some primary historical data from comparable projects may inform baselines, but all future impacts are estimated. Once fabrication starts, modules A1-A3 are progressively updated from estimated to tracked and measured information (see Figure 3). Project teams can integrate instance-specific EPDs for products and measured data for processes. In AM, machine-logged electricity consumption and material flow sensor data can replace estimates directly.

During transport and construction (modules A4-A5), IoT-based measurements replace predicted construction process impacts. Updated as-built BIM models and tracked material quantities progressively substitute simulated impacts. Exact delivery routes or reported fuel use from transport vehicles improve transport accuracy. In AM, robot sensor data provide comparable information for material use and process energy.

During the operational phase (module B), two primary information types become relevant. First, metered energy from BMS progressively replaces estimated consumption for heating, cooling, and ventilation. Continuous prospective projections remain necessary to detect deviations before targets are exceeded. Second, renovation activities must be incorporated with their specific materials, processes, and impacts.

The demolition phase (module C) determines the realized building lifetime, the key parameter assumed at the outset. Only at demolition can final accumulated impacts be determined with certainty. Although corrective action is no longer possible at that stage, stakeholders can transfer results to subsequent projects or apply financial adjustments if targets were exceeded.

DBL integration

In this work, the DBL provides the institutional and technical context for operationalizing ALCA. It functions as an authoritative repository for realized building information, including as-built products and as-performed construction processes, rather than as an analytical or predictive system. This aligns the DBL with the ex-post dimensions of ALCA: measured or reported information provides an auditable baseline of accumulated impacts, while projected future impacts are generated by external analytical services, such as digital twin tools, using DBL data as their factual basis.

The role of DPPs within this architecture is significant. DPPs, mandated under the EU Ecodesign for Sustainable Products Regulation (European Commission, 2024), enable instance-specific documentation of individual building products and components, including material composition, declared environmental impact data, and service life information. Within ALCA, DPPs serve as the link between generic EPD datasets and instance-specific LCA inputs. Where a standard building LCA uses a type-level EPD covering a class of products, DPPs enable attribution of verified environmental data to a specific installed component, e.g., a particular concrete batch used on a specific date, with documented transport distance and energy consumption. This instance-specific traceability directly improves the accuracy of product-related impacts .

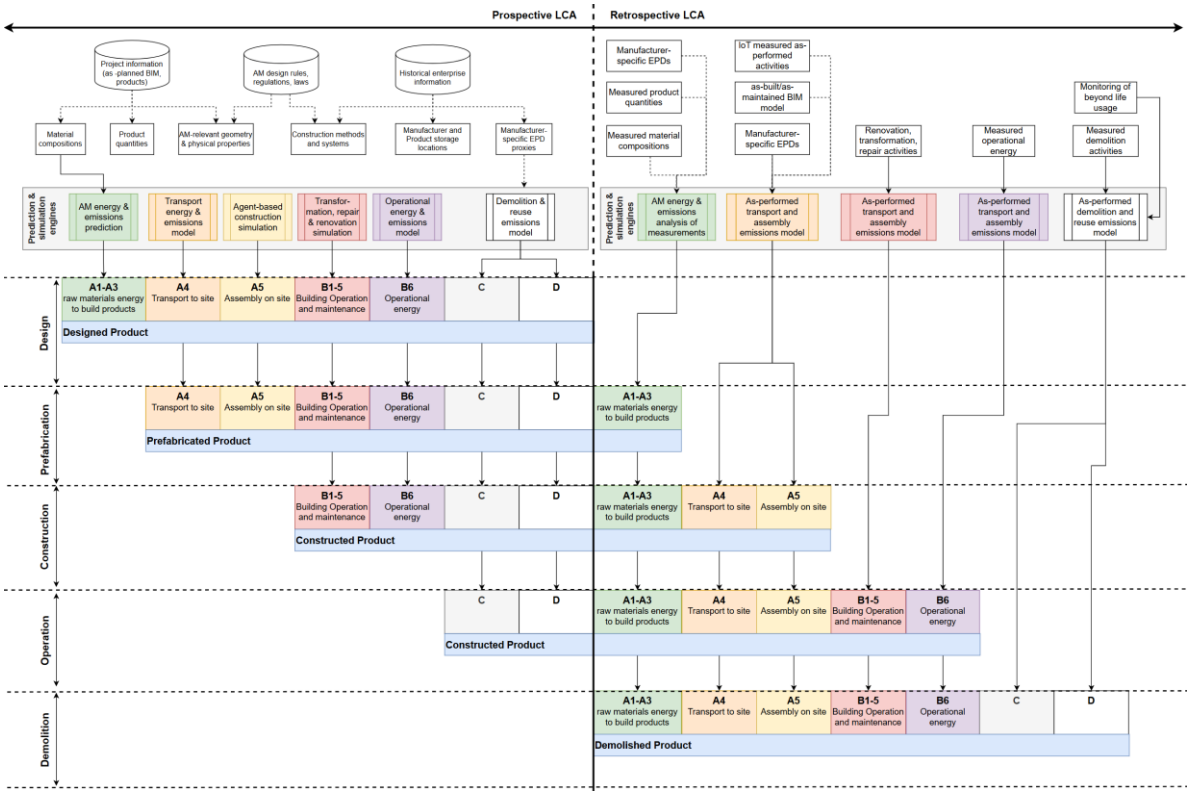


Figure 3: Instantiation of ALCA in an AM construction site.

in the LCI and reduces the uncertainty associated with generic EPD assumptions. As DPP adoption expands under EU regulation, the epistemic quality of ALCA inputs is expected to improve correspondingly, progressively replacing secondary estimates with primary, product-specific data. This connection represents the core mechanism through which the fourth contribution of this paper is realized in practice. The evolution of EPD content within DPP frameworks, particularly the shift from generic type-level declarations to instance-specific performance records, is therefore a key development that ALCA is designed to leverage.

Operationalizing ALCA through the DBL requires interoperable data structures and robust versioning. The DBL should act as an integration layer connecting DPPs, material passports, permit documentation, and LCA state records, rather than duplicating all data. Each update must include timestamps, source links, assumptions, and provenance metadata so that accumulated impacts remain auditable and suitable for regulatory compliance.

This separation clarifies the distinction between the DBL and a DTC. Although the DBL can serve as an information container within a DTC ecosystem, it does not perform reasoning, prediction, or optimization. The DBL stores and exposes information, while external tools provide analytical capabilities. Maintaining this separation preserves institutional clarity and avoids mixing authoritative records with interpretive or predictive models.

Discussion

DBLs and related passport systems should be understood as governance instruments, not merely technical data containers. They formalize documentation obligations, responsibilities, and verification mechanisms across the building lifecycle, giving ALCA the institutional continuity required for regulatory use.

An explicit comparison with DLCA clarifies ALCA's contribution. DLCA resolves time within the LCA model by defining parameter variations as predefined functions, enabling time-specific characterization of emissions and impacts (Arvidsson et al., 2024). However, once the DLCA model is constructed and executed, its results are fixed: the model does not update when actual conditions deviate from modeled trajectories. ALCA differs fundamentally in this respect. Rather than predefined trajectories, ALCA responds to actually observed data, generating new timestamped states as evidence accumulates. ALCA therefore does not replace DLCA; instead, it introduces an adaptive dimension that DLCA lacks. In practice, DLCA methods could be deployed within each prospective ALCA state, while the adaptive state-transition logic of ALCA incorporates retrospective evidence as it emerges. The two approaches are therefore complementary rather than competing.

Regarding standardization, LCA is an ISO-standardized methodology (ISO 14040; ISO 14044), and ALCA does not propose an alternative impact assessment method. Impact categories, characterization factors, and system boundary conventions remain unchanged. What ALCA introduces is a governance framework for managing the updating process: defining when a new state is generated, how prior states are preserved, and how accumulated and projected impacts are combined at any point in time. Developing normative standards for these aspects, including trigger thresholds, state management protocols, and interoperability requirements between DBL systems and LCA calculation engines, represents a necessary step toward regulatory adoption. Engagement with standards bodies such as ISO TC 207 and CEN TC 350, alongside regulatory alignment at the national and EU level, is an important direction for future work.

The conceptual relationship between ALCA and DTC reveals strong theoretical overlap. Both rely on structured information dimensions, iterative updating, and PDCA-based control logic. However, their scopes differ: DTC focuses primarily on construction process monitoring and optimization, whereas ALCA addresses lifecycle environmental impact accounting across all phases. ALCA can therefore be understood as a specialized environmental subsystem within a broader DTC ecosystem, accessing shared semantic information while specifying environmental data requirements that other subsystems must provide. Ontology-based semantic frameworks developed in DTC research (Schlenger et al., 2025; Speiser et al., 2025) offer a promising direction for future integration, particularly to enable consistent cross-system interpretation of environmental data. Extending such frameworks toward environmental impact accounting is acknowledged as future research rather than a current contribution of this paper, as it presupposes empirical validation of the ALCA framework itself. Generative and agentic AI can support experts in including relevant knowledge into such systems (Speiser et al., 2026).

The separation between DBLs as authoritative repositories and digital twin services as analytical engines has important governance implications. Verified ex-post information becomes part of the regulatory record within the DBL. Predictive simulations and scenario analyses remain external and do not automatically acquire authoritative status. This preserves institutional clarity and prevents conflation of observed performance with modeled projections. Regulators can assess compliance based on accumulated and verifiable impacts, while project teams use predictive tools to anticipate deviations and implement mitigation measures within a PDCA logic.

The framework remains conceptual and has undergone argumentative rather than empirical validation. While AM illustrates the feasibility of high-resolution monitoring, application in conventional construction, operational phases, renovation cycles, and end-of-life processes requires further investigation. Long building lifetimes limit full lifecycle validation, but prefabrication facilities and AM environments may serve as short-term testbeds. Standardization of semantic structures, interoperability requirements, and regulatory alignment remain open research challenges. Moreover, although ALCA progressively reduces epistemic uncertainty, it

cannot eliminate inherent predictive limits in future lifecycle modules.

Conclusions

This paper introduced ALCA as a time-resolved framework for treating building LCA as a continuously evolving environmental knowledge state. ALCA combines accumulated ex-post impacts with projected ex-ante impacts across lifecycle phases and structures this information through temporal, epistemic, and product-process dimensions. By generating discrete, timestamped states as verified data become available, ALCA reframes building LCA from a static reporting exercise into an ongoing process of environmental knowledge revision.

The paper further showed how ALCA can be anchored within a DBL ecosystem. The DBL provides an authoritative repository for verified ex-post information, while DPPs improve the instance-specific accuracy of product-level LCA inputs and external digital twin services support projection and decision-making. This separation preserves governance clarity while enabling continuous monitoring and PDCA-based intervention.

The framework remains conceptual and requires empirical validation in pilot settings. Future work should develop interoperable semantic structures, test ALCA across construction and operational contexts, and define normative requirements for adaptive LCA updating.

Declaration of AI in the writing process

During the preparation of this work, the authors used generative AI to proofread text and improve clarity. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgments

The research presented in this paper was funded by the European Union Horizon 2020 research and innovation program under agreement no. 101162318 (AM2PM).

References

Arvidsson, R., Svanström, M., Sandén, B.A., Thonemann, N., Steubing, B. and Cucurachi, S. (2024). Terminology for future-oriented life cycle assessment: review and recommendations, International Journal of Life Cycle Assessment, 10.1007/s11367-023-02265-8.

Bruhn, S., Gislason, S., Røgild, T., Andreasen, M., Ditlevsen, F., Larsen, J., Sønderholm, N., et al. (2024). Pioneering historical LCA - A perspective on the development of personal carbon footprint 1860–2020 in Denmark, Sustainable Production and Consumption, Vol. 46, 10.1016/j.spc.2024.03.014.

European Commission. (2024). Ecodesign for Sustainable Products Regulation, Official Journal of the European Union.

Forth, K. (2025). BIM-Based Semantic Enrichment for Environmental Analyses Using Large Language Models.

Global Alliance for Buildings and Construction. (2022). 2022 Globall Status Report for Buildings and Construction.

Hauschild, M.Z., Rosenbaum, R.K. and Olsen, S.I. (2017). Life Cycle Assessment: Theory and Practice, Life Cycle Assessment: Theory and Practice, 10.1007/978-3-319-56475-3.

International Energy Agency. (2023). Benchmarking and Target-Setting for the Life Cycle-Based Environmental Performance of Buildings.

Kristensen, T., Aaes, O. and Weisbjerg, M.R. (2015). Production and environmental impact of dairy cattle production in Denmark 1900-2010, Livestock Science, Vol. 178, 10.1016/j.livsci.2015.06.012.

Sacks, R., Brilakis, I., Pikas, E., Xie, H.S. and Girolami, M. (2020). Construction with digital twin information systems, Data-Centric Engineering, Vol. 1 No. 6, 10.1017/dce.2020.16.

Schlenger, J., Pluta, K., Mathew, A., Yeung, T., Sacks, R. and Borrmann, A. (2025). Reference architecture and ontology framework for digital twin construction, Automation in Construction, Vol. 174, p. 106111, 10.1016/j.autcon.2025.106111.

Social- og Boligstyrelsen. (2018). BR18, Bygningsreglementet.Dk.

Speiser, K., Seiß, S., Boukamp, F., Melzner, J. and Teizer, J. (2025). From fragmented data to unified construction safety knowledge: A process-based ontology framework for safer work, Automation in Construction, Vol. 176, p. 106293, 10.1016/j.autcon.2025.106293.

Speiser, K. and Teizer, J. (2023). An Ontology-Based Data Model to Create Virtual Training Environments for Construction Safety Using BIM and Digital Twins, 30th European Group for Intelligent Computing in Engineering Conference (EG-ICE), London, UK.

Speiser, K., Maciocci G., Boukamp F., Teizer J. (2026). "Agentic system for construction safety risk assessments using large language models and knowledge graphs", Advanced Engineering Informatics, Special issue: AI for the Built World, https://doi.org/10.1016/j.aei.2026.104681

Volt, J., Toth, Z., Glicker, J., De Grotte, M., Borragán, G., De Regel, S., Dourles-Quaranta, S., et al. (2020). Definition of the digital building logbook - Report 1 of the Study on the Development of a European Union Framework for Buildings' Digital Logbook, 10.2826/480977.

De Wilde, P. (2014). The gap between predicted and measured energy performance of buildings: A framework for investigation, Automation in Construction, Vol. 41, 10.1016/j.autcon.2014.02.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

NO CARBON LEFT BEHIND: TOWARDS TRACEABLE BIOGENIC CARBON IN CONSTRUCTION VALUE CHAINS

Edwin Zea Escamilla, Alexander N. Walzer, and Markus Knecht
University of Applied Sciences and Arts Northwestern Switzerland
FHNW, Muttenz and Windisch, Switzerland

Abstract

Biobased construction materials store atmospheric carbon, yet current life cycle assessment and environmental product declaration practices represent biogenic carbon as aggregated indicators with limited traceability across value chains. This limits transparency, attribution, and credibility of carbon-related claims. This paper explores how blockchain-enabled system principles can address these deficiencies by treating biogenic carbon as a persistent, traceable state linked to material flows. Through an exploratory mapping onto bamboo value chains, we identify how such ledger-based logic can formalize carbon continuity, prevent double counting, and support auditable life cycle transitions. The paper contributes a conceptual framework for carbon ledger approaches in construction.

Introduction

Biobased construction materials such as bamboo and wood are increasingly promoted as key contributors to climate mitigation strategies in the built environment (Rincón et al., 2023). Their relevance stems from their ability to sequester atmospheric carbon during growth and to temporarily store this carbon in buildings and infrastructure (Bundi et al., 2024). As embodied carbon gains prominence alongside operational emissions, the question of how biogenic carbon is accounted for has become central to sustainability assessments in construction.

Current life cycle assessment (LCA) frameworks provide standardized methods to quantify environmental impacts, including biogenic carbon flows (Hellweg & i Canals, 2014). However, in practice, biogenic carbon is typically represented as an aggregated balance at product, building, or system level (Tellnes et al., 2017). Information about origin, transformation, ownership, temporal storage, and end-of-life fate is progressively abstracted away as materials move along value chains (Sayın et al., 2025). This abstraction limits transparency, complicates verification, and creates uncertainty when biogenic carbon is used to support climate claims or compensation mechanisms (Dwyer & Mowry, 2022).

These limitations are particularly pronounced in global and heterogeneous value chains, such as those associated with bamboo-based construction materials. Bamboo is harvested, processed, transformed, and assembled across multiple actors and locations, often with limited digital continuity between stages. While the material itself is inherently traceable in physical terms, the associated carbon information is not.

In parallel, digital technologies in construction have primarily focused on design coordination, simulation, and operational performance, while the underlying logic of carbon accounting has remained largely unchanged, reflecting broader institutional misalignments between technological innovation and adoption in the construction sector (Walzer et al., 2025). However, blockchain and distributed ledger technologies have emerged in other sectors as mechanisms to ensure traceability, immutability, and transparency of data across decentralized systems (Alotaibi et al., 2024). Their potential to address persistent traceability challenges in construction has been widely discussed, yet their relevance for biogenic carbon accounting remains largely unexplored. Further, improved traceability and verification are increasingly linked to the credibility of emerging green finance and carbon market instruments (Schletz et al., 2020).

Therefore, this paper adopts an exploratory perspective to investigate how blockchain-inspired system logics could support traceability and transparency of biogenic carbon in biobased construction materials, using bamboo as a representative case. Rather than proposing a finalized technical solution, the paper examines conceptual mismatches between current carbon accounting practices and material realities, and explores how distributed ledger principles could help bridge these gaps.

This study asks: *How can blockchain-inspired system principles be operationalised to guarantee traceable, auditable biogenic-carbon accounting across heterogeneous bamboo value-chains?*

Methods

The exploration follows a scoped and purposive analytical review suitable for an emerging research field, rather than a full systematic literature review. Relevant literature is selectively sampled to identify conceptual patterns and limitations in current biogenic carbon accounting practices, and the analysis is structured around three analytical steps:

(i) identification of traceability gaps in current biogenic carbon accounting practices,
(ii) abstraction of blockchain-relevant principles applicable to these gaps,

(iii) exploratory mapping of these principles onto bamboo-based material value chains,
(iv) formalization of these insights into a conceptual framework for traceable biogenic carbon accounting

First, common biogenic carbon accounting practices were reviewed with a focus on how carbon information is generated, aggregated, and transferred across life cycle stages. Particular attention was given to points where carbon attribution becomes ambiguous, such as material transformation, component assembly, reuse, and end-of-life treatment. This review was informed by existing LCA-based studies on bamboo construction materials and by current Environmental Product Declaration (EPD) and reporting conventions.

Second, blockchain technologies were not treated as a specific software implementation, but as a set of system principles. These include distributed state management, immutable transaction histories, unique identifiers, and rule-based state transitions. The analysis deliberately avoids technological specificity in order to focus on conceptual compatibility rather than implementation feasibility.

Third, bamboo-based construction materials were used as an exploratory case to test how these principles could be mapped onto real material flows. Bamboo value chains are characterized by biological growth phases, multiple processing steps, and a wide range of end-of-life scenarios. Conceptual data entities representing biogenic carbon quantities were defined and linked to material batches or components. Hypothetical life cycle events (such as processing, assembly, reuse, and disposal) were then interpreted as state transitions that would require transparent and verifiable updates of carbon information.

Fourth, the insights derived from the exploratory mapping were formalized into a conceptual framework for traceable biogenic carbon accounting. This step translates the identified system principles and observed material-flow dynamics into a structured entity–transaction model, in which biogenic carbon is represented as a persistent accounting state linked to material representations. Core conceptual elements, including Carbon Units, Material Tokens, and a defined set of life cycle transactions, were specified to capture how carbon information is created, transformed, transferred, and retired over time.

In addition, a set of invariants was introduced to ensure internal consistency, including conservation of carbon, single attribution, traceable aggregation, explicit state transitions, and auditable history. This formalization step consolidates the preceding analytical steps into a coherent system logic that can support transparent, continuous, and verifiable carbon accounting across heterogeneous value chains.

Results analysis and discussion

Review of current biogenic carbon accounting practices and points of ambiguity

Our ongoing analysis indicates that current biogenic carbon accounting practices are internally consistent within individual assessments, yet they exhibit structural limitations once carbon information is expected to remain traceable across multiple life cycle stages and actors. In most LCA-based studies on bamboo construction materials, biogenic carbon is quantified at the point of material production using established inventory data and is reported as an aggregated balance per functional unit (Tellnes et al., 2017). At this initial stage, carbon uptake and storage are clearly defined and methodologically unambiguous.

As materials progress through processing, assembly, and use phases, however, carbon information is progressively abstracted. Processing steps redistribute material mass through yield losses, residues, and by-products, but the associated carbon is commonly reallocated using averaging or allocation rules that do not preserve links to the original biomass source. During component assembly and at building level, carbon indicators are further aggregated into single values, which are suitable for comparative reporting but obscure the provenance and also the distribution of biogenic carbon within multi-material systems. In reuse scenarios, carbon storage is typically addressed through scenario-based assumptions rather than through documented continuity of material histories, while end-of-life treatments rely heavily on generalized modeling choices that are applied retrospectively (Silvia et al., 2025).

Dynamic life cycle assessment (LCA) has been increasingly explored to account for the temporal dynamics of biogenic carbon flows. However, existing approaches vary widely in their methodological choices, including system boundaries, time horizons, and modelling of carbon sequestration and release, reflecting a lack of consensus and leading to substantial variability in results (de Jong et al., 2026). This methodological diversity, together with increased data and modelling requirements, introduces additional complexity and challenges in interpretation (de Jong et al., 2026). While dynamic LCA can provide more temporally consistent representations in certain contexts, its added value remains case-dependent and it is not part of established standardized LCA practice, which today remains predominantly static. The contribution of this paper is complementary to this insight: rather than focusing on temporal modelling, it addresses the continuity and traceability of biogenic carbon across value chains.

These patterns reveal that biogenic carbon is repeatedly recalculated at successive system boundaries rather than transferred as a conserved and persistent quantity. While this approach aligns with the logic of conventional LCA, it limits transparency when carbon information is interpreted across parallel assessments or over extended time horizons. The resulting ambiguity is not primarily a consequence of missing data, but also of accounting structures that prioritize internal coherence over cross-stage continuity. As a result, attribution becomes increasingly dependent on assumptions embedded in models rather than on verifiable material trajectories.

Our findings suggest that the principal challenge in biogenic carbon accounting lies less in improving quantification methods and more in developing

mechanisms that can preserve attribution, continuity, and auditability as materials move through complex value chains. Addressing this challenge requires system logic capable of explicitly managing carbon state transitions over time, which motivates the exploration of blockchain-inspired principles discussed in the following sections.

Blockchain-inspired system principles relevant to biogenic carbon traceability

Our analysis indicates that, when abstracted from specific software implementations, blockchain technologies can be understood as a set of system principles that directly address several of the structural weaknesses identified in current biogenic carbon accounting practices (Nair et al., 2024). When viewed through this lens, these principles show a high degree of conceptual compatibility with the traceability and transparency gaps discussed in the previous section.

We find that distributed state management is particularly relevant for biobased material value chains, where carbon-related information is generated, modified, and interpreted by multiple actors over time (Komdeur & Ingenbleek, 2021). In contrast to centralized databases or document-based reporting, a distributed state model allows all participants to reference a shared and consistent representation of carbon states. Importantly, this does not require full decentralization of governance or the absence of authoritative oversight. Rather, it ensures that updates to carbon information are synchronized across actors and assessment contexts, reducing the risk of divergent interpretations arising from parallel data silos.

Immutable transaction histories directly respond to the loss of provenance observed in conventional accounting workflows. In current LCA and EPD practices, carbon information is typically recalculated and overwritten as materials progress through successive life cycle stages. As a result, earlier assumptions and allocations are no longer accessible once new results are reported. By contrast, an immutable transaction logic preserves previous states while recording subsequent changes, enabling carbon information to remain auditable over time. This makes it possible to trace later interpretations (such as reuse or end-of-life outcomes) back to earlier decisions and methodological choices, rather than treating each assessment as an isolated snapshot in time.

Our analysis suggests that unique identifiers are a prerequisite for persistence. Assigning persistent identifiers to biogenic carbon quantities, rather than only to products, components, or documents, allows carbon information to remain associated with material flows even when physical elements are transformed, aggregated, or separated. This directly addresses the attribution loss observed during processing and component assembly, where carbon is currently redistributed through averaging and aggregation without preserving its original linkage to biomass sources.

Rule-based state transitions emerge as a further critical principle. In the reviewed accounting practices, decisions regarding carbon release, continued storage, or substitution effects are often embedded implicitly in scenario assumptions or applied retrospectively at reporting stages (He et al., 2024). Blockchain-inspired systems allow such decisions to be formalized as explicit rules governing state transitions. For example, conditions under which biogenic carbon remains stored, are transferred to a new use phase, or are considered released at end-of-life, can be defined and documented consistently. This shifts carbon accounting away from static reporting toward a dynamic, state-based representation that better reflects material trajectories over time.

Taken together, our results suggest that the relevance of blockchain technologies for biogenic carbon accounting lies less in automation or decentralization per se, and more in their capacity to formalize continuity, conservation, and transparency of carbon information, consistent with prior findings that digital construction technologies derive value primarily from decision support rather than from automation alone (Graser et al., 2021). By focusing on system logic rather than implementation, the proposed analysis illustrates how blockchain-inspired principles could complement existing LCA frameworks while addressing limitations that cannot be resolved through methodological refinement alone.

Exploratory mapping of blockchain principles onto bamboo-based material flows

Our exploratory mapping of the identified blockchain-inspired principles onto bamboo-based construction materials both illustrate the suitability and the constraints of these logics when confronted with real material flows. Bamboo value chains are particularly illustrative because they span biological growth, heterogeneous processing pathways, and highly variable use and end-of-life scenarios, which together expose where carbon information is currently lost, reinterpreted, or rendered non-verifiable. At the earliest stage, biogenic carbon is associated with biological growth in managed or semi-managed bamboo systems (Buonocore, 2023). In our mapping, this carbon is represented as a conceptual data entity linked to a defined biomass batch, for which uptake is temporally and spatially constrained. At this stage, the carbon state is relatively stable and attribution is comparatively unambiguous. Introducing a persistent identifier at this point enables the carbon quantity to remain referenceable independently of later transformations, thereby establishing a continuous accounting anchor across the subsequent value chain.

As bamboo undergoes processing such as cutting, splitting, flattening, or lamination, the physical material is divided into multiple output streams, including products, residues, and losses (Liu et al., 2024). Treating these steps as explicit state transitions makes the redistribution of biogenic carbon visible, rather than implicitly reallocating it through mass-based averaging. In our mapping, each processing event updates the carbon state by allocating portions of the original carbon entity to new identifiers associated with derived batches or components. This differs from conventional practice, where carbon is often reassigned at the level of the new product without preserving a traceable link to the originating biomass and the yield structure that produced it.

During component assembly, bamboo-derived elements are frequently combined with other biogenic and non-biogenic materials to form components and systems. In our framework, assembly is treated as a reversible aggregation of carbon entities rather than as a terminal reporting step. Carbon information remains linked to individual components even when aggregated for reporting purposes, preserving the option to disaggregate carbon states during selective disassembly, repair, or reconfiguration. This is particularly relevant for bamboo construction systems where replaceability and modular interventions are common, yet the carbon implications of partial replacement are difficult to represent within conventional building-level aggregation.

Reuse scenarios reveal an important advantage of the state-based approach: When a bamboo component is removed from its initial application and reintroduced elsewhere, conventional LCA practice typically represents this through scenario assumptions or crediting approaches that are external to the component's documented history. In our mapping, reuse is recorded as a state transition that extends the storage duration of the existing carbon entity. This avoids re-estimation and reduces the risk of re-claiming the same stored carbon across multiple assessments, while making reuse effects explicit, time-stamped, and auditable (Zea Escamilla et al., 2020).

End-of-life events further demonstrate the relevance of rule-based transitions. Disposal pathways such as incineration with or without energy recovery, landfilling, recycling, or cascading use represent qualitatively different carbon outcomes and are sensitive to timing. Our mapping shows that these end-of-life assumptions often dominate carbon conclusions, yet are rarely documented with sufficient granularity to be verifiable. Encoding end-of-life as a verifiable transition clarifies when and why biogenic carbon is considered released, substituted, or retained, and it links those outcomes to explicit events rather than to generic scenario choices applied after the fact.

Overall, the bamboo case indicates that blockchain-inspired principles can be meaningfully aligned with the material logic of biobased construction systems when life cycle events are represented as state transitions acting on persistent carbon entities. This approach makes allocations, assumptions, and boundary choices more transparent by externalizing them into recorded transitions. At the same time, our analysis suggests that the usefulness of such systems depends less on technological sophistication than on the clarity of the accounting rules, the definition of permissible transitions, and the governance mechanisms used to validate critical events, particularly those at origination and end-of-life.

Importantly, although the specific form of life cycle events such as end-of-life is not known in advance, their occurrence is inevitable. These events are treated as potential state transitions that are recorded as they occur, allowing carbon information to be updated incrementally rather than predefined through fixed scenarios.

Conceptual framework for traceable biogenic carbon accounting

Building on the ambiguities identified in current biogenic carbon accounting practices and our exploratory mapping of blockchain-inspired principles, we propose a conceptual framework intended to support traceable and transparent carbon accounting for biobased construction materials. The framework is not presented as a software architecture. Instead, we formalize a system logic that can complement LCA-based methods by preserving continuity of biogenic carbon information across life cycle stages, actors, and assessment contexts.

The central premise of the framework is that biogenic carbon should be treated as a persistent state associated with material flows, rather than as a static attribute recalculated at each assessment boundary. To operationalize this premise, we distinguish between (i) carbon information as a persistent accounting entity and (ii) physical material representation as a batch- or component-level entity. These are linked through explicit state transitions that correspond to life cycle events. In doing so, we aim to externalize allocation decisions and scenario assumptions into auditable updates rather than allowing them to remain implicit in isolated assessments.

At the core of the framework is a limited set of conceptual entities and transactions that define how biogenic carbon information is created, redistributed, transferred, and retired over time. Table 1 summarizes these elements.

Table 1: Conceptual entities and transactions in the proposed framework

Category	Concept	Description
Entity	Carbon Unit (1 tCO_2eq)	Persistent representation of a quantified amount of biogenic carbon, including metadata on origin, material type, and data quality
Entity	Material Token	Representation of a physical material batch or component, referencing one or more Carbon Units
Transaction	Origination	Creation of Carbon Units associated with biomass growth and initial material production
Transaction	Transformation	Allocation, splitting, or aggregation of Carbon Units following material processing steps
Transaction	Assembly / Disassembly	Reversible aggregation or separation of Material Tokens without loss of carbon provenance
Transaction	Transfer of custody	Change of responsible actor while preserving carbon state and history
Transaction	End-of-life	Terminal or transformative update of Carbon Units reflecting release, substitution, or continued storage

Within this framework, life cycle events are interpreted as state transitions acting on Carbon Units rather than as opportunities for re-estimation. Processing steps redistribute carbon quantities explicitly by allocating portions of existing Carbon Units to derived Material Tokens. In the context of digital ledgers, a token is a cryptographically secured data object that represents a specific right, asset, or quantity and can be transferred, subdivided, and verified without a central intermediary (Marijana & Bernhard, 2025). Assembly aggregates material representation without erasing carbon provenance, thereby preserving the possibility of later disaggregation during repair, reuse, or selective dismantling. Reuse extends the storage duration of existing Carbon Units through documented transitions rather than through retrospective scenario reinterpretation. End-of-life events close or transform carbon states based on documented treatment pathways, making the timing and nature of carbon release or substitution explicit.

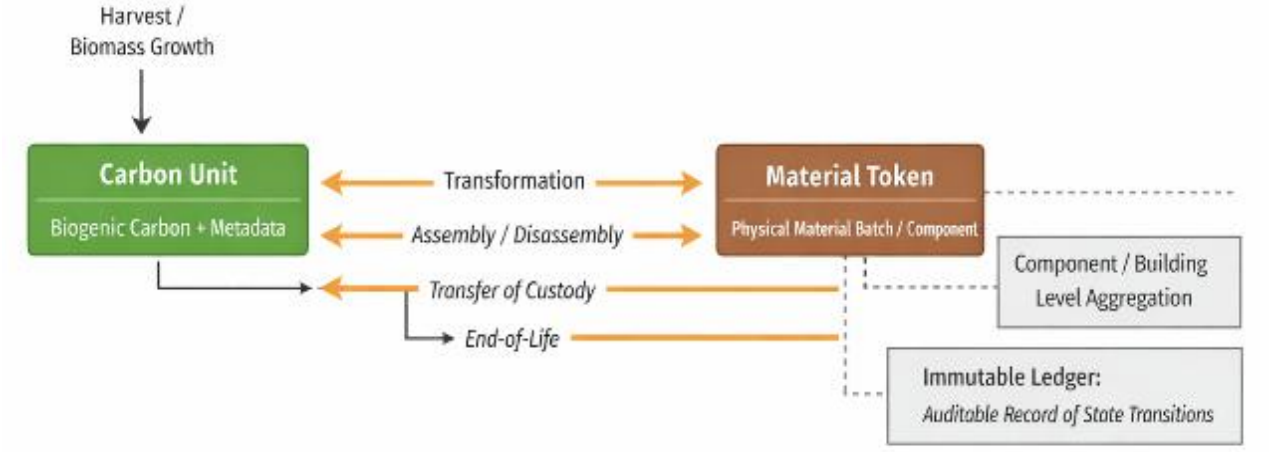


Figure 1: Conceptual entity–transaction diagram for traceable biogenic carbon accounting

To ensure internal consistency and prevent common accounting failures, we define a small set of invariants that govern all transactions (Knecht, 2025):

- First, conservation of carbon requires that the sum of biogenic carbon quantities across derived Carbon Units does not exceed the originating amount, except where methodological adjustments are explicitly recorded as such. This prevents the carbon leakage observed during processing and yield-loss allocation.
- Second, single attribution requires that a Carbon Unit can exist in only one active accounting state at a given time, preventing parallel claims on the same stored carbon. It therefore eliminates the double-counting risk highlighted when bamboo components are aggregated into downstream products.
- Third, traceable aggregation requires that aggregated indicators at component or building level are derived from referenced Carbon Units rather than entered independently. This addresses the loss of provenance that occurs when component-level data are rolled up to building-level totals.
- Fourth, explicit state transitions require that changes in carbon status (e.g., reuse, release, substitution) are represented as transactions rather than implicit assumptions. By doing so, it replaces the implicit scenario assumptions used for reuse and end-of-life modelling.
- Fifth, auditable history requires that state changes preserve previous states, enabling retrospective verification of carbon trajectories and the assumptions that shaped them. This provides the traceability needed to resolve the "black-box" recalculations typical of conventional LCA workflows.

The framework deliberately avoids prescribing specific technologies or governance models. Distributed ledger technologies are relevant insofar as they provide a shared, append-only record of state transitions across multiple actors, thereby supporting consistency and auditability in decentralized value chains. However, the conceptual contribution lies in formalizing continuity, accountability, and conservation of biogenic carbon information, rather than in decentralization or automation per se.

By framing biogenic carbon as a traceable state linked to material flows, the proposed framework addresses the transparency, double counting, and attribution challenges identified in the introduction. It also provides a structured basis for subsequent work on verification, governance, and integration with digital construction environments (for example, see Walzer et al., 2024), while remaining compatible with established LCA and EPD practices.

Figure 1 presents the framework as an entity–transaction system linking Carbon Units (persistent biogenic carbon states with metadata) to Material Tokens (physical batches or components).

Carbon Units are created at origination and persist throughout the system, while Material Tokens can be split, combined, assembled, or disassembled. Life cycle events are represented as transactions that update carbon states according to predefined rules while preserving an

auditable history. Aggregated carbon indicators are then derived from referenced Carbon Units rather than recalculated independently.

The diagram emphasizes conservation, explicit state transitions, and traceability across actors and life cycle stages, without prescribing a specific technological implementation.

Conclusions

Our objective in this paper was to explore how blockchain-inspired system principles could contribute to improved traceability and transparency of biogenic carbon accounting in biobased construction materials, with a particular focus on bamboo value chains. Rather than proposing a definitive technical solution, we set out to examine whether the structural challenges identified in current LCA- and EPD-based practices could be addressed through an alternative accounting logic that preserves continuity of carbon information across life cycle stages.

Our analysis shows that existing biogenic carbon accounting practices are effective for quantification within defined system boundaries but are limited when carbon information must remain attributable and verifiable as materials move through complex value chains. The review and discussion revealed that aggregation during processing and assembly, scenario-based treatment of reuse, and generic end-of-life assumptions progressively detach carbon indicators from their material origins. These limitations do not primarily stem from insufficient data or methodological inconsistency, but from accounting structures that repeatedly reinterpret carbon at successive assessment boundaries rather than transferring it as a conserved quantity.

By abstracting blockchain technologies into a set of system principles, we demonstrated that distributed state management, persistent identifiers, immutable transaction histories, and rule-based state transitions are conceptually well aligned with these challenges. When applied to bamboo-based material flows, these principles enable biogenic carbon to be represented as a persistent state linked to material batches and components and updated explicitly through documented life cycle events. The exploratory mapping showed that such an approach can make processing allocations, reuse effects, and end-of-life outcomes more transparent by externalizing assumptions that are otherwise embedded implicitly in conventional accounting models.

The conceptual framework proposed in this paper builds on these findings by formalizing biogenic carbon as a traceable accounting entity governed by a small set of invariants, including conservation, single attribution, and auditable history. In doing so, it outlines how a carbon ledger logic could complement existing LCA frameworks without replacing established standards. The potential of such a framework does not lie in automation or decentralization per se, but in its capacity to support consistent carbon attribution across actors and over time, thereby reducing the risk of double counting and increasing the credibility of carbon-related claims (Knecht & Stiller, 2023).

Our preliminary results suggest that ledger-based representations of biogenic carbon could provide a robust foundation for future carbon accounting and reporting mechanisms in the built environment. In particular, they offer a structured basis for component-level traceability, for handling reuse and cascading use scenarios, and for linking carbon outcomes to documented material trajectories rather than to abstract scenarios. At the same time, the exploratory nature of this work shows the need for further research on governance, verification, and integration with digital construction workflows, as well as on the institutional conditions under which such carbon ledgers could be adopted.

Overall, this study contributes to ongoing discussions on digitalization and sustainability in construction by reframing biogenic carbon accounting as a state-based, traceable process. By doing so, it opens a pathway toward carbon ledger approaches that can support transparent, verifiable, and scalable carbon accounting for biobased materials within and beyond the bamboo sector.

References

Alotaibi, E. M., Khallaf, A., Abdallah, A. A.-N., Zoubi, T., & Alnesafi, A. (2024). Blockchain-driven carbon accountability in supply chains. Sustainability, 16(24), 10872.

Bundi, T., Lopez, L. F., Habert, G., & Zea Escamilla, E. (2024). Bridging Housing and Climate Needs: Bamboo Construction in the Philippines. Sustainability, 16(2), 498.

Buonocore, L. (2023). Implementation of smart contract solutions on a distributed ledger network (blockchain) based on physiological state variables of woody plants.

de Jong, M., Corona, B., Junginger, M., & Shen, L. (2026). From static to dynamic: Reviewing the application and potential of dynamic LCA to bio-based systems. The International Journal of Life Cycle Assessment, 31(4), 56.

Dwyer, B. P., & Mowry, C. D. (2022). Minimizing Fraud in the Carbon Offset Market Using Blockchain Technologies.

Graser, K., Hunhevicz, J., Jähne, R., Walzer, A. N., Seiler, F., Wüst, R., & Hall, D. M. (2021). A qualitative technology evaluation scoreboard for digital fabrication in construction. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction,

He, Y., Wang, S., Zhou, Z., Xiao, K., Xie, A., & Wu, B. (2024). A Blockchain-based carbon emission security accounting scheme. Computer Networks, 243, 110304.

Hellweg, S., & i Canals, L. M. (2014). Emerging approaches, challenges and opportunities in life cycle assessment. Science, 344(6188), 1109-1113.

Knecht, M. (2025). A robust smart contract approach based on substructural types and on-chain code verification, University of Zurich.

Knecht, M., & Stiller, B. (2023). Cusc: Composable user-centric smart contracts. 2023 5th Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS),

Komdeur, E. M., & Ingenbleek, P. T. (2021). The potential of blockchain technology in the procurement of sustainable timber products. International Wood Products Journal, 12(4), 249-257.

Liu, K., Jayaraman, D., Shi, Y., Xiong, Z., Yang, J., Symeonidis, A., & Zea Escamilla, E. (2024). Life cycle assessment (LCA) of the industrial production of structural glued laminated bamboo. Journal of Cleaner Production, 144367.

Marijana, S., & Bernhard, W. (2025, 2025/July). The Practitioner Perspective: Tokenization of Real Estate Proceedings of the 2025 European Conference on Computing in Construction, https://ec-3.org/publications/conference/paper/?id=EC32025_447. https://doi.org/10.35490/EC3.2025.447

Nair, M. R., Bindu, N., Jose, R., & Satheesh Kumar, K. (2024). From assistive technology to the backbone: The impact of blockchain in manufacturing. Evolutionary Intelligence, 17(3), 1257-1278.

Rincón, C. E., Montoya, J. A., & Archila, H. F. (2023). Bamboo Construction Inspired by Vernacular Techniques for Reducing Carbon Footprint: A Life Cycle Assessment (LCA). Sustainability, 15(24), 16893.

Sayın, M., Sonmez, R., Ahmadisheykhsarmast, S., & Uysal, F. (2025). A DECENTRALIZED BLOCKCHAIN FRAMEWORK FOR TRACING GREEN/SUSTAINABLE MATERIALS DURING BUILDING LIFE CYCLE.

Schletz, M., Nassiry, D., & Lee, M.-K. (2020). Blockchain and tokenized securities: The potential for green finance.

Silvia, M., Lavinia Chiara Chiara, T., Stefano, R., Giovanni, M., Andrea, B., Roberto, N., Rosa, M., & Giuseppe Di, G. (2025, 2025/July). Revolutionizing Construction Procurement: Blockchain and Smart Contracts Driving Transparency and Monitoring Waste Production Proceedings of the 2025 European Conference on Computing in Construction, https://ec-3.org/publications/conference/paper/?id=EC32025_444. https://doi.org/10.35490/EC3.2025.444

Tellnes, L., Ganne-Chedeville, C., Dias, A., Dolezal, F., Hill, C., & Zea Escamilla, E. (2017). Comparative assessment for biogenic carbon accounting methods in carbon footprint of products: a review study for construction materials based on forest products. iForest-Biogeosciences and Forestry, 10(5), 815.

Walzer, A. N., Kozlova, M., & Yeomans, J. S. (2024). Printing homes: Unit cost estimation for additive manufacturing in construction. In M. Kozlova & J. S. Yeomans (Eds.), Sensitivity analysis for business, technology, and policymaking (pp. 116–138). Routledge. https://doi.org/10.4324/9781003453789-8

Walzer, A. N., Tan, T., Graser, K., & Hall, D. M. (2025). Bug or feature? Institutional misalignments between construction technology and venture capital. Construction Management and Economics, 43(2), 130-152.

Zea Escamilla, E., Klippel, M., Cakir, I., & Maniak, M. (2020). Challenges and opportunities for timber buildings in the swiss real estate market-The role of carbon credits: Financing the low-carbon economy-Instruments, Barriers and Recommendations. Swiss Sustainable Finance.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

HOUSEHOLD-LEVEL ENERGY POVERTY ASSESSMENT IN IRELAND: A COMPARATIVE ANALYSIS OF INDICATOR-BASED APPROACHES

Afshin Saeedian[1], Lorna Fitzsimons[1], James McGrath[2], Arash B. Naamrudi[1], Reihaneh Aghamolaei[1]
[1] School of Mechanical and Manufacturing Engineering, Faculty of Engineering and Computing, Dublin City University, Dublin 9, Whitehall, Ireland
[2] Department of Physics, Maynooth University, County Kildare, Maynooth, Ireland

Abstract

Energy poverty measurement at the household level is sensitive to indicator choice and data availability. This study evaluates energy poverty in Dublin using a set of expenditure-based, income-sensitive, and needs-based indicators with a Building Energy Rating (EPC) dataset. Household income uncertainty is addressed through scenario-based income assumptions, enabling estimation of indicator-specific thresholds and escape income levels. Results reveal divergence in household classification across indicators, driven largely by structurally high energy costs linked to dwelling characteristics. The findings highlight the limitations of single-indicator approaches and demonstrate the value of comparative, household-level frameworks for informing targeted energy poverty interventions in data-constrained contexts.

Introduction

Energy poverty is increasingly recognised as a multidimensional condition arising from the interaction of low household income, high energy costs, and poor dwelling energy efficiency, which together restrict access to essential domestic energy services. This concept extends beyond traditional fuel poverty that focuses primarily on heating affordability, and instead encompasses the broader set of energy services required for an acceptable standard of living, including space heating, water heating, lighting, and appliance use (Boardman 1991; González-Eguino 2015; Bouzarovski et al. 2014; Bouzarovski and Simcock 2017; Thomson et al. 2017; Scarpellini et al. 2019; Shahzad et al. 2022).

Across Europe, energy poverty has emerged as a growing policy concern, driven by a combination of rising energy prices, ageing building stock, and persistent income inequality. Recent estimates suggest that a substantial share of households experience difficulties in maintaining adequate indoor thermal comfort or meeting essential energy needs, with impacts extending beyond energy affordability to health, well-being, and social inclusion (European Commission et al. 2024; Thomson et al. 2017; 2017). In Ireland, these challenges are particularly pronounced due to a housing stock characterised by low energy efficiency, high dependence on fossil fuels for space heating, and comparatively slow rates of deep retrofit, increasing the exposure of households to energy cost shocks (ESRI et al. 2022).

Despite growing policy attention, the identification and measurement of energy poverty remain contested. Existing approaches rely heavily on single-dimensional indicators, such as fixed expenditure thresholds or income-based rules, which are often unable to capture the combined effects of affordability constraints and dwelling-related energy needs. Indicators such as the 10% rule or income-only measures may misclassify households that under-consume energy due to affordability pressures, or overlook those living in energy-inefficient dwellings with structurally high energy requirements (Hills 2011; Palma et al. 2024; Wojewódzka-Wiewiórska et al. 2024; Al Kez et al. 2024). As a result, the scale and nature of energy poverty can vary substantially depending on the indicator applied, raising concerns regarding robustness, comparability, and policy relevance.

In the Irish context, the measurement of energy poverty is further constrained by data availability and scale. While national surveys provide valuable insights into income and self-reported deprivation, they lack the spatial and technical resolution required to assess dwelling-specific energy needs. Conversely, the EPC database offers detailed information on building characteristics and modelled energy demand at the household level, but does not contain direct information on household income. This separation between economic and technical data has limited the ability to evaluate energy poverty consistently across indicators and at fine spatial scales, creating a methodological gap between policy monitoring and building-level intervention design.

Building on these conceptual and methodological debates in the energy poverty literature, the following section outlines the data sources and analytical framework used to operationalise and compare multiple energy poverty indicators at the household level in Dublin.

Methodology

The methodology consists of three main stages: (i) preparation and preprocessing of EPC data, (ii) Application of leading energy poverty indicators on a common analytical baseline, and (iii) household-level result generation and cross-indicator comparison for the Dublin residential stock.

Step 1: EPC dataset preparation and preprocessing

This study uses a cleaned and validated Irish EPC dataset for residential dwellings in Dublin as the primary technical evidence base. Each record represents an individual dwelling and includes dwelling descriptors, delivered energy by end use, heating fuel types, and energy performance ratings. The analysis is conducted at dwelling (household proxy) level.

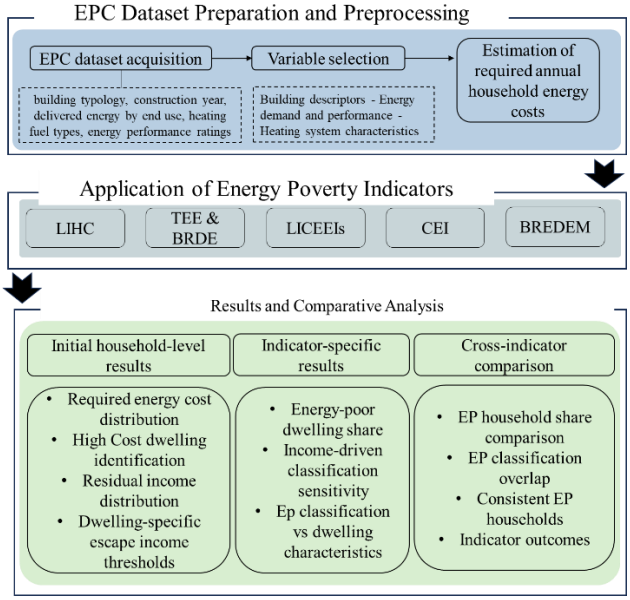


Figure 1: Methodological workflow for household-level energy poverty assessment using EPC data.

Variable selection and analytical dataset construction

To ensure methodological transparency and reproducibility, the full EPC dataset was reduced to a minimum set of variables required to apply the selected energy poverty indicators and to compute annual required energy costs. The retained variables were grouped into three categories: 1. Dwelling descriptors (unique dwelling identifier, dwelling archetype, Year of Construction, and BER Rating), 2. Delivered energy by end use (kWh/year) (space heating, water heating, lighting, fans and pumps, and other electricity uses where available), and Fuel type metadata (Main Space Heating Fuel and Main Water Heating Fuel).

These variables form the common input layer for all subsequent indicator implementations.

Estimation of required annual household energy Costs

Annual required household energy costs were estimated by converting delivered energy to cost using fuel-type-specific unit prices. Space-heating and water-heating costs were calculated by multiplying delivered energy (kWh/year) by the assigned unit price of the corresponding heating fuel. Electricity costs were computed by aggregating electricity-related end uses (lighting, pumps/fans, and recorded electricity consumption where available) and applying the electricity unit price. The total annual required energy cost for each dwelling was then calculated as:

$$C_i = C_i^{space} + C_i^{water} + C_i^{elec}$$

where C_i is the total required energy cost for dwelling i. This cost measure is treated as a required energy cost proxy derived from EPC data, enabling consistent comparison of structural energy burden across dwellings prior to introducing any income assumptions.

The resulting dataset provides a harmonised dwelling-level input layer comprising building descriptors, delivered energy by end use, fuel type information, and an estimated annual required energy cost, which together underpin the implementation of all subsequent energy poverty indicators.

Step 2: Application of energy poverty indicators

Following the preparation of a harmonised EPC-based analytical dataset and the estimation of required annual household energy costs, energy poverty was determined using a suite of leading indicators identified from the literature. To ensure comparability, all indicators were applied on a common technical baseline and evaluated consistently at dwelling (household proxy) level.

Given the absence of household income information in EPC data, income uncertainty was addressed through a scenario-based assumed-income framework, rather than income imputation or spatial proxies. This approach enables transparent sensitivity analysis and avoids introducing hidden bias into indicator application.

Income representation and poverty threshold

Household income was represented using a discrete set of assumed annual income levels spanning low to high income ranges. These income scenarios were applied uniformly across all dwellings, allowing assessment of how energy poverty classification responds to changes in income while holding building characteristics constant.

Low-income status was defined using the national at-risk-of-poverty threshold for Ireland, corresponding to 60% of the median equivalised disposable income. For the purposes of this study, the 2024 poverty line of €17,998 was adopted. This threshold was used consistently across all indicators incorporating income or residual income criteria.

Low income high costs (LIHC)

The LIHC indicator was implemented following the residual income formulation widely adopted in EU and UK methodological guidance. A dwelling was classified as energy poor if it simultaneously satisfied:

- High-Cost condition: Required annual energy cost exceeds the median energy cost of the Dublin sample (€2,129).
- Low Income condition: Residual household income after paying required energy costs falls below the poverty threshold (€17,998).

Formally, for dwelling i with assumed income Y and required energy cost C_i, energy poverty occurs when:

$$C_i > \text{Median}(C) \wedge (Y - C_i) < PL$$

where PL denotes the national poverty line.

In addition to binary classification, a dwelling-specific escape income was derived, defined as the minimum

income required for a household to avoid LIHC classification given its energy cost.

Combined TEE & BRDE indicator

To reduce misclassification associated with single-threshold indicators, a combined TEE & BRDE indicator was applied. A dwelling was classified as energy poor if it simultaneously satisfied:

1. Energy expenditure exceeding 10% of household income (TEE),
2. Required energy cost above the median (BRDE),
3. Residual income after energy costs below the poverty threshold.

This combined formulation integrates affordability stress, relative energy burden, and income vulnerability within a single classification framework.

Low income combined with energy expenditure indicators (LICEEIs)

LICEEIs were implemented to capture households experiencing energy poverty due to the interaction of low income and energy-related constraints, including hidden energy poverty. The indicator combines low residual income status with either:

- high required energy costs, or
- unusually low energy consumption relative to dwelling characteristics (used as a proxy for under-heating).

This formulation enables identification of households whose energy poverty may not be fully captured by expenditure-based thresholds alone.

CEI (composite expenditure indicator)

CEI integrates multiple expenditure and income dimensions into a single composite score. Required energy costs and income measures were normalised and combined to produce a continuous vulnerability index, allowing households to be ranked by degree of energy poverty severity rather than classified using a binary threshold.

BREDEM-type model-based indicator

A BREDEM-type indicator was applied using EPC-derived delivered energy and building characteristics as proxies for required energy needs. By focusing on modelled energy demand rather than observed consumption, this indicator captures structurally driven energy needs and enables identification of hidden energy poverty associated with constrained energy use.

Step 3: results generation and comparative analysis

Following the application of the selected energy poverty indicators, results were generated and analysed at dwelling (household proxy) level for the residential building stock of Dublin. This step focuses on three complementary result layers: (i) initial household-level insights, (ii) indicator-specific outcomes, and (iii) cross-indicator comparison.

Initial household-level results for Dublin

The first stage of analysis examines the distribution of key energy poverty-related variables across the full sample of dwellings. This includes:

- distribution of required annual household energy costs,
- identification of high-cost dwellings,
- distribution of residual income under assumed income scenarios,
- derivation of dwelling-specific escape income thresholds.

These results provide an initial overview of structural energy burden and affordability stress within the Dublin housing stock. By analysing required energy costs prior to applying any indicator-specific thresholds, this stage isolates the role of building energy performance in shaping household vulnerability.

Descriptive statistics and distributions are used to highlight heterogeneity across dwellings and to identify segments of the housing stock exposed to disproportionately high energy costs, independent of income assumptions.

Indicator-specific results

In the second stage, results are analysed separately for each energy poverty indicator. For each indicator, the following outputs are produced:

- the number and proportion of dwellings classified as energy poor,
- sensitivity of classification to assumed household income levels,
- relationship between classification outcomes and dwelling characteristics (e.g. typology, construction period, energy rating).

For expenditure-based indicators (LIHC, TEE, TEE & BRDE), particular attention is given to the income levels at which households transition into or out of energy poverty. Dwelling-specific escape income thresholds are used to quantify how much income is required for households to avoid energy poverty conditional on building energy performance.

For composite and model-based indicators (LICEEIs, CEI, BREDEM-type), results are interpreted in terms of vulnerability intensity and structural energy need, enabling identification of households experiencing hidden or borderline energy poverty.

This indicator-specific analysis allows differences in sensitivity, strictness, and diagnostic capacity across indicators to be systematically assessed.

Cross-indicator comparison

The final stage of analysis compares results across all implemented indicators. This includes:

- comparison of the proportion of households identified as energy poor by each indicator,

- assessment of overlap and divergence in household classification,
- identification of households consistently classified as energy poor across multiple indicators,
- evaluation of how indicator choice influences conclusions regarding the scale and nature of energy poverty.

By applying all indicators to a common technical and income-scenario baseline, the analysis enables direct comparison of indicator behaviour and robustness. Differences in classification outcomes are interpreted in relation to each indicator's underlying conceptual focus, distinguishing between income-driven, expenditure-driven, and structure-driven vulnerability.

This comparative analysis forms the basis for identifying strengths and limitations of existing indicators and provides empirical grounding for the future development of a refined multidimensional energy poverty indicator tailored to the Irish context.

Result

Initial household-level outputs

The initial stage of the results focuses on household-level outputs derived directly from the cleaned and validated EPC dataset for Dublin. This stage aims to characterise the distribution of required energy costs, identify structurally high-cost dwellings, and assess affordability pressures prior to the application of income-based energy poverty indicators.

Distribution of required annual energy costs

Figure 2 illustrates the distribution of required annual household energy costs derived from the cleaned EPC dataset for Dublin. The distribution is markedly right-skewed, indicating substantial heterogeneity in energy cost requirements across the housing stock. While the majority of dwellings cluster at relatively low to moderate cost levels, a pronounced upper tail is observed, reflecting a subset of dwellings with exceptionally high required energy costs.

The dashed vertical line denotes the median energy cost threshold (€2,129), which serves as the structural high-cost benchmark within the LIHC framework. A considerable proportion of dwellings lie to the right of this threshold, demonstrating that high required energy costs are a widespread structural feature of the housing stock rather than a marginal phenomenon. This pattern indicates that energy affordability pressures in Dublin are not confined to a small number of extreme cases, but instead affect a substantial segment of dwellings prior to the application of income-based criteria.

Overall, Figure 3.1a provides direct empirical evidence that required energy costs are unevenly distributed and structurally driven, justifying the subsequent identification of high-cost dwellings and the application of income-sensitive energy poverty indicators in the following analysis.

To better visualise the upper tail of the cost distribution, the same data are shown on a logarithmic scale in Figure 3. This representation reveals a long and continuous upper tail, confirming that extremely high energy cost values are structurally embedded within the dataset rather than being isolated outliers. These high-cost dwellings are likely associated with unfavourable building characteristics, inefficient heating systems, or large floor areas, underscoring the importance of building-related factors in energy poverty analysis.

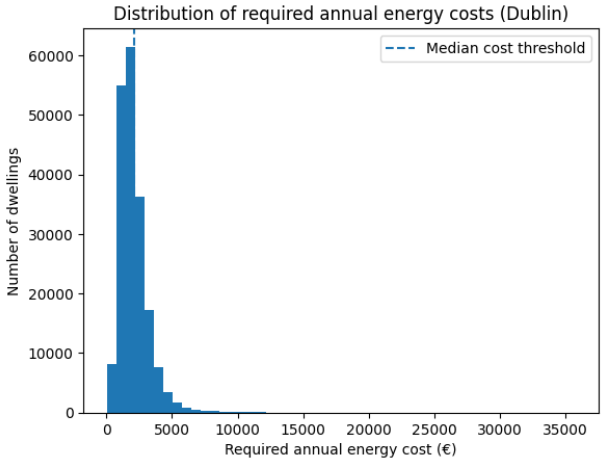


Figure 2: Distribution of required annual household energy costs (Dublin)

This pattern confirms that a subset of dwellings faces structurally elevated energy requirements spanning several orders of magnitude, reinforcing the role of building-related characteristics in shaping energy affordability pressures. The persistence of high-cost values beyond the median threshold further supports the identification of structurally high-cost dwellings prior to the application of income-based energy poverty indicators.

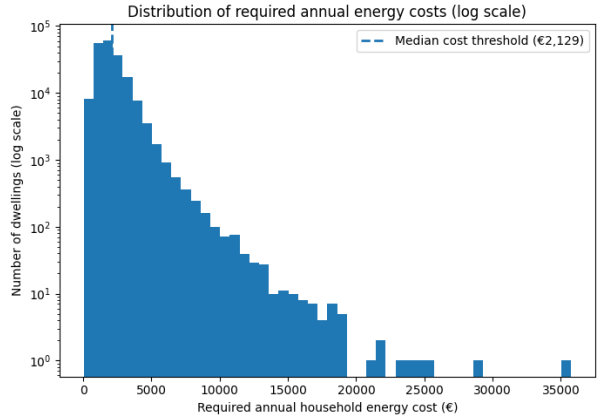


Figure 3: Distribution of required annual household energy costs (Dublin – log scale)

Identification of structurally high-cost dwellings

Using the median required annual energy cost (€2,129) as a structural threshold, dwellings were classified into high-cost and non–high-cost groups. Approximately 37% of dwellings in Dublin were identified as structurally high-cost, indicating that a substantial share of the housing stock faces elevated energy requirements independent of

household income. This classification provides the basis for subsequent residual income and escape income analyses within the LIHC framework.

Escape income distribution for structurally high-cost dwellings

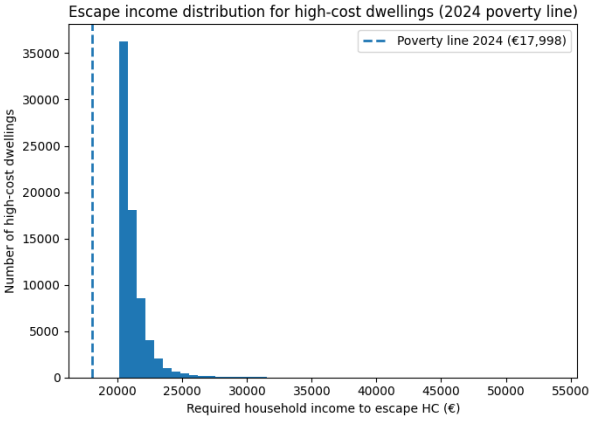


Figure 4. Distribution of escape income for high-cost dwellings

Indicator-specific results

The results reveal pronounced differences in both prevalence levels and income sensitivity across indicators. At very low assumed income levels (€10,000–€17,998), income-driven indicators such as LIHC, TEE & BRDE, and CEI identify approximately 37% of dwellings as energy poor, corresponding closely to the share of structurally high-cost dwellings. In contrast, indicators incorporating energy need and under-consumption dimensions exhibit substantially higher prevalence. LICEEI identifies approximately 52% of dwellings as energy poor, while the BREDEM-based proxy classifies up to 88% of dwellings as energy poor at the lowest income scenario. This divergence highlights the extent to which needs-based indicators capture vulnerability beyond income inadequacy alone.

As assumed household income increases, marked differences emerge in indicator sensitivity. For LIHC, TEE & BRDE, and CEI, energy poverty prevalence remains constant up to the poverty-line threshold (€17,998) and then declines sharply beyond €20,000, falling below 1% at €25,000. This stepwise reduction reflects the threshold-based nature of these indicators and their strong dependence on income adequacy.

In contrast, LICEEI and BREDEM exhibit a more gradual decline with rising income. Notably, at an assumed income of €20,000, LICEEI continues to classify 38% of dwellings as energy poor, while BREDEM identifies over 42%, indicating persistent vulnerability driven by high structural energy needs rather than low income alone. However, at higher income levels (€25,000 and above), convergence across indicators is observed, with prevalence dropping to negligible levels for all measures.

Overall, these results demonstrate that indicator choice substantially influences both the scale and interpretation of energy poverty. Income-driven indicators primarily identify affordability-driven vulnerability, while needs-based indicators reveal a broader set of households exposed to structurally driven energy hardship. This finding provides strong empirical justification for cross-indicator comparison and for the development of multidimensional approaches to energy poverty assessment.

Cross-indicator comparison

Figure 5 compares the sensitivity of energy poverty classification to assumed household income across all examined indicators. Clear and systematic differences are observed in both the level and persistence of energy poverty identified by each indicator, highlighting the strong influence of methodological choice on assessment outcomes.

At low assumed income levels (€10,000–€17,998), income-driven indicators (LIHC, TEE & BRDE, and CEI) converge at approximately 37%, closely matching the income-independent share of structurally high-cost dwellings. This indicates that, at low incomes, affordability constraints dominate classification outcomes, and energy poverty identification is primarily driven by the high-cost condition embedded within these indicators.

In contrast, indicators incorporating building energy needs and consumption constraints exhibit substantially higher prevalence. LICEEI identifies around 52% of dwellings as energy poor across low-income scenarios, while the BREDEM-based proxy reaches values as high as 88% at the lowest assumed income. This divergence demonstrates that needs-based indicators capture a broader set of vulnerable dwellings, including those facing high structural energy requirements that are not fully reflected in income-based measures.

As assumed income increases, sharp discontinuities emerge for income-sensitive indicators. For LIHC, TEE & BRDE, and CEI, energy poverty prevalence collapses rapidly beyond €20,000, approaching zero at €25,000 and above. This stepwise behaviour reflects the threshold-based structure of these indicators and underscores their strong dependence on income adequacy rather than persistent dwelling-level energy demand.

By contrast, LICEEI and BREDEM display a more gradual decline with increasing income. Notably, both indicators continue to identify a substantial share of households as energy poor at €20,000, indicating that elevated energy needs remain a significant source of vulnerability even when income approaches moderate levels. However, convergence across all indicators is observed at higher incomes, where prevalence becomes negligible, suggesting that income sufficiency eventually offsets structural energy cost pressures.

Overall, the cross-indicator comparison confirms that energy poverty is inherently multidimensional. Income-based indicators primarily identify affordability-driven vulnerability, while needs-based indicators reveal structurally driven energy hardship linked to building performance. These findings demonstrate that reliance on a single indicator risks mischaracterising both the scale and nature of energy poverty, providing strong empirical

support for comparative and multidimensional assessment approaches.

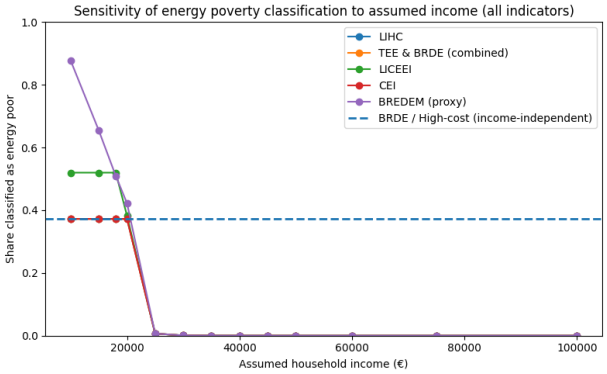


Figure 5: Sensitivity of EP share to income (all indicators)

Discussion and future directions

The results demonstrate that energy poverty in Dublin is strongly shaped by both household income and structurally driven energy costs associated with building performance. The initial household-level analysis shows a highly uneven distribution of required energy costs, with a substantial share of dwellings facing structurally high energy requirements. This structural heterogeneity directly translates into elevated escape income thresholds for many households, indicating that energy poverty cannot be explained by low income alone.

The indicator-specific and cross-indicator analyses further reveal that energy poverty prevalence and classification are highly sensitive to the choice of indicator. Income-based measures such as LIHC and TEE & BRDE primarily identify affordability-driven vulnerability and exhibit sharp threshold effects as income increases. In contrast, needs-based indicators, including LICEEI and the BREDEM-based proxy, capture a broader and more persistent form of vulnerability linked to building inefficiency and elevated energy demand. These findings confirm that reliance on a single indicator risks mischaracterising both the scale and nature of energy poverty.

From a policy perspective, the results highlight the need for integrated assessment frameworks in the Irish context. While income support measures may alleviate energy poverty for some households, the persistence of vulnerability among structurally high-cost dwellings underscores the importance of targeted building retrofit interventions. Future work in Ireland could benefit from the development of multidimensional energy poverty indicators that combine economic affordability, building energy performance, and aspects of household comfort and well-being, such as thermal adequacy and the ability to maintain healthy indoor conditions.

Several limitations should be acknowledged. The analysis relies on EPC-derived estimates of required energy costs rather than measured consumption, which may not fully capture behavioural effects such as under-heating or rebound. Household income was represented through assumed income scenarios, rather than observed microdata, which limits the ability to reflect actual income distributions. In addition, comfort-related dimensions were not explicitly quantified in the current framework.

Future research should therefore focus on integrating EPC data with household survey datasets and measured energy consumption where available, enabling more robust validation of energy poverty indicators. Expanding the framework to incorporate multidimensional well-being and comfort metrics would further enhance its relevance for policy design and evaluation, supporting more effective and equitable energy poverty interventions in Ireland.

References

Al Kez, Dlzar, Aoife Foley, Christopher Lowans, and Dylan Furszyfer Del Rio. 2024. "Energy Poverty Assessment: Indicators and Implications for Developing and Developed Countries." Energy Conversion and Management 307 (May): 118324. https://doi.org/10.1016/j.enconman.2024.118324.

Boardman, Brenda. 1991. "Fuel Poverty: From Cold Homes to Affordable Warmth." (No Title).

Bouzarovski, Stefan, Saska Petrova, and Sergio Tirado-Herrero. 2014. "From Fuel Poverty to Energy Vulnerability: The Importance of Services, Needs and Practices." SSRN Electronic Journal, ahead of print. https://doi.org/10.2139/ssrn.2743143.

Bouzarovski, Stefan, and Neil Simcock. 2017. "Spatializing Energy Justice." Energy Policy 107 (August): 640–48. https://doi.org/10.1016/j.enpol.2017.03.064.

ESRI, Michelle Barrett, Niall Farrell, ESRI, Barra Roantree, and ESRI. 2022. Energy Poverty and Deprivation in Ireland. ESRI. https://doi.org/10.26504/rs144.

European Commission, E3M., Trinomics., IEECP., and Wuppertal Institut. 2024. Study on Optimisation of Energy Poverty Indicators Collected at EU and National Level: Final Report. Publications Office. https://data.europa.eu/doi/10.2833/6627986.

González-Eguino, Mikel. 2015. "Energy Poverty: An Overview." Renewable and Sustainable Energy Reviews 47 (July): 377–85. https://doi.org/10.1016/j.rser.2015.03.013.

Hills, John. 2011. Fuel Poverty: The Problem and Its Measurement.

Palma, Pedro, Roberto Barrella, João Pedro Gouveia, and José Carlos Romero. 2024. "Comparative Analysis of Energy Poverty Definition and Measurement in Portugal and Spain." Utilities Policy 90 (October): 101770. https://doi.org/10.1016/j.jup.2024.101770.

Scarpellini, Sabina, M. Alexia Sanz Hernández, José M. Moneva, Pilar Portillo-Tarragona, and María Esther López Rodríguez. 2019. "Measurement of Spatial Socioeconomic Impact of Energy Poverty." Energy

Policy 124 (January): 320–31. https://doi.org/10.1016/j.enpol.2018.10.011.

Shahzad, Umer, Mansi Gupta, Gagan Deep Sharma, Amar Rao, and Ritika Chopra. 2022. “Resolving Energy Poverty for Social Change: Research Directions and Agenda.” Technological Forecasting and Social Change 181 (August): 121777. https://doi.org/10.1016/j.techfore.2022.121777.

Thomson, Harriet, Stefan Bouzarovski, and Carolyn Snell. 2017. “Rethinking the Measurement of Energy Poverty in Europe: A Critical Analysis of Indicators and Data.” Indoor and Built Environment 26 (7): 879–901. https://doi.org/10.1177/1420326X17699260.

Wojewódzka-Wiewiórska, Agnieszka, Hanna Dudek, and Katarzyna Ostasiewicz. 2024. “Household Energy Poverty in European Union Countries: A Comparative Analysis Based on Objective and Subjective Indicators.” Energies 17 (19): 4889. https://doi.org/10.3390/en17194889.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

QUANTIFYING THE ROLE OF ASSET MAINTENANCE FOR RESILIENT INFRASTRUCTURE

Shyamratna Yadav, Chirag Kothari, and Anushka Singh Rathore
Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur, India

Abstract

Asset maintenance and infrastructure resilience are directly related. However, most resilience assessments overlook pre-disaster asset condition. This study develops a probabilistic lifecycle framework integrating stochastic condition-index (CI) evolution, performance degradation, recovery modeling, and explicit maintenance strategies within a multi-hazard simulation environment. The framework quantifies how pre-event maintenance shapes resilience trajectories. Application to a representative bridge portfolio demonstrates that preventive maintenance consistently improves lifecycle performance and the Composite Resilience Performance Index (CRPI), while rehabilitation provides substantial but time-dependent recovery gains. The proposed framework offers a quantitative decision-support tool linking maintenance actions to resilience outcomes, supporting resilience-informed infrastructure planning and management.

Introduction

Critical infrastructure and lifelines such as bridges are fundamental to economic activity, emergency response, and community well-being (Bruneau et al. 2003, Francis and Bekera 2014). The capacity of these assets to resist, absorb and recover from disruptive events determines not only direct repair needs but also continuity of services, the integrity of value chains, and public safety (Bruneau et al. 2003, Cimellaro 2016). Modern resilience thinking therefore extends conventional engineering objectives into life-cycle goals that prioritise robustness, rapidity, redundancy and resourcefulness. Despite this conceptual progress, routine asset management rarely incorporates explicit traceable resilience objectives in maintenance decision processes (Pant et al. 2019).

During normal operating conditions (pre-event or pre-disaster), the infrastructure components often degrade through multiple mechanisms such as chloride ingress, corrosion, fatigue accumulation, scour and freeze–thaw cycles that often produce nonlinear transitions to accelerated capacity loss and discrete damage states (Li 2003, Ghodoosi et al. 2018). Thus, the performance of an asset also decreases during pre-event/pre-disaster phase. Resilience and fragility assessments that assume "as-new" properties therefore risk overestimating robustness and understating recovery needs (Yang and Frangopol 2019). The pre-event condition of an asset is not a peripheral descriptor but a primary determinant of resilience trajectories within the performance–time framework (Poulin and Kane 2021).

As far as pre-event condition is concerned, Practitioners commonly summarise structural health using composite condition indices (CIs) that aggregate inspection observations, monitoring signals and surrogate indicators into a single, interpretable metric (Li 2003, Li and Gao 2022). Treating CI as an explicit, time-evolving state variable creates a direct modelling bridge between maintenance actions, state-dependent vulnerability, and post-event recovery dynamics. Empirical and modelling studies indicate that poorer pre-event condition is associated with larger immediate performance losses and slower recovery for comparable hazard intensities, implying that maintenance which improves CI yields resilience benefits beyond mere reductions in repair effort (Salem et al. 2020, Poulin and Kane 2021). Accordingly, CI provides a practical means to connect inspection data and maintenance choices to probabilistic performance outcomes.

Further, how CI changes over time due to wear and tear dictates maintenance strategies. Maintenance strategies vary in timing, triggering mechanism, cost structure and mobilization lag. Broadly, strategies can be classified as corrective (reactive), preventive (periodic), predictive (condition-based) and rehabilitative (capacity-restoring or upgrading) (Durán and Vergara 2022, Applegate and Tien 2019). These approaches interact with resilience in different ways: preventive and predictive actions tend to preserve pre-event resilience by slowing deterioration, whereas corrective and rehabilitative interventions accelerate recovery and can modify fragility characteristics (Li and Gao 2022, Salem et al. 2020). Because mobilization lag and implementation constraints differ across strategies, resilience-aware decision support must weigh time-dependent benefits (for example, avoided downtime and reduced recovery time) alongside operational feasibility and budgetary limits.

Thus, under normal operation conditions, the maintenance decisions are driven by asset deterioration, maintenance strategies, and budget availability. Additionally, assets' health is impacted by extreme weather events. Fragility curves are used to represent this information.(Otárola et al. 2022) Fragility curves translate hazard intensity into damage probability. Classical fragility

formulations commonly assume fixed vulnerability parameters, but recent advances emphasise state-dependent fragility in which median and dispersion parameters vary with CI, cumulative damage or material degradation (Izadi et al. 2020, Otárola et al. 2022). Physically, section loss or foundation scour lowers the intensity required to reach a given damage state; conversely, CI improvements through maintenance shift fragility curves toward lower damage probabilities for the same hazard level. Including CI explicitly in fragility models thus provides a mechanistic pathway to quantify how maintenance investments alter risk exposure and expected performance under stochastic hazards (Liu et al. 2020). State-aware fragility operationalises the causal linkage between maintenance and risk.

Multi-hazard and cascading risk chains introduce further complexity. Infrastructure commonly experiences sequences of hazards—earthquakes followed by floods or landslides, storms that induce scour and amplify hydraulic vulnerability—where coupled hazards produce non-additive impacts (Ganin et al. 2017, Dong and Frangopol 2016). Prior events can increase susceptibility to subsequent hazards (for example, seismic damage that exacerbates scour), which may reprioritise maintenance when single-hazard assumptions are used Dong and Frangopol (2016). Robust resilience planning therefore requires propagation of uncertainty across multi-hazard sequences and explicit accounting for how maintenance alters residual capacity available to withstand future events. Consequently, life-cycle analyses that simulate hazard sequences and maintenance interactions are essential for realistic prioritisation.

Methodologically, several research streams are mature enough to be integrated into a unified life-cycle framework: stochastic degradation and ageing models (Liu et al. 2020); Bayesian updating of fragility from damage observations (Li and Gao 2022); Monte Carlo simulation of multi-hazard portfolios (Liu et al. 2020); and translation of probabilistic service availability into prioritisation metrics (Poulin and Kane 2021). Nevertheless, three gaps constrain uptake in practice: (1) multi-hazard, state-aware fragility models that directly ingest CI observations are not standard in maintenance portfolios; (2) non-additive coupling of sequential hazard damage within life-cycle analyses remains underexplored; and (3) end-to-end frameworks that convert CI-driven resilience metrics into quantitative decision criteria for maintenance prioritisation under uncertainty are still nascent (Ouyang 2014, Pant et al. 2019).Addressing these gaps would make resilience metrics actionable for asset managers operating under resource constraints.

This paper addresses these gaps by developing a probabilistic lifecycle framework that can be used to estimate the role of asset maintenance in improving resilience of an infrastructure system. The developed framework integrates probabilistic life-cycle methodology linking stochastic CI dynamics with CI-congruent fragility, defines decision policies for preventive, corrective and rehabilitative maintenance strategies , simulates multi-hazard and cascade scenarios via Monte Carlo sampling, and translates probabilistic performance trajectories into resilience-oriented metrics for strategy ranking and portfolio prioritisation. We demonstrate the framework on a synthetic span-bridge portfolio to illustrate how CI-aware fragility modelling combined with hazard sequencing can materially influence resilience outcomes and inform maintenance prioritisation in resource-constrained settings. The proposed approach aims to provide infrastructure operators with a practicable decision-support tool that integrates inspection and maintenance information into investment choices guided by resilience objectives and contemporary asset-management standards (International Organization for Standardization 2014, United Nations Office for Disaster Risk Reduction (UNDRR) 2022, World Bank 2021).

Methodology

This section summarises the probabilistic simulation engine used to compare corrective, preventive and rehabilitation maintenance strategies. The model couples (i) a physics-informed, condition-dependent deterioration law, (ii) log-normal fragility models conditioned on asset condition, (iii) deterministic damage-state (DS) selection via a shared event shock, (iv) loss mapping to condition index (CI), and (v) recovery dynamics. Outputs are life-cycle and event-level resilience metrics (CRPI, Robustness, Recoverability, Rapidity). The approach follows established resilience and fragility practices Bruneau et al. (2003), Cimellaro et al. (2010), Porter (2015), Committee et al. (2013).

Simulation workflow

A Monte-Carlo ensemble of N trials is executed. For each trial: The framework will be implemented using a Monte Carlo loop, where the entire simulation process will be repeated several times in order to capture the variability of hazard occurrence.

1. Pre-sample multi-hazard events (AEPs and intensities) and a uniform shock variable $u \in [0, 1)$ per event; the same u is replayed across all strategies to ensure comparability. All simulations are performed over a fixed lifecycle horizon with identical time discretization, ensuring that all Monte Carlo trials are equal in duration and directly comparable across maintenance strategies.
2. Propagate monthly deterioration (strategy-neutral physics).
3. Schedule/apply maintenance according to the tested strategy (Preventive: periodic jump; Corrective: post-event patch with delay; Rehabilitation: triggered explicit target).
4. If an event occurs at time t, evaluate fragility (CI-conditioned), determine DS via u, map DS to loss fraction, and update CI.
5. Apply recovery dynamics and record time series /

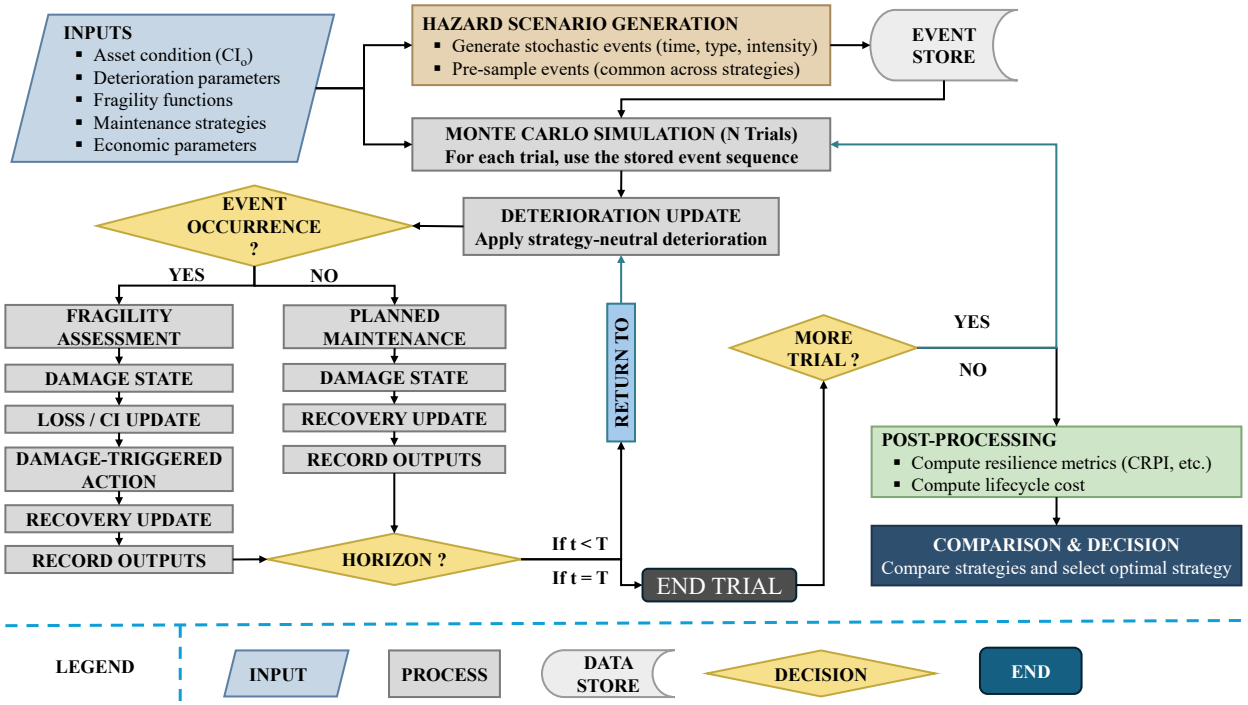


Figure 1: Methodology framework for lifecycle resilience assessment and maintenance strategy comparison, highlighting the integrated simulation process for hazard generation, condition deterioration, damage evaluation, recovery, and maintenance within a fixed time horizon, allowing for fair comparison of maintenance strategies for decision-making purposes.

event outcomes.

Deterioration model

Monthly evolution of asset condition is represented using an age- and condition-dependent deterioration process, with the Condition Index (CI) treated as a bounded state variable in the range $[0, 100]$. The formulation captures stage-wise ageing behaviour, age-related acceleration, and condition-sensitive degradation, while remaining computationally efficient for life-cycle Monte-Carlo simulation.

A smooth transition between early-life and late-life deterioration regimes is modelled using a logistic stage-weight function,

$$w(a) = \frac{1}{1 + \exp\left(-\frac{a - a_0}{s}\right)}, \tag{1}$$

where a denotes asset age (years), a_0 is the transition age, and s controls the smoothness of the transition.

The effective baseline annual deterioration rate is defined as

$$\bar{r}_{\text{base}} = r_{\text{base}}\big[(1 - w(a)) + w(a) \cdot \text{PM}\big], \tag{2}$$

where r_{base} is the nominal baseline rate and PM scales late-stage deterioration intensity. Age- and condition-dependent acceleration is incorporated through

$$r_{\text{ann}}(a, \text{CI}) = \bar{r}_{\text{base}}\,(1 + \gamma a)\big(1 + \eta(1 - \text{CI}/100)\big), \tag{3}$$

with γ governing age-related acceleration and η controlling sensitivity to declining condition.

For numerical implementation, the annual rate is converted to a monthly rate,

$$r_{\text{m}} = \frac{r_{\text{ann}}}{12}. \tag{4}$$

CI evolution over a time step Δ_t is computed using an exponential decay update,

$$\text{CI}_{t+\Delta} = \text{CI}_t \exp(-r_{\text{m}}\Delta_t), \tag{5}$$

which preserves positivity and ensures numerical stability over long simulation horizons.

This expression describes transitions from early to late ageing, rapid degradation at low CI, and continuous degradation dynamics as described in established deterioration modelling practice (McKibbins et al. 2019). The deterioration process is modelled as strategy-neutral, with maintenance activities affecting CI by discrete gains or recovery dynamics, but not affecting the deterioration physics.

Fragility and damage-state selection

For each hazard intensity I and DS threshold index j the exceedance (upper tail) probability is modeled by a log-normal CDF:

$$P(\text{DS} \geq j \mid I, \text{CI}) = \Phi\left(\frac{\ln I - \ln \mu_j^*(\text{CI})}{\beta_j}\right), \tag{6}$$

where $\Phi(\cdot)$ is the standard normal CDF, β_j is dispersion and $\mu_j^*(\text{CI})$ is the CI-adjusted median:

$$\mu_j^*(\text{CI}) = \mu_j \cdot \left[1 + \phi\frac{\text{CI} - 50}{50}\right] \tag{7}$$

Incremental probabilities for No Damage, DS1, DS2, ... are obtained from successive differences of (6) as customary Porter (2015), Committee et al. (2013). A deterministic DS label is selected by comparing the incremental CDF to the pre-sampled u (same u used across strategies for that event) — this enforces paired comparison across strategies.

Loss mapping and immediate post-event CI

Event-induced damage states (DS) are translated into fractional condition losses L_{DS} using mean loss values (or bounded ranges where uncertainty is characterised). The immediate post-event condition is computed as

$$\mathrm{CI}_{\mathrm{post}} = \mathrm{CI}_{\mathrm{pre}}\big(1 - L_{\mathrm{DS}}\big). \tag{8}$$

By ensuring that the logic of intervention actions determines the differences in condition states after events for various strategies, this deterministic transformation eliminates the impact of maintenance and recovery actions.

Recovery and maintenance models

Recovery toward a target CI is modelled by an exponential approach:

$$\mathrm{CI}(t) = \mathrm{CI}_{\mathrm{post}} + \big(\mathrm{CI}_{\mathrm{target}} - \mathrm{CI}_{\mathrm{post}}\big)\big(1 - e^{-kt}\big), \tag{9}$$

where k is a recovery rate parameter calibrated by resourcefulness and strategy ($k = k_0 \cdot \alpha_{\mathrm{strat}} \cdot \alpha_{\mathrm{res}}$). In the implementation:

- **Preventive**: scheduled periodic jumps (no change to physics): $\mathrm{CI} \leftarrow \min(100, \mathrm{CI} + G)$ at interval.
- **Corrective**: post-event patch scheduled after a resource dependent delay; gain is fraction of event loss subject to a patch cap.
- **Rehabilitation**: triggered when CI falls below a threshold; scheduled to reach an explicit $\mathrm{CI}_{\mathrm{target}}$.

Resilience metrics (CRPI and event metrics)

Normalized performance is $p(t) = \mathrm{CI}(t)/100$. Life-cycle CRPI ($\mathrm{CRPI}_{\mathrm{life}}$) is the time-average normalized area under the performance curve:

$$\mathrm{CRPI}_{\mathrm{life}} = \frac{1}{T}\int_0^T p(t)\,dt. \tag{10}$$

Event-window metrics are computed on the recovery curve following an event Poulin and Kane (2021):

- **Robustness** = $p_{\mathrm{post}}/p_{\mathrm{pre}}$.
- **Recoverability** = $p_\infty/p_{\mathrm{pre}}$ (final steady value relative to pre-event).
- **Rapidity** = $1/\tau_{95}$, where τ_{95} is the time to reach 95% of pre-event performance.
- **$\mathrm{CRPI}_{\mathrm{event}}$** = normalized AUC of the recovery curve over the event window.

Monte-Carlo design and reproducibility

The Monte-Carlo simulations are performed using either fixed or random seeds. For each simulation run, the hazard events are pre-sampled, and a uniform random variable $u \in [0, 1)$ is generated for each event. The same realization of the hazard process and the corresponding sequence of u values are used for all maintenance strategies, such that the differences between all pairs of outcomes are only due to the maintenance strategies.

The cascading effects are modeled in two ways: (i) ordered scheduling of follow-on hazards within a specified vulnerability window, and (ii) probabilistic hazard intensity amplification if the asset is still in the vulnerable post-event state. The simulation outputs for each simulation run include time series of CI, event sequences, maintenance schedules, cost data, and aggregated resilience values, which are exported in structured CSV format.

Implementation details

The simulation engine is implemented with a monthly time step, deterministic damage state (DS) choice, and explicit scheduling of maintenance events. Repair and rehabilitation events are applied when the month of their scheduled execution is reached, allowing for mobilization delays and event timing to be handled consistently across simulations.

Results from the simulation engine include pre- and post-event fragility plots (S-curves with DS indicators), resilience phase plots, resilience dividend summaries, and event-level CSV files summarizing CI trajectories, event histories, maintenance events, and costs. These results enable traceability, diagnostic analysis, and post-simulation processing of lifecycle and event-level metrics.

The simulation methodology is consistent with existing frameworks for resilience quantification and fragility modeling as reported in the literature (Bruneau et al. 2003, Cimellaro et al. 2010, Porter 2015, Committee et al. 2013). Degradation modeling is based on principles recommended in industry guidelines (Owen 2023), and Monte-Carlo simulation and metric estimation follow standard stochastic simulation procedures(Otárola et al. 2022). To ensure transparency and reproducibility, the complete implementation of the simulation framework is provided in a publicly accessible repository (Yadav 2026).

Assumptions and limitations

- Damage-state (DS) to loss mapping is done by using mean fractional losses to facilitate deterministic and repeatable paired comparisons of maintenance strategies. The variability of loss estimates is not sampled in the baseline simulations.
- The degradation process is considered strategy agnostic, where the baseline degradation physics are the same for all strategies, and the impact of maintenance is only felt after repair and recovery activities.
- The analysis is done based on resilience performance metrics or other related metrics. However, other economic valuation metrics such as resilience dividend, net present value, or life cycle cost are not included in the analysis for the current study. The consideration of these metrics, along with other consequence-based

factors such as traffic demand or network importance, is an important area for future research directions.

Results and discussion

This section presents numerical results obtained from the Monte-Carlo simulation framework and discusses their implications for maintenance-strategy selection. The number of simulations (5000 for each strategy) was chosen to ensure that there was convergence of the resilience metrics, as no significant changes were found in the results for a larger number of simulations.

Ensemble summary statistics

Table 1 summarizes ensemble lifecycle resilience metrics obtained from 5000 Monte–Carlo realizations per strategy, including mean and standard deviation of $\mathrm{CRPI}_{\mathrm{life}}$ along with robustness, recoverability, and rapidity. Figure 2 shows the corresponding $\mathrm{CRPI}_{\mathrm{life}}$ distributions. Preventive maintenance exhibits the highest median and the narrowest interquartile range, indicating both superior and stable lifecycle performance. The no-maintenance strategy shows the lowest median and a pronounced lower tail, reflecting frequent excursions into low-performance states. Corrective and rehabilitation strategies provide intermediate performance with larger dispersion than preventive maintenance. These ensemble results demonstrate that scheduled preventive actions improve average resilience while reducing variability across hazard realizations.

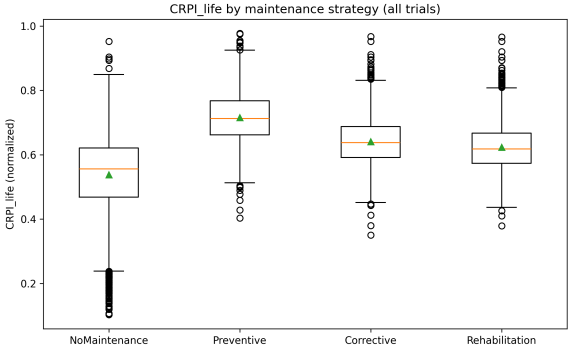


Figure 2: Distribution of lifecycle Composite Resilience Performance Index (CRPI_{life}) across 5000 Monte–Carlo trials for each maintenance strategy. Boxes indicate interquartile ranges, whiskers denote 1.5×IQR, and markers represent ensemble means.

Table 1: Ensemble lifecycle resilience metrics for maintenance strategies (5000 Monte–Carlo trials per strategy).

Strategy	Mean $\mathrm{CRPI}_{\mathrm{life}}$	Robustness	Recoverability	Rapidity
No Maintenance	0.5376	0.9474	0.9974	0.5499
Preventive	0.7158	0.9483	0.9974	0.4755
Corrective	0.6405	0.9474	0.9974	0.5499
Rehabilitation	0.6235	0.9474	0.9974	0.5417

To better understand how infrastructure behaves under each maintenance strategy, a representative trial is examined in detail. This trial run helps explain the ensemble results by showing condition changes, damage response, and recovery at the event level.

Event-level behaviour and fragility mapping

Fig. 3 shows prior vs post S-curves for a representative event. Key points:

- Pre-event condition (CI) shifts fragility medians, increasing probability of severe DS when CI is low.
- A deterministic per-event shock u is reused across strategies to isolate strategy effects.
- Post-event CI uses the mean-loss mapping $\mathrm{CI}_{\mathrm{post}} = \mathrm{CI}_{\mathrm{pre}}(1 - L_{\mathrm{DS}})$.

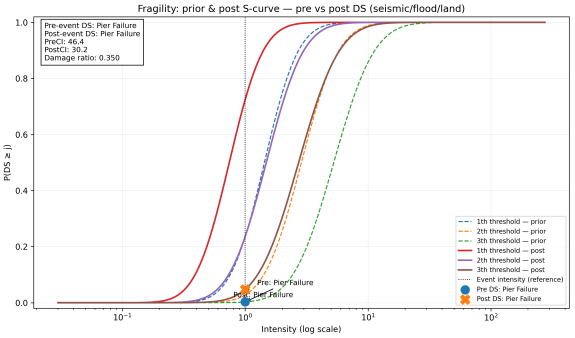


Figure 3: Condition-dependent fragility S-curves (prior and post) with realized damage states for a representative hazard, showing the effect of asset condition on damage exceedance probability.

Resilience phase diagrams (illustrative trial)

Figures (4–7) depict sample single-trial realizations to demonstrate the behavior of each maintenance strategy, whereas statistical results are based on all Monte Carlo simulation runs. Representative resilience phase diagrams illustrate the joint evolution of asset condition and performance for the illustrative trial. Monthly Condition Index (CI) trajectories are shown in the left subpanels, while the corresponding normalized performance used to compute $\mathrm{CRPI}_{\mathrm{life}}$ is shown in the right subpanels. Preventive maintenance produces frequent moderate CI restorations that stabilize baseline condition and sustain performance. Corrective maintenance results in event-triggered recovery with implementation delays and partial restoration, whereas rehabilitation generates large but infrequent CI resets governed by trigger thresholds. These trial-level dynamics provide mechanistic insight into the ensemble $\mathrm{CRPI}_{\mathrm{life}}$ trends reported in Table 1 and Figure 2.

Lifecycle dynamics and recovery mechanisms

Representative resilience phase diagrams (Figures 4–5) illustrate how ensemble trends arise from trial-level condition and performance trajectories. Under no maintenance, continuous deterioration exposes progressively weakened assets to successive hazards, leading to sustained performance decline. Preventive maintenance introduces frequent moderate CI restorations that stabilize baseline condition, thereby limiting damage severity and shortening recovery periods. Corrective maintenance relies on post-damage interventions with implementation delays, result-

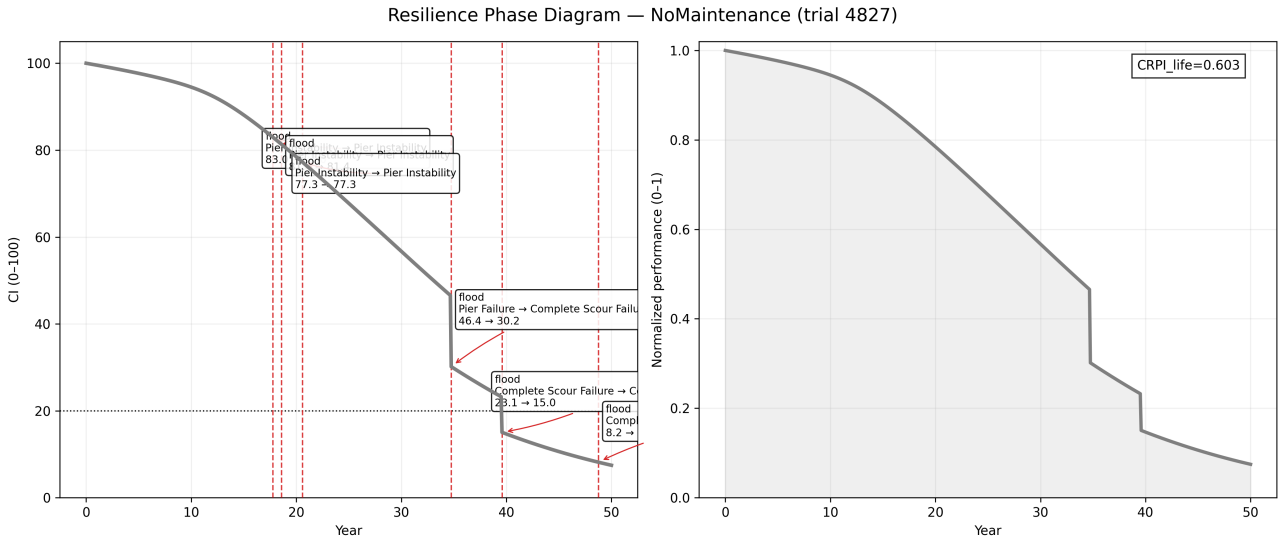


Figure 4: Representative Resilience Phase Diagram for No Maintenance Strategy, with monthly degradation of Condition Index shown in the left panel, and corresponding performance trajectory shown in the right panel.

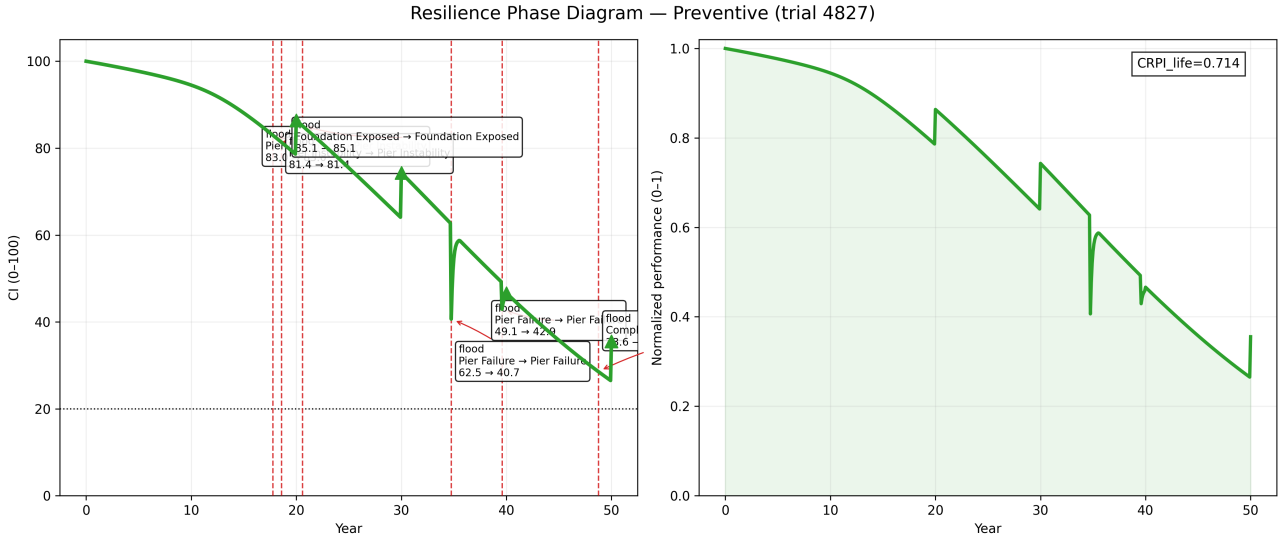


Figure 5: Representative resilience phase diagram for the preventive maintenance strategy (single trial), with the Condition Index on the left and the corresponding normalized performance used in the calculation of the CRPI on the right.

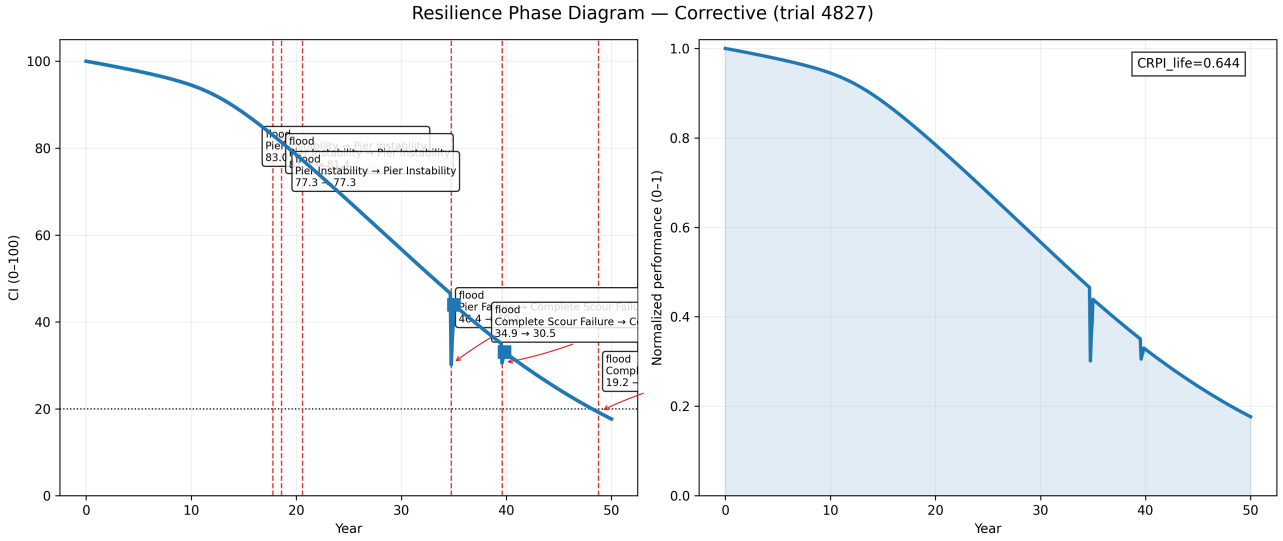


Figure 6: Representative resilience phase diagram for the corrective maintenance strategy (single trial): Condition Index (left) and normalized performance (right) used for the computation of the CRPI.

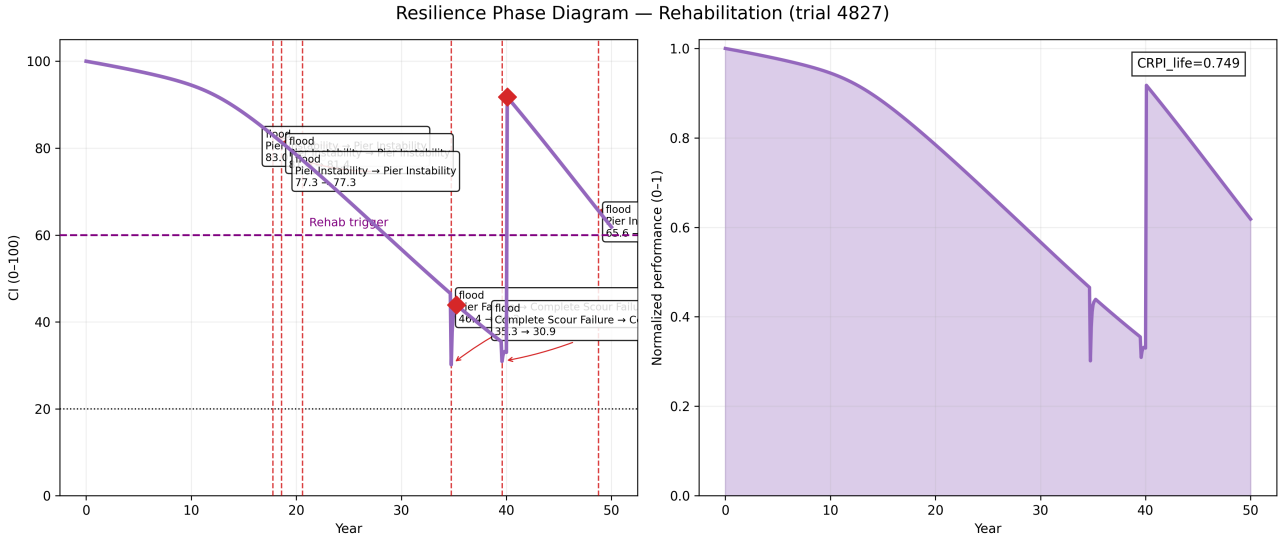


Figure 7: Representative Resilience Phase Diagram for Rehabilitation Strategy (single trial): Condition Index (left), and corresponding normalized performance for CRPI computation (right), highlighting threshold triggering and lifecycle restoration.

ing in partial and delayed recovery, while rehabilitation produces large but infrequent CI resets governed by trigger thresholds, allowing substantial degradation between interventions.

These contrasting recovery behaviors explain the ensemble outcomes summarized in Table 1 and visualized in Figure 2. Preventive maintenance achieves the highest mean $\mathrm{CRPI}_{\mathrm{life}}$ (0.716) with the lowest dispersion, whereas no maintenance exhibits both the lowest mean value (0.538) and the largest variability. Corrective and rehabilitation strategies yield intermediate life-cycle performance.

Implications for resilience-informed maintenance

Combining Table 1, Figure 2, Figure 3 and the phase diagrams shows that life-cycle resilience in this study is driven primarily by pre-event asset condition. Preventive maintenance preserves higher CI prior to hazard exposure, shifts fragility toward lower damage probabilities, reduces immediate losses and time in low-performance states, and therefore yields higher mean $\mathrm{CRPI}_{\mathrm{life}}$ and lower ensemble variability. Corrective and rehabilitation strategies provide post-event recovery but are limited by delay and infrequency, producing lower mean CRPI and greater uncertainty. Although the framework provides quantification of the impact of maintenance strategies on resilience performance, the decision-making process is also dependent on the consequences of failure. Therefore, assets with higher traffic demand or network importance may need higher prioritization even when the resilience performance is similar. The resilience indicator CRPI is a component of the decision-making process and should be used in conjunction with consequence-based measures.

Conclusions

This study shows that asset maintenance during normal operating conditions is critical to improving infrastructure resilience. Resilience thinking must extend to the pre-event stage of the lifecycle: proactive condition preservation strengthens robustness, limits damage at hazard exposure, and accelerates recovery. By embedding maintenance within a probabilistic Monte–Carlo lifecycle framework that couples condition-index (CI) evolution with CI-conditioned fragility, the analysis places pre-event condition at the center of resilience outcomes.

Across strategies, preventive (scheduled) maintenance consistently delivers the highest lifecycle resilience by reducing time spent in high-vulnerability CI ranges. Corrective maintenance provides moderate benefits but is strongly controlled by repair delay and effectiveness, while rehabilitation yields large one-time CI recovery yet is highly sensitive to cost, timing and trigger rules. A key mechanism is the condition-dependent shift in fragility: lower pre-event CI substantially increases the probability of severe damage states and collapse, amplifying the value of proactive, condition-preserving interventions.

The quantitative results should be interpreted in light of the modelling assumptions used here (simplified loss/cost representation, deterministic event impacts, strategy-independent deterioration physics and a single-asset formulation). Extensions such as stochastic loss dispersion, Bayesian fragility updating from observed damage, resource-constrained optimization with discounting, and multi-asset modelling would increase decision relevance.

Overall, the results support prioritising preventive, pre-event maintenance as a core asset-management strategy to enhance resilience under increasing hazard uncertainty.

References

Applegate, C. J. and Tien, I. (2019). Framework for probabilistic vulnerability analysis of interdependent infrastructure systems. Journal of Computing in Civil Engi-

neering, 33(1):04018058.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., and von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. Earthquake Spectra, 19(4):733–752.

Cimellaro, G. P. (2016). Urban Resilience for Emergency Response and Recovery: Fundamental Concepts and Applications, volume 1. Springer, Cham, Switzerland.

Cimellaro, G. P., Reinhorn, A. M., and Bruneau, M. (2010). Framework for analytical quantification of disaster resilience. Engineering Structures, 32(11):3639–3649.

Committee, E., Whitman, R., Borcherdt, R., Brookshire, D., Eisner, R., Holmes, W., Olson, R., O'Rourke, M., Lagorio, H., Reitherman, R., and Solutions, S. (2013). HAZUS-MH 2.1 Earthquake Model Technical Manual.

Dong, Y. and Frangopol, D. M. (2016). Probabilistic time-dependent multi-hazard life-cycle assessment and resilience of bridges considering climate change. Journal of Performance of Constructed Facilities, 30(5):04016034.

Durán, O. and Vergara, B. (2022). Maintenance strategies definition based on systemic resilience assessment: A fuzzy approach. Mathematics, 10(10):1677.

Francis, R. A. and Bekera, B. (2014). A metric and framework for resilience analysis of engineered and infrastructure systems. Reliability Engineering & System Safety, 121:90–103.

Ganin, A. A., Kitsak, M., Marchese, D., Keisler, J. M., Seager, T., and Linkov, I. (2017). Resilience and efficiency in transportation networks. Science Advances, 3(12):e1701079.

Ghodoosi, F., Abu-Samra, S., Zeynalian, M., and Zayed, T. (2018). Maintenance cost optimization for bridge structures using system reliability analysis and genetic algorithms. Journal of Construction Engineering and Management, 144(2):04017116.

International Organization for Standardization (2014). Iso 55000:2014 asset management — overview, principles and terminology. Geneva, Switzerland.

Izadi, A., Yazdandoost, F., and Ranjbar, R. (2020). Asset-based assessment of resiliency in water distribution networks. Water Resources Management, 34(4):1407–1422.

Li, C. Q. (2003). Life-cycle modeling of corrosion-affected concrete structures: Propagation. Journal of Structural Engineering, 129(6):753–761.

Li, R. and Gao, Y. (2022). On the component resilience importance measures for infrastructure systems. International Journal of Critical Infrastructure Protection, 36:100481.

Liu, X., Li, D.-Q., Cao, Z.-J., and Wang, Y. (2020). Adaptive monte carlo simulation method for system reliability analysis of slope stability based on limit equilibrium methods. Engineering Geology, 264:105384.

McKibbins, L., Spink, T., and Power, C. (2019). Deterioration modelling of civil engineering infrastructure assets.

Otárola, K., Fayaz, J., and Galasso, C. (2022). Fragility and vulnerability analysis of deteriorating ordinary bridges using simulated ground-motion sequences. Earthquake Engineering & Structural Dynamics, 51(13):3215–3240.

Ouyang, M. (2014). Review on modeling and simulation of interdependent critical infrastructure systems. Reliability Engineering & System Safety, 121:43–60.

Owen, A. B. (2023). Practical Quasi-Monte Carlo Integration.

Pant, R., Barker, K., and Ramirez-Marquez, J. E. (2019). Interdependent infrastructure resilience: A review and implications for transportation systems. Journal of Infrastructure Systems, 25(2):04019002.

Porter, K. (2015). Beginner's Guide to Fragility, Vulnerability, and Risk, pages 235–260. Springer Berlin Heidelberg, Berlin, Heidelberg.

Poulin, C. and Kane, M. B. (2021). Infrastructure resilience curves: Performance measures and summary metrics. Reliability Engineering & System Safety, 216:107926.

Salem, S., Siam, A., El-Dakhakhni, W., and Tait, M. (2020). Probabilistic resilience-guided infrastructure risk management. Journal of Management in Engineering, 36(6):04020073.

United Nations Office for Disaster Risk Reduction (UNDRR) (2022). Guidelines on resilient infrastructure.

World Bank (2021). Building back better: Achieving resilience through stronger, faster, and more inclusive post-disaster reconstruction.

Yadav, S. (2026). Quantifying the role of asset maintenance for resilient infrastructure. `https://github.com/Shyamratnayadav/resilience-maintenance-model`.

Yang, D. Y. and Frangopol, D. M. (2019). Life-cycle management of deteriorating civil infrastructure considering resilience to lifetime hazards: A general approach based on renewal-reward processes. Reliability Engineering & System Safety, 183:197–212.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATED BUILDING CLASSIFICATION FOR SCALABLE URBAN MATERIAL STOCK ASSESSMENT

Elisa Iliste, Ingrid Viskus, Ergo Pikas and Targo Kalamees
Tallinn University of Technology, Tallinn, Estonia

Abstract

Soviet-era apartment blocks in Eastern Europe are being demolished, generating mixed waste and limiting circular reuse. Existing assessments rely on generic archetypes and lack element-level material detail. This study develops an automated algorithm to detect standardised Soviet building series for scalable material stock analysis. A distance-based classification using apartment layouts, floor area, height, and staircases was validated on 337 buildings and applied to Estonia's building stock. The model achieved 70% series-level accuracy, with dominant series represent 61% of the stock. This element-level quantification enables targeted deconstruction and higher material recovery for circular economy planning.

Introduction

Across Eastern Europe, a substantial share of the housing stock consists of Soviet-era buildings constructed during periods of mass housing development. Built under conditions prioritising speed and quantity over long-term durability (Meuser and Zadorin, 2015), now exhibiting declining living conditions, deteriorating structural integrity, and architectural obsolescence.

The Soviet-era housing stock in Estonia is characterised by residential districts dominated by standardised pre-fabricated concrete apartment buildings. These structures, ranging primarily from 3- 5 to 9 and 16-storey tower blocks, were assembled using a limited number of industrial mass-produced series (e.g 1-317, 1-464, 111-121, etc.). Visually, these districts present a repetitive landscape of rectangular blocks grouped around semi-open courtyards or distributed freely to ensure better daylight conditions (Glendinning, 2019).



Figure 1: Example of a typical 5-storey 111-121 series building

Many of the soviet era buildings are facing abandonment; they no longer function as viable housing, both socially and economically, leading to deterioration and inevitable demolition (Eurofound, 2023). Traditional demolition practices generate large quantities of mixed construction and demolition waste, much of which is directed to landfilling or low-value processing, reinforcing a linear "produce-use-discard" economic model (Ilomets *et al.*, 2022). However, research highlights that demolition waste contains substantial amounts of materials—such as concrete, masonry, and timber—that have high reuse or recycling potential if handled appropriately (Sakthibala *et al.*, 2025). Selective deconstruction, as opposed to conventional demolition, enables materials to be separated at source, significantly improving their quality and recyclability while retaining economic value (Ilomets *et al.*, 2022). Timeliness is critical: early intervention allows material recovery before deterioration, maximising both environmental and economic benefits.

Despite this circular economic opportunity, current approaches to assessing urban material stocks face significant limitations. Research on material potential in existing buildings is constrained by data availability (Romero Perez de Tudela *et al.*, 2020; Tirado *et al.*, 2021). Urban analytics typically rely on building archetypes based on usage type and construction year only, without differentiating material composition, many rely only of building volume or area (Ali *et al.*, 2021; Oliveira *et al.*, 2024). In Urban Energy Modelling, for example, LOD1 or LOD2 models are used to calculate building volumes (Xu *et al.*, 2023), making them appropriate for envelope material quantities but failing to represent building interiors. From machine-learning applications (Markopoulou *et al.*, 2024) used image-based segmentation with cadastral data, and (Ghione *et al.*, 2022) used CNN-based building classification for Oslo's building stock. The main limitation of these image-based solutions is the computational efficiency.

Soviet-era architecture offers a unique opportunity: its centralised design templates and documented material specifications make it an ideal candidate for systematic detection and material quantification. Many of these buildings are remotely located, built for factory workers, and thus disproportionately affected by population decline (Glendinning, 2019). While collections of original design drawings exist (Meuser and Zadorin, 2015). Much of the documentation is largely available only in Russian,

limiting accessibility. Recent advances in AI-powered translation and computer vision now make these previously unviable data sources accessible for large-scale analysis.

If building series—can be reliably detected from readily available building data, it becomes possible to estimate material stocks at urban and national scales without costly individual building surveys. This is essential for circular economy planning, enabling municipalities to identify buildings with high reuse potential, prioritise selective deconstruction over conventional demolition, and plan material recovery operations before building deterioration advances.

This research addresses this opportunity by developing an automated detection algorithm for building series using open building registry data. The design science research methodology is employed to develop and evaluate the detection algorithm and demonstrate its application for material stock quantification. The algorithm is designed to operate with minimal input parameters, while maintaining sufficient accuracy for practical decision-making. To operationalise the main aim, the following research questions are posed:

RQ1: Can building series be automatically detected from open building registry data, and what classification accuracy can be achieved at the building stock level?

RQ2: Does this building series classification provide sufficient granularity to support circular economy planning and material stock quantification at element level?

This research makes three key contributions. First, it presents a novel distance-based classification approach for detecting building series using only four parameters (apartment configuration, floor areas, building height, and number of staircases). Second, it validates the method on Estonia's complete apartment building stock (13,200+ buildings), demonstrating up to 80% accuracy. Third, it demonstrates element-level material quantification for the 111-121 building series, revealing approximately 4.86 million m³ of concrete and 0.23 million tons of steel.

Methods

The design research methodology was employed to develop the building series and type detection algorithm. The development process comprised the following steps:

- Development criteria were defined
- Input parameters were selected based on data availability, relevance, and the objective of minimizing the number of required variables to ensure scalability and the possibility of manual data entry.
- Algorithm logic and calculation procedures were developed, including classification rules and parameter threshold definitions.
- The algorithm was implemented, tested and applied to the full Estonian apartment building dataset
- Validation was conducted using a smaller reference dataset to assess classification accuracy.
- Material quantities were calculated for one representative building series to demonstrate the applicability of the developed approach.

Data sets

A dataset of 437 buildings with expert-assigned series and types was utilised as ground truth. Out of those buildings, 65% had information about both the series and the extension. The other 35% of buildings were untypical, meaning they didn't have a series assigned to them. This set was divided into two. First, a set of 337 apartment buildings was used to validate the workflow itself. Another 100 buildings were used for additional validation for the national-level mapping.

Additionally, a dataset containing all Estonian apartment buildings was utilised for the nation-level mapping. Before applying the workflow, the dataset was pre-processed. Wooden apartment buildings (approximately 5,700) were excluded because the template library does not include wooden buildings. Buildings constructed before 1945 (approximately 1,300 buildings) were also excluded, as the template set is limited to post-war standardized designs, corresponding to the established classification of Estonian apartment building archetypes. After this filtering 13,200 buildings remained.

Data Sources

The Estonian Building Register (EBR) is a public database that contains information on all Estonian buildings. It includes various types of data, such as the year of first use, purpose of use, dimensions (footprint area, heated area, and building height), apartment-specific data (apartment number, floor of entry, area, number of rooms) and building technical data that provide information on building envelope materials. It has an open data policy and web services for data acquisition.

A collection containing general information on more than 100 Soviet-era building designs was used to develop a database of templates serving as benchmarks for the distance-based classification. The level of detail and machine-readable data varied across the designs. Only those designs containing the parameters required for the detection algorithm were included in the final template database.

Evaluation Methods

Recall was used to measure the proportion of correctly identified buildings for each series and is defined as:

$$\text{Recall}_s = \frac{TP_s}{TP_s + FN_s} \quad (1)$$

where TP_s is the number of true positives for series s, and FN_s is the number of false negatives (i.e. buildings belonging to series s that were assigned to a different series). High recall indicates that the method is effective at detecting buildings of a given series when they are present.

The overall classification error rate is defined as the proportion of incorrectly classified buildings:

$$\text{Error rate} = \frac{N_{\text{incorrect}}}{N_{\text{total}}} \quad (2)$$

where $N_{\text{incorrect}}$ is the number of buildings whose predicted series does not match the validated series, and N_{total} is the total number of evaluated buildings. In addition, a filtered error rate was computed by excluding buildings that do

not correspond to any known template ("untypical" buildings), providing an estimate of performance for standardised building types only.

The false-positive rate for a given series measures the extent to which buildings from other series are incorrectly assigned to it:

$$\text{FP rate}_s = \frac{FP_s}{TP_s + FP_s} \quad (3)$$

where FP_s is the number of buildings incorrectly assigned to series s. High false-positive rates indicate that a series acts as an attractor for structurally similar but distinct building series.

Building Series and Type Representation

General data from 75 designs across 11 series were compiled into design templates to serve as benchmarks for distance calculations. Templates were implemented as Python objects using a set of data classes. Each template encodes one building design and is described by six main attributes, sampled in Table 1.

First, the series and series extension (type) identify the building family and the specific design variant (e.g., 1-464-1KE). Together, they form a unique series identifier.

Second, the series type indicates whether the template represents a whole building (WHOLE) or a modular staircase or block (MODULE). The whole series corresponds to complete buildings with multiple staircases, whereas the modular series describes a single repeatable building staircase.

Third, the spatial pattern describes the apartment configuration on a typical floor. This is represented as a nested list. For whole buildings, each inner list corresponds to one staircase and contains the number of apartments in each segment of that staircase. For modular buildings, the pattern consists of a single list representing one module.

Fourth, the area pattern stores the floor area associated with each staircase. Instead of storing the area of individual apartments, the total area per staircase is used.

Fifth, the number of floors is represented as a list. For buildings with a fixed height, this list contains a single value. For building series occurring at multiple heights, the list includes all typical numbers of floors.

Finally, all templates are stored in a central template library that serves as a registry and allows filtering by series type or other attributes.

This representation provides a machine-readable encoding of architectural typologies, while preserving the hierarchical structure of real buildings.

Table 1: Template instance examples, Value A represents a modular type, Value B represents the whole building type

Attribute	Value A	Value B
Series	'111-121'	'1-464'
Extension	'03E'	'17'
Type	Module	Whole
Pattern	[[4, 4, 2, 3]]	[[2,1,3], [2,2,2], [2,2,2], [3,1,2]]
Area Pattern	[255.0]	[136, 136, 136, 136]
Floor nr	[9]	[4]

Developed Workflow

The developed building series detection workflow consists of three main phases: (1) Data Augmentation and Preprocessing; (2) Distance calculation; (3) Classification and Scoring. The workflow is illustrated on Figure 2.

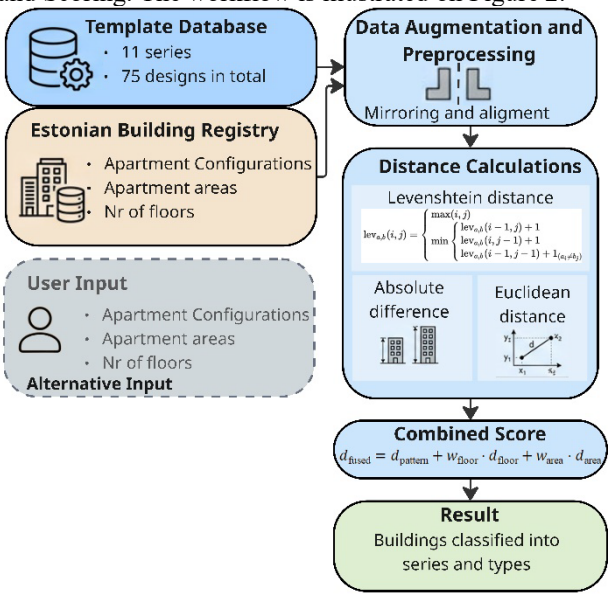


Figure 2: Series detection workflow

Three principal design criteria guided the development of the framework:

- Machine-readable and scalable representation of the building series, the data must be easily expandable with new information
- Robustness to structural variation and data inconsistencies:

Data Augmentation and Pre-Processing

To account for symmetric building layouts and staircase ordering inconsistencies, the matching procedure evaluates both the original and mirrored configurations of each building. Mirroring is implemented by reversing the order of staircases in the building representation. This accounts for cases where the same building design may be recorded with opposite orientations in the source data. For each building, two configurations are therefore considered: the original staircase sequence and its mirrored counterpart. For each configuration, the full matching workflow is executed independently.

Modular building series have to be treated differently from whole-building series. While whole-building templates represent complete buildings with a fixed configuration, modular templates represent configurable staircase units that may appear multiple times within a building and be mixed.

For modular series, templates are first grouped by their base series code. Each staircase of the subject building is then independently matched against all modular templates within the corresponding series. The best matching module for each staircase is selected using the minimum Levenshtein distance, considering both the original and mirrored module patterns.

The total distance for a modular series is computed as the sum of the per-staircase distances:

$$d_{\text{module}} = \sum_{i=1}^{n} \min\left(dL(s_i, m_j)\right) \quad (4)$$

where $s_i$is the apartment pattern of a staircase $i$in the building, and $m_j$are the available module patterns for the series. For each modular series, only the configuration with the smallest distance is selected for further calculations to minimise computational cost.

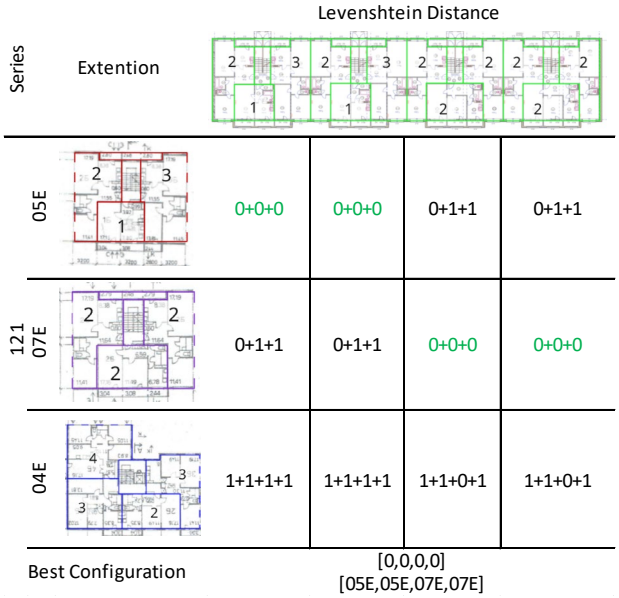


Figure 3: Example of finding the best modular configuration over a series using the Levenshtein distance

Distance Calculations

Apartment Pattern distance

Pattern similarity is computed using the Levenshtein (edit) distance between the subject building's staircase sequences and the template sequences. This metric was chosen because the objective is to identify the most similar structural configuration rather than to measure numerical proximity. The staircase sequences represent categorical configurations rather than continuous numerical variables; therefore, computing absolute or Euclidean differences would not meaningfully capture the similarity. The Levenshtein distance is calculated as shown in Equation 5.

$$\text{lev}_{a,b}(i,j) = \begin{cases} \max(i,j) & \text{if } \min(i,j) = 0, \\ \min\begin{cases} \text{lev}_{a,b}(i-1,j) + 1 \\ \text{lev}_{a,b}(i,j-1) + 1 \\ \text{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} \end{cases} & \text{otherwise.} \end{cases} \quad (5)$$

The algorithm considers both original and mirrored building configurations.

Floor distance

The floor distance quantifies the difference between building floor count and the template specifications:

$$d_{\text{floor}} = \min_{f \in F} \; | \, n_{\text{floors}} - f \, | \quad (6)$$

where $f$is the set of allowed floor counts from the template and $\boldsymbol{n}_{\text{floors}}$is the actual building floor count.

Area Distance

Area similarity is calculated using the Euclidean distance, with each staircase acting as a dimension. For building staircase areas $\mathbf{A}_B = [a_1, a_2, \ldots, a_n]$and template areas $\mathbf{A}_T = [t_1, t_2, \ldots, t_n]$:

$$d_{\text{area}} = \sqrt{\sum_{i=1}^{n} \delta_i^2} \quad (7)$$

where the percentage difference for each valid staircase pair is:

$$\delta_i = \begin{cases} \min\left(\frac{| \, a_i - t_i \, |}{t_i}, 1.0\right) & \text{if } a_i > 0 \text{ and } t_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Fused Distance Score

All distances are combined using a weighted linear combination:

$$d_{fused} = d_{pattern} + w_{floor} \cdot d_{floor} + w_{area} \cdot d_{area} \quad (9)$$

where $w_{\text{floor}}$and w_{area} control the relative importance of floor and area constraints and allow to prioritize the pattern distance. In this paper, these were chosen as 0.5.

Validation of the workflow

337 buildings were used to assess the performance of the proposed workflow. The overall classification accuracy was 60% for series and 50% for templates. When excluding the generic unclassified category, accuracy increases to 80% for series and 70% for templates, indicating a higher performance for well-defined building types. To assess scalability, the runtime of the classification workflow was measured.

The violin plot shows that while the average distance for correct predictions is 1.2, the median distance is lower (0.4), indicating that the majority of correctly classified buildings have very small distances, with the mean being influenced by a limited number of high-distance outliers (mainly the unassigned buildings, which have an average distance of 3.5). The average and median distances for incorrectly assigned templates are 2.8 and 2.7, respectively.

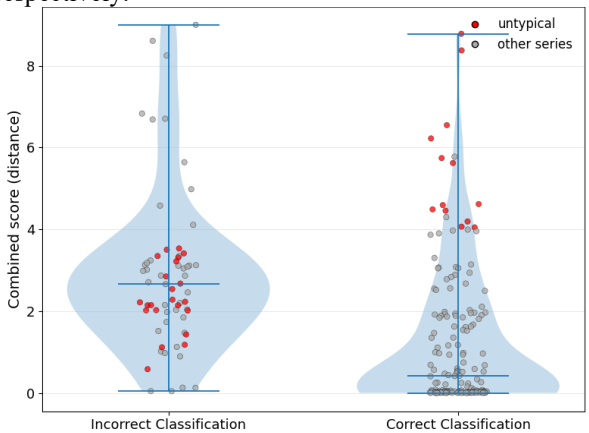


Figure 4: Distribution of the combined distance score by correctness, showing separation between correctly and incorrectly classified buildings.

To assess the performance of the proposed classification approach, we analysed the confusion matrix(Table 2) between the validated building series and the final predicted series. Columns represent the ground-truth (validated) series, while rows represent the model predictions.

The model demonstrates strong performance for several major building series. High, over 90% recall is observed for 121(100%),72057 (100%),10(93%),133(92%) and 1-464 (91%). These results suggest that the proposed features effectively capture the characteristics of the most well-defined building types.

In contrast, several series exhibit poor recall. Series 113-EKE, 66, 22 and 3F-2F all have a recall of 0. indicating that these categories are either underrepresented in the training data or insufficiently distinguishable from other series. The generic category also performs poorly, correctly classifying only 37 % of true instances.

The errors show consistent confusion between specific series. For example,

- 1Э-318 → 121 (36%), indicating that over one third of buildings from series 1Э-318 are systematically misclassified as 121.
- 113-EKE 10 → 23 (50%), suggesting strong similarity between these two series.

These results indicate families of series that are structurally similar and difficult to distinguish using a flat multi-class classification scheme.

Several series show a disproportionate number of false positives. Series 121 absorbs 36% of true 1Э-318, 11% of true Other, and 3% of true 1-317, despite having perfect recall. Similarly, the Other class accounts for 18% of true 1э-318, 10% of true 113-EKE 11, and 7% of true 133. Series 23 also generates notable false positives, including 50% of true 113-EKE 10 and 13% of true 113-EKE 11.

Nation-level mapping

The combined distance metric was computed for the 13,200 buildings. The full building stock (13,200 buildings) was processed in approximately seconds on a standard desktop machine. As a result of this classification, the most frequent series were 1-317 (14.26%), 10 (13.47%), 111-121 (13.00%), 72057 (11.29%), and 23 (9.45%). Together, these five series account for over 61% of all classified buildings.

The assigned building series was validated using a sample of 100 buildings. The overall classification error rate was 36%; excluding buildings that did not correspond to any known template reduced it to 28%. The recall remained high for 1-464 (98%); 1-317 also showed a high recall of 82%.

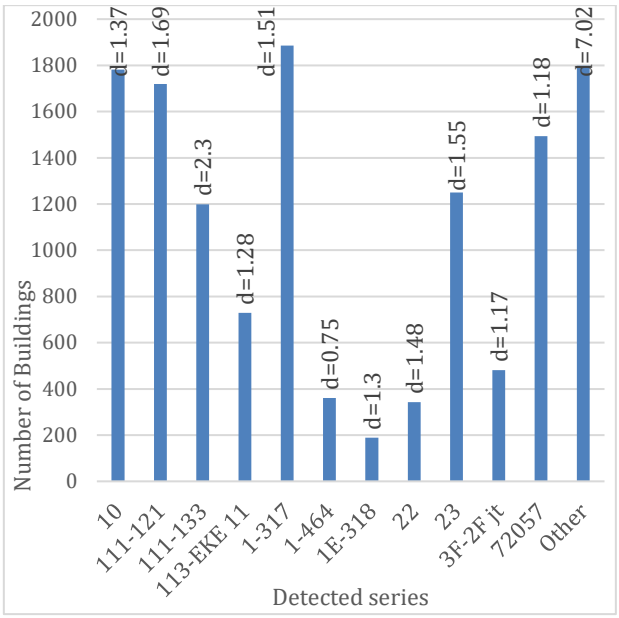


Figure 5: Building Stock Classification by Detected Series with Average Distance Score (d)

Material quantities calculation

To demonstrate the proposed workflow, a material stock calculation was performed for the 111-121 series. The series was selected for two reasons: firstly, it is one of the most prevalent building types in the national stock

Table 2: Confusion Matrix, in the column are the validated series, in the row the automatically classified series, green indicates a match and red the correct classification

Result / Validated	10	22	23	66	121	133	72057	113-EKE 10	113-EKE 11	113-EKE 12	1-317	1-464	1Э-318	Other	3F-2F
10	94%	50%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	6%	0%
22	0%	0%	6%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	33%
23	6%	0%	69%	0%	0%	0%	0%	50%	13%	0%	0%	8%	9%	3%	0%
121	0%	0%	0%	100%	100%	0%	0%	0%	0%	0%	3%	0%	36%	11%	0%
133	0%	50%	0%	0%	0%	93%	0%	0%	0%	100%	0%	0%	0%	11%	0%
72057	0%	0%	0%	0%	0%	0%	100%	0%	3%	0%	3%	0%	0%	14%	33%
113-EKE	0%	0%	17%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
113-EKE 11	0%	0%	0%	0%	0%	0%	0%	50%	71%	0%	0%	0%	0%	0%	0%
1-317	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	87%	0%	0%	11%	0%
1-464	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	92%	0%	0%	0%
1Э-318	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	36%	0%	0%
3F-2F	0%	0%	6%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%	6%	33%
Other	0%	0%	3%	0%	0%	7%	0%	0%	10%	0%	8%	0%	18%	37%	0%

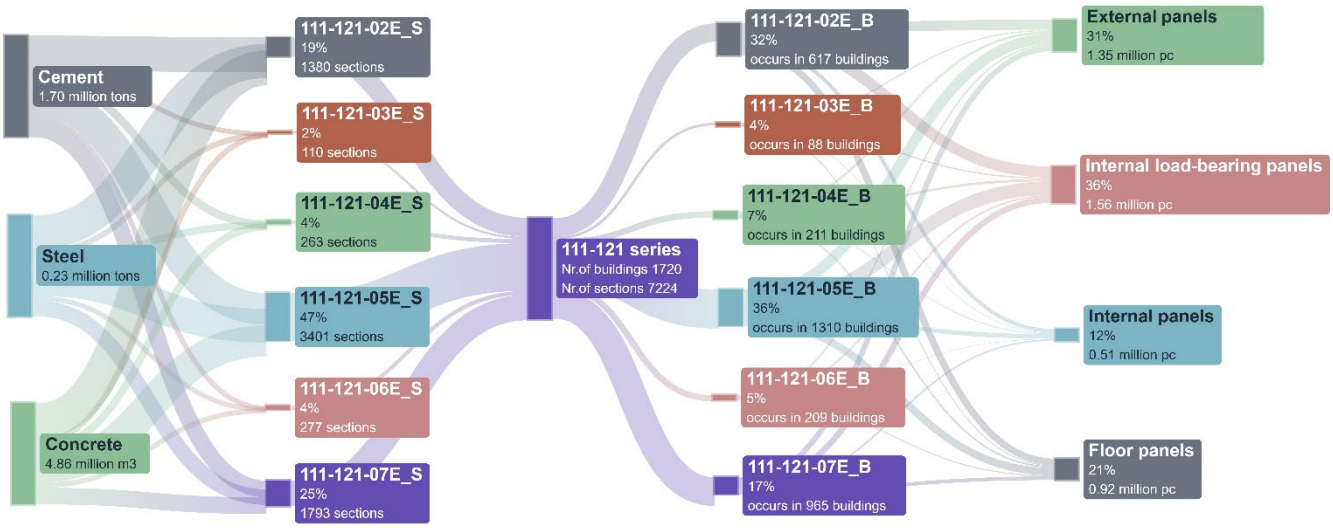


Figure 6: Material flow analysis for the 111-121 series: Central nodes represent building section templates, with flows splitting into two paths: left quantifies raw materials (concrete, steel, cement) for recycling estimates, and right classifies prefabricated components (floor and wall panels) for reuse planning.

(N=1720 detected buildings); secondly, the 111-121 series is modular, consisting of various section types assembled in different configurations. This modularity allows analysis in which material quantities are calculated for a specific section template rather than using generic building averages.

The analysis focused on the six most frequently detected types in the series, which collectively account for the vast majority of the 111-121 building stock.

Total material stock (M*total*) for the series was calculated by aggregating the specific material quantities of each identified section type:

$$M_{total} = \sum_{i=1}^{k} \left(N_i \times Q_{(i,m)} \right) \qquad (10)$$

Where:

Ni is the total count of detected sections of type *i* across the entire building stock (derived from the dataset's frequency distribution; e.g., type 111-121-05E appears 3401 times).

Qi,m is the quantity of a specific material *m* (e.g., reinforced concrete in *m*3, steel in tons) per single section of type *i*, extracted from the original design passports (reference project documentation).

k is the number of unique section templates identified (e.g., 02E, 05E, 07E, etc.)

As shown in Figure 6 The material flow analysis reveals distinct patterns. Although 5-storey building sections (Types 05E and 07E) are numerically dominant, accounting for over 70% of all sections, the 9-storey sections contribute disproportionately to the total material mass due to their larger building volume. The analysis indicates a total embodied stock for series 111-121 of approximately 4.86 million m3 of concrete, 1.70 tons of cement, and 0.23 million tons of steel for this series alone. Breaking this down by element type, the floor slabs (panels) account for the largest share at ~36%. This level of granularity is crucial for circular planning as it distinguishes between elements based on their material complexity and 'cleanness', which are the key factors in evaluating the reuse potential. Unlike the external wall panels, which in Soviet-era construction often consist of composite layers that are difficult to separate and whose insulating layers are often problematic to reuse, floor and internal load-bearing panels are typically monolithic reinforced concrete elements. That according to (Ilomets *et al.*, 2022) This structural homogeneity gives floor panels a higher potential for direct reuse, with an estimated recovery of up to 90% for floor slabs in good condition. He also confirms that internal load-bearing elements are one of the most valuable components, as they are generally well-preserved and clean.

Discussion

The results demonstrate that building series can be systematically identified from open building registry data with meaningful accuracy. For the primary validation dataset (337 buildings), the overall classification accuracy was 60% for series and 50% for templates. When excluding buildings that do not correspond to any known template (generic/unclassified category), the accuracy increases to 80% for series and 70% for templates. At the national level, validation yielded an overall accuracy of 64%, which rose to 72% when excluding unclassified buildings.

This performance is comparable to machine learning approaches reported in the literature, such as the 87% accuracy reported by Markopoulou et al. (2024). However, the distance-based approach offers distinct advantages in interpretability, minimal data requirements, and computational efficiency. Previous material stock studies have relied on statistical averages, applying material intensity coefficients (*kg*/*m*2 *or kg*/*m*3) to floor area or building volume (Oliveira *et al.*, 2024). While effective for large-scale estimates, these methods fail to capture the structural heterogeneity within building series. For example, taller sections require more reinforcement steel and concrete per m2 than lower counterparts due to structural requirements. The ability to detect these patterns enables more accurate quantification of material

properties for engineering design rather than relying on statistical averages.

Although distance-based similarity metrics perform well across several building series, the results suggest that additional descriptive dimensions are needed to improve robustness, particularly for identifying atypical buildings that do not conform to any standardised series. Such cases are inherently difficult to capture using purely rule-based or metric-based approaches.

Machine learning, therefore, constitutes a promising direction for future work. While the current number of templates is too large to support a classification model with the available data, machine learning could be employed in a hybrid framework. For instance, a predictive model could assist in ranking candidate series rather than series and extensions, resolve ambiguities when multiple templates exhibit similar distance values, and flag buildings that deviate substantially from known typologies.

The implications of the results extend beyond quantifying material stock. Building series detection enables the systematic application of any series-specific data at the urban scale. This includes thermal properties and spatial data for energy modelling, as well as structural characteristics and construction details for renovation planning.

While the proposed method is developed and demonstrated on standardised Soviet-era building typologies, the underlying framework is not unique to this context. Similar industrialised and prefabricated housing approaches were widely implemented across Europe during the post-war period. These systems share key characteristics with Soviet building series, including modular design, limited layout variation, and predefined structural and spatial configurations.

Comparable patterns can also be observed in other European contexts, such as the Swedish Million Programme (Berggren and Wall, 2019), where large volumes of residential buildings were constructed using standardised designs. Although the specific notion of “series” may not be explicitly defined in all these contexts, the underlying logic of repetition is consistent.

Several limitations remain. First, data availability and quality significantly affect applicability and scalability. While Estonia's comprehensive building registry enabled national-scale analysis, many Eastern European countries lack a comparable open data infrastructure.

Second, parameter sensitivity poses a critical challenge. Building registries often contain inconsistencies and data quality issues that directly affect large-scale assessments that rely on these sources. Errors in registry data propagate through the classification workflow, leading to potential misclassification of building series. Consequently, the classification accuracy is highly sensitive to the quality of input parameters. Small deviations in recorded apartment configurations, floor areas, or building heights can lead to incorrect template assignments. The confusion matrix revealed systematic misclassifications between specific series pairs (e.g., 1Э-318 → 121 at 36%, 113-EKE 10 → 23 at 50%), indicating that the four-parameter representation may be insufficient to distinguish closely related designs.

Third, template database coverage constrains comprehensiveness. The 75 designs represent major series but cannot capture the full diversity of Soviet-era construction. The "generic" category had only 37% recall, serving as a catch-all for buildings that did not match known templates. Expanding the database requires access to original design documentation.

Fourth, building modifications and renovations significantly impact classification accuracy. Soviet-era buildings have undergone various interventions during their 40-70-year lifespans: apartment reconfigurations that alter spatial patterns, balcony enclosures that change floor area, and roof additions that modify building height. The building registry may not consistently reflect these changes, leading to mismatches between recorded parameters and actual characteristics.

Despite these limitations, the method provides a practical foundation for circular economy planning where detailed building-level surveys are infeasible. The element-level material quantification demonstrated for the 111-121 series offers actionable information for selective deconstruction planning. Future research should expand the validation dataset, focus on hybrid classification frameworks that combine distance metrics with machine learning, conduct validation studies in other Eastern European countries, and integrate building condition assessment to account for deterioration and modifications.

Conclusions

This paper proposes an automated series of building detection for urban material stock assessment, focusing on Soviet-era standardised buildings in Estonia.

The results show that building series can be automatically detected from open building registry data with meaningful accuracy. Using four parameters: apartment configuration, floor areas, building height, and number of staircases, the distance-based algorithm achieved 60% overall accuracy for series and 50% for templates across 337 validation buildings. Excluding the generic unclassified category, accuracy increased to 80% for series and 70% for templates. National-scale classification of 13,200 buildings yielded a similar accuracy of 64% for all series, increasing to 72% when excluding non-standardised buildings.

Building series classification provides sufficient granularity for circular economy planning and element-level material stock quantification. The 111-121 series (1,720 buildings) revealed approximately 4.86 million m³ of concrete, 0.23 million tons of steel, and 1.70 million tons of cement. Floor slabs account for 36% of total material mass. This granularity is essential because it distinguishes elements by reuse potential—monolithic floor panels offer up to 90% recovery, compared with composite external wall panels with problematic insulation layers.

This work provides three key contributions: (1) Novel distance-based classification with minimal input parameters maintaining practical accuracy; (2) Validation

on Estonia's complete apartment building stock, demonstrating national-scale scalability; (3) Element-level material quantification, distinguishing components by reuse potential.

Acknowledgments

This work has been supported by the Estonian Research Council grants (PSG963 and PRG2732); by the European Commission through Horizon Europe projects SoftAcademy (EUI01- 168); and by the Estonian Centre of Excellence in Energy Efficiency, ENER (grant TK230) funded by the Estonian Ministry of Education and Research.

References

Ali, U., Shamsi, M.H., Hoare, C., Mangina, E. and O'Donnell, J. (2021), "Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis", *Energy and Buildings*, Vol. 246, p. 111073, doi: 10.1016/j.enbuild.2021.111073.

Berggren, B. and Wall, M. (2019), "Review of Constructions and Materials Used in Swedish Residential Buildings during the Post-War Peak of Production", *Buildings*, Multidisciplinary Digital Publishing Institute, Vol. 9 No. 4, doi: 10.3390/buildings9040099.

Eurofound. (2023), *Unaffordable and Inadequate Housing in Europe*, Publications Office of the European Union, Luxembourg.

Ghione, F., Mæland, S., Meslem, A. and Oye, V. (2022), "Building Stock Classification Using Machine Learning: A Case Study for Oslo, Norway", *Frontiers in Earth Science*, Frontiers, Vol. 10, doi: 10.3389/feart.2022.886145.

Glendinning, M. (2019), "Mass housing and extensive urbanism in the Baltic countries and Central/Eastern Europe: a comparative overview", *Housing Estates in the Baltic Countries: The Legacy of Central Planning in Estonia, Latvia and Lithuania*, Springer International Publishing Cham, pp. 117–136.

Ilomets, S., Kiviste, M., Tuisk, T., Paalandi, K., Kallavus, U., Hain, T., Pikas, E., *et al.* (2022), *Applied Study on Reuse and Recycling of Materials from Demolition of Vacant Apartment Buildings – Phase 1 Interim Report*, Ministry of Economic Affairs and Communications, Tallinn.

Markopoulou, A., Taut, O. and Shawqy, H. (2024), "Enhanced Databases on City's Building Material Stock. An Urban Mining Method Based on Machine Learning for Enabling Building's Materials Reuse Strategies", in Thomsen, M.R., Ratti, C. and Tamke, M. (Eds.), *Design for Rethinking Resources*, Springer International Publishing, Cham, pp. 685–701, doi: 10.1007/978-3-031-36554-6_44.

Meuser, P. and Zadorin, D. (2015), "Towards a typology of soviet mass housing", *Berlin: DOM. Alexander*.

Oliveira, J., Schreiber, D. and Jahno, V. (2024), "Circular Economy and Buildings as Material Banks in Mitigation of Environmental Impacts from Construction and Demolition Waste", *Sustainability*, Vol. 16, p. 5022, doi: 10.3390/su16125022.

Romero Perez de Tudela, A., Rose, C.M. and Stegemann, J.A. (2020), "Quantification of material stocks in existing buildings using secondary data—A case study for timber in a London Borough", *Resources, Conservation & Recycling: X*, Vol. 5, p. 100027, doi: 10.1016/j.rcrx.2019.100027.

Sakthibala, R.K., Vasanthi, P., Hariharasudhan, C. and Partheeban, P. (2025), "A critical review on recycling and reuse of construction and demolition waste materials", *Cleaner Waste Systems*, Vol. 12, p. 100375, doi: 10.1016/j.clwas.2025.100375.

Tirado, R., Aublet, A., Laurenceau, S., Thorel, M., Louërat, M. and Habert, G. (2021), "Component-Based Model for Building Material Stock and Waste-Flow Characterization: A Case in the Île-de-France Region", *Sustainability*, Vol. 13 No. 23, doi: 10.3390/su132313159.

Xu, H., Kim, J.I., Zhang, L. and Chen, J. (2023), "LOD2 for energy simulation (LOD2ES) for CityGML: A novel level of details model for IFC-based building features extraction and energy simulation", *Journal of Building Engineering*, p. 107715, doi: 10.1016/j.jobe.2023.107715.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DIGITAL CLIMATE INDICATORS FOR RESILIENT DESIGN: EXAMINING COMPLETENESS ACROSS CARBON AND CLIMATE RISK ANALYTICS

Silvia Meschini[1], Paola Gasbarri[1], Lavinia Chiara Tagliabue[1], Stefano Rinaldi[2], Viviana Vaccaro[1], and Hira Ovais[1]
[1]University of Turin, Turin, Italy
[2]University of Brescia, Brescia, Italy

Abstract

Digitalisation expanded access to design, operational and environmental data, yet whole carbon-life assessment and climate risk prediction still evolve in separate, non-interoperable workflows. This fragmentation limits climate-responsive design, especially for urban flooding and heat. The paper develops a conceptual framework for digital climate indicators that link carbon performance with predictive flood and urban heat analytics. Using a structured review of recent work on BIM-integrated life cycle assessment, digital twins, machine-learning-based hazard modelling and resilience assessment, it derives key information requirements and illustrates their application. The contribution is a set of indicator families and research directions for robust, portable climate-resilience metrics.

Introduction

Climate-responsive design in the built environment is increasingly expected to achieve two intertwined objectives: reducing greenhouse-gas emissions and maintaining acceptable performance under changing climate hazards. The IPCC AR6 Synthesis Report confirms that anthropogenic climate change is already driving widespread impacts and escalating risks for cities and infrastructures (IPCC, 2023). At the same time, buildings and infrastructure are undergoing rapid digitalisation through Building Information Modelling (BIM), Geographic Information Systems (GIS), remote sensing, Internet of Things (IoT) sensing and, more recently, digital twins (DTs) that promise continuous monitoring and predictive analytics (Zhang et al., 2025; Petri et al., 2025).

Despite this convergence, analytical workflows for mitigation and adaptation remain largely fragmented. Whole-Life Carbon (WLC) is typically assessed through life cycle assessment (LCA) and Environmental Product Declarations (EPDs), increasingly connected to BIM for automated quantity take-off (Zhang et al., 2025; Parece et al., 2025; Berges-Álvarez et al., 2024). Climate risk, by contrast, is usually analysed through hydrodynamic models, microclimate simulations, remote sensing and machine learning (ML) for hazard or susceptibility mapping (Islam et al., 2025; Ahmed et al., 2025). These workflows differ in assumptions, spatial and temporal resolution, data models and outputs, and they tend to meet, if at all, late in the process, via narrative reports rather than integrated indicators.

This separation has practical consequences. Design options optimised for WLC can inadvertently increase exposure to future flooding or overheating; conversely, adaptation measures that improve robustness may carry substantial embodied-carbon burdens that are not systematically quantified. Recent work on sustainability and resilience assessment explicitly calls for methods that link hazard, damage, recovery and environmental impacts into coherent frameworks (Sesana et al., 2024; Tanguay et al., 2024). However, there is still no shared concept of digital climate indicators that consistently connect carbon, risk and resilience analytics in digital workflows.

In this paper, a digital climate indicator is defined as a computable metric derived from integrated digital artefacts (e.g., BIM/IFC, GIS, hazard layers, scenarios) to support climate-responsive and resilient decisions.

Within the wide spectrum of climate hazards, this paper uses urban flooding and urban heat as reference, since they are high-impact in cities and supported by mature geospatial and ML ecosystems for indicator computation (Islam et al., 2025; Ahmed et al., 2025).

The aim of this paper is therefore to explore under what informational and methodological conditions indicators derived from digital artefacts, BIM/IFC models, GIS layers, sensor streams and climate-scenario data, can support climate-responsive and resilient design, with urban flooding and heat as reference hazards. The contribution is conceptual and methodological: we propose a framework for digital climate indicators, identify key information requirements, and outline research questions and directions for developing robust, interoperable metrics.

The remainder of the paper is structured as follows. Section 2 reviews recent work on BIM-integrated carbon assessment, DT-enabled dynamic LCA, ML-based flood and heat analytics, and resilience assessment methods. Section 3 introduces the indicator framework and associated information requirements. Section 4 formulates research questions and methodological directions, including a minimal operational example for flood and heat. Section 5 discusses implications and typical failure modes, and Section 6 concludes with a research agenda for digital climate indicators in the built environment.

Preliminary literature review

Digital carbon tools and BIM-integrated LCA

LCA has become the reference framework for quantifying environmental impacts of buildings and infrastructures over their life cycle and is increasingly used to support low-carbon design (Zhang et al., 2025). Recent reviews show that integrating LCA with BIM has advanced substantially in terms of automating inventory compilation. A systematic review of BIM-based LCA applications reports that BIM can automate a large fraction of life cycle inventory data extraction, reducing manual work and enabling early-stage comparisons (Parece et al., 2025). However, the same review shows that life cycle impact assessment still often requires manual mapping between BIM elements and LCA or EPD datasets, and that only a minority of studies embed results into optimisation or multi-criteria decision-support processes (Parece et al., 2025).

A complementary body of work extends the focus from environmental impacts to life cycle sustainability assessment, integrating economic and social dimensions (Berges-Álvarez et al., 2024). These studies repeatedly highlight issues of inconsistent semantics, missing metadata and limited interoperability between BIM models, classification systems and LCA datasets (Zhang et al., 2025; Berges-Álvarez et al., 2024). Even when computational pipelines are sophisticated, these informational gaps can undermine the reliability and comparability of carbon indicators. In practice, this can lead to a subtle failure mode: numerical precision without semantic robustness, where whole-life carbon values appear accurate but are conditioned on incomplete or project-specific assumptions about materials and systems. Overall, the literature suggests that digital carbon tools are moving from bespoke analyses toward more standardised workflows, but that their outputs remain sensitive to information quality and scope. For digital climate indicators, this implies that indicator design must explicitly consider semantic completeness and provenance, rather than assuming that any BIM-based LCA result is a stable quantity.

DTs and dynamic environmental indicators

DTs in the built environment are generally described as digital representations of assets or systems that are continuously updated with data from sensors, simulations and operational processes (Petri et al., 2025; Luo et al., 2024). In the last few years, DTs have been proposed as platforms for moving from static sustainability metrics to dynamic, data-driven indicators. The Building Life-cycle Digital Twin (BLDT) framework, for example, combines IoT measurements, ML models and interoperable data structures to generate time-dependent LCA outputs, enabling near real-time assessment of environmental performance and supporting predictive interventions (Petri et al., 2025). More conceptual work on DT-enabled dynamic LCA underlines the importance of embedding LCA models into broader systems of objectives, data flows and stakeholder needs, arguing that a DT can host multiple assessment lenses over time rather than producing one-off results (Luo et al., 2024).

These developments highlight an important opportunity for digital climate indicators: moving from single-snapshot values to time-series indicators that reflect changing operating conditions, retrofits and climate scenarios. At the same time, they expose new challenges. If dynamic indicators are generated without clear documentation of data provenance, model calibration, drift detection and uncertainty, DTs can become "black boxes" that amplify opacity rather than reducing it. For indicators that may be used in regulatory or financial contexts, this raises concerns about auditability and trust. Thus, any proposal for dynamic climate indicators should be accompanied by explicit information requirements and governance mechanisms.

Predictive flood analytics and risk metrics

Flood risk is a prominent climate hazard for urban areas, and data-driven approaches have proliferated alongside more traditional hydrodynamic modelling. A systematic review on urban flood susceptibility mapping synthesises remote sensing datasets, topographic and land-use variables, and ML methods, reporting that models such as random forests, gradient boosting and deep neural networks can achieve high predictive performance when sufficiently rich conditioning factors are available (Islam et al., 2025). The review also notes considerable heterogeneity in how outputs are defined (susceptibility indices, probabilities, hazard classes) and in how performance is evaluated, which complicates cross-study comparison (Islam et al., 2025).

Case studies using deep learning and hybrid schemes confirm that ML can identify dominant drivers of flood susceptibility and refine spatial resolution, but also show sensitivity to data quality, feature selection and transferability across regions (Islam et al., 2025). From the standpoint of digital climate indicators, the key point is that sophisticated hazard models do not automatically generate design-relevant risk metrics. Their outputs need to be linked to assets, functions and decision thresholds, and these links are often treated informally or in a project-specific way. Moreover, if flood susceptibility layers are not aligned with design models in terms of coordinate systems, resolution and temporal reference, there is a risk that indicators derived from them may look precise but be geometrically or temporally misaligned.

Predictive heat analytics and urban heat indicators

Urban heat and the urban heat island (UHI) effect have similarly attracted ML-based methods. A recent review of predictive approaches for UHI reports a shift from purely statistical or physics-based models towards ML, deep learning and hybrid approaches, using satellite imagery, in situ measurements and urban morphological data as inputs (Ahmed et al., 2025). Studies show that deep models can capture complex, non-linear relationships between land cover, building density, vegetation and surface temperature, yielding improved predictive accuracy compared to conventional methods (Ahmed et

Table 1: Literature-gap matrix

Literature stream	Carbon performance	Flood risk	Urban heat	Resilience	Cross-domain integration
BIM-integrated LCA	✓✓	✕	✕	✕	✕
Flood susceptibility & ML models	✕	✓✓	◐	✕	✕
Urban heat prediction models	✕	✕	✓✓	✕	✕
Resilience assessment frameworks	◐	◐	◐	✓✓	◐
Present work	✓	✓	✓	✓✓	✓✓

Legend:
✓✓ = core focus ✓ = addressed ◐ = partially addressed ✕ = not addressed

al., 2025). However, the same review emphasises limitations in generalisability, given that models trained on one urban context may perform poorly when applied elsewhere without retraining or adaptation (Ahmed et al., 2025).

For digital climate indicators, heat prediction is directly relevant to cooling loads, indoor overheating risk and outdoor comfort. Yet, as with flood models, heat predictions are typically published as thermal maps or statistics, not as indicators tied to specific buildings, spaces or design decisions. Without clear information on spatial correspondence, scenario assumptions and temporal resolution, integrating such outputs into carbon and resilience indicators becomes difficult.

Resilience assessment methods and metrics

Resilience assessment in the built environment has evolved from qualitative descriptions to more structured frameworks and tools. A critical review of resilience assessment tools for the built environment notes a proliferation of methods but also fragmentation, inconsistent definitions and limited empirical validation (Cerè et al., 2017). The authors argue that many tools remain diagnostic rather than operational, and that the relationship between resilience and sustainability is often under-specified (Cerè et al., 2017).

More recent work proposes quantitative key performance indicators (KPIs) for risk and resilience assessment of assets exposed to climatic and non-climatic hazards, combining concepts such as robustness, redundancy, exposure and recovery (Ricciardi et al., 2025). A systematic review of sustainability and resilience assessment methods in the construction sector further documents the uneven development of approaches and calls for integrated frameworks that explicitly link hazard scenarios, damage, recovery trajectories and environmental impacts (Sesana et al., 2024). In a related contribution, the sustainability of resilient built environments is examined by combining performance-based damage assessment under hazards with LCA of repair and retrofit interventions, illustrating both synergies and trade-offs between resilience and environmental performance (Tanguay et al., 2024).

Taken together, this literature indicates that resilience is conceptually rich and quantitatively expressible, but that indicators remain heterogeneous and only partially integrated into digital design and management tools. For digital climate indicators, this reinforces the need to anchor resilience metrics in clear functional definitions and information requirements.

While the reviewed literature demonstrates significant progress within individual analytical domains, it also reveals a persistent fragmentation across carbon assessment, climate risk modelling and resilience analysis. BIM-integrated life cycle assessment workflows primarily address mitigation objectives, but rarely engage with hazard exposure or recovery dynamics. Conversely, flood and urban heat modelling studies provide increasingly accurate hazard predictions, yet typically remain detached from asset-level carbon implications and resilience metrics. Resilience assessment frameworks offer conceptual and quantitative tools to describe system response and recovery but are often weakly connected to digital design and hazard analytics.

Cross-cutting gaps and integration challenges

Across the reviewed literature, a recurring limitation is the lack of integration between analytical domains that are increasingly interdependent under climate change. Carbon assessment studies typically operate at building scale and rely on static or weakly scenario-dependent assumptions, while flood and urban heat modelling often adopt district or city scales driven by climate scenarios that are not aligned with those used in mitigation analyses. This mismatch in spatial and temporal scales complicates the joint interpretation of results and limits their decision relevance.

A further cross-cutting gap concerns uncertainty communication. While uncertainty is acknowledged in both life cycle assessment and climate risk modelling, it is rarely propagated or interpreted consistently when outputs from different domains are combined. As a result, decision-makers may be exposed to indicators that appear precise but embed heterogeneous and implicit assumptions.

Finally, the mapping between hazards, asset performance and functional consequences remains weakly formalised in much of the literature. Flood and heat indicators often describe exposure or intensity, while resilience metrics attempt to capture response and recovery, yet explicit links between hazard levels, asset degradation and downtime are rarely integrated into digital workflows.

Table 2: Indicator families, information requirements, and typical gaps for flood- and heat-related applications

	Key Information elements	Alignment needs	Typical gaps
Carbon performance (WLC)	• BIM-based quantities • Material specifications • Environmental Product Declarations (EPDs) • Operational profiles • Decarbonisation scenarios	• Semantic alignment between BIM elements and LCA datasets • Temporal alignment with energy and grid scenarios	• Missing or inconsistent material metadata • Partial system boundaries • Limited transparency on scenario assumptions
Climate risk (flood & heat)	• Flood susceptibility or hazard maps • Urban heat metrics (e.g. land surface temperature, heat indices) • Exposure data • Basic vulnerability attributes	• Spatial alignment between hazard layers and asset geometries • Consistent resolution and coordinate systems • Scenario coherence	• Resolution mismatch • Limited transferability of ML models • Weak linkage between hazard scores and asset-level consequences
Resilience performance	• Functional definitions • Performance thresholds • Recovery time assumptions • Interdependency between systems	• Temporal alignment between hazard events and recovery models • Explicit performance metrics	• Qualitative or poorly validated recovery assumptions • Lack of standardised resilience metrics

These gaps motivate the need for a structured framework that explicitly connects carbon performance, climate risk and resilience through shared information requirements and transparent alignment processes

Table 2 synthesises these patterns by positioning the main literature streams against the three indicator families considered in this paper, highlighting both areas of maturity and structural gaps, and it motivates the need for an integrative indicator framework capable of bridging these domains.

Proposed framework

Concept of digital climate indicators

The aim of the present work consists in defining a digital climate indicator as a metric derived from digital representations of the built environment (e.g.. BIM and IFC models, GIS layers, sensor data and climate-scenario datasets) to support climate-responsive and resilient decision-making. Unlike generic performance metrics, digital climate indicators are explicitly conditioned on the quality, completeness and alignment of the underlying information from which they are computed.

The need for such indicators emerges from the growing availability of digital data and analytical tools, coupled with the persistent separation between mitigation-oriented assessments (e.g. whole-life carbon, WLC) and adaptation-oriented analyses (e.g. climate risk mapping). Recent literature shows that while BIM-integrated life cycle assessment (LCA) workflows and DTs enable increasingly automated and dynamic carbon assessment (Parece et al., 2025; Petri et al., 2025), climate risk analytics, particularly for flooding and heat, are often developed independently using geospatial data and machine learning (ML) models (Islam et al., 2025; Ahmed et al., 2025). As a result, outputs are difficult to combine in a coherent and decision-relevant manner. Digital climate indicators are therefore not conceived as new standalone metrics, but as integrative constructs that explicitly link carbon performance, climate risk and resilience performance under clearly defined information requirements. Their value lies not only in numerical results, but in their ability to make assumptions, dependencies and uncertainties explicit across heterogeneous digital workflows.

Indicator families and hazard focus

The development of the indicator is structured along three indicator families: carbon performance, climate risk, and resilience performance. These families correspond to distinct but interdependent decision questions: how much an asset contributes to climate change, how exposed it is to climate hazards, and how it performs and recovers when those hazards occur. Within the broad spectrum of climate-related hazards, this framework focuses explicitly on urban flooding and urban heat. This choice is motivated by three considerations. First, flooding and heat are consistently identified as dominant climate threats for urban areas, with significant impacts on infrastructure functionality, public health and economic activity (IPCC, 2023). Second, both hazards are supported by relatively mature digital ecosystems, including remote sensing, urban morphology datasets and ML-based susceptibility or prediction models, making them suitable for indicator computation at building and district scales (Islam et al., 2025; Ahmed et al., 2025). Third, flood and heat hazards exhibit strong interactions with both operational and embodied carbon, as well as with resilience metrics such as downtime and recovery, making them particularly relevant for illustrating cross-domain trade-offs. By concentrating on flooding and heat, the framework achieves methodological clarity while remaining representative of broader multi-hazard challenges faced by climate-resilient design.

Information requirements and alignment conditions

A recurring finding across the reviewed literature is that the main barriers to integrated climate assessment are not

purely computational, but informational. The computation of digital climate indicators depends on the availability, alignment and interpretability of heterogeneous data sources. Table 2 synthesises the key information requirements associated with each indicator family, with explicit reference to flood and heat-related applications. For carbon performance indicators, such as WLC, required information includes BIM-based quantities, material specifications, Environmental Product Declarations (EPDs), operational profiles and decarbonisation scenarios (Zhang et al., 2025; Berges-Álvarez et al., 2024). Semantic alignment between BIM elements and LCA datasets is critical, missing or inconsistent metadata can lead to systematic underestimation or exclusion of relevant components. For climate risk indicators, flood susceptibility maps and urban heat metrics must be spatially linked to asset geometries through consistent coordinate systems and resolution. ML-based models can provide high-resolution hazard predictions, but their outputs must be interpreted within explicit scenario and validation boundaries to avoid false precision (Islam et al., 2025; Ahmed et al., 2025).

For resilience performance indicators, information requirements extend beyond hazards to include functional definitions, performance thresholds and recovery assumptions. Literature on resilience assessment highlights that these aspects are often weakly formalised or treated qualitatively, limiting indicator comparability and robustness (Cerè et al., 2017; Sesana et al., 2024).

Across all families, spatial alignment, temporal and scenario transparency, and uncertainty representation emerge as cross-cutting requirements. Table 2 identifies the minimum information elements required for each indicator family. Figure 1 translates these requirements into an operational alignment sequence and it conceptually illustrates the proposed workflow for computing digital climate indicators by explicitly articulating how heterogeneous digital data sources are transformed into decision-relevant metrics. The workflow is structured around three main stages: data sourcing, alignment and processing, and indicator computation. Rather than assuming seamless interoperability, the figure makes visible the intermediate alignment steps (i.e.; semantic enrichment, spatial matching between assets and hazard layers, temporal and scenario harmonisation, and uncertainty characterisation) that condition the correctness and robustness of the resulting indicators. By positioning these steps explicitly between data ingestion and indicator outputs, the workflow highlights where information gaps, misalignments or implicit assumptions may propagate into carbon, flood and heat risk, and resilience metrics. This articulation provides the basis for a critical discussion of typical failure modes and limitations, which are addressed in the following section.

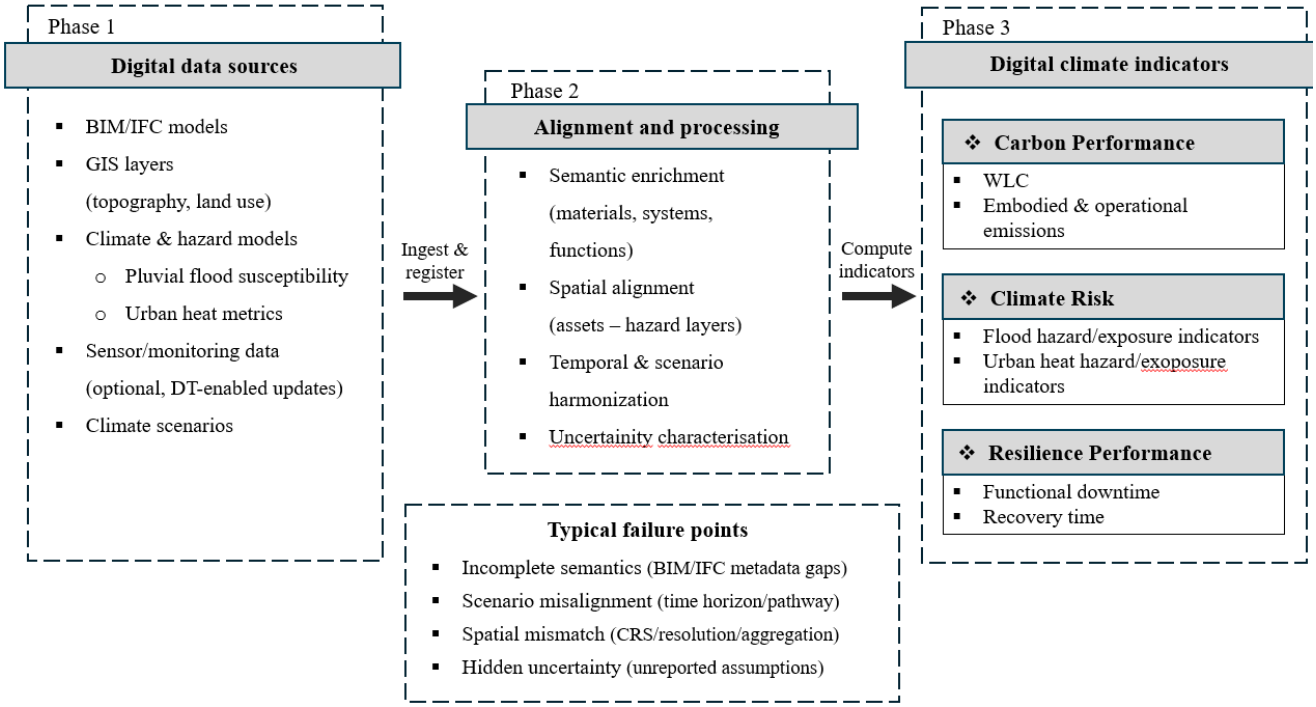


Figure 1: Conceptual workflow for digital indicators' computation in three action-oriented stages, integrating WLC, flood and urban heat risk, and resilience performance

Indicator validity, scale, and decision context

A critical aspect of digital climate indicators concerns their validity across scales and decision contexts. Indicators computed at building scale may be suitable for early design choices or retrofit comparisons, whereas district-scale indicators are more relevant for planning decisions, infrastructure prioritisation and portfolio-level assessment. The proposed framework does not prescribe a single scale of application; instead, it requires that the scale of indicator computation is explicitly aligned with the intended decision context.

In this perspective, indicator validity is not primarily defined by predictive accuracy, but by internal consistency, transparency of assumptions and sensitivity

to information completeness. Whole-life carbon indicators, for example, may support robust comparison between alternative design strategies even when absolute emission values are uncertain, provided that system boundaries, material mappings and scenarios are consistently defined.

Similarly, flood and urban heat indicators are often more informative when used to position options relative to one another rather than to predict exact hazard impacts. By making scale and decision context explicit, the framework helps prevent the misuse of indicators outside their domain of validity and supports more appropriate interpretation of results. This shift from optimisation to comparative reasoning is particularly relevant for climate-responsive design, where decisions must be made under deep uncertainty and evolving boundary conditions.

Research questions and methodological directions

The proposed framework motivates a set of research questions aimed at advancing the development and use of digital climate indicators for flood- and heat-related applications. Such questions are derived from the literature review and framework synthesis and are presented as a research agenda rather than being fully answered within the present contribution.

RQ1 - Indicator correctness

Under what information conditions can integrated carbon, flood/heat risk and resilience indicators be considered correct representations of climate performance? This question concerns the relationship between indicator definitions and the completeness and consistency of the underlying digital information.

RQ2 - Portability and interoperability

To what extent can indicators computed for flood and heat hazards be reproduced across projects, tools and local contexts? Given the documented variability in BIM semantics, LCA datasets and ML-based hazard models, this question addresses the limits of indicator transferability (Parece et al., 2025; Islam et al., 2025).

RQ3 - Uncertainty and robustness

How should uncertainty from climate scenarios, hazard models and recovery assumptions be propagated and communicated through digital climate indicators? Prior work on sustainability and resilience assessment suggests that robustness-oriented representations may be more decision-relevant than single deterministic values (Sesana et al., 2024; Tanguay et al., 2024).

RQ4 - Dynamic updating through DTs

When indicators are computed and updated within DT-based environments, how can provenance, versioning and validation be managed to avoid drift and loss of comparability over time (Petri et al., 2025; Luo et al., 2024)?

Mini operational example: flood- and heat-responsive indicators

Aiming at illustrating the applicability without introducing a full case study, a minimal operational example is presented below, applying the workflow introduced in Figure 1 to the case of urban flooding and urban heat, illustrating how the same sequence of data ingestion, alignment and indicator computation can be instantiated for concrete climate hazards. The example is intentionally procedural (computability-focused) and does not claim empirical validation of indicator values.

In this example, the digital artefacts required at each step correspond directly to the information requirements outlined in Table 2. Concretely, the BIM/IFC inputs must be enriched with the minimum attributes required for WLC mapping (e.g., construction assemblies, key system descriptors, and stable identifiers to link to EPD datasets), as well as basic functional zoning information to support resilience metrics. The flood and heat layers must then be aligned to asset geometries through explicit georeferencing and scale handling (e.g., aggregation rules when raster resolution is coarser than building/zone geometry, and consistent assumptions on reference elevation for flood susceptibility interpretation). Finally, temporal and scenario harmonisation is required to avoid false precision, ensuring that carbon and hazard indicators refer to compatible horizons and scenario choices (e.g., heat-related operational impacts under specified climate pathways).

Carbon performance indicators are derived from a BIM-based WLC assessment, including embodied emissions associated with building fabric and adaptation measures, as well as operational emissions influenced by increased cooling demand under projected heat conditions. These computations assume the availability of semantically enriched BIM models and scenario-based energy profiles, but do not require real-time monitoring or full operational calibration.

Climate risk indicators are computed by spatially linking asset geometries to pluvial flood susceptibility maps and urban heat metrics derived from remote sensing and machine-learning-based models. At this stage, the workflow relies on consistent spatial resolution and explicit scenario selection rather than on hazard model optimisation.

Resilience performance is expressed through simplified metrics such as functional downtime and recovery time following representative flood and heat events. These metrics are not intended as predictive estimates, but as comparative indicators that allow alternative design or adaptation options to be positioned relative to one another.

By following the workflow in Figure 1, the example demonstrates how carbon, flood and heat risk, and resilience indicators can be jointly computed under clearly stated assumptions. The resulting indicators support comparative reasoning across options, rather than optimisation, and make explicit where trade-offs and uncertainties enter the decision process.

Discussion

The proposed framework aims to respond to recurrent limitations observed in current digital climate assessment practices through the research questions introduced

previously. Rather than providing definitive solutions, the framework clarifies how integrated digital climate indicators can be constructed and interpreted under explicit conditions, and where their limitations remain.

Indicator correctness (RQ1) emerges primarily as an information-driven challenge. As illustrated by the workflow in Figure 1, the framework makes explicit that indicator validity depends on semantic completeness, spatial alignment and scenario consistency across heterogeneous digital inputs. This represents a clear advance over approaches in which carbon, flood or heat indicators are computed independently, often masking underlying assumptions. At the same time, the framework does not eliminate epistemic uncertainty: if information is incomplete or misaligned, indicators may still be internally consistent yet only partially representative. The contribution therefore lies in making such limitations visible rather than implicit.

Portability and interoperability (RQ2) are addressed by structuring indicators around shared information requirements rather than tool-specific implementations. Table 2 highlights how carbon, climate risk and resilience indicators rely on different datasets and modelling choices, many of which are context dependent. While this limits full transferability, the framework supports partial portability by enabling indicators to be recomputed consistently across projects, if information requirements are met. A remaining challenge concerns machine-learning-based flood and heat models, whose sensitivity to local training data constrains cross-context comparability and requires careful documentation.

Uncertainty and robustness (RQ3) are treated implicitly through the framework's emphasis on comparative rather than absolute indicator interpretation. As suggested by the operational example, integrated indicators are more effective in positioning design or adaptation options relative to one another than in providing precise predictive values. This shift aligns with recent work advocating robustness-oriented assessment in climate-resilient design but also raises challenges in communicating uncertainty without undermining decision confidence. The framework offers a structure for this communication, but methodological advances are still needed to operationalise uncertainty propagation across indicator families.

Dynamic updating through digital twins (RQ4) represents both an opportunity and a risk. DT-based environments enable indicators to be recalculated as new data become available, supporting adaptive decision-making over time. However, without explicit governance mechanisms, such as versioning, provenance tracking and validation rules, dynamic updating may compromise indicator comparability and trust. The framework highlights these risks by positioning updating as a conditional capability rather than an inherent benefit, underscoring the need for controlled integration rather than continuous recalculation.

The framework does not aim to provide empirical validation of indicator values, nor to define new standards, thresholds or regulatory metrics. Its purpose is to clarify the informational and alignment conditions under which integrated digital climate indicators can be computed and meaningfully interpreted.

Despite its potential, the framework has deliberate limitations: it does not harmonise underlying carbon, flood/heat, or resilience models, and comparability remains dependent on the transparency and context-dependence of adopted datasets and assumptions. Addressing these limitations requires empirical multi-context testing and governance mechanisms (e.g., provenance, versioning and validation rules), which are beyond the scope of this contribution.

Overall, the proposed framework enables a more transparent and structured engagement with carbon, flood and heat risk, and resilience indicators, while also exposing the practical and methodological challenges that remain. Its value lies less in resolving these challenges than in providing a coherent basis for addressing them in future empirical and methodological work.

Conclusion

This paper has addressed the challenge of linking mitigation and adaptation-oriented analyses in the built environment by proposing a framework for digital climate indicators that integrate whole-life carbon performance, flood and urban heat risk, and resilience metrics. While digitalisation has enabled unprecedented access to data and analytical tools, current workflows remain fragmented, limiting their ability to support climate-responsive and resilient design decisions.

Building on a structured review of recent literature, it is shown that the technical components required for integration, BIM-based carbon assessment, DT-enabled dynamic analysis, ML-based hazard modelling and resilience metrics, are already available, but rarely combined under explicit information requirements. By organising indicators into three families and articulating the alignment steps needed to compute them, the proposed framework clarifies where and why integration succeeds or fails. The workflow illustrated in Figure 1 and the information mapping in Table 2 provide a conceptual scaffold for moving from isolated analyses to coherent, decision-relevant indicators. The focus on urban flooding and urban heat has allowed the framework to be grounded in high-impact, data-rich hazards that are representative of broader climate challenges. The mini operational example demonstrates how integrated indicators can reveal cross-domain interactions and trade-offs that would remain hidden in siloed assessments, even without a full case study.

The lasting contribution of this work lies not in defining new metrics, but in reframing indicator development as an information-driven and methodologically explicit process. Taken together, the framework allows a first set of methodological answers to the proposed research questions to be stated. Regarding indicator correctness (RQ1), the paper clarifies that "correctness" is conditional on explicit information requirements and alignment steps: integrated indicators become fragile when BIM/IFC semantics are incomplete for WLC mapping, when hazard

layers are not spatially matched to assets, or when temporal/scenario assumptions are inconsistent. Regarding portability (RQ2), the framework supports recomputation across projects by making the required inputs and mappings explicit, but comparability remains constrained by context-dependence of datasets and by the transferability limits of flood and heat models. Regarding uncertainty (RQ3), the paper supports a robustness-oriented interpretation: the most defensible use of integrated indicators at early stages is comparative reasoning across options under stated scenarios, rather than absolute prediction. Finally, for dynamic updating (RQ4), the framework highlights that DT-enabled recalculation is useful only under governance mechanisms (provenance, versioning and validation rules) that preserve traceability and comparability over time. In this perspective, digital climate indicators should be treated as information-driven constructs whose validity depends on explicit assumptions, alignment conditions and decision context. By reframing indicator development around information requirements and integration workflows, this paper contributes a structured basis for linking carbon performance, flood and urban heat risk, and resilience within digital design and assessment processes. Future work will focus on applying the proposed framework to real urban case studies exposed to flood and heat stress, to test indicator computability, sensitivity to information completeness, and usefulness for comparing design and adaptation options. Such applications will be essential to further consolidate the framework and to support the responsible use of integrated indicators in climate-responsive and resilient design.

From a practical standpoint, these conclusions suggest that deployment-oriented work should treat information requirements and alignment procedures as first-class deliverables, alongside algorithmic components. For research evaluation, this implies reporting integration effort (e.g., BIM/IFC enrichment steps, spatial matching rules, scenario harmonisation choices, and governance overhead) as part of methodological transparency, rather than focusing solely on model accuracy.

Ultimately, advancing digital climate indicators requires not only better models, but clearer assumptions, shared information requirements and critical reflection on how indicators are used in practice.

Acknowledgements

The authors acknowledge the EC3 Data Sensing & Analysis (DSA) Committee for fostering interdisciplinary collaboration and supporting the development of integrated research perspectives on data-driven technologies in the built environment.

References

Ahmed, A.N., Bakar, A.A., Othman, F.B., Afan, H.A., El-Shafie, A. (2025). The urban heat island effect: A review on predictive methods and modelling approaches. *Results in Engineering*, 25, 101879.

Berges-Álvarez, I., Martínez-Rocamora, A., Solís-Guzmán, J., Marrero, M. (2024). BIM-based life cycle sustainability assessment for buildings: A systematic review. *Sustainability*, 16(24), 11070.

Cerè, G., Rezgui, Y., Zhao, W. (2017). Critical review of existing built environment resilience assessment tools. *Sustainable Cities and Society*, 28, 174–186.

IPCC (2023). *Climate Change 2023: Synthesis Report*. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva: IPCC.

Islam, T., Zeleke, E.B., Afroz, M., Rahman, M.A., Ahmed, A.N. (2025). Urban flood susceptibility mapping: A systematic review of remote sensing, machine learning and modelling approaches. *Remote Sensing*, 17(3), 524.

Luo, Y., Liu, Y., Sun, Y., Ding, L. (2024). Leveraging digital twins and dynamic life cycle assessment for sustainable systems. In: *Proceedings of the Sustainable Design and Manufacturing Conference 2024*. Springer, Cham.

Parece, S., Resende, R., Rato, V. (2025). BIM-based life cycle assessment: A systematic review on automation and decision-making during design. *Building and Environment*, 248, 111726.

Petri, I., Amin, A., Ghoroghi, A., Hodorog, A., Rezgui, Y. (2025). Digital twins for dynamic life cycle assessment in the built environment: The Building Life-cycle Digital Twin (BLDT) framework. *Science of the Total Environment*, 993, 179930.

Ricciardi, G., Bellagamba, S., Bolognini, G., Crotti, G. (2025). Quantitative key performance indicators for risk and resilience assessment of built environment assets under climatic and non-climatic hazards. *Sustainable Cities and Society*, 104, 105349.

Sesana, M.M., Rota, M., Poli, T., et al. (2024). Sustainability and resilience assessment methods: A systematic literature review for the construction sector. *Energies*, 17(6), 1440.

Tanguay, X., Gosselin, L., Ménard, S. (2024). Assessing the sustainability of a resilient built environment: A review. *Journal of Cleaner Production*, 421, 140437.

Zhang, Y., Wang, J., Ding, Z., Li, H. (2025). Insights from review articles on life cycle assessment for buildings. *Applied Sciences*, 15(3), 1324.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

BEYOND TRAFFIC PERFORMANCE: ASSESSING TRANSPORT NETWORK RESILIENCE THROUGH ACCESS AND EQUITY

Giuseppe Vega[1], Peter Demian[1], Georgios M. Hadjidemetriou[1,2]
[1]School of Architecture, Building and Civil Engineering, Loughborough University, UK
[2]The Bartlett School of Sustainable Construction, University College London, UK

Abstract

Weather disruptions endanger transport network operations and accessibility, hindering people's possibilities to reach opportunities and services. While most research focuses on infrastructure impacts, few studies have examined how different socioeconomic groups experience such disruptions. This paper introduces a comprehensive method for assessing transport resilience to extreme weather and includes a case study highlighting social disparities in access to healthcare services. The case study reveals that access to essential city services varies by socioeconomic strata and that extreme weather events can exacerbate inequalities.

Introduction

Maintaining physical connection between distant places has been fundamental to the functioning of societies and economies throughout history. Transport networks enable the movement of people and goods, serving as the backbone of economic activity and societal development (Zhang and Cheng, 2023). In recent decades, these infrastructure systems have faced significant challenges due to the increasing intensity and frequency of extreme weather events (Kurth et al, 2020). The disruptions caused by these events have had several implications, including economic losses, casualties, and physical and psychological health issues among affected populations (European Environment Agency, 2024).

Research on transport infrastructure resilience to extreme weather constantly operationalises resilience as a time-dependent property of service that combines (i) the ability to withstand or absorb a disruption, (ii) the capacity to adapt to new settings, and (iii) the ability to recover within acceptable time and cost envelopes (Bruneau et al., 2003; Li et al., 2020; Wassmer et al., 2024). Hence, resilience metrics are described in operational terms (e.g., speed, delays), network-topology terms (e.g., efficiency under link removal), accessibility terms (e.g., reachable opportunities), or monetised socioeconomic impacts (e.g., wage loss) (Afrin and Yodo, 2020; Faturechi and Miller-Hooks, 2014; Hadjidemetriou et al., 2022). As a consequence of this approach, the role of social disparities among affected communities is often overlooked.

For example, the operational approach to resilience treats performance metrics, such as time delays, as both the primary measure of impact and diagnostic tool for inferring management priorities. From operational approaches, a resilient transport network is resistant to disturbances and can recover rapidly in case of occurrence, ensuring traffic flow and reducing congestion (Ganin et al., 2017; Zhang et al., 2017). In this line, Rebally et al. (2026) assess the impact on traffic of the 2013 flood in Calgary (Canada) using HEC-RAS for flood modelling and SUMO for micro-scale traffic simulation, and demonstrate the extent to which the flooding event causes longer travel times and distances, and how the disruption leads to cascading impacts. Likewise, Liu et al. (2026) integrate the effects of flooding, reduced visibility, and traffic signal failures during extreme rainfall to evaluate network performance by estimating the average delay per time slice and a delay-based resilience index in a case study of Greater London. In these and similar studies, transport resilience is measured by observing or simulating operational degradation and recovery in traffic or service variables under specified meteorological conditions. This stance treats disruption impact as a departure from the original “steady-state”, and resilience as the shape of the performance-time profile, frequently summarised into a scalar “resilience index” for comparison across different scenarios (Achillopoulou et al., 2020; Zhang et al., 2024).

Scholars across fields and perspectives do not deny the importance of societies' ability to withstand shocks, but call for awareness of how that goal is pursued (Cutter, 2016). Uncritically applied, resilience may imply short-term stability at the cost of long-term vulnerability. However, if conceived critically, resilience can be aligned with transformative adaptation and social justice, thereby making communities safer and more equitable. In this line, a significant share of research in transport resilience has shifted from “traffic loss” to “activity loss”.

A major stream of research treats resilience and vulnerability to extreme weather events as changes in access (to opportunities, locations, services, etc.) rather than as changes in traffic speed, flow or time delays. In this approach, the impact of weather-related disruption is translated into the extent to which it elongates or interrupts effective connections between people and destinations, and resilience is evaluated by how rapidly and unevenly accessibility is restored or substituted (Martín et al., 2021). Accessibility metrics are frequently

integrated with link-removal experiments and hazard overlays. For example, Martín et al. (2021) compare how territorial accessibility is reduced under the cumulative exclusion of transport links through random removal, removal by criticality, and targeted removal in high-flood-risk areas. Other studies, like Farahmand et al. (2024) focus on metrics such as agency costs, user costs, and environmental costs to enable targeted investment through benefit-cost analysis. Overall, the methodological importance lies in integrating climate change factors and effectively coupling network criticality with hazard-informed failure mechanisms (Martín et al., 2021).

Further critical perspectives serve a crucial function in prompting the question: 'Resilience for whom and to what end?' (Cretney, 2014; Cutter, 2016). For example, little is said about prioritising emergency vehicles when recovering the level of service (Wassmer et al., 2024; Wu and Chen, 2023), or trips by road users trying to reach healthcare facilities.

A critical gap exists in leveraging existing methodologies to account for social disparities in access to essential services through the existing infrastructure, particularly in the context of weather-related disruptions. For this reason, this paper leverages transport modelling principles to investigate the effect of weather-related disruptions on trips to healthcare facilities across users from households with different socioeconomic strata. To achieve this, the proposed workflow, the required data and preprocessing procedures, model specifications, and the disruption experiment design are presented in the Methodology chapter. Then, a case study of a real midsize city in Colombia is presented, highlighting pre-existing disparities among households categorised as low-, mid-, and high-socioeconomic stratum households in access to healthcare services and how a flooding disruption can deepen these disparities.

Methodology

This paper leverages transport planning tools and data from Origin-Destination (OD) surveys to develop a reproducible workflow for estimating the impact of extreme-weather-related disruptions on groups of people with different income levels and trip purposes. The main steps in the workflow are (i) the construction of the model in Geographical Information Systems (GIS) of the city's road network, (ii) the simulation of the OD demand matrix distribution over the network under static traffic assignment, and (iii) the quantification of network performance and travel times for groups of interest for a baseline "as-is" scenario and a disrupted scenario.

Data and tools

The proposed workflow requires: (i) geographical network and zoning data; (ii) the demand OD matrices for the established zoning; and (iii) the disruption layer (i.e., flooding risk map). The network can be assembled from GIS vector layers representing nodes (i.e., intersections and connections between origins and destinations), links (i.e., roads), and zones (i.e., Transport Analysis Zones). This study employs AequilibraE for modelling and aligns data requirements with this tool (Camargo, 2015). However, other software can be used to build the model. Overall, network data can be accessed using OpenStreetMap (OSM), obtained from local transport authorities, or built from a basic geometric representation of the network. Table 1 summarises the network attributes that must be specified.

Table 1: Network attributes specification

Data	Format	Source(s)
Nodes (N)	Geometry	OSM
N_ID	ID number of node	
N_IsCentroid	1: is a centroid 0: is not a centroid	Computed
Links (L)	Geometry	OSM
L_NodeA	ID number of node A	
L_NodeB	ID number of node B	
L_Direction	1: from A to B 0: both ways -1: from B to A	
L_Distance	Length [m]	
L_Lanes (AB & BA)	Number of lanes	
L_Speed (AB & BA)	Speed [km/h]	Observed/ Estimated
L_Type	'Default' or 'Connector'	Computed
L_TravelTime (AB & BA)	Travel time [s]	Computed
L_Capacity (AB & BA)	Total capacity [PCU/h]	Computed
Zones (Z)	Geometry	OD survey – Transport Analysis Zones (TAZs)
Z_ID	ID number of zone	

OD demand matrices must be provided, including the origin zone (i), destination zone (j), and number of trips (from i to j). For this study, the static traffic assignment approach assumes a single network for Passenger Car Unit (PCU) assignment, so trips are computed by assigning 1.0 PCU to car trips and 0.4 PCU to motorcycle trips. Table 2 lists the fields required for the OD matrices.

Table 2: Fields required for OD matrices

Field	Format	Source(s)
Origin	ID number of zone	OD survey (individuals)
Destination	ID number of zone	
Trips	PCU/h	
Mode	'car' or 'motorcycle'	OD survey (households)
Others (e.g., income level, trip purpose, etc.)	-	OD survey

The preprocessing procedure includes verifying network connectivity (i.e., every link has a start and an end node; no nodes or links are isolated from the network; each zone has a centroid node; and each centroid node is associated with a connector link to the network). Also, some network attributes are either observed or computed from the available OSM data. Speed (free-flow) on links can be observed or estimated from the link road class (e.g., highway, local, collector, etc.), the topology of the network, and the traffic management systems (e.g., heavy signalling cities with many intersections and pedestrian crosswalks require a lower speed than metropolitan areas connected by major roads or highways). Travel time is computed based on link speed and distance, and capacity is assumed to be the standard value per lane (1800 PCU/h/lane).

The simulation requires that the OD matrices be balanced with the Production-Attraction (P-A) vectors (i.e., the total number of trips produced equals the total number of trips attracted). Trips from OD surveys are converted to PCUs using assigned mode-specific values and aggregated across fields of interest based on the available data (e.g., trip purpose, income level, etc.).

Model specification

The model follows the principles of a 4-step transport model for an urban setting (Ortuzar and Willumsen, 2024). However, the first three steps (i) trip generation, (ii) distribution, and (iii) mode choice are intentionally fixed to isolate the effects of the weather-related disruption on (iv) the assignment step. The network is represented as a graph for the car mode (or PCU), where the supply representation and cost function are given via the congested travel time on the links, which is computed post-assignment using the Bureau of Public Roads (BPR) as Volume-Delay Function (VDF) (Bureau of Public Roads, 1964):

$$t = t_0\left(1 + \alpha\left(\frac{v}{c}\right)^{\beta}\right) \quad (1)$$

Where for each link, t0 is the initial travel time defined using estimated/observed free-flow speed, v is the assigned flow, c is the capacity, and α and β are parameters of the BPR function that define how early and sharply congestion starts.

The model employs two different assignment steps. The initial assignment is made using the methodology All-or-Nothing (AoN) (Bureau of Public Roads, 1964) to ensure network connectivity. This is done by loading the entire demand onto a single least-cost path computed using the initial travel time (i.e., there is no feedback from congestion, as the paths are defined by the static cost). Once the model implements the AoN assignment, the model moves to the User Equilibrium (UE) assignment (Patriksson, 1994), which operationalises Wardrop's principle: no traveller can reduce their travel time by unilaterally changing routes, given the congestion created by everyone else. The UE assignment is an iterative procedure in which each iteration takes the current estimate of travel times on the network's links, finds the shortest path, assigns flows, updates travel times using the previously defined VDF function, and then repeats until the convergence criterion is met.

Modelling outputs

Once the OD matrix trips are assigned to the network, traffic flow (PCU/h) and travel times are obtained for each used link. Then, instead of determining resilience in terms of operational variables that average the experience across all the road users in the network, the travel time for every group defined by the previously chosen fields of interest (e.g., income level, trip purpose) can be computed using the initial OD trips matrix and the OD congested times matrix, as:

$$\bar{T}^{(g)} = \frac{\Sigma_i \Sigma_j D_{ij}^{(g)} T_{ij}}{\Sigma_i \Sigma_j D_{ij}^{(g)}} \quad (2)$$

Where g is the index for the groups made of chosen fields (e.g., higher income level travelling for work), Dij(g) is the demand for group g from the origin i to the destination j, and Tij is the congested time from i to j. Obtaining travel times for different user types is not only useful for acknowledging disparities between groups but also for understanding whether these disparities are exacerbated under disruption scenarios.

Disruption experiment design (flood scenario)

The disruption input is a polygon layer delineating high-risk areas for disruption (e.g., flooding). The disruption mechanism reduces link capacities in proportion to the length share of each link intersecting the risk areas, scaled by a scenario severity parameter. Flood share in links is computed as the percentage of the link's length that overlaps the flooding area. The capacity reduction rule induced, and the corresponding parameters are as follows:

$$cap_{flood} = cap_0(1 - severity \cdot flood_{share}) \quad (3)$$

Where cap_{flood} is the new capacity of the link to be used for the trip assignment, cap0 is the initial capacity considered for the baseline scenario, and the severity is a parameter that determines how the link's capacity is affected by the portion affected by the flood. However, a capacity floor is required to prevent convergence issues caused by ODs that are 4disconnected and cannot be assigned.

Case study

To test the proposed workflow, the case study compares a baseline assignment with a flooding scenario for Villavicencio, a mid-sized city of approximately 600.000 inhabitants and an urban centre spanning up to 70 km2, located in the eastern plains of Colombia. Using data from the Origin-Destination Survey conducted in the city in

2022, travel times for people from households in lower-, middle-, and higher socioeconomic strata whose purpose of travel is "health" are obtained. The results of the baseline scenario show the extent to which times differ across socioeconomic strata and how they change during a flooding-induced disruption across the network.

Figures 1 and 2 display the network (links only) and zoning (TAZs) of Villavicencio, highlighting flood-prone areas along the city's rivers, as well as the density of "health" trips production (Figure 1) and attraction (Figure 2).

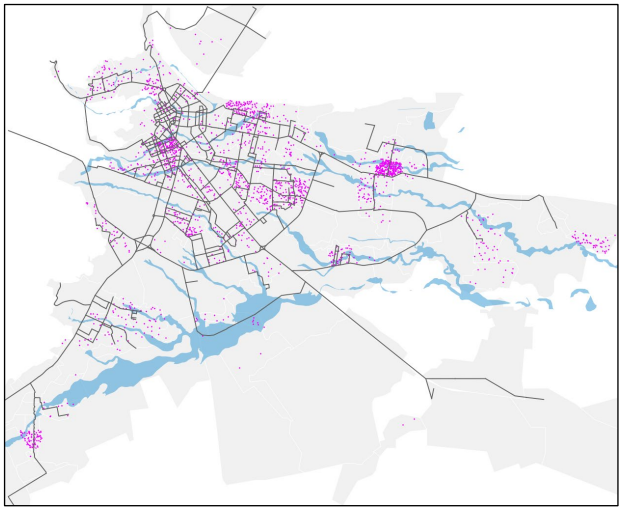


Figure 1: Villavicencio's network, flood-prone areas and "healthcare" trips production

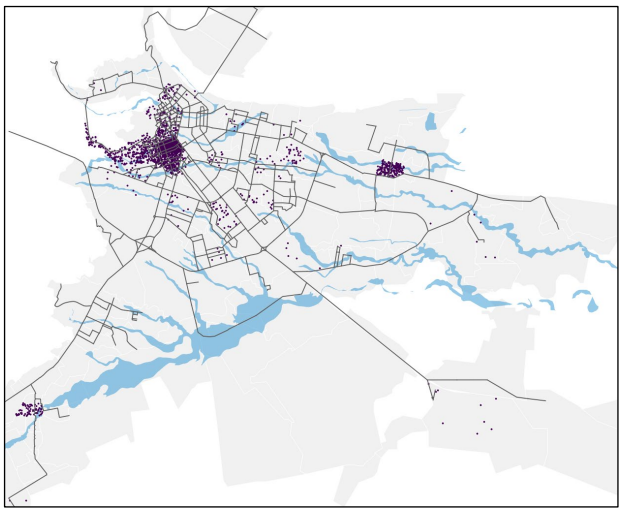


Figure 2: Villavicencio's network, flood-prone areas and "healthcare" trips attraction

The OD survey gathers information on approximately 11.500 trips that, after applying expansion factors, account for approximately 880,000 trips, of which 300,000 are by car or motorcycle (as driver or taxi user). Of the total number of trips, around 23.000 are made to access healthcare. However, despite the provision of health centres at each of the three corners of the triangular urban area, access remains unequally distributed across the population. In Colombia, households are classified in six socioeconomic strata based on their location, surrounding infrastructure, and access to public services (e.g., water and energy supply) (DANE, 2021). Table 3 shows the number of "health" trips by stratum (of the household) and mode during a day.

Table 3: Number of "health" trips by mode and level of income

Socioeconomic strata	Trips by car	Trips by motorcycle	Other modes
Low (1 & 2)	2786	1601	10244
Middle (3 & 4)	4125	1036	2870
High (4 & 6)	248	29	129

Since the methodology is based on static traffic assignment, using only trips observed during the highest-demand hour would produce a sparse OD matrix and exclude a substantial number of "health" trips. Therefore, the OD pattern was determined from a 3-hour period with the highest demand, which provides better spatial coverage of trips.

To ensure consistency with the static assignment framework, in which network capacities are represented on an hourly basis, the 3-hour OD matrix is converted into an equivalent peak-hour matrix using the expression:

$$T_{ij} = \frac{T_{1h,max}}{T_{3h,max}} \cdot T_{ij,3h} \quad (4)$$

The resultant OD matrix is converted to PCU by assigning 1.0 PCU to cars and 0.4 PCU to motorcycles. Along with the network properties, topology and scenario settings, traffic assignment is conducted to obtain congested travel times for every OD pair. The UE assignment is conducted using $\alpha = 0.3$ and $\beta = 2.0$. This procedure is carried out for the baseline and disruption scenarios, in which the link segments that intersect the flood-prone area (highlighted in Figure 3) are used to reduce capacity accordingly.

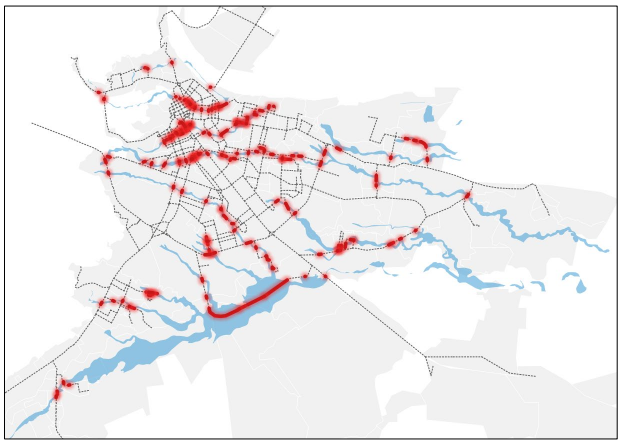


Figure 3: Network links intersecting flood-prone areas

Table 4 presents the number of health trips considered in the modelling period, and the estimated travel time across

the three socioeconomic strata groups and simulated scenarios.

Table 4: Travel time by scenario, socioeconomic stratum for "health" trips on the 3-hour peak period

Scenario	Socio-economic stratum	Health trips	Health trips avg travel time, min
Baseline	Low	1049	29.9
	Middle	950	17.9
	High	67	13.4
Disruption	Low	1049	31.8
	Middle	950	18.6
	High	67	13.9

The results of the baseline scenario show that access to health services in the city is not equally distributed among people across different socioeconomic strata. Even in the absence of disruption, the network topology and urban setting make it more difficult for most of the people in the city, who live in lower socioeconomic stratum households, to access health services in the city. People in this stratum experience travel times up to 67% longer than those in middle-stratum households and 123% longer than those in higher-stratum households.

In the flooding scenario, travel times are generally longer for all trips because congestion from reduced capacity on affected links impacts everyone. Nevertheless, the results show that the flooding does not have the same effect on everyone in the city. Travel time for trips by people in lower stratum households to access health services increases by 6.5%, by 3.9% for middle stratum households, and by 3.8% for high stratum households. Thus, the urban setting and transport network not only provide reduced access to healthcare services for people in less favourable conditions but also allow for this gap to be stretched during flooding-related disruptions.

Conclusions

This paper presents a methodology that leverages transport demand modelling tools and OD survey data to assess the impact of weather-related disruptions beyond traffic metrics. Although transport modelling is a valuable tool for understanding the observable effects of disruptions on transport operations, it is crucial to continue developing methodologies that translate traffic performance assessments into a more comprehensive resilience assessment of transport networks.

The main contribution of this paper is to advance this line by proposing a workflow that accounts for social disparities in the context of transport network disruptions. Further research can improve the methodology by incorporating tools that enable dynamic analysis of the disruption over time (e.g., agent-based modelling), while still examining the impact of these events on people rather than on assets or operations. Other disruptions can also be considered, as the methodology allows for the implementation of additional event types that ultimately affect the network's capacity.

The findings of the case study are inherently shaped by the specific socioeconomic conditions, land-use configuration, and transport network topology of Villavicencio. Accordingly, they should not be interpreted as evidence that weather-related disruptions intensify social disparities in the same way across all cities. Nevertheless, the proposed methodology is broadly transferable, as it can be implemented in many urban contexts where the input data typically used in demand estimation studies are available. Furthermore, rather than seeking to predict real-world conditions, the case study is intended to illustrate the potential of the proposed methodology to bring social disparities in accessibility to the forefront. In doing so, the methodology supports a more nuanced understanding of how inequalities influence access to vital urban services in normal conditions and how they may be amplified during disruptions such as flooding.

Finally, the policy implications for Villavicencio point to the need for further investigation into how residents from lower socioeconomic strata access healthcare facilities using modes other than private vehicles, particularly public transport, as these groups appear to depend more heavily on such modes for this type of trip. In this context, a prudent priority would be to strengthen transport connectivity under baseline conditions, since access to healthcare already appears more difficult for populations in less favourable circumstances, even in the absence of disruption. This suggests the value of complementing the city's efforts to expand the spatial provision of healthcare facilities across the urban area with a closer assessment of the transport connections serving those locations, as well as of healthcare-related services that remain concentrated in the city centre.

Acknowledgements

The author gratefully acknowledges the *Secretaría de Movilidad de Villavicencio* for authorising the academic use of the information associated with this research and for granting access to the data required to conduct the study.

References

Achillopoulou, D. V., Mitoulis, S.A., Argyroudis, S.A., Wang, Y., 2020. Monitoring of transport infrastructure exposed to multiple hazards: a roadmap for building resilience. Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.2020.141001

Afrin, T., Yodo, N., 2020. A survey of road traffic congestion measures towards a sustainable and resilient transportation system. Sustainability (Switzerland). https://doi.org/10.3390/su12114660

Bruneau, M., Chang, S., Eguchi, R., Lee, G., Rourke, T., Reinhorn, A., Shinozuka, M., Tierney, K., Wallace, W., von Winterfeldt, D., 2003. A Framework to Quantitatively Assess and Enhance the Seismic

Resilience of Communities | Enhanced Reader. Earthquake Spectra 19, 733–752.

Bureau of Public Roads, U.D. of C., 1964. Traffic Assignment Manual.

Camargo, P., 2015. AequilibraE: A free QGIS add-on for transportation modeling. Foss4g North America.

Cretney, R., 2014. Resilience for whom? Emerging critical geographies of socio-ecological resilience. Geogr. Compass 8, 627–640. https://doi.org/10.1111/GEC3.12154

Cutter, S.L., 2016. Resilience to What? Resilience for Whom? Geogr. J. 182, 110–113.

DANE, 2021. Mesa de expertos de estratificación socioeconómica.

European Environment Agency, 2024. Economic losses from weather- and climate-related extremes in Europe [WWW Document].

Farahmand, H., Yin, K., Hsu, C.W., Savadogo, I., Espinet Alegre, X., Mostafavi, A., 2024. Integrating climate projections and probabilistic network analysis into regional transport resilience planning. Transp. Res. D Transp. Environ. 133. https://doi.org/10.1016/j.trd.2024.104229

Faturechi, R., Miller-Hooks, E., 2014. Measuring the Performance of Transportation Infrastructure Systems in Disasters: A Comprehensive Review. https://doi.org/10.1061/(ASCE)IS.1943

Ganin, A.A., Kitsak, M., Marchese, D., Keisler, J.M., Seager, T., Linkov, I., 2017. Resilience and efficiency in transportation networks. Sci. Adv. 3. https://doi.org/10.1126/SCIADV.1701079

Hadjidemetriou, G.M., Herrera, M., Parlikad, A.K., 2022. Planning road network layout based on connectivity assessment of critical assets. Proceedings of the Institution of Civil Engineers: Smart Infrastructure and Construction 175, 181–189. https://doi.org/10.1680/jsmic.21.00034

Li, C., Fang, Q., Ding, L., Cho, Y.K., Chen, K., 2020. Time-dependent resilience analysis of a road network in an extreme environment. Transp. Res. D Transp. Environ. 85. https://doi.org/10.1016/j.trd.2020.102395

Liu, J., Xu, Z., Wan, L., MacAskill, K., 2026. Road transport resilience under extreme rainfall: Integrating multiple impact factors and delay propagation. International Journal of Disaster Risk Reduction 134. https://doi.org/10.1016/j.ijdrr.2026.106024

Martín, B., Ortega, E., Cuevas-Wizner, R., Ledda, A., De Montis, A., 2021. Assessing road network resilience: An accessibility comparative analysis. Transp. Res. D Transp. Environ. 95. https://doi.org/10.1016/j.trd.2021.102851

Ortuzar, J. de D., Willumsen, L.G., 2024. Modelling Transport.

Patriksson, P., 1994. The traffic assignment problem: models and methods.

Rebally, A., Valeo, C., Saidi, S., He, J. (Jennifer), 2026. Assessing the temporal qualities of indirect impacts of flooding on traffic. International Journal of Disaster Risk Reduction 135, 106035. https://doi.org/10.1016/j.ijdrr.2026.106035

Wassmer, J., Merz, B., Marwan, N., 2024. Resilience of transportation infrastructure networks to road failures. Chaos 34. https://doi.org/10.1063/5.0165839

Wu, Y., Chen, S., 2023. Traffic resilience modeling for post-earthquake emergency medical response and planning considering disrupted infrastructure and dislocated residents. International Journal of Disaster Risk Reduction 93. https://doi.org/10.1016/J.IJDRR.2023.103754

Zhang, W., Wang, N., Nicholson, C., 2017. Resilience-based post-disaster recovery strategies for road-bridge networks. Structure and Infrastructure Engineering 13, 1404–1413. https://doi.org/10.1080/15732479.2016.1271813

Zhang, X., Lu, Y., Wang, J., Qi, Y., 2024. Enhancing highway transportation safety resilience during emergencies: A network-based analysis and assessment. PLoS One 19. https://doi.org/10.1371/journal.pone.0307233

Zhang, Y., Cheng, L., 2023. The role of transport infrastructure in economic growth: Empirical evidence in the UK. Transp. Policy (Oxf). 133, 223–233. https://doi.org/10.1016/J.TRANPOL.2023.01.017

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUTOMATIC ROOF TYPE CLASSIFICATION FROM CITYJSON FOR RULE-BASED ROOF STRUCTURE'S MATERIAL QUANTITY ASSESSMENT

Dušan Isailović[1], Aleksandra Parezanović[1], Zorana Petojević[1,2], Nenad Brodić[1]
[1]University of Belgrade, Faculty of Civil Engineering, Serbia;
[2]ZHAW School of Architecture, Design, and Civil Engineering, Zurich, Switzerland

Abstract

Current city-scale material stock assessment relies on archetype-based estimation with limited geometric grounding or on city-to-BIM conversion workflows that are costly and sensitive to representation detail. Both hinder auditable roof typology and quantity derivation from open 3D city models. This paper proposes an interpretable CityJSON pipeline that merges roof meshes into planar patches, extracts roof metrics (area, projection, slopes, span), performs rule-based roof classification (flat/shed/gable/hip/complex), and computes rule-based roof-layer and timber quantities. On 88 labelled buildings, LoD2 achieves 0.716 accuracy (macro-F1 0.543); BIM comparison on six buildings supports feasibility. The contribution is a reproducible, explainable roof-to-quantities method with quantified LoD sensitivity.

Introduction

Cities are increasingly adopting open 3D city models to support planning, permitting, resilience, and environmental assessment. At the same time, circular construction and urban mining require reliable, scalable estimates of material stocks and flows in the existing building stock—especially for elements that dominate mass and replacement cycles. Roofs are a particularly relevant target: they strongly influence building performance and refurbishment needs, they exhibit recurring typologies that relate to structural systems, and they are well represented in city models through explicitly modeled roof surfaces. Yet turning city-model roof geometry into material quantities remains challenging because the data are typically triangulated, vary across levels of detail (LoD), and rarely contain construction semantics beyond coarse surface types.

Most existing city-scale quantity and stock assessment approaches rely on one of two strategies. The first is archetype-based estimation, where building age, use, and generic typologies are used to assign typical assemblies and materials. These methods can scale well, but their accuracy depends heavily on assumptions and weak geometric coupling. The second strategy is a geometry-driven pipeline that extracts measurements from 3D models, then maps them to construction rules. While promising, such pipelines must address practical issues: roof geometry is often fragmented into many small facets, different LoDs represent the same roof differently, and black-box methods (e.g., deep learning) can be difficult to audit when outputs directly drive normative structural selection and material take-offs.

This paper argues that, for city-scale roof material quantification, interpretability and robustness to representational variability are as important as raw classification accuracy. Instead of relying on end-to-end learning, we propose a deterministic workflow that explicitly links roof geometry to roof type labels and quantity-relevant metrics, enabling traceable propagation from input model to material quantities. The workflow uses CityJSON as an open and lightweight representation and focuses on (i) merging triangulated roof faces into planar roof patches, (ii) extracting roof metrics needed by rule-based quantity models, and (iii) classifying roof types using explainable geometric rules. Because LoD can substantially alter planar fragmentation and rule triggers, we explicitly evaluate the same classifier on both LoD2 and LoD3 representations of the same area to quantify typology stability and to support guidance on which LoD is more appropriate for operational quantity pipelines.

The method is demonstrated on a Belgrade test area with 88 manually labelled buildings. In addition, a subset of six buildings is reconstructed in BIM to provide baseline take-offs for validating rule-based roof-layer and timber quantities. Beyond reporting classification performance, we analyze how roof typology errors affect downstream material estimates—showing that some materials (e.g., timber) are relatively robust to subtype ambiguity, while others (notably concrete in flat-roof archetypes) are highly sensitive to flat/non-flat misclassification due to archetype switching.

Related work

3D city models as a basis for urban-scale assessment

Open 3D city models (CityGML/CityJSON) are increasingly used as a geometric backbone for city-scale analyses because they provide consistent building envelopes and semantic surface classes (e.g., RoofSurface) that can be processed automatically. CityJSON in particular has been positioned as a lightweight, developer-friendly encoding of CityGML concepts for practical processing workflows (Ledoux, et al., 2019). The Levels of Detail (LoD2, LoD3) used in such models follow the conceptual definitions of CityGML 3.0, where LoD2 represents generalized roof structures and LoD3 introduces detailed architectural

features such as dormers and roof superstructures (Open Geospatial Consortium, 2026).

Large-scale LoD2/LoD3 building models are commonly generated from airborne LiDAR and photogrammetric data using automated reconstruction pipelines that detect planar roof segments and infer roof topology (Haala & Brenner, 1999). These workflows often assign roof types explicitly as dataset attributes (e.g., *roofType* in CityGML), reflecting the inferred geometric archetype of the roof.

However, such attributes are not consistently available across datasets, are derived using dataset-specific assumptions, and are typically not designed for construction-oriented interpretation or material quantity estimation. As a result, relying directly on dataset-provided roof types may limit reproducibility and applicability across heterogeneous data sources.

Roof type inference from city models and remote sensing

Roof typology inference has been studied from multiple perspectives. A line of work targets roof shape classification for energy and solar potential assessment using machine learning on roof features, showing that roof-type prediction is feasible but sensitive to the quality and representativeness of training data and features (Mohajeri, et al., 2018). Other research infers roof type from sparse or derived building attributes, demonstrating that the "flat vs non-flat" decision can often be made robustly and that a small set of geometric attributes can be highly informative (Biljecki & Dehbi, 2019). Recent datasets and studies also highlight how roof typologies can be learned from CityJSON-labelled corpora, reporting roof classification performance in the 70–75% range across multiple European cities—comparable to the performance envelope observed in practical city-scale pipelines (Sitong, Chengzhi, Chi, & Qiuxian, 2023).

In parallel, deep learning approaches—often driven by aerial imagery—have been used to classify roof types (flat, gable, hip, etc.), benefiting from large training sets but typically offering limited traceability between the decision and the 3D geometry that will later drive construction quantities (Buyukdemircioglu, Can, & Kocaman, 2021). For applications such as material quantity estimation, where classification outcomes directly affect structural assumptions, interpretability becomes a critical requirement.

BIM–GIS integration and material stock/quantity assessment

A growing body of work addresses BIM–GIS integration for circular construction and city-scale material stock assessment. Recent studies combine remote sensing, 3D GIS and BIM to improve material estimation at building and element levels (Parezanović, Nadaždi, Isailović, Višnjevac, & Petojević, 2025). In the EC3 community specifically, approaches have also been proposed for converting 3D city models to BIM to support automated quantity take-offs (Dušan Isailović, Ana Nadaždi, Aleksandra Parezanović, Zorana Petojević, & Nenad Višnjevac, 2025). Complementary review work highlights interoperability challenges and the importance of consistent semantics in CityGML-based environments (Janisio-Pawłowska & Pawłowski, 2024).

At the urban scale, material stock datasets and frameworks commonly rely on city model geometry (footprints, heights, LoDs) combined with building attributes (age, typology, structural system) to compute material quantities (Pei, Biljecki, & Stouffs, 2022). While these approaches scale efficiently, they often rely on weak geometric coupling or require costly BIM conversion workflows.

Positioning of this paper

Across the literature, roof type inference is often treated as an intermediate step within 3D reconstruction pipelines or as a standalone classification task for applications such as solar potential assessment. In many operational datasets, roof types are already inferred during LiDAR- or photogrammetry-based reconstruction and may be available as attributes. However, these attributes are not consistently available, are dataset-specific, and are not designed for construction-oriented interpretation or material quantity estimation.

The novelty of this work lies not in roof classification alone, but in integrating roof plane extraction, geometric metric derivation, typology inference, and rule-based material quantity estimation into a single interpretable and auditable workflow. In this context, roof typology serves as a bridge between geometric representation and construction semantics, enabling traceable propagation from city-model geometry to material quantities.

Furthermore, this work explicitly investigates the sensitivity of roof typology inference to different Levels of Detail (LoD2 vs LoD3), highlighting how geometric representation affects both classification stability and downstream quantity estimation. This connection between representation, semantics, and quantities remains underexplored in existing literature.

Method

The proposed workflow is an interpretable, city-scale pipeline that derives roof typology and quantity-relevant metrics from an open 3D city model. For each building, we:

1. Extract roof surfaces from CityJSON,
2. Merge triangulated roof faces into planar roof patches ("roof planes"),
3. Derive a roof-plane adjacency representation from shared edges,
4. Compute roof metrics required for material quantity rules,
5. Classify roof type using explicit geometric rules,
6. Map the classified roof type and span proxy to a normative structural archetype and compute timber quantities, and
7. Compute additional roof-layer quantities from roof area/projection using rule tables.

The workflow is executed primarily on LoD2; LoD3 is used to discuss representation-induced fragmentation effects and the resulting differences in typology stability.

The workflow uses:

- Cityjson lod2 (primary) and lod3 (comparative) building models for the Belgrade test area;
- Rule-based roof classifier implementation developed from the original roof_classifier.py into the lod2-selected version;
- Manual roof-type labels for evaluation; and
- Normative tables and rules for timber roof structure selection and quantity factors, plus roof-layer material rules.

The workflow produces:

- Roof type label (roof_type_automatic) and QA indicators;
- Roof metrics (roof area, horizontal projection area, slope statistics, plane count, span proxy);
- Normative structural type;
- Timber volume estimate;
- Roof-layer quantities; and
- For a BIM-modeled subset, a comparison against BIM-derived baseline take-offs.

For each building, roof faces are selected primarily using CityJSON semantics (e.g., RoofSurface), ensuring the method remains consistent with semantic organization where available. When semantics are incomplete, a geometric fallback can be applied by selecting upward-facing faces within the upper elevation band of the building envelope (reported as a limitation and used only when necessary).

Planar roof patch (roof plane) extraction

City models typically represent roofs as triangulated meshes; direct triangle-level reasoning is unstable and inflates the number of planes. Therefore, roof faces are merged into larger planar patches ("roof planes") using a deterministic clustering procedure:

1. Per-face geometry: for each face, compute a unit normal vector, face area, and a representative point (e.g., centroid).
2. Plane merging: faces are clustered into planes using (a) normal similarity (angle between normals below a tolerance) and (b) coplanarity (distance from the face representative point to the candidate plane below a tolerance). This normal + coplanarity merging prevents parallel-but-offset patches from being merged, which is important for detailed models where small roof features can create offset but parallel surfaces.
3. Plane attributes: for each roof plane, compute total area (sum of merged face areas), area-weighted mean normal, and slope angle relative to horizontal.

Roof-plane adjacency representation and ridge/valley proxies

To capture roof topology beyond plane counts and slopes, plane adjacency is derived from shared edges between faces. An edge map is created over roof faces; each undirected edge is stored using a canonical orientation (e.g., smaller vertex index first) to make adjacency deterministic. Two planes are adjacent if any face from plane A shares an edge with any face from plane B. Shared-edge lengths are accumulated per plane pair and aggregated per roof.

For each shared edge between two planes, a dihedral sign proxy is computed using plane normals and a consistently oriented edge vector. Edges classified as convex act as ridge proxies and concave edges act as valley proxies. These proxies support explainable complexity assessment and can be used for detailing-related quantity rules (e.g., ridge/valley flashing length), even if the present paper focuses primarily on area-driven roof layers and structural timber.

Roof metrics extraction

For each building roof, we compute a set of metrics used both for classification and for quantity estimation:

- Roof surface area: sum of roof plane areas.
- Horizontal projection area: projection of roof surfaces onto the horizontal plane; implemented to support polygon holes (outer ring minus inner rings) where applicable.
- Slope metrics: maximum slope of the steepest plane (and optional area-weighted summaries).
- Plane count: number of roof planes after merging.
- Span proxy: derived from the building footprint, obtained by selecting the largest near-base horizontal ring and computing a conservative span proxy (typically the shorter side of the footprint bounding rectangle).

Quality indicators (QA) are computed to support interpretation (especially for LoD3): non-manifold edge counts and tiny-plane statistics (number and area ratio of very small planes). These indicators do not change the LoD2 operational method, but help explain LoD3 fragmentation effects.

Rule-based roof type classification

Roofs are classified into a construction-oriented typology set: gable, shed, complex, flat, and hip (Figure 1). The classifier is intentionally rule-based to remain interpretable and auditable at city scale and to support downstream normative mapping and material computations.



Figure 1: Roof types for automatic classification.

The concept used for the decision logic is as follows:

1. Detect flat roofs based on near-horizontal slope conditions (including robust handling of small steep facets);
2. Detect shed roofs by a single dominant pitch/orientation;
3. Detect mansard roofs by the coexistence of shallow and steep planes above area thresholds;
4. Separate gable and hip using plane orientation dominance and adjacency structure; and
5. Assign complex as a conservative fallback when rules are inconclusive or roofs are irregular/fragmented.

LoD2 is used as the primary input because its roof geometry is typically less fragmented and more stable for plane merging and rule triggers. LoD3 introduces additional small facets (dormers, chimneys, parapets), inflating plane counts and disturbing dominance logic, therefore LoD3 is treated as comparative evidence.

Mapping roof types to normative structural classes

To estimate timber quantities, each roof is mapped to a normative structural class based on roof type and the span proxy L. The mapping follows normative span intervals and a deterministic tie-breaking strategy: where multiple classes cover the same span, the class with minimum timber consumption is selected; gaps are resolved by selecting the next stronger admissible class; spans above the maximum use the strongest available class as an upper approximation.

For hip roofs, timber consumption is approximated using gable-roof normative rules at this level of abstraction. Complex roofs are treated as gable-equivalent for baseline normative selection by span, with an additional correction factor representing geometric complexity (derived from concave-edge intensity proxies).

Roof structure's material quantity estimation

Timber volume is computed as:

$$V_{timber} = A_{proj} \times r_{norm} \times k \quad (1)$$

where A_{proj} is the horizontal projection of the roof's area, r_{norm} is the normative timber consumption rate (m³ per m² of horizontal projection) for the selected class, and k is a correction factor (k = 1 by default; k > 1 for complex roofs based on complexity bins).

This formulation couples geometry-derived area and span to normative structural selection and complexity-aware adjustments.

Concrete quantity in flat roof structures is computed much more straightforward, as:

$$V_{concrete} = A_{roof} \times d_{slab} \quad (2)$$

where A_{roof} is the roof's area, and d_{slab} is the concrete slab thickness.

Pseudocode below provides the step-by-step operational detail.

FOR each Building in CityJSON:

1) Extract roof faces

- *Prefer semantic RoofSurface faces*
- *Else fallback: upward-facing faces near top elevation*

2) Compute per-face geometry

- *normal, area, centroid, horizontal projection area (holes supported)*

3) Merge faces into roof planes

- *cluster by (normal angle tolerance) AND (coplanarity distance tolerance)*
- *compute plane area + plane slope*

4) Build plane adjacency + ridge/valley proxies

- *construct deterministic shared-edge map (canonical edge key)*
- *for each shared edge between different planes:*
  - *accumulate shared-edge length*
  - *classify as convex (ridge proxy) or concave (valley proxy)*

5) Compute roof metrics + QA

- *roof area, projected area, max slope, plane count*
- *footprint span proxy from largest near-base horizontal ring*
- *QA: non-manifold edges, tiny-plane statistics*

6) Classify roof type (rule-based)

- *decide flat vs non-flat from slope*
- *detect shed / gable / hip using plane counts, dominant plane areas,*
- *slope distribution, concave ratio, and footprint aspect ratio*
- *else fallback: complex*

7) Export per-building results (roof_type + metrics)

Implementation notes and reproducibility

All processing is deterministic with explicit parameter thresholds (e.g., plane normal tolerance, coplanarity tolerance, minimum shared-edge length, and slope cutoffs). Intermediate outputs (roof planes, QA indicators, roof metrics) are exported alongside final labels and quantities to support diagnosis and sensitivity analyses, enabling reproducibility and auditability.

Data and case study

Study area and datasets

The case study is conducted on a dense, mixed building block in Belgrade (Serbia), selected as a representative urban fabric with a wide variety of roof geometries (flat roofs, mono-pitch roofs, and multi-patch roofs with local superstructures). The study area is used to demonstrate an end-to-end workflow for:

- Automatic roof type inference from an open 3D city model,
- Roof metrics extraction, and
- Rule-based material quantity assessment for roof layers and the timber roof structure.

Two CityJSON datasets are used for the same spatial extent: a LoD2 model, where roof geometry is represented with a cleaner set of roof facets, and a LoD3 model, which additionally encodes façade openings and finer roof

superstructures (Figure 2). While the classifier can be executed on both datasets, LoD2 is treated as the primary operational input, and LoD3 is used to illustrate how increased geometric detail can affect planar fragmentation and typology stability.

The evaluation dataset contains 88 buildings with manually assigned roof type labels. A subset of six buildings was further modelled in BIM to obtain baseline take-offs for validating roof structure quantities and timber volumes (Figure 3).

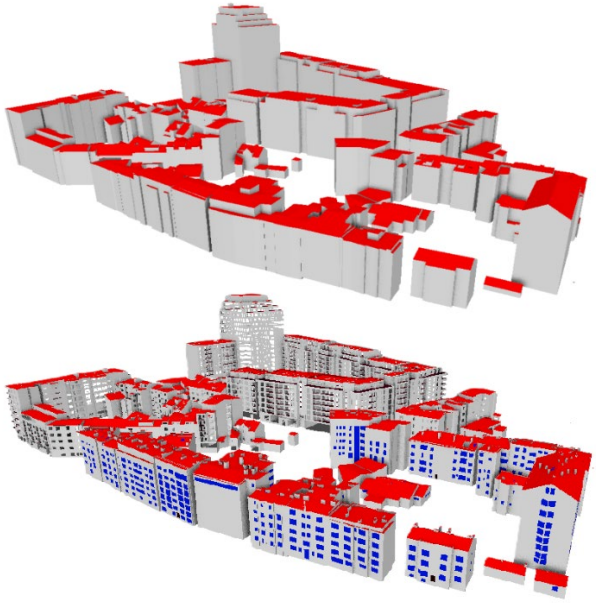


Figure 2: 3D City models of the case area at various level of detail: CityJSON models at LoD2 (top), and LoD3 (bottom).

3D city models (CityJSON LoD2 and LoD3)

Both city models follow the CityJSON data model and represent buildings as boundary representations with triangulated surfaces. Roof faces are semantically identifiable (RoofSurface) and are used as the primary input for plane merging, metric extraction, and roof type classification. The two levels of detail differ mainly in the degree of geometric fragmentation: LoD2 tends to express roofs as a smaller number of planar patches, whereas LoD3 introduces additional small facets (e.g., dormers, chimneys, parapets), which increases the number of planes after merging and may perturb dominance-based rules. This LoD effect is reported in the Results section through LoD2–LoD3 prediction agreement–

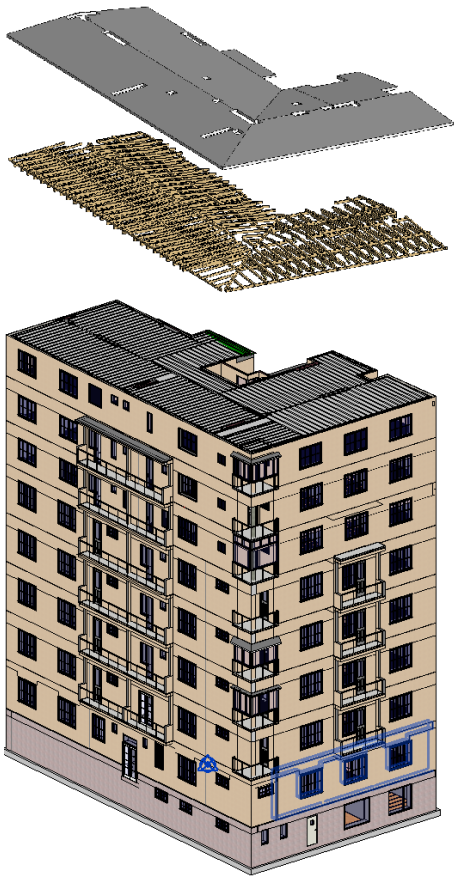


Figure 3: BIM model of a residential building from the case area needed for validation of a roof structure's material quantity computation.

Results

Roof type classification

The evaluation uses 88 labelled buildings from the Belgrade test area. Using CityJSON LoD2 as input, the classifier achieves an overall accuracy of 0.716 and a macro-averaged F1-score of 0.543. Running the same code on the LoD3 model yields an accuracy of 0.557 and macro-F1 of 0.403. Agreement between LoD2 and LoD3-based predictions is 0.534 (Cohen's $\kappa = 0.375$), indicating that the choice of LoD representation affects typology stability even when the algorithm is unchanged.

Table 1 (LoD2) and Table 2 (LoD3) report per-class precision/recall/F1 against manual labels, while Table 3 and Table 4 provide the corresponding confusion matrices. Across both LoDs, errors concentrate on (i) flat vs non-flat ambiguity and (ii) gable vs complex ambiguity.

Extracted roof metrics and their role in the pipeline

For each building, the workflow reports plane count, maximum roof slope, roof area, horizontal projection area, and a footprint span proxy. These metrics provide the inputs to the material quantity rules and enable explainable links between geometry, typology, and quantities.

Table 1: Roof type classification performance using CityJSON LoD2 as input (automatic predictions vs manual labels). Support denotes the number of labelled buildings per class; macro-averages are computed across classes.

Roof type	Support (n)	Precision	Recall	F1
flat	16	1.000	0.562	0.720
shed	24	1.000	0.833	0.909
gable	12	0.400	0.333	0.364
hip	1	0.000	0.000	0.000
complex	35	0.625	0.857	0.723
Macro avg	88	0.605	0.517	0.543
Overall accuracy	88			0.716

Table 2: Table 2. Roof type classification performance using CityJSON LoD3 as input (same code and rules; automatic predictions vs manual labels). Differences relative to Table 1 quantify sensitivity to LoD representation.

Roof type	Support (n)	Precision	Recall	F1
flat	16	0.643	0.562	0.600
shed	24	0.647	0.917	0.759
gable	12	0.000	0.000	0.000
hip	1	0.000	0.000	0.000
complex	35	0.737	0.400	0.519
Macro avg	88	0.405	0.376	0.403
Overall accuracy	88			0.557

Table 3: Confusion matrix for LoD2-based classification (rows: manual labels; columns: automatic predictions).

Pred flat	Pred shed	Pred gable	Pred hip	Pred complex
9	0	0	0	7
0	20	2	0	2
0	0	4	0	8
0	0	0	0	1
0	0	4	1	30

Table 4: Confusion matrix for LoD3-based classification (rows: manual labels; columns: automatic predictions).

Pred flat	Pred shed	Pred gable	Pred hip	Pred complex
9	5	0	0	2
0	22	0	0	2
0	9	0	0	3
0	1	0	0	0
5	6	0	0	24

Implications for material quantities

Using identical quantity rules, the mean absolute difference between automatic and manual estimates is 4.32 m³ for timber (mean relative error 0.033) and 7.80 m³ for concrete (mean relative error 0.385, computed where manual concrete > 0). Table 6 summarizes this sensitivity.

Concrete quantities are particularly sensitive to the flat/non-flat decision: 5 of 14 manually labelled flat roofs are classified as non-flat in the LoD2-based predictions and 8 of 14 in the LoD3-based predictions (Table 5). Because concrete quantities are only computed for flat-roof structural archetypes, these misclassifications force the concrete estimate to zero and therefore dominate concrete errors.

For the BIM validation subset (n=6), Table 7 compares automatic and manual rule-based quantities against BIM-derived baselines. Across buildings where a BIM baseline is available, the mean relative error of the automatic estimate is 0.647 for timber and 0.600 for concrete (where BIM baseline > 0). Remaining discrepancies are expected to reflect both typology errors and differences in modelling scope between rule-based archetypes and BIM take-offs.

Table 5: Agreement between LoD2- and LoD3-based automatic roof type predictions, including flat-roof sensitivity. Cohen's κ complements exact agreement by adjusting for chance agreement.

Metric	Value
Exact agreement (LoD2 vs LoD3 predictions)	0.511
Cohen's κ (LoD2 vs LoD3 predictions)	0.335
Flat roofs misclassified (LoD2 predictions)	1/10
Flat roofs misclassified (LoD3 predictions)	3/10

Table 6: Sensitivity of rule-based material quantities to roof type classification (automatic vs manual labels, same quantity rules). Concrete quantities are evaluated only for roofs mapped to flat-roof structural archetypes.

Material / group	MAE (m³)	Mean rel. error	Interpretation
Timber (all buildings)	3.22	0.023	Stable for pitched roofs; differences mainly driven by typology/normative-type mapping
Concrete (flat-roof archetypes only)	15.48	0.400	Highly sensitive to flat/non-flat decision; misclassification forces concrete estimate to 0

Table 7: BIM validation subset (n=6): automatic and manual rule-based quantities compared against BIM-derived baseline take-offs. Remaining differences may reflect both typology errors and modelling-scope differences (e.g., inclusion of secondary timber members).

Building ID	Manual roof type	Auto roof type (LoD2 pipeline)	# planes	Max slope (°)	Timber Auto (m³)	Timber Manual (m³)	Timber BIM (m³)	Concrete Auto (m³)	Concrete Manual (m³)	Concrete BIM (m³)
Cv_90	complex	complex	5.0	6.54	22.87	22.87	19.28	0.0	0.0	0.0
Dl_55	gable	complex	6.0	32.85	15.64	17.37	16.48	0.0	0.0	0.0
StGl_34	gable	complex	5.0	29.83	13.17	10.98	4.07	0.0	0.0	0.0
StGl_36	complex	complex	9.0	45.91	16.47	16.47	18.57	0.0	0.0	0.0
Dl_28	flat	complex	6.0	18.16	36.23	0.0	0.0	0.0	68.11	67.14
Dl_57	flat	flat	2.0	2.54	0.0	0.0	0.0	8.42	8.42	10.52

Discussion

Interpreting classification performance and error structure

The results suggest that an interpretable, rule-based classifier can support operational roof typology inference at city scale when applied to a semantically structured LoD2 CityJSON model. On the 88-building dataset, LoD2 classification achieves 0.716 overall accuracy with a macro-averaged F1 of 0.543, indicating that several roof types exhibit distinctive planar signatures, while performance remains uneven across classes.

Two ambiguity families dominate. Flat vs non-flat errors arise when small steep facets on otherwise flat roofs (e.g., parapets or technical housings) affect slope-based triggers even after robustness filtering. Gable vs complex errors occur for multi-part roofs (annexes, terraces, multiple ridges), where "complex" acts as a conservative fallback suggesting that a single roof archetype may be insufficient without segmentation.

Why LoD2 is operationally preferable, and what LoD3 adds

LoD2 is operationally preferable because it preserves the primary roof shape with less fragmentation, while LoD3 adds superstructures and small facets that perturb dominance-based rules. This representational sensitivity is visible in the LoD2–LoD3 agreement statistics (0.511 exact agreement, $\kappa = 0.335$), indicating only moderate stability when the algorithm is unchanged but the geometric representation differs. LoD3 is therefore best treated as diagnostic evidence motivating generalization/feature suppression or a "structural roof surface" abstraction.

Effects of typology errors on material quantities

Typology errors propagate asymmetrically into quantities. Timber estimates are comparatively stable because most non-flat types map to pitched-roof families and are driven primarily by projection area and span-based normative selection; this is reflected in a timber MAE of 3.22 m³ (mean relative error 0.023) between automatic- and manual-driven estimates under identical quantity rules.

Concrete is highly sensitive because it is computed only for flat-roof archetypes; misclassifying a flat roof as non-flat forces the concrete estimate to zero and creates step-change errors. Accordingly, concrete shows a larger MAE of 15.48 m³ (mean relative error 0.400, computed where manual concrete > 0). For city-scale accounting, improving the flat/non-flat decision (e.g., area-weighted flatness indicators rather than max-slope triggers) is likely to yield larger downstream gains than refining pitched-roof subtypes.

Interpretable rules vs. end-to-end deep learning

The rule-based approach is appropriate here because it provides traceability from geometry → typology → archetype → quantities, is data-efficient when labelled city-scale roof datasets are limited, and aligns naturally with normative quantity rules. Deep learning remains a useful complement, e.g., as a lightweight model applied only to ambiguous cases (near flat thresholds or "complex" candidates) or as a calibration layer for threshold tuning.

Limitations and targeted improvements

Three limitations follow from the error modes: (1) single label per building roof, which motivates roof-instance segmentation; (2) fragmentation sensitivity, suggesting area-weighted flatness indicators rather than max-slope triggers; and (3) validation scope, since the BIM subset (n=6) supports feasibility but should be expanded for stronger generalisation claims.

Conclusions

This paper presented an interpretable CityJSON-based workflow for automated roof type classification and roof metrics extraction, enabling rule-based estimation of roof material quantities at city scale. Using a Belgrade test area, the approach demonstrated that a rule-driven classifier can provide reliable roof typology inference when applied to a semantically structured LoD2 city model, while simultaneously producing the geometry-derived parameters required for downstream quantity

rules (roof areas, projected areas, slope statistics, plane counts, and a footprint-based span proxy).

The results show that classification errors are structured rather than random. Misclassifications concentrate on flat versus non-flat ambiguity caused by small steep facets on otherwise flat roofs and gable versus complex ambiguity for irregular or multi-part roofs. Importantly, the study quantified how such typology errors propagate to quantity estimation: timber volume estimates remain comparatively stable across non-flat typologies, whereas concrete quantities are highly sensitive to flat roof recall because flat/non-flat decisions control whether the flat-roof archetype branch is applied at all.

A key methodological takeaway is that LoD selection matters. While the same classifier can be executed on LoD3, LoD2 provides a more stable basis for city-scale roof typology inference due to reduced geometric fragmentation. LoD3 is therefore best interpreted as a complementary dataset for diagnosing representation-induced instability and motivating generalization and segmentation strategies.

Future work

Future work will prioritize:

1. Roof instance segmentation to handle buildings with multiple roof components (main roof vs annex), reducing “complex” fallbacks and improving quantity mapping;
2. Lod3 generalization to suppress/aggregate superstructure facets into a stable “structural roof surface” prior to classification; and
3. Improved flat-roof robustness using area-weighted flatness indicators to reduce flat/non-flat errors that dominate concrete quantities.
4. In addition, we will expand validation beyond six BIM baselines and report uncertainty via sensitivity analysis over key thresholds and normative mappings.

Acknowledgments

This research was funded by the European Union's Horizon Europe research and innovation programme under grant agreement No 101082068 (CircBoost project).

References

Biljecki, F., & Dehbi, Y. (2019, September). Raise the roof: towards generating lod2 models without aerial surveys using machine learning. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, IV-4/W8, 27–34. doi:10.5194/isprs-annals-IV-4-W8-27-2019

Buyukdemircioglu, M., Can, R., & Kocaman, S. (2021, June). Deep learning based roof type classification using very high resolution aerial imagery. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B3-2021, 55–60. doi:10.5194/isprs-archives-XLIII-B3-2021-55-2021

Dušan Isailović, D. I., Ana Nadaždi, A. N., Aleksandra Parezanović, A. P., Zorana Petojević, Z. P., & Nenad Višnjevac, N. V. (2025, July). 3d city model conversion to bim for automatic material quantities assessment. doi:10.35490/EC3.2025.378

Janisio-Pawłowska, D., & Pawłowski, W. (2024, January). Implementation of bim data in citygml—research and perspectives for creating a qgis plugin for spatial analysis: experience from poland. Sustainability, 16, 642. doi:10.3390/su16020642

Ledoux, H., Arroyo Ohori, K., Kumar, K., Dukai, B., Labetski, A., & Vitalis, S. (2019, December). CityJSON: a compact and easy-to-use encoding of the CityGML data model. Open Geospatial Data, Software and Standards, 4, 4. doi:10.1186/s40965-019-0064-0

Mohajeri, N., Assouline, D., Guiboud, B., Bill, A., Gudmundsson, A., & Scartezzini, J.-L. (2018, June). A city-scale roof shape classification using machine learning for solar energy applications. Renewable Energy, 121, 81–93. doi:10.1016/j.renene.2017.12.096

Open Geospatial Consortium. (2026, April 7). OGC City Geography Markup Language (CityGML) Part 1: Conceptual Model Standard, Version 3.0, OGC Doc No. 20-010, 2021. Retrieved from OGC: https://docs.ogc.org/is/20-010/20-010.html

Parezanović, A., Nadaždi, A., Isailović, D., Višnjevac, N., & Petojević, Z. (2025, April). Mapping the urban building stock for a circular economy by integrating GIS and BIM. A case study from Belgrade, Serbia. Resources, Conservation and Recycling, 215, 108075. doi:10.1016/j.resconrec.2024.108075

Pei, W. Y., Biljecki, F., & Stouffs, R. (2022, October). Dataset for urban scale building stock modelling: identification and review of potential data collection approaches. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4/W2-2022, 225–232. doi:10.5194/isprs-annals-X-4-W2-2022-225-2022

Sitong, L., Chengzhi, R., Chi, Z., & Qiuxian, W. (2023). Building and roof type classification on 3D city models. Building and roof type classification on 3D city models. Retrieved February 8, 2026, from https://repository.tudelft.nl/record/uuid:585100a5-f696-487f-a197-858ffa422f26

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

EXPLORING FLOOD VULNERABILITY PATTERNS AND RESIDENTIAL FLOOD IMPACTS IN URBAN ENVIRONMENT

Hyun Jeong Koo[1] and Vassiliki Demetracopoulou[2]
[1]Wayne State University, United States of America
[2]Rutgers University

Abstract

This study analyzes FEMA flood insurance claims and Zillow housing data in Detroit, United States, using ZIP-code-level metrics and damage-based quartiles to identify neighborhood-scale spatial patterns. Results show statistically significant differences in housing price and claim frequency across damage quartiles, with higher-damage ZIP codes characterized by lower housing values and more frequent flood claims. Housing characteristics showed moderate correlations with flood damage but were insufficient preditors in regression models. By jointly examining flood damage and housing characteristics at the neighborhood scale, this study advances understanding of residential flood vulnerability and provides evidence to support more informed planning and mitigation efforts.

Introduction

Urban flooding has increased in frequency and severity in recent decades, driven by climate change, urbanization, and aging infrastructure, with wide-ranging impacts on transportation systems, utilities, and housing. In the United States, older urban centers are particularly vulnerable due to legacy stormwater systems, high impervious surface coverage, and development patterns that predate modern flood management standards (US EPA, 2021; ASCE, 2025). These conditions contribute to recurrent flood damage, including roadway inundation, service disruptions, and residential property losses.

Vulnerable and low-income communities are often disproportionately affected by urban flooding, as housing characteristics such as older structures, basement units, and limited floodproofing intersect with a reduced capacity to afford repairs or mitigation measures (Cutter et al., 2003; Wing et al., 2022). As a result, urban flooding represents not only an infrastructure resilience challenge but also a mechanism through which existing social and economic inequities are intensified.

Despite increasing attention to urban flooding and social vulnerability, empirical research remains limited on how flooding impacts intersect with housing and residential characteristics at the neighborhood scale. This study addresses this gap by examining relationships between flooding indicators and housing characteristics in Detroit, Michigan, using FEMA flood insurance claims and Zillow housing data. Detroit provides a critical case for analyzing urban flooding dynamics due to the combined effects of aging infrastructure, extensive impervious surfaces, climate-driven precipitation extremes, and persistent socioeconomic vulnerability. By jointly analyzing flooding indicators, such as total flood damage derived from FEMA claims, and housing characteristics, such as average housing price and built year obtained from Zillow, this research advances neighborhood-scale understanding of urban flood vulnerability and supports more equitable flood mitigation and housing resilience strategies.

Literature review

Urban flooding is primarily driven by short-term or continuous rainfall that exceeds the capacity of the drainage system (Liu et al., 2020; Qi et al., 2020). Unlike traditional flood hazards, urban flooding is strongly shaped by land-use patterns, impervious surface coverage, drainage system design, and infrastructure condition. Climate change has further intensified urban flooding by increasing the frequency and magnitude of short-duration, high-intensity precipitation events (IPCC, 2021). Numerous studies have documented rising flood-related damages in U.S. cities, including transportation disruptions, property losses, and repeated service outages (Wing et al., 2022).

A substantial body of literature has demonstrated that flood impacts are unevenly distributed across populations. Social vulnerability theory emphasizes that socioeconomic characteristics, such as income, race, age, and housing tenure influence exposure, sensitivity, and recovery capacity during hazard events (Cutter et al., 2003; Cutter et al., 2008). In the context of flooding, lower-income and marginalized communities often face higher exposure to flood hazards and experience greater difficulty recovering from flood-related losses.

Recent large-scale studies have strong evidence of systematic inequities in flood risk and damage across the United States. Wing et al. (2022) showed that socially vulnerable populations are disproportionately exposed to flood risk, even after accounting for physical flood hazard condition. Kousky (2018) further noted that limited access to insurance, savings, and credit constrains post-flood recovery, leading to cumulative and compounding

losses over time. These inequities are particularly pronounced in cities such as Detroit, where historical disinvestment and infrastructure decline intersect with contemporary climate risks.

Housing characteristics play a critical role in shaping flood vulnerability, yet they remain underexplored relative to broader socioeconomic indicators. Prior research has shown that older housing stock is often more susceptible to flood damage due to aging materials, basement configurations, and limited incorporation of flood-resilient design features (Elliott & Clement, 2017; Tate et al., 2021). Older homes are also more likely to be served by combined sewer systems, increasing exposure to sewer backups and surface flooding during heavy rainfall events.

Housing value has been identified as an important indicator for both structural condition and adaptive capacity. Lower-value housing is often associated with deferred maintenance, limited floodproofing, and reduced access to resources for mitigation or repair (FEMA, 2023; Kousky, 2018). Several studies have documented higher relative flood losses in lower-income or lower-value housing areas, even when controlling for hazard intensity (Wing et al., 2022; Shokry et al., 2022). However, much of this work relies on modeled flood exposure or survey-based vulnerability indices rather than observed damage data.

Despite growing recognition of the role of housing in flood vulnerability, relatively few studies have empirically linked observed flood damage with neighborhood-scale housing characteristics using claims-based data. Existing research often focuses on floodplain exposure, hydrologic modeling, or county-level socioeconomic indicators, which may obscure localized patterns of residential vulnerability within cities (Tate et al., 2021). Moreover, while FEMA's National Flood Insurance Program (NFIP) data have been used extensively to study flood losses, they are less frequently combined with independent housing datasets to examine structural and economic housing attributes at spatial scales.

This study addresses this gap by examining how observed flood losses relate to housing characteristics in a legacy urban context, using FEMA flood insurance claims and housing data in the City of Detroit.

Research methodology

To achieve the goal of this study, two primary datasets were used: (1) FEMA National Flood Insurance Program (NFIP) residential claims for the City of Detroit, which provide information on total building and contents losses for flood-affected properties, and (2) publicly available housing data from Zillow that include parcel-level characteristics such as built year, estimated housing value, and selected structural attributes. These datasets were selected because they offer complementary perspectives. NFIP claims capture observed flood damage and loss outcomes, while Zillow housing data provides descriptive indicators of neighborhood housing stock, including housing prices and the year of construction. The Zillow housing data were obtained in 2024.

Both datasets underwent systematic preprocessing before analysis. NFIP records were screened to remove entries with incomplete flood damage information. Housing data were similarly cleaned by excluding parcels with missing year built, estimated value, or location attributes. Duplicate records, unmatched parcels, and entries lacking ZIP code identifiers were removed to ensure consistency and data quality across datasets.

After preprocessing, the two datasets were spatially joined using ZIP code identifiers, enabling aggregation of parcel-level housing attributes with ZIP-level flood damage data (Table 1). This process produced a unified dataset suitable for neighborhood-scale analysis while retaining comparability of housing characteristics (e.g., year built and housing value) across geographic areas. ZIP codes were then grouped into quartiles based on total flood damage, representing increasing levels of flood impact. This quartile-based stratification enabled systematic comparisons across damage-intensity groups and supported subsequent statistical analyses.

Table 1: Zip-code-level variables dataset

Variables	Definition	Source
Total Flood Damage	Sum of building and contents losses for claims in each ZIP Code	FEMA
Number of Losses	Count of residential claims per ZIP code	FEMA
Water Level	Average reported flood water depth associated with residential claims	FEMA
Average Housing Price	Mean Zillow housing value estimate per Zip code	Zillow
Average Built Year	Mean built year of residential parcels per ZIP code	Zillow
Average Frontage Length	Mean frontage length for parcels in each Zip code	Zillow

The study employed a multi-step analytical framework to examine relationships between neighborhood housing characteristics and flood losses. First, correlation analysis quantified linear associations between flood damage and key variables, including water level, housing price, and year built. Second, one-way analysis of variance (ANOVA) tests assessed whether housing characteristics differed significantly across four damage-based quartile groups. Third, regression models were estimated to evaluate the explanatory power of housing attributes in predicting total flood damage. Finally, an unsupervised machine learning approach (i.e., clustering analysis) was applied to identify data-driven neighborhood groupings based on flood damage and housing characteristics.

This integrated methodological approach supported both exploratory and inferential analyses, enabling a structured assessment of the relationships between neighborhood housing characteristics and flood impacts at the neighborhood scale.

Data analysis and results

Total flood damage and the number of NFIP residential claims varied substantially across Detroit ZIP codes, indicating pronounced spatial heterogeneity in flood impacts. Descriptive analyses indicated that ZIP codes with greater flood damage generally contained older, lower-value housing, suggesting that housing age and value may be associated with disproportionate flood impacts. These patterns were subsequently examined using correlation analysis, one-way ANOVA, regression modeling, and K-means clustering.

(1) Correlation analysis

Pearson correlation coefficients were calculated to assess the strength and direction of relationships between total flood damage and housing-related attributes (Table 2). The analysis revealed no strong linear relationship between average water level and total damage ($r = -0.148$), indicating that water depth alone does not explain ZIP-level differences in losses. This finding is consistent with prior research indicating that factors such as housing condition, age of the housing stock, drainage performance, and neighborhood infrastructure can substantially influence flood impacts beyond hydrologic intensity alone (Cutter et al., 2008; Kousky, 2018; Wing et al., 2020).

A moderate negative correlation between housing price and total damage ($r = -0.379$) indicates that lower-value housing areas experienced higher cumulative flood losses. This pattern suggests that neighborhoods characterized by older or lower-value housing may be more susceptible to flood damage, potentially due to differences in building condition, maintenance, or flood mitigation capacity.

In addition, a moderate positive correlation between year built and total damage ($r = 0.326$) indicates that areas with older housing stock tended to experience greater flood losses. Older buildings may be more vulnerable due to aging materials, basement configurations, and legacy infrastructure systems. Many Detroit neighborhoods contain pre-1950 housing and combined sewer systems, which have been associated with recurrent flooding and sewer backups. These findings highlight the importance of neighborhood housing characteristics in shaping flood impacts, alongside hydrologic factors.

Table 2: Correlation analysis results

Variables	1	2	3	4
1 Total Damage	1.00	-0.379*	0.326	-0.149
2 Housing Price		1.00	-0.109	0.330
3 Built Year			1.00	-0.422*
4 Water Level				1.00

*Significant at alpha level of 0.05.

(2) Comparative analysis across damage quartiles

To assess whether neighborhood housing characteristics differed systematically across damage levels, a one-way analysis of variance (ANOVA) was conducted comparing four ZIP code quartiles defined by total flood damage (Table 3). The results indicated statistically significant differences in average housing price across quartiles ($F(3, 26) = 4.89$, $p = 0.0079$). ZIP codes in the highest-damage quartile exhibited the lowest average housing values, reinforcing the correlation results and suggesting that economically vulnerable neighborhoods experience disproportionately high flood impacts.

There were also statistically significant differences in the number of losses per ZIP code ($F(3, 26) = 8.49$, $p = 0.0004$), with higher-damage quartiles experiencing more frequent claims. This suggests that repeated events have contributed to cumulative neighborhood-level damage patterns.

In contrast, several variables did not exhibit statistically significant variation across damage quartiles. Average water level ($p = 0.86$), year built ($p = 0.24$), and average frontage length ($p = 0.16$) did not differ significantly among groups. The lack of significance for year built, despite the moderate correlation observed earlier, suggests that housing age may be associated with flood damage but does not clearly differentiate neighborhoods when aggregated into damage-based quartiles. The nonsignificant differences in water level further indicate that structural and socioeconomic conditions may play a larger role than hydrologic indicators alone in shaping neighborhood-scale flood impacts. In summary, the ANOVA results highlight pronounced disparities in housing value across neighborhoods experiencing different levels of flood damage. Figures 2 and 3 visually present the flood level and damages per ZIP code across Detroit.

Table 3: One-way ANOVA results

Variables	df	F	p-value
Housing Price	3	4.89	0.0079
Built Year	3	1.49	0.2403
Water Level	3	0.25	0.8558
Average frontage length	3	1.89	0.1560
Number of Losses	3	8.49	0.0004

(3) Regression Modeling

To further examine predictors of total flood damage, a multiple linear regression model was estimated using average housing price, year built, average water level, and average frontage length as independent variables. The results (table 4) indicated that none of the predictors were statistically significant at conventional levels, and the overall model explained a relatively small proportion of the variance in total flood damage ($R^2 = 0.18$, $p = 0.279$). Although average housing price and year built exhibited coefficient signs consistent with earlier correlation patterns, their effects were weak and not statistically significant. These findings indicate that, at the ZIP code scale, housing-related attributes alone provide limited

explanatory power for modeling variation in total flood damage across Detroit neighborhoods.

Table 4: Regression results

Predictor	Coeffic -ient	Std. Error	t-value	p-value
Constant	22,243	-*	0.20	0.84
Housing Price	3.52	5.51	0.64	0.53
Built Year	-13.81	11.72	-1.18	0.25
Water Level	-28.87	63.00	-0.46	0.65
Average frontage length	-0.00	0.00	-0.44	0.66

* The standard error of the intercept is omitted because it does not provide meaningful interpretive value.

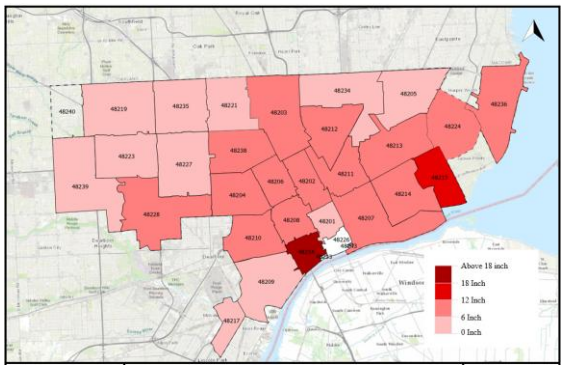


Figure 2: Detroit flood level by zip code

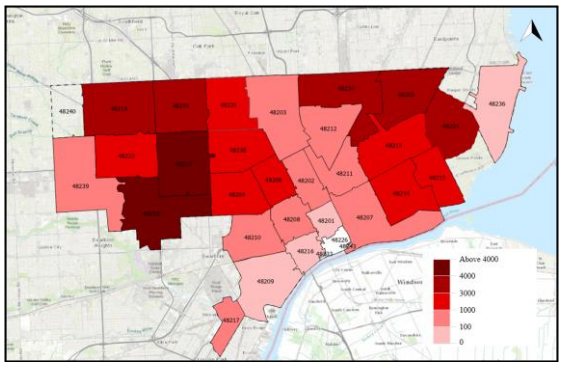


Figure 3: Detroit damage by zip code

(4) K-Means clustering analysis

To complement the quartile-based comparisons, a K-means clustering analysis (k=3) was conducted to identify data-driven neighborhood groupings based on flood damage and housing characteristics. Alternative values of *k*were explored, but k=3 yielded the most meaningful and interpretable group distinctions for the study objectives. Six variables were standardized and used in the analysis, to ensure that all variables contributed equally to the clustering process.

Table 5 presents the results of the clustering analysis, which identified three distinct flood-vulnerability profiles. Cluster 1 represents the highest-risk group, characterized by substantially higher total flood damage, lower housing values, smaller average frontage lengths, higher water levels, and the greatest number of flood loss events. The housing stock in this cluster is also older on average, indicating elevated structural vulnerability.

Clusters 0 and 2 exhibit comparatively lower flood damage and fewer loss events, along with higher housing values and larger average frontage lengths, suggesting relatively lower flood vulnerability. Although the average year built is similar across clusters, the combined economic and physical characteristics clearly differentiate the resulting risk profiles.

Notably, the highest-risk cluster closely aligns with the highest-damage quartile identified in earlier analyses, reinforcing the conclusion that older, lower-value housing areas are disproportionately affected by flooding. Overall, the clustering results demonstrate that unsupervised, data-driven classification methods can effectively identify neighborhood-level flood-vulnerability patterns and support automated flood-risk assessment and resilience planning.

Table 5. Clustering Results

Predictor	Cluster 0	Cluster 1	Cluster 2
Total Flood Damage	425	2,580	354
Housing Price	131	53	85
Built Year	1940	1942	1946
Water Level	6.0	6.5	5.5
Average frontage length	28501	1600	2600
Number of Losses	2.7	3.0	2.5

Discussion and conclusions

This study examined relationships between observed flood losses and neighborhood housing characteristics in Detroit to better understand residential flood vulnerability in a legacy urban context. The combined results of the correlation and comparative analyses indicate that neighborhood housing characteristics are associated with patterns of flood damage in Detroit, even though they do not fully explain flood losses on their own. Moderate correlations and statistically significant differences across damage quartiles show that ZIP codes with lower average housing values experienced higher cumulative flood damage and more frequent loss events. Housing age exhibited a moderate association with flood damage, but did not differ significantly across damage quartiles, suggesting that older housing stock may contribute to flood vulnerability without clearly distinguishing neighborhoods when aggregated at the ZIP-code scale. These findings indicate that economically vulnerable neighborhoods experience disproportionate flood impacts, reinforcing prior evidence that flood risk and losses are unevenly distributed. The absence of significant differences in average water levels across damage quartiles further suggests that hydrologic intensity alone

does not account for observed neighborhood-scale disparities in flood damage.

The lack of statistically significant regression results highlights the multidimensional nature of urban flood vulnerability. Flood impacts are also shaped by factors not captured in parcel-level housing datasets, including sewer system condition, drainage capacity, basement construction, neighborhood infrastructure age, and localized hydrologic pathways. In addition, aggregation at the ZIP-code scale may obscure finer-grained spatial variability in flood exposure, particularly in legacy cities where housing and infrastructure conditions vary substantially at smaller spatial scales. NFIP claims may also underrepresent total flood losses due to incomplete insurance coverage, underreporting, and socioeconomic differences in claim participation. The findings also underscore trends that warrant further exploration, such as the negative correlation between total damage and house prices in recorded flooding events. Future research should therefore explore alternative modeling approaches, including nonlinear and spatial regression models, as well as machine learning techniques that can capture complex interactions among variables. Incorporating additional covariates—such as drainage system condition, impervious surface coverage, elevation, floodplain proximity, and building-level characteristics—may also improve explanatory power. These additions clarify that the current findings are not treated as terminal, but rather as a basis for guiding more comprehensive modeling approaches in future work.

This study provides empirical evidence of neighborhood-level disparities in flood impacts using observed damage data. While the analysis is limited to the City of Detroit, similar dynamics may be present in other cities with aging housing stock, legacy infrastructure, and recurrent urban flooding. A key limitation of this study is the use of Zillow housing value estimates. Prior research has shown that automated valuation models (AVMs), including Zillow estimates, may be less reliable in areas with high housing vacancy rates, limited transaction activity, and heterogeneous property conditions (Elliott & Clement, 2017). In legacy cities, where housing markets are often fragmented and marked by significant disinvestment, estimated housing values may not fully reflect the true structural conditions or market dynamics. As a result, housing prices should be interpreted as a proxy indicator of neighborhood economic conditions rather than a precise measure of property-level value.

In addition, NFIP claims data may introduce systematic bias due to uneven participation in flood insurance programs. Enrollment in the National Flood Insurance Program is not universal and is often lower in socially vulnerable or lower-income communities due to affordability constraints, limited risk awareness, and institutional barriers (Kousky, 2018; Wing et al., 2022). Consequently, observed claims may underrepresent total flood damage in uninsured or underinsured areas, potentially biasing comparisons across neighborhoods. These limitations are particularly relevant when interpreting equity-related findings, as areas identified as lower-risk based on claims data may in fact experience substantial unreported losses. Therefore, while the results provide important insights into neighborhood-scale flood vulnerability, they should be interpreted considering the context of this work.

Future research should extend this approach to multiple cities and integrate additional variables related to infrastructure condition, drainage performance, and building-level characteristics to improve explanatory power and generalizability. In addition, future research should integrate additional data sources and finer-scale observations to more fully capture the distribution of flood impacts.

The findings presented in this study are relevant to a diverse set of stakeholders involved in urban flood risk management and housing resilience. City officials and urban planners can use the neighborhood-scale insights to prioritize flood mitigation investments, target infrastructure upgrades, and integrate housing vulnerability considerations into stormwater and resilience planning. Housing agencies and community development organizations may apply the results to identify neighborhoods where older, lower-value housing stock faces elevated flood impacts and where rehabilitation, retrofit, or assistance programs could yield the greatest equity benefits. Homeowners and residents can benefit from improved awareness of localized flood vulnerability, supporting more informed decisions regarding property-level mitigation and insurance coverage. In addition, insurers and risk analysts may use the findings to enhance neighborhood-level risk assessment and better understand the limitations of housing characteristics alone in predicting flood losses. Collectively, these results support more targeted, data-informed, and equitable approaches to urban flood mitigation and housing resilience in legacy cities.

References

ASCE. (2025). Infrastructure Report Card. https://infrastructurereportcard.org/

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. Global environmental change, 18(4), 598-606.

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. Social Science Quarterly, 84(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

Elliott, J. R., & Clement, M. T. (2017). Natural hazards and local development: The successive nature of landscape transformation in the United States. Social Forces, 96(2), 851-876.

Federal Emergency Management Agency (FEMA). (2023). National Flood Insurance Program (NFIP) Overview and Policy Details.

IPCC Sixth Assessment Report. (2021). Climate Change 2021. https://www.ipcc.ch/report/ar6/wg1/

Kousky, C. (2018). Financing flood losses: A discussion of the National Flood Insurance Program. Risk Management and Insurance Review, 21(1), 11–32. https://doi.org/10.1111/rmir.12090

Liu, J., Shao, W., Xiang, C., Mei, C., & Li, Z. (2020). Uncertainties of urban flood modeling: Influence of parameters for different underlying surfaces. Environmental research, 182, 108929.

Qi, M., Huang, H., Liu, L., & Chen, X. (2020). Spatial heterogeneity of controlling factors' impact on urban pluvial flooding in Cincinnati, US. Applied Geography, 125, 102362.

Shokry, G., Anguelovski, I., Connolly, J. J., Maroko, A., & Pearsall, H. (2022). "They didn't see it coming": Green resilience planning and vulnerability to future climate gentrification. Housing Policy Debate, 32(1), 211-245.

Tate, E., Rahman, M. A., Emrich, C. T., & Sampson, C. C. (2021). Flood exposure and social vulnerability in the United States. Natural Hazards, 106(1), 435-457.

U.S. EPA. (2021). Flood Resilience and Stormwater Management.

Wing, O. E., Lehman, W., Bates, P. D., Sampson, C. C., Quinn, N., Smith, A. M., ... & Kousky, C. (2022). Inequitable patterns of US flood risk in the Anthropocene. Nature Climate Change, 12(2), 156-162.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AN ONTOLOGY-DRIVEN FRAMEWORK FOR INTELLIGENT EMBODIED CARBON DECISION-SUPPORT IN METRO STATION DESIGN

Linghui Xie, and Xueqing Zhang
Hong Kong University of Science and Technology, Hong Kong, China

Abstract

Conventional element-centric BIM-LCA integration omits temporary works (e.g., excavation support, formwork, dewatering) and faces semantic barriers (e.g., inconsistent IFC attribute naming), causing significant carbon underreporting. A dynamic ontology-driven framework was proposed to overcome this. Its semantic core is a multi-module ontology, shifting assessment from static elements to context-aware construction processes. An automated pipeline transforms OpenBIM data into a carbon knowledge graph, capturing permanent and temporary emissions. A federated neuro-symbolic multi-agent system, combining LLM-based probabilistic reasoning with deterministic ontology constraints, enables natural language querying and scenario simulation for interactive decision support.

Introduction

As urbanization accelerates, metro systems are pivotal for sustainable mobility. In China, as of March 2025, 304 metro lines span 6,530 stations, with 157 lines under construction (He et al. 2025). While operation offers environmental benefits, construction is energy-intensive: a kilometer of open-cut metro emits 39,000–84,000 t CO_2e (Wang et al. 2023; Gan et al. 2024; Zhang et al. 2024). This intensity is comparable to constructing approximately 46 kilometers of highway paradigms y(Chen et al. 2023a). Therefore, calculating and reducing greenhouse gas emissions during the materialization stage is a critical priority for industry (Geng et al. 2022; He et al. 2025).

Early-stage design decisions lock in >80% of emission reduction potential (Gao et al. 2024b), but comprehensive LCA requires detailed data only available later (Roberts et al. 2020), leaving a gap between decision needs and data availability (Hollberg et al. 2020). BIM-LCA integration automates quantity extraction and scenario simulation (Gao et al. 2024a; Kim et al. 2024), yet remains ill-equipped for underground metro complexities (Li et al. 2024).

Current methods face three technical limitations for metro stations. First, the element-centric paradigm omits temporary works (excavation support, dewatering, earthwork logistics) rarely modeled in IFC files, causing significant underreporting. Second, tools lack domain-aware reasoning; early-stage low-LOD models cannot infer implicit construction sequences or backfill requirements. Third, semantic disconnect between design models and environmental databases forces error-prone manual mapping (Cavalliere et al. 2019; Abbasi and Noorzai 2021; Cheng et al. 2022).

To overcome these, we propose an ontology-driven OpenBIM framework that shifts from data extraction to knowledge-enabled inference. It harmonizes structural, enclosure, and geological data via formal logic rules that deduce implicit construction processes from explicit design geometry (e.g., inferring excavation zones from foundation pit design). The framework is validated on six open-cut metro stations with three contributions: (1) the Metro Station Carbon Assessment Ontology (MSCA-Onto) decoupling geometry from construction logic; (2) an automated inference mechanism reconstructing implicit work quantities from sparse geometry; (3) a neuro-symbolic cognitive interface combining LLM reasoning with deterministic ontological constraints, turning the tool into an interactive engineering copilot.

Methodology

Overview

Conventional BIM-LCA methods are constrained by an element-centric paradigm, relying on static bills of quantities from explicit geometry. This causes three shortcomings: (1) data incompleteness – omission of temporary works (formwork, excavation support, dewatering); (2) semantic interoperability barriers – inconsistent IFC naming conflicts with LCA databases, requiring manual mapping; (3) static post-hoc reports – no interactive "what-if" simulation. To address these, we propose a dynamic ontology-driven framework with three integrated layers, each targeting one gap.

The Semantic Foundation Layer (Layer I) redefines the underlying assessment logic to resolve semantic ambiguities. Instantiated by the Metro Station Carbon Assessment Ontology (MSCA-Onto) and based on a System-Element-Process-Resource-Carbon (S-E-P-R-C) architecture, this layer shifts the focus from static elements to context-aware construction processes. By establishing a unified vocabulary, it bridges the semantic gap between heterogeneous design data and environmental databases. Building upon this foundation, The Automated Computational Layer (Layer II)

operationalizes the ontology to quantify emissions comprehensively. It transforms OpenBIM data into a computable Carbon Knowledge Graph (CKG) by automating semantic alignment and geometric enrichment. Crucially, this layer utilizes ontological inference rules to deduce the consumption of unmodeled temporary works from permanent structures, thereby ensuring a complete system boundary and correcting the underreporting inherent in conventional methods.

While the first two layers establish a rigorous and complete data repository, a cognitive interaction layer (Layer III) is implemented to bridge the final cognitive gap between structured data and human decision-making. Designed as a federated neuro-symbolic Multi-Agent System (MAS), this layer synergizes the probabilistic reasoning of Large Language Models (LLMs) with the deterministic logic of the ontology. It transforms the CKG from a passive database into an active decision-support engine, enabling stakeholders to synthesize optimization strategies and perform various scenario analysis through natural language interaction. Collectively, this framework not only ensures accurate quantification during the design phase but also establishes a scalable, semantic continuum for lifecycle carbon management.

Metro station carbon assessment ontology

The MSCA-Onto serves as the semantic foundation of the proposed framework, engineered to facilitate a paradigm shift from static, element-centric accounting to dynamic, process-aware carbon assessment. The development of this ontology follows a hybrid methodology that integrates the structural rigor of METHONTOLOGY with the lifecycle flexibility of the NeON framework. Spanning six stages from specification to evaluation, the process involved a comprehensive synthesis of construction standards, IFC schemas, and project documentation. To ensure high interoperability, ontology strategically reuses established schemas such as QUDT for units and PROV-O for provenance, ultimately formalizing domain knowledge into OWL 2 DL expressions supplemented by SWRL rules for advanced automated reasoning.

The structural core of MSCA-Onto is the S-E-P-R-C framework, which establishes a unified semantic layer to decouple static geometric data from dynamic resource consumption. Within this architecture, the System and Element modules decompose the metro station into functionally distinct engineering systems and specialized subtypes, such as diaphragm walls and bracing, effectively bridging the semantic gap between generic IFC schemas and underground engineering practice. The Process module represents the framework's primary innovation by utilizing WorkPackageTemplate entities as semantic adapters. These templates function as inference engines that automatically deduce implicit construction activities, such as formwork or excavation, from sparse BIM geometry. This process-based logic subsequently connects to the Resource and Carbon modules, which map activities to material descriptors and emission factors from external databases to enable multi-level carbon aggregation across lifecycle stages A1 to A5. The overall architecture is illustrated in Figure 1. Six competence questions validated MSCA-Onto's ability to quantify material consumption, evaluate low-carbon substitution, analyze temporary structure recycling, identify carbon hotspots, and assess logistics footprints, proving its utility for early-stage carbon management.

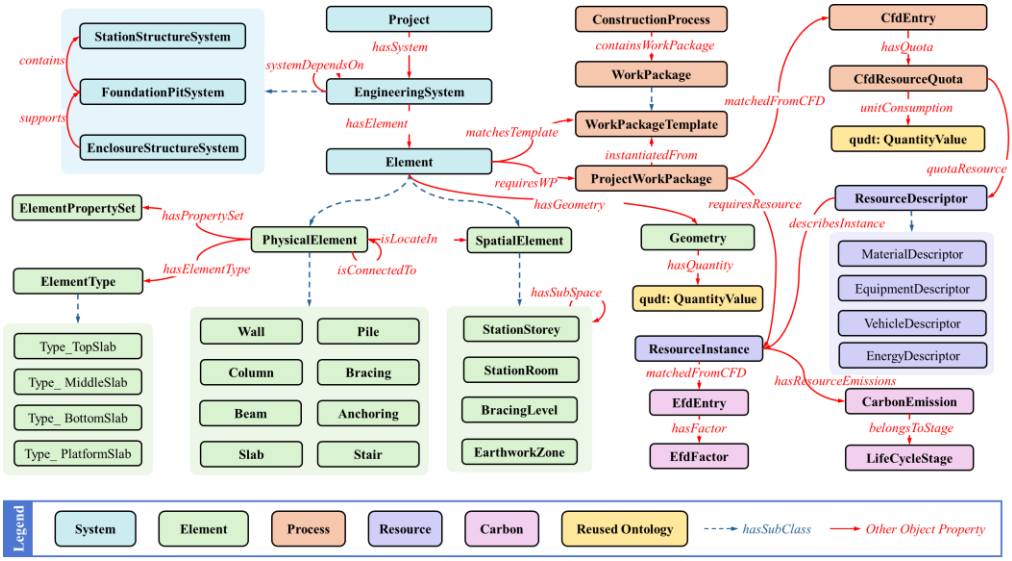


Figure 1: Overall architecture of MSCA-Onto

Table 1 summarizes the six competence questions used to validate MSCA-Onto, along with the SPARQL query patterns and expected outcomes. For instance, CQ5 identifies carbon hotspots via a grouping and ranking query; the results across six stations consistently ranked diaphragm walls and slabs as the top two contributors (42% and 28% of structural carbon, respectively). CQ6 quantifies logistics emissions by joining material supply distances with transport emission factors, yielding a contribution of 8–12% of A1–A3 emissions. The full

executable SPARQL queries and result datasets are provided in the supplementary material.

Table 1 Ontology validation competency questions

Competence Question	Validation Objective	SPARQL Pattern
CQ1: What are the material consumptions and total carbon emissions generated by waterproofing engineering in the main structure?	Material Consumption and Carbon Quantification	Process→Resource→Carbon chain with material filtering
CQ2: How much carbon reduction can be achieved by substituting conventional concrete with GGBS-based low-carbon concrete in primary structural elements?	Low-carbon material decision support	Comparative analysis with emission factor lookup from EFD
CQ3: What is the carbon impact of temporary support system demolition activities, including recycling credits?	Temporary structure lifecycle assessment	Temporal process analysis with negative emission accounting
CQ4: What is the carbon emission comparison between vertical retaining systems and horizontal bracing systems in the enclosure structure?	Comparative System Analysis	Cross-system comparison with intensity metrics calculation
CQ5: What are the top 10 carbon hotspots among structural components, and what is their carbon intensity ranking?	Component-level carbon hotspot identification	Multi-criteria ranking with statistical aggregation
CQ6: What is the carbon contribution of material transportation in construction logistics?	Construction logistics carbon analysis	Distance-weight-emission calculation per transport mode

Automated carbon assessment pipeline based on OpenBIM

Based on the MSCA-Onto, this study establishes a systematic automated pipeline designed to transform heterogeneous OpenBIM data into a comprehensive Carbon Knowledge Graph (CKG) suitable for high-precision embodied carbon assessment. The workflow begins by anchoring the assessment in two hierarchically structured reference databases: a construction process database (CPD) for resource inventory and an emission factor database (EFD) for environmental impact, which are strategically decoupled to ensure data flexibility and fitness for purpose. To resolve the semantic ambiguity inherent in raw IFC data, the pipeline employs a hybrid topological reasoning engine. By utilizing spatial indexing and geometric algorithms, such as ray casting and collision detection, the system infers critical spatial relationships to accurately classify elements based on their functional roles, such as distinguishing a generic slab from a specific station platform. For instance, an unclassified IfcSlab instance is resolved as a "Platform Slab" via three topological checks: normal vector orientation (ray casting), vertical position (R-tree query), and adjacency to passenger spaces (collision detection). Simultaneously, to address the omission of temporary earthworks in standard BIM, the module executes a rigorous geometric reconstruction of the foundation pit. Using a "profile-sweep-trim" operation, the system synthesizes horizontal enclosure envelopes with vertical structural surfaces to generate precise volumetric domains for excavation and backfill, which are then intersected with geological layers for quantification.

The pipeline culminates in an automated element-process-resource mapping mechanism that operationalizes the ontology's inference logic. Through domain-specific SWRL rules, semantically enriched elements are matched with WorkPackageTemplate adapters. For example, the following SWRL rule triggers an excavation support work package when the depth of a foundation pit exceeds 5 meters:

```
StationPit(?pit) ∧ hasDepth(?pit, ?depth) ∧ swrlb:greaterThan(?depth, 5.0) → requiresWorkPackage(?pit, ExcavationSupportWorkPackage)
```

Figure 2: An example of the SWRL rule

These adapters dynamically retrieve construction methods from the CPD and associate emission factors from the EFD, instantiating ProjectWorkPackage and ResourceInstance entities. This enables bottom-up aggregation of carbon emissions across stages A1–A5 (material production, transport, on-site construction), ensuring full traceability from individual resources to the project footprint. The complete workflow is shown in Figure 3.

A Multi-agent system for interactive carbon intelligence

To bridge the critical cognitive gap between the static Carbon Knowledge Graph (CKG) and the dynamic, unstructured nature of engineering decision-making, this study proposes a federated neuro-symbolic Multi-Agent System (MAS). Designed around a rigorous divergence - convergence workflow, the MAS strictly decouples cognitive planning from domain-specific execution across three logical layers: cognitive planning, semantic execution, and decision synthesis. This hierarchical architecture ensures that the probabilistic reasoning of Large Language Models (LLMs) is constrained by the deterministic logic of the MSCA-Onto, enabling reliable

support for information retrieval, knowledge correction, and strategic analysis entirely through natural language interaction.

The workflow is governed by the Cognitive Planning Layer, where the Orchestrator Agent (A_{orch}) functions as the central reasoning engine. Upon receiving a user query, A_{orch} utilizes a fine-tuned LLM for intent classification and employs Chain-of-Thought (CoT) reasoning to decompose the request into a machine-readable Task Dependency Graph (TDG). To constrain the LLM's probabilistic reasoning without exceeding context window limits, the system does not parse the entire OWL ontology as a raw prompt. Instead, a compressed schema dictionary—containing only the permitted S-E-P-R-C core classes, taxonomic relations, and operational properties—is dynamically injected into the LLM's system prompt to grant it structural awareness. Meanwhile, strict semantic verification (e.g., SWRL rule checking) is deferred to the downstream symbolic agents.

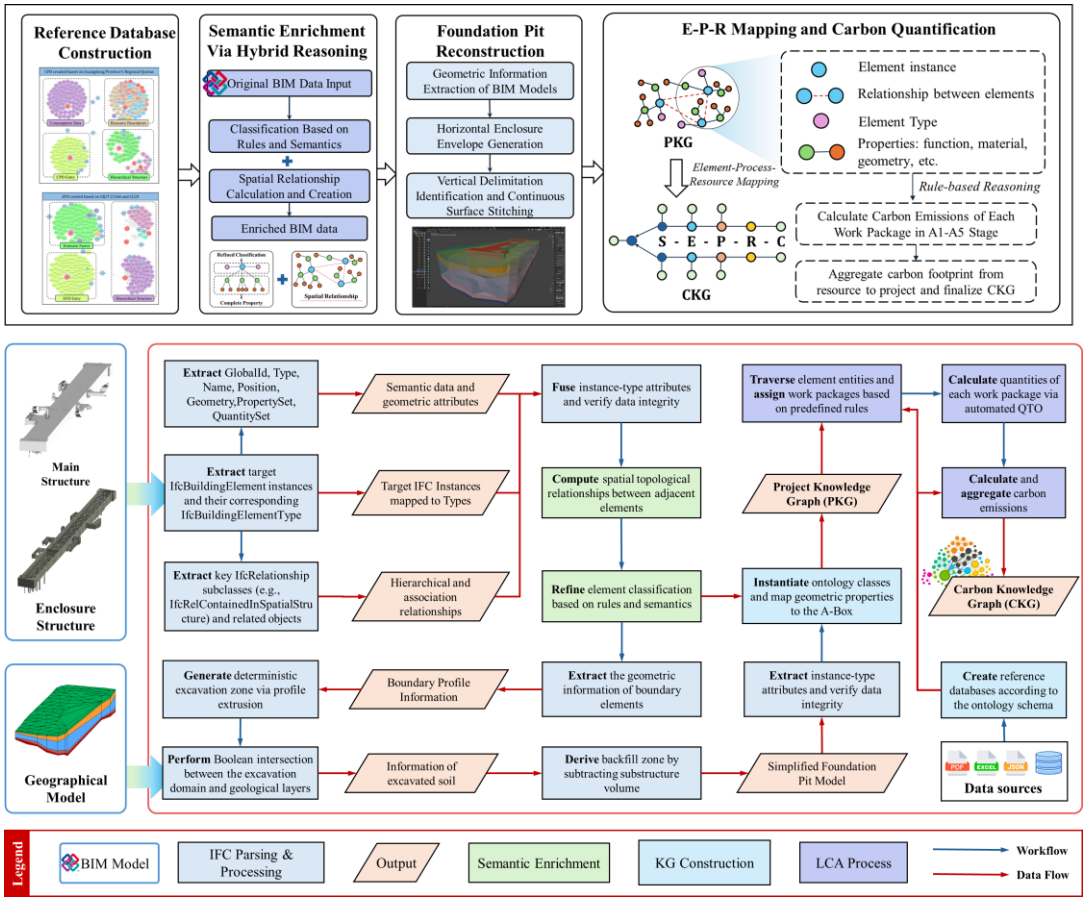


Figure 3: Workflow of automated carbon assessment pipeline

The LLM performs orchestration only: it maps natural language to a predefined set of task templates and produces a TDG. It does not generate SPARQL or numerical parameters directly. For example, given " Replace C30 with C40 concrete in the bottom slab" , CoT produces: (1) Intent: Strategic Analysis; (2) Entity: bottom slab → Type_BottomSlab; (3) Parameter: C30→ C40; (4) TDG: [Extraction→Reader→Calculation]. All formal queries and calculations are executed by symbolic agents using deterministic templates. The LLM's reasoning is confined to intent classification, ambiguity detection, and workflow planning. To optimize efficiency, the system implements a hybrid shared-state architecture: control signals are dispatched asynchronously to trigger downstream agents, while data artifacts are managed via a global "blackboard" pattern. This ensures that agents read from and write to a validated, synchronized context, minimizing communication overhead while maintaining strict data consistency.

The Semantic Execution Layer operationalizes these tasks through a suite of functionally atomized agents. The Metadata Agent ($\boldsymbol{A}_{\text{meta}}$) serves as the semantic bridge, parsing complex queries into atomic propositions and mapping ambiguous terms to precise ontology entities via a "decomposition-retrieval-verification" pipeline. For hypothetical scenarios, the Extraction Agent ($\boldsymbol{A}_{\text{extr}}$) translates user intents into a "delta graph specification" ($\boldsymbol{G}_{\text{spec}}$), a provenance annotated artifact that prioritizes user-defined parameters over system defaults while filtering out invalid structural alterations. Simultaneously, the Unified Reader Agent ($\boldsymbol{A}_{\text{read}}$) executes a dual-path retrieval strategy-using deterministic SPARQL for project

data and hierarchical semantic search for external reference databases-to construct a composite data context ($\boldsymbol{D}_{\mathbf{comp}}$). These inputs converge at the Calculation Agent ($\boldsymbol{A}_{\mathbf{calc}}$), which performs a sandboxed semantic overlay operation $R = f_{\text{template}}\left(D_{\text{comp}} \oplus G_{\text{spec}}\right)$. By dynamically superimposing user constraints onto the baseline context, this agent ensures that all quantitative derivations are mathematically rigorous and fully traceable.

Finally, the Decision Synthesis Layer ($\boldsymbol{A}_{\text{syn}}$) acts as the system's executive interface and safety valve. It adaptively formats heterogeneous outputs-ranging from visualization of carbon hotspots to comparative scenario matrices-into coherent, human-readable responses enriched with provenance annotations. Crucially, for tasks involving persistent modifications to the CKG, $\boldsymbol{A}_{\text{syn}}$ enforces a strict "human-in-the-loop" protocol. Rather than executing updates immediately, it generates a comprehensive impact assessment forecasting the cascading effects of the proposed changes. The system then pauses to solicit explicit user adjudication, executing the final SPARQL UPDATE statements only after receiving authorization, thereby guaranteeing that stakeholders retain ultimate sovereignty over data integrity. The structure of MAS is shown in Figure 4.

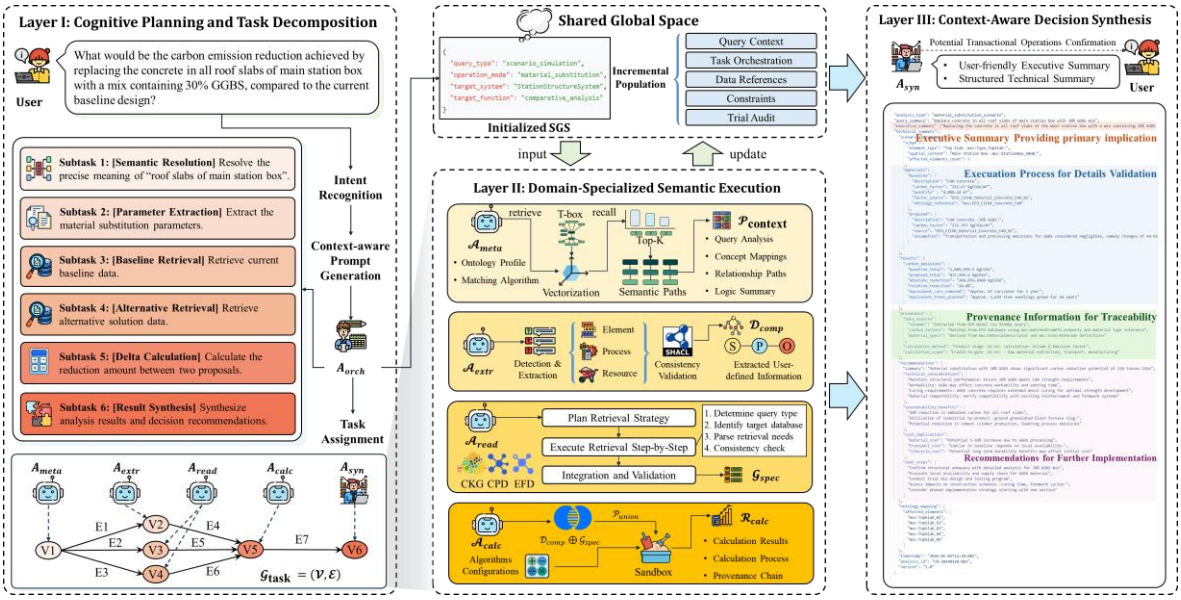


Figure 4: The workflow of MAS proposed

Implementation and results

Evaluation of the automated assessment pipeline

The empirical validity of the proposed framework was established using a diverse dataset comprising six open-cut metro station projects located in Guangdong Province, China. These projects represent a comprehensive range of structural typologies. The excavation depths vary significantly from 16.5 meters to 22.3 meters. This diversity ensures that the validation encompasses different geological conditions and construction complexities.

The input data for each project consisted of the original Industry Foundation Classes design models and the corresponding structural drawings. To rigorously test the scalability of the automated pipeline, we maintained the BIM models at a high level of fidelity. The dataset contains a total of over 44,000 individual geometric components. All three methods, including Manual Calculation (Baseline), Typical BIM-LCA Software, and the Proposed Framework, operated on these identical IFC datasets. They also referenced the exact same data infrastructure for construction processes and emission factors. This rigorous control ensures that any observed performance variance is attributable solely to methodological differences rather than data discrepancies. The CPD was populated using Guangdong Province's municipal construction quotas, covering 124 construction activities (e.g., diaphragm wall excavation, concrete pumping, formwork erection). The EFD (Emission Factor Database) combines Ecoinvent 3.8 for background processes (e.g., cement production, diesel combustion) and region-specific factors from the China Life Cycle Database (CLCD) for materials like steel and concrete. Transportation distances were set to default regional averages (50 km for concrete, 200 km for steel) unless specified otherwise in the BIM properties.

To rigorously quantify the performance of the ontology-driven pipeline, the evaluation was structured around three key performance indicators designed to assess mathematical precision, statistical correlation, and operational efficiency. The first metric, mean absolute percentage error (MAPE), measures the average magnitude of error between the automated results C_{auto} and the manual ground-truth C_{man} across the n test projects, calculated as:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{C_{man,i} - C_{auto,i}}{C_{man,i}}\right| \quad (1)$$

This provides a normalized assessment of accuracy where lower values signify higher alignment with expert standards. The second metric is the coefficient of determination (R^2), which evaluates the proportion of variance in the manual baseline that is predictable from the automated pipeline. Finally, temporal efficiency is assessed through the total end-to-end processing time required for a complete assessment.

To rigorously quantify the reliability of these generated profiles, we compared the automated results against the expert-verified manual baseline. We employed Mean Absolute Percentage Error as the primary accuracy metric. This measures the average magnitude of deviation between the automated results and the ground truth. Additionally, we used the coefficient of determination to evaluate linear consistency. The automated pipeline demonstrated exceptional alignment with the manual baseline. It achieved a project-level MAPE of 2.8% and an R^2 value of 0.983. This confirms that the rule-based, ontology-driven workflow can reliably replicate complex engineering logic without the significant deviations observed in standard BIM-LCA plugins. Standard tools often suffer from errors exceeding 23% due to their inability to account for the unmodeled auxiliary processes described above. The detailed accuracy comparison is summarized in Table 2.

Table 2: Accuracy comparison of carbon assessment methods

Metric	Manual Calculation (Baseline)	Typical BIM-LCA method	Ours
Project-Level Accuracy (MAPE)	N/A	23%	2.8%
Component-Level Accuracy (MAPE)	N/A	12%	5.5%
Correlation (R^2)	1.00	0.81	0.983

In terms of operational efficiency, the framework demonstrated a transformative reduction in assessment latency. The manual calculation process required an average of 7.5 person-hours per project. This duration was dominated by tedious quantity take-off and Excel-based mapping. In contrast, the automated pipeline completed the entire end-to-end assessment in an average of 22 minutes.

To understand the computational cost, we decomposed this 22-minute duration into its constituent phases. The semantic enrichment and reasoning phase consumed most of the time, taking approximately 12 minutes. This reflects the computational intensity of the topological algorithms used to resolve element relationships such as containment and adjacency checks. The geometric reconstruction of the braced foundation pit required an average of 8 minutes. This represents a necessary trade-off to capture the implicit earthwork quantities accurately. The final mapping and calculation phase was highly efficient and completed in under 2 minutes. Although the Typical BIM-LCA method was faster at approximately 5 minutes, this speed resulted in the significant accuracy penalty noted in the previous section. Consequently, the proposed framework offers an optimal balance. It delivers professional-grade accuracy with industrial-scale efficiency that is approximately 20 times faster than manual workflows.

Evaluation of the Neuro-Symbolic MAS

To evaluate the efficacy and safety of the proposed Neuro-Symbolic MAS, a comprehensive performance assessment focusing on its interaction reliability across diverse engineering tasks was conducted. The evaluation utilized a benchmark suite of 120 standardized engineering queries distinctively categorized into three operational modes: information retrieval (IR), knowledge correction (KC), and strategic analysis (SA). To assess the framework's adaptability and model-agnostic robustness, the experiments were replicated using three distinct LLMs as the reasoning backbone: DeepSeek-V3, Gemini 3 Pro, and Qwen 2.5-14B.

To strictly quantify the system's engineering capabilities, we introduced a "trap mechanism" within the benchmark suite. Approximately 15% of the queries were intentionally designed with missing parameters or ambiguous intents, such as requesting a material update without specifying the target component. For these queries, a correct outcome is defined not as executing an operation, but as halting execution and requesting clarification. This tests the system's meta-cognitive ability to detect ambiguity, which is a critical safety feature in engineering.

Accordingly, the system's performance was measured using a quantitative net reliability score (R). Unlike subjective fluency metrics, this score is strictly grounded in the quantitative accuracy of the output. The metric rewards correct answers and penalize unsafe errors while neutrally treating safe failures where the system correctly refuses to answer due to ambiguity. The metric is formalized as:

$$R = \frac{1}{N}\sum_{i=1}^{N}\left(I_{\text{correct}}^{(i)} - \lambda \cdot I_{\text{unsafe}}^{(i)}\right) \quad (2)$$

where $I_{\text{correct}}^{(i)}$ is a binary indicator set to 1 if the system's output strictly matches the ground truth or correctly identifies a trap. $I_{\text{unsafe}}^{(i)}$ is a penalty indicator set to 1 if the output contains a constraint violation, such as returning a physically impossible parameter or hallucinating a value for an ambiguous query. The safety penalty coefficient λ is set to 1.0, reflecting the engineering principle that providing erroneous data is as detrimental as the value of a correct insight. To ensure statistical robustness, the benchmark suite was executed three times for each configuration resulting in a total of 360 independent trials.

The experimental results, summarized in Table 3, demonstrate that the proposed neuro-symbolic architecture effectively standardizes performance output

enabling even smaller models to achieve engineering-safe reliability levels.

Table 3: Comparative Performance of LLM Backbones within the MAS Framework

Operational Mode	Metric	Qwen 2.5-14B	DeepSeek-V3.2	Gemini 3-Pro
Information Retrieval	$I_{correct}$	68.2%	85.3%	92.1%
	I_{unsafe}	3.0%	2.0%	0.0%
	R	0.65	0.83	0.92
Knowledge Correction	$I_{correct}$	58.0%	82.5%	91.3%
	I_{unsafe}	6.0%	3.5%	0.0%
	R	0.52	0.79	0.91
Strategic Analysis	$I_{correct}$	50.0%	66.6%	86.4%
	I_{unsafe}	9.5%	4.0%	3.0%
	R	0.41	0.63	0.83
Global Performance	Avg. R Score	0.53	0.67	0.86
	Avg. Latency	18.5s	28.8s	34.2s

Using DeepSeek-V3 as the backbone, the system achieved a robust global R score of 0.91. It exhibited exceptional performance in Strategic Analysis tasks with a success rate of 92.0%, successfully decomposing complex "what-if" scenarios into valid delta specifications. The constraint violation rate was kept extremely low at 1.8%, proving that the orchestrator's planning logic effectively restrains the model's generation within ontological bounds.

The integration of Gemini 3 Pro yielded the highest reliability with a global R score of 0.86. Its superior reasoning capabilities minimized the errors in complex multi-hop queries. Notably, in KC tasks, it achieved a 100% success rate demonstrating perfect alignment with the feasibility protocol when parsing complex user directives. The absence of unsafe errors across all trials highlights the model's rigorous adherence to the provided prompts and its ability to correctly identify all trap questions.

Crucially, the lightweight Qwen 2.5-14B model demonstrated the structural value of the framework. While its raw reasoning capability is lower than the flagships resulting in a lower raw success rate, the MAS architecture successfully mitigated the risks. Instead of generating plausible hallucinations for ambiguous queries, the model frequently defaulted to "safe failures" when uncertain. Although its global R score (0.67) is lower, the constraint violation rate was contained at 3.5%. This is significantly better than the typical instability of unconstrained small models confirming that the framework functions as a cognitive guardrail that makes lightweight models viable for non-critical engineering tasks.

Discussion

The empirical results show that the element-centric BIM-LCA paradigm is insufficient for underground infrastructure. About 12.4% of total embodied carbon comes from unmodeled temporary work, revealing systematic scope leakage and confirming that geometric explicitness is a poor proxy for construction reality. Our ontology-driven pipeline bridges this gap using the S-E-P-R-C model as an active inference engine. The 2.8% error rate proves that semantic reconstruction is necessary for professional-grade accuracy, suggesting that future standards should mandate procedural logic beyond geometric LoD.

A key architectural insight from the MAS evaluation is that the reliability of generative AI in engineering depends more on system architecture than on model size. Lightweight local models achieved operational safety when constrained by the framework, implying that highly reliable, privacy-compliant tools do not require expensive cloud models. The neuro-symbolic architecture decouples probabilistic intent understanding from deterministic execution and defaults to safe failures on ambiguous queries, defining trustworthiness as the ability to recognize knowledge boundaries.

The framework reduced assessment latency twenty-fold (from 7.5 hours to 22 minutes), shifting engineers' cognitive load from tedious quantity take-off to high-value strategic analysis. This enables real-time carbon budgeting alongside cost and schedule during design. Accuracy depends on database granularity; missing methods fall back to generic averages. Geometric reconstruction is optimized for cut-and-cover stations and may need adaptation for tunnels or complex nodes. The human-in-the-loop step ensures safety but adds latency. The S-E-P-R-C architecture is domain-agnostic: extending to tunnels or bridges only requires domain-specific work package templates and geometric threshold rules. The manual baseline may contain expert bias; future work will validate it with real-world construction data.

Conclusion

This research addressed the systemic limitations of conventional BIM-LCA for underground infrastructure—specifically the omission of auxiliary processes and the lack of interactive decision support. A dynamic assessment framework based on a neuro-symbolic MAS was proposed and validated, marking a paradigm shift from static element-based carbon reporting to interactive, process-aware engineering intelligence.

The primary contribution is a robust computational methodology that harmonizes sparse design data with deep engineering knowledge. The MSCA-Onto, built on the S-E-P-R-C model, decouples geometry from resource consumption, enabling dynamic inference of typically invisible work packages (e.g., excavation support). The automated pipeline transforms OpenBIM data into a computable CKG, capturing the full life-cycle inventory with near-expert precision. Most significantly, integrating

LLMs with deterministic ontologies provides a viable pathway for reliable engineering AI. The federated MAS acts as a cognitive copilot, interpreting complex intents, executing rigorous simulations, and enforcing safety via safe-failure mechanisms. Empirical results show a 12.4% completeness improvement and high reliability across diverse backbones, validating the neuro-symbolic approach as a scalable industry solution.

Future work will focus on agent learning from user corrections and integrating real-time schedule data for dynamic carbon forecasting. This framework lays the foundation for intelligent, holistic carbon management, moving sustainability assessment into an integral, effortless part of infrastructure design.

Reference

Abbasi, Saman, and Esmatullah Noorzai. 2021. "The BIM-Based Multi-Optimization Approach in Order to Determine the Trade-off between Embodied and Operation Energy Focused on Renewable Energy Use." Journal of Cleaner Production 281:125359. doi:10.1016/j.jclepro.2020.125359.

Cavalliere, Carmine, Guillaume Habert, Guido Raffaele Dell'Osso, and Alexander Hollberg. 2019. "Continuous BIM-Based Assessment of Embodied Environmental Impacts throughout the Design Process." Journal of Cleaner Production 211:941–52. doi:10.1016/j.jclepro.2018.11.247.

Chen, Kunyang, Xiangsheng Chen, Lei Wang, Wensheng Yang, Tong Qiu, Dong Su, and Huanyu Wu. 2023. "Low-Carbon Effects of Constructing a Prefabricated Subway Station with Temporary Internal Supports: An Innovative Case of Shenzhen, China." Journal of Cleaner Production 426:139023. doi:10.1016/j.jclepro.2023.139023.

Cheng, Baoquan, Kun Lu, Jianchang Li, Huihua Chen, Xiaowei Luo, and Muhammad Shafique. 2022. "Comprehensive Assessment of Embodied Environmental Impacts of Buildings Using Normalized Environmental Impact Factors." Journal of Cleaner Production 334:130083. doi:10.1016/j.jclepro.2021.130083.

Gan, Mi, Dandan Li, Zhu Yao, Hao Yu, and Qichen Ou. 2024. "Intelligent Decision Modeling for Optimizing Railway Cold Chain Service Networks under Uncertainty." Information Sciences 679:121112. doi:10.1016/j.ins.2024.121112.

Gao, Hui, Donglin Wang, Xiaoxu Du, and Zhongwei Zhao. 2024. "An LCA-BIM Integrated Model for Carbon-Emission Calculation of Prefabricated Buildings." Renewable and Sustainable Energy Reviews 203:114775. doi:10.1016/j.rser.2024.114775.

Gao, Yu, Jiayuan Wang, and Tak Wing Yiu. 2024. "Multi-Information Integration-Based Life Cycle Analysis of Greenhouse Gas Emissions for Prefabricated Construction: A Case Study of Shenzhen." Environmental Impact Assessment Review 104:107330. doi:10.1016/j.eiar.2023.107330.

Geng, Jingjing, Jiajia Wang, Jianguang Huang, Ding Zhou, Jing Bai, Jiayuan Wang, Hui Zhang, Huabo Duan, and Wenbo Zhang. 2022. "Quantification of the Carbon Emission of Urban Residential Buildings: The Case of the Greater Bay Area Cities in China." Environmental Impact Assessment Review 95:106775. doi:10.1016/j.eiar.2022.106775.

He, Qiufeng, Xiangsheng Chen, Tong Qiu, Kunyang Chen, and Aidong Li. 2025. "GHG Emission Efficiency of Prefabricated Composite Subway Stations: A Novel Case Study of Shenzhen, China." Environmental Impact Assessment Review 110:107664. doi:10.1016/j.eiar.2024.107664.

Hollberg, Alexander, Gianluca Genova, and Guillaume Habert. 2020. "Evaluation of BIM-Based LCA Results for Building Design." Automation in Construction 109:102972. doi:10.1016/j.autcon.2019.102972.

Kim, Hyunsik, Seungjun Roh, and Jeonghwan Kim. 2024. "Embodied Carbon of BIM Bridge Models According to the Application of Off-Site Prefabrication: Precast Concrete Applied to Superstructure and Substructure." Developments in the Built Environment 20:100550. doi:10.1016/j.dibe.2024.100550.

Li, Tao, Yi Rui, Hehua Zhu, Linhai Lu, and Xiaojun Li. 2024. "Comprehensive Digital Twin for Infrastructure: A Novel Ontology and Graph-Based Modelling Paradigm." Advanced Engineering Informatics 62:102747. doi:10.1016/j.aei.2024.102747.

Roberts, Matthew, Stephen Allen, and David Coley. 2020. "Life Cycle Assessment in the Building Design Process – A Systematic Literature Review." Building and Environment 185:107274. doi:10.1016/j.buildenv.2020.107274.

Wang, Qian, Yun Bai, Yao Chen, Qian Fu, and Tin Kin Ho. 2023. "Optimizing Vertical Alignment of Underground Metro for Energy Saving of Train Operation." Energy 273:127295. doi:10.1016/j.energy.2023.127295.

Zhang, Huan, Zhuoran Li, Xianwang Fan, Deyi Duan, and Wandong Zheng. 2024. "Air-Conditioning Load Characteristics and Grey Box Predicting Model in Subway Stations." Journal of Building Engineering 91:109656. doi:10.1016/j.jobe.2024.109656

Human Data Interaction

Sub-Tracks:

Theories, Methodologies & Methods in Emerging HDI

Paradigms & Frameworks for HDI Design & Analysis

Human-in-the-Loop AI & Decision Support Systems

Trust, Acceptance & Social Impacts of Digital Innovations

Augmented & Virtual Reality Interfaces for AEC

UX, Usability & Cognitive Ergonomics of AEC Tools

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AUGMENTED REALITY AND COMPUTER VISION TO SUPPORT SCHEDULING AND MONITORING IN CONSTRUCTION: A MEP CASE STUDY

Patrick Dallasega[1], Marco Lanzone[1], Andrea Revolti[1], Alessandro Sottovia[1], and Oswald Lanz[1]
[1]Faculty of Engineering, Free University of Bozen-Bolzano, Italy

Abstract

Mechanical, Electrical, and Plumbing (MEP) systems require complex spatial coordination and installation sequences, making scheduling and progress monitoring challenging. This paper presents a framework that integrates Augmented Reality (AR) and Computer Vision (CV) to support MEP scheduling and progress tracking. BIM-based plans are visualized on-site through AR to guide installers, while CV algorithms detect MEP components from images to measure progress and identify schedule deviations. A prototype developed in Unity for HoloLens 2 was tested in laboratory and construction site environments. Results demonstrate its potential to improve construction managers' situational awareness.

Introduction

Mechanical, electrical, and plumbing (MEP) construction projects are among the most complex phases of building delivery because they require the installation of numerous interdependent components within constrained spatial environments. As buildings become more technologically sophisticated, the density of MEP systems has increased, placing significant demands on coordination, sequencing, and precision during installation. Prior research highlights that MEP works often involve intricate spatial relationships and tolerance requirements, which make them particularly susceptible to coordination problems and rework if not managed carefully (Eastman et al., 2011). Moreover, these activities typically occur at later stages of a project, when structural and architectural elements are already in place, thereby limiting flexibility and amplifying the consequences of any errors or clashes discovered on-site.

Despite advances in Building Information Modeling (BIM) and digital design workflows, construction errors such as geometric clashes or incorrect installations are still frequently detected only at late stages of construction. Several studies have shown that while BIM-based clash detection is widely adopted, its benefits are reduced when updates to the model are not timely or when discrepancies between the as-designed and as-built conditions are not systematically captured (Hartmann et al., 2008). In practice, field conditions deviate from planned information for a range of reasons, for example design changes, unforeseen site constraints, or human error. These deviations are often discovered only during installation or inspection. By the time such issues emerge on-site, corrective actions may be costly, disruptive, or infeasible. This temporal gap between planning and execution remains a persistent challenge for effective project control.

The consequences of late error detection manifest in well-documented industry-wide problems of delays and cost overruns. Global construction performance studies indicate that time and budget deviations are widespread, with a significant portion of construction projects exceeding planned durations and costs (Flyvbjerg et al. 2002; Cantarelli et al., 2013). Highly integrated scopes such as MEP are particularly prone to such deviations because misalignments accumulate across trades and require iterative rework cycles. Furthermore, traditional monitoring techniques, such as manual inspections, paper-based checklists, or 2D drawings often limit the ability of project managers to obtain an accurate, real-time understanding of construction progress or emerging risks. As a result, decision-making tends to be reactive rather than proactive, reducing opportunities for early intervention.

In recent years, Augmented Reality (AR) has received growing attention to improve visualization and communication in construction environments. By superimposing digital information directly onto physical space, AR can provide workers and managers with intuitive insights into planned tasks, component locations, and spatial constraints (Billinghurst et al., 2015). Earlier research has demonstrated that AR can support construction activities such as layout, inspection, training, and safety management by making complex information easier to comprehend on-site (Kamat et al., 2011). When combined with sensor data or computer vision capabilities, AR systems can also help detect discrepancies between planned and actual conditions, thereby creating opportunities for early error identification. The integration of AR with BIM and progress monitoring tools has thus emerged as a promising direction for enhancing construction project management.

This paper proposes an integrated AR and CV approach aimed at supporting the scheduling and monitoring of construction projects, with a particular focus on MEP-intensive environments. The approach combines AR-

based visualization of the construction schedule with automatic detection of installed components through CV techniques. This integration facilitates users to compare the planned and actual states of construction, facilitating earlier detection of deviations and more informed decision-making. The objective is to move beyond traditional, document-based progress tracking toward an immersive, data-driven project control system that can reduce errors, minimize rework, and improve adherence to schedule.

To demonstrate the feasibility and usefulness of the proposed approach, a prototype system was developed and tested partly in a laboratory setting and partly on-site within the construction of a large winery production and storage facility in northern Italy. This project provided a particularly relevant testing environment due to its complex MEP installations, tight construction timelines, and stringent quality requirements typical of industrial facilities. The pilot test allowed for an evaluation of the system's usability, accuracy, and potential benefits in supporting field operations. Preliminary findings suggest that AR-CV technologies can enhance situational awareness, streamline progress verification, and support proactive project control.

Overall, this research contributes to ongoing efforts to digitalize construction management by offering a practical solution for integrating AR and CV into project scheduling and monitoring workflows. By focusing on the challenges characteristic of MEP-intensive environments, the proposed approach addresses a critical gap in current construction management practices and offers new avenues for improving productivity, reducing errors, and enhancing coordination on construction sites.

Related works

In recent years, the literature on construction progress monitoring has increasingly focused on the development of digital pipelines integrating CV, BIM, Digital Twin (DT), and AR to support near real-time project control, effectively rendering traditional manual, episodic, and subjective approaches obsolete. Recent review studies highlight that the topic is no longer limited to the automation of data capture but is increasingly oriented toward the development of closed-loop control systems, in which as-built observations directly inform the comparison with as-planned data through BIM. Within this context, traditional monitoring methods based on site inspections and manual reporting are still widely used but are considered inadequate for complex and dynamic projects, particularly when frequent and objective assessment of progress is required (Reja et al., 2022; Pal et al., 2023; Samsami, 2024).

Research on CV applied to the construction industry primarily focuses on the comparison between as-built and as-planned conditions through BIM models. Recent advancements show a shift from simple recognition of whether an activity has been completed to more granular, activity-based methods. A representative example is the work of Yang et al. (2023), who proposed a YOLOv8-based system for the automated monitoring of window installation progress. Several studies emphasize that the reliability of CV models is strongly influenced by site conditions, such as lighting variations or the presence of occluding objects that may interfere with visual inspection, as well as by the limited availability of annotated datasets and the difficulty of recognizing partially installed elements (Yu et al., 2025; Yue et al., 2025). In the MEP domain specifically, point cloud-based approaches have shown that deep learning networks can accurately segment pipe components even from poorly scanned data, provided that the domain gap between BIM generated synthetic training data and real as-built scans is explicitly addressed through feature and label alignment strategies (Yu et al., 2025). In particular, point clouds of MEP scenes are inherently affected by severe occlusions due to the dense overlapping of components, which degrades the performance of downstream recognition and classification tasks if left unaddressed (Yue et al., 2025). Both Yu et al. 2025 and Yue et al. 2025 demonstrate that BIM models can serve as a scalable and cost-effective source of synthetic annotated training data, directly mitigating the scarcity of labeled industrial datasets that represents one of the primary bottlenecks in applying CV to MEP environments (Yu et al., 2025; Yue et al., 2025).

Pal et al. (2024) propose a framework, which combines BIM, semantic segmentation, and ortho-view synthesis, also leveraging Neural Radiance Fields (NeRF), to estimate activity completion percentages with an average error below 6%, overcoming the limitations of methods that only detect the presence or absence of elements. Many studies aim to optimize the acquisition of point clouds for integration and comparison with BIM models. Some approaches involve SLAM-based mobile mapping devices to capture point clouds and panoramic images in indoor environments. Vassena et al. (2023) highlight the potential of more continuous workflows but also the persistence of issues related to semantic interpretation and automated understanding of point clouds, which negatively affect the real-time capabilities of DT systems. Other research explores neural rendering techniques and lightweight, photorealistic 3D representations. Studies on NeRF show that construction site scenes can be used not only for visualization but also for BIM comparison and progress estimation, often relying on handheld smartphone-captured video (Jeon et al., 2024). More recent work is also exploring 3D Gaussian Splatting and hybrid workflows to improve coordination, immersive visualization, and the quality of as-built versus as-planned comparison (Liao et al., 2025). These developments indicate that CV for monitoring is no longer limited to 2D image detection and segmentation but is increasingly merging with scene representation techniques, digital twin systems, and visual analytics.

Regarding AR technologies, recent literature confirms that their main value lies not only in the visual overlay of BIM models onto the real environment, but also in their ability to make project information, construction sequences, geometric deviations, and coordination issues immediately interpretable on-site (Sidani et al., 2021). The review by Bhatarai et al. (2025) describes BIM-AR

integration as a digital-physical bridge extending across the lifecycle, with expected benefits in design optimization, rework reduction, and contextual data access. At the application level, recent studies show that AR integrated with scheduling and BIM can support real-time progress tracking (Arvikar et al., 2025), while in the MEP domain AR has proven particularly promising for verifying complex installations, enabling direct comparison between the as-built condition and design geometry within the physical environment (Girgin et al., 2023).

More recently, research has addressed the convergence of CV, BIM, DT, and AR. These technologies are no longer treated as independent modules but as components of a unified monitoring and inspection architecture. For instance, Tao et al. (2024a) propose a passive AR inspection framework that integrates object detection, 3D projection, digital twin-driven recognition, and AR visualization. Tao et al. (2024b) investigate different methods and their combinations for recognizing construction progress directly through AR, by overlaying BIM models onto captured images. At a broader scale, recent studies show that the integration of CAD, BIM, immersive technologies, and advanced scanning techniques can yield measurable benefits even in highly complex structural and MEP coordination scenarios, reducing rework and review times (Liao et al., 2025).

Within this context, the present work aligns with the most recent research direction by proposing an integrated AR-CV framework, in which AR is not used solely as a visualization tool, but as an operational interface for construction progress recognition, task visualization, and real-time monitoring, contributing to a more immediate and integrated connection between site conditions, BIM-based planning, and decision-making support.

AR-CV based Construction Management

The proposed approach consists of three modules: 1) Planning with AR, where the structure and level of detail is defined. 2) Scheduling with AR, where based on module "1 Planning" the components/objects that are scheduled for installation are visualized in AR. 3) Monitoring with AR and CV where the scheduled objects/components in AR are compared with the installed objects in reality by means of CV. In the proposed approach scheduling is based on monitoring implementing the so- called "rolling scheduling" approach. In this way, the construction work is scheduled in increasing levels of detail as the project progresses, rather than defining the entire schedule in detail from the start. The system was developed in Unity 2020.3.42f1 by using the Mixed Reality Toolkit (MRTK) 2.8 and deployed on the Head Mounted Display (HMD) Microsoft Hololens 2. Figure 1 shows the AR-CV based approach for construction management.

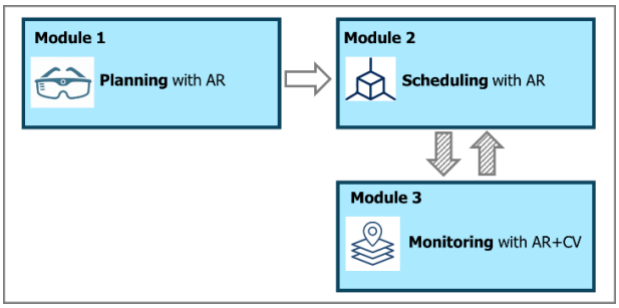


Figure 1: Schematic representation of the AR-CV based approach for construction management

Module 1 – Planning with AR

Module 1 is used to structure and define the Level of Detail (LoD) of the BIM, ensuring consistency between data acquisition, modelling, and downstream visualization processes. The definition of LoD within this module establishes clear modelling requirements aligned with the intended use of the digital model, particularly for AR visualization and CV–based monitoring. As a first step, a digital point cloud of the existing environment was generated using the Leica BLK2GO 3D laser scanner. This mobile scanning device enabled rapid acquisition of high-density spatial data while maintaining sufficient geometric accuracy for BIM modelling purposes. The resulting point cloud served as the primary geometric reference for the creation of the BIM model, which was developed using ArchiCAD version 27. The modelling process adhered to the LoD specifications defined in Module 1, ensuring that only the necessary geometric and semantic information was included.

The point cloud–based BIM model constitutes the foundation for the digital model used in AR visualization. By directly deriving the BIM geometry from the scanned data, spatial coherence between the physical environment and the virtual model is preserved, which is essential for accurate AR alignment and user interaction. This approach supports both visualization and spatial validation. In parallel, the planning module was employed to define the object structure and coding system required for AR visualization and monitoring. Each BIM object was assigned a standardized classification and unique identifiers, enabling seamless integration with the CV based monitoring approach. This structured coding framework allows selected elements to be visualized in AR and subsequently monitored in real time, supporting object recognition, status tracking, and progress analysis.

Module 2 - Scheduling with AR

Figure 2 shows the interface for scheduling and monitoring the construction project by using the HMD Microsoft HoloLens 2. The buttons on the left side are used to show the scheduling for a specific week (KW = calendar week in German), the "tig" indicates if the installation of the scheduled week had been completed and by using the "Undo" sign the completion can be revoked. An important aspect is the accurate aligning and anchoring of the digital model to the real-world environment so that virtual objects appear in the correct

position, orientation, and scale. This is also called AR registration. For this purpose, the Microsoft QR Code Tracking Package in Unity was used and 2 QR codes were set for the AR registration.



Figure 2. Interface for scheduling and monitoring by using the HMD

Module 3 - Monitoring with AR and CV

Module 3 is used to monitor the installation of the components on-site based on the scheduling. The complete system operates as a spatial correspondence estimator to quantify physical construction progress relative to the digital model through a multi-stage computation pipeline. The workflow begins with QR code detection, which establishes a spatial anchor that links Windows Perception Spatial Graph coordinate system to Unity's reference frame. This alignment ensures that the reference BIM and the spatial reconstruction produced are expressed in a common coordinate frame. Such normalization is essential as overlap estimation, and voxelization requires the same origin, orientation, and scale. To represent the real world, we used the HoloLens's 2 SLAM algorithm that generates view-dependent spatial mesh representing partial observations of the physical environment (SLAM, 2026).

In contrast, the reference BIM is deterministic and hierarchically structured, consisting of multiple geometric components that together encode construction elements, assemblies, and sections. Upon user-triggered comparison, both the reference geometry and the sensed spatial reconstruction are discretized through voxelization at a fixed spatial resolution of 5 centimeters. This process converts continuous surface meshes into discrete spatial occupancy representations that enable efficient comparison. Voxelization relies on the triangle-box intersection based on the Separating Axis Theorem (SAT), ensuring accurate surface coverage (Ericson, C., 2004). The SAT is a collision detection method, which determines intersection by testing for the existence of a separating axis such the projected intervals of the objects do not overlap. The system then computes the intersection over union (IoU) between the aggregated reference representation and the sensed environment to quantify spatial correspondence. Specifically, for each hierarchical node n , IoU is defined as the ratio between the number of voxels shared by the reference and sensed representations and the total number of voxels belonging to either representation. The reference voxel set V_n represents the aggregated spatial occupancy of node n in the reference hierarchy, while V_s denotes the voxel set corresponding to the sensed environment. Matched and total occupancy information is subsequently propagated recursively from lower-level components to higher-level nodes through bottom-up aggregation in the scene graph. Equation 1 shows the IoUn metric where n indexes a hierarchical reference node.

$$IoU_n = \frac{(|V_n \cap V_S|)}{(|V_n \cup V_S|)} \quad (1)$$

This bottom-up aggregation can be interpreted as a completion metric that estimates the degree of spatial correspondence for each construction unit. The resulting completion values are interpreted as empirical indicators of construction progress rather than exact geometric truth, explicitly accounting for the downward bias introduced by incomplete sensing. Module 3 was tested exclusively under laboratory conditions due to constraints encountered during the case study implementation. Immediately after installation, the pipes were covered as part of the standard construction process, preventing direct visual access required for the deployment and validation of the proposed approach on-site. As a result, laboratory testing was adopted to ensure controlled conditions for evaluating the functionality and performance of module 3.

Testing

The proposed approach was pilot tested partly in a laboratory setting and partly on-site within the construction of a large winery production and storage facility, where the construction intervention was focused on the realization of large underground cellar spaces dedicated to bottle storage. The case study company is a small enterprise located in Northern Italy and specialized in Heating, Ventilation and Air Conditioning (HVAC). The project considers the execution of technical installations within extensive subterranean open-plan environments. The industrial partner involved in the study was responsible for the on-site assembly of building services systems MEP, within these spaces.

Figure 3 shows the structuring of the BIM model as basis for the scheduling and monitoring of the case study. The size of section 1 is 17.80 m x 21.70 m, section 2 has a size of 13.50 m x 39.00 m and section 3 has a size of 13.00 m x 36.00 m.

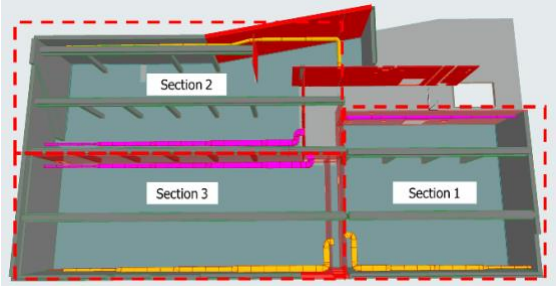


Figure 3. BIM model and structure of the case study

This case study introduced two main challenges: (i) the exceptionally large scale of the environments, characterized by open spaces, and (ii) the limited lighting conditions present during on-site testing activities. The solution deployed on site demonstrated high accuracy in hologram-to-reality alignment, maintaining geometric consistency between the physical environment and the digital model. However, the HMD initially experienced reduced spatial tracking performance due to insufficient illumination. This critical aspect was successfully addressed by installing temporary artificial lighting, which restored stable tracking conditions and enabled the continuation of the experimental activities.
Figure 4 shows module 2 and specifically the weekly scheduling in AR of the installation of air exchange pipes in sector 3 by using the Microsoft Hololens 2 HMD.

The results demonstrate that the AR–based visualization has great potential, even when applied to large-scale construction sectors. The spatial alignment between the digital model and the physical environment enables immediate identification of clashes and intersections among building elements, supporting rapid detection of potential conflicts. Furthermore, according to the foreman of the case study company, the integration of scheduling information into the visualization environment made construction sequencing tangible, allowing temporal aspects of the project to be intuitively understood and facilitating improved comprehension of planned activities and their spatial implications.



Figure 4. Module 2 weekly scheduling in AR of the installation of air exchange pipes in sector 3

Figure 5 shows the testing of module 3 in the laboratory environment.

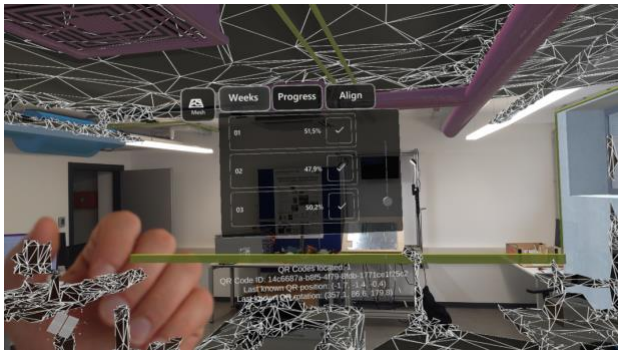


Figure 5. Testing of module 3 in laboratory settings

Figure 6 shows the 5 main component groups that were used to test module 3: A) The fan coil unit, B) the fan pipe supply connected to the unit, C) the wall channel for covering electrical wires, D) electrical pipes for the ventilation system and the writing desk plugins and E) the air exchange pipe system.

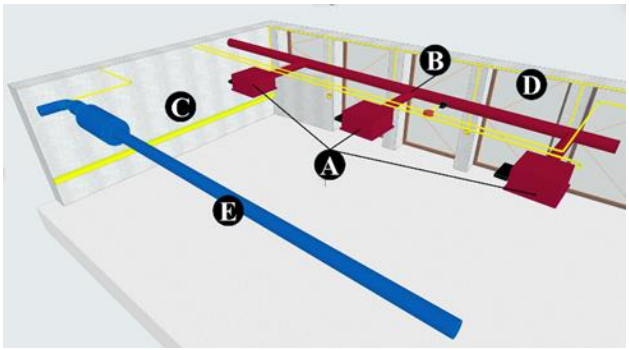


Figure 6. Component groups used for the testing of module 3

Table 1 shows the results of the lab tests where the IoU indicator was used according to Equation 1. Two tests in August 2026 were done.

Table 1: Lab test results of module 3

Object name and acronym	Intersection over Union IoU [%] – test 1	Intersection over Union IoU [%] – test 2
Fan coil unit (A)	48.00	77.80
Fan pipe supply (B)	78.10	62.30
Wall channel (C)	75.00	37.50
Electrical pipes (D)	54.7	45.60
Air exchange pipe (E)	91.4	63.60

According to the test results, electrical pipes D) and wall channels C) exhibit comparatively lower values of the Intersection over Union (IoU) ratio, indicating greater difficulty in achieving accurate spatial overlap for these elements. This behaviour can be attributed to their reduced cross-sectional dimensions and partial occlusions within the scene. In contrast, air exchange pipes E) and fan supply components B) demonstrate relatively high IoU values, reflecting more reliable detection and spatial correspondence. The larger geometric dimensions and clearer visibility of these elements contribute to improved detection, confirming the sensitivity of the IoU indicator to object scale and visual accessibility.

Discussion

The results indicate that the proposed AR–CV system enables a significantly faster detection of potential planning and execution errors when compared to traditional inspection and coordination methods. By superimposing digital information directly onto the physical environment, discrepancies between the planned and the as-built conditions can be identified in real time,

supporting prompt corrective actions and reducing the risk of error propagation during construction.

The findings further demonstrate that the visualization approach remains effective even in large-scale construction projects. Despite the increased spatial complexity and scale, the system maintained sufficient clarity and spatial accuracy, allowing users to navigate extensive construction sectors while preserving meaningful contextual information. This highlights the scalability of the AR-based visualization framework for complex construction environments.

The proposed AR–CV system addresses key limitations in current methods for scheduling and monitoring MEP construction projects. Unlike traditional approaches that rely heavily on large, annotated datasets, which are often scarce or costly to obtain in construction contexts, the proposed approach operates without the need for pre-labeled training data, enhancing its scalability and practical deployment. Furthermore, the AR–CV system tightly integrates scheduling with real-time monitoring, enabling dynamic adaptation to the inherent variability and uncertainty of construction processes. This interlinking facilitates more responsive progress tracking and decision-making. Additionally, while existing CV approaches struggle to accurately recognize partially installed components, Module 1 of the proposed system introduces a flexible level-of-detail definition, allowing robust detection and classification even in intermediate installation states. Collectively, these features contribute to a more adaptable and reliable solution for managing complex MEP construction workflows.

However, the analysis also reveals that the perspective of visualization plays a critical role in the evaluation of the IoU indicator (Table 1). Variations in viewing angle and distance influence the apparent overlap between detected objects and their corresponding digital representations, thereby affecting IoU values. This becomes also evident considering the discrepancies between test 1 and test 2 visualized in Table 1. This suggests that IoU-based performance assessment in AR–CV applications should account for viewpoint-dependent effects.

Moreover, the performance of the AR–CV construction management system was found to be highly sensitive to low-light conditions. Insufficient illumination negatively impacts image quality and feature extraction, leading to reduced detection accuracy and lower IoU values. This limitation highlights the need for controlled lighting conditions or the integration of additional sensing modalities to improve robustness in real-world construction environments.

Finally, the results indicate that the IoU indicator is particularly sensitive to objects that are not spatially free or detached on all sides, such as elements attached to walls or partially occluded components. In such cases, limited visible surfaces reduce the measurable overlap between detected and reference objects, leading to lower IoU values. This observation suggests that complementary evaluation metrics may be required to fully assess system performance for attached or constrained building elements.

Conclusion and outlook

Mechanical, Electrical, and Plumbing (MEP) systems involve spatial coordination and often complex installation sequences, making schedule adherence and progress monitoring particularly challenging in construction projects. This paper proposes a framework that leverages Augmented Reality (AR) and Computer Vision (CV) to enhance scheduling support and real-time monitoring in MEP construction. The approach combines BIM-based planning data with on-site AR visualization to provide field personnel with intuitive, context-aware guidance on installation tasks and sequencing constraints. By using CV algorithms an automatically detection of MEP components from site images and video streams becomes possible. This enables objective progress measurement and early identification of deviations from planned schedules. By fusing AR-enabled human-centric decision support with automated CV-based performance tracking, the system facilitates more accurate schedule updates, reduces rework caused by miscoordination, and improves communication on-site. The paper proposes a prototype that has been developed in Unity for the Head Mounted Display HoloLens 2. The AR-CV prototype had been pilot tested partly in a laboratory setting and partly on-site by using as case study the construction of a large winery production and storage facility in northern Italy.

Future research will focus on further enhancing the monitoring capabilities of the proposed system. First, Module 3 will be improved to understand the user's point of view, increasing robustness to viewpoint misalignment and occlusions during on-site data capture. Second, complementary evaluation metrics will be investigated to more comprehensively assess system performance, particularly for attached or geometrically constrained building elements where standard detection measures may be insufficient. Finally, extensive empirical testing of the monitoring module will be conducted on active construction sites to evaluate its accuracy, reliability, and practical usability under real-world conditions.

Acknowledgments

The authors would like to express their acknowledgement to the company Ruedl Hans OHG for valuable contributions to this research. The research was co-funded by MIMIT - Ministero delle Imprese e del Made in Italy in the context of the project Digital Innovation Hub South Tyrol (DIS-HUB 28-31), CUP B57H22006510001 - B57H22006500008.

References

Arvikar, S., Aung, P. P. W., Cha, G., & Park, S. (2025). Augmented Reality-based construction site management using optimized BIM and project schedule integration. Automation in Construction, 175, 106204. https://doi.org/10.1016/j.autcon.2025.106204

Bhatarai, R., Banihashemi, S., Shakouri, M., & Antwi-Afari, M. (2025). Integration of Augmented Reality with Building Information Modeling: Design Optimization and Construction Rework Reduction Perspective: R. Bhatarai et al. Archives of Computational Methods in Engineering, 32(4), 2233-2254. https://doi.org/10.1007/s11831-024-10211-6

Billinghurst, M., Clark, A., & Lee, G. (2015). A survey of augmented reality. Foundations and Trends in Human-Computer Interaction, 8(2-3), p.pp. 73-272. https://doi.org/10.1561/1100000049

Cantarelli, C. C., Flybjerg, B., Molin, E. J., & Van Wee, B. (2013). Cost overruns in large-scale transportation infrastructure projects: Explanations and their theoretical embeddedness. arXiv preprint arXiv:1307.2176. https://doi.org/10.48550/arXiv.1307.2176

Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). BIM Handbook. Wiley.

Ericson, C. (2004). Real-Time Collision Detection (1st ed.). CRC Press. https://doi.org/10.1201/b14581

Flyvbjerg, B., Holm, M. S., & Buhl, S. (2002). Underestimating costs in public works projects: Error or lie?. Journal of the American planning association, 68(3), p.pp. 279-295. https://doi.org/10.1080/01944360208976273

Girgin, S., Fruchter, R., & Fischer, M. (2023). A case study towards assessing the impact of mixed reality-based inspection and resolution of MEP issues during construction. Journal of Information Technology in Construction, 28. https://doi.org/10.36680/j.itcon.2023.029

Hartmann, T., Gao, J., & Fischer, M. (2008). Areas of application for 3D and 4D models on construction projects. Journal of Construction Engineering and Management, 134 (10), p.pp. 776-785. https://doi.org/10.1061/(ASCE)0733-9364(2008)134:10(776)

Jeon, Y., Kulinan, A. S., Tran, D. Q., Park, M., & Park, S. (2024, June). Nerf-con: Neural radiance fields for automated construction progress monitoring. In ISARC. In Proceedings of the International Symposium on Automation and Robotics in Construction, Lille, France (pp. 3-5). https://doi.org/10.22260/ISARC2024/0149

Kamat, V. R., Martinez, J. C., Fischer, M., Golparvar-Fard, M., Peña-Mora, F., & Savarese, S. (2011). Research in visualization techniques for field construction. Journal of Construction Engineering and Management, 137(10), 853-862. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000262

Liao, X., Liu, Y., Huang, Y., Luo, Y., Zhou, Z., & Chong, H. Y. (2025). Integrating CAD, BIM, immersive technology, and 3D Gaussian Splatting for construction model coordination with ISO 19650. Scientific Reports. https://doi.org/10.1038/s41598-025-32244-y

Pal, A., Lin, J. J., Hsieh, S. H., & Golparvar-Fard, M. (2023). Automated vision-based construction progress monitoring in built environment through digital twin. Developments in the Built Environment, 16, 100247. https://doi.org/10.1016/j.dibe.2023.100247

Pal, A., Lin, J. J., Hsieh, S. H., & Golparvar-Fard, M. (2024). Activity-level construction progress monitoring through semantic segmentation of 3D-informed orthographic images. Automation in Construction, 157, 105157. https://doi.org/10.1016/j.autcon.2023.105157

Reja, V. K., Varghese, K., & Ha, Q. P. (2022). Computer vision-based construction progress monitoring. Automation in Construction, 138, 104245. https://doi.org/10.1016/j.autcon.2022.104245

Samsami, R. (2024). A systematic review of automated construction inspection and progress monitoring (ACIPM): applications, challenges, and future directions. CivilEng, 5(1), 265-287. https://doi.org/10.3390/civileng5010014

Sidani, A., Dinis, F. M., Duarte, J., Sanhudo, L., Calvetti, D., Baptista, J. S., Martins, J.P., & Soeiro, A. (2021). Recent tools and techniques of BIM-Based Augmented Reality: A systematic review. Journal of Building Engineering,42,102500. https://doi.org/10.1016/j.jobe.2021.102500

SLAM, (2026). Microsoft Hololens Spatial Mapping. URL: https://learn.microsoft.com/en-us/windows/mixed-reality/design/spatial-mapping (accessed on the 12.01.2026)

Tao, B., Li, J., & Bosché, F. (2024a). Smart Automatic Mixed Reality-Based Construction Inspection Framework. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 41, pp. 776-783). IAARC Publications. https://doi.org/10.22260/ISARC2024/0101

Tao, B., Bosché, F., & Li, J. (2024b). Mixed Reality-based MEP construction progress monitoring: Evaluation of methods for mesh-to-mesh comparison. Automation in Construction, 168, 105852. https://doi.org/10.1016/j.autcon.2024.105852

Vassena, G. P., Perfetti, L., Comai, S., Mastrolembo Ventura, S., & Ciribini, A. L. (2023). Construction progress monitoring through the integration of 4D BIM and SLAM-based mapping devices. Buildings, 13(10), 2488. https://doi.org/10.3390/buildings13102488

Yang, J., Wilde, A., Menzel, K., Sheikh, M. Z., & Kuznetsov, B. (2023). Computer vision for construction progress monitoring: A real-time object detection approach. In Working Conference on Virtual

Enterprises (pp. 660-672). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-42622-3_47

Yu, W., Shu, J., Yang, Z., Ding, H., Zeng, W., & Bai, Y. (2025). Deep learning-based pipe segmentation and geometric reconstruction from poorly scanned point clouds using BIM-driven data alignment. Automation in Construction, 173, 106071. https://doi.org/10.1016/j.autcon.2025.106071

Yue, H., Wang, Q., Yan, Y., & Huang, G. (2025). Deep learning-based point cloud completion for MEP components. Automation in Construction, 175, 106218. https://doi.org/10.1016/j.autcon.2025.106218

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

PROPOSAL OF HUMAN-SENSORED ARCHITECTURE FRAMEWORK BY INTRODUCING MIXED REALITY

Kyung-Tae Lee[1]
[1]Kangwon National University, Chuncheon-si, Republic of Korea

Abstract

Remote work has increased the need to understand how indoor spatial design affects occupants' cognitive performance and physiological states. This study proposes a Human-sensored Architecture framework integrating Mixed Reality (HoloLens 2) for real-time spatial manipulation and EEG (Emotiv EPOC+) for physiological measurement. Following a network analysis of 124 publications, experiments with 30 participants revealed that red walls enhanced short-term executive function through stress-induced arousal, whereas preferred, low-stress settings improved long-term memory and productivity. These findings demonstrate the value of human-centered spatial design tailored to specific cognitive tasks and provide a foundation for personalized, adaptive architecture.

Introduction

Since the onset of the COVID-19 pandemic in 2020, the prevalence of remote work has increased significantly. As individuals spend more time indoors, there is a growing demand for environments that facilitate diverse tasks and activities effectively. Given that changes in the indoor environment profoundly affect the emotional and physiological states of occupants, research is continuously being conducted to identify which spatial configuration factors enhance user convenience and task efficiency (Lee et al., 2022). Historically, researchers have primarily focused on analyzing human states relative to environmental variables such as building facilities, temperature, and humidity to determine optimal conditions (Banaei et al., 2020). However, there remains a critical lack of analysis regarding the physiological responses to fundamental design elements—such as wall and floor colors, ceiling heights, and materials—that dictate the overall ambiance of a space (Lee et al., 2022). This gap is largely attributed to the technical limitations previously associated with modifying indoor designs in real-time.

Nevertheless, with the emergence of Extended Reality (XR) technologies, particularly Virtual Reality (VR) and Mixed Reality (MR), it has become possible to alter spatial designs instantaneously while incorporating real-time client feedback (Han et al., 2021). As a bridge between digital experiences and physical reality, these technologies—encompassing both AR and VR—serve as powerful tools for visualizing structures and environments (Portman et al., 2015). MR, in particular, has emerged as a significant paradigm in user interface systems, augmenting human perception by overlaying computer-generated information onto the real world. In this context, user interaction and immersion within the virtual environment are vital parameters that must be optimized for practical application (Chamilothori et al., 2019). While virtualization in MR enhances efficiency and productivity while mitigating risks (Portman et al., 2015), the most critical factor is interoperability. Since visual designs are superimposed on reality, maximizing interaction is essential. Furthermore, as MR devices often operate without external trackers or remotes, gesture-based recognition is paramount. Despite the potential of Head-Mounted Displays (HMDs) to enable accurate task performance and convenient daily life through 3D information, challenges remain. Mismatches or conflicts during the brain's information processing can induce cybersickness, leading to visual fatigue, anxiety, nausea, and disorientation (Im et al. 2022). Physical discomfort, such as the weight of the device, also necessitates further technological refinement (Rehring et al., 2019). Consequently, the MR market is expanding, with increasing adoption in the construction industry due to its interoperability and the convenience of controller-free operation.

To evaluate the physical and mental states of individuals within these spaces, psycho-physiological indicators are utilized. These indicators facilitate the assessment of human states through biological responses, identifying sympathetic and parasympathetic nervous system reactions to visual stimuli via metrics such as EEG, heart rate variability (HRV), and electrocardiogram (ECG) (Lee et al., 2022). Previous studies applying these metrics to the architectural field have found that white or green walls tend to induce a sense of psychological comfort, characterized by increased Alpha wave activity associated with reduced tension (Lee, Shin, and Lee, 2020). However, prior research has frequently focused on isolated components (e.g., a single wall or material) rather than the holistic design of a space. Moreover, while studies have analyzed physiological indicators to maintain a state of rest, research aimed at proposing optimal spatial designs for high task performance based on such indicators remains scarce.

Accordingly, this study aims to elucidate the relationship between indoor spatial design and human affective-physiological responses through empirical analysis. Since modifying physical indoor environments in real-time is practically infeasible due to construction and material constraints, Mixed Reality (MR, Microsoft HoloLens 2) was adopted as a specialized experimental apparatus. MR serves as a medium to overcome these limitations, enabling the instantaneous and controlled manipulation of spatial configurations to observe their direct impact on human states. To establish a robust theoretical foundation for the experimental design, a *bibliometric analysis* and network analysis were conducted. Based on the clusters derived from core descriptors, a theoretical framework was proposed, which subsequently guided the experimental protocol. Based on this framework, experiments were conducted with 30 participants across four spatial configurations: a baseline white space, a preferred space, a non-preferred space, and a self-selected space. During each session, two cognitive tests were administered to assess task efficiency. Concurrently, EEG data were collected and preprocessed to calculate Mental Stress and Mental Workload indices using established neuroscientific formulas. Through repeated measures statistical analysis, this study identifies fluctuations in affective-physiological indicators and task efficiency, ultimately proposing the optimal spatial configuration for specific task types.

Research Design

Methodology

This study aims to elucidate the intricate relationship between humans and spatial environments through empirical analysis. To achieve this, a theoretical framework was established via a literature review, followed by the development of an experimental environment capable of real-time modifications to indoor spatial designs. The review process involved curating scholarly publications from Web of Science and Scopus, with a subsequent network analysis performed on the extracted keywords. By identifying and synthesizing key thematic clusters, a comprehensive framework was constructed to map the connectivity between human subjects and spatial variables. Based on this framework, an experimental protocol was implemented. Technically, HoloLens-based mixed reality was utilized for spatial selection and manipulation. Human physiological responses were captured using electroencephalogram (EEG) to assess cognitive states, while spatial activities were defined and evaluated through task performance, enabling a multi-dimensional understanding of human reactions within diverse spatial configurations.

Establishment of experimental environment

Development of Theoretical Framework

To develop the proposed framework, research papers were extracted in accordance with a four-stage process. This process consists of the following: Stage 1, the selection of search engines and keywords; Stage 2, the screening and extraction of literature based on the selected keywords; Stage 3, quantitative literature analysis; and Stage 4, the analysis of results and the subsequent discussion process (Kim and Hyun, 2019). Initially, papers were retrieved from Web of Science and Scopus, which are recognized as representative academic databases. The temporal scope for the literature search was set from approximately 2000 to 2023. Specifically, in Web of Science, papers were extracted using the search string: TS=(interior AND space AND building) AND AK=(design) NOT AK=(Thermal AND Ventilation AND Energy). For Scopus, the search was conducted based on the query: ALL ("space" AND "building" AND "interior") AND KEY ("design") AND KEY ("interior") AND NOT ("Thermal") AND NOT ("ventilation") AND NOT ("energy").

Following this systematic screening as previously mentioned, a final total of 124 papers were selected, which provided the foundation for constructing the framework. In conducting the network analysis, a keyword co-occurrence network was constructed from the abstracts and keywords of the selected papers, and PageRank centrality (Brin and Page, 1998) was computed using the R programming environment. PageRank centrality was originally developed for ranking web pages based on hyperlink structures; however, it has been increasingly adopted in bibliometric research to evaluate the relative importance of keywords within co-occurrence networks, as the structural logic of interconnected nodes applies analogously to both domains (Ding et al., 2009). Through this approach, core keywords were identified based on their centrality scores, and the resulting network was subsequently visualized to delineate thematic clusters. Finally, after identifying the core values of the major clusters, the theoretical framework was established based on the synthesis of these cohesive factors.

Experiment Plan

The experimental protocol was meticulously designed to comprehensively reflect the components of the proposed framework. To investigate the fluctuations in affective-physiological indicators relative to indoor spatial design, it was essential to establish an experimental environment capable of real-time modifications. In this study, a Mixed Reality (MR) environment was introduced—a technology that enhances realism by superimposing virtual holographic information onto the physical environment (Im et al., 2022; Han et al., 2021). For the implementation of this MR setting, Microsoft HoloLens 2 was utilized, allowing for the manipulation of wall colors, materials, floor textures, and ceiling heights within a standardized indoor space measuring 5,500 mm × 7,000 mm × 3,600 mm.

Specifically, the wall color palette included white, blue, beige, green, and red. Flooring materials ranged from epoxy, stone, and wood to carpet, tile, and PVC, while wall finishes included brick, wood, paint, stone, concrete, and wallpaper. Additionally, the ceiling height was adjusted between high and low configurations. Based on a preliminary survey, participants were presented with three distinct spatial scenarios: a baseline white space, a

preferred space, and a non-preferred space. Furthermore, participants were encouraged to engage in decision-making by directly configuring their own preferred spaces. Findings indicated that participants predominantly preferred a configuration consisting of beige walls, PVC flooring, a wooden ambiance, and high ceilings, citing a sense of comfort and warmth. In contrast, a configuration featuring red walls, concrete and tile textures, and low ceilings was identified as non-preferred, with participants describing it as visually overstimulating and oppressive.

Subsequently, task efficiency within each spatial configuration was evaluated while concurrently measuring affective-physiological indicators. EEG data were collected using the Emotiv EPOC+ device, with 14-channel signals recorded in accordance with the international 10-20 system. This enabled the measurement of brainwave activity across four regions: the frontal, temporal, occipital, and parietal lobes (Lee, Shin, and Lee, 2020). Task performance and corresponding EEG patterns were analyzed through the Stroop test, which assesses short-term executive function, and the Reading Span Test, which evaluates working memory and long-term performance efficacy.

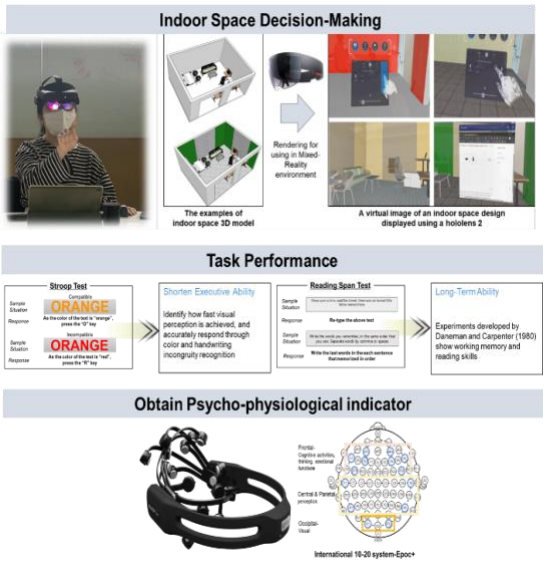


Figure 1. Process of experiment and activities

Data Collection

Experiment Participants

The experimental protocol of this study was conducted following approval from the Institutional Review Board (IRB) (Approval No. HYUIRB-2022202-005), after which participants were recruited. A total of 30 participants in their 20s and 30s were enrolled. The sample size was determined using G*Power 3.1 software and validated with reference to prior research in the field (Banaei et al., 2020). The participant pool comprised 30 Architectural Engineering students, balanced by gender (N=30; 15 males and 15 females). To eliminate potential bias arising from prior exposure, the selection was restricted to individuals who had never experienced Mixed Reality. Prior to the data collection phase, an introductory session on the MR device was conducted, followed by a 10-minute period for participants to independently adapt to the virtual environment. To ensure the integrity of the physiological data, all participants were screened to confirm they had no history of mental illness, were not suffering from sleep deprivation, and had no visual impairments. Furthermore, they were required to abstain from alcohol and smoking for 24 hours prior to the experiment. The Short Portable Mental Status Questionnaire (SPMSQ) was administered to verify that all participants possessed normal cognitive function. The SPMSQ, a 10-item screening instrument developed by Pfeiffer (1975) to detect cognitive impairment, was administered to all participants. The SPMSQ evaluates basic orientation, memory, and general knowledge through brief verbal questions of relatively low difficulty, serving as a preliminary screening tool rather than a comprehensive cognitive assessment. In this study, all 30 participants answered all items correctly (i.e., zero errors), confirming that the entire sample possessed intact cognitive function suitable for the experimental tasks. Before proceeding the experiment, participants wore the Microsoft HoloLens 2 and the Emotiv EPOC+ device, allowing them to engage in spatial decision-making tasks while their EEG signals were concurrently collected and recorded throughout the activities.

Data Pre-processing

First, to evaluate the task performance data obtained from the experiments, accuracy (%) and response time (ms) were recorded for each session. These metrics were subsequently converted into a 'Learning Ability Index' for comparative analysis, as defined in Equation (1).

$$\text{Learning Performance} = [(Accuracy)^{0.5} * \frac{1}{(Response\,time)^{0.5}}]^2 * 100 \quad (1)$$

Following this, electroencephalogram (EEG) was utilized as an affective-physiological indicator to analyze human responses. As a non-invasive method, EEG facilitates the practical measurement of neuronal activity and synaptic signals within the brain. The captured brainwaves were categorized into specific frequency domains to identify the underlying relationship between emotional states and indoor spatial configurations. For data processing, the frequency and spatial coordinates of each electrode were initially identified using the EEGLAB toolbox in MATLAB. Given that EEG signals are prone to artifacts such as eye blinks, muscle movements, and respiratory cycles, a rigorous preprocessing stage was essential to isolate the 2–39 Hz frequency range suitable for empirical analysis. As illustrated in Figure 2, this study employed the Blackman Window, characterized by low side-lobe levels and high frequency resolution. Through Independent Component Analysis (ICA) and Fast Fourier Transform (FFT), the signals were decomposed into distinct bands: Theta ($4-8$ Hz), Alpha ($8-13$ Hz), Beta ($13-30\ Hz$), and Gamma (30-39 Hz). Since lower frequencies are typically associated with resting states, these values were converted into 'Relative Power' to ensure an objective comparison of cognitive activity

(Kamzanova et al., 2014). Finally, two key indices established by previous researchers—Mental Workload and Mental Stress—were calculated to perform a spatial comparison, based on Equations (2) and (3), respectively.

$$\text{Mental workload} = \frac{(Relative\ Beta)_{Frontal}}{(Relative\ Theta)_{Frontal}+(Relative\ Alpha)_{Frontal}} \quad (2)$$

$$\text{Mental Stress} = \frac{(Relative\ Theta)_{Frontal}}{(Relative\ Alpha)_{Parietal}} \quad (3)$$

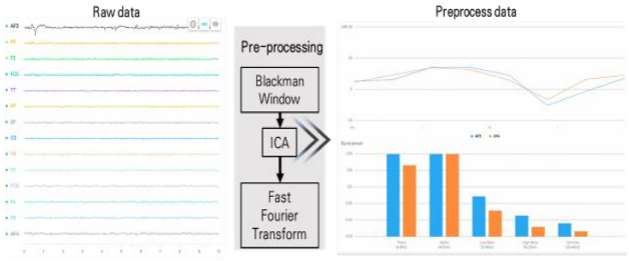


Figure 2 Data pre-processing method

Data Analysis

Statistical analysis in this study was performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA). To determine the appropriate statistical methods, the Kolmogorov-Smirnov (K-S) test was first conducted to assess the normality of the data distribution. For datasets that satisfied the normality assumption, a Repeated Measures Analysis of Variance (RM-ANOVA) was employed to compare EEG responses and task efficiency across the different indoor spatial configurations. Prior to the RM-ANOVA, the fundamental statistical assumptions were verified. First, the homoscedasticity of the dependent variables was confirmed, with all variables satisfying the requirement ($p \geq .05$). Second, Mauchly's test was employed to assess the assumption of sphericity. In instances where sphericity was violated, the Greenhouse-Geisser correction was applied to adjust the degrees of freedom and ensure the validity of the F-statistics (Yang et al., 2021). In instances where the normality assumption was violated, the non-parametric Friedman test was utilized for the comparative analysis (Han et al., 2021). To further identify specific differences between spatial conditions, post-hoc analyses were performed: Tukey's Honestly Significant Difference (HSD) test was applied following RM-ANOVA, whereas the Bonferroni correction was implemented for pairwise comparisons following the Friedman test.

Results

Proposal of theoretical Framework

As illustrated in Figure 3, the network analysis results reveal that the data are structured into four distinct primary clusters. Within each cluster, the size of the nodes (circles) corresponds to the frequency of the terms, indicating their significance as core keywords. As previously discussed, by employing PageRank centrality, the underlying main topics of each cluster were identified based on the inter-node connectivity between terms. Table 1 defines the representative keywords for each cluster, focusing on those descriptors that exhibit large node diameters and high degrees of connectivity as depicted in Figure 3.

Furthermore, terms associated with each primary keyword were systematically organized, providing a comprehensive basis for synthesizing the thematic focus and intended subjects represented by each individual cluster.



Figure 3 Result of Network Analysis

Based on a bibliometric analysis centered on Table 1, this study analyzed the interplay between humans and spatial design, called human-sensored architecture. The human component is broadly categorized into two dimensions: the cognitive dimension and the psycho-physiological reaction. Within the cognitive dimension, perception encompasses both innate temperament and accumulated experience, representing the diverse knowledge and empirical background individuals develop throughout their lives. Simultaneously, humans exhibit various psycho-physiological reactions in response to spatial configurations.

Table 1 Main topic of each cluster

Cluster	Main keyword	Other keyword	Main topic
Cluster 1 (red)	color	furniture object, light, textile, wall, height, surface, aesthetic, ceiling	Space Configuration Component
Cluster 2 (green)	experience	perception, psychology, behavior, decision, creativity, trait	Temperament of Human
Cluster 3 (blue)	technology	computer, quality, building information modeling, artificial intelligent, virtual reality technology	Smart construction technology
Cluster 4 (yellow)	human	comfort, satisfaction, safety, privacy, productivity, health	Status of human in indoor space

Specifically, as indoor environments undergo modifications, fluctuations are observed not only in physiological markers—such as electroencephalogram (EEG) and heart rate variability (HRV)—but also in psychological emotional states. Consequently, this study identified the cognitive dimension (conceptualized as the x-axis) and the psycho-physiological response (conceptualized as the y-axis) as the fundamental pillars influencing human spatial decision-making. Ultimately, the human state, which is modulated by the indoor environment, dictates the nature of human activity. These activities, typically manifested as work productivity or rest, necessitate that environmental configuration be proposed based on a comprehensive understanding of human factors. The resulting theoretical framework synthesized from these relationships is illustrated in Figure 4.

Differences in work efficiency by space

The results of the comparative analysis regarding short-term executive function and long-term working memory across the four spatial configurations (Baseline, Preferred, Non-preferred, and Self-selected/Decision-making space) are as follows. To determine the appropriate statistical method for evaluating task efficiency, the Kolmogorov-Smirnov (K-S) test was conducted. The results yielded p-values of 0.200 (D(30)=.113) for executive function and 0.178 (D(30)=.136) for working memory, confirming that both datasets satisfied the normality assumption. Consequently, a Repeated Measures ANOVA (RM-ANOVA) was performed to examine the differences based on spatial design, the results of which are summarized in Table 2. The RM-ANOVA revealed significant differences in both executive function and working memory depending on the spatial configuration. Regarding executive function, the highest performance levels were observed in the non-preferred space. Subsequent Tukey post-hoc analysis further elucidated that while performance was significantly higher in the non-preferred environment, no statistically significant differences were found between the baseline, preferred, and self-selected spaces. Conversely, the analysis of working memory showed that performance was lowest in the non-preferred space, whereas the highest performance was recorded in the self-selected (decision-making) space. Post-hoc results for working memory indicated that while no significant differences existed among the non-preferred, baseline, and preferred spaces, the self-selected configuration yielded a significantly superior performance outcome. In summary, the impact of indoor spatial design on task efficiency varied by cognitive demand: short-term executive function was enhanced in non-preferred environments, whereas long-term working memory performance was optimized in self-selected spatial configurations.

Differences in Psycho-physiological indicators by space

Following the preprocessing of EEG data, the physiological indicators derived from Equations (2) and (3) were subjected to normality testing. The Kolmogorov-Smirnov (K-S) test revealed that while Mental Stress satisfied the normality assumption ($D(30)=.109$, $p > .200$), Mental Workload violated it ($D(30)=.168$, $p = .030$). Consequently, Mental Stress was analyzed using RM-ANOVA, whereas the Mental Workload was evaluated via the non-parametric Friedman test. Regarding Mental Stress, Mauchly's test indicated that the sphericity assumption was satisfied for the executive function test. In contrast, the assumption was violated for the working memory test; consequently, the Greenhouse-Geisser correction was applied to adjust the degrees of freedom for the subsequent analysis. Figure 5 presents box plots illustrating the distribution of Mental Stress and Mental Workload across the different spatial configurations. In the case of Mental Stress during the executive function test, the highest levels were observed in the preferred space, followed by the self-selected and non-preferred spaces. Post-hoc analysis indicated no significant difference between the non-preferred and

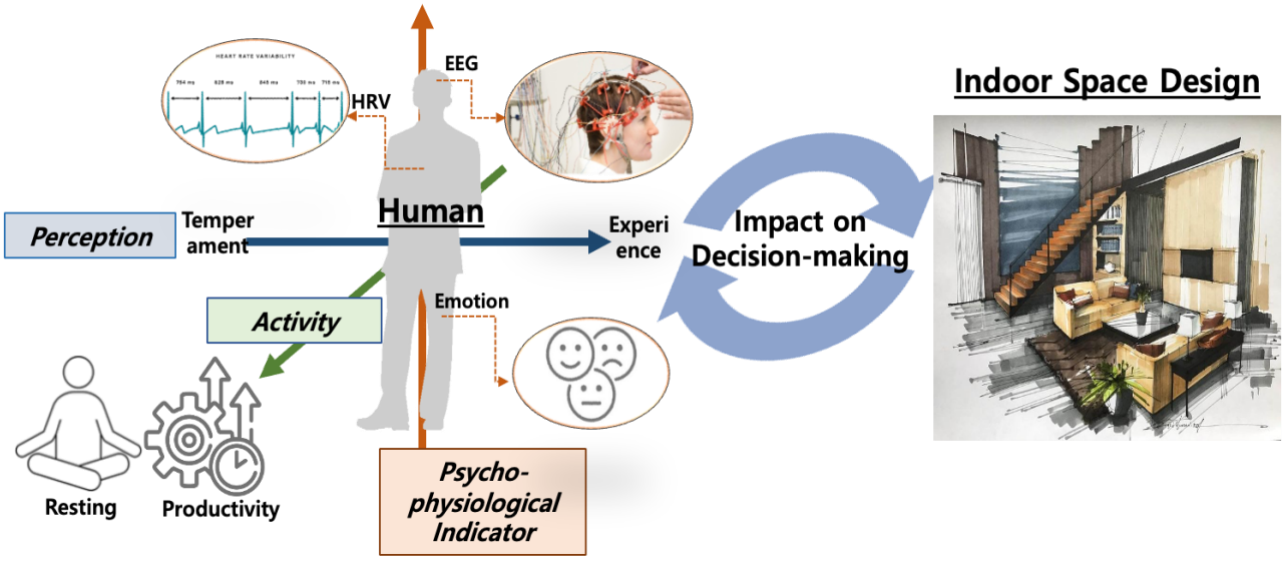


Figure 4. Proposal of Human-sensored Architecutre Framework

Table 2 Statistical analysis of taskperformance among space design

Task Performance	Interior design	Mean	Standard deviation	Tests of within subject effects					Post hoc
				Type Ⅲ Sum of squares	Df	Mean square	F-Statistics	Sig.	Tukey HSD
Executive ability	Basic	9.349	4.063	549.384	2.572	213.564	27.966	.000**	1=2=4<3
	Prefer	12.203	3.960						
	Non-Prefer	14.790	4.088						
	Personal	12.275	4.044						
Working Memory	Basic	1.046	0.753	2.319	2.600	.892	2.815	.042*	3=1=2<4
	Prefer	1.055	-.656						
	Non-Prefer	0.916	0.660						
	Personal	1.301	0.492						

Note *: p-value<0.05, **: p-value<0.001

baseline spaces; however, participants experienced significantly higher stress in the preferred space. Similarly, significant spatial variations in Mental Stress were observed during the working memory test. Post-hoc comparisons confirmed that the preferred space induced the highest stress levels. Notably, in contrast to the executive function results, stress levels in the self-selected space during the working memory task were found to be lower than those in the non-preferred space. Regarding Mental Workload, analyzed via the Friedman test, the results for the executive function test showed the most intense workload in the non-preferred space. Bonferroni-corrected post-hoc analysis confirmed that these differences were statistically significant, indicating that the non-preferred environment imposed a substantially higher cognitive load, while the workload in the self-selected space was relatively low. Conversely, although the workload during the working memory task appeared higher in the non-preferred space, this difference did not reach statistical significance.

Discussion

With the increasing emphasis on Digital Twin technology, interest in Mixed Reality (MR) has grown significantly. In the construction industry, MR has primarily been utilized for spatial visualization, often limited to safety training and education. This study, however, expands the application of MR by redirecting its focus toward spatial decision-making. A key differentiator of this research lies in its integrated presentation of fluctuating physiological indicators alongside task efficiency across diverse spatial configurations. While traditional approaches have focused on subjective impressions and qualitative expressions of space, future spatial planning will likely transition toward more complex, biometric-centered proposals (Lee et al., 2022). As evidenced by this study, the evolution of MR enhances the decision-making process for stakeholders, fostering greater coherence and streamlining communication (Zaker and Coloma, 2018). Despite current technical limitations—such as resolution constraints and sensitivity to lighting—these challenges are expected to be overcome in the long term. Given its capacity for real-time interaction and 3D visualization, MR is poised to become an efficient tool, particularly in the early stages of construction projects.

The findings confirm that visual information significantly influences biometric responses, which in turn affect behavioral and task performance. Interestingly, the preferred space showed the highest Mental Stress values. This result suggests that Mental Stress should not be interpreted solely as indicating a more adverse or uncomfortable condition. Rather, the non-preferred condition was more consistently associated with elevated Mental Workload, suggesting increased cognitive demand, whereas the preferred space may have elicited greater emotional arousal or attentional engagement. Notably, the superior task performance observed in non-preferred spaces regarding short-term concentration can be elucidated through the 'fight-or-flight response' theory (Cannon, 1915). This theory suggests that when individuals experience stress, the autonomic nervous

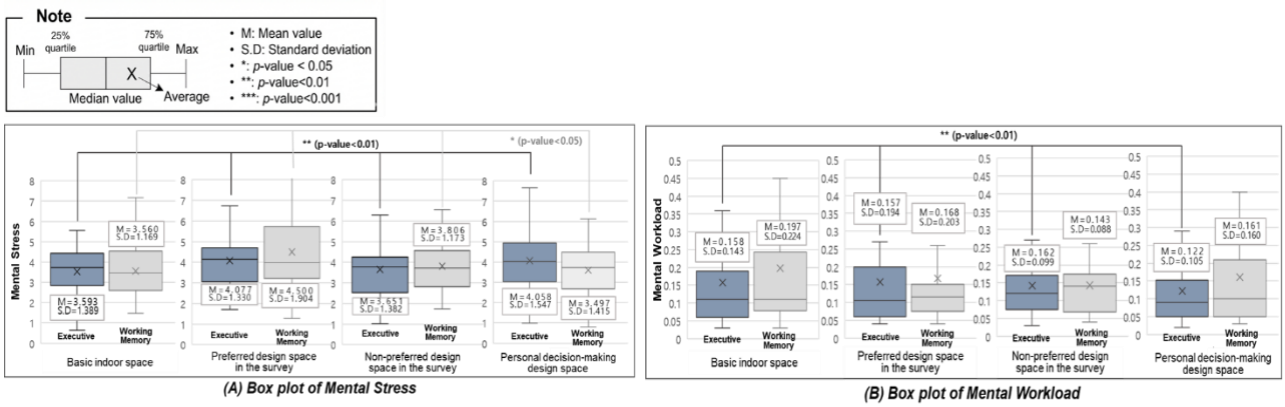


Figure 5 Boxplot of psycho-physiological index

system is activated to heighten focus on the stressor as a survival mechanism. Specifically, exposure to environments with red walls or low ceilings imposed a significantly higher cognitive workload, as evidenced by elevated Mental Workload indices. This heightened cognitive demand can be interpreted through the lens of the fight-or-flight response theory. According to this framework, the autonomic nervous system is activated under demanding conditions to sharpen focus as a survival mechanism, which may paradoxically enhance short-term executive function (Duyan & Ünver, 2016). This explains why participants exhibited improved attention despite the taxing nature of the non-preferred environment. However, considering that long-term capacity is better manifested in preferred spaces due to spatial adaptation, this research underscores the necessity of configuring environments tailored to specific individual competencies and task types. Future interior design should prioritize the emotional responses and cognitive demands of users, ensuring that individuals can successfully adapt their behavior within these optimized spatial configurations.

Conclusions

In conclusion, this study sought to analyze the variations in human affective-physiological responses and task efficiency according to real-time modifications in indoor spatial design. By implementing a Mixed Reality (MR) environment for 30 participants, the research facilitated the collection of EEG data and the comparative analysis of task performance across diverse spatial configurations. The findings revealed that in the short term, although participants experienced a higher cognitive workload in non-preferred spaces, they paradoxically reported lower stress levels and demonstrated superior task performance compared to preferred or self-selected environments. This suggests that visually stimulating or non-preferred environments may heighten acute attention, leading to increased output in brief, execution-oriented tasks.

However, the temporal dynamics of these effects shifted during long-term engagement. While no significant difference was initially observed between spatial design and workload during prolonged tasks, the non-preferred space eventually led to an accumulation of stress, resulting in a subsequent decline in task efficiency. In contrast, participants demonstrated a significant adaptation to self-selected (decision-making) spaces over time. In these configurations, stress levels decreased as individuals acclimated to their environment, which in turn facilitated higher long-term task efficiency. These results underscore that while non-preferred spaces may trigger higher short-term executive function, self-selected environments are optimized for sustained productivity and psychological well-being through spatial adaptation. Therefore, environmental configurations should be proposed dynamically, aligning with specific task types and the required levels of cognitive concentration (Cho et al., 2019).

Despite these insights, this study is subject to certain limitations, particularly its reliance on a Mixed Reality simulation. Future research should strive to validate these findings in full-scale physical environments through longitudinal action research. Furthermore, the scope of inquiry should be expanded by incorporating a broader array of affective-physiological indicators—such as heart rate variability (HRV) and electrocardiogram (ECG)—to achieve a more comprehensive understanding of human biological responses to the built environment.

Acknowledgments

This research was funded by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (grant number RS-2024-00351295).

References

Banaei, M., Ahmadi, A., Gramann, K., and Hatami, J. (2020). Emotional evaluation of architectural interior forms based on personality differences using virtual reality. Frontiers of Architectural Research, 9(1), 138-147.

Cannon, W. B. (1915). Alternative satisfactions for the fighting emotions.

Chamilothori, K., Wienold, J., & Andersen, M. (2019). Adequacy of immersive virtual reality for the perception of daylit spaces: comparison of real and virtual environments. Leukos, 15(2-3), 203-226.

Cho, S. H., Lee, K. T., Kim, S. H., & Kim, J. H. (2019). Image processing for sustainable remodeling: Introduction to real-time quality inspection system of external wall insulation works. Sustainability, 11(4), 1081.

Duyan, F., & Ünver, R. (2016). A research on the effect of classroom wall colours on student's attention. A| Z ITU Journal of the Faculty of Architecture, 13(2), 73-78.

Han, M., Baek, K., Lee, K. T., Ko, S., and Kim, J. H. (2021). A Study on Supporting Design Decision Making in Office Building Remodeling Projects by Introducing Mixed Reality. Korean Journal of Construction Engineering and Management, 22(1), 3-12.

Im, J. B., Hwang, S. J., Jang, J. H., Kim, J. H., & Lee, K. T. (2022). Introduction of mixed reality on special tree landscape design review phase. Journal of the Architectural Institute of Korea, 38(8), 71-81.

Kamzanova, A. T., Kustubayeva, A. M., & Matthews, G. (2014). Use of EEG workload indices for diagnostic monitoring of vigilance decrement. Human factors, 56(6), 1136-1149.

Lee, J. K., Shin, J., & Lee, Y. (2020). Circulation analysis of design alternatives for elderly housing unit allocation using building information modelling-enabled indoor walkability index. Indoor and Built Environment, 29(3), 355-371.

Lee, K. T., Im, J. B., Park, S. J., and Kim, J. H. (2022). Conceptual Framework to Support Personalized Indoor

Space Design Decision-Making: A Systematic Literature Review. Buildings, 12(6), 716.

Portman, M. E., Natapov, A., and Fisher-Gewirtzman, D. (2015). To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. Computers, Environment and Urban Systems, 54, 376-384.

Rehring, K., Greulich, M., Bredenfeld, L., and Ahlemann, F. (2019, January). Let's get in touch-decision making about enterprise architecture using 3D visualization in augmented reality. In Proceedings of the 52nd Hawaii International Conference on System Sciences.

Zaker, R., and Coloma, E. (2018). Virtual reality-integrated workflow in BIM-enabled projects collaboration and design review: a case study. Visualization in Engineering, 6(1), 4.

Pfeiffer, E., Médecins, U., & Pfeiffer, E. (1984). Short portable mental status questionnaire (SPMSQ). Source book of geriatric assessment, 1, 115-6.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems, 30(1-7), 107-117.

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. Journal of the American Society for Information Science and Technology, 60(11), 2229-2243.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

CHAT WITH PRE-DEMOLITION AUDITS: KNOWLEDGE-GRAPH BASED INFORMATION RETRIEVAL

Ioanna Mitropoulou[1], Beril Önalan[1], and Catherine De Wolf[1]
[1]ETH Zurich, Switzerland
Circular Engineering for Architecture.

Abstract

Although pre-demolition audits (PDAs) provide early information on building materials, they are underutilized partly because their data require domain expertise to access and interpret. This raises the question: can natural-language querying enable non-experts to retrieve information from PDA datasets? We propose a knowledge-graph-based Retrieval-Augmented Generation (RAG) framework for natural-language querying of PDAs, applied to a case study of 40 French audits. Evaluated against a semantic RAG baseline across 21 queries, Graph-RAG achieves a success rate of 90% compared to 57% using vector-based retrieval, though at 2.34× higher cost, demonstrating its potential to making PDA content accessible to non-expert stakeholders.

Introduction

Pre-demolition audits (PDAs) are increasingly recognized as a key source of information for enabling reuse (Wu et al., 2026; Davis et al., 2025). Produced before the building intervention (refurbishment or demolition), they capture materials, quantities, conditions, and potential reuse pathways. However, despite their early availability and potential to inform design, PDAs remain largely disconnected from early-stage reuse decisions and design workflows. In practice, architects and clients rarely consult PDAs during conceptual design, when material selections and building configurations are most flexible.

The primary barrier to the effective utilization of PDAs is semantic accessibility. These audits often consist of vast quantities of fragmented information spread across extensive structured tabular inventories (material quantities, dimensions, conditions) with unstructured textual reports (building context, accessibility constraints, intervention plans). Extracting insights from these heterogeneous data currently requires users to understand schema conventions and to have the technical proficiency to formulate precise filters or database queries. For non-expert stakeholders, this technical overhead acts as a significant deterrent to data-driven reuse, limiting accessibility, user agency, and the practical use of PDA data in decision-making.

This paper aims to bridge this gap by developing a natural-language interface that enables easy and intuitive access to the contents of these large and complex documents through natural-language queries. We propose a Knowledge-Graph-based Retrieval-Augmented Generation (Graph-RAG) workflow structured to address specifically the typical formats of PDAs, so their information becomes more accessible and interpretable to non-expert users. We compare it against a baseline RAG workflow, as shown in Fig. 1), and conclude the strengths and weaknesses of each system. With the proposed Graph-RAG workflow, we enable users to extract specific material details and building-level context through natural language, effectively democratizing access to reuse data for the architecture, engineering, and construction industry.

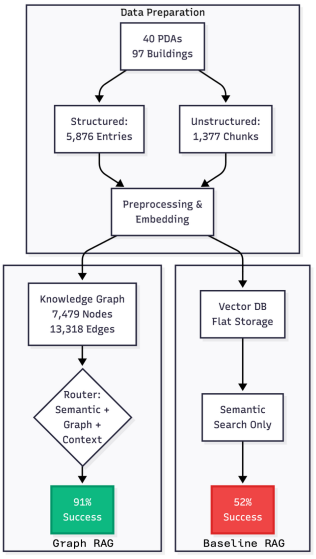


Figure 1: Overview of the proposed Graph-RAG framework for PDA data and its comparison with a baseline RAG approach.

Background

Pre-demolition audits in circular construction. PDAs have emerged as essential tools in circular construction for identifying, quantifying, and managing building materials at end-of-life stages to minimize end-of-life waste (Wu et al., 2026; Godina et al., 2025; Luciano et al., 2022). These audits systematically collect information about the type, quantity, quality, and accessibility of building materials, supporting selective demolition planning, material recovery pathways, and integration into digital building logbooks (Oluleye et al., 2023).

However, PDA datasets present significant challenges for

effective data retrieval. First, heterogeneity and inconsistency across different audits complicate data comparison, transferability, and generalization across different buildings (Akanbi et al., 2020). While recent research has addressed these issues through automated classification, structuring, and enrichment of heterogeneous data into unified schemas (Mitropoulou et al., 2026), challenges in querying remain. Second, PDAs consist of both structured data (tabular inventories with quantitative material information) and unstructured data (textual reports describing building conditions, accessibility, and contextual information), which are interlinked but difficult to query cohesively. This necessitates computational approaches that can jointly query both data modalities.

Natural-language interfaces for built-environment data. Natural-language interfaces have emerged as a critical mechanism for making complex built-environment data accessible to non-expert users. Retrieval-augmented generation (RAG) based on semantic similarity has been increasingly explored for knowledge-base querying (Li and Wang, 2026). RAG has become widely adopted in pre-trained language models to enhance performance in domain-specific tasks by dynamically accessing external knowledge bases, improving factual grounding and relevance in generated responses (Shang et al., 2025). However, standard RAG approaches face significant limitations when handling structured knowledge: they treat the data as isolated chunks of plain text, thus losing information about their relations. As a result, they fail to capture the global semantics and relational dependencies present in built-environment datasets (Procko and Ochoa, 2024; Xu et al., 2024).

To address these challenges, for tabular data specifically, recent methods have explored how LLMs can search, interpret, and reason over tables stored in comma-separated values (CSV) format. However, such LLM-to-CSV search methods suffer from similar loss of semantic information and high failure rates on complex queries (Sequeda et al., 2024). Knowledge graphs offer a promising alternative by representing heterogeneous information through semantically meaningful nodes and relationships. Knowledge graphs preserve the hierarchies and connections present in source data (Shang et al., 2025), enabling more successful retrieval of information through Graph-RAG (Li and Wang, 2026; Sequeda et al., 2024; Xu et al., 2024). Recent advances in Graph-RAG have demonstrated substantial improvements over standard RAG: the Vector-Graph RAG framework for querying semantic web building data achieved a 20-40% improvement in accuracy over state-of-the-art LLMs using traditional RAG (Li and Wang, 2026). Similarly, graph-based approaches for Building Information Modeling (BIM) data showed 75% accuracy compared to just 16.7% for direct LLM querying (Hsiu Tung et al., 2025). Additionally, Graph-RAG has been implemented to enable natural-language interaction with digital twins, demonstrating how such technologies can democratize access to complex built-environment information for non-expert users (Pan et al., 2026). However, to date, Graph-RAG has not been applied to datasets in which structured and unstructured information are semantically interdependent, as in PDA datasets.

Summary. Despite advances in structuring heterogeneous PDA data and developing natural-language interfaces for built-environment information, a gap remains in retrieval frameworks capable of jointly handling semantically linked *structured* data (e.g., tables) and *unstructured* data (e.g., textual reports). This paper addresses these limitations by proposing a Graph-RAG workflow that integrates structured and unstructured PDA data within a graph-based retrieval framework, enabling non-expert users to query this information in natural language.

Methods

We first describe the initial preprocessing of PDA data and the subsequent construction of the knowledge graph that combines structured and unstructured data. Following this, we introduce our retrieval algorithm, which uses a router to decompose the query into smaller steps. Each step is then executed through either semantic search or graph search. Finally, we present a comparative evaluation of our retrieval framework against a baseline semantic RAG interface (lacking access to the knowledge graph) by testing both strategies on a set of 21 questions.

Data preprocessing

The goal of the preprocessing step is to transform the raw PDAs into an accessible format suitable for subsequent knowledge-graph creation. Our case study consisted of a set of 40 PDAs sourced from different buildings across France that collectively describe entries from 97 distinct buildings. The information in these audits consisted of both structured data (tables), which describe building materials and products, and unstructured data (textual reports), which describe contextual information (Fig. 2).

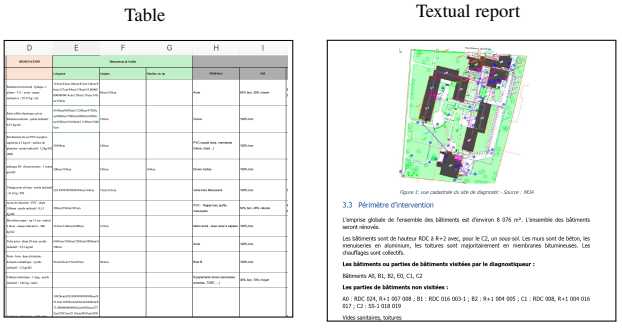


Figure 2: Example of the contents of a PDA. Left: structured input (table). Right: unstructured input (textual report).

Structured data preprocessing. To address the heterogeneity of the structured tabular data across different PDAs, we applied the methodology proposed by Mitropoulou et al. (2026). This process structured and classified the

entries according to a unified schema and classification system. This resulted in a CSV file containing over 7000 entries translated to English in a unified format. Following Mitropoulou et al. (2026), the entries were initially categorized into *Materials* and *Products*, each using a different classification system. For simplicity in the current study, we focused exclusively on the Product entries to maintain a single classification system, specifically the Uniformat II schema (Charette and Marshall, 1999), which provides a hierarchical structure consisting of three classification levels.

Unstructured data preprocessing. A subset of PDAs contained unstructured textual reports. These reports were first translated from French into English using OpenAI GPT-4o. Subsequently, the translated text was segmented into discrete text chunks, each limited to a maximum length of 1200 characters, with an overlap of 200 characters between neighboring chunks.

Knowledge-graph creation

Schema. To create the knowledge graph, we first defined the schema, which specifies the nodes, how they are connected, and which attributes each node stores. The schema includes the nodes *Project*, *Building*, *Entry*, *Classification*, *Report*, and *Text Chunk*, connected as shown in Fig. 3. We developed this model by first clarifying the purpose of the graph, its intended users, and the types of queries it should be able to answer. In our case, the dataset was designed for non-expert stakeholders seeking reliable information on documented materials to inform early-stage reuse decisions. Based on this, we manually identified the core entities and relationships present in the dataset and iteratively refined the structure by testing it on a small data sample. The attributes associated with each node type are listed in Table 1. The nodes Project, Building, Entry, and Text Chunk each store an embedding vector, enabling their contents to be retrieved through semantic search.

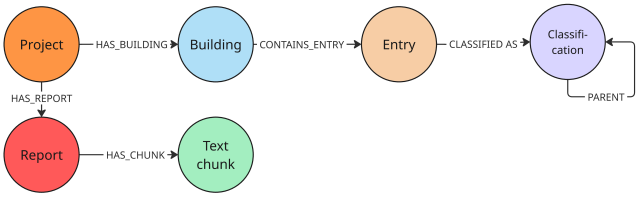


Figure 3: Schema of knowledge graph containing the following nodes: Project, Building, Entry, Classification, Report and Text chunk, connected with semantic edges.

Graph creation. The knowledge graph was constructed using the Graph Data Science Python Client (Neo4j, Inc., 2025a) to connect to Neo4j (Neo4j, Inc., 2025b), a web-based platform for building and querying knowledge graphs. We imported both the structured data (a CSV file containing the processed entries) and the unstructured data (text chunks), and created nodes for each entity defined in the schema, including their embedding vectors. We then added the semantic edges specified in the schema. An example of the resulting graph for a subset of entries belonging to a single building is shown in Fig. 4.

Table 1: Node types and their attributes in the knowledge graph. The nodes Project, Building, Entry, and Text Chunk each contain an embedding vector, so that their contents can be retrieved through semantic search.

py

Node	Attributes
Project	name, zipcode, **embedding vector**
Building	name, total area, usage, year of construction, **embedding vector**
Entry	entry id, assembly type, condition, description, notes, hazardous, quantitative data (mass, volume, dimensions, etc.), **embedding vector**
Classification	code, level
Report	report id, project name
Text Chunk	chunk id, chunk index, **embedding vector**

The full knowledge graph consisted of a total of 7,479 nodes and 13,318 relationships. Querying was performed using the OpenAI GPT-4o, while semantic representations were generated with the OpenAI `text-embedding-3-small` model, which produces 1,536 dimensional embeddings.

Knowledge-graph-based information retrieval

Our workflow for retrieving information using the knowledge graph is shown in Fig. 5 (proposed Graph-RAG) and compared with the baseline RAG workflow. The process begins with the user submitting a natural-language query. In the Graph-RAG, the query is routed through a sequence of steps, described below, aiming to assemble a comprehensive context that will produce an accurate answer.

LLM Router. The query is first handled by the router, an LLM-based function that decomposes the user query into a sequence of smaller retrieval steps. Each step is assigned to either a *graph search* or a *semantic search*, depending on which method is better suited to retrieve relevant context. The purpose of this decomposition is not to answer the question directly, but to make a plan for how to construct a maximally informative context that will enable answering the question at a later step. A few-shot router prompt instructs the LLM on the available tools, the criteria for selecting each tool, and examples of how complex queries should be broken down. For every generated sub-step, the router specifies:

- *Tool*: semantic search | graph search
- *Prompt*: a concise description of what to retrieve
- *Report context*: true | false (whether to also retrieve the most relevant linked report text chunks)

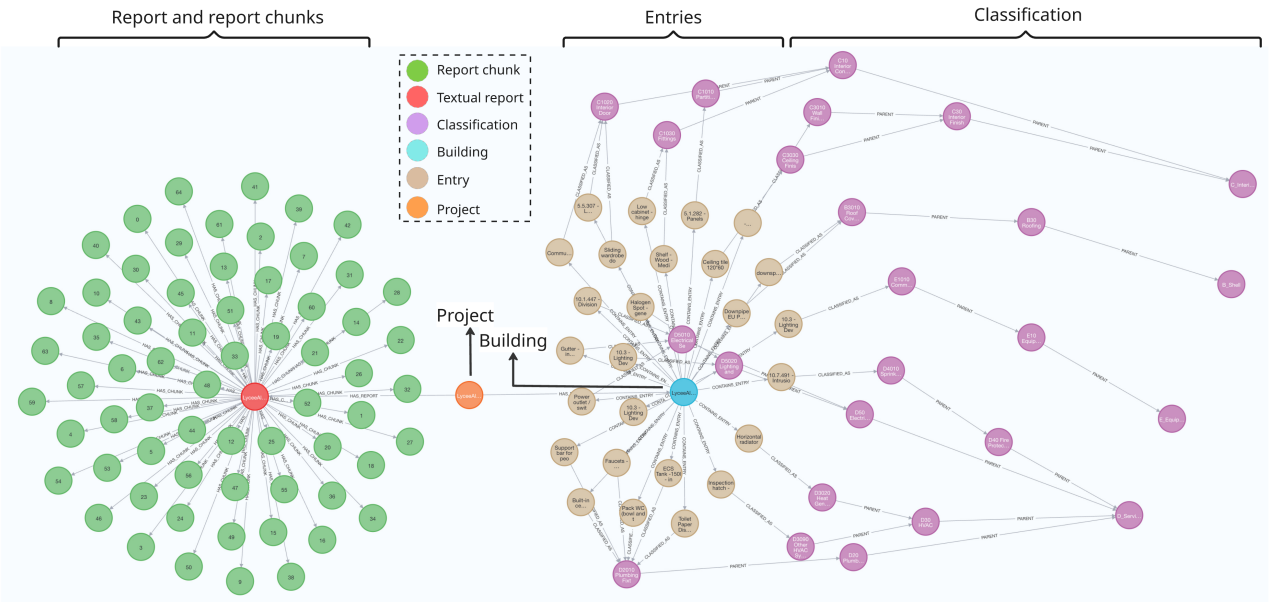


Figure 4: Example of a graph created for a subset of entries of a single building. Left: The Textual report chunk nodes (green) contain unstructured text parts of a Report node (red). Middle: A Building node (blue) is connected to a Project node (orange) and contains Entries (brown). Right: The Entry nodes are connected to their Classification nodes (purple).

Semantic search. The semantic-search component follows a standard RAG workflow. All nodes that contain embeddings (Projects, Buildings, Entries, and Textual Report Chunks) are searchable. For each step, the LLM semantic router determines which node type(s) should be queried. A similarity search is then performed, ranking nodes according to embedding similarity and selecting the top-k most relevant items, with $k = 20$.

Graph search. Graph search uses an LLM that generates Cypher code to traverse the knowledge graph, filter nodes, count elements, or retrieve specific relationships. The prompt for generating the Cypher code includes a description of the graph schema, relationships, and attributes, so that the generated queries are syntactically correct. Once generated, the Cypher query is executed to retrieve the required data.

Linked report context. For each retrieval step, the router may indicate that linked report information should also be included. In such cases, the algorithm starts from nodes obtained via semantic or graph search and traverses the graph to collect all linked textual report chunks. For example, when starting from a Building node, the traversal proceeds via the corresponding Project and its linked Report to retrieve all associated text chunks. If such chunks exist (not all projects have reports), they are semantically ranked, and the top-k most relevant fragments are added to the context.

Answering. Once the final context has been assembled by combining results from graph search, semantic search, and optionally linked report content, it is passed to the LLM along with the original user query and an answer prompt that instructs the model to base its output strictly on the provided context. The LLM then produces the final answer.

Evaluation

To evaluate the performance of the proposed approach, we designed a set of representative questions organized into five categories testing different competencies: (A) simple filtering, (B) counting, (C) more complex queries, (D) queries requiring report context, and (E) unanswerable questions (Table 2). Ground-truth answers were established through manual inspection of the dataset. The evaluation focused on end-to-end question-answering performance, that is, how well the final generated answer agreed with the established ground truth, without separately assessing the success of the intermediate retrieval steps. We compared the performance of our model against a baseline RAG system, described below.

Baseline RAG. The baseline consists of a standard RAG pipeline, illustrated in Fig. 5 left. Given a user query, the LLM semantic router determines which type(s) of searchable nodes containing embeddings to search (Projects, Buildings, Entries, or Textual Report Chunks). These nodes are then ranked according to semantic textual similarity. The top-k most relevant nodes are retrieved and incorporated into the context. For all semantic searches that retrieved the top-k most relevant nodes, we used $k = 20$.

Results

Evaluation questions

The evaluation questions, shown in Table 2, are organized into five categories that test different competencies.

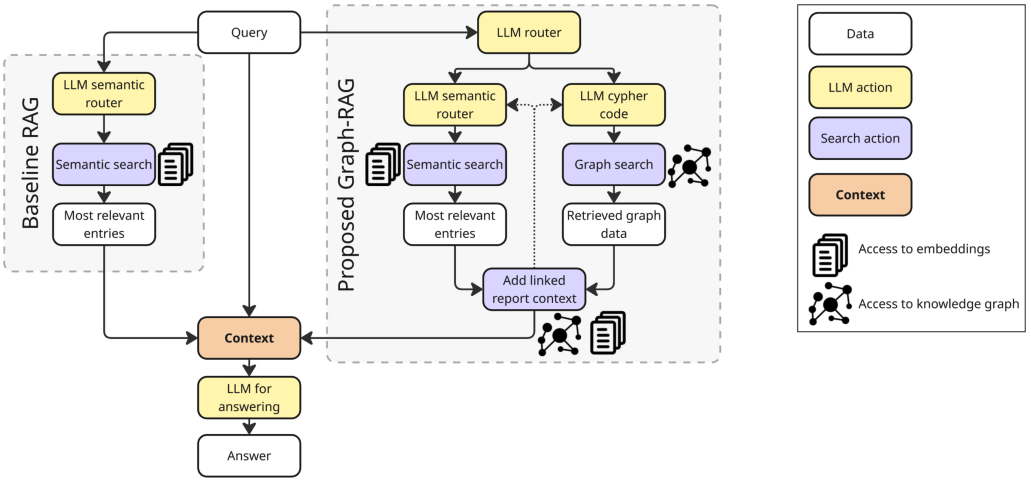


Figure 5: Diagram comparing two information-retrieval workflows for building the context that enables answering natural-language queries; baseline RAG (left), and the proposed Graph-RAG (right). The baseline RAG uses a single semantic search to retrieve relevant nodes by checking their embeddings for semantic similarity. The proposed graph-RAG first sends the query to an LLM router, which decomposes it into sub-steps executed through either semantic or graph search. The retrieved results, together with any linked report context selected by the router, are aggregated into the context.

Category A: Simple filtering. These queries involve retrieving specific entities based on explicit attribute matches, such as material condition or safety status. The baseline RAG performed well on basic listing tasks. The Graph-RAG successfully resolved these logical constraints but occasionally struggled with precise attribute filtering, such as in question A3: ‘‘Find windows in good condition”. There, it produced an invalid filter which returned zero entries, leading to failure in answering this question.

Category B: Counting. This category requires the system to aggregate and count information across the entire dataset to provide quantitative totals. Here, the baseline RAG failed in every query because it relies on local chunk similarity, which cannot capture a global count. In contrast, Graph-RAG demonstrated high competency by using graph operations to provide accurate totals.

Category C: More complex queries. These questions require both filtering and semantic search, and may also include multi-step retrieval. Graph-RAG answered all questions, whereas the baseline failed in half of the questions. This underscores a fundamental limitation of standard RAG: if the information needed is not contained within the same text chunk, the vector search cannot reliably link data from different chunks. Further, question C5, which required a global comparison of window counts across all buildings, was not successfully answered by baseline RAG. This is because this method only sees the top-k chunks it retrieves and has no way of knowing which building has the most windows. Graph-RAG successfully answered the questions by combining semantic search and graph traversal. It also handled semantic variations, such as identifying that ‘‘residential buildings” and ‘‘houses” in C1 and C2 refer to the same category. However, this accuracy came at significantly higher token consumption and cost as the retrieval required more steps and generated more extensive context.

Category D: Queries requiring report context. These queries evaluate the system’s ability to retrieve information from the reports and link it to the corresponding entries. Both methods succeeded in questions D1 and D3, where building-level attributes like construction material or future use could be found in the report text chunks. However, in D2 the baseline RAG failed to link a specific component (glazing panels) to building properties (the heating system) found elsewhere in the report, illustrating the RAG’s inability to link information stored in different chunks. In contrast, Graph-RAG successfully resolved the query by following the graph edges from the identified nodes (glazing panels) to their parent building nodes and finally to their report text chunks where the information about the heating system is stored.

Category E: Answer cannot be found in the dataset. This final category serves as a robustness check to evaluate the systems’ tendency to hallucinate when faced with queries regarding information absent from the dataset, such as market values or carbon-footprint metrics. The baseline RAG demonstrated consistent reliability; it correctly identified the absence of data. In contrast, Graph-RAG failed in E1 by confusing a different numerical property (mass) with the requested market value. Although Graph-RAG successfully provided correct refusals for E2 and E3, it did so at a significantly higher token cost since the system performed a more extensive search before concluding that the data was missing.

Table 2: Comparison of performance across 21 query types for baseline RAG and the proposed Graph-RAG approach. The symbol ✓denotes success and ✗denotes failure. The total amount of tokens used and the total cost for the creation of the full answer are reported for each approach.

Evaluation questions	Baseline RAG			proposed Graph-RAG		
	Tokens	Cost $	Result	Tokens	Cost $	Result
A. Simple filtering						
A1. List five items that are in good condition.	6,475	0.012	✓	9,262	0.021	✓
A2. List five hazardous entries.	8,433	0.024	✓	4,657	0.010	✓
A3. Find windows in good condition.	7,615	0.022	✓	14,379	0.029	✗
B. Counting						
B1. How many buildings are in the dataset?	3,332	0.009	✗	3,872	0.006	✓
B2. Count all the residential buildings in the dataset.	3,795	0.014	✗	3,999	0.007	✓
B3. Count how many entries are in each building in the dataset.	3,733	0.013	✗	8,351	0.026	✓
B4. What percentage of entries are marked as hazardous?	6,327	0.016	✗	5,531	0.013	✓
C. More complex queries						
C1. Find five windows from residential buildings	6,499	0.020	✗	20,720	0.061	✓
C2. Find five windows from houses	6,302	0.019	✗	21,072	0.072	✓
C3. Show exterior cladding options from schools in good condition.	5,627	0.017	✓	23,716	0.062	✓
C4. Find acoustic ceiling panels from office buildings.	6,550	0.018	✗	17,473	0.049	✓
C5. Which building has the most available windows and what are its details?	3,452	0.010	✗	28,802	0.072	✓
C6. We need 30 sq. meters of exterior facade cladding in good condition, which options exist in the dataset?	6,676	0.021	✓	19,263	0.038	✓
C7. Find reusable double glazing windows with more than 2 units available.	6,458	0.019	✓	20,262	0.056	✓
C8. I'm looking for 10 identical toilet separator walls that can be reused.	7,645	0.022	✓	20,217	0.052	✓
D. Queries requiring report context.						
D1. Find a residential building in the dataset whose load-bearing system is made of reinforced concrete.	8,261	0.021	✓	16,322	0.058	✓
D2. I'm looking for available glazing panels. What kind of heating system does the building have that these panels originate from?	8,557	0.022	✗	21,371	0.056	✓
D3. Which building is undergoing transformation into offices?	7,705	0.020	✓	12,761	0.035	✓
E. Answer cannot be found in the dataset.						
E1. What is the market value of the double-glazing windows recorded in the dataset?	8,227	0.025	✓	21,228	0.064	✗
E2. Which reusable material in the dataset has the highest market value?	7,634	0.022	✓	25,592	0.069	✓
E3. What is the carbon footprint of the structural steel parts in the dataset?	7,707	0.023	✓	20,678	0.053	✓

Reasoning plans

Table 3 shows examples of the reasoning plan of sub-steps created by Graph-RAG to gather the necessary context to answer the given questions.

Table 3: Step-by-step reasoning plans followed by Graph-RAG for building the context that can help the LLM answer selected evaluation questions.

ID	Question and reasoning plan
A4:	**Find three entries from buildings constructed before 1970.** 1. *Graph search.* Find buildings constructed before 1970. 2. *Graph search.* Find entries from the buildings found in the previous step.
B4:	**What percentage of entries are marked as hazardous?** 1. *Graph search.* Count all entries. 2. *Graph search.* Count all entries marked as hazardous.
C4:	**Find acoustic ceiling panels from office buildings.** 1. *Semantic search.* Find office buildings. 2. *Graph search.* Find all entries from the buildings found in the previous step. 3. *Semantic search.* Find acoustic ceiling panels from the entries found in the previous step.
C6:	**We need 30 sq. meters of exterior facade cladding in good condition, which options exist in the dataset?** 1. *Semantic search.* Exterior facade cladding entries. 2. *Graph search.* Filter entries found in the previous step to those in good condition with at least 30 square meters available.

Discussion

Overall performance. The evaluation results demonstrate a clear performance-accuracy trade-off between the baseline RAG and the Graph-RAG methods. Across 21 queries, Graph-RAG achieved 90% success rate (19/21 successful queries) compared to 57% (12/21 successful queries) for baseline RAG. Graph-RAG significantly outperformed the baseline RAG on tasks requiring global knowledge of the dataset, multi-step retrieval, and counting, while the baseline RAG proved effective for simple, localized retrieval (Category A) and exhibited higher robustness against hallucination in the absence of data (Category E). However, this improved accuracy came at a 2.34× higher cost, with Graph-RAG averaging $0.0433 per query compared to $0.0185 for baseline RAG. The main strengths and weaknesses of each method are summarized in the following:

- **Relations.** Graph-RAG successfully used graph traversal to find relations between separate nodes and address the inability of baseline RAG to link information that is stored in different text chunks.

- **Global vs. local context.** The baseline RAG is inherently limited by its top-k retrieval mechanism, making it incapable of performing global aggregations

such as counting (Category B). Graph-RAG, conversely, was successful at answering questions that require global context.

- **Efficiency and cost.** The increased accuracy of Graph-RAG comes at a substantial cost. On average, Graph-RAG came at a 2.34× higher cost than the baseline, making standard RAG the more cost-effective solution for simple keyword-based retrieval.

- **Hallucination.** As seen in Category E, the baseline RAG was more conservative and better at identifying knowledge gaps, whereas Graph-RAG occasionally became confused by the retrieved information and failed to identify the lack of the requested data.

- **Flexibility.** Graph-RAG's modular approach, which builds a series of sub-steps, enables adaptive query decomposition, providing flexibility that baseline RAG lacks.

Retrieval methods overlap. Semantic search and graph search have overlapping capabilities. As a result, most queries can be answered through multiple retrieval paths, though with varying degrees of reliability. For instance, identifying a specific material category can be achieved either through vector-based semantic similarity or by querying specific nodes within the graph.

Reliability. When Graph-RAG fails, errors typically originate from the stochastic nature of the query-decomposition phase, inaccuracies in automated Cypher generation, or system implementation issues. For instance, Graph-RAG was unable to answer question A3 because the graph-search query for windows in good condition produced the invalid filter `entry.condition = "Good condition"`, whereas the correct encoding is `entry.condition = "good"`. This illustrates that, while the approach provides a robust foundation for PDA information retrieval, its practical performance remains dependent on the accuracy with which the LLM translates natural-language inputs into structured, executable operations.

Evaluation limitations. The question-based evaluation cannot capture nuances such as incomplete answers, varying response quality, or incorrect confidence levels in generated outputs. Additionally, the manual ground-truth verification introduces potential subjectivity for queries with multiple valid interpretations. Furthermore, the evaluation does not assess user-experience factors such as query formulation difficulty or result interpretability, which are critical considerations given that the intended users are non-experts. Finally, the study focuses on a single dataset of French PDAs with specific structural characteristics. Applicability to PDAs from other regulatory contexts or formats remains to be validated.

Outlook

Several directions could advance this research beyond the current prototype. First, enriching the knowledge graph with additional semantic relationships, such as explicit MENTIONS edges linking Report nodes to referenced Entry or Building nodes and implicit edges for relations such as spatial proximity, material compatibility, and dimensional similarity, would facilitate multi-step queries. Second, extending the system to support continuous conversations with memory would enable iterative query refinement (for example, ‘‘show me more details about that building,’’ ”now filter those to only non-hazardous items”) without re-specifying context. Third, incorporating multimodal data such as photos from PDA site visits would provide richer context; photographic documentation that many PDAs already contain could be leveraged but remains unexploited in text-only frameworks. Fourth, integrating external knowledge sources such as material databases, reuse marketplaces, or embodied carbon calculators would enable queries combining PDA inventory with market availability, pricing, or environmental impact assessments. Fifth, implementing explainability mechanisms that visualize which graph paths, semantic similarities, or retrieval steps contributed to answers would build user trust and enable systematic debugging of retrieval errors. Finally, human-centered deployment studies in real design workflows are needed to evaluate the system as a decision-support tool, including its effects on user trust, understanding, perceived agency, decision quality, decision-making time, and stakeholder engagement.

Conclusion

This paper presented a Graph-RAG framework to enable natural-language querying of PDA data. The approach integrates structured tabular inventories with unstructured textual reports in a knowledge graph to enable non-expert users to retrieve information using natural-language queries without requiring schema knowledge or technical skills. In this way, it improves the accessibility of PDA data and provides a foundation for more human-centered interaction with reuse information.

The evaluation across 21 queries demonstrated that Graph-RAG achieved a higher success rate than baseline RAG, with the performance advantage most pronounced in queries requiring counting, global aggregation, and multi-step reasoning. Both methods performed equally well on purely semantic queries, confirming that semantic similarity-based retrieval remains effective for concept matching when graph structure is not required. However, Graph RAG's higher accuracy comes at a higher cost, as it requires significantly more tokens to answer queries. The study demonstrates that knowledge-graph representations of PDA data enable more flexible and context-aware information retrieval compared to semantic search over unstructured text chunks. The framework's design provides a foundation for applying natural-language interfaces to heterogeneous built-environment datasets where structured

and unstructured information coexist.

Acknowledgments

Research reported in this publication was supported by Bouygues construction R&D and Innovation Department. The authors also wish to thank Dr. Roxanne Goldberg for her valuable feedback and insightful advice. The authors further acknowledge Thibaut Menny and Agyre for facilitating access to PDA data.

References

Akanbi, L. A., Oyedele, A. O., Oyedele, L. O., and Salami, R. O. (2020). Deep learning model for Demolition Waste Prediction in a circular economy. Journal of Cleaner Production, 274:122843.

Charette, R. and Marshall, H. (1999). UNIFORMAT II Elemental Classification for Building Specifications, Cost Estimating and Cost Analysis. NIST Interagency/Internal Report (NISTIR) 6389, National Institute of Standards and Technology, Gaithersburg, MD. Accessed June 6, 2025.

Davis, A., Audí, N. M., and Hall, D. M. (2025). A review of circular industrialised construction for sustainable and affordable housing: Towards a process-driven framework. Sustainable Cities and Society, 133:106837.

Godina, M., Gowler, P., Rose, C. M., Wiegand, E., Mills, H. F., Koronaki, A., Ramage, M. H., and Shah, D. U. (2025). Strategies for salvaging and repurposing timber elements from existing buildings in the UK. Journal of Cleaner Production, 489:144629.

Hsiu Tung, Y., Rezvani, S., Asgharzadeh, F., Neumann, M., and Hartmann, T. (2025). IFC Whisperer: Querying Building Information Models with Large Language Models and IFC-based Knowledge Graphs. Journal of Physics: Conference Series, 3140(16):162007. Publisher: IOP Publishing.

Li, M. and Wang, Z. (2026). BuildingGPT: Query building semantic data using large language models and vector-graph retrieval-augmented generation. Building and Environment, 287:113855.

Luciano, A., Cutaia, L., Altamura, P., and Penalvo, E. (2022). Critical issues hindering a widespread construction and demolition waste (CDW) recycling practice in EU countries and actions to undertake: The stakeholder's perspective. Sustainable Chemistry and Pharmacy, 29:100745.

Mitropoulou, I., Forth, K., Vangelova, S., Menny, T., and De Wolf, C. (2026). Automated structuring, classification, and thermal properties enrichment of pre-demolition audits using llms. Artificial Intelligence for a Sustainable Built Environment. Under review.

Neo4j, Inc. (2025a). Neo4j graph data science python client. `https://neo4j.com/docs/graph-data-science-client/current/`. Accessed: 2025-12-11; Official documentation of the `graphdatascience` Python package for Neo4j Graph Data Science.

Neo4j, Inc. (2025b). Neo4j graph database platform. `https://neo4j.com/`. Accessed: 2025-12-11; official website of the Neo4j graph database platform, a native graph database for managing and querying interconnected data.

Oluleye, B. I., Chan, D. W. M., Olawumi, T. O., and Saka, A. B. (2023). Assessment of symmetries and asymmetries on barriers to circular economy adoption in the construction industry towards zero waste: A survey of international experts. Building and Environment, 228:109885.

Pan, Y., Wang, M., Lu, L., Lamsal, R., Pärn, E., Zlatanova, S., and Brilakis, I. (2026). Llm-enabled multi-agent framework for natural language interaction with graph-based digital twins. Automation in Construction, 183:106791.

Procko, T. T. and Ochoa, O. (2024). Graph Retrieval-Augmented Generation for Large Language Models: A Survey. In 2024 Conference on AI, Science, Engineering, and Technology (AIxSET), pages 166–169, Laguna Hills, CA, USA. IEEE.

Sequeda, J., Allemang, D., and Jacob, B. (2024). A Benchmark to Understand the Role of Knowledge Graphs on Large Language Model's Accuracy for Question Answering on Enterprise SQL Databases. In Proceedings of the 7th Joint Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA), pages 1–12, Santiago AA Chile. ACM.

Shang, Y., Ke, Z., Lin, P., Ren, Q., Zhang, W., Wang, X., Li, X., Gong, F., Wang, S., Wang, B., Xu, Z., Sun, M., and Tan, S. (2025). Empowering knowledge graphs with hybrid retrieval-augmented generation for the intelligent mix scheme of mass concrete. Case Studies in Construction Materials, 23:e04979.

Wu, P.-Y., El-Assady, M., and De Wolf, C. (2026). A unified intelligence-augmented framework for building audits via physical and virtual inspection. Expert Systems with Applications, 302:130514.

Xu, Z., Cruz, M. J., Guevara, M., Wang, T., Deshpande, M., Wang, X., and Li, Z. (2024). Retrieval-Augmented Generation with Knowledge Graphs for Customer Service Question Answering. In Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 2905–2909, Washington DC USA. ACM.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

COLLABORATIVE FACADE DESIGN WITH LLM-BASED MULTI-AGENT SYSTEMS USING RECLAIMED BUILDING COMPONENTS

Pritha Jacob[1], Ioanna Mitropoulou [1], Pei-Yu Wu[2] and Catherine De Wolf[1]
[1]ETH Zurich, Zurich, Switzerland
[2]TU Delft, Delft, Netherlands

Abstract

The potential for material reuse is restricted by the lack of methods for integrating reclaimed materials into facade design workflows. This paper presents a multi-agent pipeline using LLMs to support integrating reclaimed materials into facade design. Tested through a web application in two design modes: a *collaborative* mode, where human and LLM collaborate, and an *independent* mode, without human guidance; the users reported the collaborative mode to be more useful. The proposed AI-human collaboration framework improved the tasks of searching for materials, evaluating options, and visualizing design iterations, highlighting the value of AI-driven design frameworks in encouraging reuse in architecture.

Introduction

Designing with reclaimed components remains largely a task of trial and error, with designers manually assembling designs from available inventories, which makes them more hesitant to implement reclaimed materials in their design (Kawa et al., 2022). Much of the research to date aimed at facilitating circular design frames the integration of reclaimed materials as an optimization problem. As a result, researchers are pursuing fully automated deterministic methods to generate feasible design solutions (Moussavi et al., 2022). However, facade design, in particular, cannot be reduced to optimization alone: it involves aesthetic judgment, contextual reasoning, and qualitative considerations that are difficult to formalize computationally. When the human designer is incorporated into the loop, these qualitative dimensions can be addressed through interpretation, judgment, and situated decision-making, leveraging human expertise alongside computational reasoning. The present research aims to develop such a collaborative design workflow that integrates human expertise with generative AI methods, specifically large language models (LLMs), to support design with reclaimed materials. In doing so, it is guided by two key Human-Data Interaction (HDI) principles: *legibility*, meaning that the system's reasoning, suggestions, and use of data remain understandable to the designer, and *agency*, meaning that the designer retains informed control over decisions rather than being reduced to validating automated outputs (Durango et al., 2024).

We propose a workflow based on an AI multi-agent system, composed of multiple interacting AI agents capable of adaptation, learning, and reasoning. As such, this research advances the use of AI agents in construction as HDI mechanisms. It aims to support designers in real-world practice within the context of circular construction by facilitating material selection, application on the facade, and visualization of the result. This enables more iterations during early design stages.

Background

The background for this research was done using a scoping review to understand AI agents, their capabilities, types, and application in various fields. An AI agent is a system or program that is capable of autonomously performing tasks on behalf of a user or another system by designing its workflow and utilizing available tools (Gutowska, 2025). Core components of AI agents include capabilities for planning and task decomposition, memory, reasoning, decision-making, action execution, and tool utilization. AI agents can be classified into different types according to their level of autonomy and intelligence, decision-making processes, and interaction with the environment to achieve desired outcomes (Stryker, 2025). They are also distinguished by their reasoning methods, such as using conditional statements, ReAct (Reason + Act), ReWOO (Reasoning Without Observation), or self-reflection (Caballar and Stryker, 2025). In this study, we use ReAct agents that integrate LLMs, tools, and prompt design, and also take feedback from humans to achieve adaptive intelligence.

AI agents are increasingly used in business, personal, and specialized domains (e.g., healthcare, finance, research) to improve productivity, support decision-making, and enable innovation (Krishnan, 2025). A notable example is the Gemini 2.0–based AI co-scientist, which helps researchers generate knowledge and hypotheses aligned with their goals (Gottweis et al., 2025). In construction, autonomous, distributed multi-agent systems integrated into design software utilize stochastic algorithms (Cousin et al., 2023), performance metrics, and environmental data to collaboratively generate, evaluate, and optimize adaptive, performance-driven facade geometries (Pantazis and Gerber, 2018; Gerber et al., 2017). Today, construction-related research also demonstrates the use of AI applications centered on HDI principles (Pal et al., 2025). While these approaches do not explicitly rely on Generative AI (GenAI) techniques, recent studies highlight the signifi-

cant potential of GenAI methods such as LLMs in architectural design. Tools such as ChatGPT and Bard are transforming architectural workflows by supporting ideation, optimization, and collaboration (Rane et al., 2023). Research also shows the examples of AI-enhanced workflows for façade design. Hoeng et al. (2025) integrate computer vision, large language models, parametric modeling, and reinforcement learning with human expertise to optimize design outcomes and reduce planning time. Cheung et al. (2025) propose a conversational AI-enhanced design framework, featuring a multimodal agentic AI bot integrated with WeChat for design ideation, optimization, evaluation, and style testing. Building on this line of work, the present study adopts a conversational chatbot paradigm to develop an LLM-based multi-agent workflow that supports iterative concept development in façade design.

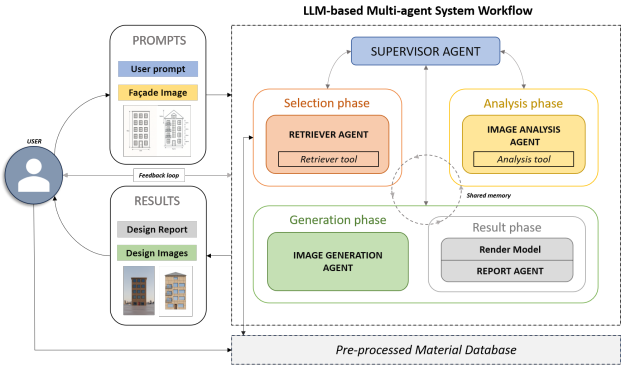


Figure 1: Overview of the workflow detailing the inputs, outputs, phases, and agents.

Methodology

Conceptual Model

We developed our proposed LLM-based multi-agent system using the AI agent framework LangGraph (LangChain, 2024). Our system takes as inputs, first, *a prompt*, which is a text prompt with an image of the design (sketch or render), and, second, *a database* of reclaimed materials. This database is pre-processed and converted into JSON files detailing the dimensions, quantity, material type, and other specifics of reclaimed components.
Our workflow, described in Fig. 1, includes:

Prompts: The prompts needed as inputs are text-based user prompts to assign tasks to the agents and a facade image to convey the user's design concepts.

Results: Aimed at designers and stakeholders, the results include a visual representation of the designs through generated renders (images) and compiled reports.

Specialized phases: The workflow is divided into phases, with each phase having its own specialized agents.

- *Selection phase*: This phase involves an agent that selects materials from the database based on user prompts.
- *Analysis phase*: This phase aims at analyzing the input facade sketch or image to extract its key features and dimensions.
- *Generation phase*: This phase integrates the selected materials into the façade to create new façade designs and reports according to the user's preference. It also generates renderings of the proposed designs.

Agent Types

Based on the conceptual model, the various agents that work in the respective phases of the workflow are as follows:

Supervisor agent: This agent is an LLM-based router that directs tasks to different agents in the workflow based on user prompts.

Retriever agent: As part of the Selection Phase of the workflow, this agent is a pre-built LLM-based LangGraph ReAct agent. It uses a retrieval augmented generation tool to semantically search the material database to retrieve elements that match the user prompts.

Image analysis agent: Part of the Analysis Phase of the workflow, this agent is also an LLM-based ReAct agent that analyses the input facade image to extract details such as the number of floors, windows, etc.

Image generation agent: This agent, part of the Generation Phase, creates N Matplotlib codes that describe design variations, using the materials from the retriever agent and facade analysis from the Image analysis agent. Its prompt includes example code templates, details from the other agents, and user instructions to generate multiple façade designs.

Report agent: Part of the result phase, a subset of the Generation Phase, this LLM-based agent compiles the results of all other agents into a structured technical PDF report.

Agent Evaluation

To ensure that the three main phases (Selection, Analysis, and Generation) of the workflow function as intended, evaluations were integrated into the workflow. These evaluations are essential to validate the agent's working and verify the results produced by the agents. We implemented the following evaluation strategies:

LLM-as-a-judge: With their ability to process diverse data formats and generate scalable, adaptable assessments, LLMs offer a strong alternative to traditional expert-based evaluations. The LLM-as-a-judge paradigm emphasizes that LLMs are auto-regressive generative models that produce follow-up content from prior context and derive evaluations from this generated output (Gu et al., 2025).

In the context of this workflow, this evaluation strategy involves the use of an LLM to validate the agent's performance by checking, first, that its answers are rooted in data generated by its tools rather than hallucinations, and, second, that it responds to the requirements of the prompt.

Human-in-the-loop (HITL): HITL systems enable integrating feedback from the nuanced contextual understanding of human users (Natarajan et al., 2024). In this context, HITL is the main way in which collaboration between humans and AI agents is introduced. Its goal is to ensure that the agents' outputs align with human expectations. Furthermore, it enforces explicit user validation of the generated content prior to its finalization. In this workflow, the HITL mechanism is implemented through the LangGraph "Interrupt" feature, which pauses LLM reasoning until feedback is given by the user.

Unit Testing

The development of the agent-based workflow started with testing each phase (unit) separately. The unit testing involved determining the most suitable LLMs for each agent and the best evaluation strategy for each of the three main phases; *Selection, Analysis, and Generation*.

Selection phase with retriever agent: The retriever agent often hallucinated in its final responses despite correct outputs of the retrieval-augmented generation-tool. Adding LLM-as-a-judge to the pipeline reduced these hallucinations by about 80%. Following this, a HITL evaluation step was introduced. Requiring explicit user approval of each final response enforced relevance and precision. Finally, both strategies, LLM-as-a-judge and HITL,were combined, as shown in Fig. 2. Each addresses complementary aspects of reliability and quality control.

The final agent framework operates as follows. The retriever first generates a response to the user prompt, which is checked by the LLM-as-a-judge for hallucinations. If the response is accepted, the HITL node requests user approval or triggers regeneration. If the response is rejected, the retriever produces a revised response. Feedback from the HITL node is incorporated as a new prompt, and the resulting output is again assessed by the LLM-as-a-judge and HITL nodes. This iterative workflow was tested using `Google's Gemini 2.5` and `OpenAI's GPT-4.0 mini`.

Analysis phase with image analysis agent: An analogous procedure was implemented for the image analysis agent, which likewise interfaced with an image-analysis tool, as shown in Fig. 2. This configuration minimized tool-agent inconsistency and mitigated hallucinations while simultaneously enabling human reviewers to act as the ultimate verification layer. The LLMs tested included `Gemini 2.5` and `GPT 4.0 mini`.

Generation phase with image generation agent: Since this agent's task involved comparatively complex reasoning, yielded inherently subjective outputs, and did not require the use of external tools, the inclusion of an LLM-as-a-judge component in the evaluation pipeline was deemed unnecessary. Instead, only the HITL evaluation was employed. The HITL process ensured that the design artifacts produced by the agent could be iteratively improved through user feedback, including targeted suggestions and corrective interventions. In this case, the tested LLMs include `Gemini 2.5 Pro, GPT 4.0 mini, Qwen 2.5 Coder 0.5B` and `Gemini 2.5 Pro, GPT 4.0 mini, Qwen 2.5 Coder 0.5B, and Qwen 2.5 Coder 7B`.

System Architecture and Testing

In this context, the term *system architecture* denotes both the way the various agents are connected to each other and how they interact with the user. The overall architecture (Fig.2) exhibits a graph-like structure in which the various agents and their associated tools, additional functions (routing and execution), and the evaluation strategies are encapsulated within *nodes*. These nodes comprise the agents mentioned previously as well as a non-agentic node called the "stylization node," as seen in Fig. 2. This node is part of the *Generation phase* and consists of a vision language model (VLM) that takes as input the plot images and creates photorealistic façade renderings. Two VLMs, `Google's Imagen 3.0` (capability 001) and `Google's Gemini 2.5 flash` (image-preview), were evaluated for this task.

The intermediate and final outputs from the different nodes are stored in a shared data structure, termed *State*, and inter-agent communication (i.e., state updates) is orchestrated via *Edges* as defined by LangGraph. These edges can be either conditional or fixed. Fixed edges (displayed with continuous arrows) connect a node to a single subsequent node, whereas conditional edges (displayed with dotted arrows) may route execution to multiple candidate nodes. With conditional edges, the specific node is selected according to the evaluation of the condition associated with that edge. For this workflow, two main system architectures were explored:

Collaborative workflow: As illustrated in Fig.2a, the graph structure of the system is organized around a supervisor agent, which is connected via conditional edges to four specialized agents: retrieval, image analysis, image generation, and report.

Operationally, when a user submits a query, the START node is triggered, which in turn activates the supervisor node. The supervisor node is responsible for determining the appropriate routing of the request within the graph. For example, for material-related queries, such as "Could you select some wooden flooring materials?", the supervisor node routes the request to the retriever node, which encapsulates the retriever agent and its associated tool. The retriever node then produces an output that updates the state. This output is subsequently evaluated

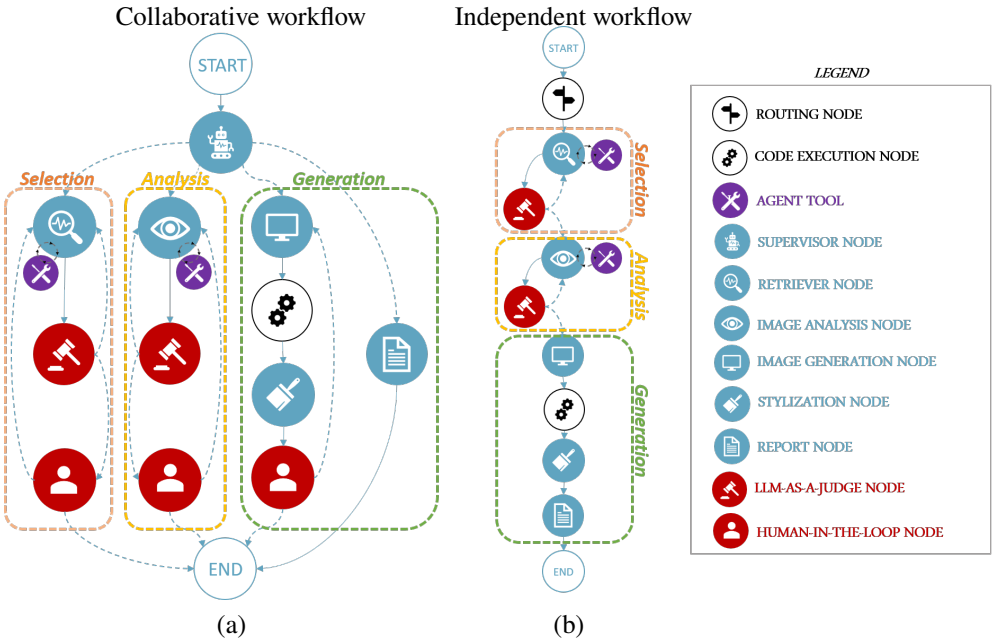


Figure 2: System architecture: (a) collaborative workflow, (b) independent workflow. The Selection, Analysis, and Generation phases each have their dedicated agent nodes (in blue), tools (in purple), assisting code-nodes (uncolored), and evaluation nodes (in red). The dotted blue arrows represent conditional edges, which may or may not be executed depending on the state of the system, while the solid blue arrows represent fixed edges that are always executed. The uncolored nodes (code execution node and routing node) consist of deterministic functions that do not include an LLM.

for potential hallucinations and must be validated by an LLM-as-a-judge module. Upon approval, the process advances to the HITL node, where the user may either approve the generated results, thereby terminating the process at the END node, or provide additional suggestions, which causes the entire workflow to recur. An analogous procedure is followed in the image analysis node, with the key difference that the user input in this case consists of both a text-based query and an uploaded image of the facade. In contrast, for the image generation agent, once the relevant code is produced by the image generation node, this code is executed in the execution node, resulting in generating the facade plots. These plots are then passed to the stylization node, where they are transformed into photorealistic renderings. The pipeline subsequently proceeds to the HITL node, where user approval or suggestion again determines whether the graph terminates or recurs. After all primary agents have been invoked and the user has expressed satisfaction with all intermediate results, the user may finally trigger the report node, which compiles and generates the corresponding PDF report.

Independent workflow: The system architecture depicted in Fig.2b represents a unidirectional pipeline in which each processing stage is executed sequentially. It focuses on minimizing human interaction by only including LLM-as-a-judge evaluation and no HITL. In this configuration, the routing function is implemented as a dedicated mechanism that decomposes the user query to update specialized keys in the state. These keys are subsequently used by the corresponding downstream nodes in the graph to perform their designated operations. Operationally, when a user query, such as retriever agent: "Could you select some wooden flooring materials?, analysis agent: analyze this image, generation agent: generate a facade design.", is submitted to the system, the routing function decomposes the input into its constituent segments. Subsequently, the graph-based workflow is initiated with the retriever agent as the first node in the sequence, as illustrated in Fig. 2. Upon completion of its task, the retriever agent's output is evaluated by the LLM-as-a-judge module. Once accepted, the graph updates the results in the state. The graph then proceeds sequentially through each subsequent node, with their respective tasks executed in turn, analogous to the architecture of the collaborative workflow. This process ultimately culminates in the generation of a consolidated PDF report.

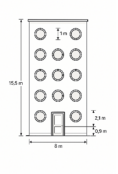
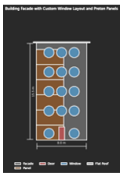
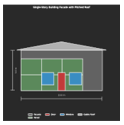
These two architectures were specifically selected due to their modularity (facilitating the straightforward addition and removal of nodes), robustness (reducing the number of bottlenecks and potential points of failure), and their flexibility with respect to integrating agent evaluation mechanisms. Their suitability was subsequently examined through user testing, which was carried out to investigate user perceptions of the two fully implemented system architectures.

User Testing

The two system architectures were tested to gauge user perspectives. To gain user insight for this research, a web ap-

[3] This prototype can be shared for testing purposes under reasonable request to the author.

Table 1: Workflow showing retrieval, analysis, and generation prompts with base images, coded plots, and renders.

	Inputs			Outputs		
Trial	**Retriever agent prompt**	**Analysis agent prompt**	**Generation agent prompt**	**Base Image**	**Coded Plot**	**Render**
1	Can you suggest some wall cladding materials that have the highest CO_2 saved and are preferably wood?	What are the details of the façade in the given image? (*upload base image*)	Use the material details for the panelling and façade details for the overall façade. Using this information, generate the façade image.			
2	Can you suggest a lattice material that can allow ventilation within the interior spaces, with a height of 2.53m?	Describe the façade in the image? (*upload base image*)	Use the material details for the panelling and façade details for the overall façade. Using this information, generate the façade image.			

plication, *reFAC*[3], was created as a proof-of-concept for the research workflow. *reFAC* and its user interface (UI) were designed through *Streamlit*, a Python open-source application framework well-suited for hosting applications that involve AI agents and facilitate meaningful human interaction. This application has two pages: one that gives access to the collaborative workflow and one that gives access to the independent workflow, as shown in Fig. 3.

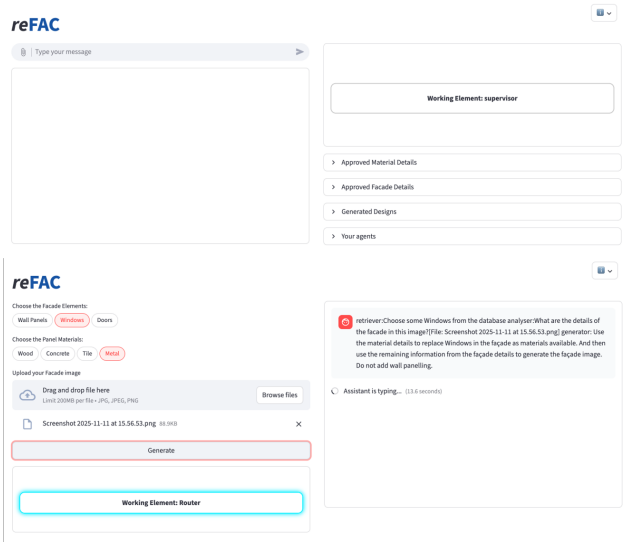


Figure 3: Pages of the UI corresponding to the different system architectures. Top: Page that gives access to the collaborative workflow. Bottom: Page that gives access to the independent workflow.

Results & Discussion

LLM models: The unit and integrated tests indicate that `Gemini` is the most suitable LLM for the agents' use cases, while for vision–language tasks the best-performing model was `Google's Gemini 2.5 flash` (image-preview). These models were used for the user testing.

Comparison of workflows: The collaborative workflow differs from the independent workflow primarily through its integration of the HITL node, which enables individual agent results to be verified and refined. In contrast, the independent workflow lacks such checks and risks propagating errors or requiring complete restarts. While the independent workflow offers a highly automated and rapid solution for design suggestions when the user has no specifications, the collaborative approach enables an iterative process where specific agents can be re-invoked with new feedback from the user without rerunning the entire pipeline. This modularity makes the collaborative system more token-efficient and cost-effective for targeted modifications. However, the flexibility of a supervisor-led architecture can occasionally lead to routing errors if user queries are ambiguous. For example, instead of the systematic querying carried out, if the user directly asks: "I would like to design a building with wooden slabs from the database," the supervisor may incorrectly route the request to the image-generation rather than the retriever node, interpreting it as a primarily generative task instead of first recognizing the need to retrieve relevant materials from the database. This limitation is largely avoided by the independent workflow's use of predefined, standardized prompt categories. The choice between these architectures involves a trade-off between the speed of autonomous execution and the granular control and resource optimization provided by a collaborative framework.

A trait common to both workflows is the low reproducibility of the results produced by them due to the stochastic nature of the LLMs, meaning that identical inputs will generally not produce the same outputs. This variability can be beneficial during conceptual design, as it supports rapid exploration of alternative ideas. However, it also

poses challenges for comparison and validation. To integrate such tools into systematic design processes, where iteration on the same design is needed, additional control strategies should be developed.

User testing: Both workflows, developed as a web application, were tested by five researchers in the field of architecture and engineering. User testing revealed that 60% of the users believe this application improves task completion, with 20% reporting that the generation is good but may require cross-verification of the results with the material database. 60% of the users report it took a lot more effort with respect to clarifying and repeating requests to the agents than expected, as shown in Fig. 4. This perception primarily arose from the heterogeneity of requirements across agents. The necessity for repeated clarification with these agents prompted the users to want an autonomous assistant that proactively orchestrates and manages a process, as opposed to one that executes externally specified instructions. This aligns with literature findings (Cheung et al., 2025) that emphasize the need for a conversational ideation process over directive interactions. To increase autonomy within the workflow, further improvements are needed, including refinements to agent-specific prompts and inter-agent coordination, in order to strengthen supervisor–agent communication and enhance the agents' ability to operate more independently. However, users did appreciate having the final say, indicating the value of collaboration in design. Future developments should aim for a balanced user-agent control ratio to foster co-designing. Fig. 4 also shows that 40% found the application to be overwhelming. These users felt like they were speaking to multiple agents as opposed to a system performing a single task. The workflow was intended to make users feel as if they are conversing with a unified Multi-agent System, with the supervisor agents assigning tasks to its worker agents. However, because each agent requires specific inputs from the users, the entire process can occasionally be experienced as question-answering rather than conversational.



(a) (b) (c)

Figure 4: Summary of the prototypes' user experience, with (a) human-AI collaboration, (b) expectation on effort, and (c) agent-human collaboration reception.

Fig. 5 shows the performance of the three major agents based on four main categories: independence, creativity, accuracy, and reliability. Among the agents, the retriever agent ranks the highest in accuracy, reliability, and independence. The image analysis agent ranks the lowest in all categories, while the image generation agent ranks the highest in creativity and independence.

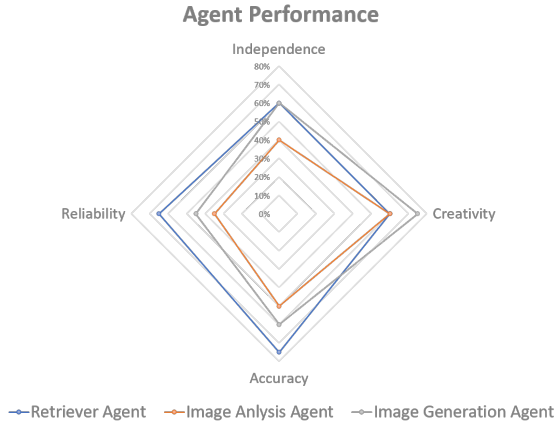
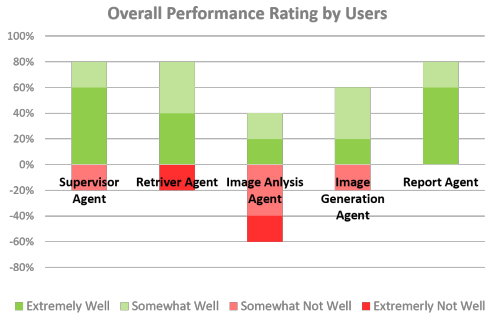
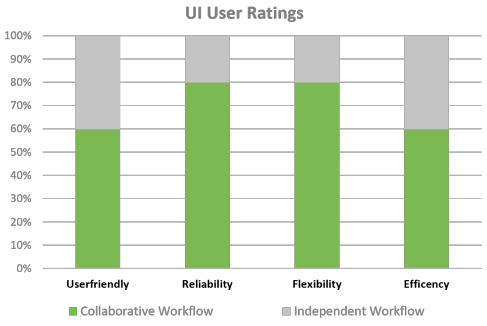


Figure 5: Comparison of Agent performance on various categories.

Fig. 6 shows that 80% of the users consider the collaborative interface to be reliable and flexible, with 20% preferring the independent interface. It is clear from the user ratings of the specific agents (see Fig. 6) that most of them perform quite well. Overall, the results show that the best performance is seen in the report and supervisor agent, while the least performative agent is the image analysis agent.



(a)



(b)

Figure 6: Summary of user rating for agent performance and UI experience, with (a) overall agent performance rating by users, (b) UI comparison rating by users.

Limitations

Geometric fidelity. A current limitation of the system is the absence of hard geometric constraints that enforce strictly following the dimensions of the façade and the components used in its design. As a result, generated façades may appear to respect geometric proportions of the facade and the retrieved materials, but in reality they contain dimensional inaccuracies. Introducing explicit geometric constraints, rather than relying solely on the judgment of the LLM, could improve geometric fidelity.

Design flexibility. The system can exhibit reduced design flexibility in certain cases. Due to the prompts that guide code generation for façade synthesis, some outputs become oversimplified, resulting in generic or standardized façade typologies. While for some outputs the workflow is able to adapt well to the input (e.g., Fig. 7 top), other cases (e.g., Fig. 7, bottom) show excessive simplification, leading to the loss of salient architectural features.

Supervisor agent. This agent is more of an LLM router than a brain. If a query does not fit neatly in a specific phase, then the supervisor has difficulty assigning it to an agent. This limitation could be mitigated by giving further context, for example, the entire conversation history. The problem we observed when giving additional context to the supervisor is that the agent tries to answer the query by itself without routing to the other agents and using the available tools, despite being prompted to do so. Boundaries and guardrails should therefore be added to manage the additional context while fulfilling the supervisor's role.

Sensitivity to varied prompting. User testing indicates that the system's performance varies across different types of prompts. For instance, the function of the retriever agent is to choose a material from the database based on a user prompt. However, if the user prompt asks for a different type of retrieval, such as the retrieval of multiple materials, then the agent may fail to give a result. This is because it may or may not choose more than one material from the inventory based on requests relating to material types and properties, such as related emissions, type, color, or function, like windows or doors. A similar lack of flexibility was observed with the image analysis agent, which works best with elevation rather than render images, which should include dimensions. There were instances, as shown in Fig. 7, where the user input perspective images or sketches that did not have legible dimensions, leading to a poor analysis and evident optical character recognition errors by the agent. These issues can be mitigated by extending the system prompts for those agents to make them more flexible when dealing with more variable inputs, or by splitting their functionality down to more granular tasks that are carried out by separate agents.

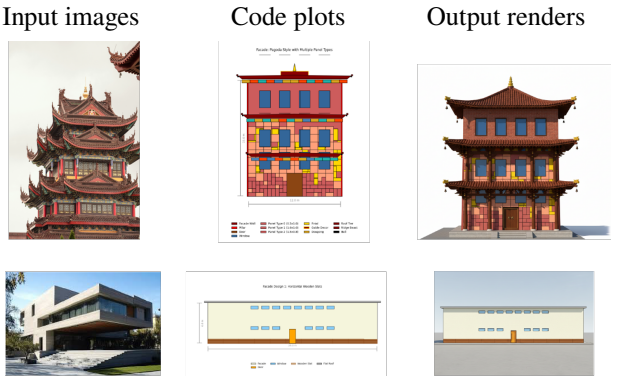


Figure 7: Comparison of final results: The top set shows a successful facade generation that follows the input image, while the bottom set represents an unsuccessful facade generation due to the unclear input image, where the details of the resulting facade have been overly simplified.

Outlook

This research demonstrates the potential of collaborative AI-based workflows. The proposed framework and its individual agents, such as data retrieval, image analysis, and generative design, can be adapted as versatile ideation tools for architectural offices exploring design alternatives, including variations in proportion, rhythm, and fenestration patterns. With further development, the system could also generate CAD- or BIM-compatible geometry, enabling a smoother transition from conceptual design to detailed documentation. Achieving this would require extending the current approach with parametric definitions of building elements, constraint-based geometric representations, and validation mechanisms to ensure compliance with CAD/BIM schemas. The current framework addresses conceptual design and ideation rather than detailed model generation. Such human-AI collaborative workflows have the potential to become integral to architectural design processes that incorporate reclaimed materials. They might support tasks such as efficient inventory search, rapid generation of design variants, and the structured arrangement of reused components on façades, while maintaining key HDI principles such as designer agency and the legibility of AI-supported decision-making.

Conclusion

This paper presented a collaborative LLM-based multi-agent system workflow for façade design with reclaimed materials, combining human expertise with generative AI tools. The proposed framework integrates LLM-powered agents and explores complementary validation strategies, including LLM-as-a-judge and human-in-the-loop approaches, to support accuracy, transparency, and effective collaboration. In doing so, it emphasizes HDI principles, especially legibility, by making the system's outputs and reasoning easier to understand, and agency, by keeping the designer actively involved in evaluating and guiding the design process. Through a comparative

user evaluation, the collaborative workflow was consistently perceived as more useful than a fully independent workflow, highlighting the importance of keeping designers actively involved in the design process. The proof-of-concept system, *reFAC*, demonstrates the potential of AI agents to operate with a degree of autonomy while remaining responsive to human input. It was positively received by users. Overall, this work establishes a modular and extensible foundation for future research on AI-assisted architectural design, particularly in the context of LLM applications, circular construction, and design with reclaimed materials.

Acknowledgments

This paper presents the results of the author's Master's thesis conducted within the Integrated Building Systems program at ETH Zurich. We gratefully thank the five participants from ETH Zurich for their valuable feedback in the user study, as well as Dr. Roxanne Goldberg for her insightful guidance and advice. This research was supported by the Bouygues Construction R&D and Innovation Department.

References

Caballar, R. and Stryker, C. (2025). What is agentic reasoning? https://www.ibm.com/think/topics/agentic-reasoning#498277083. IBM.

Cheung, L. H., Wang, L., and Lei, D. (2025). Conversational, agentic ai-enhanced architectural design process: Three approaches to multimodal ai-enhanced early-stage performative design exploration. Architectural Intelligence, 4(1).

Cousin, T., Marshall, D., Pearl, N., Alkhayat, L., and Mueller, C. (2023). Integrating irregular inventories: Accessible technologies to design and build with non-standard materials in architecture. In Journal of Physics: Conference Series, volume 2600. Institute of Physics.

Durango, I., Gallud, J. A., and Penichet, V. M. R. (2024). Human-Data Interaction Framework: A comprehensive model for a future driven by data and humans. arXiv (Cornell University).

Gerber, D. J., Pantazis, E., and Wang, A. (2017). A multi-agent approach for performance based architecture: Design exploring geometry, user, and environmental agencies in façades. Automation in Construction, 76:45–58.

Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Palepu, A., Sirkovic, P., Myaskovsky, A., Weissenberger, F., Rong, K., Tanno, R., Saab, K., Popovici, D., Blum, J., Zhang, F., Chou, K., Hassidim, A., Gokturk, B., Vahdat, A., Kohli, P., Matias, Y., Carroll, A., Kulkarni, K., Tomasev, N., Guan, Y., Dhillon, V., Vaishnav, E. D., Lee, B., Costa, T. R. D., Penadés, J. R., Peltz, G., Xu, Y., Pawlosky, A., Karthikesalingam, A., and Natarajan, V. (2025). Towards an ai co-scientist.

Gu, J., Jiang, X., Shi, Z., Tan, H., Zhai, X., Xu, C., Li, W., Shen, Y., Ma, S., Liu, H., Wang, S., Zhang, K., Wang, Y., Gao, W., Ni, L., and Guo, J. (2025). A survey on llm-as-a-judge.

Gutowska, A. (2025). What is a multi-agent system? https://www.ibm.com/think/topics/multiagent-system#276684308. IBM.

Hoeng, S. K. H. S. K., Wiederer, J. W. J., Eder, F. E. F., Obergriesser, M. O. M., and Linner, T. L. T. (2025). Towards AI-enhanced Facade Planning: Integrating Human Expertise with Machine Learning-driven Parametric Modeling. Computing in construction, 6.

Kawa, G., Galle, W., and Temmerman, N. D. (2022). Developing a 'design with reuse' guide. case study analysis from a design procedural perspective for the integration of reclaimed wood in façade architecture. In IOP Conference Series: Earth and Environmental Science, volume 1122. Institute of Physics.

Krishnan, N. (2025). Ai agents: Evolution, architecture, and real-world applications.

LangChain (2024). Langgraph overview. https://docs.langchain.com/oss/python/langgraph/overview. LangGraph documentation.

Moussavi, S. M., Svatoš-Ražnjević, H., Körner, A., Tahouni, Y., Menges, A., and Knippers, J. (2022). Design based on availability: Generative design and robotic fabrication workflow for non-standardized sheet metal with variable properties. International Journal of Space Structures, 37:119–134.

Natarajan, S., Mathur, S., Sidheekh, S., Stammer, W., and Kersting, K. (2024). Human-in-the-loop or ai-in-the-loop? automate or collaborate?

Pal, A. P. A., Murguia, D. M. D., and Middleton, C. M. C. (2025). AI-POWERED APPLICATION FOR CONSTRUCTION SCHEDULE MANAGEMENT USING NATURAL LANGUAGE. Computing in construction, 6.

Pantazis, E. and Gerber, D. (2018). A framework for generating and evaluating façade designs using a multi-agent system approach. International Journal of Architectural Computing, 16(4):248–270.

Rane, N., Choudhary, S., and Rane, J. (2023). Integrating chatgpt, bard, and leading-edge generative artificial intelligence in architectural design and engineering: Applications, framework, and challenges. SSRN Electronic Journal.

Stryker, C. (2025). Types of ai agents. https://www.ibm.com/think/topics/ai-agent-types#578379079. IBM.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

SERIOUS CONSTRUCTION GAME IN VIRTUAL REALITY: EMBEDDING HIGH-RESOLUTION TELEMETRY FOR LEAN PERFORMANCE ASSESSMENT

Kilian Speiser and Jochen Teizer
Technical University of Denmark, Denmark

Abstract

Existing Lean construction simulation games rarely provide detailed personalized feedback for participants but rely on coarse metrics like completion time or flow visualization. This paper investigates how high-resolution telemetry in active multi-role Virtual Learning Environment (VLE) supports lean assessment on teams in a staged introduction of process and automation improvements. We developed a VLE for 11 building trades using role-specific input devices and a telemetry pipeline that captures interactions to compute Lean indicators (flow discontinuity, production rate, waste). We found high potential of embedding high-resolution telemetry into VLE of data-driven evaluation of Lean principles, offering value to educators, trainers, and practitioners.

Introduction

Lean construction builds on concepts from the Toyota Production System, later described as the Lean Production System (Ohno, 1988). It aims to increase value while reducing waste, thereby improving productivity (Koskela, 1992). Lean construction emphasizes several principles: respect for people, value, flow, waste reduction, and continuous improvement. A central school within Lean is pull planning, in which work is released by downstream demand rather than pushed by a central schedule, most prominently operationalized through the Last Planner System (Ballard, 2000). To teach Lean principles, serious games have emerged as an active method, in both analog and virtual forms (Bhatnagar et al., 2023). Serious games enhance learning through structured challenges and feedback (Clark, 1970; Muntean, 2011) (Domínguez et al., 2013).

An early implementation of such lean construction games included the assembly of a building using Lego bricks, with and without pull planning (Warcup and Reeve, 2014). Such games introduce Lean principles over several rounds, and improvements between rounds highlight how users can improve through the adoption of Lean. However, quantifying these improvements requires clear performance indicators. In analog games, this is difficult to achieve based on objective data. For instance, players must record time and manually assess how many tasks were completed correctly. Virtual Environments (VEs) overcome this by automatically capturing task duration or errors at high resolution (Speiser and Teizer, 2024a; Wolf et al., 2022; Wolf and Teizer, 2022), and by visualizing progress or flow lines (Jacobsen et al., 2021; Sacks et al., 2007). However, while serious games in Lean construction enable automated data collection, a crucial gap remains: sub-second logs of user inputs, location updates, tool and material interactions, zone entries/exits, and task events are seldom used to compute Lean predictive Key Performance Indicators (KPIs). The remainder of this paper uses high-resolution telemetry to refer to such logs.

The idea investigated in this paper is the embedding of high-resolution, cross-role telemetry into a multi-user VR game to compute KPIs, thereby enabling actionable feedback at individual and team levels to enable cross-role and cross-round comparison. This work sits within Human–Data Interaction (HDI). The telemetry pipeline operationalizes three HDI principles for construction training: *legibility* (raw behavioral traces are rendered as indicators non-experts can read), *agency* (workers see indicators about their own work and can act on them within the round), and *negotiability* (cross-role indicators such as EWPT make coordination losses visible to the team that caused them rather than logged against the trade that absorbed them).

This paper addresses the following research questions based on the gaps identified before.

RQ1: How can role-specific lean performance be quantified and assessed in real time using in-game telemetry data?

RQ2: How can cross-role telemetry data be used to analyze collaboration patterns and detect coordination breakdowns in VR-based construction training?

A design science research (DSR) methodology iteratively develops and evaluates a networked VR artifact and its telemetry pipeline. The artifact derives a conceptual indicator framework to assess lean performance. In a study combining hoisting and apartment finishing in a high-rise building with 11 roles, role-specific interactions, and staged interventions across rounds, the research results in the following contributions:

C1: A theoretically grounded telemetry-based indicator framework with standardized lean performance indicators suitable for VR telemetry.

C2: An instantiated VR prototype and modular architecture demonstrating how realistic tasks, roles, and indicators can be adapted across multi-round scenarios to operationalize Lean principles.

Lean Performance Indicators

In construction projects, Lean performance is often assessed through outcome-focused indicators such as cost, schedule adherence, or percent complete, which serve as lagging indicators rather than true Lean metrics (Koskela, 1992; Ballard, 2000). Recent reviews highlight inconsistencies in their use that hinder comparison across projects and countries (Asmone et al., 2025). Process-oriented measures that capture flow, waste, and value have long been proposed in Lean research but are rarely applied systematically in practice, partly because collecting the fine-grained data required is difficult on active sites (Jacobsen et al., 2023). At its core, this is a HDI problem: data is fragmented across humans, devices, and trades, and translating it into legible indicators for workers and managers remains unresolved. The barriers are human-related (workers' attention demands), data-related (sparse, noisy site sensing), process-related (variable task definitions), and organizational-related (limited incentives to record). VLEs remove all four by capturing them automatically, offering an opportunity to measure Lean principles in new ways.

Flow is central to Lean thinking. Four indicators are particularly relevant (Rathnayake et al., 2023): (1) Level of Work Discontinuity (LWD), which quantifies crew downtime relative to work progress; (2) Mean Trade Production Rate (MTPR), which reflects the average completion rate of a trade; (3) Excess Work-in-Progress time (EWPT), which highlights inefficiencies between trades in the same location; and (4) Mean Location Production Rate (MLPR), which aggregates progress across trades at a given location (Rathnayake et al., 2023). Together, these provide diagnostic insights than higher-level indicators such as Percent Plan Complete (PPC), which is widely used but heavily dependent on the quality of the baseline construction schedule (baseline plan) against which progress is measured (Asmone et al., 2025).

Waste is another essential dimension of Lean, representing non-value-adding activities. The eight wastes are waiting, motion, inventory, extra processing, defects, overproduction, underutilised talent, and transport (Ohno, 1988; Liker, 2004). While it is difficult to measure waste on-site (Jacobsen et al., 2023), VR environments can automatically capture waste, linking low-level inefficiencies to overall productivity.

Value in Lean was originally defined as what the client is willing to pay for (Koskela, 1992). This concept has since broadened to include stakeholder benefit, end-user wellbeing, and quality-in-use (Salvatierra-Garrido and Pasquire, 2011). In VEs without client in the loop, quality and reliability serve as measurable proxies for the value the client would otherwise judge.

Research methods

The research presented in this paper follows a DSR methodology (Hevner et al., 2004; Kirk, 2013). DSR is appropriate here because the research question is not whether telemetry-derived Lean indicators correlate with productivity in the abstract, but whether such indicators can be constructed at all in a multi-role VR setting and made useful to learners and trainers. This is a design problem in the sense of Hevner et al.: it requires building an artefact, exercising it, observing where it breaks, and iterating. Pure observational or experimental designs cannot produce the artefact itself. The paper aims to produce a conceptually viable artifact as an instantiation of a serious Lean game in VR. The artifact is based on a relevant problem: Existing Lean games do not provide performance indicators across the different Lean principles from low-level data collection.

Recreation of the VR environment

The scene generation follows a mixed approach based on previous research. Two Building Information Modelling (BIM) models were obtained from contractors. The first project is a construction site in Munich, and the BIM model has been included using the IFC Importer (Speiser and Teizer, 2024a). The second BIM model represents a completed student dorm project in Denmark. The BIM model was included using the Unity Export plugin available in the Autodesk store. Using real projects rather than fabricated scenarios reduces authoring effort and avoids questions about ecological validity (Speiser and Teizer, 2024b).

The BIM models miss the construction site layout and machinery. Hence, these elements are added in Unity using Assets. Site equipment, including cranes, forklifts, guardrails, barriers, and materials, is manually added. On top of that, the elements that the users interact with are added manually, and scripts are assigned, like previously published research (Jacobsen et al., 2022). Generally, however, the materials that each trade needs are added. To further strengthen the real-world relevance, trajectory information for a forklift is included using an algorithm developed in previous work (Speiser et al., 2023).



Figure 1. Scene overview of the virtual construction site in Unity with zones highlighted.

The construction sequence integrated in this training environment is based on a handbook shared by a Finnish contractor. It outlines the construction of bathroom pods in residential housing in detail. The handbook was selected because it documents the bathroom-pod sequence at task-level granularity, which the German and Danish project documents lacked; pod-assembly logic is essentially country-agnostic, so the sequence transfers. This sequence was used to recreate the sequence. This sequence was also the baseline in identifying the trades required for the serious game. For simplicity, each trade represents one role, and each role is implemented as one user. Technically, there could be more users for each trade, and it would enhance the freedom during gameplay, but it has been chosen to stick with one per trade to enable smoother gameplay and implementation.

As the user study, which will be described later, includes three different rounds, the scene is also adapted for each of the rounds. The changes include moving the position of equipment, adding or removing safety equipment, or changing tools and interaction possibilities. These changes were chosen based on role-practicability. The objective was that each role had technological enhancements in each round and that Lean principles were included in each role. Hence, each trade receives information on how to reduce waste within their task, how to improve flow (for all trades), and how to improve quality as a proxy for value.

The virtual training environment combines observations from two real projects into one synthetic site encompassing four interdependent activities with eleven

Table 1. Role overview with tasks, required tools, and interaction setup.

Trade	Task description	Required tools and materials	Changes over rounds	Input devices and immersive displays
Crane Operator	Hoist bathroom pods to target units; coordinate with Riggers via signals	Tower crane, rig, bolts/rigging hardware, bathroom pods	R2: Crane repositioned; overhead-load proximity notifications; reminders to clear under load	HMD (optional desktop); Farmstick
Rigger Ground	Attach pod to rig; signal crane; prep/load boxes for forklift	Rigging points; two bolts/connectors; material boxes/pallets; ladder/ platform	R1: Ladder next to bathroom assembly R2: Safe platform replaces ladder; reminder to exit under-load zone	HMD (optional desktop); XR controllers (optional data gloves)
Rigger Top	Prepare the bathroom location with heel tabs and drainage; signal the final position; and disconnect the pod.	Heel tabs for placement; hand tools for disconnect (two bolts); ladder	R1: Manual material collection, no Lean board; R2: Overhead-load notifications; Lean board progress reporting	HMD (opt. desktop); XR controllers (opt. data gloves)
Plumber	Install four pipe runs (hot, cold, wastewater, ventilation); close opening with plasterboard.	Pipes; wrench for connecting pipes; plasterboard	R1: Manual material collection, no Lean board; R2: Materials pre-positioned; Lean board enabled	HMD (opt. desktop); XR controllers (opt. keyboard/mouse)
Screed Layer	Place self-leveling screed; level to finish	Screed; screed pump; levelling tool	R1: Manual material collection, no Lean board; R2: Materials pre-positioned; Lean board enabled	HMD (opt. desktop); XR controllers (opt. keyboard, mouse)
Glazier	Remove guardrail; install window; tidy debris	Window unit; guardrail components; self-retracting lifeline	R1: Manual material collection, no Lean board; R2: Materials pre-positioned; Lean board enabled	HMD (opt. desktop); XR controllers (opt. keyboard, mouse)
Carpenter	Cut and install floor elements; tidy waste	Pre-cut floor elements (by round); miter saw	R1: Cutting & fitting; R2: Pre-cut elements on pallets; Lean board	HMD (opt. desktop); XR controllers (opt. keyboard, mouse)
Painter	Paint two walls; dispose of trash	Paint buckets; brush/paint roller	R1: Bring/return bucket; R2: Bucket pre-staged; Lean board	HMD (opt. desktop); XR controllers (opt. keyboard, mouse)
Electrician	Verify power is off; Install two outlets, two switches, and one wall lamp;	Sockets; switches; lamp; drill	R2: Materials pre-positioned; Lean board	HMD (opt. desktop); XR controllers (opt. keyboard, mouse); 3d-printed drill
LEAN and safety manager	Inspect site; report defects/hazards; support coordination	N/A	R2: Progress visible on Lean board; dashboards between rounds	HMD (opt. desktop); XR controllers (opt. keyboard, mouse)
Autonomous forklift	Transport material boxes along predefined trajectories	Forklift; fork; material boxes; storage area	Trajectory constant; interacts with Rigger Ground workflow	N/A (autonomous)

roles. Table 1 summarizes each trade's tasks, tools, input devices, and round-specific changes. Together, these define the game structure and performance assessment.

Flow indicators as a proxy for productivity

Trade flow is measured through LWD and MTPR. LWD is defined as the ratio of total time spent in between locations to the floor area of the building (12 m² per apartment). It captures discontinuity in a trade's workflow as time per unit area. Since this indicator rewards consistency, even at a slow pace, MTPR is added to represent the average completion time of a trade across all locations. Together, LWD and MTPR provide a trade-level assessment of flow performance as an evidence-based proxy for productivity.

Team performance at the location level is measured through EWPT and the MLPR. EWPT quantifies the time lost due to unnecessary accumulation between crews, with lower values indicating smoother flow. However, as with LWD, steady but slow work is rewarded. To address this limitation, the mean location production rate is added, capturing the aggregated completion time of all trades working in the same location.

Waste indicators as a proxy for productivity

The VE tracks five waste types: motion, extra processing, inventory, defects, and waiting. Underutilized talent, overproduction, and transport are excluded due to their limited relevance in the developed game. For each user role and task, different aspects of waste are monitored. As with the flow indicators, waste is assessed both at the trade (individual) and location (team) level. To maintain consistency with flow terminology, these are referred to as trade waste and location waste.

In this paper, waiting is defined as time spent outside the apartments, which is treated as non-productive time. More advanced tracking methods could include continuous monitoring of worker movement, but here, waiting is only considered when users are not inside the apartments. Waiting is also a key input for the discontinuity indicator. For the tower crane operator, waiting is measured as the time between the Rigger's signal to hold and the subsequent signal to move the load after connection or disconnection. Motion waste captures unnecessary movement that results in lost time. Inventory waste is defined as carrying too many or too few elements. Too few elements result in additional trips, which are already classified as motion waste. Hence, only surplus inventory is considered, and a 10-second penalty is assigned for each excess item. Extra processing waste is defined as repeating activities without added value. In this game, it is only assessed for the Painter and screed layer. The Painter may repaint the same area multiple times, while the screed layer may repeatedly level the surface without improvement. The time spent on such redundant actions is aggregated as extra processing waste.

By combining these low-level contributors into higher-level indicators, the framework enables a new perspective on Lean performance in VR training.

Data collection and analysis

Unity provides robust data collection features. Specialized scripts are created and added to game objects, and for each role, the collected events are stored in the *UserDataCollector* component. Each script was tested individually during early tests of the training scene (Speiser and Teizer, 2024b).

Zones are modelled as colliders so that each avatar's entering and exiting can be detected. These events are sent to the *UserDataCollector*. The content of the virtual toolbox is also sent to analyze whether users brought the correct amount of materials to the rooms. Task completion is recorded based on the installation of the objects. Each trade has a certain number of elements to install. Elements snap into place once the user holds the correct item near its final position.

User studies

The study validates the technical and operational feasibility of the developed artifact to collect high-resolution telemetry data and convert it into performance indicators that enable comparison across roles. Based on previous experiences (Teizer et al. 2017, Teizer et al. 2019, Tagliabue, et al. 2020, and Teizer et al. 2020) the evaluation is structured in three parts. First, each role was tested individually with researchers through think-aloud sessions supported by expert feedback. Two evaluation sessions were then conducted with researchers as participants: Study 1 (n = 3, roles: Painter, Glazier, Plumber) and Study 2 (n = 3, roles: Rigger Top, Plumber, Crane Operator). The aim of this stage was to verify the data collection mechanisms and to confirm that the captured telemetry can be operationalized into indicator-based feedback using established Lean metrics. At this stage, verification and validation are limited to the Riggers, Crane Operator, Plumber, Electrician, Glazier, and Painter, while the Screed Layer and Carpenter roles are still under development.

Computed Role-level indicators

Figure 2 compares the trade-specific results for all indicators. LWD is an indicator of trade flow. At the trade level, the Glazier, Rigger, and Plumber in Study 1 showed much higher discontinuity compared to the Painter and Plumber in the Study 2. However, the Painter had a much lower MTPR. This highlights two findings: the Painter worked at a consistent pace, but at the lowest rate. The choice of combining indicators is therefore essential, since a pure flow metric cannot capture productivity.

Figure 2 illustrates this issue. The flow charts alone do not make the differences immediately clear. It shows the flow over time, and it's also visible that the trade spent most time in unit C3. Nevertheless, quantifiable units are more usable at first sight. While the graphs help, high-level KPIs provide quantification and decision support. Only with the numerical value discontinuity of 66 is an objective flow performance comparison possible.

MTPR for the Glazier and Plumber was comparable, while the Painter and Rigger lagged. Read alongside their LWD, this distinguishes two failure modes that flow

charts cannot: the Painter is consistent but slow (low LWD, low MTPR), whereas the Rigger is fast but discontinuous (high LWD, moderate MTPR). The two cases require different interventions, namely pacing support for the Painter and dependency smoothing for the Rigger. Because the Painter was scheduled last in the work sequence, downstream trades were not affected any more.

Waste indicators provide insights into non-value-adding activities across trades. In Study 1, extra processing dominated the Painter's waste profile (678 s of 848 s). Without task-level guidance, learners optimized for visible coverage rather than first-pass quality, identifying this as the primary lever for round-2 improvement. The Glazier's waste (466 s) was driven by motion (320 s), indicating that material staging rather than execution was the bottleneck. The Plumber's profile (367 s) combined inventory (190 s) and defects (100 s), pointing to logistics and standardization as separate addressable issues.

Study 2 totals were lower across all trades (Crane Operator 255 s; Rigger Top 318 s; Plumber 458 s), but the composition matters more than the total. The Rigger Top's profile is dominated by waiting (218 s), which reflects coordination loss propagated from the Crane Operator rather than Rigger performance. Trades downstream of slow signalers absorb their inefficiency, and only telemetry surfacing this dependency makes the loss attributable to its source. The Plumber's recurring inventory (210 s) and defect (100 s) waste across both studies indicates the issue lies in the task rather than the participant, with material logistics and installation standardization as the relevant levers. Figure 3 illustrates these differences. Charts help to show the spread of waiting, motion, inventory, defects, and extra processing, but numerical KPIs allow for easier comparison. Listing the waste types is important for identifying potential improvements in the next round. The Glazier mainly suffered from motion waste caused by repeated trips to storage. If they did not need to carry the window, they could transport more inventory in one trip and reduce this waste. The Painter's waste was driven mostly by extra processing, which is hard to see in flow graphs. The Rigger experienced the longest waiting time due to slow crane operations. The crane operator also showed high extra processing waste, which points to possible improvements through automated trajectory planning instead of manual operation, since they often deviated from the shortest path.

Computed Team-level indicators

Study 1 shows lower waste and shorter completion time, resulting in a higher production rate. The main waste contributor is extra processing, which is caused exclusively by the Painter (see Figure 4). The second-largest waste category is motion, driven mainly by the Glazier. In contrast, Study 2 eliminates motion waste but shows a higher waiting waste. This waiting time could be reduced if the crane operator worked faster. Inventory and defect waste are comparably high in both groups. The higher waste observed in Study 2 also results in longer completion time per unit.

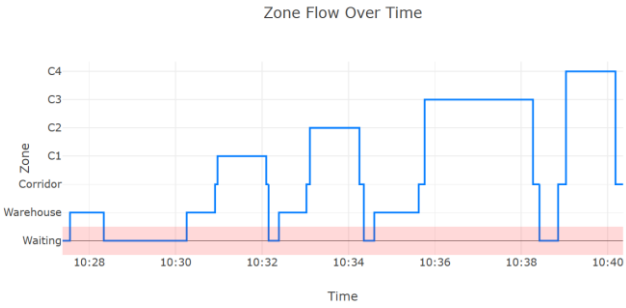


Figure 2. Plumber trade flow over four units.

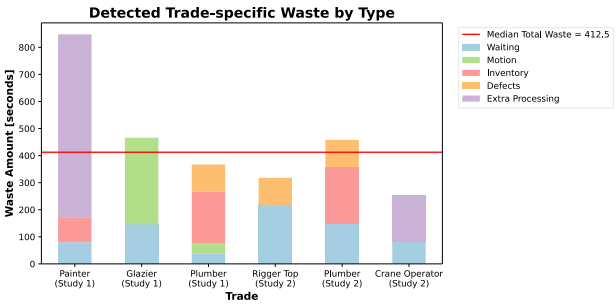


Figure 3. The detected waste for all trades by waste type.

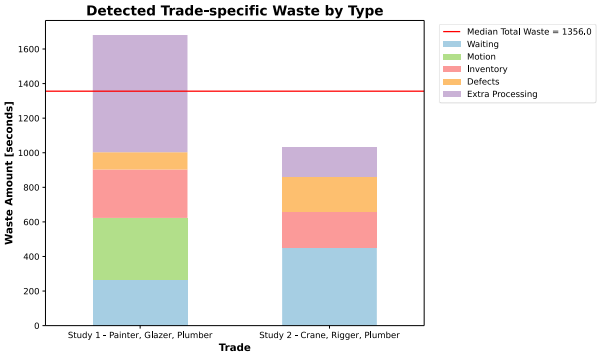


Figure 4. Aggregated team waste by waste type comparing Study 1 with Study 2.

EWPT quantifies what Figure 5 shows visually. Idle times between trades grow as the sequence progresses, most sharply between the Glazier and the Painter, totaling 389 s across locations (Figure 6). Three implications follow. First, EWPT exposes a lagging-trade penalty that PPC and completion time cannot, since locations sit unused even when each individual trade appears on track. Second, the metric is round-comparable: learners can be tasked with reducing EWPT and measure their own improvement quantitatively, which turns the dashboard into a learning instrument rather than a report. Third, the C1 value functions only as a relative baseline; calibrated thresholds remain a field-validation task (see Limitations). EWPT and LWD are interdependent, since a discontinuous trade upstream creates EWPT downstream.

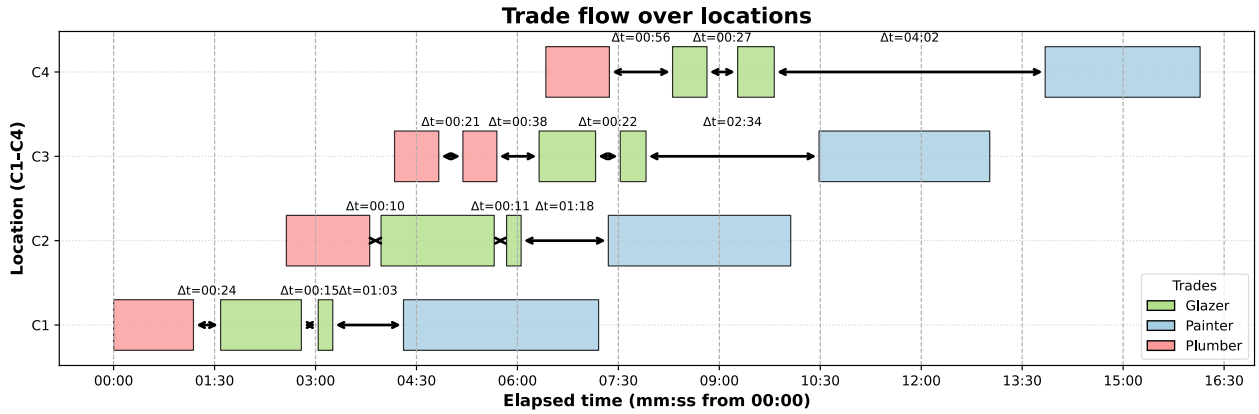


Figure 5. Location flow for group 1, where both motion waste and EWPT per location are included.

Bottlenecks and identified improvements

The Painter consistently worked at a much slower pace than the other trades. Unless their performance improves, subsequent activities will be delayed. A possible solution would be to either add a second Painter or redistribute the workload to balance the overall schedule. The Crane Operator's slower response times also created waiting waste for the Rigger. Improving crane operation speed, for example, through additional training or automation of trajectory planning, could reduce idle times and increase throughput. In addition, workers who arrived on site but could not start immediately contributed to unnecessary waiting and overcrowding. Adjusting arrival times or sequencing work differently would minimize idle presence and improve location utilization.

The Glazier experienced high motion waste due to frequent trips to storage. Improved material staging or delivery batching would reduce unnecessary movement. The Plumber repeatedly generated inventory and defect waste, suggesting systematic issues in material quality or installation practices. Standardizing installation procedures and improving material logistics could help mitigate these recurring inefficiencies. The Painter's extra processing waste points to inefficiencies in task execution. Clearer task planning or defined performance benchmarks could help prevent overworking the same areas.

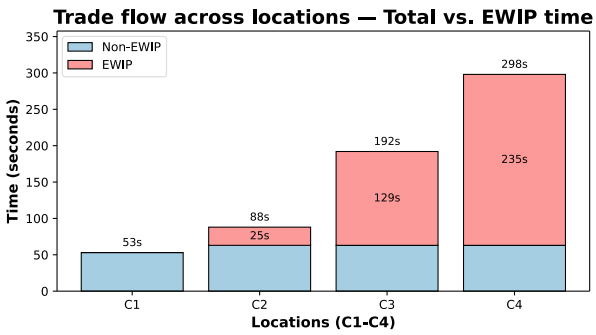


Figure 6. EWPT time by location.

The EWPT analysis revealed significant idle times between the Glazier and the Painter. This indicates a misalignment in trade sequencing that could be improved by staggering tasks or adjusting work areas. Monitoring and reducing EWPT times in subsequent rounds would allow teams to better visualize and quantify improvements in location flow. Moreover, providing real-time feedback on discontinuity (LWD) could enable trades to adjust pacing immediately and reduce downtime.

Conclusions

This paper developed a multi-user VLE that integrates lean indicators through high-resolution telemetry. The system shows how telemetry can turn low-level interactions such as signals, movements, and errors into actionable indicators that quantify flow and waste on both an individual and a team level. Linking role-specific and team-level data, the framework revealed coordination bottlenecks such as crane–rigger timing gaps and excess work in progress that remain hidden in single-role analyses. These results show that the telemetry pipeline enables a more comprehensive evaluation of construction workflows compared to pure completion tracking.

The study makes two main contributions. First, it provides an indicator framework to track lean performance in serious games. Second, it provides methods to collect data in VEs and compute the determined indicators, which support the assessment of both individuals and teams.

Compared to prior VR Lean games (Jacobsen et al., 2021; Sacks et al., 2007), which report aggregate completion or flow visualizations, this paper expands the knowledge through nuanced performance evaluation and eases the identification of coordination breakdowns or weak performance.

The HDI contribution lies in *who* the indicators are computed for. Conventional construction analytics produces dashboards for managers from data workers cannot see; here the same telemetry stream feeds back to the workers themselves during play. This shifts the indicator's function: LWD becomes a signal the trade can act on, not a post-hoc score; EWPT becomes a coordination signal visible to the team that caused it, not a report on them. The Rigger's 218 s of waiting in Study 2 is the clearest example: under conventional reporting, the waste is logged against the Rigger; under telemetry-mediated feedback, it is attributable to the Crane Operator's trajectory and visible to both in real time. This

is what makes pull planning measurable rather than aspirational: downstream readiness, the quantity the Last Planner System asks teams to commit to, becomes a continuous signal rather than a weekly self-report.

This paper is subject to several limitations. Interactions were simplified and limited to zone entries, tool pickups, install snaps, and signaling to ensure reliable measurement, which reduced ecological validity. The evaluation involved a small number of participants and only a subset of trades. The indicators do not yet have calibrated thresholds, so results can only be interpreted in relative terms. The framework has not yet been thoroughly tested in the field with construction workers, which restricts external validity. Telemetry collection from trades-people also raises privacy and ethical questions that this study does not yet address: data minimization, informed consent, and which indicators are appropriate to expose to whom remain open design questions for field deployment.

Practical implications arise for both education and industry. Trainers can use the framework to provide learners with immediate and data-driven feedback on lean performance. Cross-role visualizations, such as location-based flow charts and EWPT, also act as game mechanics that help players and trainers understand Lean principles in an active and engaging approach.

A priority future direction concerns HDI-aligned indicators for the respect-for-people principle. The current framework demonstrates that lean indicators can be computed from telemetry, but does not yet test whether they are meaningful to the people they describe. Future work will operationalize three HDI indicators: *trust calibration* (do users act on the indicator they see?), *perceived agency* over personal data, and *cognitive load* imposed by dashboards during play. These will be measured by pairing telemetry with short in-VR self-reports and post-session interviews. This direction also frames future integration of safety telemetry (Speiser et al., 2024, 2025; Speiser and Teizer, 2024c) and field validation as steps toward calibrated thresholds, benchmarks, and longitudinal evidence of behavioral change.

Declaration of AI in the writing process

During the preparation of this work, the authors used generative AI to proofread text and improve clarity. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Acknowledgments

The research presented in this paper was funded by the European Union Horizon 2020 research and innovation program under grant agreement no. 101058548 (BEEYONDERS) and Erhvervsfyrtårn Femern (FF6S). The Business Lighthouse Femern is co-financed by the European Union and the Danish Business Development Agency.

References

Asmone, A.S., Pal, A., Murguia, D., Rathnayake, A. and Middleton, C. (2025). Where is the evidence of performance improvement in the lean construction literature?, Annual Conference of the International Group for Lean Construction, IGLC, Vol. 33, International Group for Lean Construction, pp. 1–12, 10.24928/2025/0128.

Ballard, G. (2000). The Last Planner System of Production Control. PhD thesis, University of Birmingham.

Bhatnagar, S., Jacob, G., Devkar, G., Rybkowski, Z.K., Arefazar, Y. and Obulam, R. (2023). A systematic review of lean simulation games in the construction industry, Architectural Engineering and Design Management, Taylor and Francis Ltd., 10.1080/17452007.2022.2155604.

Clark, C.C. (1970). Serious games, Business Horizons, Vol. 13 No. 3.

Domínguez, A., Saenz-De-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C. and Martínez-Herráiz, J.J. (2013). Gamifying learning experiences: Practical implications and outcomes, Computers and Education, Vol. 63, 10.1016/j.compedu.2012.12.020.

Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004). Design science in information systems research, MIS Quarterly: Management Information Systems, Vol. 28 No. 1, 10.2307/25148625.

Jacobsen, E.L., Solberg, A., Golovina, O. and Teizer, J. (2022). Active personalized construction safety training using run-time data collection in physical and virtual reality work environments, Construction Innovation, Vol. 22 No. 3, pp. 531–553, 10.1108/CI-06-2021-0113.

Jacobsen, E.L., Strange, N.S. and Teizer, J. (2021). Lean construction in a serious game using a multiplayer virtual reality environment, IGLC 2021 - 29th Annual Conference of the International Group for Lean Construction - Lean Construction in Crisis Times: Responding to the Post-Pandemic AEC Industry Challenges, 10.24928/2021/0160.

Jacobsen, E.L., Teizer, J. and Wandahl, S. (2023). Work estimation of construction workers for productivity monitoring using kinematic data and deep learning, Automation in Construction, Vol. 152, 10.1016/j.autcon.2023.104932.

Kirk, R. (2013). Experimental Design: Procedures for the Behavioral Sciences, 4th ed., SAGE Publications, Inc., Thousand Oaks, 10.4135/9781483384733.

Koskela, L.J. (1992). Application of the New Production Philosophy to Construction.

Liker, J.K. (2004). *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer.* McGraw-Hill.

Muntean, C.C.I. (2011). Raising engagement in e-learning through gamification, The 6th International Conference on Virtual Learning ICVL 2011, No. 1.

Ohno, T. (1988). Toyota Production System Beyond Large -Scale Production Taiichi Ohno, Production, Vol. 1999 No. December.

Rathnayake, A., Murguia, D. and Middleton, C. (2023). Analysing the impact of construction flow on productivity, Annual Conference of the International Group for Lean Construction, IGLC, Vol. 31, International Group for Lean Construction, pp. 1510–1521, 10.24928/2023/0172.

Salvatierra-Garrido, J. and Pasquire, C. (2011). Value theory in lean construction. *Journal of Financial Management of Property and Construction*, 16(1), 8–18.

Sacks, R., Esquenazi, A. and Goldin, M. (2007). LEAPCON: Simulation of Lean Construction of High-Rise Apartment Buildings, Journal of Construction Engineering and Management, Vol. 133 No. 7, 10.1061/(asce)0733-9364(2007)133:7(529).

Speiser, K., Golovina, O. and Teizer, J. (2025). Behavioral safety performance assessment in construction: a data driven approach in virtual reality, Annual Conference of the International Group for Lean Construction, IGLC, Vol. 33, International Group for Lean Construction, pp. 1114–1125, 10.24928/2025/0169.

Speiser, K., Hong, K. and Teizer, J. (2023). Enhancing the Realism of Virtual Construction Safety Training: Integration of Real-Time Location Systems for Real-World Hazard Simulations, Proceedings of the 23rd International Conference on Construction Applications of Virtual Reality, Florence, pp. 155–166, https://doi.org/10.36253/979-12-215-0289-3.15.

Speiser, K. and Teizer, J. (2024a). Formalizing virtual construction safety training: A schematic data framework enabling real-world hazard simulations using BIM and location tracking, ITcon, Speiser & Teizer, Vol. 29, 10.36680/j.itcon.2024.043.

Speiser, K. and Teizer, J. (2024b). Automatic creation of personalised virtual construction safety training using digital twins, Proceedings of Institution of Civil Engineers: Management, Procurement and Law, ICE Publishing, 10.1680/jmapl.23.00104.

Speiser, K. and Teizer, J. (2024c). Worker-Centric Path Planning: Simulating Hazardous Energy to Control Construction Safety Using Graph Theory, 41st International Symposium on Automation and Robotics in Construction (ISARC), Lille, France, pp. 505–512, 10.22260/ISARC2024/0066.

Speiser, K., Teizer, J. and Ali, G.M. (2024). Smart signalperson for safe tower crane operations on a future digital construction site, Proceedings of CIBW099W123, Kumasi, Ghana, pp. 353–365.

Tagliabue, L.C., Ventura, S.M., Teizer, J. and Ciribini, A.L.C. (2020). A Serious Game for Lean Construction Education Enabled by Internet of Things, 5th Intl. Conference on Smart Learning Ecosystem and Regional Development, Bucharest, Romania, 10.1007/978-981-15-7383-5_19.

Teizer, J., Melzner, J., Wolf, M., Golovina, O. and König, M. (2017). Automatisierte 4D-Bauablaufvisualisierung und Ist-Datenerfassung zur Planung und Steuerung von Bauprozessen, VDI-Bautechnik, Jahresausgabe 2016/17, *Bauingenieur*, Springer, October 2017, pp. 129-135, ISSN 0005-6650.

Teizer, J. and Melzner, J. (2019). Neue Wege zur Integration von BIM, IoT und Lean Construction in der Bauingenieurausbildung, VDI-Bautechnik, Jahresausgabe 2019/20, *Bauingenieur*, Springer, October 2019, pp.19-25, ISSN 0005-6650.

Teizer, J., Embers, S., Golovina, O. and Wolf, M. (2020). A serious gaming approach to integrate BIM, IoT and Lean Construction in Construction Education, Construction Research Congress, Tempe, Arizona, USA, 10.1061/9780784482889.003

Warcup, R. and Reeve, E. (2014). Using the Villego® Simulation to Teach the Last Planner® System, Lean Construction Journal, pp. 1–15.

Wolf, M. and Teizer, J. (2022). Serious Game Design Framework to Teach Hazard Awareness and Avoidance on Construction Sites using Pre-defined Hazard Definitions, 22nd Conference on Construction Application of Virtual Reality (CONVR), Seoul, Korea, pp. 348–358.

Wolf, M., Teizer, J., Wolf, B., Bükrü, S. and Solberg, A. (2022). Investigating hazard recognition in augmented virtuality for personalized feedback in construction safety education and training, Advanced Engineering Informatics, Vol. 51, 10.1016/j.aei.2021.101469.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DEVELOPING A LINKED-DATA-BASED FRAMEWORK FOR REGULATION COMPLIANCE CHECKING IN ARCHITECTURAL DESIGN

Yutian Yi[1], Davide Lombardi[1], Zhelun Zhu[1]and Theodoros Dounas[2]
[1]Xi'an Jiaotong-Liverpool University, Suzhou, China
[2]Heriot-Watt University, Edinburgh, Scotland

Abstract

Manual regulation compliance checking remains common in architectural design, while BIM-based and semantic approaches operate outside the authoring environment, requiring specialist expertise, delaying feedback in iterative work. This paper presents a lightweight compliance-checking framework implemented as a pyRevit extension and prototyped in Autodesk Revit. The framework extracts model parameters, converts them to RDF aligned with ifcOWL and a project namespace, and validates constraints using Shapes Constraint Language. The case study shows non-compliances can be detected and reported via both documentation outputs and in-model graphical feedback. The contribution is an auditable linked-data mechanism for extensible regulation checking within BIM authoring tools.

Introduction

The Architecture, Engineering and Construction (AEC) industry is characterised by increasing project complexity, multidisciplinary collaboration, and fragmented information flows across design, approval, and construction stages (Michaud et al., 2019). Within this context, building regulations play a critical role in guaranteeing construction quality and safety. However, regulation compliance checking during architectural design remains largely manual and error-prone (Preidel & Borrmann, 2016). Architects frequently encounter situations where design proposals fail to satisfy specific code requirements because of misinterpretation, omission, or inconsistent misunderstanding of regulatory clauses. In addition, interpreting building regulations can often be subjective, as different practitioners may interpret the same clause in varying ways (Zheng et al., 2022). Such subjectivity often leads to drawings being returned by local authorities for rectification, triggering multiple cycles of revision. Furthermore, due to multiple revisions, data loss and format error might also occur. This extends the construction schedule and incurs additional consumption of labour and materials, as well as costs (Mäki & Kerosuo, 2015). Such ambiguity increases the risk of inconsistent compliance decisions and potentially opens up space for corruption. As a result, under such circumstances, designers struggle to enhance their work efficiency. Recently, smart supervision, which is defined as the application of digital technologies such as sensors and automation to monitor construction activities, has been increasingly adopted to enhance transparency and control (Mhmoud Alzubi et al., 2023). However, research and implementations keep focusing on construction and operation stages, leaving the design stage insufficiently supported. This gap highlights the need for a digital workflow that embed regulation checking directly into architectural design environments, enabling proactive and consistent compliance verification.

This study aims to build a direct link between architectural models and regulatory clauses by representing them in object-oriented and machine-readable forms. These encoded rules can be processed automatically, ensuring consistent and objective compliance checking directly within the Building Information Model (BIM) environment. Thus, this study develops a linked-data-based workflow that digitizes local building codes, automatically extracts corresponding element data from building model, and cross-checks design information against regulatory constraints. The resulting data is structured in a machine-readable format, facilitating the information interoperability and process automation (Zhong et al., 2019). Moreover, the framework is designed without requiring the users to possess advanced computer skills. All processing procedures are completed by providing the appropriate input data. The overarching goal of this research focuses on the minimisation of human errors, the prevention of wasted efforts from information discrepancies, and shortening the pre-construction process.

The remainder of the paper is organised as follows. In section two a brief overview of related work is presented so as to clearly identify the research gap. In section three, the linked-data based framework is introduced in detail. The framework is structured in four parts: code conversion, information extraction and semantic expression, and user-oriented performance. Section four validates the feasibility of the proposed framework by a real case, and section five provides the conclusion and highlights the future work emerging from this study.

Related Work

This study focuses on linked-data and automated compliance checking during architectural design phase.

The following sections elaborate on previous research of the relevant aspects.

Automated compliance checking

Automated Compliance Checking (ACC) is the process of using digital asset models and encoded contractual rules to automatically verify that the design model and as-built asset satisfy commissioning and operational readiness requirements, thereby enabling rapid and consistent compliance assurance (Dimyadi & Amor, 2013). Eastman et al. (2009) first summarised the rule checking process into four phases: Rule Expression Phase, Model Data Preparation Phase, Rule Enforcement Phase, and Check result reporting phase. Two main data sources are indispensable during a typical ACC process, namely the building model and normative knowledge (Amor & Dimyadi, 2021).

Recent papers show ACC progressing from rule execution prototypes toward more robust rule interpretation and institutional adoption. Early work demonstrated ACC for topologically complex requirements by combining model preparation with automated rule evaluation and clear reporting of non-compliances (Bloch et al., 2019). Building on this idea, Bloch et al. (2024) proposed graph-based semantic enrichment to improve robustness beyond hard-coded rule scripts. Doukari et al. (2021) proposed object-centred checking workflows and plug-in implementations that structure the pipeline from requirement definition through execution and result presentation. Furthermore, the recent growth trend indicates that many studies have applied artificial intelligence technology to compliance checks. Zhang & El-Gohary (2019) extract hierarchy information from building codes by dependency parsing based on deep neural network model training. Chen et al., (2024) leverages deep learning and Large Language Models to categorize categorise regulatory documents and conduct a preliminary screening of the text. From a regulatory standpoint, Fuchs et al., (2025) recognises that fully automated decisions may not be acceptable in regulatory contexts, and the need for human oversight still remains, thus, there is need for traceability and transparency. Being able to explain why a certain element passed or failed compliance plays a pivotal role particularly in audits. For better scalability and information sharing, Zhong et al., (2018) developed specific ontologies to represent varying building environmental information and interact it with SPARQL formed clauses.

Linked-data model

The vision of automated compliance checking in the AEC industry has long been limited by interoperability issues. In this vein, the Linked Data paradigm, built on Semantic Web technologies, presents a promising solution for meaningful data integration and machine interpretation (Boje et al., 2020). A key development is ifcOWL, a web ontology language of the IFC schema, which allows BIM models to be represented as Resource Description Framework (RDF) graphs, transitioning data from a closed format to an open, queryable knowledge graph (Pauwels & Terkaj, 2016; Tomczak, 2025). This semantic enrichment facilitates precise representation and reasoning about building information, enabling regulatory rules to be expressed in complex terms. The power of linked data is further demonstrated by its ability to integrate diverse data sources. For instance, the BRICK ontology offers a standardised schema for representing building equipment and their relationships within building automation systems, facilitating interoperable data exchange across platforms (Balaji et al, 2018; Brick Consortium, n.d.). Cerovšek & Omar (2025) illustrate how IFC-based semantic data can be aligned with operational datasets structured using BRICK, highlighting the role of RDF as an architectural layer for interoperability. Donkers et al. (2023) illustrate the dynamic interconnection of RDF data across knowledge domains through web-based applications.

Querying and validating these interconnected RDF graphs require specialised languages. For the specific task of regulation checking, which is inherently a validation process, the Shapes Constraint Language (SHACL) is applied as a standard for validating asserted RDF data, and was published as a W3C (World Wide Web Consortium, the main international standards body for the Web) Recommendation in 2017 (W3C, 2017). SHACL is designed for validating "what must be true" (Heimsbakk & Torkelsen, 2023). This provides a more declarative and direct method to formalise building codes as constraints against the ifcOWL representation. However, a significant gap remains between linked-data-based compliance checking and everyday architectural practice, as many research still depend on manual export-conversion-validation pipelines which interrupt iterative modelling and discourage routine use by design teams. In this study, lightweight refers to an operationally lean workflow that avoids full-model exchange and instead extracts only the minimum set of code-relevant elements and attributes required for checking selected quantitative clauses. In other words, the challenge is no longer how to represent codes and BIM data as knowledge graph, but how to configure and operationalise them as a live, in-context compliance helper inside BIM authoring environments. This leads to two research questions:

RQ1: What lightweight knowledge-graph representation and constraint configuration is sufficient to support reliable checking of quantitative residential-code clauses during design iterations?

RQ2: How can this linked-data checking pipeline be embedded into a BIM authoring workflow to provide real-time feedback for architects without requiring expert coding skills?

To answer these questions, this study bridges the gap by integrating ifcOWL-based data representing and SHACL validation directly within Autodesk Revit through a pyRevit extension which is an open-source add-in framework that enables custom Python-based tools to run directly inside Autodesk Revit, thereby turning semantic checking into a designer-centric tool rather than a separate post-processing task.

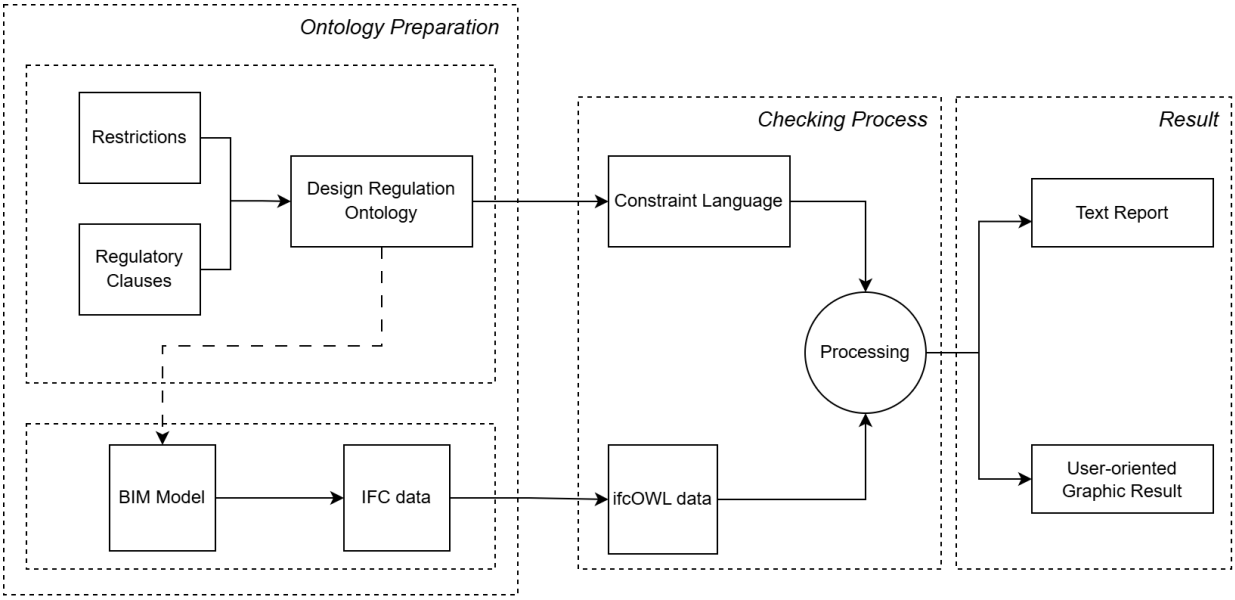


Figure 1: Proposed linked-data-based framework of automated compliance checking

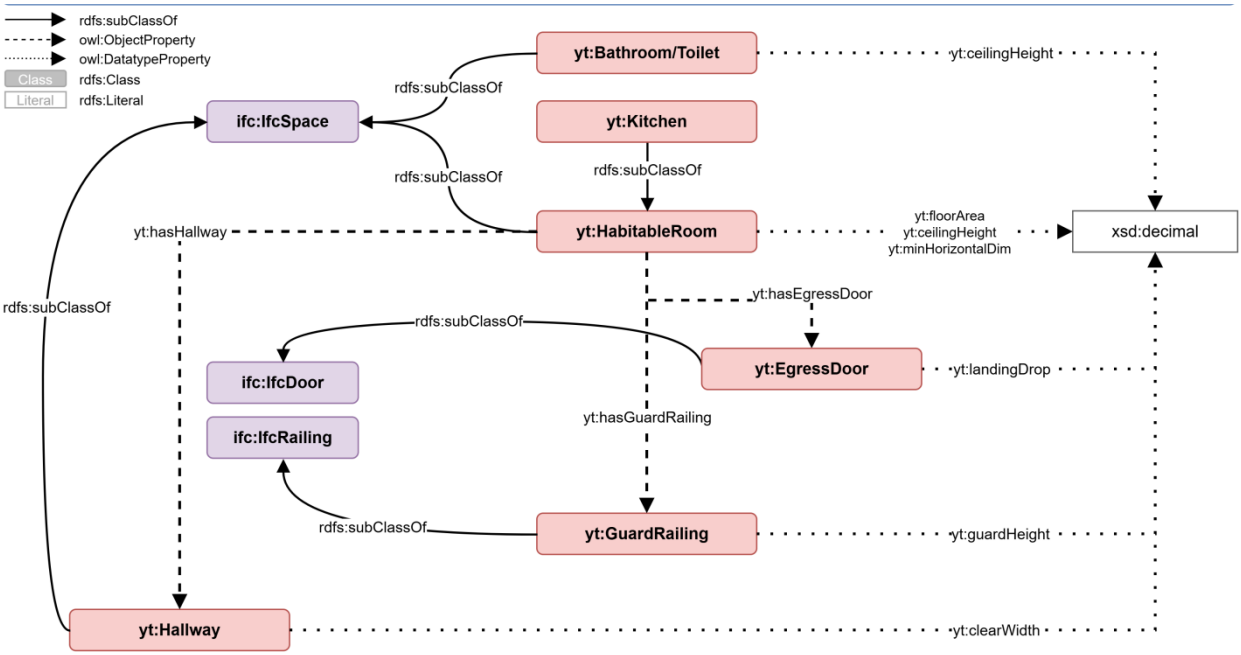


Figure 2: Snippet of the building regulatory ontology schema aligned with ifcOWL

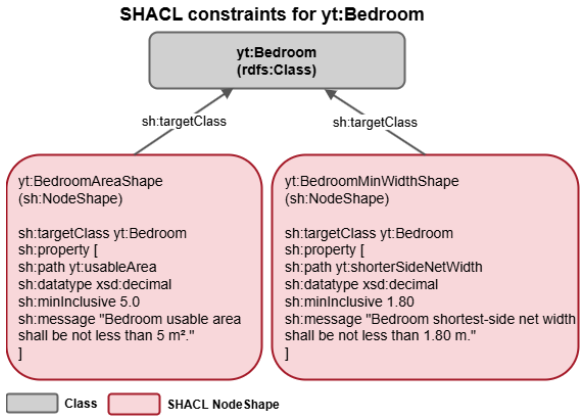


Figure 3: Example of SHACL rule definitions for compliance checking

Methodology

Scope definition and code sources

This study develops a flexible regulatory-checking framework aimed at the architectural design phase, where design teams need real-time, in-context feedback without relying on external consultants or expert programmers. To demonstrate feasibility of embedding code reasoning into the BIM operational environment, a subset of clear and quantitative requirements was extracted and encoded from the 2025 Edition China Residential Code (CRC2025), and it can be found on Github[1].

The framework is intentionally modular. The normative knowledge base acts as a replaceable code database which can be expanded or localised to other jurisdictions by updating the SHACL constraints and the associated regulatory terms. Each rule is expressed in a machine-readable format, enabling scalable adaptation as additional requirements are introduced and/or use of the framework for other regulations. By limiting the initial implementation to well-defined, quantitative provisions, the framework validates the workflow logic while ensuring its utilisation for non-technical design practitioners. This infrastructure provides a foundation for future inclusion of more complex and qualitative compliance criteria through the same linked-data structure.

Linked-data-based framework

This study proposes a linked-data based framework (shown as Fig 1.) for regulation compliance checking that is conceptually independent of any single BIM authoring platform, while being implemented and validated through a Revit-based prototype. The core idea is to decouple normative knowledge from software-specific model representations by first translating regulatory clauses into a regulatory ontology and then aligning model data with this ontology through shared semantic structures.

In the proposed framework, only the minimum required semantic information is extracted from the active BIM model and transformed into RDF with explicit mappings to ifcOWL and a project-specific regulatory ontology. This semantic abstraction enables regulation checking to be executed through SHACL validation at the ontology level, rather than being hard-coded to a specific BIM API. While the current prototype realises the workflow through a pyRevit extension, providing visual feedback, tabular summaries, and auditable reports within Revit, the same semantic rule set and validation logic can, in principle, be applied to models organising from other BIM environments that support IFC-based data exchange. Therefore, the use of IFC/ifcOWL and SHACL in this study is not intended to demonstrate cross-platform interoperability empirically, but to establish a software-agnostic regulatory reasoning layer, with Revit serving as a representative design environment for proof-of-concept validation.

[1] https://github.com/Yutian0425/BuildingRegulations-Ontology/tree/main

Encoding the regulatory clauses with ontology

The proposed framework translates selected quantitative clauses from the China Residential Code into a machine-interpretable regulatory ontology aligned with ifcOWL. Each clause is formalised as constraints over (i) a target building element class (e.g., space, door, railing) and (ii) measurable attributes (e.g., area, clear width, ceiling height). In this way, legal requirements are expressed as explicit semantics that can be consistently validated over BIM-derived data.

Figure 2 presents a subset of the developed ontology. The visual schema in Figure 2 follows the Ontology Design Template by Donkers (2022) to ensure consistent notation and readability. The schema reuses ifcOWL classes to represent BIM entities (e.g., ifc:IfcSpace, ifc:IfcDoor, ifc:IfcRailing) and extends them through a lightweight project namespace "yt:" (Yutian Yi), publicly released on GitHub for reuse, extension, and transparent review. Domain-specific regulatory concepts such as "yt:HabitableRoom", "yt:EgressDoor", and "yt:GuardRailing". They are defined as subclasses of their closest IFC counterparts, preserving interoperability while allowing code-oriented distinctions. Quantitative requirements are then attached through datatype properties (e.g., "yt:floorArea", "yt:clearWidth", "yt:ceilingHeight") whose values are represented as xsd:decimal. Encoding the regulatory knowledge in RDF triples ensures the resulting rule vocabulary is linked, queryable, and reusable across projects, enabling a traceable connection between legal clauses and digital building entities.

In-model data extraction

Within Autodesk Revit, a dedicated pyRevit extension automates extraction of compliance-relevant information directly from the authoring model. Using the Revit API, the extension identifies target element categories (e.g., rooms/spaces, doors, railings) and retrieves only the parameters required by the rule set, avoiding unnecessary geometry export and reducing operational overhead during iterative design.

To support rule execution and semantic alignment, each exported field is annotated with two layers of correspondence. First, an ifcOWL-oriented mapping associates the extracted elements with established AEC semantics (e.g., Room is mapped to ifc:IfcSpace, BalconyRailing is mapped to ifc:IfcRailing, Door is mapped to ifc:IfcDoor). Second, a regulatory mapping links the same elements and attributes to the project's "yt:" ontology introduced above (e.g., habitable-space attributes to "yt:floorArea", "yt:minHorizontalDim", "yt:ceilingHeight"; egress-related attributes to "yt:landingDrop"; guard attributes to "yt:guardHeight"). This dual annotation enables reuse of standardized IFC semantics while preserving the code-specific vocabulary needed for regulatory checking. The extracted data are serialized as a compact JSON-LD payload, where numeric values are preserved alongside an explicit

semantic context that binds each field to ifcOWL and the "yt" regulatory vocabulary. This lightweight linked-data serialization avoids exporting a full IFC model while maintaining machine-interpretable interoperability across tools and rule sets.

Conversion to RDF graphs

The extracted dataset is first serialised as JSON-LD, where every Revit element assigned a stable identifier (e.g., "yt:Room_101") and an explicit semantic context that maps fields to the "yt" regulatory vocabulary and the ifcOWL schema. This means the payload is already a Linked Data representation rather than a purely syntactic JSON export. The workflow then materialises the JSON-LD into an RDF graph (and, when required by downstream tools, into a Turtle .ttl file). In this graph, element instances are represented as resources, attributes are expressed as RDF triples, and their conceptual types are connected to ifcOWL via rdf:type assertions. The conversion step automatically binds the "ifc:" and "yt:" namespaces and produces a standards-compliant RDF dataset for subsequent SHACL validation.

This semantic representation merges two domains that are usually separated in practice: BIM authoring data and rule-based knowledge graphs. By doing so, the building model can participate in the web of data—each wall, door, or space is no longer merely a 3D object but a node in a logically connected graph that can be validated, queried, and extended. The RDF layer therefore acts as a bridge between design tools and semantic reasoning technologies, while JSON-LD provides a lightweight and interoperable carrier format for moving data across components of the workflow.

Rules checking process via SHACL

Compliance checking is executed using the SHACL, which interprets the ifcOWL-encoded model as building model and normative knowledge as shape constraints. Each regulatory clause corresponds to a SHACL shape. Fig 3. presents an example, a "yt:HabitableRoomShape" enforcing sh:minInclusive 6.5 for "yt:floorArea", which means habitable room area must be greater than or equal to 6.5 square meters.

The SHACL engine (PySHACL library) parses the RDF graph, matches each instance of the target class, and then checks the relevant properties against the specified numeric thresholds. Non-conforming elements generate validation reports with precise pointers to both the violating node and the specific rule.

As both the data and the rules share the same ifcOWL backbone, reasoning occurs at the semantic level rather than through file-specific attributes. This alignment allows the framework to extend seamlessly to other jurisdictions or to add higher-order logic by simply appending new SHACL shapes.

Result Presentation

After validation, results are returned through two synchronised channels. On one hand, a textual report and an Excel summary are generated, listing all items and highlighting non-compliant element with its property values and the corresponding rule citation. These documents are automatically saved to the designer's local workspace and can be used for later verification or regulatory submission.

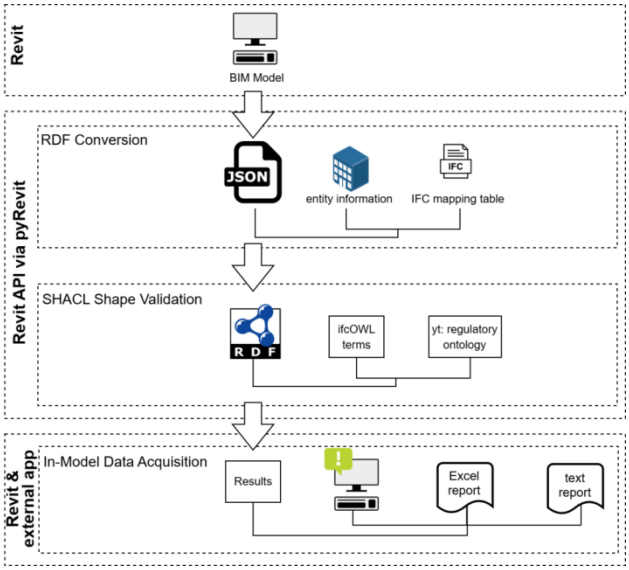


Figure 4: Overall data flow path of the prototype

More importantly, the pyRevit interface presents the feedback directly within the Revit user-side model. The script parses the SHACL report, identifies the offending element IDs, and visually highlights them in the active view using colour coding and annotations. This live feedback loop turns regulation checking into an interactive design activity, where architects can correct issues while modelling rather than after exporting or outsourcing verification. Fig 4. illustrates the data path across the whole framework. Through this integration, the traditionally disconnected processes of design and compliance merge together into a single continuous workflow governed by linked data semantics.

Case study

To demonstrate the feasibility and practical value of the proposed framework, a case study was conducted on a residential project comprising two-family dwellings currently under design development. The project type aligns directly with the applicability scope of the CRC2025, which governs low-rise residential buildings. The model was developed in Autodesk Revit 2026, and the pyRevit extension created in this study was used to perform data extraction, ontology conversion, and regulatory validation within the same digital environment.



Figure 5: The button of the developed pyRevit plug-in and the chosen case model

The chosen project represents an early-stage design model that is still undergoing iterative refinement, allowing the proposed framework to demonstrate its usefulness in real-time design evaluation rather than retrospective verification. Through the developed pyRevit plug-in shown in Fig 5., spatial and dimensional attributes of each building component were exported, converted into an RDF graph based on the ifcOWL schema and the extended "yt" ontology, and validated against the SHACL constraint set derived from local codes.

By suitable building model and normative knowledge input, the validation process successfully identified non-compliance instances, confirming the framework's capability to automatically trace code violations. For example, warnings in the SHACL validation report indicated in Fig 6.: usableArea: Missing or 0.

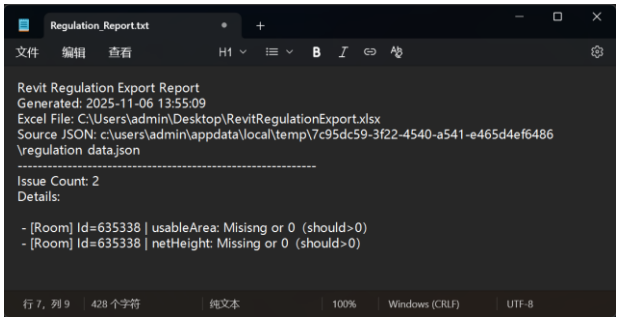


Figure 6: The generated result of regulation report text

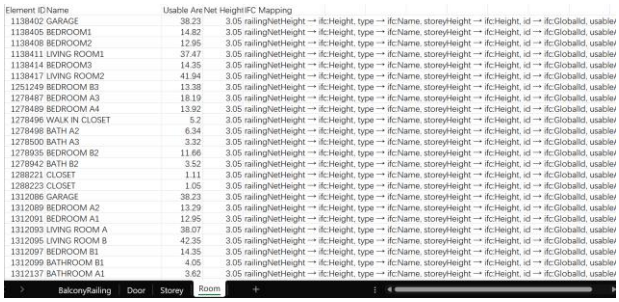


Figure 7: Example figure layout

Beyond textual validation, the framework produced two additional outputs that enhance its practical utility. First, it automatically generated a structured Excel file (Fig 7.), organising all model elements and their compliance status for item-by-item inspection by designers or reviewers. This tabular format enables traceability, documentation, and manual double-checking of design changes over time. Then the system delivered graphical feedback within Revit. Non-compliant elements were automatically selected and highlighted in the model view, enabling designers to visually identify and address each issue directly. If no violations were detected, a pop-up message appeared within Revit, indicating full compliance with the current rule set. Those are shown in Fig 8. This immediate interaction between reasoning output and model visualization underscores the system's lightweight, user-friendly nature.

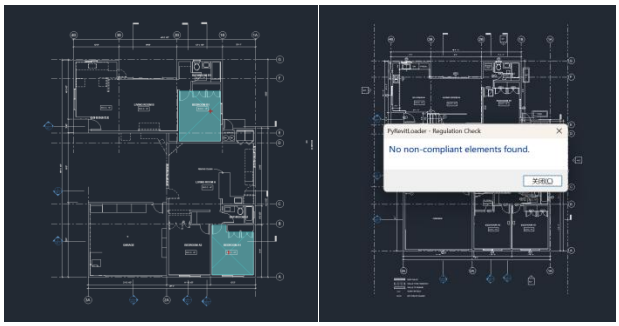


Figure 8: The graphical result shown inside Revit

Conclusions

This study presented a lightweight regulation compliance framework integrated within Autodesk Revit through a custom pyRevit extension. By linking BIM data with semantic web technologies with ifcOWL and SHACL, the system enables architects to perform specific regulation checks intuitively without leaving their design environment. The workflow automates data extraction, RDF conversion, and rule validation while providing both tabular and visual feedback, significantly improving accessibility for non-programming users.

Beyond its practical value, the study offers a theoretical implication by clarifying a reference architecture for "in-model semantic compliance": regulatory knowledge is treated as an explicit, modular layer which includes ontology and constraints that is conceptually separable from any particular BIM platform, while user interaction and feedback remain embedded in the authoring environment. This framing contributes to the Human-Data Interaction literature by positioning compliance checking as a human–data interaction problem where transparency, traceability, and actionable feedback are as critical as rule correctness.

A key limitation is that, at the current stage, a single prototype implementation was developed and validated in Revit; cross-platform interoperability across multiple BIM authoring tools will be empirically evaluated as primary goal in next step of the study. The current implementation is also restricted to quantitative and spatial clauses and still relies on manual encoding of regulatory knowledge. Future work will extend the rule set and test the workflow across additional BIM environments, as well as building on recent advances in AI-assisted code translation, natural language processing and large language models. Such integration of AI-driven knowledge modelling and semantic reasoning could significantly accelerate the creation of digital regulatory databases, improving adaptability, scalability, and adoption of compliance automation in the AEC industry.

Acknowledgments

The authors gratefully acknowledge BIM Engineer Hanqin TANG (WRNS Studio) for generously providing the Revit model used in the case study which enabled the practical validation of the proposed workflow within a real design project context.

References

Amor, R., & Dimyadi, J. (2021). The promise of automated compliance checking. Developments in the Built Environment, 5, 100039. https://doi.org/10.1016/j.dibe.2020.100039

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Bergés, M., Culler, D., Gupta, R. K., Kjærgaard, M. B., Srivastava, M., & Whitehouse, K. (2018). Brick: Metadata schema for portable smart building applications. Applied Energy, 226, 1273–1292. https://doi.org/10.1016/j.apenergy.2018.02.091

Bloch, T., Katz, M., Yosef, R., & Sacks, R. (2019). Automated model checking for topologically complex code requirements – security room case study. In Proceedings of the 2019 European Conference on Computing in Construction (EC3) (pp. 48–55). https://doi.org/10.35490/EC3.2019.157

Bloch, T., Wu, J., Alnuzha, M., & Borrmann, A. (2024). Leveraging graph based semantic enrichment for enhanced automated code-checking. In Proceedings of the 2024 European Conference on Computing in Construction (EC3).

Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. Automation in Construction, 114. Scopus. https://doi.org/10.1016/j.autcon.2020.103179

Brick Consortium. Brick: A uniform metadata schema for buildings. Brick Ontology. Retrieved January 12, 2026, from https://brickschema.org/

Cerovšek, T., & Omar, M. (2025). Advancing Semantic Enrichment Compliance in BIM: An Ontology-Based Framework and IDS Evaluation. Buildings, 15(15), 2621. https://doi.org/10.3390/buildings15152621

Chen, N., Lin, X., Jiang, H., & An, Y. (2024). Automated Building Information Modeling Compliance Check through a Large Language Model Combined with Deep Learning and Ontology. Buildings, 14(7), 1983. https://doi.org/10.3390/buildings14071983

Dimyadi, J., & Amor, R. (2013). Automated building code compliance checking – Where is it at? In Proceedings of CIB World Building Congress 2013: Construction and Society (pp. 172–185). Brisbane, Australia

Donkers, A. (2022). Ontology Design Template. https://doi.org/10.5281/zenodo.681689

Donkers, A., Yang, D., Vries, de V., & Baken, N. (2023). A Visual Support Tool for Decision-Making over Federated Building Information. In C. Andriotis, M. Turrin, & A. Rafiee (Eds), Computer-Aided Architectural Design. INTERCONNECTIONS: Co-computing Beyond Boundaries: 20th International Conference, CAAD Futures 2023, Delft, The Netherlands, July 5–7, 2023, Selected Papers (Vol. 1819, p. 485~500). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-37189-9

Doukari, O., Greenwood, D., Rogage, K., & Kassem, M. (2021). Object-oriented compliance checking: An automated approach and a case study. In Proceedings of the 2021 European Conference on Computing in Construction (EC3) (pp. 293–302). https://doi.org/10.35490/EC3.2021.152

Eastman, C., Lee, J., Jeong, Y., & Lee, J. (2009). Automatic rule-based checking of building designs. Automation in Construction, 18(8), 1011–1033. https://doi.org/10.1016/j.autcon.2009.07.002

Fuchs, S., Fauth, J., Boden, M., & Amor, R. (2025). The challenge of automated compliance checking: A regulatory view. In Proceedings of the 2025 European Conference on Computing in Construction (EC3 2025) & CIB W78 Conference on IT in Construction (Porto, Portugal, July 14–17, 2025).

Heimsbakk, V., & Torkelsen, K. (2023). Using the Shapes Constraint Language for modelling regulatory requirements (No. arXiv:2309.02723). arXiv. https://doi.org/10.48550/arXiv.2309.02723

Mäki, T., & Kerosuo, H. (2015). Site managers' daily work and the uses of building information modelling in construction site management. Construction Management and Economics, 33(3), 163–175. https://doi.org/10.1080/01446193.2015.1028953

Mhmoud Alzubi, K., Salah Alaloul, W., Malkawi, A. B., Al Salaheen, M., Hannan Qureshi, A., & Ali Musarat, M. (2023). Automated monitoring technologies and construction productivity enhancement: Building projects case. Ain Shams Engineering Journal, 14(8), 102042. https://doi.org/10.1016/j.asej.2022.102042

Michaud, M., Forgues, E.-C., Carignan, V., Forgues, D., & Ouellet-Plamondon, C. (2019). A lean approach to optimize BIM information flow using value stream mapping. Journal of Information Technology in Construction (ITcon), 24, 472–488.

Pauwels, P., & Terkaj, W. (2016). EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. Automation in Construction, 63, 100–133. https://doi.org/10.1016/j.autcon.2015.12.003

Preidel C, Borrmann A (2016). Towards code compliance checking on the basis of a visual programming language, ITcon Vol. 21, Special issue CIB W78 2015 Special track on Compliance Checking, pg. 402-421, https://www.itcon.org/2016/25

World Wide Web Consortium. (2017, July 20). Shapes Constraint Language (SHACL) (W3C Recommendation). https://www.w3.org/TR/shacl/

Zhang, R., & El-Gohary, N. (2019, June). A machine learning-based method for building code requirement hierarchy extraction. 2019 Canadian Society for Civil Engineering Annual Conference, Laval, Canada.

https://legacy.csce.ca/elf/apps/CONFERENCEVIEWER/conferences/2019/pdfs/PaperPDFversion_147_0423081134.pdf

Zheng, Z., Zhou, Y.-C., Lu, X.-Z., & Lin, J.-R. (2022). Knowledge-informed semantic alignment and rule interpretation for automated compliance checking. Automation in Construction, 142, 104524. https://doi.org/10.1016/j.autcon.2022.104524

Zhong, B., Gan, C., Luo, H., & Xing, X. (2018). Ontology-based framework for building environmental monitoring and compliance checking under BIM environment. Building and Environment, 141, 127–142. https://doi.org/10.1016/j.buildenv.2018.05.046

Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H., & He, L. (2019). A scientometric analysis and critical review of construction related ontology research. Automation in Construction, 101, 17–31. https://doi.org/10.1016/j.autcon.2018.12.013

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

BETWEEN ARCHITECTURAL DESIGNERS AND DEVELOPERS: EMPIRICAL DESIGN RESEARCH FOR COMPUTATIONAL DESIGN SUPPORT

Elien Vissers-Similon[1] and Theodoros Dounas[2]
[1]University of Antwerp, Belgium
[2]Heriot-Watt University, United Kingdom

Abstract

This paper addresses the gap between architectural designers and computational design support developers by proposing a Lakatosian framework grounded in empirical design research. Combining Peirce's reasoning modes with Schön's reflective practitioner, we extend Pauwels et al.'s (2013) framework by revising the hypotheses on design thinking and tool adoption. We derive three guidelines: computational design support as aides of reasoning modes, as artefacts for experimenting, and as collaborative reasoning agents. For each guideline, we formulate a primary empirical research question. We also propose two empirical instruments to help answer those questions.

Introduction

The architectural design process continuously becomes more complex due to an increase of collaborators, stakeholders, regulations, and client demands. This in turn leads to a growth of computational design support tools development. However, the adoption rate of such tools is rather low in architectural design practice. This is especially true for small and medium-sized architectural design firms, who do not tend to set the narrative on emergent technologies but do represent most European firms. (The Architects' Council of Europe, 2024) Surveys show that parametric modelling tools have particularly low adoption rates for small and medium-sized firms (Stals et al., 2018) despite having access to user-friendly tools such as Grasshopper since 2007. (Tedeschi, 2010) Other surveys show that even building information modelling (BIM) software faces low adoption rates in such firms (Burgess et al., 2026) despite BIM's proven value for the AEC sector. The common denominator for these surveys is that they capture the *designers' perspective* on these computational design support tools. The barriers and limitations observed in the surveys unveil that architectural designers and developers are lost in translation. On one hand, designers assume that computational design support restricts their creativity, they expect expensive implementation difficulties and fail to see sufficient added value for their designerly processes. On the other hand, developers approach design processes as homogeneous processes rather than heterogeneous thinking processes and fail to understand layered decision-making processes within a design task.

To bridge this gap, this paper argues for more empirical design research into architectural design processes, to develop computational design support that is compatible with design thinking. By critically extending Pauwels et al.'s (2013) theoretical framework into a Lakatosian research programme, we establish practical guidelines for design support developers and formulate relevant empirical research questions.

Our argument is grounded in an ongoing empirical research programme studying the adoption of diffusion models in small and medium-sized architectural practice (Vissers-Similon et al., 2026) and educational settings (Vissers-Similon and Dounas, 2026A), which provides situated evidence of how reasoning modes, design artefacts and design features interact in real design processes.

Specifically, we address four research questions:

1. Which Lakatosian framework on computational design support can bridge the gap between architectural designers and developers?
2. Which resulting guidelines are relevant for developers?
3. Which resulting empirical design research questions are relevant for future research on computational design support?
4. Which instruments can we use to pursue these research questions?

Traditionally, computational design support is either a means of representation, a tool, or both. We consider computational design support to be a tool for decision-making within design processes.

State-of-the-art review

Design thinking and computational design support

The relationship between design and technology has been theorised through multiple paradigms. Simon's (1969) *Sciences of the Artificial* first established design as a legitimate scientific inquiry. Schön's (1983) *Reflective Practitioner* introduced a paradigm shift by challenging the technical-rational model, arguing that designers engage in 'reflection-in-action': a process fundamentally different from technical knowledge application. This catalysed subsequent research by Dorst (2011) and Cross (2011) establishing "design thinking" as a distinct mode of cognition operating through iterative cycles of problem-framing and solution-finding. Contemporary

understanding of design thinking draws significantly on Peirce's (1931) triadic framework: abduction as generating hypotheses, deduction as deriving consequences, and induction as testing through observation. Dorst (2011) demonstrates how design problems involve 'frame creation': the abductive leap of simultaneously determining problem and solution approach. Kolko (2010) shows designers rely heavily on abduction during design synthesis, indicating that reasoning mode importance shifts during a design process. Garbuio and Lin (2021) further distinguish explanatory from innovative abduction, mapping abductive activities to problem search frames, hypothesis generation, and hypothesis evaluation.

When considering "computational design thinking", Oxman's (2017) *Thinking Difference* establishes parametric thinking as a distinct form enabling fundamentally different modes of design cognition. Terzidis (2006) extends this by distinguishing parametric from algorithmic thinking. These frameworks reveal a consistent pattern: new computational capabilities enable new forms of design thinking, not merely more efficient existing processes.

Alexander et al.'s (1977) *A Pattern Language* focused on "design knowledge" rather than design thinking, proposing codification through patterns. This work followed an unexpected trajectory: initially developed for architecture, pattern languages were adopted by software engineering (Gamma et al., 1994) and are still relevant for software developers today. From software engineering, pattern languages circled back to influence computational architectural design, at first through Stiny and Gips's (1972) shape grammars and their extensions (Stouffs and Krishnamurti, 1996) and through many applications that have followed since. This crosspollination reveals both the portability and context specificity of design knowledge.

We note that, despite the increasing theoretical understanding of design thinking, it is the more tangible and specific design knowledge that generally shapes the development of new computational design support tools.

Empirical evidence from protocol analysis and practice studies

The above claims on design thinking, that reasoning modes shift and tools shape cognition, are not merely theoretical. Goldschmidt's (2014) linkography method demonstrates that design thinking operates through networks of 'design moves' with identifiable critical moments. Her work shows measurable patterns of fore-linking (divergent thinking) and back-linking (convergent thinking) that fluctuate over design sessions, providing quantitative evidence for dynamic reasoning mode dominance. Gero and Kannengiesser's (2014) situated Function- Behaviour-Structure (FBS) framework has generated extensive empirical evidence through protocol studies comparing design cognition across tools and media, showing statistically significant differences when using hand sketching versus parametric modelling. Suwa and Tversky (1997) identified four cognitive levels (physical, perceptual, functional and conceptual) showing expert architects make significantly more functional inferences from sketches than students. Suwa et al. (1998) extended this, identifying distinct design phases with different cognitive profiles. Dorst and Dijkhuis (1995) empirically compared Simonian and Schönian descriptions of the same protocol, finding reflection-in-action better preserves the relationship between process and content. Cramer-Petersen et al. (2019) demonstrate that abductive-deductive patterns dominate early-stage ideation. Critically, recent empirical work by the authors' research programme provides situated evidence from contemporary practice. Vissers-Similon et al. (2026) conducted a year-long living lab with a small Flemish architectural design practice, studying the adoption of diffusion models across complete design processes. The analysis represents design data on three encapsulating levels: *design features* (such as style, texture, geometry, spatial relationships, feeling, lighting, details); *design artefacts* (such as floorplans, sketches, mood boards, 3D models, generated images) as sequences representing design processes; and *design thinking* (specifically Peirce's reasoning modes) representing reasoning between artefacts. The results show that: (a) reasoning mode dominance shifts measurably across design stages, with abductive reasoning dominant during presentation sketch creation and annotation/discussion moments, deductive reasoning during cost estimation and mood board assembly, and inductive reasoning during 3D modelling and floorplan iteration; (b) despite diffusion model adoption, design decisions were predominantly situated in analogue moments, through manual annotation and discussion, and coincide with abductive reasoning; (c) generated artefacts function as effective design artefacts in Schön's sense, 'talking back' to the designer. Short empirical experiments in education (Vissers-Similon and Dounas, 2026A) also corroborate this, showing students make design decisions based on 'moments of interpretation' that are textual, visual or spatial in nature.

Information systems and reasoning mode alignment

Pauwels et al. (2013) provide the key synthesis linking computational design support to design thinking through Peirce's reasoning modes. Their framework connects cognitive reasoning modes to practical requirements of digital tools, arguing effective support must align with natural reasoning patterns. They propose three application development approaches: computational design support as surrogates for reasoning modes, as tools for experimenting, and as autonomous reasoning agents. They then argue that traditional information systems typically target only one reasoning mode at a time, which may explain limited adoption.

Research gaps

The state-of-the-art review establishes crucial principles but also reveals critical gaps. First, computational design support is most often based on design *knowledge* rather than design *thinking*. Second, design thinking operates through reasoning modes that shift in dominance, as demonstrated by empirical evidence from both protocol

and practice studies. Third, Pauwels et al.'s (2013) framework treats the reasoning cycle as static rather than dynamic and does not accommodate design artefacts as catalysts for reflection-in-action. Fourth, no formalised representation maps Peirce's reasoning modes onto concrete design activities in a way that is both empirically grounded and machine-readable. These gaps motivate a revised Lakatosian framework and accompanying empirical instruments.

Methodology

This paper employs a dual methodological approach. Its primary method is *critical conceptual analysis*: we systematically extend an existing theoretical framework through gap identification and hypothesis revision, following Lakatos's (1970) methodology of scientific research programmes. Its secondary method draws on *design science research* (Hevner et al., 2004): we treat the revised framework and its accompanying empirical instruments as design artefacts, subject to evaluation against utility criteria. The complete methodological approach is visualised in Figure 1.

Selection and analysis of base framework

We select Pauwels et al.'s (2013) framework because it explicitly links computational design support to *design thinking* through Peirce's reasoning modes, rather than to design knowledge. Alternatives were considered: Hevner et al.'s (2004) design science framework focuses on information system artefacts rather than design cognition; Oxman's (2017) parametric thinking framework addresses a specific tool typology; Gero and Kannengiesser's (2014) Function- Behaviour-Structure ontology provides an empirical coding scheme but not a normative framework for tool development.

Gap identification

The base framework rests on two major hypotheses, regarding design thinking and tool adoption. We interrogate each hypothesis against two evidence sources: (a) established protocol analysis studies showing reasoning mode shifts (Goldschmidt, 2014; Suwa et al., 1998; Cramer-Petersen et al., 2019) and (b) situated empirical evidence from the living lab study (Vissers-Similon et al., 2026) mapping reasoning mode transitions across complete design processes in a real architectural practice. A hypothesis is revised when it cannot accommodate documented variability (criterion of flexibility) or when it treats support as replacement rather than augmentation (criterion of compatibility with reflective practice).

A Lakatosian framework for design thinking support

Lakatos (1970) introduced a structure for scientific research programmes distinguishing between a 'hard core' of fundamental commitments and a 'protective belt' of auxiliary hypotheses subject to modification. Applied to our case, this suggests computational design support should be developed as design research programmes: maintaining core commitments while allowing implementations to evolve based on empirical evidence. To accommodate empirical modification of the auxiliary hypotheses in the protective belt, our Lakatosian framework includes practical guidelines for developers of computational design support, and theoretical research questions to facilitate empirical design research.

Design science component: from framework to empirical instrument

Following Hevner et al.'s (2004) design science framework, which itself adopts Lakatos's structure, we propose an empirical instrument as a design artefact: a structured protocol-analysis coding instrument mapping Peirce's reasoning modes onto observable design activities. This instrument operationalises the revised framework and can be evaluated through the empirical studies we propose. Its design draws on Gero and Kannengiesser's (2014) Function- Behaviour-Structure coding scheme, Goldschmidt's (2014) linkography, and Vila-Henninger et al.'s (2022) abductive coding methodology, calibrated against Vissers-Similon et al.'s (2026) three encapsulated levels: design features, design artefacts and design thinking. The empirical instrument provides a formal structure into which findings can be encoded, closing the loop between the empirical observation of design thinking and machine-readable applications of design knowledge in computational design support.

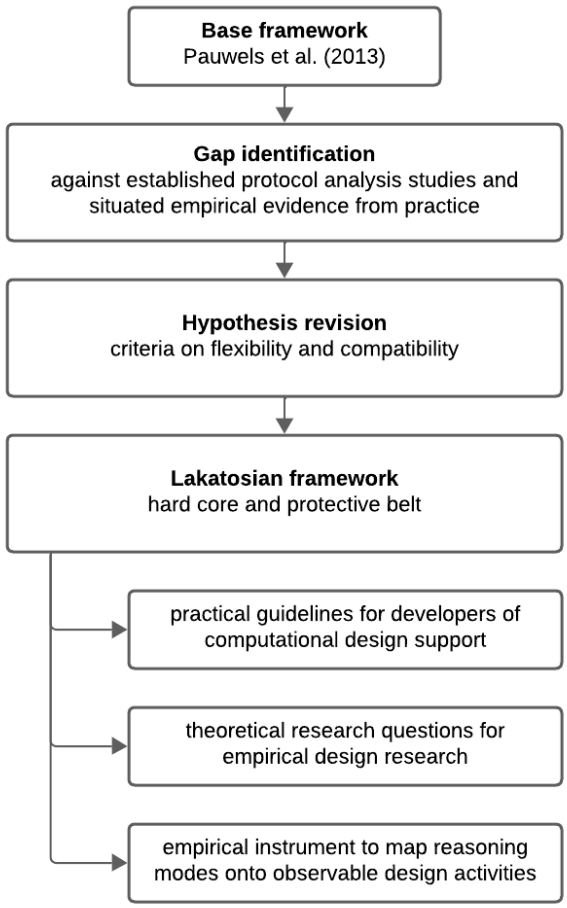


Figure 1: Methodological flowchart

Revised hypothesis on design thinking

The original hypothesis from the base framework states: "Creative designing rests on a cyclic combination of abductive, deductive and inductive reasoning processes." (Pauwels et al., 2013) We support Peirce's reasoning as foundation, because reasoning and decision-making is at the core of designing. However, the original hypothesis leaves no flexibility for varying importance of reasoning modes along a design process.

Protocol analysis and empirical evidence support this: Goldschmidt's (2014) linkography shows fore-linking and back-linking patterns shift measurably. Cramer-Petersen et al. (2019) demonstrate abductive-deductive patterns dominate early ideation. Vissers-Similon et al.'s (2026) living lab provides direct situated evidence: abductive reasoning dominates during presentation sketch creation and manual annotation of generated images; deductive reasoning during cost estimation and mood board assembly; inductive reasoning during 3D modelling and floorplan iteration. Design decisions, the moments that actually advance the project, coincide predominantly with abductive reasoning, conducted through manual annotation and collegial discussion, even when diffusion models are actively adopted.

The living lab thus confirms Schön's (1983) reflecting-in-action with contemporary evidence: generated artefacts from diffusion models succeed 'just as much as traditional artefacts' at 'talking back' to the designer.

We rephrase reasoning modes within the context of design thinking: abduction yields a plausible conclusion based on established design artefacts and background knowledge; deduction deducts a logical conclusion from established artefacts; induction infers a general theory from observations of established artefacts. This leads to an amended hypothesis: "Design thinking rests on a combination of abductive, deductive and inductive reasoning modes. At certain moments during a design process, design thinking adapts so that some reasoning modes are more dominant than others."

Revised hypothesis on tool adoption

The original hypothesis from the base framework states: "Traditional information systems typically target only one of these reasoning processes at a time, which might explain their limited applicability and usefulness." (Pauwels et al., 2013) We question whether targeting a single mode is inherently problematic. What truly matters is understanding which modes are dominant at which stages. The living lab illustrates this: diffusion models function effectively when targeting the right moment, but their outputs require human abductive interpretation to yield design decisions.

The amended hypothesis is thus: "Design support systems typically target only one of the reasoning modes at a time, without understanding which reasoning mode is dominant during the specific design task at hand, which might explain their limited applicability and usefulness."

Guidelines for computational design support

Based on the revised hypotheses, we propose three non-exclusive practical guidelines for developers of computational design support.

Guideline 1: Aides of reasoning modes

Developers should focus on tools that aide reasoning, targeting the right mode for the task at hand. Vissers-Similon et al.'s (2026) living lab shows that during sketch design, when abductive reasoning dominates, designers create hand drawn presentation sketches over printed 3D model views. A mode-aware tool could detect this abductive phase and offer precedent images as prompts, rather than imposing deductive constraint-checking. However, developers should be cautious of the epistemic mismatch in how humans and LLMs produce judgement. (Quattrociocchi et al., 2025)

Guideline 2: Artefacts for experimenting

Developers should focus on tools providing experimental artefacts. The critical question is whether tool outputs function as design artefacts in Schön's (1983) sense: objects provoking reflection-in-action. The living lab provides direct evidence: generated images from diffusion models functioned as effective Schönian artefacts when designers manually annotated them to extract design features (style, colour scheme, materials) and discussed annotations with colleagues. The artefact quality that mattered was *interpretive affordance* - the capacity to provoke functional inference, which Suwa and Tversky (1997) identified as the hallmark of effective design artefacts.

Guideline 3: Collaborative reasoning agents

Developers should focus on collaborative rather than autonomous reasoning agents. Vissers-Similon et al.'s (2026) living lab highlights the necessary collaboration between human designers and diffusion models to make effective design decisions: diffusion models did not generate the project's final façade design, but the human designers did considerably speed up their decision-making process in collaboration with diffusion models.

Empirical research questions

The framework's flexibility depends on empirical evidence to inform the hypotheses. For each guideline, we define one primary research question, which empirical instruments help to answer – and in turn, those new empirical findings might adjust the protective belt's hypotheses. The following empirical design research questions align with the established guidelines:

1. **RQ-E1:** Which reasoning mode is dominant at which moment in the design process?
2. **RQ-E2:** Which types of artefacts facilitate design thinking at which moment in the design process?
3. **RQ-E3:** Which type of interaction facilitates design thinking at which moment in the design process?

Empirical instruments

We propose empirical instruments to help answer the empirical research questions above.

First, a structured coding instrument operating at three levels, which links findings from established protocol-analysis instruments to findings from architectural practice, at each of those three levels. At the micro level, Gero and Kannengiesser's (2014) Function- Behaviour-Structure coding scheme provides the theoretical, ontological base. To map these acts of design, we use "design features". At the meso level, we use abductive coding (Vila-Henninger et al., 2022), to map the "design artefacts" onto their design processes. At the macro level, we relate to Goldschmidt's (2014) linkography method to document how designers think, to map "design thinking". A complete methodology of mapping design features, artefacts and thinking on these three levels is described in Vissers-Similon et al.'s (2026) *Mapping Design Features, Artefacts, and Thinking: Developing a Method for Assessing the Impact of Diffusion Models' Implementation in Interior Design*. By mapping reasoning modes onto observable design activities, we create an empirically grounded instrument for protocol-analysis coding.

Second, for a machine-readable representation, which addresses the fourth and final gap from the "research gaps" section in this paper, we propose an ontology that can be navigated by collaborative reasoning agents. This ontology uses design features, artefacts and thinking as building blocks. The theoretical foundation of such an ontology is described in Vissers-Similon and Dounas' (2026B) *OntoCAAD: an empirically grounded Ontology for Collaborative Agentic Architectural Design*. The OntoCAAD ontology provides the conceptual infrastructure which collaborative reasoning agents could navigate: it formalises design knowledge into classes (Feature, Artefact, Actor, DesignDecision) and formalises design thinking through class interactions (reframe, reason, make, implement) that align with the epistemic pipelines of both human and machine reasoning. (Quattrociocchi et al., 2025) A collaborative reasoning agent built on this ontology could reframe a design problem by identifying relevant features, generate artefacts through artificial intelligence models, and support the designer's reasoning while leaving the actual design decision with the human. The ontology's Actor class distinguishes Person and Agent subclasses, allowing for a precisely calibrated distribution of agency and authorship.

Results: a Lakatosian framework for computational design support

The resulting framework is visualised in Figure 2.

Discussion: empirical research agenda

Regarding RQ-E1, Vissers-Similon et al.'s (2026) living lab provides preliminary answers based on one case study. The empirical mapping instrument can now be applied across additional design processes to test generalisability.

Regarding RQ-E2, Vissers-Similon et al.'s (2026) living lab provides preliminary answers on generated artefacts by diffusion models, which facilitate design thinking through annotation and discussion. OntoCAAD's Artefact class provides the structured vocabulary for further large-scale comparative analysis.

Regarding RQ-E3, Vissers-Similon and Dounas' (2026A) educational experiments show decisions occur at 'moments of interpretation', and OntoCAAD's interact() function and AgentRole enumeration provide categories for evaluating interaction modalities in further large-scale comparative analysis.

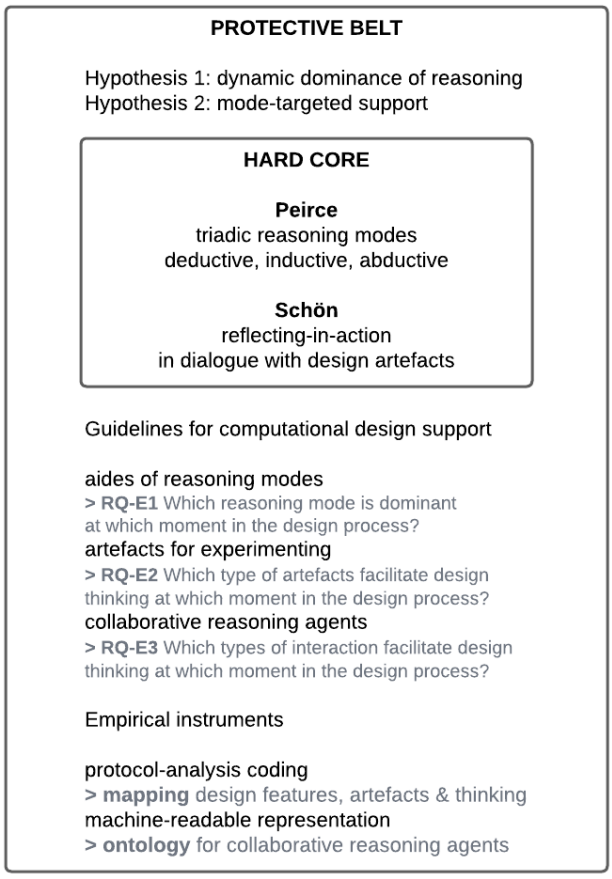


Figure 2: Lakatosian framework for computational design support

Validation roadmap

To provide substantiated answers to the empirical design research questions, and thus to solidify the Lakatosian framework for computational design support, we develop a research programme with the following roadmap.

Phase 1 (completed): establishing the framework. The educational experiments (Vissers-Similon and Dounas, 2026A) and year-long living lab (Vissers-Similon et al., 2026) provided foundational data to establish the Lakatosian framework.

Phase 2 (ongoing): initial calibration of the framework. The empirical mapping instrument as described in Vissers-Similon et al. (2026) will be applied to empirical observations of four additional living lab studies, with small and medium sized architectural firms. The empirical machine-readable instrument, OntoCAAD – as described

in Vissers-Similon and Dounas (2026B), will be built and tested to (i) finetune the ontology and (ii) define potential collaborative reasoning agents.

Phase 3: carrying out the research programme. The OntoCAAD ontology will be integrated in a digital platform, alongside prototypes of useful collaborative reasoning agents. This empirical infrastructure will then be used for testing design sessions across multiple small and medium sized architecture firms to answer RQ-E1, RQ-E2 and RQ-E3 for each collaborative reasoning agent.

Conclusions and contribution

This paper advocates for more empirical architectural design research to bridge the gap between designers and developers of computational design support. We propose a Lakatosian framework based on Peirce's triadic reasoning modes and Schön's reflection-in-action, to substantiate two hypotheses: on dynamic dominance of reasoning modes, and on mode-targeted support.

The framework produces three nonexclusive guidelines for computational design support and two empirical instruments to help achieve these guidelines. The guidelines are computational design support as aides of reasoning modes, artefacts for experimenting and collaborative reasoning agents. The empirical instruments are protocol-analysis coding through mapping, and machine-readable representation through an ontology.

The key contribution lies in recognising that the development of computational design support constitutes applied design research. Each tool instantiates hypotheses about design thinking and creates opportunities for empirical validation, with actual adoption in practice as the end goal. The framework's guidelines provide three primary empirical research questions, which can be answered with the help of two empirical instruments. First, the protocol-analysis coding instrument allows us to map reasoning modes onto observable design activities. Second, the machine-readable representation by the OntoCAAD ontology translates validated findings into structured data, bridging design cognition research and computational agent development.

Limitations

This paper is a conceptual extension grounded in, but not constituting, new empirical data collection. The protocol-analysis instrument is established but not yet calibrated. The ontology instrumented is described but not yet built or calibrated. The empirical findings which serve as the foundation for this research programme are all collected in the Flemish architectural design context; thus, generalisability remains to be demonstrated. The authors bring combined experience in architectural practice, computational design research and empirical design cognition studies, which informed gap identification but may introduce biases toward particular tool typologies.

Acknowledgments

The research on this Lakatosian framework, as well as the completed and ongoing phases of the validation roadmap, was done primarily by Elien Vissers-Similon, under the supervision of Theodoros Dounas. In this paper, Dounas contributed to the state-of-the-art review, research gaps, methodology, discussion and conclusions sections.

References

Alexander, C., Ishikawa, S. & Silverstein, M. 1977. A Pattern Language, New York, Oxford University Press.

Burgess, M., Wilson, C. & Fan, Y. V. 2026. Adoption drivers and barriers of BIM in Europe. Energy and Buildings, 116953.

Cramer-Petersen, C. L., Christensen, B. T. & Ahmed-Kristensen, S. 2019. Empirically analysing design reasoning patterns. Design Studies, 60, 39–70.

Cross, N. 2011. Design Thinking, Understanding how designers think and work, Berg Publishers

Dorst, K. 2011. The core of 'design thinking' and its application. Design Studies, 32, 521–532.

Dorst, K. & Dijkhuis, J. 1995. Comparing paradigms for describing design activity. Design Studies, 16, 261–274.

Gamma, E. et al. 1994. Design Patterns, Elements of Reusable Object-Oriented Software, Pearson Publishers.

Garbuio, M. & Lin, N. 2021. Abductive reasoning in problem finding. Academy of Management Proceedings.

Gero, J. S. & Kannengiesser, U. 2014. The FBS Ontology of Design. In: An Anthology of Theories and Models of Design, Springer, 263–283.

Goldschmidt, G. 2014. Linkography: Unfolding the Design Process, MIT Press.

Hevner, A. R. et al. 2004. Design science in IS research. MIS Quarterly, 28, 75–105.

Kolko, J. 2010. Abductive thinking and sensemaking. Design Issues, 26, 15–28.

Lakatos, I. 1970. Falsification and the Methodology of Scientific Research Programmes. In: Criticism and the Growth of Knowledge, Cambridge University Press.

Lawson, B. 2005. Oracles, draughtsmen, and agents. Automation in Construction, 14, 383–391.

Oxman, R. 2017. Thinking difference, theories and models of parametric design thinking. Design Studies, 52, 4–39.

Pauwels, P., De Meyer, R. & Van Campenhout, J. 2013. Design thinking support: IS versus reasoning. Design Issues, 29, 42–59.

Peirce, C. S. 1931. Collected Papers, Harvard University Press.

Quattrociocchi, W. et al. 2025. Epistemological fault lines between human and AI. arXiv:2512.19466.

Schön, D. A. 1983. The Reflective Practitioner, London.

Simon, H. A. 1969. The Sciences of the Artificial, MIT Press.

Stals, A., JANCART, S. & ELSEN, C. 2018. Influence of parametric tools on complexity in everyday work of SMEs. Archnet-IJAR, 12, 206–227.

Stiny, G. & GIPS, J. 1972. Shape grammars. IFIP Congress, Ljubljana, 1460–1465.

Stouffs, R. & Krishnamurti, R. 1996. On a query language for weighted geometries, Third Canadian Conference on Computing in Civil and Building Engineering (eds. O. Moselhi, C. Bedard, S. Alkass), Canadian Society for Civil Engineering, Montreal, 783- 793.

Suwa, M. & Tversky, B. 1997. What do architects and students perceive in their design sketches? Design Studies, 18, 385–403.

Suwa, M., Purcell, T. & Gero, J. S. 1998. Macroscopic analysis of design processes. Design Studies, 19, 455–483.

Tedeschi, A. 2010. Interview with David Rutten. MixExperience.

Terzidis, K. 2006. Algorithmic Architecture, London.

The Architects' Council of Europe 2024. The Architectural Profession in Europe 2024.

Vila-Henninger, L. A. et al. 2022. Abductive coding. Sociological Methods & Research, 51, 968–1005.

Vissers-Similon, E., De Walsche, J. & Dounas, T. 2026. Mapping Design Features, Artefacts, and Thinking: Developing a Method for Assessing the Impact of Diffusion Models' Implementation in Interior Design. Journal of Interior Design.

Vissers-Similon, E. & Dounas, T. 2026a. The Human Designer in Times of Artificial Intelligence: Diffusion-Driven Architectural Design Explorations. In: Agkathidis, A., Hudert, M. & Medel-Vera, C. (eds.) Architecture in the AI Era for Research, Practice, and Pedagogy. Singapore: Springer Nature Singapore.

Vissers-Similon, E. & Dounas, T. 2026b. OntoCAAD: an empirically grounded Ontology for Collaborative Agentic Architectural Design. In: Proceedings of the 44th Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe). 7-11 September 2026. Lübeck.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

SUPPORTING CONSTRUCTION SITE H&S INSPECTIONS WITH COMPUTER VISION: A BPMN PROCESS FORMALISATION

Federica Madaschi[1], Alessandro Bruttini[2], and Marco Lorenzo Trani[1]
[1]Department of Architecture, Built Environment, and Construction Engineering (DABC), Polytechnic of Milan, Milan, Italy
[2]Department of Architecture (DIDA), University of Florence, Florence, Italy

Abstract

Routine construction site Health and Safety (H&S) inspections remain manual, experience-driven, and weakly digitised. While Computer Vision (CV) has advanced construction safety, particularly for real-time monitoring, it has rarely been investigated to support everyday inspection practices. This paper addresses this gap by analysing a routine inspection process and formalising current practice using Business Process Model and Notation (BPMN) 2.0 to make roles, decisions, and information exchanges explicit and expose key limitations. The study then proposes a CV-enabled inspection process that structures the collection and interpretation of visual evidence and suggests compliance-related observations to the H&S Manager, while preserving their assessment responsibility.

Introduction

In recent years, rapid advances in Artificial Intelligence (AI) have accelerated its adoption in the Architecture, Engineering, Construction, and Operations (AECO) sector, with applications demonstrating practical benefits across a wide range of use cases. However, despite the growing availability of mature technologies, their implementation in everyday workflows often remains limited (Abioye et al., 2021; Khan et al., 2024). This is partly due to the construction industry's fragmented nature, which hinders the acquisition and retention of the data required for AI solutions, thereby constraining effective deployment. Furthermore, evidence indicates that adoption among Small and Medium-sized Enterprises (SMEs) remains comparatively limited and slower than in other sectors (Khan et al., 2024; Regona et al., 2022).

This implementation gap is particularly evident during the construction phase in site Health and Safety (H&S) management, where most applications currently focus on hazard detection and real-time monitoring (Pandithawatta et al., 2024), leaving AI's potential to support other routine activities of H&S managers unexploited. In this regard, construction site H&S inspections remain largely manual and weakly digitised, despite their central role in monitoring on-site conditions (Lin et al., 2025; Natasha Mohamad Shubqizam et al., 2025). In practice, effective AI support is constrained by the limited structure and poor quality of available data and the costs of integrating AI tools into established procedures. Among these, Computer Vision (CV) stands out as one of the most mature, accessible, and readily applicable options (Liu et al., 2021). Nonetheless, its adoption in H&S inspection is still hindered by the absence of shared, clearly defined, and actionable processes that translate CV outputs into operational decision support for H&S managers (Daniel et al., 2025), as well as by socio-legal constraints related to liability and on-site data collection and use (Agapiou, 2024). H&S inspections rely on the professional's experience of the individual manager or organisation, with compliance assessment and information management heavily dependent on unintegrated textual documentation (Forteza et al., 2023; Hussamadin et al., 2023). Without a structured understanding of inspection practices, efforts to develop and integrate AI tools into these workflows risk producing solutions that do not address real operational needs. To this end, this study aims to explore and provide a structured basis for the use of innovative CV-based approaches in construction site H&S inspections. Specifically, current inspection processes are analysed and modelled using the Business Process Model and Notation 2.0 (BPMN) to formally represent activities, decisions, and interactions among the involved actors (OMG, 2013). By clarifying how information is currently generated and shared, and by identifying bottlenecks and repetitive manual tasks, the paper discusses where and how CV can be integrated to systematise the collection, interpretation, and management of visual evidence. The proposed perspective is intended to support compliance assessment and reporting activities while accounting for the H&S manager's responsibility. On this basis, the study provides a process-oriented and structured interpretation of H&S inspections to guide the development and adoption of CV-based tools that support H&S managers' everyday operational practice on construction sites.

Background

This study bridges current manual H&S inspection practices and CV-enabled decision support by synthesising inspection limitations, BPMN workflow modelling, and AI/CV applications, moving beyond hazard detection toward process-driven support.

Site H&S inspection: current practice and limitations

Internationally, Health and Safety (H&S) management is increasingly recognised as a cornerstone of sustainable work, with Occupational Health and Safety Management

System (OHSMS) standards such as OHSAS 18001 and ISO 45001:2018 promoting structured cycles of planning, implementation, monitoring, and review to ensure compliance and continuous improvement, particularly in high-risk sectors such as construction (Arévalo Sarrate et al., 2024; Marhavilas et al., 2022). In the European Union, this systemic approach is embedded in a comprehensive regulatory framework centred on the Framework Directive 89/391/EEC and the construction-specific Directive 92/57/EEC for temporary or mobile construction sites, which establish core obligations for risk assessment, prevention, coordination, and the allocation of responsibilities among project actors.

Comparative studies consistently characterise construction as a high-risk industry that requires robust preventive measures, formal H&S planning, and continuous supervision to ensure that dynamic site conditions remain aligned with regulatory requirements (González García et al., 2021). Within this context, the H&S manager assumes a central coordinating role, bearing significant responsibility for organising prevention activities, assessing compliance, and monitoring site conditions through routine inspections. However, these inspections largely rely on manual, experience-based practices: checks are predominantly visual and non-standardised, and observations are often recorded inconsistently across individuals and organisations, reflecting tacit knowledge rather than a structured process logic (Ramos-Hurtado et al., 2022; Saurin, 2016). Fragmented or paper-based data collection further weakens reporting, traceability, and follow-up. At the same time, inspection programmes are frequently irregular and insufficiently risk-based, leading to under-coverage of recurrent high-severity hazards and delayed corrective actions (Rey et al., 2021). Moreover, inspection outputs are rarely leveraged as leading-indicator datasets; reports are commonly archived rather than analysed to identify patterns, systemic weaknesses, or recurring causal mechanisms, limiting their contribution to prevention and continuous improvement (Saurin, 2016).

BPMN for process analysis and improvement

Despite their central role, H&S inspections have so far benefited only marginally from the broader digitalisation efforts underway in the AECO sector, where BIM, digital twins, and AI-based solutions are increasingly adopted to support design coordination, information management, and operational decision-making. Adequate digital support for routine practices, however, requires explicit formalisation of the underlying processes, so that activities, responsibilities, information flows, and decision points are transparent, comparable, and amenable to systematic assessment and continuous improvement (Häußler et al., 2021). This need is particularly acute for H&S inspection workflows, where inconsistent practices and fragmented documentation hinder accountability and limit learning across projects.

Process modelling approaches offer a structured means to capture such workflows and identify opportunities for standardisation and automation. Among these, BPMN provides a widely adopted, domain-agnostic standard for representing processes in a shared and interpretable form (OMG, 2013) and is increasingly supported by executable workflow engines that enable process enactment and monitoring (Geiger et al., 2018). In this sense, BPMN can act as a bridge between manual inspection practices and future digital solutions by supporting both the analysis of current workflows and the design of improved scenarios.

Several studies have applied BPMN to formalise construction inspection workflows, primarily for quality control and information exchange. Tsai and Hsieh (2016) use BPMN in combination with BIM to plan and rehearse inspection tasks in advance, structuring on-site activities and data capture. Similarly, Comai et al. (2022) adopt BPMN to model document- and data-intensive inspection and quality assessment processes, positioning process formalisation as a prerequisite for coherent digital information exchange and workflow execution. Complementarily, Shin and Cha (2023) highlight the lack of actionable process models as a barrier to the adoption of smart inspection technologies and propose a BIM-based quality inspection workflow supported by smart devices. However, these approaches largely focus on quality inspections and do not address routine H&S inspections in a way that clarifies how AI-based solutions can be integrated or how their outputs support H&S managers' decision-making.

AI/CV-based approaches in construction site H&S

In dynamic, visually complex construction site environments, images and videos provide a rich source of safety-relevant evidence that can be automatically analysed, making CV one of the most mature and widely investigated AI-based approaches in this domain (Fang et al., 2020; Guo et al., 2021). As a result, extensive research has focused on the use of CV to enhance situational awareness and automate parts of safety monitoring that are traditionally performed through first-hand visual assessment. Most CV applications address routine tasks such as worker and equipment detection, personal protective equipment (PPE) compliance checking, and recognising unsafe behaviours or hazardous interactions on site (Paneru and Jeelani, 2021; Seo et al., 2015). Recent advances in deep learning have significantly improved detection accuracy, enabling near real-time identification of safety-related conditions even in unstructured construction environments. Accordingly, CV is mainly applied for continuous or real-time safety monitoring via fixed cameras, mobile devices, or surveillance systems, focusing on detecting safety-relevant objects, conditions, and behaviours to improve awareness of on-site risks (Zhang et al., 2020). Although these approaches demonstrate the technical feasibility of CV for construction safety, they largely remain monitoring- and detection-centric, with limited use of CV outputs to support routine inspection workflows. As a result, CV is often deployed as an isolated solution rather than embedded in coherent, process-driven H&S practices, reinforcing the gap between advanced monitoring systems and established inspection routines. In response,

recent research has combined CV with Large Language Models (LLMs) to translate visual evidence more effectively into inspection-ready narratives and decision support. In this regard, Hussain et al. have reported promising results, combining image-based object detection with regulatory rule checking and then feeding the resulting classification and rationale to an LLM for narrative generation and dialogue (Hussain et al., 2026). However, in their approach, safety knowledge is hard-coded as general regulatory rules rather than retrieved from project-specific safety documentation, so inspection outputs are not grounded in the safety requirements defined during project design and planning.

Research Objective & Methodology

Building on the gap identified in the previous section, this study focuses on defining a structured and generalizable representation of routine H&S inspections. Rather than introducing new technologies, the paper aims to organise inspection activities, information flows, and decision logic into a coherent process that makes implicit practices explicit and operationally interpretable. This structured representation is intended to enable a systematic linkage between on-site observations, relevant safety requirements, and subsequent actions, thereby supporting more consistent inspection outcomes and facilitating the integration of digital support mechanisms. Importantly, the proposed approach is conceived as a foundation for the future development of CV-based inspection-support systems, providing a structured reference framework that can be operationalised through different technological implementations.

To achieve this goal, this study adopts a process-modelling methodology grounded in the BPMN standard (OMG, 2013) to make routine H&S inspection work explicit, comparable, and ready for digital support. The modelling effort is informed by a focused review of regulatory requirements, established best practices, and typical inspection deliverables (e.g., reports, issues, and corrective-action records), with primary reference to the Italian construction context while maintaining alignment with internationally recognised H&S management system principles and regulatory frameworks. The resulting workflow is further informed by direct observations conducted at residential and infrastructural construction sites within the broader research framework underpinning this study, providing empirical grounding for the proposed process structure. BPMN is selected not only for its ability to communicate roles, tasks, information exchanges, and decision points in a shared and interpretable form, but also for its suitability as a foundation for implementation. In particular, BPMN models can be operationalised through workflow engines (i.e., process orchestrators) to coordinate activities across heterogeneous tools and information repositories, supporting traceable and consistent execution in fragmented settings such as construction sites (Dumas et al., 2018; van der Aalst, 2013).

The paper proceeds in three steps, following the methodology adopted in the study. First, it models the end-to-end current construction site H&S inspection process, explicitly capturing the involved actors, core activities, exchanged information, and key decision points. Second, it zooms into a recurring inspection task—i.e., the systematic checking of site features of interest such as ongoing activities, workers and equipment, and layout elements—analysing it at a finer level of detail to highlight friction points, bottlenecks, and repetitive manual operations. Third, building on these insights, it proposes an enhanced process model in which CV-enabled capabilities are integrated at well-defined points to support the structured collection and interpretation of visual evidence alongside other H&S information, and to streamline reporting while preserving the H&S manager's central role and responsibility. Together, these three steps provide a process-oriented basis to guide the informed development and adoption of CV-enabled decision support for routine construction site H&S inspections.

Construction site H&S inspection process

This section presents the analysis and BPMN formalisation of a routine construction site H&S inspection process (Figure 1), modelled from the perspective of the H&S manager. The model focuses on the on-site phase of the inspection and considers the two main participants: the H&S manager, acting as the inspector and responsible for safety coordination, verification, and control during the construction phase, and the contractor, representing one or more contracting organisations operating on site. While the model covers only on-site inspection activities, these are situated within the broader construction site H&S management process, which, prior to the start of site operations, includes planning and verification steps involving both participants and the H&S manager's review of the contractor's documentation. To this end, the contractor is required to submit detailed, site-specific documentation, including Safety Operating Procedures (SOPs) for the activities to be performed, records of workers' general and task-specific training, and technical and administrative specifications for equipment and other planned resources. The construction site safety plan, developed by the H&S manager, defines the overarching site organisation and preventive measures. The contractor's documentation is assessed against this plan for compliance, complementarity, and formal validity—particularly with respect to regulatory requirements, safety procedures, and the validity of training and authorisations. Where gaps or inconsistencies are identified, revisions are requested from the contractor; otherwise authorisation to commence works is issued, concluding this preliminary phase.

Although heterogeneous and largely unstructured, the safety plan, together with the contractor's submission, is archived in a local or shared repository and provides the

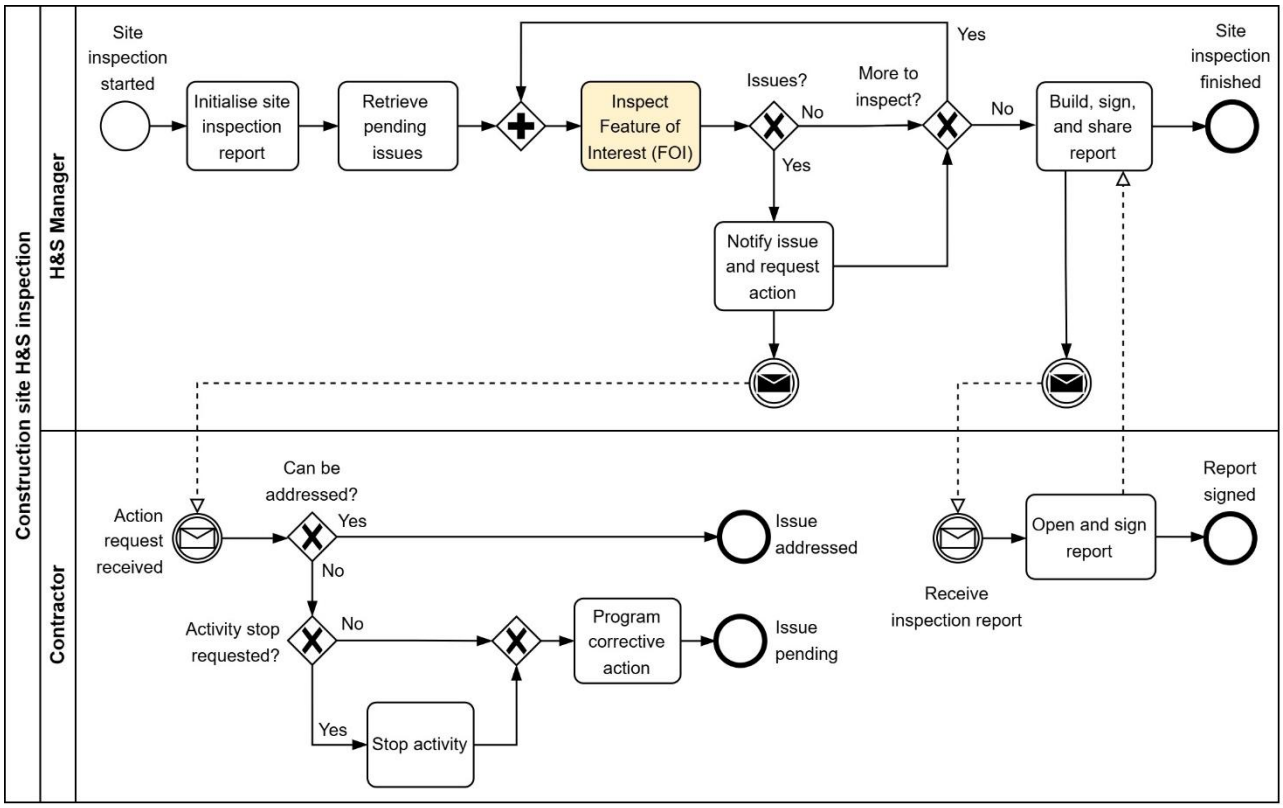


Figure 1: Routine construction site H&S inspection process model (BPMN).

basis for understanding site organisation and involved resources, and for defining the scope of safety inspection activities.

As shown in Figure 1, these are conducted by the H&S Manager as site walkthroughs and begin with creating a new inspection record, including entering site identification information and listing workers and resources (e.g., machinery) present on site. During this formal initialisation, the H&S manager also reviews previously recorded pending issues—especially missing or non-compliant documentation from earlier inspections—to assess corrective actions and identify items that require follow-up during the walkthrough.

Following these preliminary steps, the inspection proceeds to the practical on-site walkthrough, carried out zone by zone across the construction site in accordance with the zonal organisation specified in the construction site safety plan. For each zone, the H&S manager evaluates the compliance of the ongoing activities, relevant site layout elements (e.g., temporary works, access routes, protections, signage), and the workers and equipment present (see Figures 2–3).

These three categories constitute the H&S Features of Interest (FOIs) considered in the process model. For each encountered FOI, the H&S manager directly assesses the observed conditions against the prescribed safety requirements and documents the outcome as a recorded observation, indicating whether they are compliant (e.g., prescribed protections and signage in place), partially non-compliant (e.g., minor deviations from requirements), or critically non-compliant (e.g., severe violations requiring immediate attention). This evaluation and recording represent a core, non-delegable responsibility of the H&S manager, who subsequently manages identified issues by requesting corrective actions from the contractor. Where an observation indicates non-compliance, the process branches to involve the contractor. The contractor is notified of the identified issue and is requested by the H&S manager to implement a corrective action. If the correction is feasible and carried out immediately, the resolution is recorded in the inspection report; otherwise, the contractor programs its implementation, and the issue is registered as a pending item to be verified during subsequent inspections. In cases where the non-compliance is critical—i.e., it constitutes a serious breach of the prescribed safety requirements or poses an immediate risk to workers—the H&S Manager may issue a stop-work order for the affected activity or restrict access to the relevant area until the condition is resolved (e.g., missing mandatory training or authorisation, non-compliant equipment, absence of required permits/ documentation). Once all FOIs within a given area have been assessed, the process iterates to the next zone until the entire construction site has been inspected. The inspection report is drafted and signed by the H&S Manager, then shared with the contractor for acknowledgement and signature. Following this, the on-site inspection concludes. The final report is archived, serving as a formal record of the site's safety at the time and guiding future inspections.

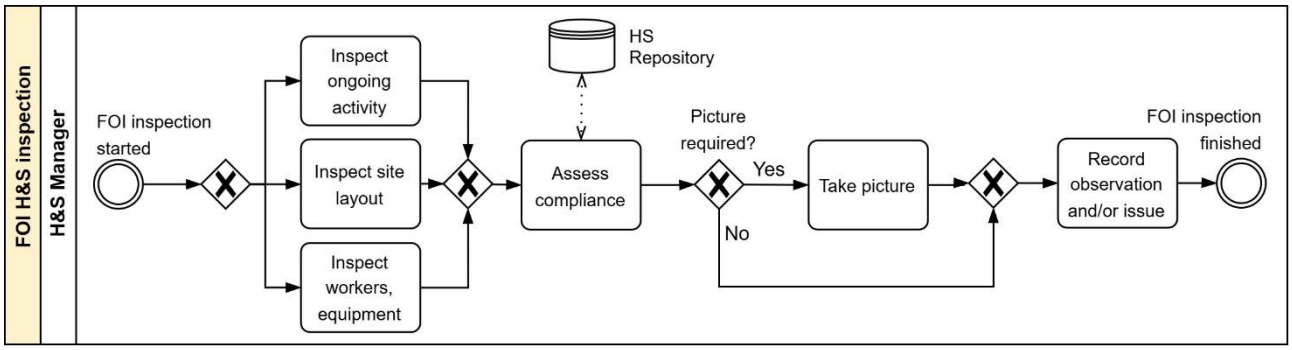


Figure 2: Routine H&S feature of interest inspection sub-process model (BPMN).

Inspection of H&S Features of Interest (FOIs) sub-process

While the overall process includes record initialisation and the management of non-compliances, the most recurrent and information-intensive part of the workflow is the repeated inspection of individual FOIs during the walkthrough. Accordingly, the "Inspect Feature of Interest" sub-process is further detailed and analysed (Figure 2) to clarify current practices and to identify points of friction and opportunities for improvement.

Upon reaching a designated construction site area, the H&S manager assesses compliance with the safety requirements prescribed in the relevant safety documentation by examining FOIs across three complementary categories.

First, ongoing activities are assessed, including the evaluation of operations currently underway in the area and the verification of the authorisations or permits required to perform the specific tasks. Second, the site layout is inspected, focusing on the presence and adequacy of collective protective measures and the implementation of the site organisation and arrangements prescribed in the construction site safety plan and further specified in the contractor's SOP documentation. Third, the workers and equipment present in the area are checked by verifying the validity and completeness of their associated documentation, including workers' general and task-specific training records and machinery inspection and maintenance documentation.

By integrating on-site observations with the relevant safety documentation, the H&S Manager assesses safety compliance for each FOI and, where appropriate, captures photographic evidence to be attached to the recorded observation in support of the inspection outcome. The outcome of the FOI assessment determines the subsequent process flow previously described (Figure 1).

Limitations of current construction site H&S inspection practices

The analysis of the current FOI inspection process highlights several limitations that constrain its efficiency, consistency, and potential for systematic improvement.

Most notably, inspection outcomes remain strongly dependent on the H&S manager's individual practices, expert knowledge, and field experience. While the main tasks and decision points can be identified at the process level, their execution is largely situational and weakly standardised, which reduces replicability and comparability across sites and over time. Supporting the H&S manager in handling this operational complexity motivates the exploration of AI-assisted solutions to structure evidence collection and facilitate timely interpretation and reporting. However, the current process has structural and data-related limitations that complicate this approach, as discussed below. At FOI level, inspections are usually conducted separately: features of interest are evaluated individually without connections to other FOIs or past inspection data, limiting the ability to identify recurring patterns or systemic issues that could guide preventive actions in future inspections or projects. This limitation is reinforced by current recording practices, as FOI assessments are frequently expressed as unstructured natural-language statements, hindering interoperability, structured querying, and integration with other site- and project-related information. Similarly, visual evidence is often collected opportunistically and stored separately, rather than being embedded in a structured workflow and consistently associated with the corresponding observations. Reporting practices represent an additional bottleneck. Inspection reports are typically drafted after the walkthrough through manual consolidation of field notes and documentation, a time-consuming step that can lead to information loss and delayed communication. Reports are commonly archived as PDF documents, which limits their reuse as data sources for analysis and continuous improvement. Finally, documentation management is often fragmented across multiple repositories and formats, further reducing the efficiency and coherence of inspection activities.
Overall, while the inspection process remains central to site safety management, it remains mostly unstructured and poorly suited to systematic data exploitation and AI-assisted support.

Enhanced inspection of H&S Features of Interest (FOIs) with a CV-enabled decision support system

To address these limitations, the enhanced process retains the same underlying inspection workflow while introducing the intervention of a CV-based decision-support system intended to support the user in activities that, in the traditional process, are typically performed manually, interpretatively, and with limited operational structure.

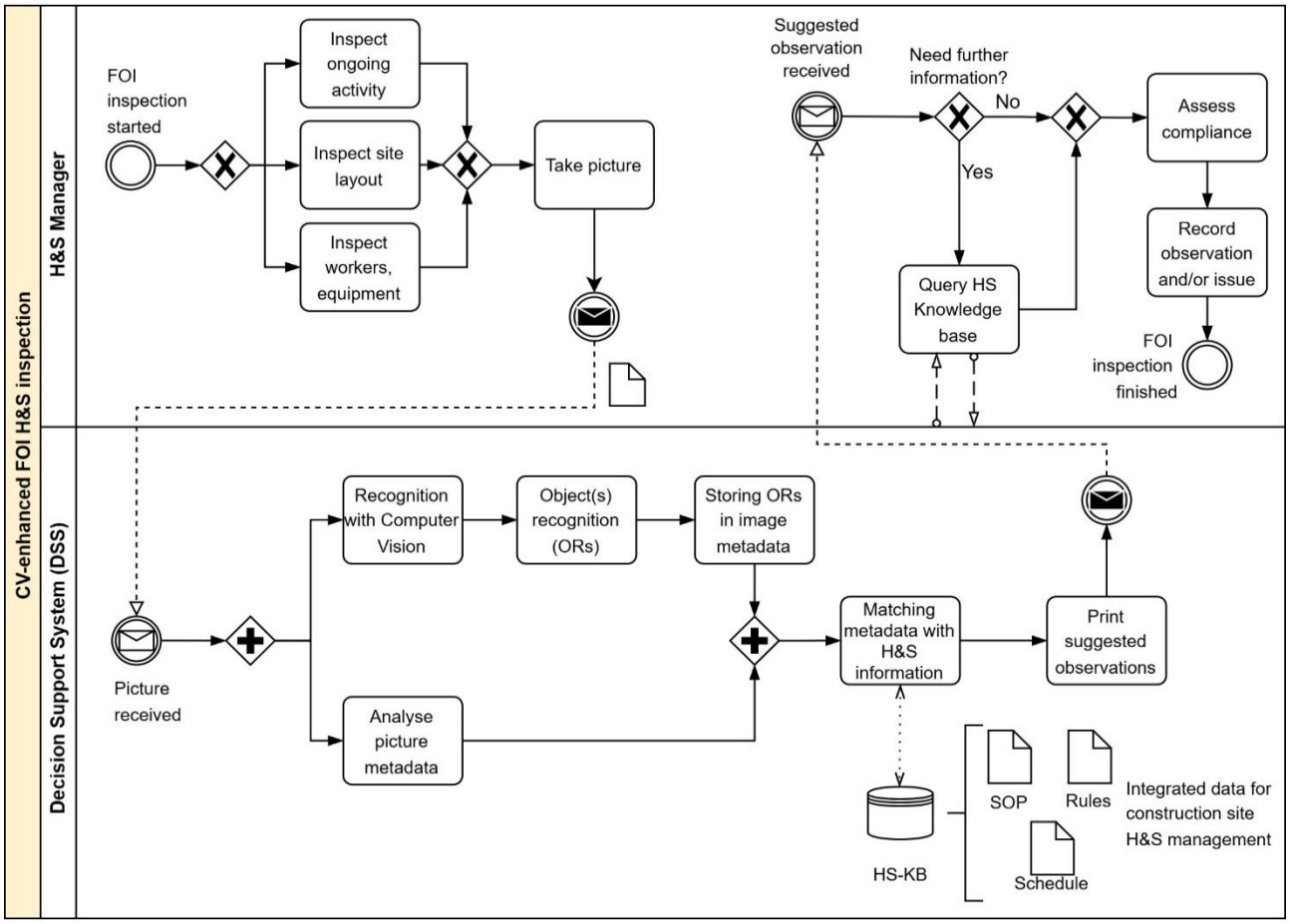


Figure 3: Sub-process model of H&S feature of interest inspection supported by a CV-based Decision Support System (BPMN).

In particular, it enables the real-time validation of on-site observations against project-specific safety requirements, establishing a more direct and actionable link between visual evidence and inspection criteria. Within this perspective, the enhanced workflow is conceived as part of a broader framework connecting inspections with safety planning and knowledge reuse.

As illustrated in Figure 3, the proposed approach intends to leverage existing safety design documents and contractors' safety documentation as primary knowledge sources for inspection planning and on-site control. Since these inputs are predominantly document-based and weakly machine-interpretable, they are first processed to transform H&S-relevant content into structured, queryable data suitable for integration within the proposed decision-support workflow.

The extracted information is stored in a Health and Safety Knowledge Base (HS-KB), integrating safety design requirements, contractors' documentation, inspection observations, and their associated corrective actions. On this basis, a CV-enabled Decision Support System (DSS) is instantiated on demand to cross-reference inspection evidence with applicable safety prescriptions and project-specific requirements. Operationally, the enhanced FOI inspection process embeds this decision-support layer into the routine walkthrough by enabling the H&S Manager to capture photographic evidence via a mobile device and upload it to the DSS. Here, the FOI image is analysed using CV techniques to recognise relevant elements (e.g., guardrails/edge protections), whose labels are stored as photographic metadata together with contextual information such as timestamp and location.

These enriched metadata are used to query the HS-KB and retrieve the corresponding safety prescriptions defined during the design phase, which are then matched with the contractor's SOPs. Based on this matching process, the DSS provides the H&S manager with pre-compiled, context-specific observation suggestions aligned with the applicable safety and project documentation, thereby supporting professional judgement without replacing it. In fact, after receiving the DSS's suggested observations, the final compliance assessment remains entirely under the responsibility of the H&S Manager. The H&S Manager may interrogate the HS-KB in real time to validate the information provided by the system or retrieve additional specifications. FOI-related observations and any encountered issues are then recorded, supporting subsequent communication with the contractor, the request, and the management of corrective actions, as previously shown in Figure 1. Ultimately, this approach enables the production of structured inspection-report data alongside the conventional unstructured document form, supporting not only the H&S Manager's day-to-day inspection activities while enabling systematic reuse of inspection outcomes. The resulting information is stored in the HS-KB, where it can be queried and aggregated

across inspections and construction sites to support cumulative learning and continuous improvement of H&S management practices.

Conclusions and future developments

Despite the growing availability of AI-based technologies, they have not yet been systematically introduced to support routine construction site H&S inspections, which remain weakly structured and largely manual. By adopting a process-modelling approach based on the BPMN 2.0 standard notation, this study formalised current inspection workflows from the H&S Manager's perspective, explicitly capturing roles, activities, decision points, and information flows. The resulting model also highlights key limitations of current practice, including non-standardised observation recording, the need to retrieve relevant H&S information from heterogeneous, fragmented documentation, and the limited reuse of inspection data. Building on this analysis, the paper proposes a structured foundation to guide the development and informed adoption of CV-enabled decision support for routine construction site H&S inspections, integrating a CV-enabled DSS at defined points to support the structured collection and interpretation of visual evidence. Rather than automating compliance assessment, the proposed approach embeds CV within a coherent inspection workflow that preserves the H&S Manager's non-delegable responsibility while improving the consistency and efficiency of inspection activities. From a Human–Data Interaction (HDI) perspective, the proposed approach improves the legibility of inspection processes by structuring the collection and processing of visual evidence within an explicit workflow, supports the H&S Manager's agency by augmenting rather than replacing compliance assessment, and enhances negotiability by allowing system outputs to be interpreted, assessed, and acted upon in light of project-specific requirements and site conditions. These aspects motivate further investigation in future developments.

The process-based approach proposed in this study provides a general foundation for integrating a CV-enabled DSS into routine construction site H&S inspections, but it still requires further specification toward a viable conceptual and technical system architecture, followed by implementation and evaluation.

First, CV algorithms must be further tested and benchmarked in the context of construction site inspections to identify the most suitable approaches.

Future work will examine deep learning models against few-shot and zero-shot approaches, accounting for variability in construction environments, site elements, and routine walkthroughs. Developments will also include structured management of visual evidence and metadata in a HS-KB, such as extracting, storing, and persisting CV outputs and photographic metadata in a machine-interpretable form for integration with other project domains. In this perspective, the system is expected to process visual data potentially including workers present on site, thereby raising issues of data governance, privacy, and responsible data use that future developments will need to address explicitly.

Linking inspection data with safety design documentation, construction planning information, and broader project management data is expected to improve interoperability, support cross-domain reasoning, and promote reuse of inspection knowledge across projects. In this regard, significant challenges remain. A key open issue is defining data structures that align with existing construction and safety documentation practices while interfacing effectively with regulatory requirements.

Addressing these challenges will be essential to ensure that CV-enabled inspection support can be adopted in practice, remain interoperable, and remain aligned with regulatory frameworks and professional responsibilities in construction H&S management.

References

Abioye, S.O., Oyedele, L.O., Akanbi, L., Ajayi, A., Delgado, J.M.D., Bilal, M., Akinade, O.O. and Ahmed, A., 2021. Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, *44*. DOI: 10.1016/j.jobe.2021.103299

Agapiou, A., 2024. A systematic review of the socio-legal dimensions of responsible AI and its role in improving health and safety in construction. *Buildings*, *14*(5), p.1469. DOI: 10.3390/buildings14051469

Arévalo Sarrate, C., Tarín Martínez, J., Lara Galera, A.L. and Galindo Aires, R.Á., 2024. Effect of health and safety management systems in the construction sector.*Buildings*,*14*(1),p.167.DOI: 10.3390/buildings14010167

Comai, S., Mastrolembo Ventura, S., Costa, S. and Ciribini, A., 2022, January. Process modelling for managing digitalised information exchanges in construction inspections and quality checks. *Conference Series: Materials Science and Engineering* (Vol. 1218, No. 1). DOI: 10.1088/1757-899X/1218/1/012030

Daniel, E.I., Oshodi, O.S., Nwankwo, N.I., Emuze, F.A. and Chinyio, E., 2025. The use of digital technologies in construction safety: A systematic review.*Buildings*, *15*(8), p.1386. DOI: 10.3390/buildings15081386

Dumas, M., La Rosa, M., Mendling, J. and Reijers, H.A., 2013. *Fundamentals of business process management* (Vol. 1, p. 2). Heidelberg: Springer. DOI: 10.1007/978-3-662-56509-4

Fang, W., Love, P.E., Luo, H. and Ding, L., 2020. Computer vision for behaviour-based safety in construction: A review and future directions.A*dvanced Engineering Informatics*, *43*, DOI: 10.1016/j.aei.2019.100980

Forteza, F.J., Carretero-Gómez, J.M., Estudillo, B. and Sesé, A., 2023. From Risk Assessment on Site to How to Improve Safety: An Easy "App" to Control

Construction Site Conditions. *International journal of environmental research and public health*, *20*(5), DOI: 10.3390/ijerph20053954

Geiger, M., Harrer, S., Lenhard, J. and Wirtz, G., 2018. BPMN 2.0: The state of support and implementation. *Future Generation Computer Systems*, *80*, pp.250-262. DOI: 10.1016/j.future.2017.01.006

García, M.G., Cañamares, M.S., Escribano, B.V. and Barriuso, A.R., 2021. Constructioń s health and safety Plan: The leading role of the main preventive management document on construction sites. *Safety science*, *143*. DOI: 10.1016/j.ssci.2021.105437

Guo, B.H., Zou, Y., Fang, Y., Goh, Y.M. and Zou, P.X., 2021. Computer vision technologies for safety science and management in construction: A critical review and future research directions. *Safety science*, *135*. DOI: 10.1016/j.ssci.2020.105130

Häußler, M., Esser, S. and Borrmann, A., 2021. Code compliance checking of railway designs by integrating BIM, BPMN and DMN. *Automation in Construction*, *121*. DOI: 10.1016/j.autcon.2020.103427

Hussain, R., Lee, D., Abbas, M.S., Zaidi, S.F.A., Pedro, A. and Park, C., 2026. Vision-language model-based intelligent assistant for onsite construction safety inspection. *Automation in Construction*, *182*. DOI: 10.1016/j.autcon.2025.106728

Hussamadin, R., Jansson, G. and Mukkavaara, J., 2023. Digital quality control system—A tool for reliable on-site inspection and documentation. *Buildings*, *13*(2), p.358. DOI: 10.3390/buildings13020358

Khan, A.N., Mehmood, K. and Soomro, M.A., 2024. Knowledge management-based artificial intelligence (AI) adoption in construction SMEs: the moderating role of knowledge integration. *IEEE transactions on engineering*,*71*. DOI: 10.1109/TEM.2024.3403981

Lin, P.H., Lin, J.J. and Hsieh, S.H., 2025. Construction Site Scaffolding Completeness Detection Based on Mask R-CNN and Hough Transform. DOI: *arXiv:2503.14716*.

Liu, W., Meng, Q., Li, Z. and Hu, X., 2021. Applications of computer vision in monitoring the unsafe behavior of construction workers: Current status and challenges. *Buildings*,*11*(9),p.409.DOI: 10.3390/buildings11090409

Marhavilas, P.K., Pliaki, F. and Koulouriotis, D., 2022. International management system standards related to occupational safety and health: An updated literature survey.*Sustainability*,*14*(20). DOI: 10.3390/su142013282

Shubqizam, N.N.N.M., Zulkurnain, N.F., Gunawan, T.S., Yusoff, N.M., Mansor, H. and Ashraf, A., 2025, September. Real-Time PPE Compliance Monitoring on Construction Sites Using YOLOv7 and YOLOv8 Object Detection Models. In *2025 IEEE 11th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)* (pp. 316-321). DOI: 10.1109/ICSIMA66552.2025.11233761

OMG, O.M.G., 2013. Business Process Model and Notation (BPMN), Version 2.0.2.

Pandithawatta, S., Ahn, S., Rameezdeen, R., Chow, C.W. and Gorjian, N., 2024. Systematic literature review on knowledge-driven approaches for construction safety analysis and accident prevention. *Buildings*, *14*(11). DOI: 10.3390/buildings14113403

Paneru, S. and Jeelani, I., 2021. Computer vision applications in construction: Current state, opportunities & challenges. *Automation in Construction*,*132*,p.103940. DOI: 10.1016/j.autcon.2021.103940

Ramos-Hurtado, J., Muñoz-La Rivera, F., Mora-Serrano, J., Deraemaeker, A. and Valero, I., 2022. Proposal for the deployment of an augmented reality tool for construction safety inspection. *Buildings*, *12*(4), p.500. DOI: 10.3390/buildings12040500

Regona, M., Yigitcanlar, T., Xia, B. and Li, R.Y.M., 2022. Opportunities and adoption challenges of AI in the construction industry: A PRISMA review. *Journal of open innovation: technology, market, and complexity*, *8*(1), p.45. DOI: 10.3390/joitmc8010045

Rey, R.O., de Melo, R.R.S. and Costa, D.B., 2021. Design and implementation of a computerized safety inspection system for construction sites using UAS and digital checklists–Smart Inspecs. *Safety science*, *143*, p.105430. DOI: 10.1016/j.ssci.2021.105430

Saurin, T.A., 2016. Safety inspections in construction sites: A systems thinking perspective. *Accident Analysis & Prevention*, *93*, pp.240-250. DOI: 10.1016/j.aap.2015.10.032

Seo, J., Han, S., Lee, S. and Kim, H., 2015. Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, *29*(2), pp.239-251. DOI: 10.1016/j.aei.2015.02.001

Shin, H.J. and Cha, H.S., 2023. Proposing a quality inspection process model using advanced technologies for the transition to smart building construction. *Sustainability*, *15*(1), p.815. DOI: 10.3390/su15010815

Tsai, Y.H. and Hsieh, S.H., 2016, October. Process modeling of a BIM-enabled construction inspection approach with BPMN. In *Proceedings of the international conference on innovative production and construction, Darwin, Australia* (pp. 20-22).

Van der Aalst, W.M., 2013. Business process management: a comprehensive survey. *International Scholarly Research Notices*, *2013*(1), p.50798. DOI: 10.1155/2013/507984

Zhang, M., Shi, R. and Yang, Z., 2020. A critical review of vision-based occupational health and safety monitoring of construction site workers. *Safety science*, *126*, p.104658. DOI: 10.1016/j.ssci.2020.104658

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A DECOUPLED ARCHITECTURE FOR DIGITALLY ACTIONABLE AND TRACEABLE REGULATORY WORKFLOWS IN THE AEC SECTOR

Hans Kristjan Aljas, Ergo Pikas, and Martin Thalfeldt
Tallinn University of Technology, Tallinn, Estonia

Abstract

The Architecture, Engineering, and Construction (AEC) sector faces complex regulatory frameworks like building energy performance certification under the European Energy Performance of Buildings Directive (EPBD). Traditional approaches suffer from fragmented documentation, procedural ambiguities, and poor traceability and auditability, causing outcome inconsistencies and automation challenges. This paper proposes a decoupled conceptual architecture for digitally actionable and traceable regulatory workflows, leveraging compliance modelling, semantic web technologies, and linked data. By separating semantic representation, reasoning, execution, and interaction layers, it improves compliance, interoperability, and adaptability. Demonstrated on Estonian Energy Performance Certificates (EPCs), it addresses key gaps and supports broader application beyond energy efficiency.

Introduction

Regulatory workflows in the AEC sector are characterised by high complexity. They simultaneously involve physical building elements and procedures, formal norms (directives, standards, laws), institutional and professional practices (national methodologies, auditing bodies, firms), and heuristic expert judgment (soft law, socio-cognitive interpretation).

A representative example is the Energy Performance of Buildings Directive (EPBD) and its national transpositions (European Parliament, 2024). Despite a common European framework, these produce poorly comparable Energy Performance Certificates (EPCs) (Jenkins et al., 2024; Sayfikar and Jenkins, 2025) riddled with methodological non-compliance and inconsistencies (Aljas et al., 2025; Carvalho et al., 2024; Schaffer et al., 2026), and create maintenance burdens for public authorities responsible for custom software pipelines (buildingSMART International, 2023). These phenomena reveal a fundamental tension between:

(1) the hybrid socio-technical nature of regulatory norms and
(2) the current paradigms used to digitise and automate them.

Importantly, this tension cannot be resolved through comprehensive standardisation or exhaustive semantic formalisation alone. Regulatory methodologies are shaped by institutional, geographical, and methodological heterogeneity, alongside practical and organisational constraints that limit the feasibility of total harmonisation (Jenkins et al., 2024; Sayfikar and Jenkins, 2025). As a result, regulatory workflows must remain operable under conditions of partial standardisation, evolving methodologies, and uneven digital maturity.

Existing approaches treat regulations either as static rules to be checked post-hoc aka “Compliance by Detection” or as logic to be programmed into custom software aka “Compliance by Design” (Casanovas, 2017). Both paradigms inevitably produce tightly coupled, brittle systems that struggle to evolve harmoniously with the norm, the context, or the technology. Furthermore, (Feng and Wang, 2025) note that overlapping stakeholder pressures in complex construction workflows expose the limitations of post-hoc compliance mechanisms.

This work advances Compliance through Design (CtD) — the governance-oriented, interpretive paradigm introduced by (Casanovas, 2017)—by providing a systematic operational framework for digitally actionable and traceable regulatory workflows.

The core artefact is a decoupled conceptual architecture with four layers (Semantic Workflow Model, Reasoning and State Engine, Execution and Orchestration Layer, Interaction Layer). The artefact functions as a knowledge-modelling framework that treats regulatory workflows as evolvable, linked-democratic ecosystems rather than monolithic code bases or conventional data pipelines. In practice, such an approach aims to reduce data and logic fragmentation, allowing stakeholders to focus on their core responsibilities or interests by minimising routine administrative tasks and extraneous cognitive load, and thereby reducing the potential for errors.

Taxonomy of compliance

In regulatory contexts, compliance encompasses adherence to laws and standards while extending to ethical and societal norms (Governatori, 2018; Hashmi et al., 2018). Compliance paradigms form a progression of increasing proactivity, addressing the limitations of reactive methods in complex environments.

Compliance by Detection operates reactively, identifying non-conformity during or after execution, which often necessitates costly redesign. Conversely, Compliance by Design shifts to prevention, incorporating rules into the design stage to ensure conformity in advance (Casanovas, 2017; Governatori, 2018). Legal Compliance by Design further refines this by addressing value-based requirements and the legality of the entire process rather than simple formal rules (Hashmi et al., 2018).

Compliance through Design (CtD) constitutes the most holistic evolution; it moves beyond rule-embedding to encompass the systemic interoperability required to model affordances—the action possibilities enabled by a system—alongside broader socio-legal conditions (Poblet et al., 2019). The paradigm acknowledges that compliance is not merely a technical constraint but a social process involving legal interpretation, institutional coordination, and the relationships between regulated entities, citizens, and the law (Casanovas, 2017; Hashmi et al., 2018). From this viewpoint, regulatory requirements operate within an ecosystem of practices and responsibilities, rather than being fully reducible to technical checks.

The shift toward greater proactivity is technically enabled by declarative paradigms. Unlike imperative approaches that specify rigid, step-by-step procedures, declarative modelling focuses on what must occur by defining relationships, preconditions, and effects (Governatori, 2018). This allows regulatory requirements to be expressed without prematurely committing to a fixed resolution strategy or technical details.

Linked Data ecosystems complement this approach by supporting norm sharing and knowledge aggregation across institutional boundaries, enabling compliance mechanisms to remain interpretable, adaptable, and aligned with broader governance structures (Poblet et al., 2019).

Structural mismatches in regulatory automation

Despite increasing digitalisation, regulatory workflows in the AEC sector remain difficult to automate in a traceable and sustainable manner. This is not primarily a data availability problem but a structural mismatch between regulatory intent, information representation, and technical execution.

Regulatory intent is expressed through natural-language instruments designed to establish legal authority and support professional interpretation. In technical domains, this intent is commonly conveyed through condensed, presentation-oriented formats—such as tables, annexes, and figures—that facilitate communication while remaining aligned with legal drafting practices. Empirical evidence highlights the limits of treating these representations as directly formalisable: (Hettiarachchi et al., 2025) show that, from an initial set of 20 674 sentences in English and Finnish building regulations, only 6% could be annotated as self-contained rules suitable for machine-readable formalisation. This observation is consistent with legal informatics research, which distinguishes between sources of authority and documentary resources, noting that regulatory meaning is not fully embedded in such artefacts and must be reconstructed through interpretive knowledge-acquisition processes (Casanovas, 2017).

Empirical analysis of design-stage EPCs in Estonia shows how this complex interpretation process propagates into methodological inconsistencies and mismatches between declared design solutions, assumed input parameters, and practical outcomes (Aljas et al., 2025).

EPBD provides an example of how fixed technical thresholds are combined with heuristic judgement, spanning multiple knowledge domains:

"Where technically and economically feasible, non-residential buildings with an effective rated output for heating systems, air-conditioning systems, systems for combined space heating and ventilation, or systems for combined air conditioning and ventilation of over 290 kW, must be equipped with building automation and control systems." - (European Parliament, 2024).

This hybrid formulation leaves essential aspects of compliance underspecified, including actor responsibilities, data sources, and the sequencing of decisions and practical actions.

Current digital tools primarily focus on operationalising these static methodologies by automating calculations, enforcing parameter constraints, and guiding data entry and reporting. While this improves usability and administrative efficiency, it also embeds a single, implicit interpretation of the regulation within individual software systems, resulting in tightly coupled implementations with limited transparency, adaptability, and traceability (buildingSMART International, 2023; Sayfikar and Jenkins, 2025).

Semantic approaches, including Linked Building Data, Building Information Graphs, regulatory-focused initiatives such as Regulatory Information Requirements, and building permitting-oriented ontologies such as OBPA or AEC3PO, contribute valuable partial formalisations of regulatory-relevant data, structures, and artefacts (buildingSMART International, 2023; Dridi et al., 2024; Hagedorn et al., 2025; Lygerakis et al., 2022; Wang and Sacks, 2025). However, these approaches do not explicitly model the procedural aspect that translates regulatory intent into executable workflows. As a result, compliance logic is either applied post-hoc through validators (Compliance by Detection) or embedded imperatively into software (rigid Legal Compliance by Design), which are commonly implemented across fragmented or siloed toolchains, with associated maintenance and coordination costs reported in the literature (buildingSMART International, 2023; Governatori, 2018).

At the workflow level, these fragmented implementations produce overlapping responsibilities and disconnected steps, imposing manual coordination and repeated data entry on users, while obliging software developers to replicate procedural logic and manage fragile integrations across tools. Such challenges persist even where extensive standardisation or semantic modelling is applied, reflecting well-documented institutional, societal, and

technological heterogeneity in regulatory practice (Carvalho et al., 2024; Jenkins et al., 2024).

Taken together, these observations reveal a procedural blind spot: the absence of an explicit, declarative representation of how regulatory intent is translated into executable workflows. Addressing this gap requires an architectural separation that preserves interpretive logic, supports traceability across abstraction levels, and enables adaptable automation.

The decoupled, goal-driven architecture

To address the procedural blind spot identified in regulatory automation, this work proposes a decoupled conceptual architecture that restructures regulatory workflows around explicit procedural representation. The architecture separates regulatory intent, logical reasoning, technical execution, and agentic interaction into distinct but linked layers, as shown in Figure 1.

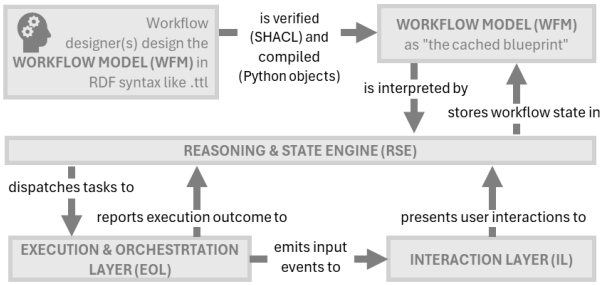


Figure 1: The 4-layer decoupled architecture

The workflow model: a design-time semantic blueprint

At the core of the architecture is a semantic Workflow Model (WFM) that serves as a design-time blueprint defining how regulatory completion can be achieved. The model is goal-driven: it defines the space of possible pathways through which explicit completion goals can be satisfied. Rather than prescribing a fixed execution sequence, the model specifies the set of steps, logic, and information dependencies that lead to these goals.

These elements are modelled as Resource Descriptive Framework (RDF) objects in a knowledge graph, making execution-relevant structure and meaning explicit through the use of two fundamental abstractions – Activities and Entities (*core layer* in Figure 2). Entities represent planned data items, which Activities can declare as inputs or outputs, forming a directed dependency structure that guides the execution, as exemplified in Figure 3. We modeled this core using a domain-independent semantic layer based primarily on the P-PLAN ontology.

The core model is complemented by our custom lightweight control vocabulary (*control layer* in Figure 2), which allows completion criteria and other regulatory logic to be explicitly represented within the Workflow Model as first-class, i.e., regulatory constraints and decision points are subclasses of as Activities and Entities.

Where beneficial, the core model can be extended with domain-specific knowledge graphs, such as Linked Building Data (LBD) vocabularies or local regulatory reference data. The semantic depth and operational specificity (as noted in Figure 2) of the WFM is thus configurable: different deployments can provide only the level of detail that execution handlers and users can interpret. Increased semantic depth enables tighter alignment between domain data and regulatory logic, allowing constraints and dependencies to be expressed explicitly where needed, but without requiring uniform detail across all deployments.

Scope of activities and hybrid control

The notion of an Activity is intentionally broad, reflecting the hybrid socio-technical nature of regulatory workflows. Operationally, Activities may be fully digital—such as automated data acquisition from APIs, parameterised calculations, consistency checks (e.g. BIM/LBD-aligned post-hoc conformity checks), logic resolution—or physical–digital, where a physical activity or professional judgment is acknowledged and documented through the system. Examples include confirmation of an on-site inspection, upload of supporting documents, or an explicit declaration that a regulatory requirement has been satisfied or acknowledged. Activities also distinguish between steps that can be executed automatically and those that require agentic input, while modeled uniformly within the same dependency structure.

Decoupled architectural layers

The architecture operationalises the WFM through four coordinated layers, following annotations in Figure 1.

The workflow model (WFM), as discussed, provides the design-time blueprint by encoding Activities, Entities, including goals and control points as a declarative semantic knowledge graph. At runtime startup, these semantic definitions are loaded, enriched with lightweight reasoning, and validated against design-time constraints (via Shapes Constraint Language aka SHACL) to ensure model conformity; they are then compiled into cached blueprint structures in two steps (the following terminology is aligned with Figure 2): first, the foundational workflow meta-model is superimposed on the project-layer blueprint instances to generate an in-memory workflow graph—i.e., cached Activity/Entity records with a compiled dependency map. Second, the *design & specialisation vocabulary* is retained as lightweight annotations (semantic types and dispatch configuration) that enable generic routing to execution and interaction handlers. However, *domain vocabulary* and *executable bindings* from the *project layer* are not required for the generic reasoning and orchestration mechanisms, but can be leveraged to implement custom functionalities (e.g., domain-specific functions, queries, or service adapters). In this way, the WFM functions as the authoritative specification of the workflow's structure and meaning, while remaining decoupled from the concrete mechanisms that execute steps, solicit inputs, and perform external tasks.

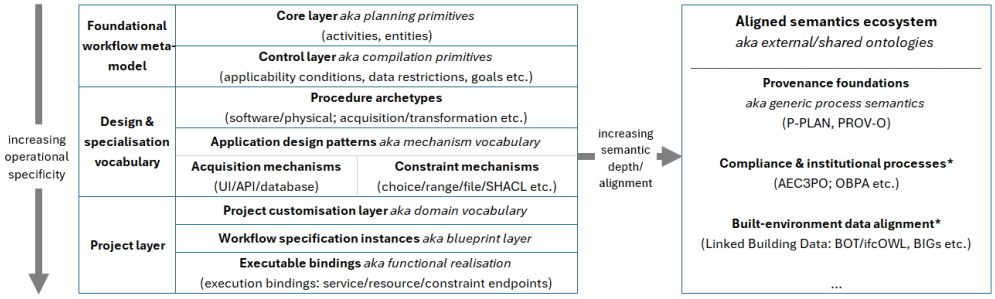


Figure 2: The Workflow model's semantic decomposition. (*) marks frameworks not implemented in the current instantiation.

The reasoning & state engine (RSE) represents the stateful operational structure of a workflow run. It operates over a directed dependency model whose first-class elements are Activities and Entities of the WFM. During execution, the RSE may further extend this structure by materialising latent semantic conditions—such as unresolved alternative production paths or exceptional situations—into explicit control-point Activities (e.g. *:MultipleProducerChoice* on Figure 3) whose outcomes gate subsequent progression. Therefore, all state transitions are exposed as structured events and can be synchronised to provenance records, enabling traceable reconstruction of how outcomes were produced and which decisions governed the execution path.

The execution & orchestration layer (EOL) carries out the Activities that the RSE has determined to be ready for execution. It receives an Activity together with its corresponding semantic type(s) and configuration, routes it to an appropriate execution mechanism via pluggable handlers, and either executes the Activity and commits its outcomes back into the runtime state, or emits an explicit interaction request and transitions the Activity into a waiting state until the Interaction Layer fulfills it. By standardising execution behind handler contracts, the EOL provides a stable execution interface that accommodates both internal functions and external services, while preserving decoupling between workflow semantics and execution mechanisms.

The interaction layer (IL) functions as the unified presentation boundary for the system. It consumes execution-state and interaction events forwarded by the EOL, and renders them through interaction primitives—such as data entry requests, explicit pathway choices, and progress events—so the interface adapts to different workflow configurations. For human experts, the IL can procedurally generate appropriate UI elements like fields, dropdowns, choice dialogs and file uploads. Alternatively, for machine agents, these primitives can be exposed through structured interfaces for programmatic consumption (e.g., an API or an MCP-style server), enabling artificial agents to consume required input/choice requests and submit decisions against the same underlying workflow logic.

Instantiation of the architecture: Estonian EPC workflow

This section presents a proof-of-concept instantiation of the proposed architecture using the Estonian Energy Performance Certificate (EPC) methodology for new buildings, grounded in the EPBD framework and ISO 52000-series standards (Minister of Economic Affairs and Infrastructure of Estonia, 2023; Dijk and Hogeling, 2019). The methodology requires prediction of primary energy use based on typical building energy modelling components such as envelope characteristics, HVAC systems, ventilation rates, infiltration rate, component efficiencies, lighting, renewable energy contributions etc. It allows using defaults from national tables or documented values from design/measurements, with professional validation or additional proof required where substitutions occur.

The implementation focuses on modelling local regulatory *domain vocabulary*, procedural dependencies, and decision logic, rather than on full numerical coverage or exhaustive system integrations. In particular, the current scope emphasises conditioning of calculations, alternative compliance pathways, and explicit control points that govern how EPC calculations may be completed at building design stage.

While the implementation covers all four layers of the architecture, the WFM's *operational specificity* (vertical arrow on Figure 2) is currently limited to non-functional placeholders for integrable data sources (e.g. *:GetBimModelFromBuildingRegistry* on Figure 3) such as the Estonian National Building Registry (later potentially via the E-ehituse OpenAPI services), or for Semantic BIM pipelines (Land and Spatial Development Board of Estonia, 2026; Pauwels and Terkaj, 2016; Wang and Sacks, 2025), with parallel options to explicitly re-declare such data entities (e.g. *:UploadBimModel*). The result is a dynamically conditioned, form-like execution environment composed of dropdowns, numerical inputs, checkboxes, acknowledgements, and calculation steps.

At runtime, available fields, selectable options, and required acknowledgements are progressively introduced or constrained as upstream Entities become resolved. Early building-level declarations—such as building category, size class, and primary construction type—function as conditioning inputs that determine which regulatory requirements apply, which calculation methods are admissible, and which exceptions or simplifications may be used. For example, small buildings may qualify for simplified calculation procedures, certain building categories apply adjusted EPC thresholds (e.g. multiplicative factors for specific construction types), and building type selections govern lookup strategies for default parameters such as ventilation rates.

This conditioning is reflected directly in the user interface. Selections made at higher levels constrain which subsequent inputs are exposed and which options are selectable, ensuring that only methodologically admissible choices are presented. Data restrictions—such as expected data type, allowed value lists, or numeric ranges—are materialised in the front-end as input field configurations (e.g. dropdown enumerations, bounded numerical fields, or categorical selectors).

At points where the Estonian methodology permits alternative means of satisfying a requirement, the workflow introduces explicit procedural choices (*:MultipleProducerChoice*) that resolve how an Entity is grounded. A common pattern is the substitution between normative defaults and documented values. Selecting a default immediately resolves the Entity using standard tables (*:LookupInfiltrationCoeff*), while inputting a custom value (*:InputInfiltrationCoeff)* may introduce additional dependent steps, such as acknowledgements or the provision of supporting material (*:InfiltrationTest-Report*). The method by which the value was obtained is preserved in Entity state and provenance records, allowing downstream logic to access previous compliance-relevant decisions.

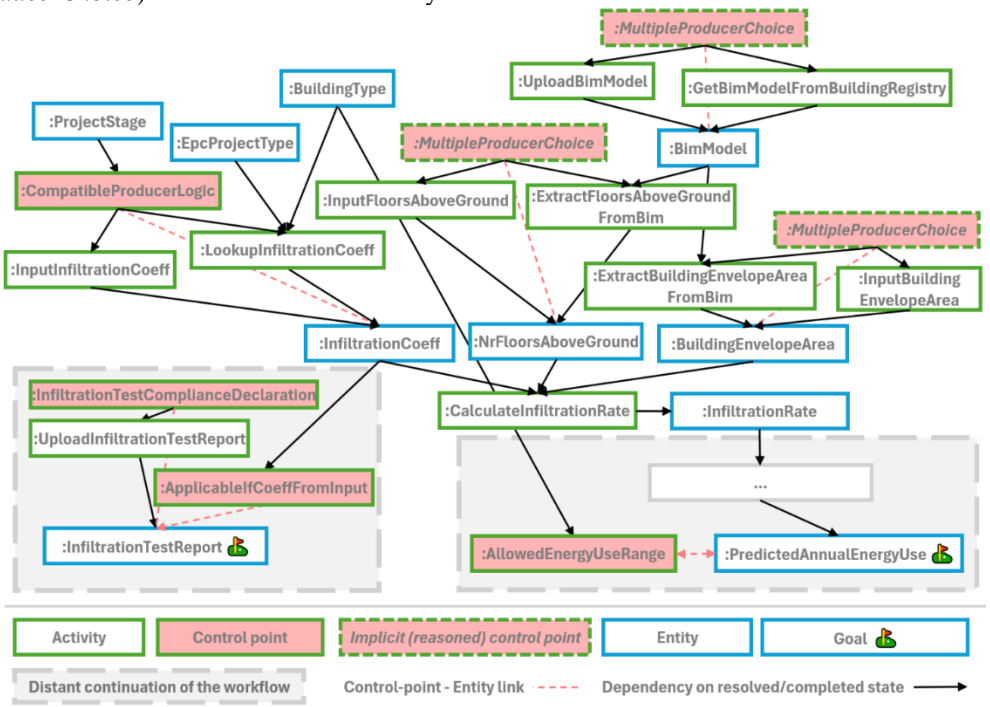


Figure 3: Simplified excerpt of the Estonian EPC Workflow Model showing main structural elements and their relations.

Not all regulatory requirements are digitally verifiable within the current scope. Where compliance depends on professional judgement, certification, or conformance to external standards that cannot be computationally assessed, explicit acknowledgement events are modelled in the Workflow Model and instantiated by the RSE as first-class Activities (*:InfiltrationTestComplianceDecla-ration)*. Although their outcomes are not derived numerically, they are committed to the runtime state and exposed through the Interaction Layer as explicit confirmations, preventing implicit assumptions and preserving traceability of regulatory decisions.

The implementation further addresses regulatory logic that applies across multiple instances of similar elements, such as technical systems or energy carriers. Their compatibility and co-dependence rules are expressed as aggregate conditions over sets of Entity instances in the Workflow Model. This constrains available options consistently across the workflow, preventing incompatible configurations at the point of selection. If upstream conditions change, previously valid choices are invalidated and surfaced through the Interaction Layer with clear visual feedback.

As the graphical interface is not the only possible data entry surface, the same restrictions are also enforced through implicit validation Activities attached directly to data Entities. Values are checked against their declared constraints regardless of how they are introduced, allowing inconsistencies to be detected even when upstream changes invalidate previously accepted inputs. This dual use of semantically declared data rules establishes a single source of truth, while supporting both immediate user-facing guidance and backend consistency validation at different stages of execution.

Where calculations are delegated to external procedures, such as *:CalculateInfiltrationRate*, these are represented as pluggable Activities bound to execution handlers in the EOL. Each such Activity consumes shared input Entities—such as building type (*:BuildingType*) or HVAC parameters—and produces a single declared output Entity, preserving the one-activity–one-output constraint of the Workflow Model. This keeps the dependency graph explicit for the RSE, while allowing external implementations to be introduced or replaced without affecting surrounding procedural logic or interaction behaviour. Figure 4 shows illustrative sections

from the Estonian EPC workflow expert system's UI in its present form (in development).

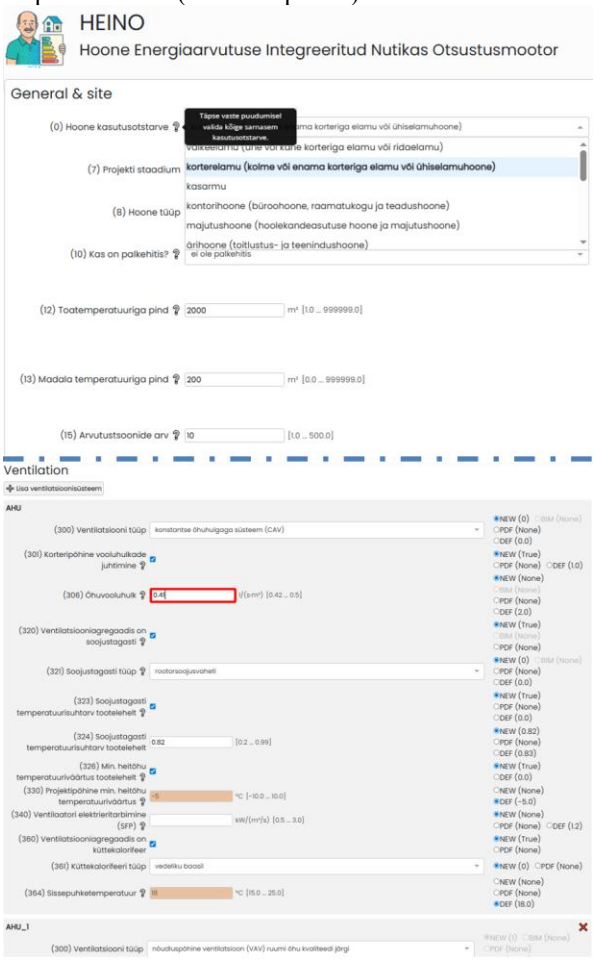


Figure 4: Screenshots from the current Estonian EPC workflow user interface (in development).

Discussion

This work frames regulatory automation as a problem of procedural articulation rather than document processing or isolated rule extraction. From this view, the central challenge is not digitising regulatory content, but coordinating intent, interpretation, and execution as workflows evolve, branch, and interact with external actors and systems. The proposed framework tackles this coordination structurally, though it cannot resolve every tension in regulatory and AEC digitalisation.

A recurring challenge is deciding how much formal structure is practical and desirable. Many regulatory steps involve physical actions, professional judgement, or context that resist full specification. The framework deliberately leaves such elements underspecified, treating them as obligations to acknowledge rather than compute, which prompts the question: how much structure supports traceability and coordination without over-constraining practice? Thus, workflow design remains an agent-guided task, with models reflecting institutional choices on where precision matters and where judgement should stay flexible.

Questions of maintenance and evolution follow naturally. Regulations change, interpretations shift, and data infrastructures mature unevenly. The Workflow Model does not erase these realities; it offers a stable reference point for managing them. Changes still demand agentic oversight and governance, but their effects become more localised and visible—underscoring that long-term viability depends as much on organisational processes and knowledge transfer as on technical design.

The framework delivers greatest value in workflows with strong interdependencies, sophisticated logic, and interaction with semantic or non-tabular data—the complex, evolving processes typical of energy performance certification or building permitting. In contrast, for workflows dominated by independent steps with minimal downstream impact—such as uploading parallel documents, running a fixed linear data pipeline, or handling predominantly physical sequences with little digital conditioning—the semantic modelling effort may yield limited incremental gains. Yet truly dependency-free cases are uncommon in regulated domains, where even seemingly separate or physical actions often carry traceability requirements, conditional applicability, or audit consequences that justify structured coordination.

The framework also highlights the limits of specialised, handcrafted solutions. Richard Sutton's "bitter lesson" from 70 years of AI research offers a useful lens: systems built around general, scalable structures or „meta-methods" ultimately outperform finely tuned domain heuristics, as the „contents of minds are tremendously, irredeemably complex" (Sutton, 2019). In regulatory automation, this suggests that durable progress more likely comes from simple, composable procedural primitives rather than from bespoke software. Early results with our framework—effective despite modest semantic depth—reinforce this view and motivate continued emphasis on foundational abstractions over narrowly optimised implementations.

The design remains intentionally open-ended, prescribing no single execution environment, interaction style, or data standard. This flexibility accommodates diverse implementations but raises challenges in interoperability, semantic alignment, and trust. Our semantic modelling (built on P-PLAN, PROV-O, a custom control vocabulary, and preliminary design & specialisation elements, as outlined in Figure 2) is still evolving; more sophisticated ontology alignment and refinement will appear in forthcoming work. Establishing and functionalising shared conventions through ontologies and data dictionaries remains an active area of research (buildingSMART International, 2023; Hjelseth et al., 2025; Utkucu and Sacks, 2025; Wang and Sacks, 2025). While the architecture supports incremental semantic enrichment, determining when and how to invest in deeper alignment is as much a policy decision as a technical one.

We see many directions demanding further exploration: real-world testing of long-running workflows to refine error handling and richer user-facing elements; semantic integration of permitting/compliance ontologies, including overlap resolution and cross-system

governance; and direct RDF graph interaction (designing/querying workflows) via natural-language and augmented retrieval tools (e.g., GraphRAG) to reduce technical overhead for non-experts. Future work is also expected to provide insights into the dilemma of deeper semantic standardisation vs reliance on self-contained tools or services, and explore semantic modelling strategies beyond the core layer that can most optimally cater for practical needs across the 4-layer architecture.

Ultimately, the framework is best viewed as scaffolding for ongoing exploration rather than a packaged product. By making procedural structure explicit, shareable, and inspectable while preserving institutional flexibility and technological diversity, it addresses—rather than ignores—the fragmented, dynamic, digital-physical realities of regulatory contexts. In such environments, this openness is likely essential for automation that proves meaningful and sustainable.

Conclusion

This paper addressed a persistent gap in the digitalisation of regulatory workflows: the absence of a stable structure for coordinating intent, interpretation, and execution across diverse actors, tools, and lifecycle stages. In fields like AEC, where compliance involves conditional procedures, partial automation, and professional judgement, existing approaches often fragment regulatory logic across documents, custom software, and informal practices—undermining traceability, adaptability, and sustainability.

To tackle this, the work introduced a declarative workflow model that reframes compliance as a procedural knowledge problem rather than a document- or rule-processing task. By encoding activities, entities, dependencies, and control points as semantic structures, the model defines what must be achieved and under what conditions—without dictating how steps are executed. This explicit, inspectable representation preserves regulatory meaning while keeping execution, interaction, and domain implementations flexible.

The four-layer architecture operationalises this model by decoupling semantic specification, state reasoning, execution orchestration, and interaction. The separation is practical: regulatory interpretation can evolve independently of software, execution mechanisms can change without rewriting semantics, and agentic judgement can be acknowledged as a first-class element rather than hidden or flattened into technical artefacts. The framework does not demand full formalisation—ambiguity, physical actions, and discretion are treated as legitimate procedural features, not automation failures.

Taken together, the contribution is not a finished system but a conceptual backbone for curating complex compliance processes. It enables incremental semantic enrichment, cross-domain coordination, and clearer accountability while remaining open to institutional practices and technological evolution. By prioritising simple, composable procedural abstractions over bespoke solutions, the work aligns with a broader shift toward scalable, general-purpose digital structures—offering a foundation for future research at the intersection of regulation, semantic technologies, and human–AI collaboration, where meaningful progress depends on making interpretation explicit, governable, and traceable rather than treating it as isolated engineering problems.

Acknowledgements

The authors acknowledge the use of large language models—including Grok 4.1 (Fast, Expert, Thinking modes), ChatGPT 5.1 & 5.2, and Google Gemini 3 (Pro & Flash)—used independently, and also within the Cursor software environment. These tools assisted in synthesising and communicating authors' original ideas and notes, referenced literature, codebase, and drafts.

All ideas, arguments, and conclusions remain the original work of the named authors. The text underwent rigorous manual review and iterative rewriting to ensure precise alignment with the authors' intent, cited sources, and scholarly standards, leaving no sloppy logic, over-interpretations or hallucinations. The models served only as aids for expression and refinement; full responsibility for content rests with the authors.

This work has been supported by the Estonian Centre of Excellence in Energy Efficiency, ENER (grant TK230) funded by the Estonian Ministry of Education and Research, European Commission through the European Union's Horizon Europe research and innovation program under grant agreement no. 101192930 (the ENTRANCE Project), and by the Estonian Research Council through the grant PSG963. This study was also co-funded by the European Union and Estonian Research Council via project TEM-TA78.

Abbreviations

AEC – Architecture, Engineering and Construction industry
AEC3PO – Architecture Engineering and Construction Compliance Checking and Permitting Ontology (Dridi et al., 2024)
BIGs – Building Information Graphs (Wang and Sacks, 2025)
BIM – Building Information Modelling
CtD – Compliance through Design (Casanovas, 2017)
EOL – Execution and Orchestration Layer (in proposed architecture)
EPBD – Energy Performance of Buildings Directive in European Union
EPC – Energy Performance Certificate
HVAC – Heating, Ventilation and Air Conditioning
IL – Interaction Layer (in proposed architecture)
LBD – Linked Building Data
MCP – Model Context Protocol by Anthropic
OBPA – Ontology for Building Permit Authorities (Hagedorn et al., 2025)
RDF – Resource Description Framework
RSE – Reasoning and State Engine (in proposed architecture)
WFM – Workflow Model (in proposed architecture)

References

Aljas, H.K., Sempelson, K.-S., Pikas, E., Thalfeldt, M., 2025. Framing Needs and Opportunities for Automating Energy Performance Certificates Validation. Presented at the 15th REHVA HVAC World Congress (CLIMA 2025), Milan, Italy.

buildingSMART International, 2023. Regulatory Information Requirements: Project status and findings from the perspective of RAVA3pro project (No. RR-2023-0001-PR). buildingSMART International.

Carvalho, C., Cleto, J., Monteiro, C., Fernandes, J., Fragoso, R., 2024. Energy Performance Certificates: policy needs and best practices | BUILD UP. European Commission.

Casanovas, P., 2017. Legal Compliance by Design (LCbD) and through Design (LCtD): Preliminary Survey.

Dijk, D. van, Hogeling, J., 2019. The new EN ISO 52000 family of standards to assess the energy performance of buildings put in practice. E3S Web Conf. 111, 04047. https://doi.org/10.1051/e3sconf/201911104047

Dridi, A., Patlakas, P., Lefrancois, M., Beach, T., Petr, H., Vakaj, E., 2024. Aec3po: A Knowledge Model for Machine Executable Construction Regulations. https://doi.org/10.2139/ssrn.4784123

European Parliament, 2024. Directive (EU) 2024/1275 of the European Parliament and of the Council of 24 April 2024 on the energy performance of buildings (recast) (Text with EEA relevance).

Feng, L., Wang, Z., 2025. Developing Proactive Compliance Mechanisms for Chinese International Construction Contractors: A PLS-SEM Analysis. Buildings 15, 1478. https://doi.org/10.3390/buildings15091478

Governatori, G., 2018. Declarative Approaches for Compliance by Design. Springer EBooks. https://doi.org/10.1007/978-3-319-76587-7_6

Hagedorn, P., Fauth, J., Zentgraf, S., Seiß, S., König, M., Brilakis, I., 2025. OntoBPR: An ontology-based framework for performing building permit reviews using standardized information containers. Adv. Eng. Inform. 66, 103369. https://doi.org/10.1016/j.aei.2025.103369

Hashmi, M., Koker, L. de, Casanovas, P., 2018. Legal compliance through design: preliminary results of a literature survey. CEUR Workshop Proc. 59–72.

Hettiarachchi, H., Dridi, A., Gaber, M.M., Parsafard, P., Bocaneala, N., Breitenfelder, K., Costa, G., Hedblom, M., Juganaru-Mathieu, M., Mecharnia, T., Park, S., Tan, H., Tawil, A.-R.H., Vakaj, E., 2025. CODE-ACCORD: A Corpus of building regulatory data for rule generation towards automatic compliance checking. Sci. Data 12, 170. https://doi.org/10.1038/s41597-024-04320-x

Hjelseth, E., Labonnote, N., Luczkowski, M., Teclaw, W., 2025. Enabling a fully integrated BIM environment using Semantics and Artificial Intelligence.

Jenkins, D.P., Sayfikar, M., Gomez, A., Fueyo, N., 2024. A Comparative Study of Energy Performance Certificates across Europe. Buildings 14, 2906. https://doi.org/10.3390/buildings14092906

Land and Spatial Development Board of Estonia, 2026. E-ehituse API services in Openapi format [WWW Document]. URL https://swaggerui.ehr.ee/3_d_kaksiku_api (accessed 1.28.26).

Lygerakis, F., Kampelis, N., Kolokotsa, D., 2022. Knowledge Graphs' Ontologies and Applications for Energy Efficiency in Buildings: A Review. Energies 15, 7520. https://doi.org/10.3390/en15207520

Minister of Economic Affairs and Infrastructure of Estonia, 2023. Methodology for Calculating Building Energy Efficiency (Hoone energiatõhususe arvutamise metoodika).

Pauwels, P., Terkaj, W., 2016. EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. Autom. Constr. 63, 100–133. https://doi.org/10.1016/j.autcon.2015.12.003

Poblet, M., Casanovas, P., Rodríguez-Doncel, V., 2019. Linked Democracy: Foundations, Tools, and Applications, SpringerBriefs in Law. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-13363-4

Sayfikar, M., Jenkins, D.P., 2025. Energy performance certificate calculation methodologies across Europe and accommodating new performance indicators. Build. Serv. Eng. Res. Technol. 46, 45–59. https://doi.org/10.1177/01436244241282076

Schaffer, M., Marszal-Pomianowska, A., Wittchen, K.B., Kragh, J., Pomianowski, M.Z., Provera, A., 2026. Assessing Energy Performance Certificate Data over Time: Evidence from Denmark. Proc. 15th REHVA HVAC World Congr. - CLIMA 2025, Lecture Notes in Civil Engineering 2. https://doi.org/10.1007/978-3-032-06810-1_116

Sutton, R., 2019. The Bitter Lesson. incompleteideas.net. URL http://www.incompleteideas.net/IncIdeas/BitterLesson.html (accessed 2.7.26).

Utkucu, D., Sacks, R., 2025. Ontology for holistic building performance modelling and analysis. Autom. Constr. 175, 106197. https://doi.org/10.1016/j.autcon.2025.106197

Wang, Z., Sacks, R., 2025. Building Information Graphs (BIGs): remodelling building information for learning and applications. Data-Centric Eng. 6. https://doi.org/10.1017/dce.2025.10024

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

OPERATIONALISING HEALTH AND SAFETY REGULATIONS THROUGH DESIGN SCIENCE RESEARCH

Nicholas Nisbet and Zijing Zhang
University College London, London, UK,

Abstract

Investigations into compliance checking have paid little attention to how process-oriented regulations can be made accessible and operable. This paper explores ways that these expectations can be met. The study considers the macro level requirements enforced by a government agency, as exemplified by the UK Construction (Design and Management) Regulations (CDM 2015) which assigns responsibilities and duties clients, designers, and contractors, many of whom lack confidence in fulfilling these obligations or in evidencing their discharge. The findings highlight an opportunity for guidance that is firmly grounded in regulatory requirements yet tailored to the distinct roles of industry participants.

Introduction

Compliance checking has usually been associated with the systematic comparison of regulations against a building proposal, typically structured building models (BIM) (Eastman et al., 2009; Nawari, 2012). Previous work has shown how the semantic mark-up of Requirements, Applicability, Selection and Exceptions (RASE) can support compliance checking, whether applied to design models (BIM) (Nisbet et al., 2023) or to site monitoring information (Nisbet et al., 2024). However, regulatory obligations that prescribe required behaviours, rather than physical or spatial performance, are qualitatively different. While the built-environment literature has devoted substantial attention to product-based regulation, far less attention has been paid to the regulation of processes, even though process-based duties are often invoked when product information is incomplete or unavailable. Health and safety regulation exemplifies this challenge: although trailing indicators such as incidents and near-misses are well understood, there are few leading indicators that can be assessed in advance (Zhang et al., 2024a). This paper therefore examines a situation in which evidence of compliance is unlikely to exist in either structured or unstructured form, and where the discharge of continuing obligations depends on duty holders who are often uncertain about what those obligations entail.

Highlighting the logical semantics of a regulation (its hierarchy of objectives, associated duties and specific metrics) offers a way to investigate the structures and behavioural patterns implied by the text (Nisbet et al., 2023). By systematically seeking hypernyms for the objects and properties named in the regulation, patterns of expected behaviour emerge, illuminating the intent and making the structures in the regulation explicit. These insights extend beyond checklists or decision trees and point toward the possibility of generating interactive teaching and learning materials that help embed compliant behaviour. In the current example, the outcome of the systematic analysis can be compared to the informal 'manage, maintain and improve, document' model used by UK Health and Safety Exective (HSE) HSE Inspectors.

Context and literature review

Regulatory context and application of CDM 2015

The Construction (Design and Management) Regulations 2015 (CDM 2015) represent the third iteration of the UK's CDM regulations (HSE, 2015a), introduced to align with the EU Directive on Temporary or Mobile Construction Sites (EU, 1992). Although construction fatalities have decreased since the first CDM regulations were introduced, the construction sector continues to experience the highest number of work-related deaths (HSE, 2025). These persistent health and safety (H&S) challenges prompted the revisions of the CDM regulations in 2007 and 2015, aiming to address shortcomings in coordination, clarity of duties, with the intention of improving H&S outcomes.

CDM 2015 places strong emphasis on accountability. The regulations are structured around five key duty holders: client, principal designer (PD), principal contractor (PC), designer and contractor. A key idea of the 2015 revision was to highlight the influence of design decisions on construction H&S, placing greater responsibility on designers and PDs to identify and manage risks early in the project lifecycle. This reflects a broader shift toward improving H&S performance through more effective design risk management, also referred to as Construction Hazards prevention through design (CHPtD) (Toole & Gambatese, 2008). However, despite this improved clarity of responsibility, the CDM 2015 regulations remain process focused and outcome oriented, offering limited detail on the specific procedures required to achieve compliance (Zhang et al., 2024a). As a result, practitioners often struggle to interpret what compliance looks like in practice, particularly during the design stage.

To address these gaps, significant effort has been made to interpret and operationalise CDM 2015. The HSE's official guidance (L153) provides definitions, clarifications and examples of what constitutes compliance, including expectations for key documents such as the health and safety file and the construction phase plan (HSE, 2015b). The HSE has also been delivering the "Discovering Safety" programme, which includes a range of projects focused on improving health and safety in construction. This has involved several 'SafetyTech' regulatory sandbox initiatives that bring in industry start-ups to trial innovative solutions and develop case studies supporting CDM compliance and enhanced H&S outcomes and construction risk library which mapped out risk scenarios and mitigation measures.

A range of industry and standards bodies have also produced supplementary guidance to support practitioners. For example, the Institution of Civil Engineers (ICE) has published guidance on Design Risk Management (DRM), a core requirement under the CDM regulations (ICE, 2020). The document provides clear explanations of risk concepts, the fundamentals of DRM, and the respective responsibilities of different stakeholders throughout the process. The Construction Industry Training Board (CITB) has published six guides, five for duty holders and one for workers. The duty-holder guides outline the key tasks, the information required or to be provided, practical suggestions of how responsibilities can be carried out, and guidance on applying these duties when working with domestic clients. The worker guide explains what the regulations mean on site, emphasises consultation and communication, sets out the information workers should give and receive, highlights the need for cooperation and coordination, clarifies expectations around competence and training, and illustrates what good health and safety practice looks like in construction. Despite the comprehensive list of actions, the guides do not always distinguish between what is essential and what is desirable; applying CDM requires judging what is reasonably practicable in context, and not all of the listed measures are proportionate or necessary on every project. In addition, without some example case studies from real life projects, the list can feel theoretical. In addition, national and international standards such as BSI PAS 1192-6 (BSI, 2018) and ISO 19650-6 (ISO, 2025)focus specifically on health and safety information management, offering guidance on how hazard and risk information can be structured, shared, and maintained throughout a project. Both standards promote treating risk information and mitigation measures as shared project resources that can later be generalised into organisational learning. These two standards have inspired the development of a Construction Risk Library (CRL) (Osorio-Sandoval et al, 2023), as an open resource to guide the identification of design risks and the selection of appropriate treatments.

Along from these general guidelines, there are also efforts to clarify key concepts such as "so far as is reasonably practicable" (SFARP). The ICE published a commentary for SFARP, which outlines the legal summary and guidance and a compliant design approach (ICE, 2010). However, it lacks concrete examples and only touched on high-level evidence for a compliant design approach. In addition, the Construction Industry Research and Information Association (CIRIA) attempted to unpack SFARP for designers by explaining how to balance risk, proportionality and practicability in design decisions, emphasising cautious, team-based judgement and the need to eliminate or reduce foreseeable risks through the design process (CIRIA, 2015).

CDM in practice

More recent research has moved beyond interpreting the CDM regulations themselves to examining how they shape practice. Mzyece et al. (2019) approached CDM regulatory compliance from a BIM-integrated perspective, proposing a BIM–CDM interoperability framework to support the discharge of CDM duties. Their framework focused on three key CDM documents (pre-construction information, construction phase plan and H&S file) and explored how BIM and common data environment could support managing these documents. An ICE review conducted three years after the regulations came into force found that, although the industry had broadly adapted to the new framework, many practitioners continued to struggle with ambiguous clauses and terminology (ICE, 2019). The report highlighted persistent uncertainty around how to interpret certain duties, particularly at the design stage. This is where misunderstanding is most evident: many designers still under apply design risk management principles, especially the requirement to eliminate, reduce and control risks through design. Responding to the outcome-oriented nature of CDM 2015, Zhang et al. (2024b) developed a process map of the H&S by design process based on interview results with practitioners, providing a practical and detailed workflow to support designers and PD to fulfil their duties. Also focusing on the design stage, Hare et al. (2020) developed a multimedia digital tool to help designers identify and mitigate common design-related hazards, offering practical support for design-stage CDM compliance. Adaku et al. (2021) focused on clients, identifying seven key practices they use to achieve CDM compliance. The findings highlighted that some these practices depend on the PD's arrangements, underscoring the importance of understanding the PD's role in coordinating pre-construction H&S. Shifting the focus to the construction stage, Nisbet et al. (2024) used a semantic annotation approach (RASE mark-ups) to support live CDM compliance checking on construction sites, integrating site diary entries and site-monitoring data to evaluate compliance in real time.

Industry working groups have also played a significant role in supporting CDM compliance on projects through collective knowledge sharing. Workshops that bring multiple duty holders (e.g., clients, PDs) were organised and lessons learned documents and best practice guidance were published addressing specific aspects of CDM

compliance. Notable examples include the BIM for Health and Safety (BIM 4 H&S) group, chaired by the HSE, and the Highways Safety Hub, underpinned by the National Highways' Principal Designer Working Group (Highways Safety Hub, 2023). Together, these initiatives provide a forum for cross-organisational learning and help translate regulatory requirements into practical approaches for improving design risk management.

The principal designer role

The studies above highlight both the importance of the Principal Designer role and the persistent uncertainty surrounding its practical interpretation. Introduced in CDM 2015 to replace the CDM Coordinator, the PD was intended to lead and coordinate pre-construction H&S matters. This shift reflected the HSE's aim to embed H&S considerations at the centre of the design process by placing responsibility with the lead designer rather than peripheral H&S consultants (Jolly, 2023). However, architects and lead designers have often been reluctant to take on the PD role, with findings indicating that many still view H&S as a contractor responsibility, resulting in the role being frequently undertaken by H&S consultants, client advisors and project managers (MPW R&R Ltd, 2023).

In addition, a key requirement of CDM 2015 is that PDs must be competent, yet the regulations offer little clarity on what this competence entails in practice. Addressing this gap, Adaku et al. (2022) developed conceptual models for assessing PDs' skills, knowledge and experience, expanding CDM competence across multiple levels, such as occupational and task-specific dimensions. Ndekugri et al. (2022) investigated the coordination role of the PD during the pre-construction stage. They found that while CDM 2015 has improved early PD appointment, significant challenges remain around assessing PD competence, ensuring adequate resourcing, and overcoming barriers such as limited awareness, misunderstandings of duties and intentional non-compliance. These issues have direct implications for how design risks are identified, coordinated and managed, which is the focus of the next section.

Design risk management

Design risk management is a core process undertaken by designers and the PD. It is an iterative process focused on identifying and managing risks during the design stage and forms an integral component of both overall project risk management and routine design practice (ICE, 2020). The central principle is that an inherently healthier and safer design can improve H&S performance across the entire project lifecycle, benefiting both construction workers and end users. Design risk management is guided by the principles of prevention, most commonly expressed through the "ERIC" hierarchy: eliminate, reduce, inform and control (HSE, 2015b). This hierarchy prioritises eliminating risks through design or design changes wherever reasonably practicable, with control measures such as reliance on personal protective equipment (e.g., using earplugs or protective gloves) considered only as a last resort. Although construction projects may involve hundreds or even thousands of potential hazards and risks, design risk management puts more emphasis on a smaller subset of significant, unusual or difficult-to-manage risks (CIRIA, 2015; ICE, 2019). This focus highlights the need for designers to address these risks proactively and early in the design process to achieve effective H&S outcomes. Design risk management, despite its focus on the design stage, also emphasises the need for residual risks to be appropriately managed and handed over to those who will control the subsequent stages of the project, typically the PC (or the contractor where no PC is appointed) (ICE, 2019).

However, despite the aspirations set out in CDM 2015, studies consistently show that designers and PDs often struggle to apply these principles consistently in practice, particularly when interpreting what constitutes a 'significant' risk or determining how far mitigations are reasonably practicable (ICE, 2019). This persistent challenge reflects a broader pattern identified across the literature, where understanding and operationalising CDM duties remains inconsistent, especially during the design stage.

Research gap

Despite extensive guidance and a decade of implementation, the literature consistently shows the varied levels of understanding and ongoing challenges in fulfilling CDM duties, particularly in relation to design risk management. Practitioners in industry still lack operable, practice-ready knowledge on how to fulfil CDM duties at the design stage. In particular, the PD role and design risk management processes remain poorly understood. This requires an iterative investigation framed to answer whether the absence of an operable target (analogous to BIM or site logs) invalidates the use of semantic annotation and whether gap between legislative text and operable knowledge can be closed by other means.

This paper addresses this gap by using semantic annotation (mark-up) to expose the logical intent of the clauses relating to the PD and then classifying those actions by examining the operative verbs. This ensures that the outcome of four classes of action is learnable and teachable whilst remaining evidently grounded in the legislation.

Research methodology

This paper is presenting a case study based on the existing UK regulations around health and safety, examined from the point of view of one of the nominated 'duty holders', namely the 'Principal Designer' (PD). Design Science Research (DSR) (Hevner et al., 2004; Peffers et al., 2007) allows a progressive sequence of iterative steps to be taken to move from the regulation to a concise summary of the duties, meeting the criteria of arriving at a learnable and teachable understanding. Figure 1 shows a DSR dashboard based on Takeda et al (1990). The knowledge base is UK legislation and regulations, and the practice context is the uncertainty around CDM 2015 duties. The problem is to produce accessible guidance with quality

around learnability and teachability, as a grounded presentation of CDM 2015. The iterative research process needs to test for clarity and simplicity.

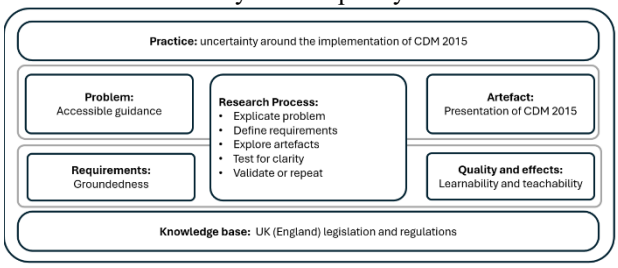


Figure 1: DSR approach (based on Takeda et al 1990)

Investigation

The investigation began with applying the RASE annotation (mark-up) methodology to the regulation. This involves identifying the hierarchy of objectives as boxing which in this case is mostly identical with the sectional hierarchy of clauses and sub-clauses. In one case a clause '8(1)' acts as a sub-clause of another '8(2)'. Mark-up is also applied to the terms that identify any testable proposition, such as the 'role' being taken, using underlining. These metrics can be annotated with their implied property and target. Some propositions may require further discussion such as 'reasonably practical' but in principle these can be answered. All the objectives and metrics are then assigned a logical role, chosen from 'Requirement' (blue), 'Application' (green), 'Selection' (purple) or 'Exception' (orange), indicating how they contribute to the evaluation of their immediate containing section. This mark-up is sufficient to make the document machine operable. For example, by specifying that the actor of interest for this case study is the 'Principal Designer', the document is immediately shortened to reflect only the clauses that refer to that actor. Figure 2 shows the mark-up following this choice.

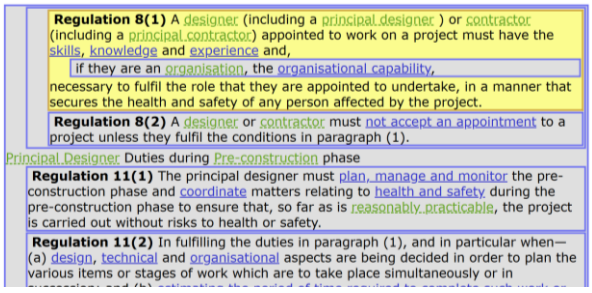


Figure 2: CDM 2015 with RASE mark-up (extract)

However, this still contained a wide variety of tests. Whilst these can be answered individually, given our aim of making the regulation more comprehensible, we can apply a systematic approach to identify not just the test such as '(has) skills' but to group this test with others such as '(has) knowledge' and '(has relevant) experience'. By consulting an online dictionary service (such as Wordnet or WordHippo) and or informally, the concept of 'skill' can be found to have the hypernym 'capability'. Capability is also a good hypernym for 'knowledge' and 'experience'. The metric '(has relevant) skills' can be revised to be 'property=capability comparator=includes target=skills'. Table 1 shows five examples.

Table 1: Regulation terms and hypernyms (extract)

Terms	Hypernym
accept, manage, monitor, coordinate	consider
plan, comply	specify
inform, liaise, ensure, cooperate, assist, provide to	communicate with
report, review	share
reasonably practical, relevant, foreseeable, appropriate	scoping

Having performed this substitution, the number of different properties mentioned in the metrics is considerably reduced. Some comments relating to their evidencing in practice are given in Table 2.

Table 2: Hypernyms with possible evidence (extract)

Hypernym	Possible evidence
consider	memos, design file notes or other records
specify	changes affecting the design and specification of the works
communicate with	confirmation of both sending and receipt.
share	information exchange records

The selection of hypernyms allows us to see a pattern in the kinds of properties involved in the metrics present in the text. Some properties remain categoric, such as whether the PD is an organization (as opposed to an individual). Functional properties tend to be numerical, but there are none in the current example. Lastly, many properties are tests for membership of classes and sets, which may require background knowledge of any hierarchy and synonyms.

Having created a rationalised mark-up for the regulation, a number of presentations can be generated automatically. For example, a conceptual graph can represent the logical operability of the Regulation. Other decision-based presentations are possible including decision trees and flow diagrams. However, these are not shown here because the regulations are not applied at a single point but are intended to be applied continuously.

A presentation derived from the marked-up regulation (Figure 3) summarizes the selected hypernyms as a form allowing the input of values for the properties involved. The text of the regulation is repeated below offering the alternative of the input of values in context. Where the range of possible answers is known, the form offers a drop-down selection. Selecting or providing a value fills in all synonymous slots.

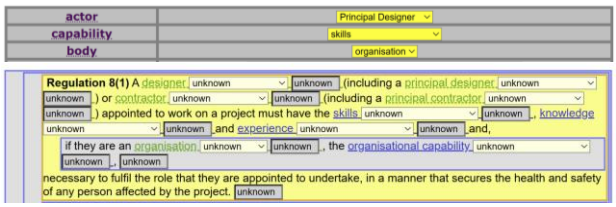


Figure 3: CDM 2015 as linked form and text (extract)

Results

An alternative presentation offers a matrix of the clauses (vertically) and the property hypernyms (horizontally), with each cell including all the target values. Figure 4 shows selected cells from the matrix. This simplifies the PD duties once appointed into a set of four manageable activities: consider/specify/communicate-with/share.

clause	consider	specify	communicate with	share	measure
R11-01	• health and safety	• plan, manage and monitor • coordinate			• reasonably practicable
R11-02	• general principles of prevention • construction phase plan • health and safety file	• design • technical • organisational • estimating the period of time required to complete such work or work stages			• relevant
R11-03	• carrying out or liable to be affected by construction work • maintaining or cleaning a structure • using a structure designed as a workplace	• risks to the health or safety			• reasonably practicable • foreseeable
R11-04	• duties in regulation 9		• designers		
R11-05		• cooperate	• client • each other		
R11-06a			• assist the client	• pre-construction information	
R11-06b			• every designer and contractor appointed, or being considered for appointment	• pre-construction information	• control
R11-07			• liaise with the principal contractor	• planning, management and monitoring of the construction phase • coordination of health and safety matters during the construction phase	

Figure 4: CDM 2015 as a matrix (extract)

This allows us to highlight the lists of key targets the PD has duties around (see Table 3). Under the duty to 'consider', we can note the inclusion of 'health and safety' and the risk register. The PD is required to 'communicate with' all of the other roles identified in the regulation. The PD is required to 'specify' (by affecting the drawings, specifications and risk registers) design, technical and organizational matters. 'Sharing' of the risk register(s) in the form of the 'pre-construction information', 'construction phase plan' and the 'health and safety file' is required.

Table 3: Hypernyms with topics (extract)

Hypernym	Topics
consider	Health and safety
	General principles of prevention
	Construction phase plan
	Health and safety file
specify	Planning
	Coordination
	Design
	Technical (matters)
	Organizational (matters)
	Time for work (estimate)
	Risks to health and safety
	Cooperation
communicate with	Designers
	Client
	Principal contractor
	All
share	Pre-construction information
	Construction phase plan
	Health and safety file
scoping	Reasonably practical
	Relevant
	Appropriate
	Foreseeable

Since these duties are open-ended in terms of their scope and in terms of their duration, it is important to note the qualifications offered for several of these duties. Table 3 summarizes some scoping filters applied to these duties. These apparently ambiguous filters are the subject of specific legal and regulatory guidance. Phrases such as 'reasonably practical' have specific meanings. These frame compliance in terms of normal practice, best practice and other considerations of proportionality. The recently developed Uniclass 'RK' risk classification table (NBS, 2026) and the publication of a 'Construction Risk Library' (CRL) may inform this judgement. The Uniclass table creates a reference list of many types of risk with RK_40 focused on risks to health and safety. This CRL library provides lists of common construction risks and possible treatments and mitigations. These two resources may help judge relevance and appropriateness. The publication of BSI PAS1192 Part 6 and its subsequent development into ISO 19650-6 has explored how the risk register can be treated as a single shared resource, through all three versions in the UK regulations 'Pre-construction information', Construction Phase Plan' and 'Health and safety file' (in-use). The forthcoming guidance to Part 6 in the Information Management Initiative (IMI, 2026) will suggest the filters (views) applicable to a risk register to obtain each of these documents.

Discussion

Regulatory compliance can be effected through changes to the physical and spatial aspects of the built environment

or through governance over the actors and information involved. Regulating actors can focus on requiring levels of competence including skill, knowledge, and experience. Alternatively it can prescribe programmes of actions that must be carried out. In either case there is unlikely to be direct evidence that can be checked systematically in real time. Instead, it becomes important that the expectations are embedded in the level of understanding of the actors so as to inform their behaviour continuously. A simple but justifiable model can be trained and learnt. The technique will be presented to a focus group later this year which may lead to the deployment of the learning/teaching materials. This should produce the evidence to support the validation of the approach.

UK CDM regulations are an example of secondary legislation that imposes expectations on process, not product, in the expectation that the desired outcomes will be achieved. The requirements are open-ended in the sense of not specify the timing of actions and in the sense of no limitation to the amount of effort and output required but for the limiting phrases. This makes the legislation daunting for practitioners and hard for third parties such as insurers to evaluate.

RASE mark-up has been used to expose the objectives and metrics used in defining one actors duties. A repeatable process has then been used to extract the verb-noun predicates. The verbs have been rationalised through the use of hypernyms and the nouns (and scoping adverbs) have been rationalised by reference to emerging classification and information management principles. The same process can be applied to any regulatory document in any language. Exceptionally, if no dictionary of hypernyms is available for a particular language, then manual intervention will be required. Filtering for any particular role, stage or duty or any other attribute is non-specific and is a feature of the generic generation of the presentations.

Conclusion and future work

This paper illustrates that benefits from RASE mark-up can exist even when there is not likely to be a target database for systematic compliance checking. It has addressed a knowledge gap by using RASE mark-up to expose the logical intent of the clauses relating to the PD and then classifying those actions by examining the operative verbs to generate a concise conceptual model of the duties imposed by the regulation. This conceptual model can inform teaching and learning material and act as a simple tacit framework to guide practitioners in their work, without having to refer to the legislation repeatedly and without the potential for scope creep induced by anxiety about under-performing the satisfaction of the obligations. This is significant for bringing 'process' based regulations into the realm of operable knowledge.

Future work can look at how online dictionaries and Natural Language Processing (NLP) tools can support Large Language Models (LLM) in automating the addition of RASE mark-up. With this in place learning/teaching material could be rapidly generated from source documents.

References

Adaku, E., Ankrah, N.A. and Ndekugri, I.E., 2021, July. Prevention Through Design: Clients' pre-construction health and safety arrangements. In Advances in Safety Management and Human Performance: Proceedings of the AHFE 2021 Virtual Conferences on Safety Management and Human Factors, and Human Error, Reliability, Resilience, and Performance, July 25-29, 2021, USA (Vol. 262, p. 264). Springer Nature.

Adaku E, Ankrah NA, Ndekugri IE, 2022, Prevention through design: conceptual models for the assessment of a principal designer's skills, knowledge and experience. Journal of Engineering, Design and Technology, Vol. 20 No. 3 pp. 595–623, doi: https://doi.org/10.1108/JEDT-07-2020-0278

BSI (2018). PAS 1192-6:2018 Specification for collaborative sharing and use of structured Health and Safety information using BIM. [online] British Standards Institution. Available at: https://knowledge.bsigroup.com/products/specification-for-collaborative-sharing-and-use-of-structured-health-and-safety-information-using-bim [Accessed 6 Feb. 2026].

CIRIA (2015). CDM 2015 - construction work sector guidance for designers, fourth edition (C755). [online] CIRIA. Available at: https://www.ciria.org/CIRIA/CIRIA/Item_Detail.aspx?iProductcode=C755&Category=BOOK [Accessed 6 Feb. 2026].

Eastman, C., Lee, J.M., Jeong, Y.S. and Lee, J.K., 2009. Automatic rule-based checking of building designs. Automation in construction, 18(8), pp.1011-1033.

European Union. 1992. Directive 92/57/EEC - temporary or mobile construction sites. Available at: https://eur-lex.europa.eu/eli/dir/1992/57/2019-07-26 [Accessed 6 Feb. 2026].

Hare, B., Kumar, B. and Campbell, J., 2020. Impact of a multi-media digital tool on identifying construction hazards under the UK construction design and management regulations. J. Inf. Technol. Constr., 25, pp.482-499.

Hevner, A.R., March, S.T., Park, J. and Ram, S., 2004. Design science in information systems research. MIS quarterly, pp.75-105..

HIGHWAYS SAFETY HUB. (2023). HIGHWAYS SAFETY HUB. [online] Available at: https://www.highwayssafetyhub.com/.

HSE. 2015a. Construction (Design and Management) Regulations 2015. In: HEALTH AND SAFETY EXECUTIVE (ed.).

HSE. 2015b. Construction (Design and Management) Regulations 2015. Guidance on Regulations [Online]. Available: https://www.hse.gov.uk/pubns/books/l153.htm [Accessed 6 Feb. 2026]

HSE (2025). Work-related fatal injuries in Great Britain, 2025. [online] Health and Safety Executive, pp.1–27. Available at: https://www.hse.gov.uk/statistics/assets/docs/fatalinjuries.pdf [Accessed 6 Feb. 2026].

ICE. 2010. So Far As Reasonably Practicable (SFARP) - A commentary for designers in construction. [online] Institution of Civil Engineers, pp.1–35. Available at: https://www.ice.org.uk/areas-of-interest/infrastructure-delivery/so-far-as-reasonably-practicable-sfarp-a-commentary-for-designers-in-construction [Accessed 6 Feb. 2026].

ICE. 2019. CDM 2015: 3-years on. [online] Institution of Civil Engineers, pp.1–27. Available at: https://www.ice.org.uk/media/ixshhezu/cdm-2015-3-years-on.pdf [Accessed 6 Feb. 2026].

ICE. 2020. Guidance for design risk management. [online] Institution of Civil Engineers, pp.1–23. Available at: https://www.ice.org.uk/media/usuhymf3/drm-guidance-version-2-march-2020.pdf [Accessed 6 Feb. 2026].

IMI 2026 Information Management Initiative Framework IMI Framework – BIM Standards, Guides & Resources [Accessed 6 Feb. 2026].

ISO (2025). ISO 19650-6:2025. [online] ISO. Available at: https://www.iso.org/standard/82705.html [Accessed 6 Feb. 2026].

Jolly, P. 2023. Navigating the Building Safety Act's role of Principal Designer. [online] Available at: https://www.ribaj.com/intelligence/navigating-the-role-of-principal-designer-building-safety-act/.

MPW R&R Ltd. 2023. Implementation of the Principal Designer role within CDM 2015 Technical Annex. [online] MPW R&R Ltd, pp.1–153. Available at: http://mpwrandr.co.uk/download/18682/?tmstv=1770410093 [Accessed 6 Feb. 2026].

Mzyece, D., Ndekugri, I.E. and Ankrah, N.A., 2019. Building information modelling (BIM) and the CDM regulations interoperability framework. Engineering, Construction and Architectural Management, 26(11), pp.2682-2704.

Nawari, N.O., 2012, March. BIM-model checking in building design. In Structures congress 2012 (pp. 941-952).

NBS 2026 Uniclass RK table https://uniclass.thenbs.com/taxon/rk [Accessed 6 Feb. 2026].

Ndekugri, I., Ankrah, N.A. and Adaku, E., 2022. The design coordination role at the pre-construction stage of construction projects. Building Research & Information, 50(4), pp.452-466.

Nisbet N, Ma L. 2023 Using RASE semantic mark-up for normative, definitive and descriptive knowledge. InECPPM 2022-eWork and eBusiness in Architecture, Engineering and Construction 2022 2023 Mar 29 (pp. 436-442). CRC Press.

Nisbet, N., Zhang, Z. and Çidik, S. 2024. Real-time assessment of regulatory compliance of construction sites, Proceedings of the 2024 European Conference on Computing in Construction (EC3), European Council on Computing in Construction. doi: 10.35490/EC3.2024.215.

Osorio-Sandoval, C.A., Crick, G., Collinge, W.H., Farghaly, K., Mosleh, M.H., Manu, P. and Cheung, C.M., 2023. Dataset of characterised construction safety risks and related treatments. Data in brief, 48, p.109293.

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S., 2007. A design science research methodology for information systems research. Journal of management information systems, 24(3), pp.45-77.

Takeda, H., Veerkamp, P. and Yoshikawa, H., 1990. Modeling design process. AI magazine, 11(4), pp.37-37.

Toole, T.M. and Gambatese, J., 2008. The trajectories of prevention through design in construction. Journal of safety research, 39(2), pp.225-230.

Zhang, Z., Kasim, T., Antwi-Afari, M.F., Tezel, A. and Potter, D., 2024a. Mapping the processes and developing the rule sets for automated compliance checking of health and safety regulations in UK's infrastructure projects. Digital Building Permit conference 2024. European Network for Digital Building Permits.

Zhang, Z., Tezel, A., Antwi-Afari, M.F., Kasim, T. and Potter, D., 2024b, Automated Compliance Checking Situation in Health and Safety Management in UK's Infrastructure Sector. Proceedings of the 2024 European Conference on Computing in Construction (EC3), European Council on Computing in Construction. doi:10.35490/EC3.2024.285

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A SEMI-AUTOMATED BIM-TO-VR WORKFLOW FOR TEACHING: ROBUST MODEL REVIEW IN CAVE AND HMD ENVIRONMENTS

Hu Hanwen, Emara Merna, and Menzel Karsten
Technical University Dresden, Germany

Abstract

Current BIM-to-VR approaches enable immersive visualization of buildings but often require extensive manual preprocessing and rely on rigid modeling conventions, limiting scalability and reuse. This paper presents a semi-automated BIM-to-VR workflow that enables immersive interaction with building data derived from Autodesk Revit in a VR environment implemented in Unity. The method automatically parses model hierarchies, derives building-level and component groupings, assigns colliders, and enables group-based visibility control and navigation without additional preprocessing. The workflow is demonstrated in both room-scale VR and head-mounted display setups and validated in teaching scenarios, showing reduced preparation effort and robust reuse across VR platforms.

Introduction

The architecture, engineering, and construction (AEC) industry has undergone a gradual transition from two-dimensional drafting toward three-dimensional digital modeling, with Building Information Modeling (BIM) becoming a core method for representing geometry and technical information across the building lifecycle (Sacks et al., 2018). However, interacting with complex BIM data through conventional desktop-based tools can be cognitively demanding, often limiting spatial understanding and collaborative discussion at building scale (Whyte, 2003). Immersive Virtual Reality (VR) has therefore gained attention as a complementary medium that enables intuitive exploration of BIM models, enhancing spatial cognition and improving the comprehension of architectural spaces compared to traditional visualization approaches (Samarasinghe et al., 2019).

In addition to improved visualization, VR systems support real-time interaction with building models. This capability helps designers and stakeholders identify design issues, respond to feedback, and refine solutions at early stages, thereby reducing the risk of costly changes later in the project lifecycle (Kamari et al., 2021). Due to these advantages, VR has been increasingly applied across different stages of the building lifecycle to enhance collaboration, decision-making, and design communication. Several studies have demonstrated the value of VR in stakeholder-driven sustainable design decision-making. For example, Atwa et al. (2019) used VR to collect feedback from users and experts during the development of sustainable landscape design principles for green business parks, enabling more informed and inclusive decision-making processes. Similarly, Tseng and Giau (2022) applied VR as an assessment tool for elderly housing, suggesting that immersive VR experiences offer a more realistic sense of presence and improved spatial understanding prior to occupancy.

In parallel with professional practice, significant efforts have been made to integrate BIM into AEC education curricula worldwide (Arundhati et al., 2013; Sacks and Pikas, 2013). BIM-based models and associated databases offer an enhanced educational platform by presenting complex construction information in a visual and accessible format (Clevenger et al., 2012). In particular, BIM's 4D capabilities enable students to better understand construction sequencing and assembly processes through animated simulations (Gledson and Dawson, 2017). In parallel, immersive technologies such as VR have been increasingly adopted in education, including gamification, game-enhanced learning, and experiential learning environments (Shavinina, 2013). Research on VR in education has increased substantially over the past two decades, with studies consistently reporting improved student understanding through direct interaction within realistic environments (Blazauskas et al., 2017).

Nevertheless, prior research emphasizes that VR applications must be grounded in appropriate pedagogical frameworks and learning theories to ensure alignment with educational objectives (Liu et al., 2017). This aligns with core Human-Data Interaction (HDI) principles, which suggest that for building data to be effectively utilized, the workflow must ensure legibility—making complex data understandable to the human observer—and provide the user with sufficient agency to control and customize the review process (Davila-Delgado et al., 2024; Kassem and Kifokeris, 2025).

Recent studies have specifically explored the integration of BIM and VR for AEC education. Elgewely et al. (2021) developed a BIM-integrated VR platform for construction detailing education and demonstrated that immersive, collaborative interaction with BIM models significantly improved students' engagement, motivation, and learning outcomes compared to traditional paper-based study. Similarly, Sanchez et al. (2021) reported on an intuitive and accessible BIM–VR educational framework

embedded within an AEC course curriculum. Their work illustrated how BIM–VR can be leveraged as a data-rich learning infrastructure to support complex design understanding, teamwork, remote collaboration, and client-oriented project presentation, reinforcing the value of BIM–VR as an effective educational medium in higher education.

Despite these promising results, Radianti et al. (2020), in a systematic mapping study of immersive VR applications in higher education, concluded that many VR-based educational systems remain experimental, lack grounding in learning theories, and focus primarily on usability rather than learning outcomes. Nevertheless, their findings confirm the strong potential of immersive VR as an effective and affordable pedagogical tool in university-level education.

In response to the limitations identified in existing BIM–VR applications, this paper presents a semi-automated BIM-to-VR workflow that enables immersive interaction with building data while minimizing manual preprocessing and avoiding reliance on proprietary VR middleware. The approach relies on hierarchy-based parsing of BIM-derived models after export, without requiring additional naming conventions or model-specific preparation. Building structures and component groupings are derived automatically from the existing model hierarchy, with interaction logic applied at the group level rather than per object. The workflow supports group-based visibility control, access to material information, automatic collider assignment, and navigation, and is robust to variations in model quality.

System Capabilities

The proposed workflow enables the immersive exploration of BIM data in VR environments, focusing on minimal manual preprocessing and broad accessibility. The system operates independently of proprietary VR middleware and requires only a standard export in the FBX format—the primary geometry exchange format utilized by Unity—from BIM authoring tools such as Autodesk Revit. Supporting both room-scale immersive

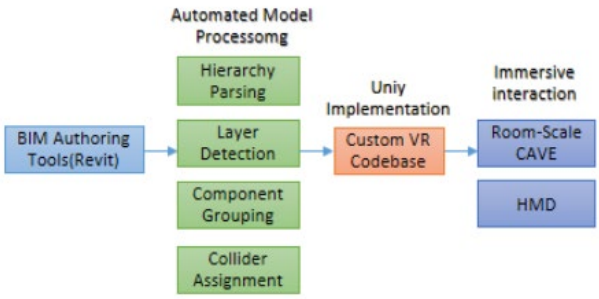


Figure 1: BIM-to-VR workflow

CAVE systems and personal head-mounted displays (HMDs), the implementation ensures a consistent user experience across varied hardware configurations. Figure 1 provides an overview of the processing stages, illustrating how the workflow enhances the legibility of complex building data—a core HDI principle—by ensuring that model hierarchies are automatically structured into a format that is understandable to non-expert users.

Upon importing a BIM-derived model, the system performs several automated operations to prepare it for interaction. Hierarchy parsing traverses the FBX structure to identify structural groupings and assemblies, detecting vertical levels and horizontal component clusters based on geometry distribution and node organization. Layer detection then infers building levels using geometry-aware height clustering, producing an ordered sequence of floors without relying on manual naming conventions or semantic annotations. Component grouping organizes subcomponents within each layer according to their hierarchical position, typically corresponding to major systems such as walls, slabs, or equipment assemblies, while collider assignment automatically equips mesh components with physics colliders to enable navigation without manual setup.

Building on this automated structuring, the interaction logic supports intuitive exploration, preserving spatial continuity through continuous, grounded movement. To ensure the system remains accessible to novice users,

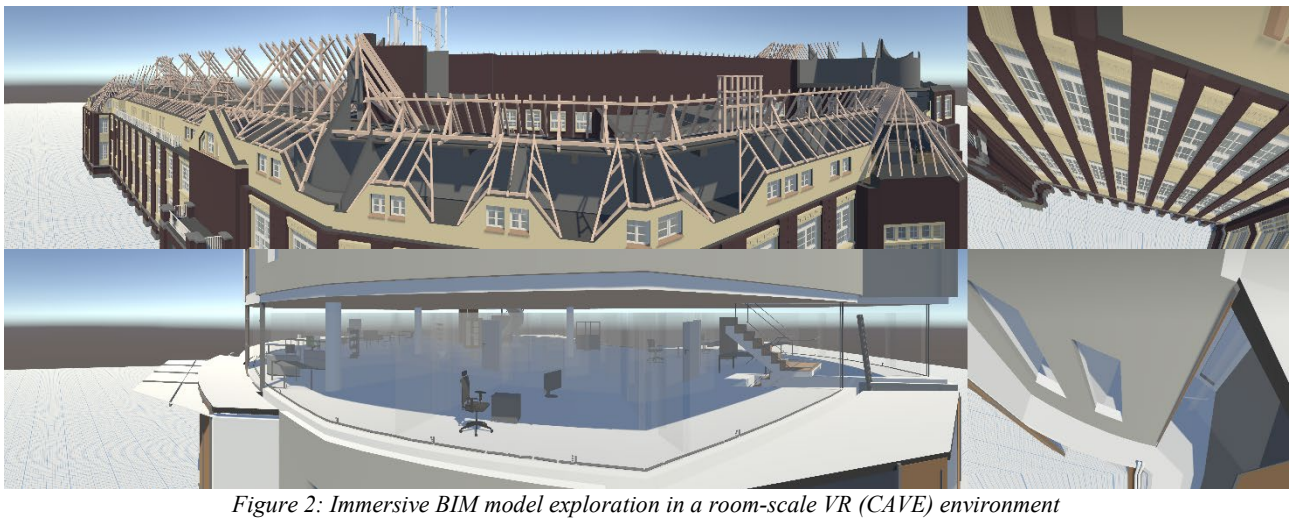


Figure 2: Immersive BIM model exploration in a room-scale VR (CAVE) environment

Table 1: Comparison of BIM-to-VR workflows for educational and exploratory use

	Proposed Workflow	Unity Reflect	Twinmotion
Required Export	fbx	Direct link	Datasmith exporter
Group/Layer Automation	yes	limited	limited
Interaction flexibility	high	low	moderate
CAVE support	yes	limited	limited
Primary use	Education&discussion	presentation	visualization

navigation is mapped to a standard Xbox controller. This choice of input device provides a familiar interface for students, allowing them to move through the building model with minimal technical training. The system's unified interaction logic ensures that the controller functions identically across both the room-scale CAVE and HMD-based setups, providing a consistent sense of agency as users choose their own paths through the digital environment.

To contextualize the proposed approach within the broader landscape of BIM-to-VR solutions, Table 1 compares the workflow with commonly used commercial and engine-based pipelines from the perspective of educational deployment. The comparison highlights differences in required preprocessing effort, dependence on proprietary middleware, and interaction flexibility.

Methodology

This study adopted an experimental research methodology conducted through a field-based teaching deployment to evaluate the real-world utility of the semi-automated BIM-to-VR workflow. The research was designed as a single-cohort study without a control group, focusing on the use of BIM models authored by participants in Autodesk Revit to isolate the specific impact of the automated processing from the general cognitive benefits of VR visualization. The deployment involved 40 students from architecture and engineering programs, organized into 11 groups of 5–6 participants. To facilitate the technical onboarding and ensure consistent system usage, the students were guided by one expert facilitator.

The VR application was deployed exclusively in a room-scale, multi-wall CAVE system to support shared immersive experiences. To facilitate navigation within this space, each student group utilized an Xbox controller as their primary interaction tool. During the standardized 30–45 minute introductory session, the expert facilitator trained the 40 students on how to use the controller to traverse the model, toggle layer visibility, and perform "peeling" interactions. By using a familiar consumer-grade controller, the study aimed to reduce the cognitive load on participants, allowing them to focus on the architectural review tasks rather than the mechanics of navigation.

The study is grounded in Human-Data Interaction (HDI) principles, specifically focusing on how the workflow enhances the legibility of complex building data and provides users with the agency to control and customize their review process. To emphasize educational application rather than tool demonstration, the VR sessions were structured as model-based review tasks aligned with established BIM learning objectives:

- **Spatial understanding and scale:** Participants explored their own partial BIM models at a 1:1 scale to assess proportions, spatial relationships, sightlines, and potential bottlenecks. This task leveraged VR's documented strengths in conveying scale and spatial detail, which are often lost on 2D screens.
- **Structural and system understanding:** Layer- and group-based visibility controls, including the progressive "peeling" interaction, were used to analyze complex relationships between internal building systems, such as the coordination between structural slabs and MEP equipment.
- **Data reflection and model quality:** Students performed material plausibility checks and metadata inspections to identify inconsistencies, such as missing or implausible material assignments. This process was designed to foster trust in the digital model by allowing students to verify informational accuracy directly within the 3D environment.
- **Collaboration:** Teams engaged in group-based discussion and joint critique of model segments directly within the immersive environment. This supported collective reasoning and peer-based feedback, as the shared viewpoint in the CAVE enabled members to ground their dialogue in the same visual evidence.

To reduce entry barriers and ensure consistent use across the 11 groups, a standardized training procedure was implemented. At the beginning of the study, participants attended an introductory session lasting approximately **30–45 minutes**, covering essential aspects of VR usage, including safety considerations, navigation principles, group-based visibility controls, and the use of material plausibility checks. Following this introduction, student groups were granted repeated, independent access to the CAVE system over a period of **2–3 weeks**. During this deployment phase, participants independently imported their own BIM models, performed the assigned review tasks, and documented their observations and discussion outcomes.

Data collection followed a mixed-methods approach, utilizing a structured questionnaire (n=40) and short reflective interviews to capture both measurable perceptions and contextualized user experiences. Quantitative data from Likert-scale and categorical items were analyzed using descriptive statistics, including

frequencies and percentages, to measure system usability and spatial comprehension. Qualitative data from open-ended questions were analyzed using thematic analysis following Braun and Clarke's inductive approach to identify recurring themes in the users' experiences. This rigorous approach allowed the study to distinguish between the inherent effectiveness of VR and the specific performance of the automated BIM-to-VR pipeline.

Discussion and Result Analysis

The evaluation of the system, based on 40 completed questionnaires, indicates high usability and effective pedagogical support despite the participants' limited technical background. As illustrated in Figure 3, which summarizes participants' prior experience with BIM tools, game engines, and VR, the study sample predominantly consisted of novice users. Specifically, nearly half of the participants (47.5%) reported no prior experience with Autodesk Revit, while 95% had no prior exposure to Unity and 87.2% had never used VR before. Despite this limited background, quantitative results indicate overall positive system usability, with 65% of participants expressing a desire to use the system frequently and 56.4% reporting confidence in its operation. While some participants initially faced a learning curve, qualitative feedback suggested these challenges diminished quickly with the guidance of the single expert facilitator who supported the 40 students during their initial onboarding. This suggests that the usability challenges were primarily related to first-time navigation rather than the fundamental design of the automated workflow.

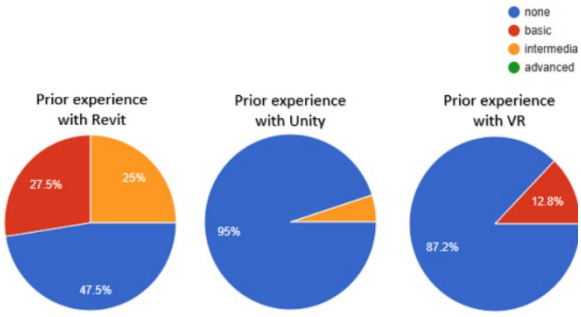


Figure 3: Prior Experience with BIM tools and VR

To address the distinction between general VR effectiveness and the proposed BIM-to-VR workflow, it is important to note that while VR inherently aids spatial reasoning, the automated hierarchy parsing was the specific factor that enabled these results for novice users. Without the semi-automated assignment of colliders and grouping, students would have spent the majority of their time on technical model preparation rather than analytical review. Spatial understanding represented the strongest observed effect, with 74.2% of participants reporting a better grasp of proportions and 61.3% finding spatial problem detection easier in VR than in desktop BIM tools. As illustrated in Figure 4, which displays the quantitative responses to spatial understanding and learning-related items, the results are strongly skewed toward agreement. By reducing technical friction, the workflow allowed participants to focus on high-level cognitive tasks, such as system-level reasoning and collaborative critique, which are often hindered in manual or proprietary pipelines.

From a Human-Data Interaction (HDI) perspective, the use of the Xbox controller was a key factor in promoting user agency and reducing the initial learning curve. Quantitative data showed that 56.4% of participants felt confident using the system, a result that qualitative feedback linked to the familiarity of the controller interface. While 20% of participants initially found navigation challenging, the "game-like" nature of the input device allowed them to overcome these hurdles quickly with the help of the expert guide. This choice of hardware, combined with the automated model preparation, effectively bridged the gap between complex architectural data and the user, ensuring that the legibility of the BIM model was matched by an intuitive means of exploration.

Following this intuitive navigation, the role of the workflow is further reflected in how participants interacted with structured building information. Quantitative results show that 50% of participants found layer and group visibility controls essential for understanding building organization, while 57.5% reported benefits from metadata inspection. The "peeling" interaction provided users with the agency to isolate specific systems, making sense of internal relationships that are typically obscured. Furthermore, the inclusion of material checks fostered a sense of trust, as participants could verify the accuracy of their models in real-time, identifying inconsistencies that were previously invisible in the desktop view.

Ultimately, 75% of participants agreed that the VR experience enhanced their learning compared to traditional methods. By removing the need for manual preparation and empowering students to manipulate data directly, the semi-automated workflow successfully bridged the gap between complex digital twins and human understanding. In addition to individual interaction, collaborative use emerged as a vital dimension, with 70.0% of participants perceiving improved team performance. These results suggest that the workflow is a robust tool for educational deployment, fostering improved individual and team performance while upholding the core HDI principles of empowerment and data transparency.

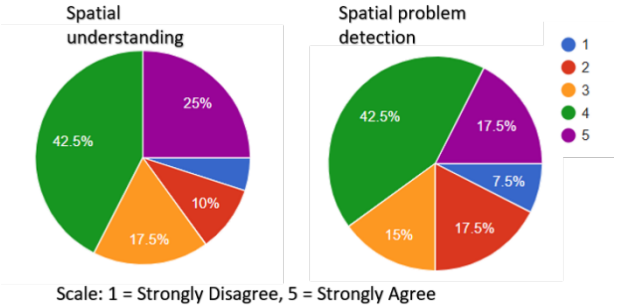


Figure 4: Spatial understanding in VR

Conclusion & Limitations

This research has presented and evaluated a semi-automated BIM-to-VR workflow specifically engineered to bridge the technical gap between complex architectural data and immersive educational review. By implementing an automated pipeline for hierarchy parsing, layer detection, and collider assignment using the FBX format within the Unity engine, the system successfully democratizes access to Virtual Reality for users without game-engine expertise. The field-based deployment, involving 40 students and guided by a single expert facilitator, demonstrates that the workflow is not merely a visualization tool but a functional interface for architectural analysis. A critical finding of this study is the distinction between the inherent cognitive affordances of VR—such as 1:1 scale perception and spatial reasoning—and the operational accessibility provided by the workflow. While the virtual medium provides the "where" for spatial understanding, it is the automated processing of the BIM data that provides the "how," allowing novice users to transition from desktop authoring to immersive critique without the burden of manual model conditioning.

From a theoretical standpoint, this study offers a significant contribution to the framework of Human-Data Interaction (HDI) in the built environment. By structuring interaction at the level of building systems and assemblies rather than granular geometric nodes, the workflow fundamentally enhances the legibility of complex BIM data. This structural clarity allows students to perceive a building not as a collection of meshes, but as a coherent system of floors, walls, and equipment. Furthermore, the provision of visibility controls and the progressive "peeling" interaction grants users a high degree of agency, empowering them to manipulate the digital environment to suit specific analytical tasks. This sense of control is essential for deep learning, as it moves the user from a passive observer to an active investigator of the building's internal logic. Additionally, the integration of material plausibility checks and metadata inspection fosters a vital sense of trust, ensuring that the immersive experience remains a high-fidelity reflection of the underlying engineering data.

The results from the room-scale CAVE deployment underscore the workflow's efficacy in fostering collaborative learning and peer-based critique. With 70% of participants reporting improved team performance and 75% noting an enhanced learning experience, the system proves to be a robust pedagogical aid. While the study identified certain limitations—specifically regarding the initial onboarding curve and the requirement for specialized CAVE infrastructure—the overall success of the deployment suggests that semi-automated BIM-to-VR transitions are a scalable solution for the AEC industry. Future research will focus on extending the system's capabilities to include bi-directional data synchronization, allowing users to annotate and modify BIM data directly within the VR environment. By continuing to refine the interface through the lens of HDI, this work aims to further empower the next generation of AEC professionals to interact with, understand, and trust the increasingly complex digital twins of our built environment.

References

Arundhati, G., Allan ,D. Chasey, Ph.D, PE and Scott, R. (2013) 'Industry and Academia: A partnership to VDC Curriculum'.

Atwa, S. M. H., Ibrahim, M. G., Saleh, A. M. and Murata, R. (2019) 'Development of sustainable landscape design guidelines for a green business park using virtual reality', Sustainable Cities and Society, vol. 48, p. 101543.

Blazauskas, T., Maskeliunas, R., Bartkute, R., Kersiene, V., Jurkeviciute, I. and Dubosas, M. (2017) 'Virtual Reality in Education: New Ways to Learn', in Damaševičius, R. and Mikašytė, V. (eds) Information and Software Technologies, Cham, Springer International Publishing, pp. 457–465.

Clevenger, C., Glick, S. and Del Puerto, C. L. (2012) 'Interoperable Learning Leveraging Building Information Modeling (BIM) in Construction Education', International Journal of Construction Education and Research, vol. 8, no. 2, pp. 101–118.

Davila-Delgado, J. M. et al. (2024) 'Paper Title Corresponding to ID 282', Proceedings of the 2024 European Conference on Computing in Construction (EC3).

Elgewely, M. H., Nadim, W., ElKassed, A., Yehiah, M., Talaat, M. A. and Abdennadher, S. (2021) 'Immersive construction detailing education: building information modeling (BIM)–based virtual reality (VR)', Open House International, vol. 46, no. 3, pp. 359–375.

Gledson, B. J. and Dawson, S. (2017) 'Use of Simulation Through BIM-Enabled Virtual Projects to Enhance Learning and Soft Employability Skills in Architectural Technology Education', in Dastbaz, M., Gorse, C. and Moncaster, A. (eds) Building Information Modelling, Building Performance, Design and Smart Construction, Cham, Springer International Publishing, pp. 79–92.

Kamari, A., Paari, A. and Torvund, H. Ø. (2021) 'BIM-Enabled Virtual Reality (VR) for Sustainability Life Cycle and Cost Assessment', Sustainability, vol. 13, no. 1, p. 249.

Kassem, M. and Kifokeris, D. (2025) 'Human-Data Interaction (HDI) in the Built Environment', EC3 White Paper 1.

Liu, D., Bhagat, K., Yuan, G., Huang, R. and Chang, T. (2017) 'The Potentials and Trends of Virtual Reality in Education', in, pp. 105–130.

Radianti, J., Majchrzak, T. A., Fromm, J. and Wohlgenannt, I. (2020) 'A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and

research agenda', Computers & Education, vol. 147, p. 103778.

Sacks, R., Eastman, C., Lee, G. and Teicholz, P. (2018) BIM Handbook, Wiley.

Sacks, R. and Pikas, E. (2013) 'Building Information Modeling Education for Construction Engineering and Management. I: Industry Requirements, State of the Art, and Gap Analysis', Journal of Construction Engineering and Management, vol. 139, no. 11.

Samarasinghe, D. A. S., Latif, S. A. and Baghaei, N. (2019) 'Virtual Reality Models for Promoting Learners Engagement in Construction Studies', 2019 IEEE Global Engineering Education Conference (EDUCON). Dubai, United Arab Emirates, 08.04.2019 - 11.04.2019, IEEE, pp. 1331–1335.

Sanchez, B., Ballinas-Gonzalez, R. and Rodriguez-Paz, M. X. (2021) 'Development of a BIM-VR application for e-learning engineering education', 2021 IEEE Global Engineering Education Conference (EDUCON). Vienna, Austria, 21.04.2021 - 23.04.2021, IEEE, pp. 329–333.

Shavinina, L. V. (2013) The Routledge International Handbook of Innovation Education, Routledge.

Tseng, K. C. and Giau, D. T. N. (2022) 'A feasibility study of using virtual reality as a pre-occupancy evaluation tool for the elderly', Automation in Construction, vol. 134, p. 104037.

Whyte, J. (2003) 'Industrial Applications of Virtual Reality in Architecture and Construction', vol. 8.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DEVELOPMENT OF AN IFC-BASED CONVERTER FOR BIM INTEGRATION INTO AUGMENTED REALITY APPLICATIONS FOR BUILDING CONSTRUCTION TRADES

Christian Annen, Hendrik Morgenstern, and Katharina Klemt-Albert
RWTH Aachen University, Aachen, Germany

Abstract

To address the shortage of skilled labor in the construction industry, the application of augmented reality on construction sites offers substantial potential to enhance coordination and communication among different trades, thereby reducing labor-intensive processes. This requires the reliable transfer of geometric and semantic information from Building Information Models, including all data relevant to construction execution, through suitable data exchange formats. This paper defines information requirements for the IFC schema for the BIM-to-field use case in the construction trades, discusses appropriate exchange formats for geometric and semantic data, and presents the development of an IFC-to-Unity converter.

Introduction

A key challenge currently facing the construction industry is the shortage of skilled labor. Numerous studies (Klaus *et al.*, 2025; PwC, 2025; Emanuel *et al.*, 2023) indicate that this shortage is particularly pronounced in construction and craft-related occupations.

The implementation of digital technologies can enhance process efficiency in the construction sector (PwC, 2025), thereby contributing to the alleviation of skilled labor shortages.

At the same time, the gap between digitalization potential and actual digital capabilities in the construction industry continues to widen (PwC, 2025).

According to Klaus *et al.* (Klaus *et al.*, 2025) the gap between digitalization potential and actual capabilities is especially pronounced in the areas of simulation and visualization, cloud technologies and platforms, and Internet of Things solutions.

When the potential of digital solutions is examined across different user groups in the construction sector, it becomes evident that construction trades, particularly construction workers, play a central role.

According to the Bitkom study *Digitalization in Skilled Trades* (Geffert, 2026), only 18% of the surveyed construction and finishing trades currently use BIM. At the same time, 76% of all surveyed skilled trade companies report that their employees require stronger digital skills (Geffert, 2026). These findings point to limited BIM adoption and a continued need for workforce-oriented digital support.

In this context, the BIM-to-field concept offers a promising approach. Its objective is to make BIM-based information accessible and usable on construction sites by delivering execution-related information in a user-centered and role-specific manner. Augmented reality (AR) represents a particularly suitable interface for transferring such information from the BIM model to the site in a form tailored to the specific requirements of different user groups and supportive of their on-site activities.

Against this background, the BIM-to-field concept offers considerable potential to improve on-site coordination, support craft workers during execution, and help mitigate the shortage of skilled labor.

Correspondingly, previous studies have shown that the use of AR on construction sites can accelerate on-site decision-making (Marinho *et al.*, 2026), improve collaboration, and enhance overall project performance (Alizadehsalehi *et al.*, 2020).

The BIM-to-field concept enables the positioning of building information models from the planning phase directly on the construction site using AR, allowing relevant information to be visualized via a device such as a smartphone, tablet or AR headset. The workflow from the building information model to the construction site is illustrated in Figure 1. In particular, the underlying data and its preparation represent a major challenge.

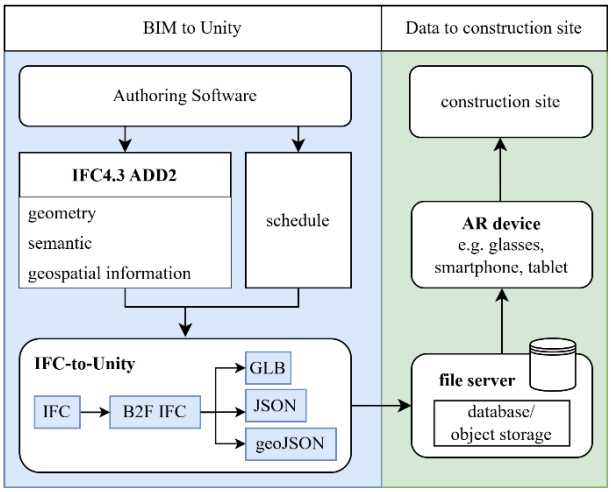


Figure 1: BIM-to-field concept within this research

Research Gap

Several AR applications already exist for both research and professional practice. These applications describe the BIM-to-field process in detail or provide functional solutions for a specific use case. However, most existing AR applications are based on the principle that information from the Building Information Model is used primarily for inspection and verification purposes. Consequently, these systems are not yet suitable for providing complex, role-specific action instructions for the various trades involved in construction execution. To sum up, most current AR applications are designed to support supervisory and managerial roles on construction sites, rather than operative roles within the construction trades.

At the same time, it is necessary to address how the different BIM standards adopted by various companies can be handled consistently. In this context, the varying information requirements and divergent coordinate reference systems (CRS) used during the planning phase must be considered. In OpenBIM, Industry Foundation Classes (IFC) is used as an open standard for exchanging model information. For this reason, the focus of this research is placed on the IFC-to-Unity process illustrated in Figure 1.

This paper describes how planning information for on-site execution can be systematically extended within an IFC model and defines the basic information requirements necessary to enable actionable, role-specific guidance for construction trades in an AR environment. Rather than extracting model data in isolation, the approach focuses on IFC-standard-compliant enrichment of existing model information, ensuring semantic consistency and interoperability.

In addition, the paper demonstrates how schedule-related information can be integrated into the IFC data structure in order to explicitly address individual trades as well as distinct role profiles within each trade. This enables the derivation of task-oriented, temporally structured instructions suitable for execution-level workflows on construction sites.

Furthermore, approaches are presented to address challenges arising from heterogeneously georeferenced IFC files and from common modelling inconsistencies and errors. Together, these measures ensure a robust, reliable and standards-based BIM-to-Unity data integration pipeline that supports construction workers beyond purely supervisory use cases.

Methodology

This study presents a structured approach to defining information requirements, analyzing the IFC schema and developing a converter for the BIM-to-field use case. The aim is to integrate BIM-to-field-specific information into the IFC schema to enable context-sensitive, AR-based work instructions for craft workers on construction sites.

The process begins with defining the problem and delimiting the research scope, driven by the absence of interoperable, IFC-compliant representations of the information required for field-oriented BIM applications.

To identify relevant concepts, existing approaches, and technical implementation strategies, an exploratory, non-systematic literature search was conducted using backward and forward snowballing, following Wohlin (Wohlin, 2014).

The literature search started from an initial set of key publications on construction scheduling, construction-specific information requirements, approaches for integrating such information into the IFC schema, georeferencing in IFC, digital data exchange and AR-supported construction workflows. It was iteratively expanded by reviewing the reference lists of relevant papers and identifying citing publications, thereby capturing relevant concepts, existing approaches and technical implementation strategies that informed the subsequent methodological steps.

On this basis, the study addresses four interrelated objectives:

- to define core information requirements for representing the BIM-to-field use case in IFC
- to examine how use-case-specific information, including schedule-related information, can be integrated into the IFC schema
- to implement these requirements in a technically robust and IFC-compliant manner
- to optimize the IFC-to-Unity information flow for downstream visualization and AR-based task guidance

These objectives are addressed through a detailed analysis of the IFC schema based on the IFC4X3_ADD2 documentation (buildingSMART International, 2026). The analysis focuses on the entities, attributes, and semantic relationships required to represent, link, and transfer the relevant information.

The resulting requirements are then translated into the development of a Python-based converter, including the selection of suitable Python packages for IFC processing, data handling, and workflow implementation.

Information Requirements

In current practice, information on building elements is commonly embedded directly within the IFC model, whereas scheduling information is usually generated in a separate planning environment and only subsequently linked to the building information model. Nevertheless, IFC provides several mechanisms for incorporating such information and representing schedule structures.

In addition to model information and schedules, supplementary information is sometimes also required. For example, tradespeople need access to assembly instructions for different products.

Therefore, an AR application for use in different construction trades must be able to interpret information from both the building model and the schedule. It must also be able to supplement information that is not part of the building information model, if necessary.

To support this objective, the following section presents a concept that defines the minimum structural requirements of an IFC model for consolidating the essential information in a single IFC file.

In line with the taxonomy proposed by Liu (Liu, 2021), which distinguishes the main categories of *product*, *organization*, *process* and *external factors* for construction workers, the discussion focuses primarily on what is being built and who is responsible for carrying out the work at a given time and place.

IFC Structure requirements

This section defines the minimum IFC data structure requirements for building construction that are necessary to support the coordination of complex trade structures and interdisciplinary interfaces through an AR-based application. The information requirements and relationship definitions developed in this work are based on the IFC4X3_ADD2 standard (DIN Deutsches Institut für Normung e. V., 2024). Additionally, this section illustrates the foundational concepts of the IFC schema that facilitates the integration of 4D information and responsibility assignments into the IFC data structure.

To ensure interoperability across different BIM software environments, the proposed extensions are realized through an entity-based modelling approach that is fully aligned with the IFC schema. Rather than introducing proprietary data structures, this approach uses existing IFC entities and standardized relationships. This preserves schema conformity and enables consistent interpretation across different BIM authoring and analysis tools.

Consequently, no new attributes are introduced in the form of *IfcPropertySingleValue* unless they are explicitly required as an extension. In this paper, existing IFC entities are systematically reused. These entities are interconnected via *IfcRelationship* instances. These relationships are defined according to the mapping shown in Figure 3.

This establishes a semantically consistent link between structures and responsible actors within the AR application. In this context, *IfcCrewResource* is used to represent a specific construction trade, to which different IfcLaborResources are assigned.

These *IfcLaborResources* are specified by an *IfcActor* which is defined by *IfcOrganization* (organization name) and *IfcPerson* (first name, last name).

Huymajer et al. (Huymajer *et al.*, 2024) present a similar approach. However, their work focuses on construction scheduling in tunnel engineering. Applications in building construction, as well as interfaces to augmented reality systems, are not addressed. Nevertheless, *IfcTask*, *IfcLaborResource* and *IfcPerson* are also used to represent IFC-based construction processes.

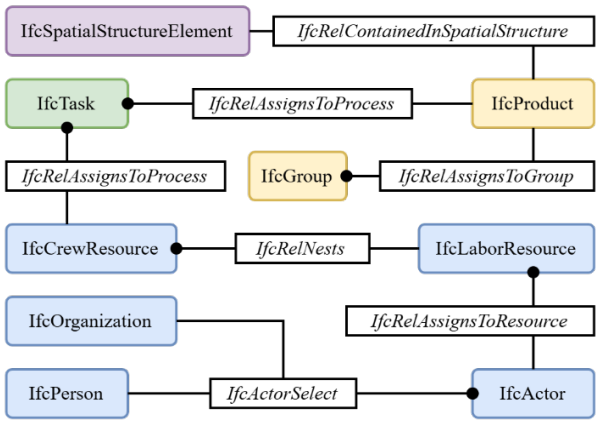


Figure 3: Entities and Relationships used in IFC

The construction trade (*IfcCrewResource*) can be associated with *IfcTask* entities, which represent construction activities. These tasks are, in turn, linked to *IfcBuiltElement* instances, which constitute a subtype of *IfcProduct* and describe the physical building elements involved in the execution of the tasks. *IfcMaterial* is used to describe the material characteristics, while *IfcPhysicalSimpleQuantity* entities are used to specify geometric properties, such as volume.

An *IfcBuiltElement* can be spatially located through its assignment to an *IfcSpatialStructureElement*. For instance, *IfcBuildingStorey* defines the corresponding floor level, whereas *IfcSpace* specifies the room or functional area in which the element is situated.

The defined entities are illustrated in Figure 2. In the figure, the information associated with the masonry trade and a sand-lime brick wall is highlighted in green for illustrative purposes.

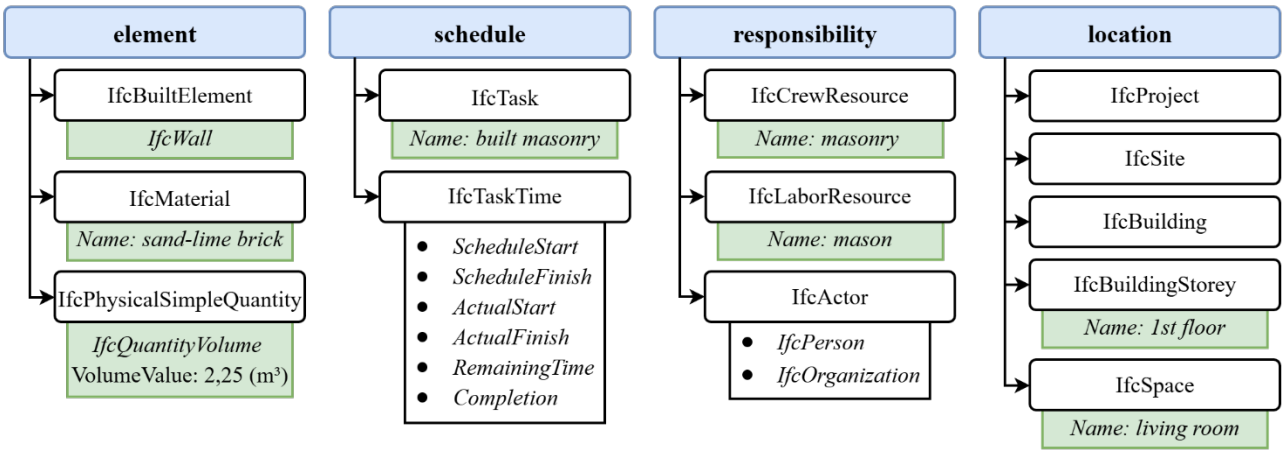


Figure 2: IFC entities for coordinating construction trades in building construction - exemplified by the trade masonry

In addition to *IfcTask*, scheduling information is enriched using selected *IfcTaskTime* attributes (Figure 2), which capture planned durations derived from the BIM model, as well as on-site execution states. Schedule deviations are documented using the *ActualStart* and *ActualFinish* attributes, the *RemainingTime* attribute shows the remaining time, and the Completion attribute indicates task completion, encoded as an *IfcPositiveRatioMeasure* (0.0-1.0).

Geometric and spatial reference Information

In addition to the semantic requirements imposed on an IFC file, particular attention must also be paid to the geometric modelling with respect to the heterogeneous BIM standards applied by industry practitioners. For the concept presented in the previous section to function as intended, building components must be modelled as instances of *IfcElement*.

With regard to georeferencing, it must be considered that different coordinate reference systems are commonly used in practice. This heterogeneity poses challenges for data integration in Unity, as models must be correctly aligned in a common coordinate reference system while simultaneously ensuring that their spatial extent remains sufficiently close to the origin. The latter is essential due to performance constraints and spatial jitter caused by floating-point precision limitations due to the large distance of the objects from the world origin (Thome, 2005).

The problem can be mitigated by positioning the model coordinates close to the origin. If necessary, they can be converted to a global CRS using the transformation parameters defined in *IfcMapConversion*.

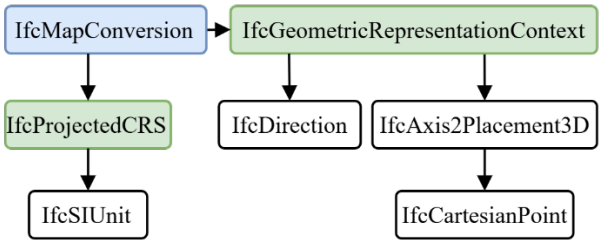


Figure 4: IFC entities linked to IfcMapConversion

IfcMapConversion stores the transformation parameters of the 2D Helmert transformation (buildingSMART International, 2020) and references *IfcProjectedCRS*, which defines the target CRS, as well as the applied length unit (*IfcSIUnit*). All entities associated with *IfcMapConversion* are illustrated in Figure 4.

To improve the interoperability of geospatial data and allow it to be integrated into Geographic Information Systems (GIS), the origin of the Cartesian model is converted into geodetic coordinates using the 2D Helmert transformation.

To preserve the associated geospatial information, these data are stored separately in a GeoJSON file. As GeoJSON relies on the World Geodetic System 1984 (WGS84) datum as its underlying CRS (Butler et al., 2016), the developed converter evaluates the EPSG code specified within *IfcProjectedCRS* to identify the coordinate reference system in which the IFC model is defined and to ensure a consistent transformation into the global reference frame. The coordinate transformation workflow is illustrated in Figure 5.

Development of an IFC-to-Unity Converter

During the development of the converter, the focus is placed on transmitting the previously defined semantic, 4D and geospatial information requirements for the BIM-to-field use case to the construction site.

Additionally, as the AR application runs natively on AR-capable end devices and is Unity-based, the building information model must be converted into Unity-compatible formats.

To validate the converter's functionality during development, two different IFC files are used as input datasets. Table 1 summarizes the key characteristics of these files.

Table 1: IFC Test Files used for Development

Test IFC	1	2
Size	80217 KB	286 KB
Nr. of Objects	1591	15
Source	Revit 2025	(buildingSMART, 2026)
Length Unit	m	mm
EPSG-Code	EPSG:25832	EPSG:32760
File Schema	IFC4X3_ADD2	

An AI assistant was used to support parts of the programming work (code suggestions and refactoring). All methodological decisions, result validation and final implementation approval were performed manually.

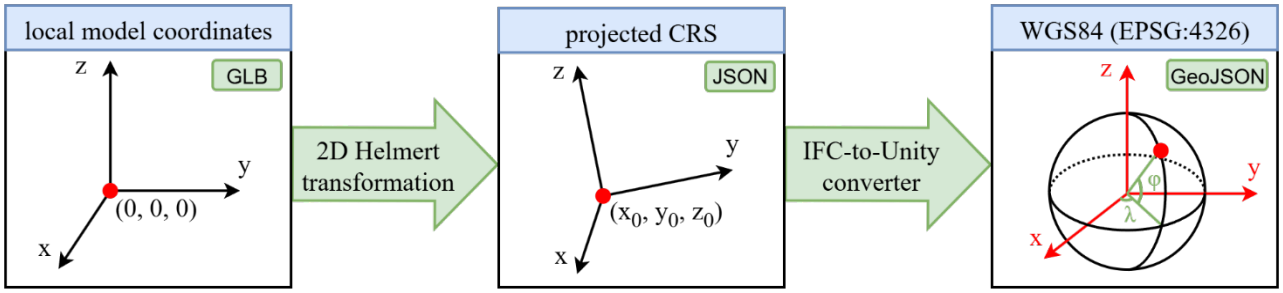


Figure 5: workflow for coordinate transformation to increase interoperability

Transmission of Geometric Information

As Unity cannot read the IFC schema, an appropriate 3D file format is selected to minimize data volume and loading times. Lee et al. (Lee *et al.*, 2019) compare the performance characteristics of commonly used 3D file formats, such as OBJ, STL, FBX, glTF and GLB. The analysis shows that glTF and GLB exhibit the highest rendering efficiency based on measured loading times within a web application.

A glTF asset is represented by a JSON-formatted (.gltf) file that contains a complete scene description, accompanied by separate binary files that store geometry data and image files that contain textures. If all this information is encapsulated within a single container file, the format is referred to as GLB. (Khronos Group, 2021)

To minimize loading times in the AR application while reducing the number of files to be managed, GLB is selected as the export format.

To transform the data from IFC to GLB, Python is used in combination with the *ifcopenshell* package (*IfcOpenShell*, 2024). The *ifcopenshell.geom* extension is used to control the geometric properties.

Since length units may be defined differently in different IFC files (see Table 1), the converter verifies whether the coordinates are defined in the length unit metre. If this is not the case, the coordinates are converted to metres before the transformation is applied.

Test IFC 1 (see Table 1) contains building openings that are modeled using void elements (*IfcOpeningElement*). To prevent these void elements from being exported as visible 3D objects into the GLB file, an auxiliary function is implemented within the codebase that collects the IDs of all objects of the type *IfcOpeningElement* and enforces the creation of openings in the GLB file by explicitly enabling opening subtractions. For this purpose, the software package *ifcopenshell* (*IfcOpenShell*, 2024) is utilized. During the export process, the filtered void elements are excluded.

Transmission of Semantic and 4D Information

The GLB format is geometry- and material-centric and does not provide a standardized way to natively represent IFC semantic information such as entity types, property sets or relational structures.

Various file formats are used in the literature to store semantic data for AR applications. Common options include CSV (Arvikar *et al.*, 2025), XML (Rehman *et al.*, 2025) and JSON (Lieblich & Weise, 2013). As XML (Extensible Markup Language) and JSON (JavaScript Object Notation) are both object-based, they can represent hierarchies and relationships derived from the IFC structure. This makes them particularly well-suited to the BIM-to-field use case.

As buildingSMART provides a dedicated ifcXML standard (Lieblich & Weise, 2013), XML is the preferred choice for representing the pure IFC structure. However, the performance of the AR application on the construction site poses a significant challenge within the BIM-to-field context. In this regard, the JSON format offers a significant advantage as it can be parsed more efficiently (Afsari *et al.*, 2017), resulting in improved runtime performance. Consequently, JSON is used as exchange format.

Before the properties are written to the JSON file, the converter first enriches a duplicated IFC model (B2F IFC) with the required entities.

At this stage, supplementary information – specifically scheduling and responsibility data (see Figure 2) – is extracted from a construction schedule. For validation purposes, a simplified schedule in JSON format is used, incorporating information on *IfcTask*, *ScheduleStart*, *ScheduleFinish*, *IfcCrewResource*, *IfcLaborResources*, *IfcActor*, *IfcPerson and IfcOrganization.*

Once the properties have been added to the B2F IFC, the BIM-to-field–relevant information is stored in an object-based manner in JSON format using Python. The exported properties are mapped to the corresponding 3D object using the GUID as the linking key.

Transmission of Geospatial Information

Geospatial information is stored in accordance with the workflow illustrated in Figure 5. Local model coordinates are stored in the GLB file to improve rendering performance and avoid visual artefacts. The projected coordinate reference system used during the planning phase is stored as an object in a JSON file. This enables spatially consistent, project-specific alignment and assembly of multiple models.

To further enhance interoperability, particularly with GIS applications, the coordinates of the projected coordinate reference system are also transformed into the WGS84 reference system and stored in GeoJSON format.

The converter uses the Python package *pyproj* (*pyproj*, 2026) to transform the coordinates into the target system (WGS84; EPSG:4326). As illustrated in Figure 5, the local model coordinates are transformed into global coordinates, and the EPSG code defined in IfcProjectedCRS is used to determine the source CRS for transforming the model origin into WGS84.

The geospatial coordinates stored within the IFC model itself remain unchanged.

Graphical User Interface

Figure 6 illustrates the graphical user interface (GUI) of the developed IFC-to-Unity Converter. To provide planning data for use in a BIM-to-field application in the most accessible and flexible way possible, the GUI has been designed so that the core functionalities can be controlled via the GUI. Accordingly, the converter functions described in the preceding sections can be selectively enabled via the GUI. The default configuration is shown in Figure 6.

Geometry-related functions are not selected by default in the GUI as these operations are only required when the underlying IFC file does not conform to the modelling standard. In such cases, in addition to the geometry functions already described, the GUI provides further options to rotate the model around the Z-axis and redefine

the Z-axis (force Z-axis), should the coordinate system be misinterpreted.

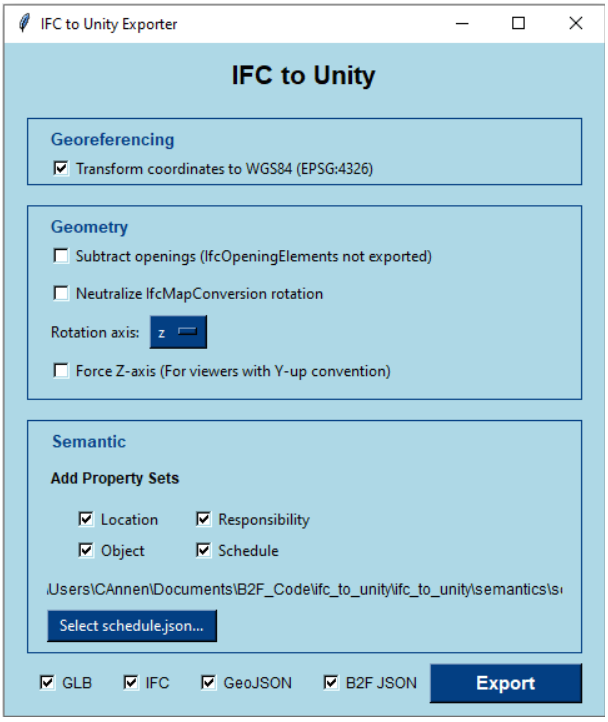


Figure 6: Graphical User Interface of the Converter

Validation

The validation of the IFC-to-Unity converter was carried out in two phases:

(1) preparation of the input data and
(2) execution of the validation.

Twelve IFC models were used as input data. Additional model information is summarized in Table 2. As no construction schedules were available for these models, model-specific schedule files were generated automatically from the IFC data. For this purpose, a Python-based schedule generator was developed to identify *IfcProduct* entities with geometric representations, group them by building storey and element category, and assign them to the trade groups structural works, finishing works, building services and other works. Based on this, deterministic synthetic actors and sequential task schedules were created.

The implementation was carried out in Python using *IfcOpenShell* for IFC parsing and geometry-related model queries. Standard Python libraries were employed for grouping, identifier generation, schedule logic and serialization. The objective was not to create technically realistic construction schedules, but to provide formally complete schedule data in JSON format, containing all input parameters required by the converter and thereby covering the previously defined minimum information requirements.

After a suitable schedule had been generated for each IFC file listed in Table 2, the resulting IFC-schedule pairs were processed by the IFC-to-Unity converter and transformed into the target formats GLB, IFC, GeoJSON, and B2FJSON. All input files were successfully transformed and thus formed the basis for the subsequent validation phase.

Table 2: IFC Files used for Validation

File	Model	Nr. of Objects	Size
1.1	structural	693	73058 KB
1.2	architectural	380	912 KB
2.1	structural	1138	49974 KB
2.2	architectural	469	922 KB
3.1	structural	435	79480 KB
3.2	architectural	322	1044 KB
4.1	structural	1016	80595 KB
4.2	architectural	639	1840 KB
5.1	structural	778	77369 KB
5.2	architectural	698	1897 KB
6.1	structural	890	17205 KB
6.2	architectural	383	972 KB

During validation, particular attention must be paid to the quality of the input data. The converter fully populates the entities assigned to the categories *Schedule* and *Responsibility* shown in Fig. 2, whereas those assigned to *Element* and *Location* depend on the completeness of the information contained in the input data. For example, if no material is assigned to an element, the exporter cannot infer the missing information automatically.

Accordingly, the validation is performed as follows. The properties assigned to the categories *Element* and *Location* are checked by means of scripts such that missing properties are reported only as warnings during export. By contrast, the properties assigned to *Schedule* and *Responsibility* must be transferred completely and correctly. Any deviation detected in these categories is therefore reported as an error during export. The whole validation process is shown in Figure 7.

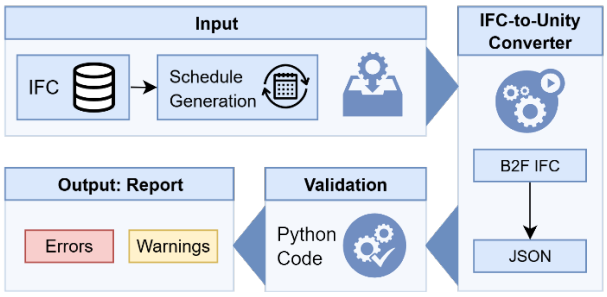


Figure 7: Validation Workflow

The validation is considered successful if the reports for the IFC files listed in Table 2 contain no errors and if all warnings can be logically explained.

The validation is performed on the exported JSON file, as this file is generated at the final stage of the IFC-to-Unity workflow. If a given piece of information is present in the JSON file, it can be assumed that it has also been successfully transferred to the B2F IFC.

The results of the validation are shown in Table 3.

Table 3: Validation Results

File	Attributes Checked	Errors	Warnings	Warning Ratio
1.1	8494	0	638	7.5%
1.2	15882	0	1526	9.6%
2.1	10601	0	654	6.2%
2.2	26916	0	2484	9.2%
3.1	7094	0	627	8.8%
3.2	9507	0	935	9.8%
4.1	14584	0	1132	7.8%
4.2	23968	0	2302	9.6%
5.1	15758	0	1374	8.7%
5.2	17377	0	1667	9.6%
6.1	8650	0	632	7.3%
6.2	21086	0	2072	9.8%
Total	**179917**	**0**	**16043**	**8.7%**

The warnings can be attributed to the following issues:

(1) the IFC source data do not provide values for either *Name* or *LongName* of *IfcBuilding* and
(2) some objects within the spatial IFC structure are not assigned to an *IfcBuildingStorey*.

As both warning types result from inconsistent input data, they are considered acceptable within the scope of the pipeline validation. Since the cross-model verification reveals no errors, the validation of the IFC-to-Unity converter is therefore considered successful.

Conclusions

Within the scope of this paper, various approaches are investigated to transform information from a building information model into a format readable by a game engine (Unity). It is demonstrated how both geometric data and semantic information can be transferred in a structured and interoperable manner. In accordance with the principles of OpenBIM, IFC version 4.3 ADD2 is employed as the source format for data transformation.

This paper formulates semantic requirements for building information models and the underlying IFC schema in order to enable the BIM-to-field use case for trades and construction workers in the construction industry. A set of relevant entities is defined based on IFC4X3_ADD2 to support targeted coordination among different trades. This includes information on building element properties, responsibilities, construction schedules and spatial location on the construction site. By relying on standardized IFC entities, interoperability between heterogeneous BIM authoring tools is ensured, while simultaneously establishing a foundation for enriching IFC models with additional information acquired during construction in future developments.

For geometric representation, GLB is identified as a suitable target format, while JSON is proposed for the transmission of semantic information. With regard to the export of 3D data, particular attention is given to the handling of IFC models that have already been georeferenced during the planning phase using *IfcMapConversion*.

While Unity operates using local model coordinates, BIM-based building planning typically relies on cartesian CRS. In contrast, data integration into GIS environments requires the use of global coordinate reference systems. To enhance interoperability in the context of data exchange, a workflow for handling these differing geospatial requirements is presented.

Furthermore, strategies are presented to prevent the export of non-essential 3D objects (e.g., *IfcOpeningElement*) by selectively filtering them out of the resulting GLB file.

Based on these findings, an IFC-to-Unity converter was developed and successfully validated.

In the context of further research, the defined information requirements will be refined and extended with trade-specific attributes. In addition, an interface for integrating construction scheduling data will be developed. This includes the specification of a linkage to scheduling tools commonly used during the planning phase, enabling the consideration of complex temporal dependencies within a BIM-to-field application. Moreover, future work will investigate methods for integrating newly acquired on-site information into the planning model. Within the field-to-BIM approach, particular emphasis will be placed on integrating schedule updates and construction progress data into the BIM environment.

Acknowledgments

This research paper was developed within the scope of the research project BIM2Field, which is funded by “Federal Ministry for Economic Affairs and Energy” under the funding program “ZIM – Central Innovation Programme for small and medium-sized enterprises (SMEs)”.

References

Afsari, K., Eastman, C.M. & Castro-Lacouture, D. (2017) JavaScript Object Notation (JSON) data serialization for IFC schema in web-based BIM data exchange. Automation in Construction, 77, 24–51. Available from: https://doi.org/10.1016/j.autcon.2017.01.011.

Alizadehsalehi, S., Hadavi, A. & Huang, J.C. (2020) From BIM to extended reality in AEC industry. Automation

in Construction, 116, 103254. Available from: https://doi.org/10.1016/j.autcon.2020.103254.

Arvikar, S., Aung, P.P., Cha, G.C. & Park, S. (2025) Augmented Reality-based construction site management using optimized BIM and project schedule integration. Automation in Construction, 175. Available from: https://doi.org/10.1016/j.autcon.2025.106204.

buildingSMART (2026) buildingSMART/Sample-Test-Files: Sample files of various formats and schema versions. Available from: https://github.com/buildingSMART/Sample-Test-Files/blob/main/IFC%204.3.2.0%20(IFC4X3_ADD2)/PCERT-Sample-Scene/Building-Structural.ifc [Accessed 23 January 2026].

buildingSMART International (2020) User Guide for Geo-referencing in IFC: How to Setup Geo-referencing in a Building or Linear Infrastructure Model, 2nd edition. Available from: https://www.buildingsmart.org/wp-content/uploads/2020/02/User-Guide-for-Geo-referencing-in-IFC-v2.0.pdf.

buildingSMART International (2026) IFC Documentation (IFC4X3_ADD2). Available from: https://ifc43-docs.standards.buildingsmart.org/ [Accessed 6 April 2026].

DIN Deutsches Institut für Normung e. V. (2024) Industry Foundation Classes (IFC) für den Datenaustausch in der Bauwirtschaft und im Anlagenmanagement – Teil 1: Datenschema.

Emanuel, B., Klinge, S., Neu-Yanders, N., Leber, U. & Schwengler, B. (2023) Auswirkungen des Angriffskrieges auf die Ukraine auf die Betriebe in Deutschland und weitere Ergebnisse des IAB-Betriebspanels 2022. Available from: https://doku.iab.de/forschungsbericht/2023/fb1523.pdf.

Geffert, A. (2026) Digitalisierung des Handwerks. Available from: https://doi.org/10.64022/2025-digitalisierung-handwerk.

Huymajer, M., Paskaleva, G., Wenighofer, R., Huemer, C. & Mazak-Huemer, A. (2024) IFC concepts in the execution phase of conventional tunneling projects. Tunnelling and Underground Space Technology, 143, 105368. Available from: https://doi.org/10.1016/j.tust.2023.105368.

IfcOpenShell (Version Version: 0.8.4.post1) (2024). Available from: http://ifcopenshell.org/.

Khronos Group (2021) glTF™ 2.0 Specification: Version 2.0.1. Available from: https://registry.khronos.org/glTF/specs/2.0/glTF-2.0.pdf.

Klaus, A., Beckmann, R. & Deyerler, Y. (2025) Fachkräfteengpassanalyse 2024. Available from: https://statistik.arbeitsagentur.de/.

Lee, G., Choi, P., Nam, J., Han, H., Lee, S. & Kwon, S. (2019) A Study on the Performance Comparison of 3D File Formats on the Web. The International Journal of Advanced Smart Convergence, (1), 65–74. Available from: https://doi.org/10.7236/IJASC.2019.8.1.65.

Lieblich, T. & Weise, M. (2013) ifcXML4Specification Methodology: Specification of the mapping methodology from EXPRESS to XML schema definition for IFC4. Version 1.1. Available from: https://standards.buildingsmart.org/documents/Implementation/ifcXML4%20specification%20methodology%20V1-1.pdf.

Liu, R. (2021) Impact of enhanced context-awareness for construction field crews on task cycle time.

Marinho, J., Sá, F., Durães, J., Fonseca, I. & Martins, N.C. (2026) A Systematic Literature Review of Augmented Reality's Development in Construction. Electronics, 15(1), 225. Available from: https://doi.org/10.3390/electronics15010225.

PwC (2025) Die Bauindustrie weiter im Umbruch: Fehlendes Know-how und Bürokratie bremsen: Die PwC-Studie zum Umgang der Baubranchemit den aktuellen Herausforderungen.

pyproj (2026). Available from: https://github.com/pyproj4/pyproj.

Rehman, S.U., Hwang, K.-E. & Kim, I. (2025) Semantic-enriched framework for asynchronous bidirectional BIM data exchange with immersive visualization platform. Automation in Construction, 178, 106409. Available from: https://doi.org/10.1016/j.autcon.2025.106409.

Thome, C. (2005) Using a floating origin to improve fidelity and performance of large, distributed virtual worlds. In: 2005 International Conference on Cyberworlds (CW'05), 2005 International Conference on Cyberworlds (CW'05), 11/23/2005 - 11/25/2005, Singapore. IEEE, 8 pp-270.

Wohlin, C. (2014) Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Shepperd, M., Hall, T. & Myrtveit, I. (Eds.) Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, EASE '14: 18th International Conference on Evaluation and Assessment in Software Engineering, 13 05 2014 14 05 2014, London England United Kingdom. ACM: New York, NY, USA, pp. 1–10.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

BIM–AR INTEGRATION IN PRACTICE: OPPORTUNITIES, CHALLENGES, AND TRADE-OFFS FROM A LIVE CONSTRUCTION CASE STUDY

Avar Almukhtar, Muhammad Shahzad, Marri Deepak Kumar, Joseph H.M Tah
Oxford Brookes University, Oxford, United Kingdom

Abstract

Building Information Modelling (BIM) and Augmented Reality (AR) integration has been widely proposed to improve construction coordination, yet evidence from live site deployments remains limited. This paper presents a qualitative case study of BIM–AR use with Gamma AR on an active university accommodation project in the UK. Site observations and interviews were used to examine workflow integration, benefits, and challenges. The findings indicate that BIM–AR enhances visualisation, coordination, and early error detection when applied selectively, but is constrained by alignment reliability, site conditions, and adoption factors. The study contributes practical deployment insights for BIM–AR implementation in construction.

Introduction

The Architecture, Engineering and Construction (AEC) industry continues to face persistent inefficiencies arising from fragmented information flows, bespoke project delivery practices, and limited integration between design intent and site execution (Wuni et al., 2024; Mujahed et al., 2025; Farea et al., 2026). Building Information Modelling (BIM) has emerged as a central response to these challenges by enabling shared digital representations of built assets and supporting coordination across disciplines and project lifecycle stages (Waqar et al., 2023; Rezgui et al., 2025). Through standardised data structures and interoperable formats such as Industry Foundation Classes (IFC), BIM provides a common information backbone that improves data consistency and cross-disciplinary communication (Gerbino et al., 2021; Lee et al., 2021; Chatsuwan et al., 2025).

In parallel, Augmented Reality (AR) has gained increasing attention as a means of improving how digital construction information is accessed and interpreted on site. By overlaying virtual content onto the physical environment via mobile or head-mounted devices, AR enables more intuitive, spatially contextualised interaction with BIM data compared to conventional screen-based tools (Alizadehsalehi et al., 2020; Sidani et al., 2021; Chen and Xue, 2022). Prior research reports that AR can enhance spatial understanding, support design review and progress monitoring, and facilitate early detection of discrepancies between planned and as-built conditions (Amin et al., 2023; Girgin et al., 2023; Tan et al., 2024; Pan and Isnaeni, 2024). AR-based visualisation has also been associated with improved task accuracy and reduced error rates through just-in-time guidance and in-situ verification (Dallasega et al., 2023; Marino et al., 2024; Chaudhari et al., 2025).

The integration of BIM and AR therefore holds promise for narrowing the long-recognised gap between digital project models and physical site realities. Empirical and review studies have demonstrated potential productivity and accuracy gains from BIM–AR integration across activities such as design review, assembly, discrepancy checking, and hazard identification (Fazel and Adel, 2024; Tan et al., 2024; Arvikar et al., 2025). However, despite these reported benefits, adoption of BIM–AR on live construction projects remains limited. Prior work consistently highlights challenges related to model preparation and interoperability, alignment and tracking accuracy, hardware constraints, environmental interference, and the need for user training and workflow adaptation (Schiavi et al., 2022; Kolaei et al., 2022; Amin et al., 2023; Jin et al., 2026). Importantly, these challenges are increasingly recognised as socio-technical in nature, with organisational readiness, staff capability, and compatibility with existing information management practices playing a decisive role in determining whether AR solutions deliver value in practice (El Asmar et al., 2021; Wang et al., 2024).

A key limitation in the current body of knowledge is the scarcity of detailed, in-situ evidence documenting how BIM–AR integration is deployed on active construction sites using commercial platforms, and how such deployments interact with established project workflows. Much of the existing literature focuses on prototype systems, controlled experiments, or performance metrics such as alignment accuracy, offering limited insight into real-world deployment conditions, calibration practices, issue-management integration, and day-to-day usability trade-offs faced by site teams (Cheng et al., 2020; Girgin et al., 2023; Bhatarai et al., 2025). This gap constrains both practitioner decision-making and the translation of research outcomes into scalable industry practice.

This paper addresses this gap through a qualitative case study of BIM–AR integration using Gamma AR, a commercial AR platform deployed on a major construction project in Oxford, UK. The study examines

how Gamma AR was configured to access BIM models hosted on Autodesk Construction Cloud (BIM 360), how AR visualisation was used on site via handheld devices, and how issues identified in AR were linked back to established digital coordination workflows. Drawing on site observations and semi-structured interviews with project staff, the research investigates both the opportunities created by BIM–AR integration, such as enhanced visualisation, improved coordination, and early error detection, and the challenges encountered during live deployment, including alignment reliability, environmental and hardware limitations, data-update frictions, and user adoption barriers. Some of these barriers, particularly around usability, trust, and model interaction, can also be interpreted through the lens of Human–Data Interaction (HDI), where constraints in *legibility*, *agency*, and *negotiability* (Mortier et al., 2014) may help explain the limited mainstream adoption of BIM–AR technologies in construction settings.

The contribution of this paper is threefold. First, it provides an empirically grounded account of an end-to-end BIM–AR deployment using a commercial toolchain under live construction conditions. Second, it synthesises practitioner experiences into a structured analysis of observed benefits and challenges, complementing prototype-focused research with deployment-level evidence. Third, it derives practical implications for both construction teams considering AR adoption and developers seeking to reduce friction in BIM–AR integration, particularly in relation to workflow fit, usability, and calibration robustness.

The remainder of the paper is structured as follows: the next section presents the research methodology and case study context; this is followed by results and discussion; and the paper concludes with key findings and implications for future research and practice.

Methodology

This research adopted a qualitative case study approach to investigate the deployment of a BIM-integrated Augmented Reality (AR) solution on an active construction site. A case study method was selected as it is well suited to investigating emerging digital technologies in real-world contexts, where implementation conditions, workflows, and user interactions cannot be meaningfully isolated or controlled. Rather than seeking to quantify performance metrics, the study focuses on understanding how BIM–AR integration is configured, used, and experienced in practice.

Case Study Context

The case study was conducted on the Clive Booth Student Village redevelopment project at Oxford Brookes University, UK; a large-scale student accommodation development delivered by a major UK contractor. The project employed a coordinated BIM model hosted on Autodesk Construction Cloud (BIM 360) to support design coordination and construction management. Gamma AR, a commercial AR application designed for construction use, was introduced on the project to enable on-site visualisation of BIM models and support spatially contextualised issue identification.

Gamma AR was used via handheld tablet devices to overlay BIM content onto the physical construction environment in real time as shown in Figure 1. The application was configured to access the latest project models from BIM 360 and to synchronise AR-based issues capture back into the project's established digital coordination workflows. The deployment therefore represented a realistic, industry-facing implementation rather than a prototype or experimental system.

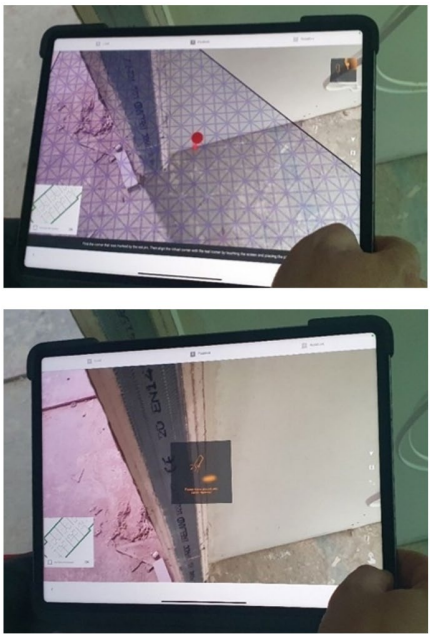


Figure 1: Identifying the purple mesh on a flat surface before restarting the Gamma Application

BIM-AR Workflow Configuration

To enable AR visualisation, the project's BIM model authored in Autodesk Revit, was exported in formats compatible with Gamma AR (including RVT, IFC, and NWC) and made available through the cloud-based BIM collaboration platform. On site, users selected relevant floors or zones within the Gamma AR application and aligned the virtual model to the physical environment using identifiable reference features such as structural gridlines and slab edges. Once aligned, the application maintained model registration using marker-less tracking based on the tablet's onboard sensors and camera.

A key aspect of the workflow was the integration of AR-based issue identification with the project's existing issue management processes. Discrepancies observed through the AR overlay could be logged directly within Gamma AR, including descriptive text and images, and synchronised back to BIM 360. This enabled issue identified in the field to be reviewed and addressed by the wider project team using familiar coordination tools. Figure 2 illustrates this BIM–AR workflow, from BIM

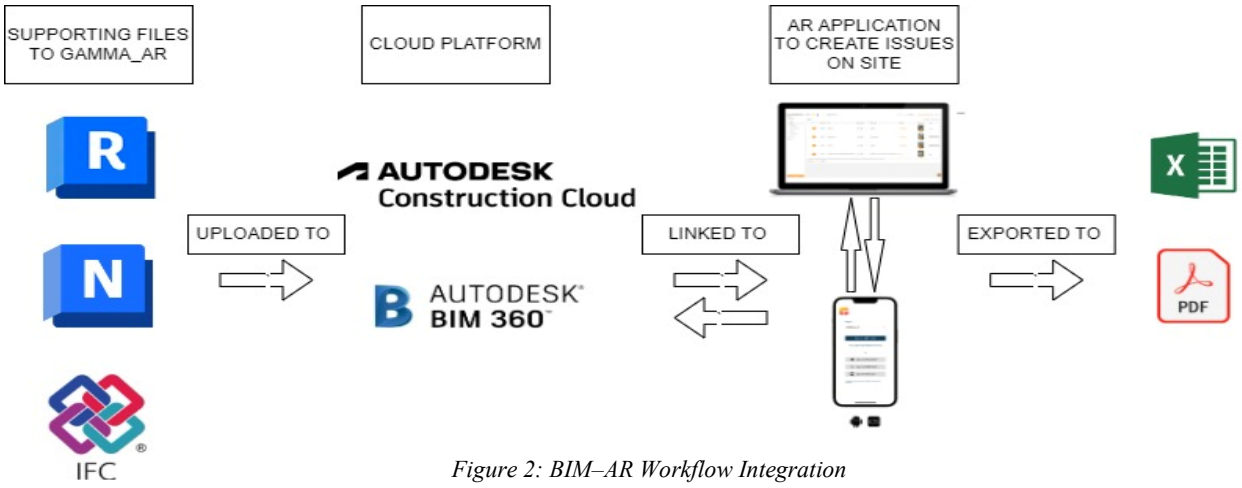


Figure 2: BIM–AR Workflow Integration

authoring and cloud-based model management to on-site AR visualisation and issue synchronisation.

Data Collection

Data collection combined direct site observations with semi-structured interviews involving members of the project team who actively used Gamma AR. Multiple site visits were conducted during which the researcher observed how site engineers and managers used the AR application for tasks such as installation verification, coordination discussions, and progress checks. Observational notes focused on alignment procedures, usage patterns, perceived benefits, and practical difficulties encountered during use.

Semi-structured interviews were conducted with six construction professionals involved in the project, including the site manager, BIM coordinator, a digital construction advisor, and site engineers. Participants were selected based on direct involvement with the Gamma AR deployment and their varied roles in on-site decision-making and BIM coordination. Interview questions explored participants' experiences with the AR tool, perceived value for construction coordination, challenges related to accuracy, usability, and workflow integration, and views on adoption and training requirements. A short questionnaire with 10 site-based staff members who had trialed the AR tool was used to supplement interview data and capture recurring themes across participants.

Data Analysis

The collected qualitative data were analysed using a thematic approach. Observation notes and interview transcripts were coded to identify recurring patterns related to (i) opportunities created by BIM–AR integration, such as enhanced visualisation, coordination, and error detection, and (ii) challenges encountered during deployment, including technical limitations, environmental constraints, and user adoption issues. The analysis focused on identifying practical trade-offs and deployment considerations rather than measuring system performance quantitatively. The overall data collection and analysis process adopted in this study is summarised in Figure 3.



Figure 3: Data collection and thematic analysis process adopted in the Gamma AR case study

Scope and Limitations

The study is based on a single case study and is not intended to provide statistically generalisable results. Instead, it offers empirically grounded insights into BIM–AR deployment under real construction conditions. The findings should therefore be interpreted as illustrative of practical opportunities and challenges, rather than definitive performance evaluations of AR technology.

Results and Discussion

The deployment of Gamma AR on the Clive Booth Student Village project provided insight into how BIM–AR integration functions under real construction conditions. The findings are organised around three themes: (i) opportunities observed during on-site use, (ii) challenges encountered during deployment, and (iii) practical trade-offs that shaped how the technology was ultimately used in practice.

Opportunities of BIM-AR Integration

The Gamma AR deployment revealed several opportunities where BIM–AR integration added value to construction coordination and decision-making. Importantly, these opportunities emerged not from automation or measurement precision, but from enhanced contextual access to BIM information within the physical site environment. One site engineer remarked, *"It's easier to understand what you're building when the model is literally on the slab. It helps reduce confusion from the drawings."* Similarly, the BIM coordinator observed, *"Having the model in the space gave the client a clearer picture — they could point out issues straight away."*

Enhanced visualisation of design intent in context

The most consistently reported opportunity was improved spatial understanding of design intent on site. By overlaying BIM models directly onto partially constructed spaces, Gamma AR enabled site personnel to visualise upcoming works in context, particularly during early fit-out stages where two-dimensional drawings were difficult to interpret. AR was frequently used to preview partition layouts, service routes, and spatial clearances prior to installation, supporting more intuitive comprehension of scope and sequence. From a Human–Data Interaction (HDI) perspective, this enhanced the *legibility* of BIM data by making abstract digital information more comprehensible in its physical context, particularly for those without prior technical training.

This finding aligns with prior research showing that AR enhances spatial cognition and design comprehension by reducing the mental effort required to translate abstract representations into physical form (Sidani et al., 2021; Chen and Xue, 2022). However, the case study suggests that these benefits are most pronounced when AR is used selectively for spatially complex tasks, rather than as a general-purpose visualisation tool. In this sense, AR acted as a cognitive aid, supporting shared understanding rather than replacing conventional drawings.

Improved coordination and communication workflows

BIM–AR integration supported stronger on-site coordination by enabling discrepancies between the BIM model and site conditions to be identified and discussed directly within the physical environment. In several observed instances, AR overlays revealed potential clashes, such as conflicts between services and structural elements that had not been immediately apparent in drawings or model reviews conducted off-site.

The ability to capture issues directly from the AR interface and synchronise them with BIM 360 shortened feedback loops between site and design teams, reducing reliance on post-hoc interpretation of photographs or annotated drawings. This supports findings from prior studies reporting that BIM–AR integration can improve coordination efficiency by anchoring discussions around a shared, spatially contextualised representation (Garbett et al., 2021; Girgin et al., 2023). Notably, the effectiveness of this workflow depended on tight integration with existing issue-management processes; AR added value by augmenting, rather than bypassing, established coordination practices.

Error reduction and quality assurance support

AR functioned as a practical quality assurance aid by enabling rapid "overlay checks" between planned and installed elements. In one example, a minor misalignment in pipework was detected when the virtual model did not align with the physical installation, prompting immediate correction. While AR was not treated as a precision measurement tool, it provided a fast and visually intuitive means of identifying potential issues early, reducing the likelihood of downstream rework.

This use of AR as a screening mechanism rather than a substitute for formal inspection or surveying was viewed positively by site staff and is consistent with prior experimental findings that AR can reduce human error in construction tasks by improving perceptual alignment between design intent and execution (Fazel and Adel, 2024; Surri and Kuncoro, 2025). The case study reinforces that AR's value in quality assurance lies in early detection and visual confirmation, rather than in guaranteeing geometric accuracy.

Time savings in issue identification and resolution

Although time savings were not formally quantified, interviewees consistently reported that AR-based issue identification reduced time spent interpreting drawings and clarifying spatial problems. The integration of issue capture within the AR environment allowed problems to be communicated with fewer clarification cycles, particularly for spatially complex issues.

Site managers estimated that certain issues were resolved in approximately half the time compared to conventional workflows, largely due to improved shared understanding rather than process automation. This aligns with findings by Wang et al. (2024), who reported that AR-supported workflows reduced selected process durations by up to 50% in comparable construction contexts. However, the case study suggests that such time savings are highly task-dependent and should be interpreted as indicative rather than guaranteed performance gains.

Stakeholder engagement and confidence

An indirect but notable opportunity was improved engagement with clients and non-technical stakeholders. During site visits, AR visualisations helped communicate layouts, progress, and design intent more effectively than drawings or screen-based models, increasing stakeholder confidence in coordination decisions. This immersive mode of communication reduced misunderstandings and supported more informed discussions around design changes.

While difficult to quantify, these effects echo prior research suggesting that AR's experiential "walkthrough" capability can enhance stakeholder engagement and support transparency in construction projects (Mutis and Ambekar, 2020; Wang et al., 2013). In the present case, AR also contributed to a perception of innovation among site staff, supporting openness to digital tools. However, such engagement benefits were contingent on usability and reliability; technical friction quickly diminished perceived value.

Challenges in BIM-AR Deployment

While the Gamma AR trial demonstrated clear value, it also revealed a range of challenges that reflect the current limitations of BIM–AR integration under live construction conditions. These challenges are discussed below, drawing on both empirical observations from the case study and findings reported in recent literature.

Model alignment and registration reliability

Achieving and maintaining reliable alignment between the BIM model and the physical environment was one of the most persistent technical challenges as shown in Figure 4. Users reported occasional drift or offset in the AR overlay, typically attributable to imprecise initial calibration, cumulative tracking error, or changes in the surrounding environment. Similar alignment instability has been widely reported in prior studies, particularly in markerless AR systems deployed on construction sites (Surri and Kuncoro, 2025; Schiavi et al., 2022).

To mitigate this issue, site staff relied on known reference features (e.g., slab edges, gridlines) and periodically rechecked alignment during use. While these workarounds were effective in maintaining acceptable visual correspondence, they introduced additional setup effort and reduced confidence in using AR for fine-grained verification tasks. Minor alignment errors also risked false interpretations, such as suggesting an installation was incorrect when the overlay itself was inaccurate, highlighting the need for more robust and transparent registration mechanisms. This reinforces findings by Cheng et al. (2020) and Girgin et al. (2023), who emphasis that registration reliability remains a critical barrier to dependable AR deployment in construction.



Figure 4: Aligning the level manually

Environmental and site condition constraints

Construction site conditions introduced several practical limitations for AR use. Bright daylight caused screen glare that reduced visibility on tablet displays, while cluttered or visually homogeneous environments occasionally disrupted feature detection and tracking stability. Unlike controlled indoor settings, lighting and surface conditions on site cannot be regulated, constraining when and where AR can be used reliably.

These challenges align with previous research noting that environmental variability significantly affects AR performance in construction contexts (Machado and Vilela, 2020; Kolaei et al., 2022). In practice, AR use was often restricted to shaded areas, interior spaces, or specific times of day to ensure acceptable visibility. This highlights the importance of hardware and visualisation modes optimised for outdoor use, as well as tracking algorithms resilient to changing lighting and surface conditions.

Hardware and usability limitations

The choice of hardware strongly shaped the AR experience. Using tablets provided accessibility and a low barrier to entry, consistent with industry preferences for familiar devices. However, prolonged use resulted in rapid battery drain and user fatigue from holding the device for extended periods. While head-mounted displays were discussed, they were considered impractical for routine site work due to known limitations such as restricted field of view, added weight, and user discomfort.

These observations are consistent with recent studies reporting that extended AR use, particularly with head-mounted devices, can cause physical discomfort and reduced situational awareness, potentially increasing safety risks on construction sites (Chen et al., 2024; Sidani et al., 2021). Although Gamma AR's compatibility with standard mobile devices is a strength, effective deployment still depends on access to relatively recent hardware with sufficient processing power and battery capacity, which introduces procurement and cost considerations.

Data management and integration frictions

Maintaining smooth BIM–AR data flow presented additional challenges. Large or highly detailed BIM models required simplification to ensure acceptable performance on mobile devices, necessitating deliberate model optimisation strategies. Furthermore, model updates introduced workflow friction: exporting and reloading revised models into Gamma AR took time and depended on reliable site connectivity.

On one occasion, limited Wi-Fi connectivity delayed synchronisation of newly logged issues with BIM 360, echoing findings from prior studies that highlight connectivity and data transfer as persistent constraints for BIM-based AR in construction environments (Cheng et al., 2020; Bhatarai et al., 2025). While Gamma AR's integration with BIM 360 was broadly effective, not all metadata were consistently accessible through the AR interface, limiting the depth of information available in the field. These findings underscore the dependence of BIM–AR workflows on robust digital infrastructure and carefully planned update strategies, including offline or pre-download options where network coverage is unreliable.

User adoption and learning curve

Beyond technical constraints, human factors emerged as a significant barrier to adoption. User acceptance varied across roles and experience levels, with initial scepticism particularly evident among staff accustomed to traditional drawings and paper-based workflows. Survey and interview feedback highlighted limited awareness of AR

capabilities, resistance to changing established practices, and concerns about implementation cost and learning effort.

Adoption improved when the technology was actively championed by site leadership and supported by hands-on demonstrations and vendor assistance. However, participants consistently noted that effective use "takes time to learn," reinforcing findings from Wang et al. (2024) and El Asmar et al. (2021) that training, managerial commitment, and usability strongly shape AR uptake in construction. These observations indicate that ease of use, intuitive interfaces, and minimal interaction overhead are critical for sustaining long-term adoption on site.

Table 1 summarises the key opportunities and challenges observed during the Gamma AR deployment, providing an overview of the empirical findings discussed in this section.

Table 1: Summary of key opportunities and challenges observed during the Gamma AR case study.

Opportunities (Benefits Observed)	Challenges (Issues Encountered)
Enhanced visualisation of design intent in real spatial context, improving comprehension of complex layouts.	Alignment and registration issues - occasional model drift requiring manual recalibration and confidence checks
Improved coordination and communication through shared AR views and faster on-site issues identification	Lighting and environment interference - screen glare and tracking instability under bright or cluttered site conditions.
Error reduction and quality assurance support by identifying discrepancies early and reducing downstream rework	Hardware limitations - battery drain, device fatigue, and dependence on relatively recent mobile hardware
Time savings in issue verification and resolution through streamlined AR-based logging and visual explanation	Data integration frictions - model preparation effort, update delays, and reliance on reliable site connectivity
Stakeholder engagement via intuitive 3D visualisation that improved confidence during site discussions	User adoption barriers - learning curve, resistance to change, and need for training and managerial support

Trade-offs and Practical Implications

The Gamma AR case study highlights several trade-offs inherent in BIM–AR deployment on live construction projects. While AR improved visualisation, coordination, and issue communication, these benefits were accompanied by additional effort related to model preparation, alignment, device management, and user training. Understanding these trade-offs is critical for determining when and how AR should be deployed to deliver net value in practice.

A central trade-off observed was productivity versus preparation. AR supported faster inspections, coordination discussions, and issue resolution once operational; however, achieving this reliability required upfront investment in model optimisation, calibration, and troubleshooting. For brief or low-complexity tasks, this overhead could outweigh the immediate benefits. In contrast, for projects characterised by spatial complexity and repeated coordination demands, the investment was more likely to pay off. A practical implication is that AR should be targeted at high-value use cases, such as scheduled coordination walks, stakeholder reviews, or verification of critical installations rather than applied indiscriminately across all site activities.

A second trade-off concerned immersion versus practicality. Although head-mounted AR systems can offer hands-free and immersive experiences, the project team preferred tablet-based AR due to comfort, familiarity, and ease of integration into everyday site routines. This suggests that the most effective AR solutions for construction are not necessarily the most immersive, but those that fit seamlessly into existing workflows. From a design perspective, construction-oriented AR tools should prioritise usability, rapid set-up, outdoor-readable displays, and one-handed operation over advanced visual realism.

The study also revealed a trade-off between accuracy and effort. High-precision alignment typically demands additional setup steps, such as multiple calibration points or physical markers, whereas faster registration can reduce confidence in positional accuracy. In this case, AR was used pragmatically as a screening and communication tool, with apparent discrepancies verified using conventional methods such as laser measurement. This approach helped maintain trust in technology while avoiding over-reliance. Design implications include the need for clearer uncertainty cues, faster recalibration mechanisms, and training that reinforces AR as a decision-support aid rather than an infallible measurement system. From a Human–Data Interaction (HDI) perspective, Gamma AR supported *agency* by enabling users to interrogate and act on real-time overlays. *Negotiability* was evident in on-site discussions where planned elements were confirmed or challenged in context. However, these affordances also raised *privacy* concerns around the passive capture of site conditions and worker activity, pointing to the need for ethical data handling and user awareness.

From an organisational perspective, the trade-off between innovation and disruption was evident. Introducing AR altered established routines for issue reporting and coordination, requiring staff to adapt to new information flows. Adoption improved when AR use was visibly supported by site leadership and introduced incrementally with clear purpose. This reinforces prior findings that managerial commitment, training, and structured protocols are decisive factors in successful AR deployment in construction (Wang et al., 2024). Without

attention to these human factors, even technically robust AR solutions risk limited uptake.

Several broader design implications emerge from the case. First, tight integration with existing common data environments is essential; AR tools should operate as extensions of established platforms rather than standalone systems. Second, interface simplicity and low interaction overhead are critical, as on-site users have limited tolerance for complex configuration. Third, future AR platforms could further extend value through real-time collaboration features, such as multi-user sessions or remote expert support, which were not available in this deployment but were strongly desired by participants.

Overall, the findings point to a pragmatic principle for BIM–AR deployment: flexibility over prescription. Gamma AR was most effective when treated as an optional support tool rather than a mandated process. Site teams used AR when it added clear value and reverted to conventional methods when these were faster or more reliable. This coexistence suggests that AR does not need to replace established practices to be useful; instead, it can enhance them when deployed selectively and thoughtfully.

Conclusion

This paper investigated the deployment of a BIM-integrated Augmented Reality (AR) solution on an active construction site through a qualitative case study of Gamma AR on a large-scale university accommodation project in the UK. By focusing on real-world use rather than controlled experimentation, the study provides empirical insight into how BIM–AR integration operates in practice and interacts with established construction workflows.

The findings demonstrate that BIM–AR integration can deliver tangible benefits when applied selectively to coordination, visualisation, and quality assurance tasks. AR-supported overlays improved spatial understanding of design intent, facilitated clearer communication between site and office teams, and enabled earlier identification of discrepancies between planned and as-built conditions. These benefits were driven not by automation or precision measurement, but by enhanced shared understanding and contextualised access to BIM information. When integrated with existing issue-management workflows, AR reduced interpretation effort and shortened feedback loops.

At the same time, the study highlights practical constraints on BIM–AR deployment, including alignment reliability, environmental conditions, hardware limitations, model maintenance effort, and user adoption factors. The findings indicate that AR is most effective when treated as a complementary tool rather than a replacement for established practices, supporting early detection and discussion while relying on conventional methods for precise verification.

The contribution of this paper lies in documenting an end-to-end BIM–AR deployment using a commercial platform under live construction conditions and articulating the trade-offs that shape practical use. While limited to a single case study and handheld AR configuration, the findings provide deployment-level evidence that can inform both practitioners considering AR adoption and developers seeking to improve BIM–AR integration. Overall, the study reinforces that realising the potential of BIM–AR depends less on technological sophistication than on alignment with site realities, user practices, and existing digital workflows.

References

Alizadehsalehi, S., Hadavi, A. and Huang, J.C. (2020). From BIM to extended reality in AEC industry. *Automation in construction*, *116*(103254), pp.1-13.

Amin, K., Mills, G. and Wilson, D. (2023) 'Key functions in BIM-based AR platforms', *Automation in Construction*, 152, 104816.

Arvikar, S., Aung, P.P.W., Cha, G. and Park, S. (2025) 'Augmented Reality-based construction site management using optimised BIM and project schedule integration', Automation in Construction, 175, 106204.

Bhatarai, R., Banihashemi, S., Shakouri, M. and Antwi-Afari, M.F. (2025) 'Integration of Augmented Reality with Building Information Modeling: design optimisation and construction rework reduction perspective', *Archives of Computational Methods in Engineering*, 32, pp. 2233–2254.

Chatsuwan, M., Moriwaki, A., Ichinose, M. and Alkhalaf, H. (2025). BIM–FM interoperability through open standards: a critical literature review. *Architecture*, *5*(3), p.74.

Chaudhari, R., Goodrum, P., Brady, N., Jones, M., Hallowell, M., Yeh, T. and Bhandari, S., (2025). Exploring the Impact of Augmented Reality on Work Performance in a Full-Scale MEP Assembly Task: Study of Industry and Novice Populations. *Journal of Construction Engineering and Management*, *151*(12), p.04025189.

Chen, K. and Xue, F. (2022) 'The renaissance of augmented reality in construction: history, present status and future directions', *Smart and Sustainable Built Environment*, 11(3), pp. 575–592.

Cheng, J.C.P., Chen, K. and Chen, W. (2020) 'State-of-the-Art Review on Mixed Reality Applications in the AECO Industry', *Journal of Construction Engineering and Management*, 146(2), Article 03119009.

Dallasega, P., Schulze, F. and Revolti, A. (2023). Augmented Reality to overcome Visual Management implementation barriers in construction: a MEP case study. *Construction management and economics*, *41*(3), pp.232-255.

El Asmar, P.G., Chalhoub, J., Ayer, S.K. and Abdallah, A.S. (2021) 'Contextualising benefits and limitations reported for augmented reality in construction research', *Journal of Information Technology in Construction (ITcon)*, 26, pp. 720–738.

Farea A, Otreba M, Ullah R, McKenna T, Carroll S, Harrington J (2026;), "Digital transformation of the AEC industry: a review of BIM implementation toolkits from leading countries". *Smart and Sustainable Built Environment*, Vol. ahead-of-print No. ahead-of-print.

Fazel, A. and Adel, A., 2024. Enhancing construction accuracy, productivity, and safety with augmented reality for timber fastening. *Automation in Construction*, *166*, p.105596.

Garbett, J., Hartley, T. and Heesom, D., 2021. A multi-user collaborative BIM-AR system to support design and construction. *Automation in Construction*, *122*, p.103487.

Gerbino, S., Cieri, L., Rainieri, C. and Fabbrocino, G. (2021). On bim interoperability via the ifc standard: An assessment from the structural engineering and design viewpoint. *Applied Sciences*, *11*(23), p.11430.

Girgin, S., Fruchter, R. and Fischer, M. (2023) 'A case study towards assessing the impact of mixed reality-based inspection and resolution of MEP issues during construction', *Journal of Information Technology in Construction*, pp. 555–570.

Jin, Y., Seo, J. and Han, S., 2026. Spatial Augmented Reality (SAR)-based 3D BIM Visualization through Semi-Automated Geometric Calibration of a Projector-Camera System. *Journal of Computational Design and Engineering*, p.qwag009.

Kolaei, A.Z., Hedayati, E., Khanzadi, M. and Amiri, G.G. (2022) 'Challenges and opportunities of augmented reality during the construction phase', *Automation in Construction*, 143, 104586.

Lee, Y.C., Eastman, C.M. and Solihin, W. (2021). Rules and validation processes for interoperable BIM data exchange. *Journal of Computational Design and Engineering*, *8*(1), pp.97-114.

Marino, E., Barbieri, L., Bruno, F. and Muzzupappa, M. (2024). Assessing user performance in augmented reality assembly guidance for industry 4.0 operators. *Computers in Industry*, *157*, p.104085.

Mortier et al. (2014) Mortier, R., Haddadi, H., Henderson, T., McAuley, D. and Crowcroft, J. (2014). *Human-Data Interaction: The Human Face of the Data-Driven Society*. SSRN / University of Cambridge.

Mutis, I. and Ambekar, A., 2020. Challenges and enablers of augmented reality technology for in situ walkthrough applications. *Journal of Information Technology in Construction*, *25*.

Pan, N.-H. and Isnaeni, N.N. (2024) 'Integration of Augmented Reality and Building Information Modeling for Enhanced Construction Inspection—A Case Study', *Buildings*, 14(3), 612.

Rezgui, H.; Ait Haddou, H.; Camilleri, G. (2025) Application of Collaborative Engineering to Manage the Work of BIM Construction Stakeholders (Design Stage). *Architecture*, *5*, 19.

Schiavi, B., Havard, V., Beddiar, K. and Baudry, D. (2022) 'BIM data flow architecture with AR/VR technologies: use cases in architecture, engineering and construction', *Automation in Construction*, 134, 104054.

Sidani, A., Dinis, F.M., Duarte, J., Sanhudo, L., Calvetti, D., Baptista, J.S., Martins, J.P. and Soeiro, A. (2021) 'Recent tools and techniques of BIM-based augmented reality: a systematic review', *Journal of Building Engineering*, 42, 102500.

Surri, D.A. and Kuncoro, T., 2025. Integration of BIM and AR in Supporting Construction Project Management Performance: A Systematic Literature Review. *G-Tech: Jurnal Teknologi Terapan*, *9*(3), pp.1289-1296.

Tan, Y., Xu, W., Chen, P. and Zhang, S. (2024) 'Building defect inspection and data management using computer vision, augmented reality, and BIM technology', *Automation in Construction*, 160, 105318.

Wang, X., Love, P. E. D., Kim, M. J., Park, C.-S., Sing, C.-P., & Hou, L. (2013). A conceptual framework for integrating building information modeling with augmented reality. *Automation in Construction*, *34*, pp.37- 44.

Wang, K., Guo, F., Zhou, R. and Qian, L., 2024. Implementation of augmented reality in BIM-enabled construction projects: a bibliometric literature review and a case study from China. *Construction Innovation*, *24*(4), pp.1085-1116.

Waqar, A., Othman, I., Hayat, S., Radu, D., Khan, M.B., Galatanu, T.F., Almujibah, H.R., Hadzima-Nyarko, M. and Benjeddou, O (2023(. Building information modeling—empowering construction projects with end-to-end life cycle management. *Buildings*, *13*(8), p.2041.

Wuni, I.Y., Abankwa, D.A., Koc, K., Adukpo, S.E. and Antwi-Afari, M.F. (2024) 'Critical barriers to the adoption of integrated digital delivery in the construction industry', *Journal of Building Engineering*, Article 108474

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

SPACE-DRIVEN KNOWLEDGE: A TOPOLOGICAL FRAMEWORK FOR SEMANTIC INFERENCE IN BUILDINGS

Isabelle Fitkau, Max Hammer, and Maximilian Sternal
Technische Universität Berlin, Germany

Abstract

Current building modeling approaches treat geometry as primary and semantics as secondary, limiting knowledge-integrated building design where spatial reasoning informs iterative design decisions. This paper presents a space-driven knowledge framework that integrates topological space partitioning with semantic triple generation, enabling immediate inference over spatial relationships. The workflow progresses through iterative stages of partitioning, algorithmic triple generation from topological queries and quantity calculations, axiomatic reasoning, and rule-based inference for requirements. A proof of concept demonstrates how storey classification, spatial adjacency, and fire safety requirements can be inferred during the modeling process using the BOT ontology extended with fire safety concepts from FiSa.

Introduction

Building design is inherently iterative (Pauwels and Bod, 2014). Regulatory requirements arise from building-level classifications to component-level specifications, depending on geometric and functional properties that emerge during design. However, current workflows separate modeling from compliance checking, meaning designers receive feedback on regulatory implications only after committing to spatial configurations.

This paper argues for a different approach in which semantic knowledge is assigned simultaneously with geometric modeling, allowing immediate reasoning about classifications and requirements. Rather than treating semantics as metadata attached after modeling geometry, we propose treating every modeling operation also as a knowledge-modeling event. When a designer partitions a building envelope into storeys, this operation should not only create geometric objects, but also instantiate semantic resources, compute quantities, and trigger inference rules that classify those objects.

The technical foundation combines two domains. Topology-aware modeling kernels, such as Topologic (Jabi and Chatzivasileiadi, 2021) and the Partition Platform (Sternal et al., 2025; Sternal, 2025; Huhnt et al., 2023), provide space partitioning operations that inherently compute spatial relationships. Semantic web technologies such as triple stores, ontologies, SPARQL, and rules provide mechanisms for representing, querying, and reasoning over knowledge. This paper presents a framework coupling these into a ”space-driven knowledge” workflow: a topology-aware kernel performing space partitioning, a triple store persisting semantic knowledge using BOT (Rasmussen et al., 2017) and FiSa (Fitkau and Hartmann, 2024), and an inference engine deriving new knowledge. The workflow proceeds through iterative cycles of **model**, **assign**, **compute**, and **infer**. Triples are generated algorithmically from topological queries, axiomatically through OWL reasoning, or through custom rules.

With spatial relationships and derived classifications explicitly represented in real-time throughout the architectural design process, changes to geometry automatically derive to updated inferences, enabling exploration of alternatives, e.g., with awareness of their regulatory consequences. Recent frameworks for the evaluation of early-phase design variants integrate spatial, structural, and environmental feedback in iterative design loops (Napps et al., 2026). Our work complements such approaches by providing the topological and semantic basis from which variant-specific classifications and requirements can be derived continuously during modeling. This paper makes two contributions: (1) a framework coupling geometric and topological modeling with semantic inference, and (2) a proof of concept demonstrating storey classification and fire safety requirement inference. We focus on the building scale while indicating how the approach extends to multi-scale contexts.

Background

First, a theoretical background introduces the main concepts of the proposed design workflow.

Design as an Iterative Process

Current CAD and BIM tools support geometric iteration and semantic attribute management, but evaluative knowledge remains disconnected from the modeling process. Requirements checking typically occurs after design decisions are made (Eastman et al., 2009; Pauwels et al., 2011), despite design being an inherently iterative, model-based reasoning process (Pauwels and Bod, 2014).

Coupling semantic inference with geometric modeling allows designers to observe how spatial configurations affect classifications and requirements during iteration (Pauwels et al., 2024). This requires information systems in which

spatial operations produce both geometric objects and semantic assertions.

Space Partitioning

Traditional geometric modeling represents shapes through boundary representations (B-rep) or constructive solid geometry (CSG), focusing on the geometry itself rather than its spatial relationships. Topology-aware modeling approaches instead emphasize the relational structure of space through the use of topologically connected data structures, creating a consistent hierarchical partition of space through subdivision (Huhnt et al., 2023). Therefore, such data structures offer direct topological queries without geometric computations (Massafra et al., 2024). However, these relationships typically remain within the geometric representation rather than being exposed as querable semantic triples (Donkers and Petrova, 2025). It should be noted that the modeling operations employed in topology-aware kernels, e.g., extrusion, slicing, and placement of primitives, are not fundamentally different from the established parametric design gestures in current CAD and BIM tools (Sternal et al., 2025).

The key distinction is not in the operators themselves but in the underlying paradigm. Many tools follow an object-centric approach, constructing components individually and validating consistency post-hoc. Space partitioning follows a space-centric approach in which the entirety of Euclidean space is subdivided into non-overlapping domains with inherently consistent topology. This guaranties manifold consistency by construction and provides a reliable geometric model for the semantic enrichment proposed in this paper.

The idea of a space-centric approach dates back to the 1970s, when soap-bubble and grid-bubble space management systems were explored (Eastman, 1999), yet adoption declined in favor of component-based tools that prioritize individual building elements over spatial coherence. Our framework addresses this gap by synchronizing topological computations with semantic knowledge graphs, going beyond what both early space management systems and current component-based tools offer.

Knowledge Management in Building Modeling

Building information workflows typically separate geometric modeling from semantic enrichment. This separation means that feedback on regulatory implications or other design related decision support arrives after spatial configurations are already established (Pauwels et al., 2024). Industry Foundation Classes (IFC) provide a unified data model that integrates geometry and semantics. However, file-based exchange treats semantic assignment as a pre-export step rather than a part of the modeling process itself. Authoring tools default to generic element types, and adding semantic knowledge requires effort disconnected from geometric manipulation (Beetz et al., 2009).

Semantic web technologies enable multiple domain-specific descriptions to coexist within a linked web of data, motivating coupling between modeling environments and semantic representations (Pauwels et al., 2012). Linked Building Data approaches represent building information as semantic graphs using the Resource Description Framework (RDF), which allows SPARQL queries and reasoning (Pauwels et al., 2011). Lightweight domain ontologies, such as the Building Topology Ontology (BOT) (Rasmussen et al., 2017), provide vocabularies for spatial structures corresponding to how designers progressively subdivide space. These representations can participate in design iterations when coupled with modeling environments.

Triple stores persist RDF data and support SPARQL queries. The deployment ranges from embedded libraries (RDFLib, RDF4J, Apache Jena) as shown by Elshani et al. (2026), to server-based endpoints (Fuseki, GraphDB) and federated architectures that distribute data between stakeholders (Werbrouck et al., 2024). Embedded approaches reduce latency for design-time integration, while federated architectures support collaboration but introduce consistency challenges during iteration.

Whereas some CAD tools are limited by restricted programmatic access to geometric and topological data through proprietary APIs, custom topology-aware kernels compute adjacency, containment, and connectivity as by-products of partitioning. When a horizontal plane subdivides a building volume into storeys, the kernel reports the resulting zones and which zones share faces or contain others. These facts map directly to BOT relationships (`bot:adjacentZone`, `bot:containsZone`), enabling triple generation from geometric operations. This provides a foundation for inference-guided design: the topology-aware kernel produces spatial relationships, the semantic model assigns and persists triples, and the inference engine derives classifications and properties. Computed quantities and extracted relationships feed classification rules that derive regulatory implications while geometry remains open to change. At the same time, data remain in their originating source.

Ontologies and Inference

Ontologies formalize domain knowledge as concept hierarchies and relationships interpretable by machines (Gruber, 1993). RDF represents knowledge as graphs of subject-predicate-object statements. OWL extends RDF with description logic semantics (Pauwels et al., 2011), which supports automated reasoning. Class membership can be inferred from property values, and property characteristics (transitivity, symmetry, inverse) derive knowledge through the graph. However, using OWL does not require utilizing its reasoning capabilities. Leveraging inference requires triggering a reasoner (Pellet, HermiT) that evaluates axioms against asserted facts. When reasoning is active, two mechanisms derive knowledge:

Axiomatic reasoning derives class memberships through necessary and sufficient conditions. OWL provides restrictions that constrain what holds for instances of a class.

Reasoners evaluate these definitions and automatically assign instances to their respective classes. This suits classifications derivable from properties, such as storey type from height and containment from transitivity. The same restrictions that enable classification also support consistency checking: a reasoner can detect when asserted facts contradict class definitions.

Rule-based inference handles conditions beyond OWL's expressivity, particularly arithmetic comparisons and conditions spanning multiple objects and properties. Custom rules can be written in rule languages, such as SWRL (Horrocks et al., 2004), Jena's rule syntax, Prolog, or native code. The mechanism often matters more than the syntax. When antecedent conditions match, consequent triples are asserted. Work on automated code checking (Eastman et al., 2009; Pauwels et al., 2011) demonstrated that regulatory logic can be encoded as executable rules in various formalisms.

We want to incorporate semantic inferences directly into the design process. We are not arguing against the continued use of rule-based applications for automatic compliance checking. However, an important distinction concerns missing information. OWL operates under the Open World Assumption (OWA), where absent statements are unknown, not false. Rather than lacking a fire stair, a building without an asserted fire stair might simply lack that information. OWL restrictions enable classification and consistency checking but cannot verify completeness. Constraint languages like SHACL operate under the Closed World Assumption (CWA), where absent information indicates a violation. Shapes verify that required properties exist and satisfy cardinality or value constraints, similar to OWL restrictions, but with different semantics suited for validation rather than classification. For iterative design, the OWA inference derives classifications and requirements from available information without treating incompleteness as an error. CWA validation fits later stages when completeness is expected before submission and permitting processes.

Our proposed framework uses OWL axioms and restrictions for classification and consistency, as well as custom rules for requirement property inference.

Method

The framework couples three components (Figure 1): a topology-aware modeling kernel, an RDF triple store, and an inference engine. Each component has distinct responsibilities and user inputs in the triple generation process.

Topology-Aware Modeling Kernel

The modeling kernel handles geometric operations: space partitioning through parametric design operations (extrude, slice, subdivide) and geometric checks. It maintains topological relationships between cells, including which faces are shared, which cells are adjacent, and which cells contain others.

The kernel exposes these relationships through a query interface. When partitioning creates new cells, the kernel reports not only the resulting geometry but also the topological facts: *Cell A and Cell B share Face F and are therefore adjacent,* which feeds the semantic model in the triple store. Example technologies include Topologic, which provides cellular complex operations via Python/C++ for non-manifold modeling, and the Partition Platform, which offers stepwise partitioning with history tracking via a custom Java kernel for manifold modeling. The presented framework is kernel-agnostic: any system that provides partitioning operations with topological relationship extraction can serve this role.

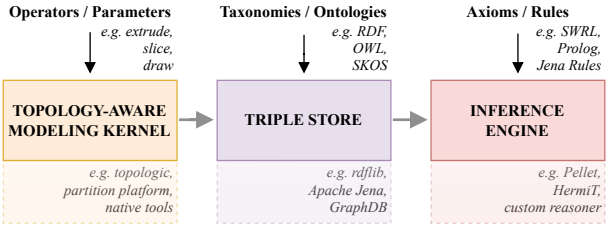


Figure 1: Framework architecture: topology-aware modeling kernel, triple store, and inference engine with their respective inputs and example technologies.

Triple Store

The triple store persists the semantic model as RDF graphs and provides query access to the knowledge base. It serves as the central repository where modeling operations generate triples, spatial relationships are stored, and inferred knowledge is materialized and maintained. RDF represents knowledge as serializable subject-predicate-object statements. Triple stores enable SPARQL queries over persisted graphs, allowing for the retrieval of objects, traversal of relationships, and filtering by properties. Updates to the triple store are managed through an inference model using a reasoner. This way, updates remain synchronized with the underlying storage, coordinating both asserted and inferred triples in their respective graphs.

Multiple named graphs can separate asserted facts (user-specified and algorithmically generated) from inferred knowledge for provenance tracking and inference validation. Implementation options range from embedded libraries to server-based systems. Embedded approaches minimize latency for design-time operations, while server-based deployments enable multi-user access and federated architectures (Werbrouck et al., 2024); however, synchronization complexity increases during iterative modeling.

Inference Engine

Inference engines derive new knowledge from asserted facts by applying logical rules and axioms over RDF triples (Eastman et al., 2009; Pauwels et al., 2011). Deductive reasoners apply description logic axioms to classify instances and derive properties through characteristics such as transitivity, symmetry, and inverse relationships. While axioms work well for property-based classifications, rule-based engines cover conditions beyond the expressiveness

of description logic. Forward chaining materializes all derivable triples upfront, while backward chaining can derive them on demand.

In building modeling, inference engines enable automated regulatory compliance checking, where geometric changes trigger the re-evaluation of derived classifications and requirements (Pauwels et al., 2024). This enables the exploration of regulatory consequences during design iterations. Domain ontologies such as ifcOWL (Beetz et al., 2009), BOT, and FiSa encode spatial relationships and component specifications that inference engines can operate on incrementally (triggered by triple store updates) or on-demand (invoked before queries). Incremental inference maintains a consistent knowledge state but incurs overhead with every update, whereas on-demand inference delays computation but may cause query latency.

Knowledge Derivation Mechanisms

Triple generation occurs through three distinct mechanisms, each suited to different knowledge derivation patterns.

Algorithmic generation produces triples directly from modeling operations and topological queries. Space partitioning operations enable object type assignment and hierarchical instantiation. Topological queries, e.g., extracting adjacency from shared faces, and geometric computations yield quantitative properties from spatial dimensions and positions. This mechanism handles knowledge derivable through direct computation on geometry and topology without requiring logical inference.

Axiomatic reasoning leverages description logic semantics to derive class membership. Reasoners evaluate class restrictions and property characteristics to materialize implicit knowledge. Classification emerges through restriction-based reasoning when instances satisfy class definitions.

Rule-based inference addresses conditions that exceeds the expressiveness of description logic. Custom rules encode arithmetic comparisons, multi-object properties, and regulatory threshold evaluation. This mechanism can handle procedural knowledge and numeric reasoning.

The framework's effectiveness derives from this separation, preserving each layer's responsibilities while leveraging the strengths of each formalism.

Space-Driven Knowledge Design Workflow

The workflow proceeds through iterative cycles, progressing to finer levels of detail (Figure 2). Architectural design development follows different spatial strategies: subdivision decomposes larger volumes into smaller ones, while aggregation composes spaces from individual units. The workflow presented here follows a subdivision strategy, progressively partitioning space from coarse to fine granularity. While aggregation-based and hybrid strategies exist in design practice, subdivision aligns naturally with the space partitioning paradigm, where the entirety of space is recursively divided into non-overlapping domains. At each stage, four activities occur: modeling (geometric partitioning), semantics (type assignment), compute (quantity extraction and topological queries), and inference (classification and reasoning). Each geometric cell created through partitioning receives a unique URI, initially asserting minimal type information. Semantic assignment then specifies object types, while computed quantities and queried topological relationships enrich the object's semantic description through subsequent workflow stages. As properties accumulate through stages, more rules apply, and further knowledge becomes inferable. The following description uses the BOT ontology extended by FiSa concepts, although this approach transfers to other ontology-based knowledge formalizations.

Stage 1: Partitioning by Terrain

The workflow begins with site context. Partitioning space into *'air'* and *'earth'* establishes the ground plane and terrain geometry. Semantically, this creates a `bot:Site` instance. Extracted quantities include terrain height, geographic coordinates, parcel boundaries, as well as relationships to adjacent zones. Inference at this stage determines the applicable regulatory framework based on geographic location through jurisdiction-specific variants, establishing the assessment basis for subsequent classifications.

Stage 2: Partitioning by Building

Within the site, further subdivision partitions 'air and earth' into building volumes. A building becomes a `bot:Building` instance with the sites containment relationship (`bot:hasBuilding`). Adjacent sites establish `bot:adjacentZone` relationships relevant to regulatory requirements. Computed quantities include gross floor area and the number of utilization units. Topological queries determine whether the building is freestanding (no adjacencies to other buildings), while geometric computation can further derive whether special building conditions prevail, such as sales premises or high-rise building classification.

Stage 3: Partitioning by Storey

Horizontal planes partition buildings into storeys. Each storey is instantiated as `bot:Storey` with hierarchical relationships (`bot:hasStorey`). Semantic assignment specifies usage types (`fisa:hasUsage`) and further defines utilization units within the building hierarchy. Computed quantities extract ceiling heights (`fisa:hasUpperCeilingHeight`) and average floor elevations relative to ground (`fisa:hasAverageHeight`). Queries determine the vertical position within the building and count the residential units per storey. These quantities enable storey classification through axiomatic reasoning, which derive storeys as `fisa:AboveGroundStorey` or `fisa:Basement` based on height and elevation properties, and further sub-classification into `fisa:UpperStorey` and `fisa:GroundFloor` according to position. Building class inference also occurs at this stage through rule-

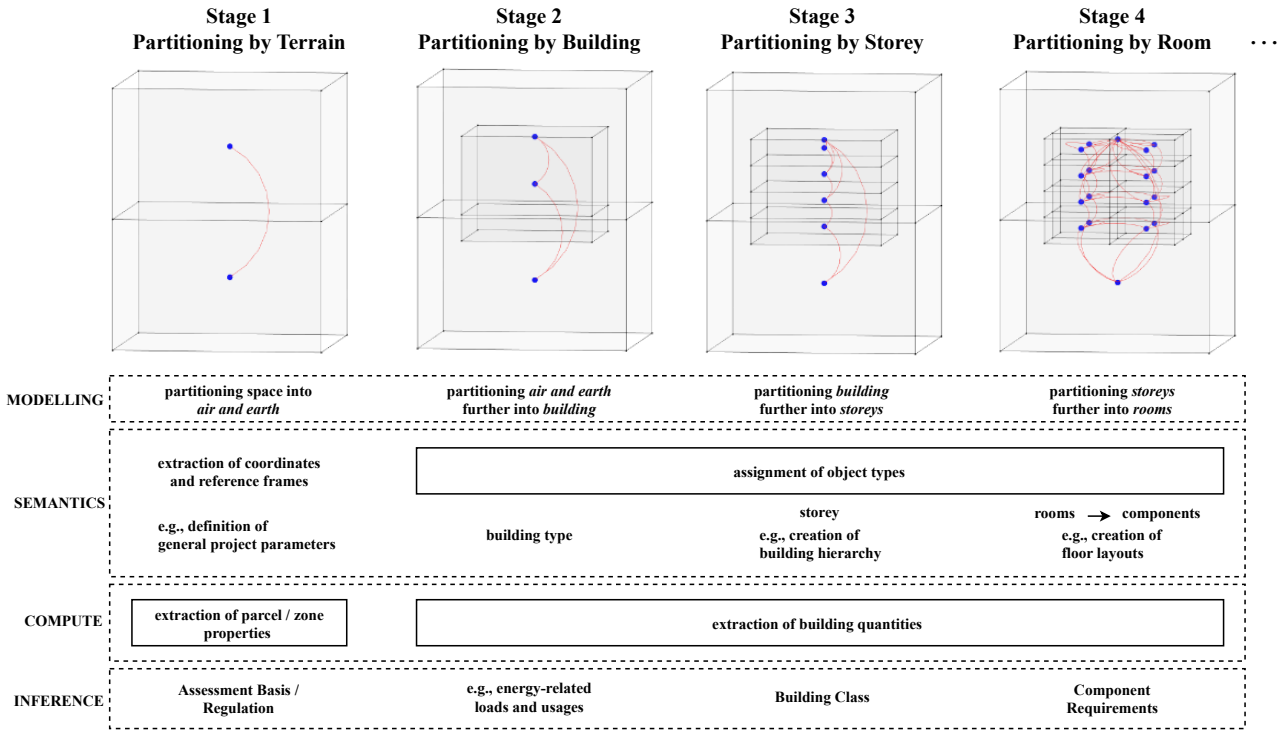


Figure 2: Iterative workflow through four stages of partitioning. Each stage involves geometric modeling, semantic assignment, quantity computation and topology querying, and inference, with results feeding back into subsequent design iterations.

based reasoning. Rules evaluate accumulated properties, such as building height, freestanding status, number of utilization units, and gross floor area, to derive a `fisa:BuildingClass`. This class inference determines component fire resistance and behavior requirements applied in subsequent stages.

Stage 4: Partitioning by Room

Further partitioning operations subdivide storeys into spaces, here also termed rooms. Each space is assigned a `bot:Space` instance. Walls separating rooms instantiate as `fisa:Wall` elements with relationships to adjacent spaces (`bot:adjacentElement`). Semantic assignment specifies space usage, necessary stairs, corridors, and other relevant types. Topological queries directly instantiate semantic knowledge, such as detecting faces with adjacency to outside zones identifies external walls (`fisa:isExternal`), while containment analysis conclude space-enclosing components (`fisa:isSpaceEnclosing`). Inference at this stage derives component requirements. Based on the building class (inferred in Stage 3), component position and properties (external, load-bearing), and adjacency relationships, rules infer the required fire resistance (`fisa:hasRequirementOfFireResistance`) and fire behavior (`fisa:hasRequirementOfFireBehaviour`) for each component.

Iterative Design Refinement

The workflow operates through feedback loops. Geometric modifications trigger quantity recomputation, which affects requirement inference. For example, when storey height changes, the system updates derived properties, re-evaluates the building class, and derives component requirements accordingly. The iterative stages, from geometry change through quantity update to classification inference and requirement derivation, provide feedback on the regulatory consequences of spatial decisions, enabling informed exploration of design alternatives.

Proof of Concept

A proof-of-concept implementation demonstrates the framework's space-driven knowledge generation capabilities. The implementation generates triples from spatial operations for space classification, as well as regulatory requirement inference at the component-level.

Implementation Overview

The Partition Platform (Sternal, 2025) used in this proof of concept serves as the topology-aware modeling kernel, providing stepwise space partitioning from coarse (air/earth, parcel, building envelope) to fine (storeys, rooms, walls) granularity. Each partitioning operation subdivides geometric cells and stores their topological relationships, e.g., the adjacency of each cell through shared faces and containment through hierarchical nesting. The platform's GUI is built using JavaFX, enabling interactive visualization and manipulation of the partitioning process. The semantic model builds on BOT for spatial structure and extends it with FiSa for fire safety concepts. The implementation uses Apache Jena 5.6.0 for RDF graph management and SPARQL querying, with Openllet 2.6.5 pro-

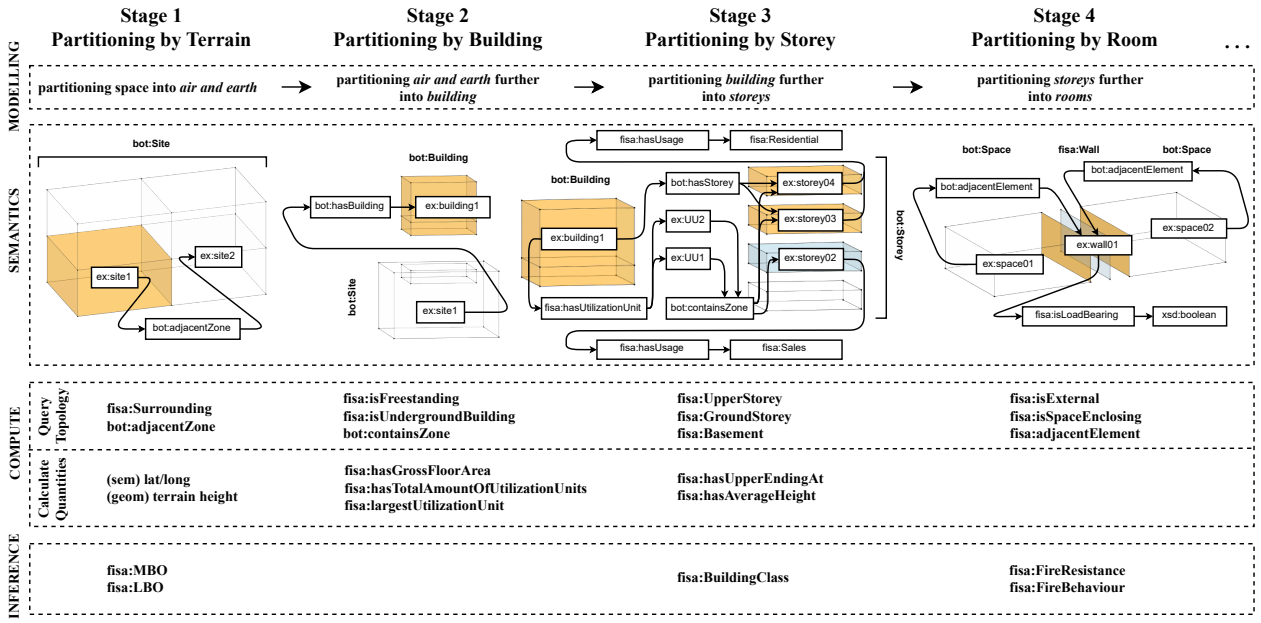


Figure 3: Proof of concept demonstrating semantic triple generation at each partitioning stage, from site context through room subdivision, with corresponding inferences.

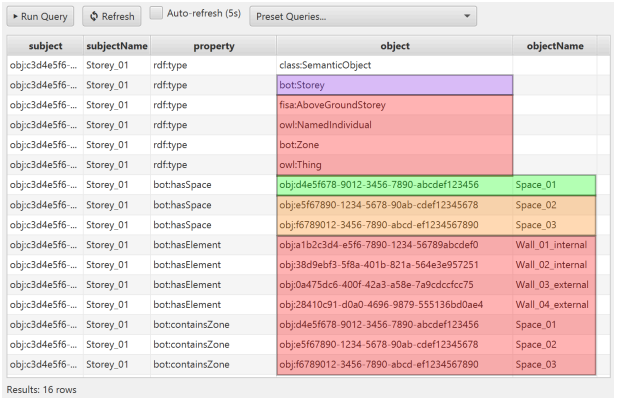
viding OWL reasoning capabilities. Algorithmic triple generation translates partitioning operations into object instantiations and BOT object properties directly in Jena's TDB2 Triple Store. Axiomatic reasoning derives classifications from geometric and topological properties. The rule-based inference engine Openllet derives knowledge written in custom SWRL rules within FiSa. Notably, the ontology was verified using the Pellet reasoner during development. The custom inference engine used in this proof of concept only implements the reasoning features required for the present scope, namely transitive property characteristics and SWRL rule execution. Apache Jena's InfModel coordinates the reasoning process, managing the interaction between the TDB2 triple store and the inference engine while maintaining synchronization between asserted and inferred triples.

Demonstration Cases

The following three cases demonstrate the triple generation process across successive partitioning stages, from storey classification (Stage 3) through spatial adjacency to space classifications and component-level inference of fire resistance requirements (Stage 4). Each case illustrates a different source of semantic enrichment, such as topological queries derived from the modeling kernel, computed properties from geometric analysis, and inferred properties from ontology reasoning (see Figure 3).

Case 1: Classification

When a storey is created by partitioning a building in the geometric model and asserted as a `bot:Storey`, properties are automatically derived from geometric analysis and topological queries, triggering classification within the FiSa storey hierarchy. As shown in Figure 4, `Space_01` is automatically classified as `fisa:AboveGroundStorey`. This classification follows from FiSa's axiom defining `fisa:AboveGroundStorey` as storeys whose floor surface has no earth contact and lies at a certain height above terrain height.



▸ Run Query | Refresh | Auto-refresh (5s) | Preset Queries...

subject	subjectName	property	object	objectName
obj:c3d4e5f6-...	Storey_01	rdf:type	class:SemanticObject	
obj:c3d4e5f6-...	Storey_01	rdf:type	bot:Storey	
obj:c3d4e5f6-...	Storey_01	rdf:type	fisa:AboveGroundStorey	
obj:c3d4e5f6-...	Storey_01	rdf:type	owl:NamedIndividual	
obj:c3d4e5f6-...	Storey_01	rdf:type	bot:Zone	
obj:c3d4e5f6-...	Storey_01	rdf:type	owl:Thing	
obj:c3d4e5f6-...	Storey_01	bot:hasSpace	obj:d4e5f678-9012-3456-7890-abcdef123456	Space_01
obj:c3d4e5f6-...	Storey_01	bot:hasSpace	obj:e5f67890-1234-5678-90ab-cdef12345678	Space_02
obj:c3d4e5f6-...	Storey_01	bot:hasSpace	obj:f6789012-3456-7890-abcd-ef1234567890	Space_03
obj:c3d4e5f6-...	Storey_01	bot:hasElement	obj:a1b2c3d4-e5f6-7890-1234-56789abcdef0	Wall_01_internal
obj:c3d4e5f6-...	Storey_01	bot:hasElement	obj:38d9ebf3-5f8a-401b-821a-564e3e957251	Wall_02_internal
obj:c3d4e5f6-...	Storey_01	bot:hasElement	obj:0a475dc6-400f-42a3-a58e-7a9cdccfcc75	Wall_03_external
obj:c3d4e5f6-...	Storey_01	bot:hasElement	obj:28410c91-d0a0-4696-9879-555136bd0ae4	Wall_04_external
obj:c3d4e5f6-...	Storey_01	bot:containsZone	obj:d4e5f678-9012-3456-7890-abcdef123456	Space_01
obj:c3d4e5f6-...	Storey_01	bot:containsZone	obj:e5f67890-1234-5678-90ab-cdef12345678	Space_02
obj:c3d4e5f6-...	Storey_01	bot:containsZone	obj:f6789012-3456-7890-abcd-ef1234567890	Space_03

Results: 16 rows

*Figure 4: Results show relationships for `Storey_01` for a SPARQL-Query for storeys in custom Database Viewer. (**purple**: user-asserted; **orange**: computed in kernel; **red**: inferred; **green**: `Space_01`)*

Case 2: Adjacency

Spatial adjacency relationships are derived by topological queries. The kernel computes `bot:adjacentZone` relations between spaces sharing boundaries, as shown for `Space_01`'s connections to `Space_02` and `Space_03`, as well as their `bot:adjacentElement` relations to the bounding walls (see Figure 5).

Case 3: Component Requirements

Fire resistance requirements are derived from custom rules. When a wall is asserted with boolean datatypes in both `fisa:isLoadBearing` and

Listing 1: Sparql-Query for Spaces and their inferred triples

```
PREFIX bot: <https://w3id.org/bot#>
PREFIX prop: <http://bitub.de/semantic/property/>

SELECT ?subject ?subjectName ?property ?object ?
    objectName
WHERE {
 GRAPH <http://bitub.de/semantic/inferred> {
   ?subject a bot:Space.
   ?subject ?property ?object .
 }
 OPTIONAL {?subject prop:name ?subjectName .}
 OPTIONAL {?object prop:name ?objectName .}
}
```

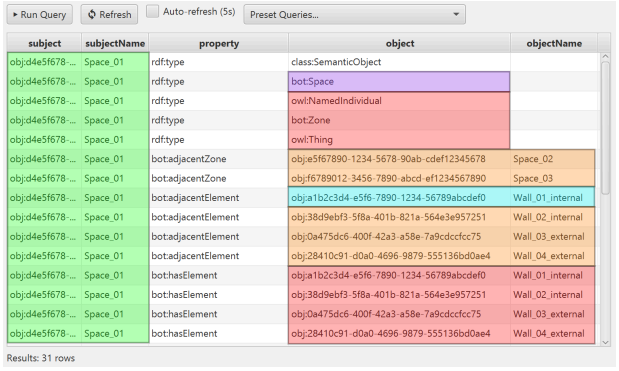


*Figure 5: Results show relationships for `Space_01` for a SPARQL Query for spaces (Listing 1) in custom Database Viewer (**purple**: user-asserted; **orange**: computed in kernel; **red**: inferred; **green**: `Space_01`; **lightblue**: `Wall_01`)*

`fisa:isSpaceEnclosing`, those rules infer the object property `fisa:hasRequirementOfFireResistance`. Consequently, `Wall_01_internal` is inferred to require `fisa:feuerhemmend` (Figure 6), following regulatory requirements for load-bearing internal walls in the considered spatial configuration.

Discussion

This paper presents a framework that couples topology-aware modeling with semantic triple generation, classification and inference, enabling real-time feedback on the regulatory implications of geometric decisions. The proof of concept demonstrates that models can be semantically enriched during design stages through the combination of topological queries, axiomatic reasoning, and custom rules.

The current implementation demonstrates feasibility rather than production readiness. Several technical limitations remain unaddressed. The modeling kernel handles straightforward partitioning operations but does not address complex parametric operations or a merging of existing geometric models. Performance at scale, involving thousands of spaces and millions of triples, remains unevaluated. The proof-of-concept implementation operates on a JavaFX 21 UI for the Database View and the Partition Platform, utilizing Apache Jena's TDB2 triple store and the Openllet reasoner, rather than being a fully integrated or web-based application.

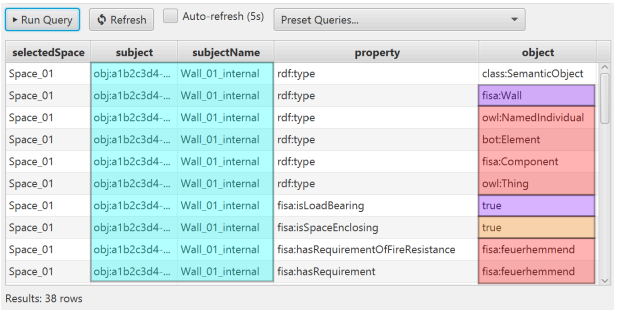


*Figure 6: Results show relationships for `Wall_01`, bounding the selected `Space_01`, in custom Database Viewer (**purple**: user-asserted; **orange**: computed in kernel; **red**: inferred; **lightblue**: `Wall_01`)*

Another trade-off concerns graph management during iteration. Updating existing graphs maintains entity identity but requires synchronization logic, while regenerating graphs sacrifices continuity for simplicity. The framework currently regenerates graphs after each modeling operation, accepting this limitation for implementation tractability. Similarly, the semantic assignment step requires further development. Requiring type assignment during modeling eliminates the generic model problem but may conflict with the exploratory early design, where flexibility takes precedence over fixed commitments. Progressive object type assignment, involving partial assignments that mature with the design, offers a possible resolution but remains unexplored.

Beyond technical constraints, questions remain about how designers will adopt this workflow. Designers who are accustomed to choosing specific types and properties at the end of modeling may resist doing so during the modeling process. Whether handling geometry and semantics at the same time demands too much attention remains untested. The framework assumes that designers benefit from immediate regulatory feedback. However, confirming this would require a carefully developed and constructed user study with designers that compare traditional workflows directly with the proposed approach.

Future work targets full integration with the Partition Platform user interface for parametric operations, which are currently under development. The inference mechanisms demonstrated here for fire safety extend to other regulatory domains where classifications derive from spatial properties and lead to component requirements. Extending the framework to multi-building and urban scales, where site adjacencies and district-level regulations apply, presents additional opportunities for space-driven knowledge approaches.

Conclusion

This paper presents a framework in which modeling operations generate geometric and topological representations and semantic knowledge simultaneously. Partitioning space creates manifold geometries and topological graphs, producing queryable semantic triples, with inferences used to derive classifications and requirements iter-

atively. The proof of concept demonstrated storey classification and the inference of fire safety requirements using the FiSa ontology, thereby establishing the foundations for workflows in which spatial reasoning informs iterative design decisions.

The methodological contribution is a staged workflow that couples topology-aware modeling with inference, applicable beyond fire safety to other regulatory domains where continuous knowledge derivation is preferred over plain validation checking.

References

Beetz, J., Van Leeuwen, J., and De Vries, B. (2009). Ifcowl: A case of transforming express schemas into ontologies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 23(1):89–101.

Donkers, A. and Petrova, E. (2025). Adding geometric functionalities to linked building data using ifcopenshell. In Proceedings of the 2025 European Conference on Computing in Construction, Porto, Portugal.

Eastman, C., Lee, J.-m., Jeong, Y.-s., and Lee, J.-k. (2009). Automatic rule-based checking of building designs. Automation in Construction, 18(8):1011–1033.

Eastman, C. M. (1999). Building Product Models: Computer Environments Supporting Design and Construction. CRC Press, Inc., USA, 1st edition.

Elshani, D., Hernandez, D., Nakhaee, A., Arrascue, A. A., Staab, S., and Wortmann, T. (2026). geof3d: Sparql geometric functions for co-designing buildings. Advanced Engineering Informatics, 71:104261.

Fitkau, I. and Hartmann, T. (2024). An ontology-based approach of automatic compliance checking for structural fire safety requirements. Advanced Engineering Informatics, 59:102314.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, 5(2):199–220.

Horrocks, I., Patel-Schneider, P., Boley, H., Tabet, S., Grosof, B., and Dean, M. (2004). Swrl: A semantic web rule language combining owl and ruleml. Technical report, W3C Member Submission 21 May 2004.

Huhnt, W., Vetter, J. Z., and Sternal, M. (2023). Space partitioning as a holistic alternative to traditional geometric modeling workflows in the aec industry. In Advances in Information Technology in Civil and Building Engineering, pages 415–426. Springer Nature Switzerland.

Jabi, W. and Chatzivasileiadi, A. (2021). Topologic: Exploring spatial reasoning through geometry, topology, and semantics. In Formal Methods in Architecture, pages 277–285. Springer International Publishing.

Massafra, A., Jabi, W., and Gulli, R. (2024). Topological bim for building performance management. Automation in Construction, 166:105628.

Napps, D., Staudt, J., Saluz, U., Chen, X., Steiner, D., Zong, C., Deghim, F., Lang, W., Geyer, P., Schnellenbach-Held, M., Petzold, F., Borrmann, A., and König, M. (2026). Evaluation of building design variants in early phases on the basis of adaptive detailing strategies. Buildings, 16(4):685.

Pauwels, P. and Bod, R. (2014). Architectural design thinking as a form of model-based reasoning. In Model-Based Reasoning in Science and Technology, volume 8, pages 583–608. Springer Berlin Heidelberg, Berlin, Heidelberg.

Pauwels, P., De Meyer, R., and Van Campenhout, J. (2012). Information system support in construction industry with semantic web technologies and/or autonomous reasoning agents. In eWork and eBusiness in Architecture, Engineering and Construction, pages 643–652. CRC Press.

Pauwels, P., van den Bersselaar, E., and Verhelst, L. (2024). Validation of technical requirements for a bim model using semantic web technologies. Advanced Engineering Informatics, 60:102426.

Pauwels, P., Van Deursen, D., Verstraeten, R., De Roo, J., De Meyer, R., Van De Walle, R., and Van Campenhout, J. (2011). A semantic rule checking environment for building performance checking. Automation in Construction, 20(5):506–518.

Rasmussen, M. H., Pauwels, P., Hviid, C. A., and Karlshøj, J. (2017). Proposing a central aec ontology that allows for domain specific extensions. In 2017 Lean and Computing in Construction Congress, volume 1, pages 237–244.

Sternal, M. (2025). Polyhedral partition of the euclidean space for consistent spatial modeling of the built environment. PhD thesis, Technische Universität Berlin.

Sternal, M., Oswald, S., and Huhnt, W. (2025). First investigations of advanced operators for consistent spatial modeling the built environment. In ICCCBE 2024, Lecture Notes in Civil Engineering, volume 628, pages 54–67. Springer.

Werbrouck, J., Pauwels, P., Beetz, J., Verborgh, R., and Mannens, E. (2024). Consolid: A federated ecosystem for heterogeneous multi-stakeholder projects. Semantic Web, 15(2):429–460.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

THE POTENTIAL OF DIGITAL TECHNOLOGY TO ENGAGE CHILDREN WITH HOSPITAL EXPERIENCE AND DESIGN

Laura Maftei
University of Reading, Reading, UK

Abstract

This study examines how virtual reality (VR) supported co-design methods can help recover children's perspectives in shaping hospital environments. Building on NHS priorities for child-centred healthcare design, the project developed immersive digital replicas of key hospital spaces through 360° video, 3D models, and VR simulations. These were tested with children from local schools to explore their perceptions, reduce hospital related anxiety, and gather design preferences. Findings show that VR supported children engagement, familiarity with clinical spaces, and insights into sound, colour, layout, and wayfinding. The study shows the potential of VR as a participatory tool for inclusive paediatric healthcare design.

Introduction

Creating strategies to educate and prepare children for hospital visits as patients and designing hospital services for children from the child's point of view has been for some time part of the NHS framework (Department of Health, 2003; NHS England, 2014). Recently, increased recognition has been given to the requirement to include the views of child patients in the planning of health care services so that hospitals can be designed to enhance future experience. However, the voices of children remain often under-represented or represented through adult proxies in projects to design, develop and improve healthcare service delivery. Co-design methodologies (Bratteteig and Wagner, 2016; Luck 2018) and digital technology such as VR bring the potential to support user empowerment and participation in the design of buildings and services (Sateei et al., 2025; Roupé et al 2020; Maftei et al. 2018).

This study draws on a co-design approach to explore the use of digital technology such as VR (virtual reality) for engaging children with hospital spaces and experiences, and recovering children's perspectives of what constitutes ideal hospital design. The study is based on a community engagement project drawn on collaboration between the author's University, the local NHS hospital and local schools. The project aimed to: 1) develop a friendly but realistic immersive experience of hospital through digital simulations (including 3D photographic visualisation and surround sound); 2) examine the use of these virtual simulations to familiarise children with hospital spaces they might encounter, and to reduce the apprehension experienced by children as reported anecdotally by the hospital; and 3) use these virtual simulations to understand children's perceptions and preferences to inform the co-design of children hospital spaces.

In collaboration with the local hospital including the Children Strategy, Health Inequalities, and Patients Experience teams, we identified that the project would be particularly useful in understanding the experience of children with special educational needs and disabilities. We ran a demonstration session with the hospital Youth Forum, a group of 16-24 year old young adults who work with the hospital to improve services for young patients.

We co-developed digital replicas of three children's hospital spaces based on advice from the hospital team. This included a journey from the car park to main reception, the children's clinic waiting area, and the X-Ray room. Using a 360° video camera different types of digital replicas were created such as immersive VR films, 2D films, and interactive 3D models developed with laser scanning techniques.

The digital replicas were shown to children in three local schools, including the Junior Carers (an initiative to engage primary school children with the hospital and healthcare services). They were able to experience the spaces virtually in a classroom environment on VR headsets, iPads and laptops. The schools were chosen to represent a diverse range of children from different backgrounds, some with special educational needs and disabilities. The project team asked children for their thoughts and ideas, while showing them what happens in hospital. Through a series of fun story-telling classes developed with teachers and the hospital, the children could ask questions and tell what they think, 'before-they-go' and 'after-they've-been' to hospital (in the virtual simulations), then what they would like to see in the future.

In workshops conducted at the schools, children were able to see parts of the hospital they had not seen before, and reported that they would not be scared if they needed to visit them. They provided useful insights into the design and experience of children's hospital spaces talking about sounds, colours, lighting, seating areas, waiting, and wayfinding. They also enjoyed the experience of using digital technologies, especially the VR headsets and often said it helped them to feel immersed in the space in a 'very real' way.

This paper reports on this project and takes a focus on the participating children's perspectives on the technology and of the hospital virtual simulations for engaging children with hospital experience and co-design.

Background and related literature

Hospitalisation is frequently accompanied by elevated anxiety in children, arising from disruption to routine, exposure to unfamiliar sensory cues, and shaped by the experience of the hospital environment (Signorelli et al., 2023). The design literature similarly recognises that paediatric environments can either amplify or reduce stress, with supportive design elements (e.g., positive distraction, sense of control, and social support) linked to improved affective outcomes for young patients (Sathyanarayanan & Caldas, 2025). Yet, despite the centrality of child experience in paediatric care, children's perspectives remain under represented in both healthcare research and design decision making. Methodological and structural barriers often lead to reliance on adult proxies, resulting in partial accounts of children's priorities.

Co-design and participatory design (PD) with children have gained traction within healthcare facilities planning, but adoption remains uneven and methods are often limited to short, one off sessions that do not translate into buildable specifications (Sathyanarayanan & Caldas, 2023; Schepers et al., 2020).

Architectural engineering and construction management research indicates that conventional tools (plans, renders, physical mock-ups) can fail to capture embodied, affective, and developmental nuances of paediatric users, thereby constraining the fidelity of user requirements that propagate through briefing, design development, and delivery (Sfandyarifard & Tzortzopoulos, 2011; Jurk, 2023).

Against this backdrop, immersive digital technologies—particularly virtual reality (VR)—present the potential to both alleviate anxiety and support co-design processes. On the clinical side, VR can reduce pre procedural anxiety and improve the hospital experience for children and families (Oh et al., 2023; Hu et al., 2024). On the design/construction side, studies in the architecture/engineering domains indicate that immersive HMD VR may enhance stakeholders' spatial understanding, sightline appraisal, logistical flow testing, and identification of latent design conflicts. When VR supports interaction and multi user collaboration, end users can iteratively adjust layouts in real time, accelerating decision cycles from weeks to hours (Sateei et al., 2025). In combination with other digital technologies, VR can support participatory design, user empowerment and better management of user requirements (Maftei et al., 2018). Other healthcare focused AEC work further explores the role of VR for simulating "in between" spaces (waiting rooms, corridors) and diverse user conditions to inform design decisions that address stress and inclusivity, thereby improving experiential quality where anxiety often accumulates (Harkness et al., 2018). For paediatrics specifically, emerging research demonstrates how immersive tools can enable child centred engagement beyond verbal self-reports. For example, architectural experiments pairing VR with eye tracking and physiological measures capture moment to moment emotional responses to spatial variables in paediatric rooms, offering developmentally appropriate pathways to elicit preferences about privacy, control, and positive distraction (Sathyanarayanan & Caldas, 2024). These methods address known shortcomings of traditional PD with children—notably, the difficulty of translating tacit or affective responses into design parameters—by supporting evidence based design (Jurk, 2023; Sateei et al., 2025).

In construction management terms, the capacity of immersive technology such as VR to host multi disciplinary, multi user reviews directly within a navigable, to scale prototype is salient for briefing, options appraisal, and stakeholder sign off—especially when paediatric participants may struggle to read drawings or comprehend scale (Sateei et al., 2025; Sfandyarifard & Tzortzopoulos, 2011). Finally, children's co-design literature highlights the need for deeper participation models—moving beyond episodic ideation toward ongoing roles in defining processes and decisions—suggesting that immersive platforms can enable such "deep participation" by making options experientially comparable and discussions evidence rich (Kheirollahi et al., 2021; Schepers et al., 2020).

Overall, this literature indicates that integrating digital technology such as VR within participatory design frameworks can both mitigate hospital related anxiety and amplify children's voices in the planning and delivery of paediatric facilities. This study builds on this literature by taking a qualitative interpretivist approach to explore how digital technology such as VR can support engaging children with hospital experience and design.

Methods

The study followed a qualitative interpretivist research design consistent with established participatory and co-design protocols used in child-centred healthcare design research (Bratteteig & Wagner, 2016; Schepers et al., 2020). This approach emphasises eliciting children's lived experiences through multimodal, developmentally appropriate formats, aligning with guidance on participatory methods for child participants in both design research and architecture/engineering contexts (Sfandyarifard & Tzortzopoulos, 2011; Sathyanarayanan & Caldas, 2023).

In line with this approach, we adopted a co-design approach to explore children's perspectives on the technology and the hospital virtual simulations for engaging children with hospital experience and co-design. The study aimed to develop a friendly but realistic immersive experience of hospital through digital simulations (including surround image and sound), to familiarise children with hospital spaces they might encounter, and to understand children's perceptions and preferences to inform the co-design of children hospital spaces.

Technology used and content creation

The equipment used comprised an Insta360° surround image and sound camera, Oculus/Meta Quest VR headsets, a 2D camera, a laser scanner, iPads, and laptops. In the first phase, the 360° camera and headsets were tested by filming the author's department building. Based on these tests and demo sessions with hospital collaborators and the Youth Forum (YF), we decided to augment 360° recordings with 2D films and laser scanned interactive models for creating journeys and higher interactivity with the virtual models. The final hospital simulations recreated a child patient's visit from the car park to main reception, through the children's clinic waiting area, and into the X ray room. These virtual simulations were co-created with volunteer actors (child patients, parents, hospital staff).

Participating groups

A purposive sampling strategy was adopted to involve groups for whom the intervention would be most relevant. In collaboration with the local NHS hospital (Community Engagement, Health Inequalities, Patient Experience and Children Strategy teams), we prioritised groups for whom the intervention might be particularly useful, notably children with Special Educational Needs and Disabilities (SEND). This sampling approach aligns with participatory design guidance encouraging recruitment of participants with lived or contextual experience relevant to the design domain (Bratteteig & Wagner, 2016; Schepers et al., 2020).

Following this sampling strategy, we worked with three schools: two mainstream primary schools in the Junior Carers programme (about 30 children total, ~10 years old, split into two workshops, W1 and W2) and one highly special educational needs school (12 children, 13–16 years, mixed disabilities and neurodivergence; W3). The Junior Carers children had ongoing, non-patient engagement with the hospital. The SEND group included children with frequent or extended hospital visits. This sampling strategy mirrors participatory approaches in healthcare design that emphasise working with children who are either direct stakeholders or frequent service users (Sfandyarifard & Tzortzopoulos, 2011).

Workshop design and flow

Each workshop followed a structured sequence. First, children developed short written stories about hospital "before" the virtual experience. Next, they experienced the simulations across three display modalities—VR headset, iPad, and laptop—rotating through sub areas of the room arranged to follow the hospital journey. This was followed by mini focus groups (≈30 minutes per subgroup) and a whole group debrief (≈15 minutes). In W1 and W2 (≈1.5 hours each), groups of ~15 were divided into subgroups of five to balance group dynamics and avoid unification and standardisation of opinions, while remaining practical for facilitation time. In W3 (~2.5 hours), 12 children worked in pairs, with ~1h40m hands-on time across devices, followed by a ~30 minute whole group debrief.



Figure 1: Image from the school-based workshops

This workshop design followed established child-centred co-design procedures. Warm up activities (in our case written "before" stories) bring the potential to activate prior knowledge and reduce cognitive load, a method routinely employed in PD with children (Schepers et al., 2020). Multimodal interaction -in our case with three display types (VR headset, iPad, laptop)- has been recognized to be useful for inclusive technology mediated engagement (Sathyanarayanan & Caldas, 2023; Roupé et al., 2020). Organizing small-group mini focus groups using interactive prompts reflects PD protocols that minimise peer influence and support diverse participation (Bratteteig & Wagner, 2016). Whole-group debriefs as employed in our study are consistent with iterative reflection cycles common in co-design and evidence-based design (Sateei et al., 2025; Jurk, 2023). Subgroup sizes (in our case, 5 in W1/W2 and pairs in W3) were selected to align with guidance on balancing conversational richness with facilitation manageability (Schepers et al., 2020).

Data collection and analysis

In line with qualitative and PD research methods of data collection, we collected: (i) written "before" stories; (ii) audio recorded mini focus groups on "after" perceptions, including what children would like to change; and (iii) whole group debriefs. Interview prompts focussed on likes/dislikes of hospital experience and space/design before and after the simulations, and explicitly invited reflections on the technology and films (in terms of usefulness for potentially helping other children familiarise with the hospital).The written stories provided pre-experience framing, aligned with narrative elicitation protocols used in paediatric design studies (Sathyanarayanan & Caldas, 2023). The audio recorded

mini focus groups drew on semi structured thematic protocols common in architectural PD research (Sfandyarifard & Tzortzopoulos, 2011; Jurk, 2023). The whole group debriefs enabled collective reflection, frequently recommended in co-design workflows (Schepers et al., 2020). Prompts were designed to capture perceptual, emotional, and spatial responses—an approach supported by VR assisted evidence-based design methods (Sateei et al., 2025).

Audio recordings were transcribed and thematically analysed. In this paper, particular attention is given to the analytical themes around the technology perception and use (device preference; immersion vs distraction; usability; comfort and accessibility) and perceived usefulness (familiarisation, anxiety reduction, wayfinding confidence).

The study complied with University ethics and hospital Information Governance approvals.

Analysis and findings

Engagement, preference, and perceived usefulness

Across the Junior Carers workshops, many children identified VR headsets as the preferred display, often highlighting the 360° look around affordance: "I liked it when we could look around through the X ray room" (Participating Child in W2). The interactivity of the laser scanned model was also appreciated—"I liked the one where you can move around… so you can, yeah, feel like you're there, instead of watching someone else do something" (Participating Child in W1)—suggesting agency and navigation mattered at least as much as immersion. Children repeatedly proposed the simulations as tools to prepare and reassure peers: the films "can help the children about knowing… what's going to happen and calm, cool them down… if they did know what was going to happen they're like, 'all right, it's going to be fine…' I think it's good to let them know what's going to happen" (Participating Child in W1). Some noted the simulations resembled real visits: "I have been to X ray before… It was the same as in the films" (Participating Child, W2).

SEND participating children and school staff framed the simulations as a useful alternative/complement to existing preparation practices (e.g., social stories, photos), emphasising choice of device: "Some people need the VR headset, but some people don't need it… it could be on a phone… an iPad… or a VR [headset]" (Participating Child, W3). For this group, the value of multiple modalities (2D, 3D, interactive, non-interactive) was explicit, reinforcing that iPads and laptops were "as useful as the VR headset" (Participating Child, W3), and sometimes more comfortable or accessible.

Immersion, novelty, and distraction

The same features that made VR compelling could disrupt focus. In W2, the headset became the attraction of the workshop, with competition for turns—"How did she have the long go? She had longer. I'd like to have another go too" (Participating Child in W2)—shifting attention away from hospital design reflections. Some children focussed on production technicalities (e.g., people disappearing and reappearing in Matterport style models) rather than spatial qualities. Others encountered navigation issues: "I didn't really get to see much of the hospital, I just went straight into this other world, and I didn't get any time in the actual hospital" (Participating Child in W2). These moments illustrate how medium driven attention capture can overshadow the message without allowing the potential to support the focus of the workshop.

Perceived usefulness: familiarity with clinical processes and spaces

Children engaged closely with clinical procedures—notably X ray—and reported recalibrated expectations from "scary" to "not scary." Seeing "what they're doing" and "what the facilities are about each room" was valued, with previewing activities in medical rooms recommended to build understanding (Participating Child in W1). This aligns with comments that simulations may reduce anxiety by clarifying who does what, where parents wait, and how the process unfolds. In Junior Carers groups, prior visits (e.g., cardiology) blended with the virtual encounter (X ray) to form an overall familiarity that children found useful.

Comfort, accessibility, and session design (SEND insights)

For SEND participants, personalisation was key. Some found VR engaging "partly because of its novelty," while others were uninterested—"I'll never get used to it… you have to want to do it" (Participating Child in W3). Practical recommendations included shorter VR experiences to avoid dizziness and pre familiarisation: "with VR you have to actually get used to it" (Participating Child in W3). Children also articulated desired technical refinements—"I wish the lens didn't have to be circular, wish I could fully immerse myself"—and affirmed that iPads/phones were equally useful, reinforcing the need for optional pathways through the content.

Children's design insights through digitally mediated participation

Across workshops, children indicated design perceptions (e.g., sounds, light, wayfinding, waiting), often proposing adjustments —quieter/ambient soundscapes; more colour; the provision of more engaging activities in the waiting area ("add iPads and books for 8–10 years old"). These place-based concerns often interacted with technology use: when audio in films included a child crying, some found it disturbing; conversely, appropriate sound design enhanced presence. For SEND participants, discussions expanded to outdoor/open air spaces, polysensory seating (softer, "hugging" seats alongside standard chairs), handrails along long routes, and strategies to mitigate waiting time challenges for autistic children who cannot physically wait. Technology was thus perceived not just as a medium, but as part of a broader toolkit and methodology for inclusive preparation and design feedback.

Discussion

Using digital technology to improve the design and children's experience of hospitals?

The technology was widely perceived as useful for engaging children with hospital environments, increasing familiarity, and "calming worries" before visits. Interestingly, the same features that make VR compelling—immersion and novelty—could also be disruptive, pulling focus away from the workshop's purpose of discussing hospital experience and design. Instances of children competing for VR time, becoming preoccupied with production technicalities, or getting "lost" in navigation illustrate how the medium may overshadow the message/purpose if not carefully used. These observations point to the importance of facilitation strategies (e.g., clear turn-taking, guided tasks, and wayfinding prompts within the VR interface) to keep attention on the spatial and experiential content rather than the device itself. At the same time, the SEND group's emphasis on choice demonstrates that, although VR was preferred by some, other children found iPads and laptops equally useful, and in some cases preferable for comfort, accessibility, or sustained focus. Designing technology-mediated participation should therefore provide multiple display options, flexible session lengths, and staged familiarisation.

Overall, the findings show that VR-supported digital simulations can meaningfully engage children in exploring hospital environments, yet the immersive qualities of VR can both support and hinder design-focused dialogue. While many participants valued the 360° affordance and sense of presence, others became preoccupied with the novelty of the device or encountered navigation challenges. These dual effects highlight the importance of structured facilitation—guided tours, turn taking protocols, and clear task prompts—to ensure that immersion enhances rather than distracts from design discussion.

Across all workshops, children articulated design-relevant preferences relating to colour, sound, lighting, waiting experiences, and wayfinding. SEND participants emphasised sensory comfort, predictable routes, and the usefulness of multiple device modalities. These observations demonstrate how multimodal, digitally enabled simulations can reveal tacit experiential knowledge difficult to express using traditional PD tools.

Implications for hospital design

The study contributes evidence that VR-supported co-design can uncover child centred insights into hospital environments. Children's suggestions—ranging from quieter soundscapes and clearer wayfinding to softer seating and reduced visual clutter—align with recognised paediatric design principles and offer specific, actionable recommendations for design teams. As shown in our study, interactive models and VR combined with 2D animated options, also support accessibility and differentiated sensory needs. Overall, the findings demonstrate how VR-supported sessions can reveal design aspects that traditional participatory tools may not capture.

Relation to Prior Literature

The findings support AEC research indicating that immersive VR enhances spatial understanding and supports stakeholder dialogue (Roupé et al., 2020; Sateei et al., 2025). At the same time, the need to manage distraction aligns with literature emphasising careful facilitation in participatory sessions with children (Sfandyarifard & Tzortzopoulos, 2011). The multimodal digital simulations preference of SEND children extends emerging work calling for flexible, accessible co-design tools in paediatric environments.

Limitations

Many participating children—particularly those in the Junior Carers programme and the SEND school—had prior or frequent hospital experience. In combination with the digital hospital simulations, this familiarity enabled children's rich insights of hospital experience and design, and it supported useful background for the participating children assessment of the realism of the hospital virtual simulations. However, this familiarity may have influenced the participating children's perceptions of the simulations, their reported levels of anxiety, and the specificity of their design suggestions. Future work could include children with little or no prior hospital exposure to explore whether the technology supports first-time preparation differently.

Conclusions and Future Work

This study shows that multimodal digital simulations—360° VR films, 2D videos, and interactive laser-scanned models—can meaningfully engage children, including SEND participants, in understanding hospital environments and expressing design-relevant preferences. VR supported participation and familiarisation, yet immersion also introduced distractions, highlighting the need for clear facilitation, guided navigation, and multimodal display options.

As in all research, there are limitations to this study. Many child participants had prior hospital experience, which may have shaped perceptions and reduced anxiety. Future work could include children with little or no prior hospital experience to examine whether the technology supports first-time preparation differently. Group-based school workshops may have also limited the depth of individual feedback. Relatedly, future work might focus on individual sessions to capture more fully personal feedback from child participants. Further avenues for future work may also explore iterative co-design with children, and may include longitudinal studies examining anxiety reduction ahead of real hospital visits.

Acknowledgments

The author gratefully acknowledges the support of the University of Reading for funding this project. Special acknowledgements are made to the local NHS hospital (Royal Berkshire NHS Foundation Trust), and the local schools teachers and children for their collaboration in

this project. Special acknowledgements are also made to Dr Ian Ewart, the co-investigator on this project, for the support and inspirational drive on the study.

References

Bratteteig, Tone and Wagner, Ina (2016). Unpacking the Notion of Participation in Participatory Design. Computer Supported Cooperative Work (CSCW), 25(6). DOI: 10.1007/s10606-016-9259-4. Springer. PISSN: 1573-7551. pp. 425-475.

Department of Health (2003) Getting the right start: National Service Framework for Children, Standard for Hospital Services. London: Department of Health.

Harkness, R., McIntosh, J. & Marques, B. (2018) 'Using virtual reality and participatory processes to design interstitial healthcare places', in Engaging Architectural Science: 52nd International Conference of the Architectural Science Association. RMIT University, pp. 657–664.archscience

Hu, Z., Yao, J., He, L., Li, X. & Guo, Y. (2024) 'The impact of virtual reality exposure on anxiety and pain levels in pediatric patients: a systematic review and meta analysis', Journal of Pediatric Nursing, 78, e364–e374.https://doi.org/10.1016/j.pedn.2024.07.027.pediatricnursing

Jurk, L.A. (2023) 'Documenting end user needs in an interactive virtual reality prototyping environment for a VR PD approach in architecture', EKSIG 2023: Experiential Knowledge Conference. Design Research Society. https://doi.org/10.21606/eksig2023.103.designresearchsociety

Kheirollahi, M., Behnia, B., Sahragard, M. & Soltanifar, A. (2021) 'A conceptual model for children's participation in the architectural design process: exploring deep participation', Architectural Engineering and Design Management. https://doi.org/10.1080/17452007.2021.1970505.academia

Luck, R. (2018) 'Inclusive design and making in practice: bringing bodily experience into closer contact with making', Design Studies, 54, pp. 96–119.

Maftei, L., Roupé, M., Johansson, M. and Harty, C. (2018) Power plays: designers, users and interactive digital technology. In: 34th Annual ARCOM Conference, 3-5 September 2018, Belfast, UK.

NHS England (2014) Health Building Note 23: Hospital accommodation for children and young people. Available at: https://www.england.nhs.uk/publication/designing-hospital-accommodation-for-children-hbn-23/

Oh, N., Parrish, N., Lee, I.W., Temple, S., Perkins, O. & Kokkinakis, M. (2023) 'Using virtual reality to reduce anxiety and improve hospital experience in paediatric orthopaedic patients and their parents', Children, 10(8), 1409. https://doi.org/10.3390/children10081409.mdpi

Roupé, M., Johansson, M., Maftei, L., Lundstedt, R., & Viklund-Tallgren, M. (2020). Virtual Collaborative Design Environment: Supporting Seamless Integration of Multitouch Table and Immersive VR. Journal of Construction Engineering and Management, 146(12), 04020132. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001935

Sateei, S., Roupé, M. & Johansson, M. (2025) 'From informative to co design: the role of immersive virtual reality for user involvement in healthcare facility design', Journal of Information Technology in Construction (ITcon), 30, Article 32. https://doi.org/10.36680/j.itcon.2025.032.itcon

Sathyanarayanan, H. & Caldas, L. (2023) 'Co designing with children: innovating patient engagement and participation in pediatric healthcare design research with immersive technology and affective interactions', The Academy Journal of the AIA Academy of Architecture for Health, 24, 30–47.

Sathyanarayanan, H. & Caldas, L. (2024) 'Pediatric inpatient room design: insights from studies using VR, eye tracking and biofeedback', Proceedings of the XLab Annual Conference. University of California, Berkeley.

Schepers, S., Schoffelen, J., Zaman, B. & Dreessen, K. (2020) 'Children's roles in participatory design processes: making the role of process designer "work"', IxD&A – Interaction Design & Architecture(s) Journal, 41, 130–154.ixdea

Sfandyarifard, E. & Tzortzopoulos, P. (2011) 'Children's hospital design through participatory approaches: a methodological perspective', in COBRA 2011 – RICS Construction and Property Conference Proceedings. Royal Institution of Chartered Surveyors, pp. 1022–1031.

Signorelli, C. et al. (2023) 'A review of creative play interventions to improve children's hospital experience and wellbeing', Hospital Pediatrics, 13(11), e355- e364

2026 European Conference on Computing in Construction
Corfu, Greece
July 1215, 2026

LOW-CODE TOOLCHAIN DEVELOPMENT TO ACCELERATE DATA-DRIVEN ENERGY MODELLING

Mariam Ali[1], Cathal Hoare[1], Gabriel Oyeyemi[1], Divyanshu Sood[2], Tiziana Margaria[1] and James O'Donnell[3]
[1]Department of Computer Science and Information Systems, University of Limerick, Ireland
[2]School of Material and Mechanical Engineering, University College Dublin, Ireland
[3]School of Engineering, University of Limerick, Ireland

Abstract

Building operations account for 30\% of global energy consumption, necessitating performant, multi-scale modelling to optimize carbon reduction. While Machine Learning (ML) offers significant promise, significant issues prevent its full adoption. This paper proposes a specific Low-Code/No-Code (LCNC) approach as a solution to democratize analytics software development, allowing stakeholders to focus on their operation needs while removing the programming syntax barrier. We show how we use the WAVE platform, a cloud-based, collaborative LCNC platform based on formal models, for the graphical composition of workflows for multi-scale energy modelling. A LCNC solution to national scale energy modelling is described and compared to traditional implementations.

Introduction

Building operations accounted for 30% of global energy consumption and 27% of energy-sector emissions in 2021 (Skillington et al., 2022). Data-driven decisions are now vital to address the issues of global warming, building comfort and stock availability (Katavoutas et al., 2022). Developing performant methods of modelling and analyzing the built environment across multiple scales, from individual buildings to national building stocks, is essential. Machine Learning (ML) offers a reliable and affordable mechanism to achieve this (Manfren et al., 2022). By fusing various information sources, such as building data, energy infrastructure and population data, and applying ML, planners can perform sophisticated analyses, to identify energy performance trends and optimize large-scale carbon reduction strategies (Ali et al., 2024).

Creating data-driven models using machine learning is non-trivial, particularly within the built environment where data are often fragmented and heterogeneous. As in any domain, a machine learning deployment workflow consists of four main stages: data management, model learning, model verification, and model deployment (Paleyes et al., 2022). In the context of energy modelling, these stages are complicated by the need to integrate disparate sources such as building models, demographic data and utility topologies.

Each stage requires a specific intersection of practitioner skill sets. For instance, data management in this sector requires not only data cleaning but an understanding of physical building properties to ensure that feature selection is based on domain fundamentals. Stakeholders in residential energy modelling report a significant AI Skills Gap (OECD, 2025). This is not merely a lack of coding ability, but a shortage of professionals who possess both energy-modelling knowledge and AI technical proficiency. For SMEs and Government agencies, this gap creates a bottleneck where the potential of data remains unfulfilled, leading to a disadvantage as they plan carbon reduction strategies (Muminova et al., 2024).

Low-Code/No-Code (LCNC) is a visual programming paradigm where developers define an application's logic through workflows, palettes of reusable component, and data models rather than text-based scripts. This paradigm emphasises the business case, which is assembled from pre-configured modules (O'Brien et al., 2025). LCNC provides several advantages for enterprises, including:

- *Democratisation of software development* - LCNC approaches remove the syntax barrier, the need to master programming syntax and, by extension, software engineering skills (Denny et al., 2011). Removing this barrier allows stakeholders in the energy modelling domain to build bespoke solutions. This paradigm shift results in applications and tools that are designed by those who understand the operational needs best.
- *Toolchain development* - by extension, LCNC facilitates the creation of toolchains and workflows where small data manipulation software components trans-form data as it is passed through a chain of established software tools.
- *Lower development costs, higher speed* - the cost of creating bespoke tools is reduced both by improving the ability to combine and reuse code, as well as overcoming the skills gap. This also brings much faster turn around times, as the design happens at the logical level, and changes are quickly made and quickly operational.

Typically, LCNC software is developed in platforms that provide development tools and frameworks (Prinz et al., 2021). Many flavours of LCNC aim at more or less professional outcomes. The WAVE platform is an example of a LCNC platform that follows the eXtreme Model Driven Development approach (Margaria and

Stef-fen, 2020). WAVE addresses many of these needs at the professional level needed for high assurance software. It is a cloud-based, collaborative application development environment that adopts a LCNC paradigm founded in formal models, with graphical composition of hierarchical workflows (control and data flow) on the basis of libraries of components that express the *skills* (or capabilities, or actions, as community-specific terminology) of the domain-specific entities to be coordinated. WAVE is domain-agnostic and employs formal methods to ensure workflow correctness, a capability absent in other platforms. This contrasts with other workflow orchestration tools such as Node-RED, which is limited to event-driven applications and KNIME which focuses on data science applications.

The paper introduces WAVE as a LCNC platform that can enable domain experts to build multi-scale energy models via graphical workflows instead of complex scripts. By showcasing specific workflows for archetype calculation and Building Energy Ratings (BER) prediction, we show that data-driven decarbonization efforts can be accelerated by abstracting from the technical MLOps' complexity. Having explored the background to this work, we introduce the WAVE platform and associated concepts. Finally, we present a use-case that focuses on the development of a multi-scale energy modelling solution.

Background

The need to decarbonize the built environment requires efficient methods for modeling building stocks across multiple scales, from individual thermal components to national energy infrastructure (Ali et al., 2024; Hoare et al., 2022). Historically, this field relied on physics-based approaches that use heat-transfer and thermodynamic equations to simulate physical behavior (Ahmad et al., 2018). However, these models require high-granularity data such as specific geometry and thermal properties (U-values), while being also computationally expensive (Ali et al., 2024; Ahmad et al., 2018).

To address these data gaps, data-driven models have gained prominence. These models leverage machine learning (ML) algorithms to predict energy use without requiring a deep understanding of modelled system (Ahmad et al., 2018; Manfren et al., 2022). To achieve multi-scale capabilities, researchers use Building Energy Modelling (BEM), which employ building archetypes, i.e., representative baseline models capturing typical characteristics of building groups, to scale individual estimates to regional or national levels (Ali et al., 2024). Furthermore, Linked Data approaches admit the integration of heterogeneous information sources, such as energy infrastructure and population data (Hoare et al., 2022). Recent advances have also integrated Geographic Information Systems (GIS) into these workflows to enhance spatial analysis of energy consumption patterns (Ali et al., 2024).

Despite the potential of ML, significant obstacles hinder its widespread adoption, particularly for Small and Medium Enterprises (SMEs) and government agencies. A primary barrier is the AI Skills Gap; firms often struggle to find personnel who possess both domain expertise and the specialized technical skills required to implement complex ML algorithms for energy models (Carioli et al., 2024; OECD, 2025). Smaller organizations often lack the financial resources to compete for data analysts familiar with energy modelling, leading to a competitive disadvantage and unfulfilled potential for the technology (Muminova et al., 2024; Dzhusupova et al., 2022).

Beyond personnel, the Machine Learning Deployment Workflow (MLOps) presents significant architectural and process challenges. Practitioners report that data management, including data discovery, cleaning, and joining dispersed datasets, consumes the vast majority of time in ML projects (Paleyes et al., 2022; Dzhusupova et al., 2022). Data often exist in *functional silos* making large-scale integration difficult (Paleyes et al., 2022; Dzhusupova et al., 2022). Additionally, the opaque nature of ML algorithms creates a *trust deficit* among stakeholders who require interpretability and formal verification to comply with high safety and regulatory standards (de Marco et al., 2025). *Low-Code/No-Code* (LC/NC) development is an emerging technology paradigm that enables individuals with no advanced computing or coding skills, collectively referred to as *Citizen Developers*, to create custom applications through visual drag-and-drop interfaces and pre-configured modules (El Kamouchi et al., 2023; Martinez and Pfister, 2023). By removing the *syntax barrier* these platforms allow domain experts to be closer and participate to code development on the basis of their own professional skills, thereby democratizing software development (Sundberg and Holmström, 2023).

The LC/NC paradigm has demonstrated success across several domains. It has been used to develop solutions for healthcare, for example diagnostic systems and virus tracking applications (Margaria and Schieweck, 2019). Similarly in manufacturing, the digital thread (Margaria and Schieweck, 2019) and automation of inventory management have been developed using LC/NC (El Kamouchi et al., 2023). It has also been used to support the development of decision support systems (Karusseit and Margaria, 2006) and for educational training at many levels (Gossen et al., 2018; O'Brien et al., 2025).

LC/NC offers a mechanism to overcome the skills gap by facilitating rapid tool-chain development (Sundberg and Holmström, 2023). These tools reduce implementation costs because organizations have their own staff to build bespoke models, bypassing the need for expensive IT consultants (Muminova et al., 2024). Platforms such as WAVE, and before it DIME (Boßelmann et al., 2016) and the jABC (Margaria and Steffen, 2009) address multi-scale modeling needs by utilizing hierarchical workflows that coordinate domain-specific skills into complex automated pipelines.

The WAVE architecture

The WAVE platform is described next. The architecture is summarized in Fig. 1. *The Formal Modelling Language* (orange layer in Fig. 1), provides the modelling languages

used in the platform. It is provided and maintained by the WAVE delveloper and support team. The Application Design layer (in blue in Fig. 1) is where the researchers from various communities use the WAVE facilities to design and execute their applications.

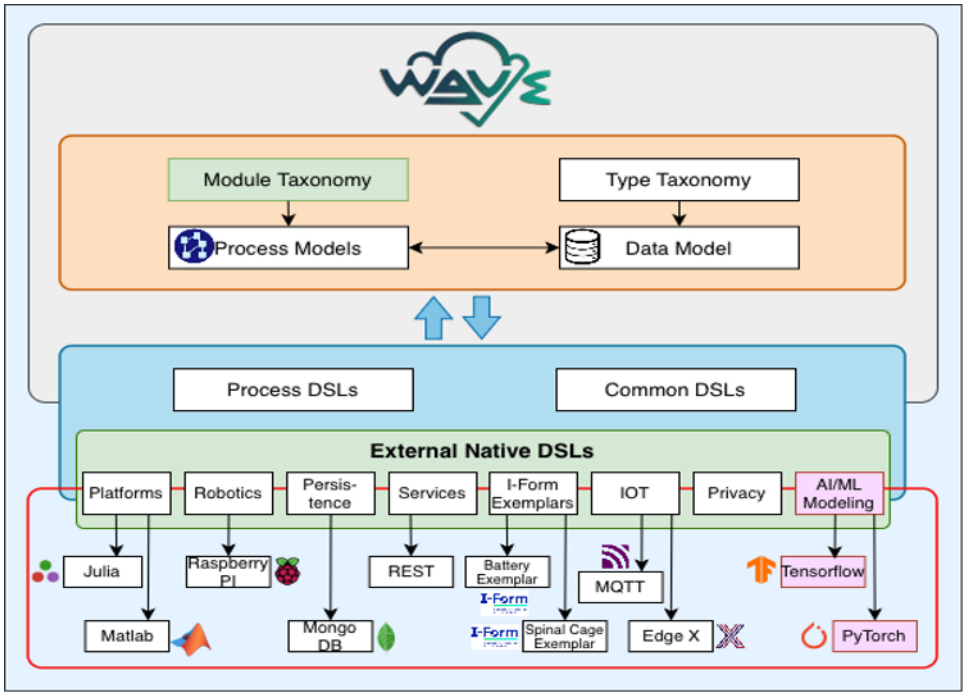


Figure 1: Architecture of the WAVE platform: the layers correspond roughly to different professional figures found in many communities

When LCNC is used to develop code, the final end-to-end logic that solves a business problem (traditionally coded as a script or program) is commonly called the workflow or process of the application under development and is modelled using compositions of components. This workflow describes how the control and the data flows from one step of the workflow to the other. The steps are designed using components called Service Independent Building Blocks (SIBs), each one embedding a domain specific functionality that executes a specific task. The task can range from simple operations, such as converting file formats, to complex ones like connecting to external APIs or training ML models. The complexity of API development, writing scripts for Python libraries such as Keras, scikit-learn, and TensorFlow, is fully abstracted. Users interact only with visual representations that expose the necessary configuration parameters. Built using the *Correctness by Construction* (CoC) approach, each SIB in the WAVE is a pre-validated block with established functional correct-ness and an associated model (Margaria et al., 2005) akin to a digital shadow. As shown in Figure 2, each SIB has well-defined, typed inputs and outputs. For each potential execution outcome there is a separate outgoing branch, labelled accordingly. The label also includes the outputs produced in that case. So even a simple visual inspection identifies what is produced (data, messages) for each specific outcome. Next steps are determined by the outcome of the current SIB, thus the outgoing branches are connected to the ”suitable” subsequent SIBs, defining the control flow. SIBs are thematically organized in palettes of related functionalities.

The collection of SIB palettes pertaining to a given application domain constitutes a Domain Specific Language (DSL) for that application domain. In our case, the DSL concerns the domain of energy consumption analysis, pre-diction and mapping.

The WAVE uses Software Engineering methods and techniques to enable researchers to design, implement and execute verified and tested applications by modelling their workflows using application-domain DSLs. If the DSL and SIBs for a particular workflow are available, designing new applications simply reuses those ready SIBs: a no-code activity. If functions are missing, they need to be either integrated from existing libraries or tools, or programmed afresh. This is the *External Native* DSL, in green in Fig. 1. There, existing or new software components, hardware, APIs or other resources (red layer in the Figure) are made available to the WAVE users as collecitons of SIBs. This is a low-code task, as it concerns a specific, well defined component to be added. This step requires some programming skills, but not complex architectural knowledge, and it is the step that produces the palettes of reusable components.

WAVE will employ formal methods like the GEAR model checker (Steffen et al., 2007) and the MC3 infinite state model checker (Steffen and Murtovi, 2018) to ensure workflow correctness at design time. Those model checkers are specialized for deep diagnostic support and procedural (i.e, hierarchical) systems respectively. While users are modelling their workflow, syntactic and static semantics check will ensure that workflows are valid. For example, verifying the absence of deadlocks and

guaranteeing that critical steps are always executed, or guaranteeing precedences. Additionally, the platform will offer support for integration of new components, execution of the workflows and adequate monitoring and logging mechanisms.

Using the WAVE

When researchers from various communities use the facilities offered by the WAVE platform to design and execute their applications, they reuse over and over again the general purpose components of the Common DSLs and the domain specific DSLs. This way, users produce the do-main and application specific workflows of the Process DSLs. As WAVE has the same common modelling language for its components and workflows, that is shared across any application domain, diverse research communities can share components with other communities, speed-ing up significantly the work of other users. For example, data analytics and visualization components and even workflows are frequently shared across several communities active in diverse research domains such as bioinformatics, history and advanced manufacturing.

Any applications, and many of the features of the WAVE platform itself, are implemented as workflows that use ap-plication domain specific libraries of components called AD-DSLs. The specialization to a domain does not require a new workflow language or composition model, but it simply happens by adding collections of skills through libraries of components that provide access to (in-terface) and execution (on an adequate runtime) of those skills. Robotics applications, data transcription applications, biomedical imaging applications (Brandon et al., 2024) have recently been realized this way.

A slightly more technical view of the components and workflows will be illustrated in the case study.

The feedback from the different communities of practice drives the growth and evolution of the platform itself. This happens through requests for additions, like recently the need of support for larger data files.

Integrating existing capabilities

The encapsulation of external resources providing data (databases, input streams), communication (http, wifi, 5G networks, ORAN), hardware (sensors, actuators, computation, robots), software (analytics libraries, graphics, simulators, geovisualisation systems) and models (AI models, recommenders, ontologies, classifiers, simulation models, digital twins) and other services follows API integration, or uses higher level community-driven services like e.g. the EdgeXfoundry services for edge IoT peripherals and computing tools.

In this way, it is possible to grow the capability of each community in according to their domain's needs, and thus happens in the context of ensured compatibility across communities. This is the key design decision that allows sharing and reuse across the entire WAVE.

Learning Support

LCNC is a new concept for many. The WAVE developers are designing and delivering training materials targeted to specific communities. For example, with education experts, they have produced both a *train the trainer program* (O'Brien et al., 2025) as well as training material on technical skills needed to use the platform. A MOOC on LCNC is currently under production. For example, topics such as hardware integration skills for robotics and advanced manufacturing researchers and workflow modelling skills for application designers have been developed.

Case study: predicting energy performance

A typical prediction methodology for energy performance of urban buildings involves the following steps:

1. Collecting data from various sources such as building stock, census, weather, and geographical data.
2. Developing building archetypes using building stock data to identify representative baseline models.
3. Parametric simulation is then conducted to develop appropriate synthetic data, followed by
4. the development of machine learning models to predict building energy performance.
5. Finally, the urban building energy performance analysis step analyzes the modeling process results for planning and decision-making purposes.

The data collection phase is non-trivial. In this country, it involves ingesting and harmonizing approximately 2.4 mil-lion residential records (in this instance, a GeoDirectory). Other heterogeneous data sources (census data from the national Central Statistics Office and Energy Performance Certificate (EPC) data from the national Sustainable Energy Authority (SEA)) are also ingested and fused with the building records to produce an overarching knowledge graph that captures both the individual data records and their relationships with other records in the graph.

When a similar methodology is applied to other jurisdictions, different data sources are ingested and fused, but the software processes for doing this are similar and suggest that code should be extensively reused with minor modification. However, in a traditional software environment, implementing this ingestion would require coordination between a domain expert (e.g., a housing analyst) and a software engineer to produce the modification. The WAVE platform replaces this requirement with the design of a workflow orchestrating SIBs according to shared, reusable components defined in a domain specific palette of SIBs. We show how to develop these workflows and related SIBs to demonstrate how the WAVE platform can be applied to develop software for multi-scale energy modelling.

Figure 2 represents (part of) the SIB library for multi-scale energy modelling within the WAVE platform. The library is organized into five distinct palettes: *Common*, *Data Ingestion*, *File Parsing and Formatting*, *Arch*

Count, and *Predict BER*. The *Common* palette contains general purpose SIBs: they are not specific to a given domain. The *Data Ingestion* palette contains SIBs to retrieve valid datasets from AWS, Azure, REST APIs and SFTP servers. Once the data is successfully stored into the local workspace, *File Parsing and Formatting* SIBs are used to homogenize data across file encodings and formats. During data engineering phase, *Arch Count* palette SIBs are used to query the Neo4J graph database and retrieve building counts whose features match those specified by the *ArcheTypesCombination* parameter. Finally, *Predict BER* SIBs are used in workflows that utilize pretrained ML models for predicting Building Energy Ratings.

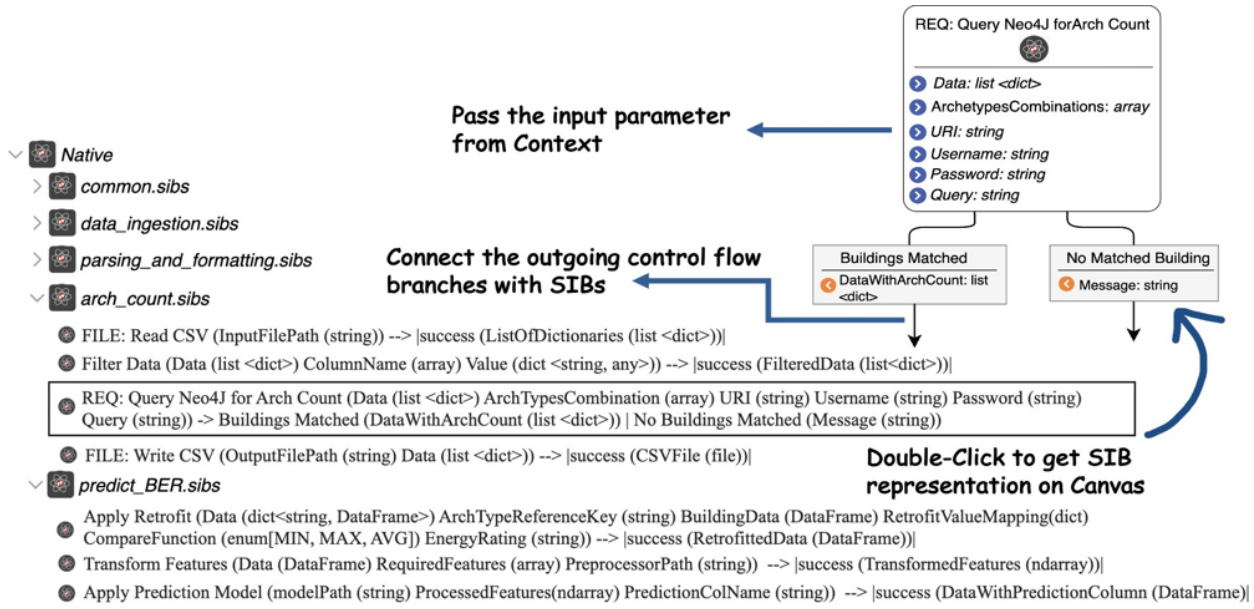


Figure 2:SIB transition from the SIB Library for multi-scale energy modelling into the WAVE Canvas

To create custom data ingestion pipelines, stakeholders select the appropriate SIBs from the palettes and use them to develop workflows. We show and discuss here two of the workflows defined for this case study: (a) Calculate Arch count (Neo4j connection and building count retrieval) and

(b) part of Predict BER (Data transformation and prediction using a pre-trained Machine Learning model). These workflows can become themselves process SIBs, and be reused for similar tasks, for example, for energy modelling in other jurisdictions.

We now show and briefly discuss two of the energy modelling workflows to demonstrate how the WAVE LCNC approach can support the full ML lifecycle.

Anatomy of a workflow

A well-formed workflow is composed of a *Start* node, one or more *End* nodes, one or more Service Independent Building Blocks (SIBs), and both the control flow of the application (solid line connectors) and its data flow (dashed line connectors). The Start node marks the initialization of the workflow with the input parameters passed from the context. Each SIB in the workflow has a name that reflects its purpose, and one or more outgoing control flow labelled *branches* that lead to the successive SIBs. Each branch corresponds to an outcome of the execution (e.g. the branches of the *REQ: Query Neo4J for Arch Count* SIB in Fig.3 are labelled *Buildings Matched*, and *No Matched Building*, respectively. In terms of the data flow, the SIB inputs are indicated in blue, and the outputs are placed on the branches, as they are typically specific to the execution outcome and thus branch specific.

Predict building energy ratings (BER)

This workflow predicts Building Energy Ratings by applying a pre-trained ML model to simulated envelope retrofits. It combines domain independent data-preprocessing SIBs, such as Load Datasets, Filter Data, and Dictionary Mapping, with application-specific ones, like Apply Retrofit. Technical complexities, including file I/O, data transformation, and running ML models via libraries like *os, pan-das*, and *joblib*, are fully abstracted. The visual modules accelerate model comparison and selection, as users can experiment with different ML models on the same or heterogeneous datasets simply by switching configurations in the Apply Prediction Model and Load Dataset SIBs. Additionally, the entire workflow can be encapsulated as a Process SIB, enhancing collaboration and reproducibility.

Calculate arch count workflow

This workflow processes a dataset that contains various archetype combinations such as Building Type, Building Energy Rating, Occupancy and Heating System, and queries a national-scale graph for the count of buildings with similar characteristics. By abstracting the technical details of connecting to Neo4j hosts, the workflow enables straightforward retrieval of building counts. The Neo4j query is passed as a parameter, allowing the matching criteria to be modified without requiring changes to the code files.

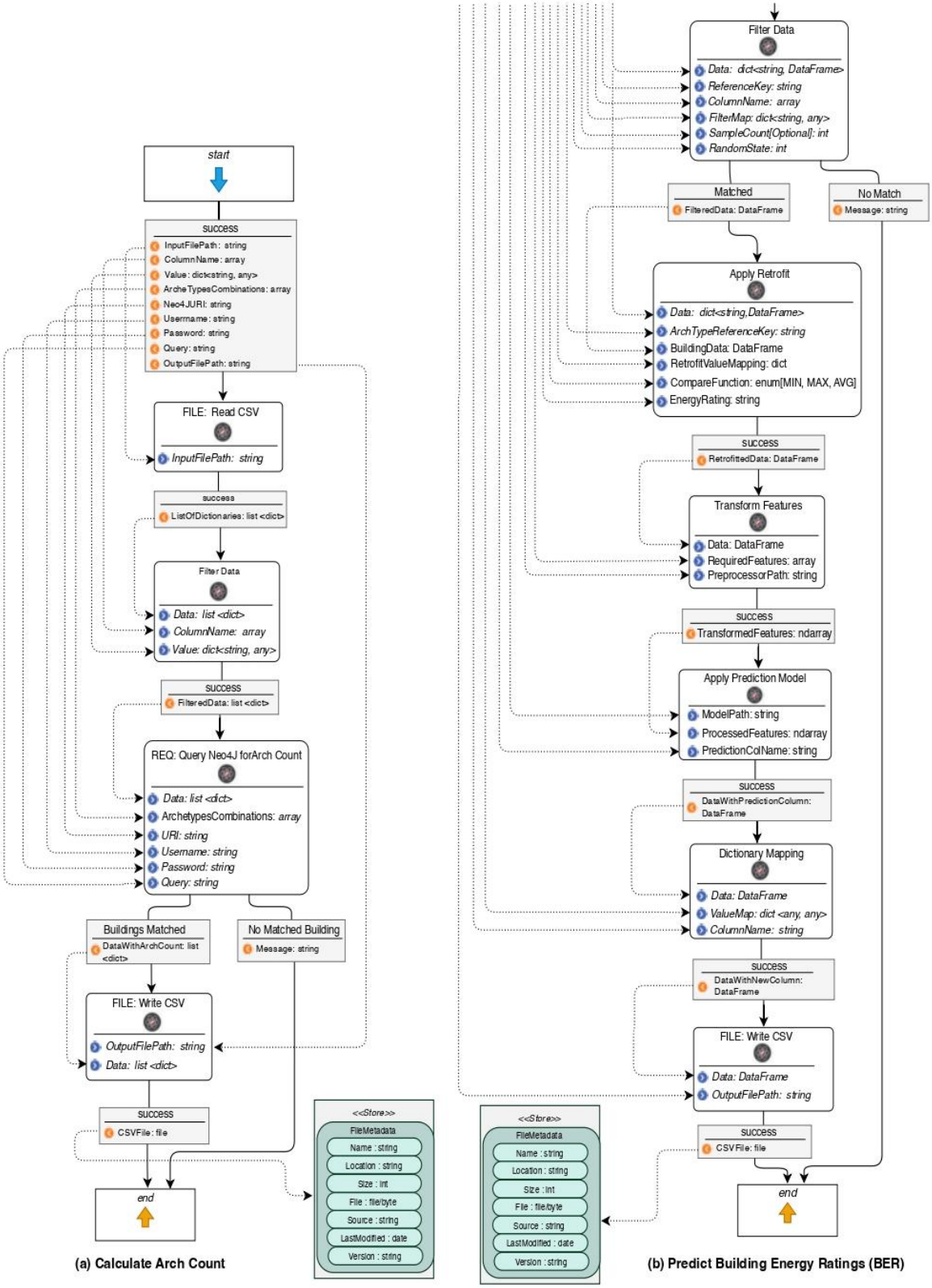


Figure 3:The WAVE workflows for (a) Calculating Archetype Counts, (b) (core part of) Predicting Building Energy Ratings (BER)

Similarly, parameters can be adjusted to calculate Arch counts for datasets filtered by a combination of columns (e.g., where ”BER” = ”A” and ”Occupancy” = 2).

Comparing traditional and LCNC approaches

While comparing the WAVE LCNC approach with a script-based pipeline, an assessment of their respective development time was done. The developer of the original

code scripts is both an expert in energy modelling and a competent programmer. In total, approximately 2000 lines of code were developed, talking approximately 3 days to develop and validate the code. A second developer created the LCNC workflows. Each one was developed and validated in approximately 4 hours. In addition to the reduced development time, the maintainability and reusability of the LCNC code also improved, as the corresponding code is componentized at a fine granularity and complies with the architectural characteristics of SIBs, and the control and data flows are explicitly modelled and thus easy to understand and to debug.

Workflows will be developed to support full electrification scenario modelling such as the effect of heat-pump adoption and EV charging within the housing stock. The resulting demand will be spatially allocated to low-voltage sub-stations, allowing feeder peak load and transformer over-load to be calculated. This will be displayed in multiscale geographical reports such as those in Fig. 4. This demonstrates how graphical workflows enable non-programmers to move from heterogeneous datasets to infrastructure-level planning insights across multiple modelling scales. Moreover, WAVE decouples workflow design from the computation, thereby enhancing execution performance. Workflows are stored in Common Workflow Language for-mat1 and executed on high-capacity server clusters, allowing users with low-specification laptops to initiate large-scale data-processing and machine learning pipelines. The backend manages metrics such as memory usage and CPU cycles through an orchestrator, eliminating the need for users to oversee these technical aspects.

Conclusions

The need to decarbonize the built environment requires a shift toward data-driven modeling. Although machine learning (ML) provides a powerful mechanism for optimizing energy performance across multiple scales, the skill gap around implementation remains a primary barrier to widespread adoption. This gap is characterized not just by a lack of technical coding ability, but by the difficulty of merging domain-specific energy expertise with complex software engineering.

LCNC platforms such as WAVE offer a convenient solution by democratizing software development. By re-moving the syntax barrier yet increasing checks and the use of domain knowledge, Wave allows domain experts who best understand building physics and operational needs - to lead the development of bespoke energy models. As demonstrated, this paradigm reduces development costs, accelerates toolchain creation, and bridges the professional skills gap. Ultimately, leveraging platforms like WAVE ensures that SMEs, policymakers and other stakeholders can fulfill the potential data-driven energy modelling.

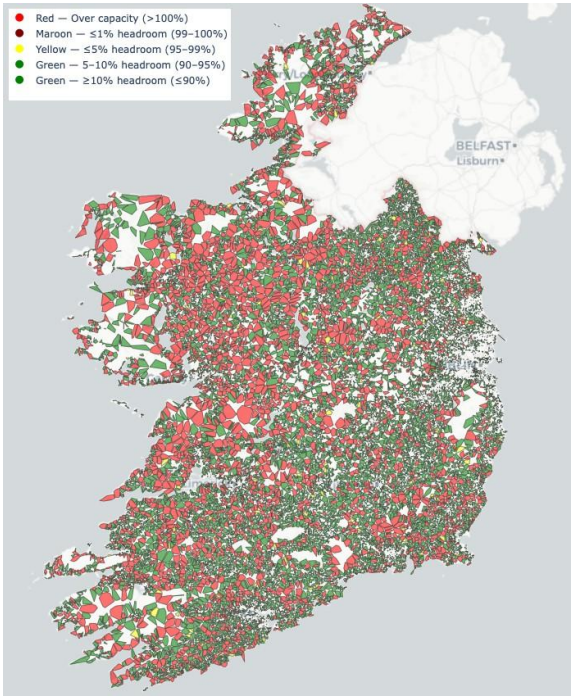


Figure 4: Results of multiscale energy modelling for scenarios including heatpump and EV adoption by residential households.

Acknowledgements

This publication has emanated from research supported by Wave project (ULF-Wave-2025) and partly by a grant from the Sustainable Energy Authority of Ireland (SEAI) un-der the Research, Development and Demonstration Funding Programme Grant No. 22/RDD/822. The opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the SEAI. Grammarly and ChatGPT were used to improve spelling, grammar, punctuation and clarity of some text here-in.

References

Ahmad, T., Chen, H., Guo, Y., and Wang, J. (2018). A com-prehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A re-view. Energy and Buildings, 165:301–320.

Ali, U., Bano, S., Shamsi, M. H., Sood, D., Hoare, C., Zuo, W., Hewitt, N., and O'Donnell, J. (2024). Urban building en-ergy performance prediction and retrofit analysis using data-driven machine learning approach. Energy and Buildings, 303:113768.

Boßelmann, S., Frohme, M., Kopetzki, D., Lybecait, M., Nau-jokat, S., Neubauer, J., Wirkner, D., Zweihoff, P., and Steffen, B. (2016). DIME: A programming-less modeling envi-ronment for web applications. In Proc. ISoLA 2026, volume LNCS 9953, pages 809–832, Cham. Springer International Publishing.

[1] CWL standard: https://www.commonwl.org

Brandon, C., Boßelmann, S., Singh, A., Ryan, S., Schieweck, A., Fennell, E., Steffen, B., and Margaria, T. (2024). Cinco de bio: A low-code platform for domain-specific workflows for biomedical imaging research. BioMedInformatics, 4(3):1865–1883.

Carioli, P., Czarnitzki, D., and Fernández, G. P. (2024). Evidence on the adoption of artificial intelligence: The role of skills shortage. ZEW-Centre for European Economic Research Discussion Paper, (24-013).

de Marco, G., Niederwieser, E., and Siegele, D. (2025). Explor-ing a multimodal conversational agent for construction site safety: A low-code approach to hazard detection and compliance assessment. Buildings, 15(18):3352.

Denny, P., Luxton-Reilly, A., Tempero, E., and Hendrickx, J. (2011). Understanding the syntax barrier for novices. In Proc. ITiCSE '11, page 208–212, New York, NY, USA. ACM.

Dzhusupova, R., Bosch, J., and Olsson, H. H. (2022). Challenges in developing and deploying ai in the engineering, procurement and construction industry. In Proc. COMPSAC 2022, pages 1070–1075.

El Kamouchi, H., Kissi, M., and El Beggar, O. (2023). Low-code/no-code development: A systematic literature review. In 2023 14th Int. Conf. on Intelligent Systems: Theories and Ap-plications (SITA), pages 1–8. IEEE.

Gossen, F., Kühn, D., Margaria, T., and Lamprecht, A.-L. (2018). Computational thinking: Learning by doing with the cinco adventure game tool. In Proc. COMPSAC 2018, volume 01, pages 990–999.

Hoare, C., Aghamolaei, R., Lynch, M., Gaur, A., and O'Don-nell, J. (2022). A linked data approach to multi-scale energy modelling. Advanced Engineering Informatics, 54:101719.

Karusseit, M. and Margaria, T. (2006). Feature-based modelling of a complex, online-reconfigurable decision support service. ENTCS, 157(2):101–118. Proc. Int. Workshop on Automated Specification and Verification of Web Sites (WWV 2005).

Katavoutas, G., Founda, D., Varotsos, K. V., and Giannakopoulos, C. (2022). Climate change impacts on thermal stress in four climatically diverse european cities. Int. Journal of Biometeorology, 66(11):2339–2355.

Manfren, M., James, P. A., and Tronchin, L. (2022). Data-driven building energy modelling – an analysis of the potential for generalisation through interpretable machine learning. Renewable and Sustainable Energy Reviews, 167:112686.

Margaria, T., Raffelt, H., and Steffen, B. (2005). Knowledge-based relevance filtering for efficient system-level test-based model generation. Innovations in Systems and Software Engineering, 1(2):147–156.

Margaria, T. and Schieweck, A. (2019). The digital thread in industry 4.0. In Proc. iFM 2019, Integrated Formal Methods, pages 3–24, Cham. Springer International Publishing.

Margaria, T. and Steffen, B. (2009). Business Process Model-ing in the jABC: The One-Thing Approach. In Handbook of Research on Business Process Modeling, pages 1–26. IGI Global, Hershey, PA.

Margaria, T. and Steffen, B. (2020). Extreme model-driven development (XMDD) technologies as a hands-on approach to software development without coding. In Encyclopedia of Education and Information Technologies, pages 732–750. Springer.

Martinez, E. and Pfister, L. (2023). Benefits and limitations of using low-code development to support digitalization in the construction industry. Automation in Construction, 152:104909.

Muminova, E., Ashurov, M., Akhunova, S., and Turgunov, M. (2024). Ai in small and medium enterprises: Assessing the barriers, benefits, and socioeconomic impacts. In Proc. ICK-ECS 2024, volume 1, pages 1–6.

OECD (2025). Bridging the AI skills gap: Is training keeping up? OECD Publishing, Paris.

O'Brien, S., Brandon, C., Busch, D., Krumrey, M., Mitwalli,

D. S., Singh, A., Teumert, S., and Margaria, T. (2025). Spe-cialized training in LCNC and AI: from the pedagogical concept to the experience. In Proc. COMPSAC 2025, pages 2394–2403. IEEE.

Paleyes, A., Urma, R.-G., and Lawrence, N. D. (2022). Challenges in deploying machine learning: A survey of case stud-ies. ACM Comput. Surv., 55(6).

Prinz, N., Rentrop, C., and Huber, M. (2021). Low-code devel-opment platforms-a literature review. In AMCIS.

Skillington, K., Crawford, R. H., Warren-Myers, G., and David-son, K. (2022). A review of existing policy for reducing embodied energy and greenhouse gas emissions of buildings. En-ergy Policy, 168:112920.

Steffen, B., Margaria, T., Nagel, R., Jörges, S., and Kubczak, C. (2007). Model-driven development with the jabc. In Bin, E., Ziv, A., and Ur, S., editors, Hardware and Software, Verification and Testing, pages 92–108, Berlin, Heidelberg. Springer Berlin Heidelberg.

Steffen, B. and Murtovi, A. (2018). M3c: Modal meta model checking. In Howar, F. and Barnat, J., editors, Formal Methods for Industrial Critical Systems, pages 223–241, Cham. Springer International Publishing.

Sundberg, L. and Holmström, J. (2023). Democratizing artificial intelligence: How no-code AI can leverage machine learning operations. Business Horizons, 66(6):777–788.

Digital Twins & Asset Intelligence

Sub-Tracks:

Building / Infrastructure Digital Twins

Smart Infrastructure Monitoring & Sensor Networks

Safety Innovations (AI Vision, VR, Proactive Systems)

Urban-Scale Digital Twins & City Analytics

Linked Data & Semantic Web for Digital Twin

Data Interoperability, Governance & Standardization

Twin-Driven Simulation & Optimization

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

ASSET-CENTRIC KNOWLEDGE SHARING FOR PERMITTING AND COMPLIANCE WITH DIGITAL TECHNOLOGIES

Alhadi Alsaffar[1], Professor Thomas Beach[1], and Professor Yacine Rezgui[1]
[1]Cardiff University, Cardiff, UK

Abstract

Digital transformation in construction asset management is advancing through Building Information Modelling (BIM) and emerging tools such as Digital Twins (DT), Blockchain (BC), and Artificial Intelligence (AI), yet knowledge sharing in permitting and compliance remains fragmented. A survey of 110 professionals, including 30 open-ended responses from the same sample, was analysed using descriptive statistics, chi-square tests, ordinal logistic regression, reflexive thematic analysis, and concept mapping. Email and collaboration platforms dominate (70%), while external sharing remains limited. BIM (93%) and AI/ML (79%) show the highest adoption. The study develops a refined hierarchical conceptual map for digital permit certification and compliance checking.

Introduction

The construction industry faces critical challenges in which effective Knowledge Management (KM) is essential for successful asset management (Rezgui, 2001; Boamah et al., 2022; Yepes and López, 2021; Rezgui, 2011), especially in permitting and compliance checks. Regulatory complexity increases uncertainty and costs due to evolving codes and lengthy permit reviews involving multiple steps and stakeholders, resulting in inefficiency, rework, and delays (Yang and Zhang, 2024; Bloch and Fauth, 2023; Ataide et al., 2023; Fauth et al., 2023). Fragmented information silos, project data trapped in disconnected teams and proprietary formats, undermine traceability; without unifying standards or integrated technologies, teams may apply different or out-of-date rules, making compliance reviews inconsistent and complicating constraint management and decision-making (He et al., 2025; Beach et al., 2024). Skills gaps and cultural resistance, including continued reliance on paper/email workflows and limited Construction 4.0 competencies, further hinder the adoption of advanced digital technologies such as Building Information Modelling (BIM), Digital Twins (DT), Blockchain (BC), and Artificial Intelligence (AI) (Alsaffar et al., 2025; Siddiqui et al., 2023; Rinchen et al., 2024).

KM is vital for organisational success because it ensures that knowledge is effectively stored, shared, and applied to achieve strategic impact (Yeboah, 2023). In construction, this underpins traceability of decisions, supports legal defensibility, and reduces redundant rework (Wang and Meng, 2021; Sik-wah Fong et al., 2005; Xiahou et al., 2025). As projects grow in complexity and regulatory requirements become more demanding, effective knowledge sharing, accessibility, and utilisation across organisational boundaries become increasingly crucial (Maheshwari et al., 2025). Knowledge sharing also requires balancing explicit knowledge (documents and reports) with tacit knowledge (experience-based know-how), which is harder to transfer due to structural differences and expression challenges (Chumg et al., 2014; Wang et al., 2014). A widely used lens for understanding this process is the SECI model, shown in **Figure 1**, which explains how knowledge cycles between tacit and explicit forms through socialisation, externalisation, combination, and internalisation (Nonaka and Takeuchi, 1995; Aldmour, 2012; Gandomani and Tavakoli, 2016; Shull et al., 2004).

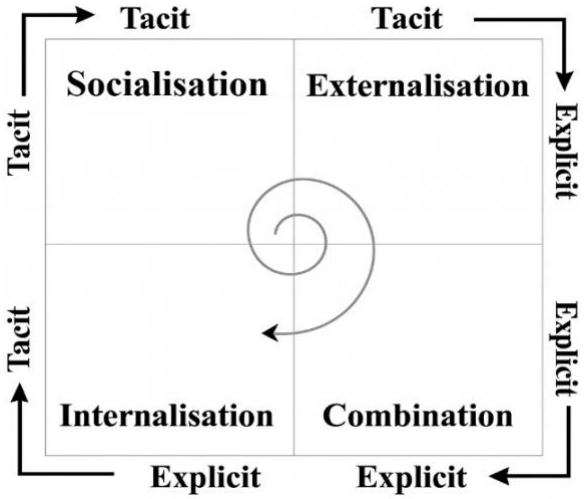


Figure 1: SECI knowledge creation model (Nonaka and Takeuchi, 1995).

The SECI model provides a valuable framework for analysing knowledge sharing in permitting and compliance processes. This is particularly relevant in construction, where knowledge sharing must account for both documented information and experience-based judgement (Chumg et al., 2014; Wang et al., 2014).

The ongoing Fourth Industrial Revolution (4IR), described as a "*fusion of technologies blurring the lines between physical, digital, and biological spheres*," has transformed how people connect and access knowledge (Klaus, 2016). In construction, this convergence is

reflected in technologies such as BIM and platform-based collaboration, where increased BIM implementation has been shown to enhance knowledge sharing and inter-organisational cooperation (Cheng et al., 2023). DT technologies have also been associated with improved efficiency and operational transparency in construction and asset management (Alsaffar et al., 2025). BC technologies further contribute to this convergence by providing immutable, decentralised records that enhance trust, traceability, and secure information exchange across project stakeholders (Alsaffar et al., 2024). These tools can support knowledge-creation and utilisation activities broadly consistent with SECI (Nonaka and Takeuchi, 1995). BIM makes design intent explicit and shareable (externalisation/combination), while DTs may, in later lifecycle stages, support forms of learning and reuse that are broadly aligned with internalisation. Importantly, BIM-based as-designed information can support later operational DT development across the asset lifecycle, offering new opportunities for integrated, knowledge-rich approaches to the built environment, although this role remains emerging (Cheng et al., 2023; Alsaffar et al., 2025; Peng and Liu, 2023).

Despite progress, the construction industry remains fragmented, with knowledge confined to specific trades, organisations, or project phases (Daniel et al., 2024). Although there have been notable advancements in capturing and sharing construction knowledge (Deng et al., 2022), permitting and compliance processes remain knowledge-intensive and still frequently rely on manual expert review (Peng and Liu, 2023). Recent research also reports that building-permit processes remain manual and fragmented (Hagedorn et al., 2025) and that traditional permit-management systems often lack transparency, thereby complicating certification verification and regulatory compliance (Alsaffar et al., 2024). This lack of transparency underscores the need to consider Blockchain-enabled certification mechanisms alongside BIM and Digital Twins to support verifiable, trustworthy permitting and compliance workflows (Alsaffar et al., 2024). In practice, permitting involves multiple stakeholders, including owners, designers, engineers, contractors, and authorities, and requires extensive coordination of technical and procedural knowledge. However, barriers such as technical compatibility issues, costs and limited digital literacy continue to constrain the adoption of digital permitting solutions (Daniel et al., 2024).

Nevertheless, empirical evidence remains limited regarding how knowledge-sharing practices and enabling digital technologies are used together in permitting and compliance checking, and on the relationships between key concepts involved in digital permit certification and compliance checking. Accordingly, building on the SECI knowledge lens and the context of digital transformation, this study examines current knowledge-sharing practices and evaluates the enabling digital technologies, including DT, BIM, BC, and AI. It is guided by the following research questions:

RQ 1: *"What knowledge-sharing practices and enabling digital technologies are used in permitting and compliance checking for construction asset management?"*

RQ 2: *"What are the relationships between key knowledge-sharing concepts in digital permit certification and compliance checking?"*

Here, "*digital permit certification and compliance checking*" refers to digital workflows for sharing, checking, and verifying permitting and compliance evidence across stakeholders in construction asset management. This paper makes two contributions. First, it reports empirical findings on knowledge sharing and enabling digital technologies in permitting and compliance checking. Second, it synthesises these findings into a refined hierarchical conceptual map for digital permit certification and compliance checking, linking DTs as as-designed information extensions of BIM, BC-enabled certification, AI-enhanced knowledge management, and organisational/regulatory enablers. The remainder of the paper presents the survey-based methodology, discusses the results, and concludes with implications and directions for future research.

Research Methodology

The research design follows the Research Onion framework (Saunders, 2019) and adopts a mixed-methods, cross-sectional survey strategy. The quantitative component provides the primary empirical evidence, while the qualitative component adds contextual depth to interpret patterns in the numerical results. **Figure 2** summarises the overall research workflow, from survey development and data collection to analysis and hierarchical conceptual mapping.

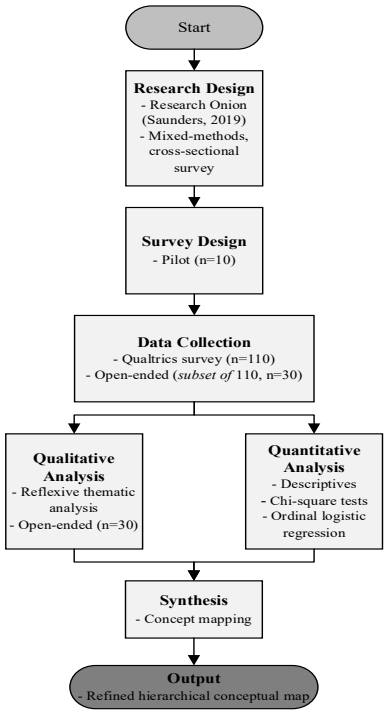


Figure 2: Research workflow summarising the mixed-methods design and analysis

Figure 2 also reflects the paper structure, in which survey design and data collection are described in this section; the quantitative and qualitative findings are reported in the "Discussion and Result Analysis" section; and the concept mapping synthesis and refined hierarchical conceptual map are presented in the final part of that section (**Figure 4** and **Table 7**).

Data were collected through an online questionnaire administered via the Qualtrics XM platform, targeting construction professionals involved in asset management, permitting, and compliance. The instrument comprised 25 questions: multiple-choice items (e.g., "Which of the following digital technologies do you use to support knowledge sharing in compliance and permitting?"), 5-point and 6-point Likert-scale evaluations (e.g., rating knowledge-sharing effectiveness), and open-ended prompts (e.g., exploring barriers to knowledge sharing). A pilot study (n=10) improved clarity and reliability before full distribution (Dudovskiy, 2024). The final dataset comprises 110 valid responses for quantitative items, with a subset of 30 respondents also completing qualitative open-ended prompts. The sample includes respondents with substantial domain experience, with over 87% reporting more than three years in compliance and permitting roles. Participants represent engineers (62%), project managers (52%), compliance officers (43%), and regulatory authority staff (40%).

While the quantitative component primarily captures what practices are used and to what extent technologies are adopted, the concept-mapping synthesis identifies and represents the relationships between key knowledge-sharing concepts in digital permit certification and compliance checking in a structured conceptual form.

Quantitative analysis used descriptive statistics (frequencies, percentages, and means) to summarise knowledge-sharing tools and practices, perceived effectiveness, adoption levels of digital technologies, and perceived barriers/benefits. Chi-square tests were applied to explore whether selected perceptions differed across respondent groups (e.g., by experience) (Kremelberg, 2010). To compare internal and external knowledge-sharing frequency, an ordinal logistic regression was conducted in R using the *polr* function from the MASS package for ordered response categories (Agresti, 2019; Chihara, 2011).

Qualitative responses were analysed using reflexive thematic analysis (Byrne, 2022), in which data were manually coded in Excel across six stages: familiarisation, generating initial codes, searching for themes, reviewing themes, defining themes, and reporting. This approach enriched contextual insight and strengthened the quantitative results, thereby providing deeper analysis and a clearer understanding of the use of digital technology in construction compliance.

Finally, concept mapping was used to synthesise the survey findings into a refined hierarchical conceptual map. Following Cañas and Novak (2008), the process involved defining the domain, listing key concepts ("parking lot"), organising concepts hierarchically, constructing a preliminary map, identifying cross-links, iteratively refining the structure, and checking it against the empirical evidence. The central domain is "Asset Management in Permitting and Compliance." Key concepts include Digital Technologies (BIM, DTs, BC, and AI), KM (Knowledge Management System (KMS), knowledge sharing, and documentation consistency), Compliance Processes and Assurance, and Strategic Barriers (communication gaps, training/technical expertise gaps, cultural resistance, infrastructure deficiencies, and regulatory barriers).

To ensure methodological rigour and trustworthy findings, the study adhered to ethical practices, including obtaining informed consent, maintaining anonymity and confidentiality, and handling data securely (Saunders, 2019; Creswell, 2023).

Discussion and Result Analysis

This section presents and analyses the findings in two parts. First, it examines current knowledge-sharing practices and enabling digital technologies in permitting and compliance checking. Second, it presents a refined hierarchical conceptual map of the relationships between key knowledge-sharing concepts.

To address **RQ1**, the current practices and gaps analysis shows that the 110 valid respondents hold roles directly relevant to permitting and compliance workflows, including Engineers (62%), Project Managers (52%), Compliance Officers (43%), Regulatory Authority staff (40%), and IT/BIM specialists (29%). More than 87% report at least 3 years of compliance/permitting experience. This high level of domain-specific experience and the balanced representation across key stakeholder roles (design, management, compliance, and IT) indicate that the sample comprises highly qualified professionals actively engaged in construction asset management and regulatory processes.

As shown in **Table 1**, compliance-related knowledge sharing is dominated by email, repositories (e.g., SharePoint), and collaboration platforms (each at 70%). KMS and Common Data Environments (CDEs) are used by around half of respondents, while paper-based exchange persists (31%), indicating continued reliance on a mix of digital and traditional approaches.

Table 1: Knowledge-sharing methods in use

Knowledge-sharing method	Count (n=89)	%
Email	62	70%
Document repositories (SharePoint)	62	70%
Collaborative platforms (Teams/Zoom)	62	70%
KMS (e.g., Primavera, Confluence)	50	57%
Common Data Environments (CDEs)	52	59%
Paper-based	27	31%

However, perceived effectiveness is mixed, as shown in **Table 2**, where 22% rated them extremely effective, 18% rated them very effective, while 22% rate them ineffective or very ineffective. The weighted average of 3.91/6 (≈65.2%) indicates only moderate performance, while the remaining 34.8% highlights unrealised potential due to inconsistent knowledge-sharing practice.

Table 2: Effectiveness Ratings of Knowledge-Sharing Practices

Effectiveness Level	Count (n=87)	%
1 (Very Ineffective)	7	8%
2 (Ineffective)	12	14%
3 (Somewhat Effective)	17	20%
4 (Effective)	16	18%
5 (Very Effective)	16	18%
6 (Extremely Effective)	19	22%

Internal exchange is substantially more frequent than external exchange, as shown in **Figure 3**. External sharing is "*rarely*" or "*never*" for a large share of respondents, indicating that compliance-related knowledge is more effectively circulated within organisations than across stakeholder boundaries. Ordinal logistic regression indicates higher-frequency sharing in internal channels than in external channels (log-odds ≈ 1.12; odds ratio ≈ 3.07), highlighting a persistent constraint in cross-organisational knowledge flow, which is critical for permitting and compliance where coordination with external stakeholders is essential.

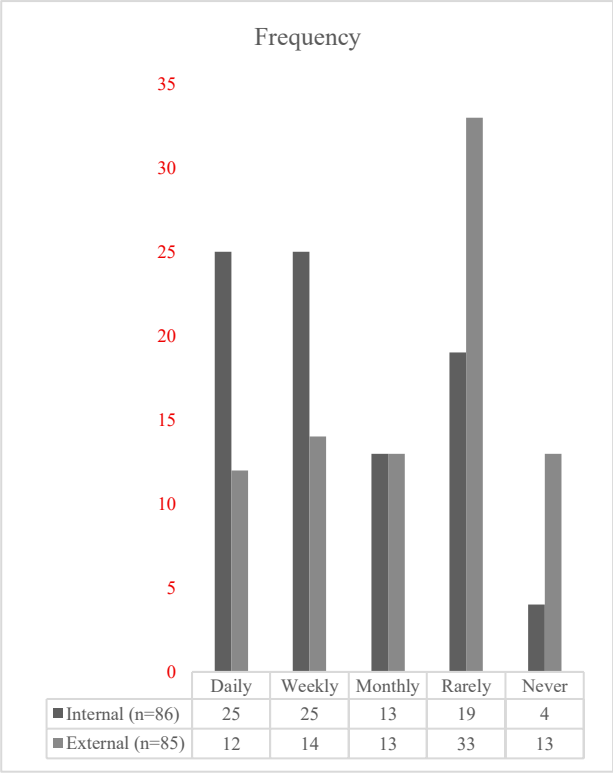


Figure 3: Comparison of the Frequency of Internal vs. External Compliance Knowledge Sharing

This disparity indicates strong but isolated internal communication and reveals significant gaps in external engagement, underscoring the need for better integration and more structured communication with external stakeholders.

The most frequently reported barriers, as shown in **Table 3**, were captured through a multiple-choice question in which respondents could select all applicable options.

Table 3: Common Barriers to Effective Knowledge Sharing

Barrier	Count (n=87)	%
Lack of communication between departments	79	91%
Lack of technology	76	87%
Lack of a formal knowledge-sharing process	69	79%
Inconsistent documentation	57	66%

In addition, a chi-square test indicates no statistically significant association between the experience group and the main barrier category ($\chi 2(3) = 4.78$, $p = 0.189$), suggesting that these barriers are widely shared across respondents.

Separately, when asked about challenges to implementing digital knowledge-sharing tools specifically, respondents identified a lack of technical expertise (87%), resistance to change (76%), high cost (71%), inadequate infrastructure (69%), limited management support (58%), and regulatory restrictions (43%). These results help explain why reliance on email exchange remains high despite interest in more advanced workflows.

In the adoption of digital technologies, as shown in **Table 4**, the findings indicate a layered digital landscape in which mature tools such as BIM coexist with emerging solutions, while widespread AI/ML uptake underscores an increasing reliance on data-driven systems.

Table 4: Digital Technology Adoption

Technology	Count (n=75)	%
BIM	70	93%
AI/ML	59	79%
Blockchain	38	51%
Digital Twins	33	44%
IoT	33	44%

Respondents also report positive operational impacts: 49% report significant improvements in permitting and compliance processes, and 43% report slight improvements. However, gap ratios (considering/using) indicate that DT (0.74) have the greatest near-term scaling potential, whereas BIM appears comparatively mature, with a gap ratio of 0.055.

As shown in **Table 5,** the most valued benefits centre on transparency/traceability and cost reduction, with data security and operational efficiency also ranking highly.

Table 5: Perceived Benefits of Digital Technologies

Benefit	Count (n=86)	%
Enhanced transparency & traceability	63	73%
Cost reduction	62	72%
Increased data security	58	67%
Improved operational efficiency	54	63%
Real-time monitoring/decision making	50	58%
Better collaboration & communication	45	52%

These findings confirm that the industry recognises cost, speed, and security as top priorities, reinforcing the strategic value of digital adoption. Nevertheless, to successfully implement these technologies, respondents highlighted several key requirements. As shown in **Table 6**, the most critical needs are greater cross-departmental integration and better staff training.

Table 6: Key Requirements for Digital Adoption

Requirement	Count (n=82)	%
Greater integration across departments	61	73%
Better staff training on digital tools	59	71%
Clearer regulatory guidelines/acceptance	60	71%
Stronger leadership & stakeholder support	57	68%
Increased investment in digital infrastructure	56	67%

This data confirms that success requires not just the right software but also supportive leadership, cohesive organisational processes, and a friendly policy environment. In addition, willingness to adopt is generally positive, with responses indicating somewhat/very willing (60%), neutral (29%), and unwilling (10%), suggesting readiness if capability and governance constraints are addressed.

In the qualitative analysis, open-ended responses from a subset of the same survey sample (n=30) were synthesised into four themes:

- Staff Training and Knowledge-Building: emphasising skill gaps, need for training courses, and even calls for dedicated KM departments.
- Centralised Systems: advocating for centralised platforms and integrated systems to reduce fragmentation.
- Leveraging New Technologies: highlighting AI-driven systems, advanced KMS, blockchain-based solutions, and the use of modern communication tools.
- Leadership Support and Cultural Adoption: stressing that top management support and cultural buy-in are essential for sustainable change.

These themes indicate that technology adoption must be matched with skills development, process formalisation, and organisational support. Overall, these findings answer **RQ1** by characterising current knowledge-sharing practices and the use of enabling digital technologies in permitting and compliance checking, and by quantifying key gaps (e.g., moderate effectiveness and lower external knowledge sharing).

To address **RQ2**, concept mapping was used to identify and represent the relationships between key knowledge-sharing concepts in digital permit certification and compliance checking within a refined hierarchical conceptual map. The synthesis integrates four connected areas: (i) Digital Technologies (BIM, DT, BC, and AI/ML), (ii) KM (KMS, documentation consistency, and knowledge-sharing practices), (iii) Compliance and Assurance Processes (permit certification, automated compliance checking, continuous compliance/operational feedback, and data security), and (iv) Strategic Barriers (communication gaps, cultural resistance, regulatory barriers, infrastructure deficiencies, and training and technical expertise gaps). The resulting conceptual components, grounded in the survey evidence, are summarised in **Table 7**, providing traceability from the empirical findings to the structure shown in **Figure 4**.

Table 7: Concept-mapping components and survey evidence

Conceptual component	Empirical justification (survey evidence)
BIM for design-stage compliance checking; DTs for lifecycle monitoring	BIM 93%; DTs 44% (gap ratio 0.74)
BC for data security, traceability, and auditable permit certification	51% adoption; data security 67%; transparency/traceability 73%
AI-driven knowledge extraction and KMS enhancement	79% adoption; qualitative theme on AI-driven systems and advanced KMS
KMS, documentation consistency, and integrated cross-department knowledge sharing	KMS 57%; CDE 59%; inconsistency 66%; effectiveness gap 34.8%; paper-based 31%
Organisational and regulatory enablement	Communication gaps 91%; expertise deficits 87%; resistance 76%; infrastructure 69%; regulation 43%

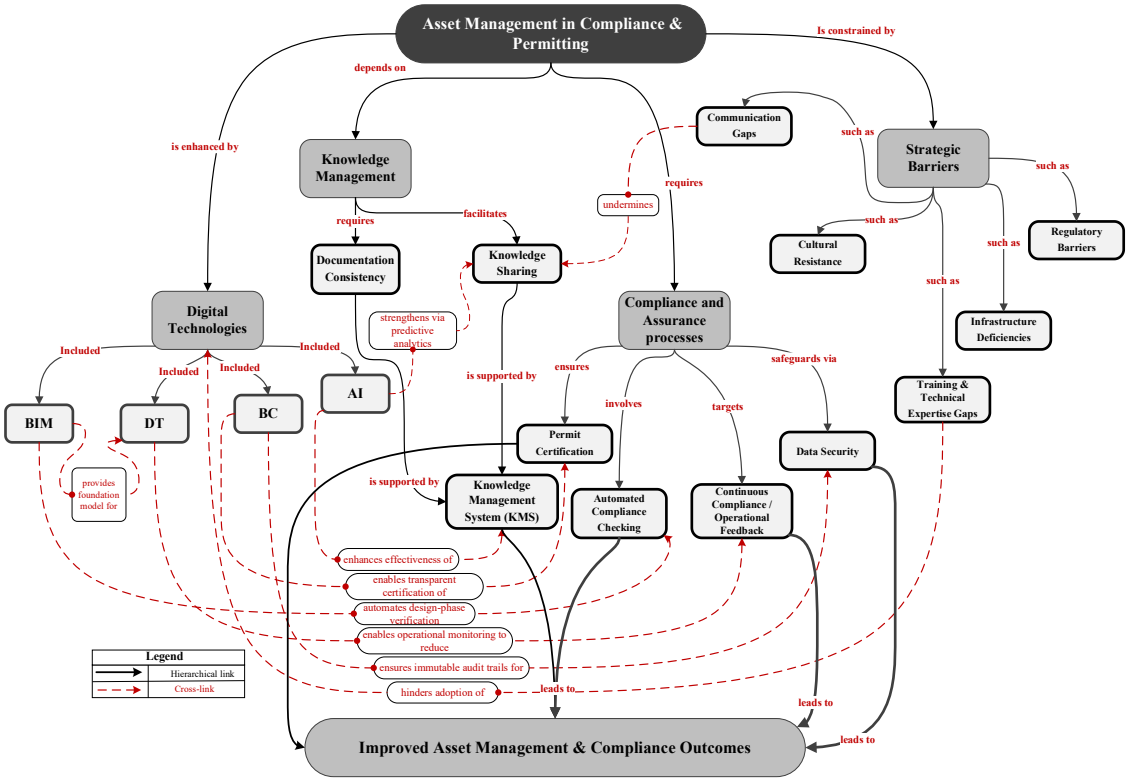


Figure 4: Refined hierarchical concept map for compliance-oriented asset management

Together, these components form an empirically grounded conceptual mapping linking technologies, knowledge processes, and implementation conditions. BIM's design-stage role is distinguished from DT's lifecycle-oriented monitoring role, with knowledge sharing being re-established as a cross-cutting theme. A lifecycle cross-link connects BIM-based design information to an emerging operational DT beyond the permit stage, and in this context, BC supports immutable audit trails and auditable permit certification, while AI contributes to knowledge extraction and KMS enhancement.

Within permitting, DTs are treated as as-designed information structures derived from BIM and documentation, aligned with ISO 19650's information-management principles (BS ISO 19650-1, 2018), rather than as mature permit-stage substitutes. They provide a practical foundation for later operational DT development and continuity from design information to inspection and monitoring. Although not yet widespread, survey results indicate emerging interest, with DT adoption at 44%. **Figure 4** integrates these relationships.

Overall, this synthesis answers **RQ2** by representing the relationships between key knowledge-sharing concepts in digital permit certification and compliance checking through the refined hierarchical conceptual map (**Figure 4**), supported by the survey evidence in **Table 7**.

Conclusions

This study examined how knowledge sharing and digital technologies are currently used for permitting and compliance, and how these findings inform a hierarchical conceptual map of knowledge-sharing practices for permitting and compliance checking in construction asset management. The analysis draws on a survey of 110 professionals, with 30 respondents from the same sample providing open-ended responses for reflexive thematic analysis. In answer to **RQ1**, knowledge sharing on permitting and compliance remains reliant on conventional tools such as email and document repositories, with external sharing notably limited. Digital technology adoption is uneven: BIM and AI/ML are well-established, while BC and DTs remain emerging, despite recognised benefits in transparency, cost reduction, and data security. These patterns are constrained by interdepartmental communication gaps, limited technical expertise, and cultural resistance.

To address **RQ2**, the empirical findings were synthesised through concept mapping to represent the relationships among key knowledge-sharing concepts in digital permit certification and compliance checking within a refined hierarchical conceptual map.

The study has limitations, as the sample may not reflect global construction practice, self-reported data may introduce subjectivity, and the cross-sectional design cannot capture temporal changes in adoption. Correlations between knowledge-sharing effectiveness and mechanism type, or between sharing frequency and organisational size, were not examined. As DT use in permitting is still emerging, its internalisation benefits remain prospective. Future work should operationalise the conceptual components in prototype solutions, validate them through case-based and longitudinal studies, and explore such correlations to provide more granular insights into optimal knowledge-sharing configurations. Overall, the study clarifies current practice and provides a

structured conceptual map to guide digital permit certification and compliance checking.

Acknowledgments

I gratefully acknowledge Professor Thomas Beach for his guidance and Professor Yacine Rezgui and the School of Engineering at Cardiff University for their support. I also extend my thanks to the General Secretariat of the Supreme Council for Planning and Development in Kuwait for its sponsorship.

References

Agresti, A. (2019) An introduction to categorical data analysis, EBSCOhost (ed), Wiley series in probability and statistics, book, Third edition., Hoboken, NJ, John Wiley & Sons.

Aldmour, R. (2012) Comparative study: The Nonaka model of knowledge management, International Journal of Engineering and Advanced Technology (IJEAT), no. 1 [Online]. Available at https://www.researchgate.net/publication/352365513.

Alsaffar, A., Beach, T. and Rezgui, Y. (2024) "Revolutionising Digital Construction: A Conceptual Framework from the Integration of DTs and BC for Enhanced Assets Management to Share Knowledge," 2024 IEEE Conference on Engineering Informatics, ICEI 2024, Institute of Electrical and Electronics Engineers Inc. [Online]. DOI: 10.1109/ICEI64305.2024.10912145.

Alsaffar, A., Beach, T. and Rezgui, Y. (2025) "Enhancing asset management: Integrating digital twins for continuous permitting and compliance - A systematic literature review," Journal of Building Engineering, Elsevier Ltd, vol. 99 [Online]. DOI: 10.1016/j.jobe.2024.111515.

Ataide, M., Braholli, O. and Siegele, D. (2023) "Digital Transformation of Building Permits: Current Status, Maturity, and Future Prospects," BUILDINGS, vol. 13, no. 10 [Online]. DOI: 10.3390/buildings13102554.

Beach, T., Yeung, J., Nisbet, N. and Rezgui, Y. (2024) "Digital approaches to construction compliance checking: Validating the suitability of an ecosystem approach to compliance checking," Advanced Engineering Informatics, vol. 59 [Online]. DOI: 10.1016/j.aei.2023.102288.

Bloch, T. and Fauth, J. (2023) "The unbalanced research on digitalization and automation of the building permitting process," ADVANCED ENGINEERING INFORMATICS, vol. 58 [Online]. DOI: 10.1016/j.aei.2023.102188.

Boamah, F. A., Zhang, J., Wen, D., Sherani, M., Hayat, A. and Horbanenko, O. (2022) "Enablers of knowledge management: practical research-based in the construction industry," International Journal of Innovation Science, Emerald Group Holdings Ltd., vol. 14, no. 1, pp. 121–137 [Online]. DOI: 10.1108/IJIS-09-2020-0142.

BS ISO 19650-1 (2018) Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM). Information management using building information modelling. Concepts and principles. BSI

Byrne, D. (2022) "A worked example of Braun and Clarke's approach to reflexive thematic analysis," Quality & quantity, article, Dordrecht, Springer Netherlands, vol. 56, no. 3, pp. 1391–1412 [Online]. DOI: 10.1007/s11135-021-01182-y.

Cañas, A. J. and Novak, J. D. (2008) The Theory Underlying Concept Maps and How to Construct and Use Them 1, [Online]. Available at http://cmap.ihmc.us/Publications/.

Cheng, J., Huang, L., Jiang, L., Chen, J., Chen, W. and He, Y. (2023) "Fostering Knowledge Collaboration in Construction Projects: The Role of BIM Application," Buildings, MDPI, vol. 13, no. 3 [Online]. DOI: 10.3390/buildings13030812.

Chihara, L. (2011) Mathematical statistics with resampling and R, Hesterberg, T. (ed), New York Academy of Sciences, book, 1st edition., Hoboken, New Jersey, Wiley.

Chumg, H. F., Cooke, L., Fry, J. and Hung, I. H. (2014) "Factors affecting knowledge sharing in the virtual organisation: Employees' sense of well-being as a mediating effect," Computers in Human Behavior, Elsevier Ltd, vol. 44, pp. 70–80 [Online]. DOI: 10.1016/j.chb.2014.11.040.

Creswell, J. W. (2023) Research design : qualitative, quantitative, and mixed methods approaches, Creswell, J. D. (ed), book, Sixth edition., Los Angeles ; London, SAGE.

Daniel, E. I., Oshodi, O. S., Nwankwo, N., Emuze, F. A. and Chinyio, E. (2024) "Barriers to the Application of Digital Technologies in Construction Health and Safety: A Systematic Review," Buildings, Multidisciplinary Digital Publishing Institute (MDPI), vol. 14, no. 8 [Online]. DOI: 10.3390/buildings14082386.

Deng, H., Xu, Y., Deng, Y. and Lin, J. (2022) "Transforming knowledge management in the construction industry through information and communications technology: A 15-year review," Automation in Construction, Elsevier B.V., vol. 142 [Online]. DOI: 10.1016/j.autcon.2022.104530.

Dudovskiy, J. (2024) The Ultimate Guide to Writing a Dissertation in Business Studies: A Step-by-Step Assistance,.

Fauth, J., Pasetti Monizza, G. and Malacarne, G. (2023) "Understanding processes on digital building permits–a case study in South Tyrol," Building Research and Information, Routledge, vol. 51, no. 5, pp. 518–532 [Online]. DOI: 10.1080/09613218.2023.2178372.

Gandomani, T. J. and Tavakoli, Z. (2016) Adaptation of Nanoka's and Takeuchi's knowledge management model to Scrum roles, International Journal of Software Engineering and Technology [Online]. Available at https://www.researchgate.net/publication/311404583.

Hagedorn, P., Fauth, J., Zentgraf, S., Seiß, S., König, M. and Brilakis, I. (2025) "OntoBPR: An ontology-based framework for performing building permit reviews using standardized information containers," Advanced Engineering Informatics, Elsevier Ltd, vol. 66 [Online]. DOI: 10.1016/j.aei.2025.103369.

He, C., Liu, M., He, W., Wang, Z. and Hsiang, S. (2025) "Addressing the information silo of construction planning constraints – a graph retrieval augmented generation approach," Engineering, Construction and Architectural Management, Emerald Publishing [Online]. DOI: 10.1108/ECAM-08-2024-1066.

Klaus, S. (2016) The Fourth Industrial Revolution: what it means and how to respond.

Kremelberg, D. (2010) "PEARSON'S R, CHI-SQUARE, T-TEST, AND ANOVA," in Practical Statistics, bookitem, United States, SAGE Publications, Incorporated, pp. 119–204 [Online]. DOI: 10.4135/9781483385655.n4.

Maheshwari, S., Ochoa, J. J., Gu, N., Rameezdeen, R., Mayer, W. and Doe, R. M. (2025) "Improving Information Sharing in Offsite Construction (OSC): A Systematic Literature Review," Buildings, Multidisciplinary Digital Publishing Institute (MDPI), vol. 15, no. 5 [Online]. DOI: 10.3390/buildings15050698.

Nonaka, I. and Takeuchi, H. (1995) The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation, Oxford University Press [Online]. DOI: 10.1093/oso/9780195092691.001.0001.

Peng, J. and Liu, X. (2023) "Automated code compliance checking research based on BIM and knowledge graph," Scientific Reports, Nature Research, vol. 13, no. 1 [Online]. DOI: 10.1038/s41598-023-34342-1.

Rezgui, Y. (2001) "Review of information and the state of the art of knowledge management practices in the construction industry," Knowledge Engineering Review, vol. 16, no. 3 [Online]. DOI: 10.1017/S026988890100008X.

Rezgui, Y. (2011) Harvesting and managing knowledge in construction : from theoretical foundations to business applications, Miles, J. (ed), book, Abingdon, Oxon ;, Spon Press.

Rinchen, S., Banihashemi, S. and Alkilani, S. (2024) "Driving digital transformation in construction: Strategic insights into building information modelling adoption in developing countries," Project Leadership and Society, Elsevier Ltd, vol. 5 [Online]. DOI: 10.1016/j.plas.2024.100138.

Saunders, M. N. K. (2019) Research methods for business students, Lewis, P. and Thornhill, A. (eds), book, Eighth edition., Harlow, Pearson.

Shull, F., Mendonç, M. G., Br, A. M., Carver, J., Maldonado, J. C. and Jeffery, R. (2004) Knowledge-Sharing Issues in Experimental Software Engineering, Empirical Software Engineering, Kluwer Academic Publishers, vol. 9.

Siddiqui, F. H., Thaheem, M. J. and Abdekhodaee, A. (2023) "A Review of the Digital Skills Needed in the Construction Industry: Towards a Taxonomy of Skills," Buildings, Multidisciplinary Digital Publishing Institute (MDPI), vol. 13, no. 11 [Online]. DOI: 10.3390/buildings13112711.

Sik-wah Fong, P., Fong, P. S. and Tong Levett, I. N. (2005) Managing Knowledge in Construction Disputes, [Online]. Available at https://www.researchgate.net/publication/283257047.

Wang, H. and Meng, X. (2021) "BIM-Supported Knowledge Management: Potentials and Expectations," Journal of Management in Engineering, American Society of Civil Engineers (ASCE), vol. 37, no. 4 [Online]. DOI: 10.1061/(asce)me.1943-5479.0000934.

Wang, Z., Wang, N. and Liang, H. (2014) "Knowledge sharing, intellectual capital and firm performance," Management Decision, Emerald Group Publishing Ltd., vol. 52, no. 2, pp. 230–258 [Online]. DOI: 10.1108/MD-02-2013-0064.

Xiahou, X., Chen, G., Li, Z., Xu, X. and Li, Q. (2025) "Knowledge management in construction quality management: current state, challenges, and future directions," IEEE Transactions on Engineering Management, Institute of Electrical and Electronics Engineers Inc. [Online]. DOI: 10.1109/TEM.2025.3550354.

Yang, F. and Zhang, J. (2024) "Prompt-based automation of building code information transformation for compliance checking," Automation in Construction, Elsevier B.V., vol. 168 [Online]. DOI: 10.1016/j.autcon.2024.105817.

Yeboah, A. (2023) "Knowledge sharing in organization: A systematic review," Cogent Business and Management, Cogent OA, vol. 10, no. 1 [Online]. DOI: 10.1080/23311975.2023.2195027.

Yepes, V. and López, S. (2021) "Knowledge management in the construction industry: Current state of knowledge and future research," Journal of Civil Engineering and Management, VGTU, vol. 27, no. 8, pp. 671–680 [Online]. DOI: 10.3846/jcem.2021.1600.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

BIMSTRUCT3D: A FULLY AUTOMATED HYBRID LEARNING SCAN-TO-BIM PIPELINE WITH INTEGRATED TOPOLOGY REFINEMENT

Mahdi Chamseddine[1,2], Fabian Kaufmann[2], Marius Schellen[2], Christian Glock[2],
Didier Stricker[1,2], and Jason Rambach[1]
[1]German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
[2]RPTU Kaiserslautern-Landau, Kaiserslautern, Germany

Abstract

Automatic generation of Building Information Models (BIM) from building scans is a key challenge in architecture and construction. We present a modular pipeline for generating IFC-compliant BIM from 3D point clouds. The hybrid approach combines learning-based semantic segmentation with topology-aware geometric reconstruction to model structural elements accurately. We propose vIoU, adapting voxel-based overlap evaluation to Scan-to-BIM by enabling holistic, instance-matching-free comparison of reconstructed and ground-truth models. We release the German Hospital dataset (DeKH), including high-resolution point clouds, ground truth BIMs, and semantic annotations. Experiments on DeKH and CV4AEC datasets show significant improvements over a RANSAC-based baseline, demonstrating robustness and scalability.

Introduction

Capturing spatial data of existing built environments can be efficiently achieved using mobile (Kanayama et al., 2025) and terrestrial scanning systems (Chang et al., 2017; Armeni et al., 2017). To leverage such data in construction (Wu et al., 2021) and facility management (Xu et al., 2021), it must be converted into semantic models in accordance with the Building Information Modelling (BIM) methodology. Given the volume of data and the complex steps required to transform raw 3D point clouds into BIM models, full automation of this process is highly desirable. However, this transformation remains a challenge.

Despite the importance of the task, scan-to-BIM is rarely treated as a comprehensive, end-to-end problem with standardized benchmarks; however, some effort is being made through initiatives like the CV4AEC challenges (Armeni et al., 2024). In this work, we introduce BIMStruct3D, one of the first comprehensive frameworks for automating the conversion of noisy and incomplete 3D point clouds into IFC-compatible BIM models. Our BIMStruct3D pipeline employs a hybrid approach that integrates both learning-based and classical point cloud processing techniques. It includes modules for 3D point cloud segmentation, geometric reconstruction and primitive extraction, as well as model post-processing, as illustrated in Figure 1.

BIMStruct3D is designed to process data from multi-storey buildings such as offices, hospitals, and residential complexes, and generate BIM representations of architectural

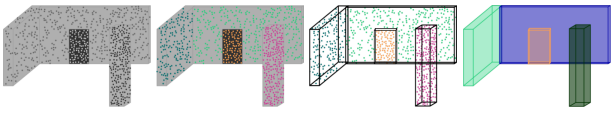


Figure 1: BIMStruct3D is a hybrid pipeline for generating IFC BIM models from 3D point cloud scans

elements including walls, doors, and columns. Accompanying our framework is `pystruct3d`, a modular library tailored for processing 3D point cloud data of buildings.
To support reproducibility and facilitate benchmarking in the scan-to-BIM domain, we provide a comprehensive evaluation of our method at both modular and end-to-end levels. This includes tests on publicly available datasets and newly acquired data, featuring high-precision, handcrafted ground truth BIMs. We also propose volumetric Intersection over Union (vIoU), which adapts voxel-based IoU evaluation to the Scan-to-BIM domain. Unlike per-object voxel IoU used in 3D shape reconstruction, vIoU operates at the class level without instance matching, making it robust to the fragmented and variably segmented wall representations common in BIM reconstruction.
In real-world scenarios, building point clouds are often incomplete due to factors such as the limitations of scanning technology, human error and environmental conditions like lighting, occlusion, and material reflectivity (e.g. glass surfaces). BIMStruct3D is built to handle such noisy and incomplete data to create usable BIM representations.
In summary, our contributions are as follows:

- BIMStruct3D, a fully automated hybrid pipeline generating IFC BIM models from 3D point clouds using semantic segmentation and geometric reconstruction.
- Specialized algorithms for primitive extraction and BIM model generation.
- vIoU, an adaptation of voxel-based IoU to Scan-to-BIM evaluation, enabling class-level comparison without instance-level matching.
- `pystruct3d`[1], an open-source library for processing 3D building point clouds, released with DeKH[2], a new dataset including high-resolution scans, semantic annotations, and handcrafted ground truth BIMs.

[1]`https://github.com/humantecheu/pystruct3d`
[2]`https://huggingface.co/datasets/RPTU-FGMB/DeKH`

Related Work

Point Cloud Semantic Segmentation

Semantic segmentation of point clouds is a fundamental computer vision task, currently dominated by deep neural networks. Early approaches include projection-based (Chen et al., 2017), voxel-based (Graham et al., 2018), and point-based networks (Qi et al., 2017), each trading off between computational cost, geometric fidelity, and flexibility. More recent developments have incorporated self-attention mechanisms, demonstrating strong performance in large-scale 3D scene understanding (Wang et al., 2019). Point Transformer by Zhao et al. (2021) advanced point-based segmentation by integrating local self-attention (Vaswani et al., 2017), vector attention (Zhao et al., 2020), and appropriate positional encodings. Subsequent iterations (Wu et al., 2024) introduce architectural improvements leading to better segmentation accuracy and model efficiency.

Scan-to-BIM Approaches

To align with the scope of this work, we focus on research targeting fully automated (end-to-end) scan-to-BIM conversion from point clouds. Gourguechon et al. (2022) and Bassier and Vergauwen (2020) distinguish between room-based and wall-based (or structural component-based) reconstruction approaches. The former identifies rooms and then reconstructs surrounding elements, while the latter directly targets structural components such as walls. Unfortunately, to this day, there are no open-source implementations of complete scan-to-BIM pipelines or benchmark datasets for evaluation.

Room-based reconstruction

Room-based methods segment indoor spaces and reconstruct surrounding elements. Representative approaches address plane fitting and wall estimation (Ochmann et al., 2016, 2019), 2D region growing with slab reconstruction (Macher et al., 2017), comprehensive element reconstruction including windows and fixtures (Xiong et al., 2023), hybrid IFC wall reconstruction with topology refinement (Bassier et al., 2018), and learning-based segmentation followed by space decomposition (Tang et al., 2022; Hu et al., 2020). While these methods offer good indoor segmentation, Bassier and Vergauwen (2020) note their limitations in modelling non-room structures and handling clutter. Thus, component-based approaches targeting structural elements have gained traction.

Structural component reconstruction

Component-based reconstruction identifies and models structural elements like walls and columns through clustering, semantic segmentation, and primitive fitting. Bassier and Vergauwen (2020) focused on extracting standard wall objects with topological rules for merging, while follow-up work (Bassier et al., 2020) compared 2D and 3D reconstruction methods. Other approaches combine 2D and 3D processing with point cloud upsampling (Gankhuyag and Han, 2021), use RANSAC and convex hull-based boundary extraction (Thomson and Boehm, 2015), reconstruct walls and spaces from enclosed rooms (Anagnostopoulos et al., 2016), or segment concrete components via concavity/convexity criteria (Son and Kim, 2017). The closest work to ours is by Kim and Kim (2021), which uses synthetic data to train a segmentation model, followed by instance clustering and planar patch segmentation. Element relationships are modelled with graph networks to infer missing links and extend incomplete elements.

Our approach differs in that we do not rely on synthetic data and instead address several common limitations, including limited generalization and reliance on manual parameter tuning. We integrate semantic segmentation, geometric reconstruction, and topological reasoning into a unified scan-to-BIM pipeline.

Methodology

Our method, outlined in Figure 2, consists of semantic and instance segmentation, geometry reconstruction, topology refinement, and BIM object creation from a 3D point cloud. IFC export is handled via *IfcOpenShell*[3], generating a project–site–building–storey hierarchy and assigning reconstructed geometry as corresponding IFC entities (`IfcWall`, `IfcDoor`, `IfcColumn`). During the reconstruction process, we adopt the Manhattan world assumption (Coughlan and Yuille, 1999), which holds for the majority of structural objects in typical building environments.

Initially, levels/storeys are identified by analysing vertical point density distributions, using surfaces like floors and ceilings to detect significant point concentrations (Bassier et al., 2020; Macher et al., 2017). Semantic segmentation using state-of-the-art deep learning models (Wu et al., 2024) serves as the basis for reconstructing walls, doors, and columns. The reconstruction pipeline follows a common structure: instance segmentation, geometry reconstruction, topology refinement, and IFC BIM object creation.

Wall reconstruction is based on axis-aligned point grouping. DBSCAN clustering (Ester et al., 1996) is applied per axis to form spatial clusters, discarding sparse noise as seen in Figure 3. HYSAC, our RANSAC (Schnabel et al., 2007) based plane segmentation algorithm with optimized seed selection, is used to extract wall surfaces. These are enclosed using Horizontal Oriented Bounding Boxes (H-OBB), followed by topological refinements to ensure geometric correctness.

Door points are initially identified via semantic segmentation and then treated as child objects of walls, relying on the reconstructed wall geometry as a starting point. The process first identifies door points within the wall bounding boxes, which typically correspond to frames, linings, or closed door leaves captured during scanning. If such points are found, the bounding box is expanded to include additional door geometry, such as open door leaves, ensuring doors are correctly associated with their respective parent walls. Our two-step approach reduces errors caused by

[3] `https://ifcopenshell.org/`

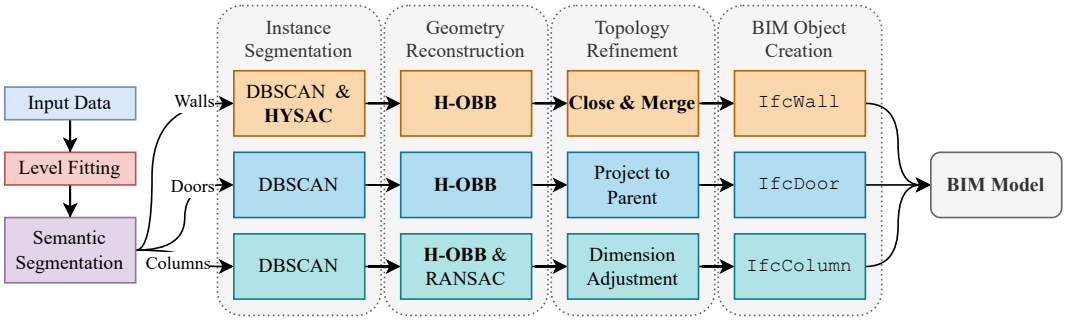


*Figure 2: Overview of the reconstruction pipeline showing the different reconstruction stages for walls, doors, and columns. In **bold** are our main contributed components.*

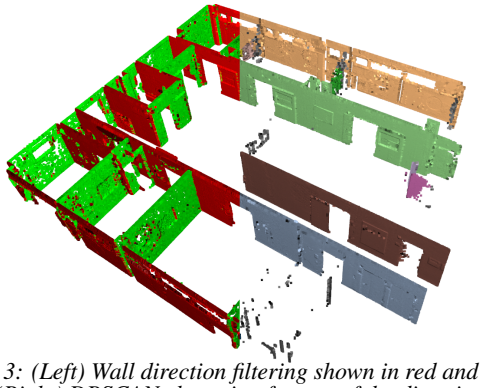


Figure 3: (Left) Wall direction filtering shown in red and green and (Right) DBSCAN clustering for one of the directions in multiple colours.

open doors adjacent to perpendicular walls.

Column points are clustered using DBSCAN (Ester et al., 1996) and filtered by minimum point thresholds to reduce artifacts. Shape classification is achieved through curvature analysis based on eigenvalue decomposition from k-d tree neighbourhood searches. Curvature distributions are modelled with Maximum Likelihood Estimation to distinguish round (low standard deviation) from square or rectangular columns (high standard deviation). We then use RANSAC cylinder fitting for round columns and H-OBB fitting square/rectangular columns.

Semantic Segmentation

Semantic segmentation is essential for producing accurate IFC models, as it forms the foundation for all subsequent steps. High-quality segmentation ensures later geometry reconstruction stages are based on reliable object labels.

We use Point Transformer V3 by Wu et al. (2024), trained on a combined dataset from three public datasets: S3DIS (Armeni et al., 2016), Structured3D (Zheng et al., 2020), and ScanNet (Dai et al., 2017). Training on a joint dataset prevents the model from over-fitting to the data and sensor used in one dataset. To further improve generalization, additional data augmentation techniques were used, such as removing color information to simulate scans from grayscale 3D sensors. This is particularly relevant because the CV4AEC datasetused for the evaluation contains colored and grayscale point clouds. This segmentation pipeline ensures robust performance across diverse scan types and scene complexities, enabling reliable downstream reconstruction of BIM elements.

Topology aware geometry reconstruction

The reconstruction pipeline leverages both state-of-the-art techniques and our novel contributions. The following sections describe our key innovations in wall reconstruction, oriented bounding box fitting, and topology correction.

HYSAC wall reconstruction

The RANSAC-based plane fitting methods (Schnabel et al., 2007) are highly sensitive to parameter settings like the minimum number of inliers and distance thresholds. Improper parameter tuning often results in under-segmentation, where a single plane, instead of two, is incorrectly fitted to an entire wall instance.

To address this, we propose Hypothesis-based Sample Consensus (HYSAC), a modified RANSAC approach that improves seed point selection. Instead of sampling seed points randomly, HYSAC benefits from the point density distribution perpendicular to the primary wall orientation. Inspired by Armeni et al. (2016), we compute a 1D density histogram along the axis orthogonal to the wall surface and select seed points from regions of highest density. Planes are fitted using Singular Value Decomposition (SVD), which provides a least-squares optimal plane fit robust to point cloud noise, and accepted based on a minimum inlier ratio. Algorithm 1 shows a pseudo code implementation of HYSAC.

As shown in Figure 4, HYSAC reliably detects multiple planes within a wall cluster, distinguishing between valid wall surfaces and outliers. The resulting inlier points are grouped and forwarded for further geometric processing.

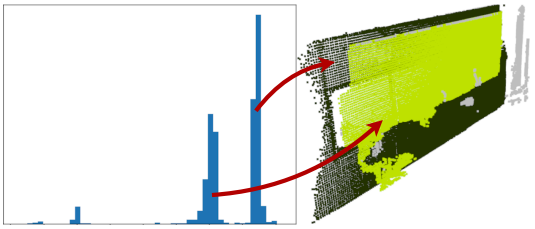


Figure 4: HYSAC plane fitting. Result of plane fitting (plane points coloured, outliers grey) with corresponding histogram of point distribution.

Algorithm 1: HYSAC plane fitting pseudocode.

```
input  :wall_points # points of a wall cluster
input  :n_bins      # number of histogram bins
input  :n_seeds     # number of seed points
input  :min_points  # minimum points per plane
output :planes      # equations of found planes
output :inliers     # points belonging to the planes
def histogram_seeds(wall_points, n_seeds):
    # generate the histogram perpendicular to the
      wall direction
    hist = histogram(cluster_points, n_bins)
    peak = find_peaks(hist)
    # select random points from peaks
    seed_points = random_sample(peak, n_seeds)
    return seed_points
end
def hysac_plane(wall_points, n_seeds):
    points = wall_points
    # loop until all planes are found
    while number_of(points) ≥ min_points :
        seeds = histogram_seeds(points, n_seeds)
        plane, inliers = fit_plane_svd(seeds)
        if number_of(inliers) ≥ min_points :
            # remove inliers of the plane from the
              wall points
            points = remove(points, inliers)
            yield plane, inliers
        end
    end
end
planes, inliers = hysac_plane(wall_points, n_seeds)
return planes, inliers
```

Horizontal Oriented Bounding Boxes Fitting

To approximate the geometry of wall instances, we fit Horizontal Oriented Bounding Boxes (H-OBB) to the previously segmented plane clusters. Given that architectural structures typically conform to horizontal and vertical alignments, we restrict the bounding boxes to axis-aligned orientations within the XY-plane.

The H-OBB fitting procedure follows a 2D projection strategy. All segmented points are projected onto the XY-plane, and a convex hull is computed. The convex hull is restricted to a quadrilateral, and each edge is evaluated as a potential candidate for the bounding box length. We rotate each edge to align with the X-axis, compute the minimum-area axis-aligned bounding box, and then select the configuration with the smallest area. The inverse rotation is applied to return the bounding box to its original orientation. This process is visualized in Figure 5.

Topology Correction

Even with high-quality plane and bounding box fitting, errors in wall topology often persist, typically caused by 3D scanning noise. We implement a topology-aware correction stage to ensure consistent and realistic wall geometries. Three main correction operations are introduced:

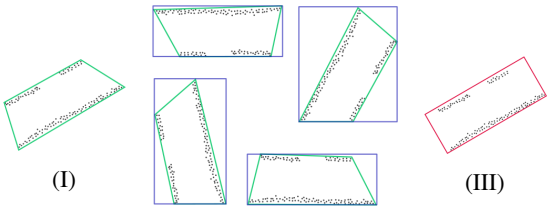


Figure 5: H-OBB fitting procedure on 2D projected data. Left: 2D projected input data with convex hull. Centre: Axis-aligned bounding box fitted to each edge rotated parallel to the X-axis. Right: Inverse rotation to minimal bounding box, final result.

- **Intersection Correction:** Perpendicular bounding boxes that intersect must be clipped or extended to form clean corner connections.
- **Merging:** Collinear bounding boxes with adjacent or overlapping baselines must be merged into longer walls to reduce redundancy.
- **Redundancy Removal:** Smaller boxes fully enclosed by larger ones are treated as artifacts and removed.

Corrections are based on geometric analysis (adjacency, parallelism, orthogonality) as well as domain knowledge. Perpendicular pairs are found by analysing centrelines of all bounding boxes and computing intersection points. If the endpoint of a bounding box lies within a certain distance r of an intersection, the box is clipped or extended accordingly as shown in Figure 6a.

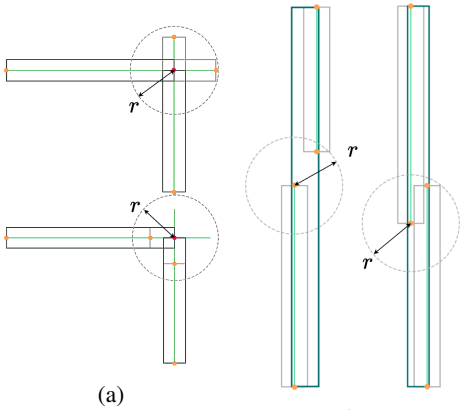


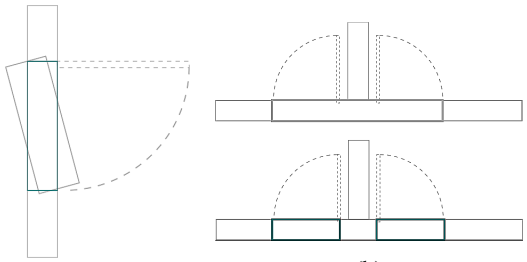
Figure 6: Wall bounding box refinement. (a) Extending and clipping bounding boxes. (b) Merging bounding boxes.

For merging, bounding boxes with endpoints within a distance threshold are joined by averaging the height and Z-coordinates of the endpoints while preserving horizontal alignment as shown in Figure 6b. Intersection correction and merging procedures are applied iteratively, as each operation may introduce new merge or correction candidates. The distance thresholds for these operations (e.g., intersection radius r, merge distance) are set once per project based on domain knowledge such as typical wall widths and corridor dimensions.

Finally, to refine door geometry, we project door bounding boxes into their parent wall to ensure alignment as seen in Figure 7a. For oversized doors, which may arise from

clustering errors, we apply a width threshold and split them into multiple instances. Figure 7b shows the corrected doors after offsets were applied to ensure a realistic spacing between the door instances.

These algorithms yield a geometrically and topologically coherent representation of walls and doors, forming a robust foundation for accurate IFC models.



(a) (b)

Figure 7: Bounding box refinement. (a) projecting the bounding box into the parent geometry. (b) splitting bounding boxes too wide.

Evaluation

The primary method for assessing reconstruction accuracy is to compare reconstructed BIMs against ground truth models. While simple cuboid representations can be directly compared, walls and doors may be reconstructed as more complex geometries, making standard geometric comparison difficult. Therefore, we use several evaluation metrics to quantify reconstruction accuracy.

Metrics

The main metric used is the 3D-Intersection over Union (3D-IoU). However, due to its limitations in handling small spatial misalignments, we introduce the Volumetric Intersection over Union (vIoU) as an alternative.

3D-Intersection over Union

The 3D-Intersection over Union (3D-IoU) is calculated as the ratio of the volume of the intersection to the volume of the union of the reconstructed and ground-truth bounding boxes. This metric captures both shape and spatial alignment, providing a compact representation of how well two volumes overlap. However, 3D-IoU is highly sensitive to minor deviations in size or alignment: for thin elements such as walls, an offset on the order of the smallest dimension (e.g., wall thickness) already drives the intersection-to-union ratio well below 0.5, even when the reconstruction is otherwise correct. Furthermore, 3D-IoU requires instance-level assignment between predicted and ground-truth elements, which becomes ambiguous when walls are fragmented or merged during reconstruction.

Volumetric Intersection over Union

To address these limitations, we propose Volumetric Intersection over Union (vIoU), computed on a voxel grid. Voxel-based IoU is well established in 3D object reconstruction and occupancy prediction Agnew et al. (2021), where it evaluates per-object shape accuracy. However, its direct application to Scan-to-BIM is non-trivial, as standard per-object voxel IoU inherits the same matching dependency as 3D-IoU. Our vIoU addresses this by operating at the class level: all reconstructed geometry of a given type (e.g., walls) and all corresponding ground-truth geometry are voxelized jointly, and overlap is computed holistically without instance assignment. This is conceptually analogous to how semantic segmentation IoU evaluates per-class predictions in 2D, but applied to 3D volumetric BIM evaluation. We use a voxel size of $5\ cm$, corresponding to a tolerance of $\pm 2.5\ cm$, which we consider reasonable for the structural elements addressed here. The 3D space is discretized, and a voxel is marked as occupied if its centroid lies inside a bounding box. For both ground truth and reconstructed geometries, we count all occupied voxels and compute vIoU as the ratio of their intersection and union. An example is illustrated in Figure 8.

This voxel-based approach simplifies comparison by avoiding complex geometric overlap calculations, provides a sub-voxel tolerance to small misalignments, particularly relevant for thin elements such as walls, and eliminates the need for instance-level bounding box matching. The latter is particularly important for walls, which may appear as a single bounding box, be split into segments, or be fragmented due to incomplete scans. Unlike 3D-IoU, which would assign ground truth to only one predicted instance, vIoU evaluates overlap holistically at the voxel level.

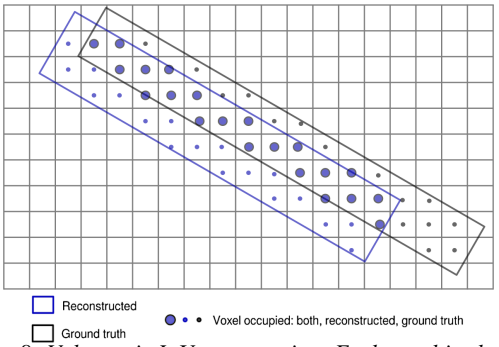


Figure 8: Volumetric IoU computation. Each voxel is classified as belonging to the reconstructed or ground-truth bounding box if its centroid lies within the corresponding geometry.

Results

We evaluate our reconstruction pipeline on two datasets: Our newly proposed German Hospital dataset (DeKH) and the CV4AEC Scan-to-BIM Challenge dataset. While DeKH features largely empty interiors, CV4AEC is fully furnished featuring a lot of clutter and complex occlusions. Furthermore, an ablation study is presented to investigate the impact of individual components in our pipeline and assess the effect of semantic segmentation accuracy.

CV4AEC

The CV4AEC dataset is more challenging due to its cluttered, incomplete scans and presence of interior furniture and occlusions. Six scans from the test set are used for evaluation, with ground-truth BIMs manually created from the input point clouds. Results are presented in Table 1.

Table 1: Results of the evaluation on the CV4AEC dataset test scenes. The vIoU is calculated for a 5 cm voxel size.

Metrics	Office_1	Office_2	Office_3	Office_4	Parking_1	Parking_2	***Mean %***	Segmentation IoU
Wall 3D-IoU [-]	0.347	0.358	0.261	0.297	0.280	0.318	*31.0*	0.672
Wall vIoU [-]	0.437	0.383	0.259	0.302	0.386	0.420	*36.5*	
Door 3D-IoU [-]	0.388	0.452	0.157	0.182	0.192	0.000	*22.9*	0.586
Door vIoU [-]	0.367	0.407	0.150	0.149	0.247	0.000	*22.0*	
Column 3D-IoU[-]	0.382	0.400	0.011	0.033	0.197	0.105	*18.8*	0.293
Column vIoU [-]	0.404	0.547	0.025	0.057	0.250	0.125	*23.5*	
Mean 3D-IoU [-]	0.372	0.403	0.143	0.171	0.223	0.141	*24.2*	0.517
Mean vIoU [-]	0.403	0.446	0.145	0.169	0.294	0.182	*27.3*	

In the first two scenes, (Office_1 and Office_2) with standard wall and door configurations, our pipeline performs best due to alignment with our assumptions. Scenes Office_3 and Office_4 contain open spaces, cubicles, and complex facade elements, making the reconstruction more challenging. The last two scenes (Parking_1 and Parking_2) are from underground parking structures. Although some improvement in column detection is observed, results remain lower than in the office scenes. Notably, no doors are detected in (Parking_2), likely due to closed doors during scanning and segmentation failure suggesting that a dedicated detection algorithm for doors might be beneficial. The segmentation IoU presented in the last column gives an indication of the challenge these types of data are posing to state-of-the-art point cloud segmentation methods.

German Hospital Dataset (DeKH)

Our DeKH dataset is a new public dataset including four scans from three buildings in an unused, largely unfurnished hospital building in Germany. Building **A**, over 100 years old, features structural walls, central corridors, and a symmetrical layout. Building **B** includes an Intensive Care Unit (ICU) area with integrated equipment, while building **C** contains empty surgical rooms with built-in furniture. In addition to the point clouds annotated following a construction ontology-based guideline (Kaufmann et al., 2023), we provide the manually created BIMs for all of the scans. The results of the reconstruction are presented in Table 2.

Table 2: Results of the evaluation on the DeKH dataset. The vIoU is calculated for a voxel size of 5 cm.

Metrics	**A** 1st floor	**A** 2nd floor	**B** ICU	**C** Surgery	***Mean %***
Wall 3D-IoU [-]	0.395	0.219	0.270	0.328	*30.3*
Wall vIoU [-]	0.553	0.463	0.388	0.451	*46.4*
Door 3D-IoU [-]	0.417	0.104	0.193	0.223	*23.4*
Door vIoU [-]	0.402	0.103	0.219	0.236	*24.0*
Mean 3D-IoU[-]	0.406	0.162	0.232	0.275	*26.9*
Mean vIoU [-]	0.478	0.283	0.304	0.344	*35.2*

Performance in the **A** building's second floor is lower than the first floor. The pipeline faced challenges in reconstructing smaller wall compartments. The door reconstruction had issues due to sparse segmentation. In the **B** ICU building, performance improves. Door reconstruction is accurate but scored poorly due to a mismatch in door representations between the ground truth of thin door leaves and the reconstructed geometry including frame and lining. In building **C**, some furniture surfaces were misclassified as walls, and some open doors lead to double reconstructions.

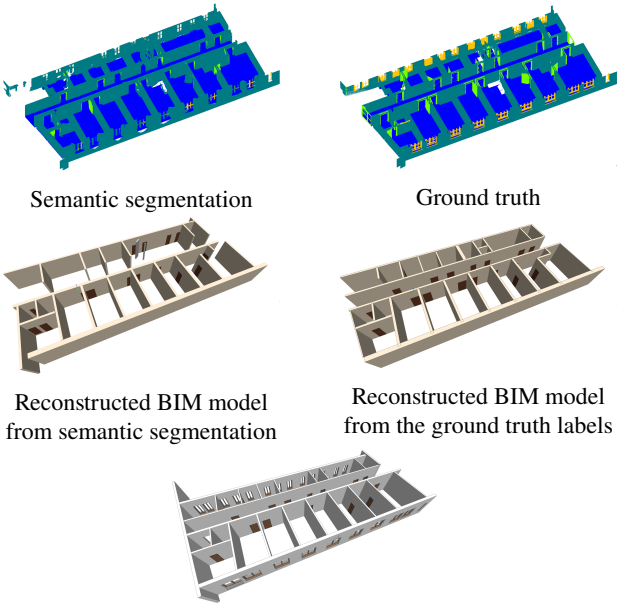


*Figure 9: Comparison of reconstructed BIM models from semantic segmentation vs ground truth in DeKH–building **A**.*

The ablation study in Table 3 presents three configurations: **Baseline**, which follows the reconstruction approach of Kaufmann et al. (2022) using RANSAC for primitive fitting and H-OBB estimation without topology correction; **Ours**, representing the full pipeline with HYSAC, H-OBB fitting, and topology refinement; and **GT labels**, which applies the same method as **Ours** but uses ground truth semantic labels to isolate the effect of segmentation quality. The results confirm our pipeline significantly improves wall and door reconstruction accuracy compared to the baseline. Notably, even with ground truth labels, wall reconstruction does not improve. Our approach achieves a 4.9% higher bounding box IoU while volumetric IoU for walls shows only a slight advantage with ground truth labels. The door reconstruction improves considerably with ground truth segmentation. These results highlight that, while our method can tolerate some incomplete segmentation especially for large structures like walls, accurate semantic segmentation remains crucial for smaller elements like doors. The

qualitative results in Figure 9 also support these findings.

*Table 3: Ablation study on the first floor of the DeKH **A** building.*

Metrics	Baseline	Ours	GT Labels
Wall 3D-IoU [-]	0.371	**0.395**	0.346
Wall vIoU [-]	0.466	**0.553**	0.571
Door 3D-IoU [-]	0.322	**0.417**	0.519
Door vIoU [-]	0.391	**0.402**	0.521
Mean 3D-IoU[-]	0.346	**0.406**	0.432
Mean vIoU [-]	0.429	**0.478**	0.546

Discussion and Conclusion

This work presents a robust and modular pipeline that advances the state of scan-to-BIM automation, generating IFC-compliant BIM models from 3D point clouds. The proposed pipeline achieves competitive reconstruction accuracy across diverse scenes, demonstrating its practical viability and scalability. We contribute to the topic with methodological advances and a new benchmark dataset.

Our approach delivers accurate reconstructions of structural elements using a hybrid methodology of learning-based and geometric processing techniques. Evaluation across two datasets confirms the pipeline's robustness and effectiveness, particularly for walls and doors, where reconstruction quality exceeds a RANSAC-based baseline. Even with incomplete or noisy data, the system maintained reliable performance. Direct comparison with existing scan-to-BIM methods is currently not feasible due to the absence of open-source implementations and standardized evaluation protocols. The CV4AEC challenge provides the closest available shared benchmark, where the proposed method was evaluated competitively. The release of DeKH aims to further support reproducible comparison in future work.

A major contribution is the introduction of the **DeKH** benchmark dataset. The release of DeKH, with high-resolution scans, ground truth BIMs, and semantic annotations, provides a new benchmark for scan-to-BIM.

Additionally, we propose vIoU, which adapts voxel-based IoU to class-level evaluation without instance matching, providing more robust comparison in scenes with misaligned, fragmented, or variably segmented elements.

To strengthen the pipeline, future efforts should expand beyond the current object classes toward more functional BIMs. The reconstruction algorithms are adaptable and could extend to elements such as windows, stairs, and mechanical, electrical, and plumbing (MEP) systems.

The topology refinement procedures showed sensitivity to parameter selection. For instance, these operations can introduce inaccuracies when small segments are wrongly merged. Thus, adaptive, context-aware refinement strategies or learned alternatives could address this vulnerability.

Our pipeline adopts the Manhattan world assumption, which constrains reconstructed geometry to predominantly orthogonal orientations. Consequently, it is best suited to buildings with rectilinear layouts, such as offices, hospitals, and residential complexes, which constitute the majority of typical building stock. Structures with significant non-orthogonal elements (such as curved walls, angled façades, or irregular floor plans) would not be handled correctly, as the axis-aligned clustering would either discard or misassign their points, and the bounding box fitting would produce poor approximations. Relaxing this assumption, for instance through data-driven orientation estimation or by supporting polygonal wall primitives, is an important direction for future work.

Looking ahead, extending the pipeline to support a broader range of object classes and incorporating subcomponent-level reasoning will be important steps toward producing complete and functionally rich BIMs. In parallel, research into end-to-end learning approaches that can infer BIM semantics and geometry directly from raw point cloud data and other sources will help move the field forward.

Acknowledgements

This research was funded by the European Union as part of the projects: HumanTech (Grant Agreement 101058236) and ShieldBOT (Grant Agreement 101235093).

References

Agnew, W., Xie, C., Walsman, A., Murad, O., Wang, Y., Domingos, P., and Srinivasa, S. (2021). Amodal 3d reconstruction for robotic manipulation via stability and connectivity. In Conference on robot learning. PMLR.

Anagnostopoulos, I., Belsky, M., and Brilakis, I. (2016). Object boundaries and room detection in as-is bim models from point cloud data. In ICCCBE.

Armeni, I., Che, E., Fischer, M., Hall, D., Jung, J., Li, F., Olsen, M., Pollefeys, M., Turkan, Y., Rastiveis, H., and et al. (2024). Computer vision in the built environment.

Armeni, I., Sax, S., Zamir, A. R., and Savarese, S. (2017). Joint 2d-3d-semantic data for indoor scene understanding. arXiv:1702.01105.

Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., and Savarese, S. (2016). 3d semantic parsing of large-scale indoor spaces. In CVPR. Ieee.

Bassier, M., Klein, R., Van Genechten, B., and Vergauwen, M. (2018). Ifc wall reconstruction from unstructured point clouds. ISPRS Annals.

Bassier, M. and Vergauwen, M. (2020). Unsupervised reconstruction of building information modeling wall objects from point cloud data. Automation in Construction.

Bassier, M., Yousefzadeh, M., and Vergauwen, M. (2020). Comparison of 2d and 3d wall reconstruction algorithms from point cloud data for as-built bim. ITcon.

Chang, A., Dai, A., Funkhouser, T., Halber, M., Niebner, M., Savva, M., Song, S., Zeng, A., and Zhang, Y. (2017).

Matterport3d: Learning from rgb-d data in indoor environments. In 3DV. IEEE.

Chen, X., Ma, H., Wan, J., Li, B., and Xia, T. (2017). Multi-view 3d object detection network for autonomous driving. In CVPR.

Coughlan, J. M. and Yuille, A. L. (1999). Manhattan world: Compass direction from a single image by bayesian inference. In ICCV.

Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., and Nießner, M. (2017). Scannet: Richly-annotated 3d reconstructions of indoor scenes. In CVPR.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In KDD. AAAI Press.

Gankhuyag, U. and Han, J.-H. (2021). Automatic bim indoor modelling from unstructured point clouds using a convolutional neural network. Intelligent Automation & Soft Computing.

Gourguechon, C., Macher, H., and Landes, T. (2022). Automation of as-built bim creation from point cloud: An overview of research works focused on indoor environment. ISPRS Annals.

Graham, B., Engelcke, M., and van der Maaten, L. (2018). 3d semantic segmentation with submanifold sparse convolutional networks. In CVPR. Ieee.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., and Markham, A. (2020). Randla-net: Efficient semantic segmentation of large-scale point clouds. In CVPR. Ieee.

Kanayama, H., Chamseddine, M., Guttikonda, S., Okumura, S., Yokota, S., Stricker, D., and Rambach, J. (2025). Tof-360-a panoramic time-of-flight rgb-d dataset for single capture indoor semantic 3d reconstruction. In CVPRW.

Kaufmann, F., Chamseddine, M., Guttikonda, S., Glock, C., Stricker, D., and Rambach, J. (2023). Ontology-based semantic labeling for rgb-d and point cloud datasets. In EC3. European Council on Computing in Construction.

Kaufmann, F., Glock, C., and Tschickardt, T. (2022). Scalebim: Introducing a scalable modular framework to transfer point clouds into semantically rich building information models. In EC3. European Council on Computing in Construction.

Kim, H. and Kim, C. (2021). 3d as-built modeling from incomplete point clouds using connectivity relations. Automation in Construction.

Macher, H., Landes, T., and Grussenmeyer, P. (2017). From point clouds to building information models: 3d semi-automatic reconstruction of indoors of existing buildings. Applied Sciences.

Ochmann, S., Vock, R., and Klein, R. (2019). Automatic reconstruction of fully volumetric 3d building models from oriented point clouds. ISPRS Annals.

Ochmann, S., Vock, R., Wessel, R., and Klein, R. (2016). Automatic reconstruction of parametric building models from indoor point clouds. Computers & Graphics.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In CVPR. Ieee.

Schnabel, R., Wahl, R., and Klein, R. (2007). Efficient ransac for point-cloud shape detection. Computer Graphics Forum.

Son, H. and Kim, C. (2017). Semantic as-built 3d modeling of structural elements of buildings based on local concavity and convexity. Advanced Engineering Informatics.

Tang, S., Li, X., Zheng, X., Wu, B., Wang, W., and Zhang, Y. (2022). Bim generation from 3d point clouds by combining 3d deep learning and improved morphological approach. Automation in Construction.

Thomson, C. and Boehm, J. (2015). Automatic geometry generation from point clouds for bim. Remote Sensing.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. NeurIPS.

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., and Solomon, J. M. (2019). Dynamic graph cnn for learning on point clouds. ACM ToG.

Wu, C., Yuan, Y., Tang, Y., and Tian, B. (2021). Application of terrestrial laser scanning (tls) in the architecture, engineering and construction (aec) industry. Sensors.

Wu, X., Jiang, L., Wang, P.-S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., and Zhao, H. (2024). Point transformer v3: Simpler faster stronger. In CVPR.

Xiong, B., Jin, Y., Li, F., Chen, Y., Zou, Y., and Zhou, Z. (2023). Knowledge-driven inference for automatic reconstruction of indoor detailed as-built bims from laser scanning data. Automation in Construction.

Xu, J., Chen, K., Xue, F., and Lu, W. (2021). 3d point cloud data enabled facility management: A critical review. In CRIOCM. Springer.

Zhao, H., Jia, J., and Koltun, V. (2020). Exploring self-attention for image recognition. In CVPR.

Zhao, H., Jiang, L., Jia, J., Torr, P., and Koltun, V. (2021). Point transformer. In ICCV. Ieee.

Zheng, J., Zhang, J., Li, J., Tang, R., Gao, S., and Zhou, Z. (2020). Structured3d: A large photo-realistic dataset for structured 3d modeling. In ECCV. Springer.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

INFORMATION REQUIREMENTS FOR DIGITAL TWIN INTEGRATION IN FIREFIGHTING: A SURVEY STUDY

Iryna Osadcha[1], Caroline McMullan[2], Andrius Jurelionis[1], Robert Howell[3], and Muhammad Intizar Ali[2]
[1]Kaunas University of Technology, Kaunas, Lithuania
[2]Dublin City University, Dublin, Ireland
[3]Dublin Fire Brigade, Dublin, Ireland

Abstract

Firefighting operations increasingly use digital tools for situational awareness, yet current systems often provide fragmented information that does not meet operational needs. Digital Twin integration appears to be a potential solution to address this gap. This study reports a survey of Irish firefighters examining information needs, data sources, Digital Twin functionality, and user readiness for adoption. The results show demand for real-time sensing and prediction, while operational decisions still rely mainly on experience and direct observation. The findings highlight the need for information prioritisation and federated rather than centralised data integration, contributing practical requirements for Digital Twin-enabled firefighting support systems.

Introduction

The evolution of the built environment is reshaping the nature of fire hazards, creating new challenges for firefighting operations. Beyond extinguishing the fire, first responders are often tasked with locating and evacuating occupants before conditions deteriorate due to heat and smoke (Khan et al., 2022). On arrival, the Incident Commander must quickly decide whether to take an offensive approach (entering the building to save occupants or property when the expected benefits outweigh risks to firefighter safety) or to remain outside in a defensive mode. This choice is never final and must be updated through dynamic risk assessment as the situation evolves (Han et al., 2010).

In practice, however, fire-ground conditions are often chaotic and unclear, making them difficult for first responders to interpret. This contributes to continued injuries and fatalities. The US Fire Administration reported 72 firefighter duty deaths in 2024, with 42 linked to stress or overexertion, and fatality numbers have fluctuated little over the past decade (ranging from 65 in 2019 to 141 in 2021) (USFA, 2024). In the UK, 677 on-duty injuries were recorded in 2024-2025, including 18 classified as major according to the RIDDOR specification (MHCLG, 2024). These figures underline the need for better situational awareness and more informed risk assessment during incidents.

Information systems have long been used in the fire safety domain and can be broadly grouped into: systems for fire safety equipment inspection and management (e.g., Almatared et al., 2023); systems designed to support occupant evacuation (e.g., Kim et al., 2024); and systems targeting firefighters directly. The latter include GIS-based systems (Tashakkori et al., 2015), sensor-driven and simulation-coupled systems (Han et al., 2010), BIM and IoT integration approaches (Rezaeifam et al., 2023), Digital Twin-based systems (Zhang et al., 2022), and immersive training environments (Chen et al., 2021). The standards NFPA 950 and 951 distinguish software used by firefighters into two main clusters: (1) incident command and situational awareness software and (2) analytical and decision support software. The standards also specify that software should be provided with an API or other standardised means of data exchange to enable integration with discrete systems (NFPA, 2020, 2022).

Research on firefighter information requirements to inform system development has led to several additional classifications. One of the earliest was a NIST workshop that identified approximately 100 information items needed during building emergencies, broadly divided into static (pre-incident building data) and dynamic (real-time sensor data) categories (Jones et al., 2005). Later, Guyo et al. (2023) distinguished between building and environmental data within each category. Tashakkori et al. (2015) grouped requirements into indoor, outdoor, and dynamic building information. Across these studies, the consistent finding is that fire response requires the integration of indoor and outdoor information from multiple sources, with information needs evolving across the phases of an incident and operational roles.

Emergency management activities are broadly categorised into four phases: preparedness, mitigation, response, and recovery (Larsson, 2008). Within fire response, Jones et al. (2005) identified three operational time periods: from dispatch to arrival, initial on-site assessment, and active mitigation. On receiving an alert, the fire team is provided with basic incident information (type, address, and dispatched units) and consults pre-incident plans en route to gain initial familiarity with the building layout and likely occupancy. On arrival, the first task is to size up the situation, covering life hazards and rescue requirements, fire accessibility, and the equipment needed. From that point, roles diverge. Frontline firefighters focus on search, rescue and fire attack, requiring information on occupancy, building hazards, and on-scene conditions. The Incident Commander, responsible for overall coordination, resource

deployment, and crew safety, additionally requires information on fire progression, resource status and tracking, personnel accountability, and inter-agency communication (Li et al., 2014; NIST, 2015).

To develop information systems for fire response operations, it is therefore necessary to identify what information matters most, when it is needed, and how it should be presented in relation to incident phases and operational roles. Existing research often relies on qualitative data gathered from interviews with firefighters, which limits the ability to prioritise information for system design (Guyo et al., 2023; Rezaeifam et al., 2023). This study addresses this limitation by conducting a structured survey that captures firefighters' perspectives on the types and priority of information, relevant data sources, Digital Twin scale, and required system features. It also collects information about users' digital skills and familiarity with relevant technologies as a proxy to assess the acceptable complexity of the future system.

The directly comparable prior work is the survey by Li et al. (2014), which examined firefighter information needs in the context of situational awareness. However, it was conducted more than a decade ago and was not oriented toward any specific technology. Similarly, the NIST workshop by Jones et al. (2005), which provided one of the first structured lists of firefighter information needs, predates the widespread adoption of digital tools in the fire service. The present survey revisits many of these information categories and frames them in relation to Digital Twin capabilities. It aims to define and prioritise the information requirements for integrating Digital Twins into fire response operations. The findings will support the specification of functional and nonfunctional system requirements and inform the design of the system architecture.

Methodology

The respondents were recruited through email invitations sent to the fire authorities in Ireland and the Dublin Fire Brigade. Eligible participants included individuals who had served as professional firefighters, regardless of their organisational role or rank. Data were collected over seven months, from May to December 2025. The survey received ethical approval from the Kaunas University of Technology Research Ethics Commission on 22 May 2025 (Approval No. M6-2025-15).

Survey design and structure

Some of the survey themes were informed by previous studies, including Jones et al. (2005), Li et al. (2014), Guyo et al. (2023), and Rezaeifam et al. (2023). Jones et al. (2005) documented the outcomes of a workshop with 25 experts, mostly at the strategic level, producing the lists of information items needed by firefighters. Li et al. (2014) conducted a USA-wide survey in 2012 with 197 responses, predominantly from command-level officers (69%), analysing information needs in the context of situational awareness. Guyo et al. (2023) conducted interviews with four experts with different operational roles from Finland, Germany, and Norway to validate an ontology of firefighter data requirements. Rezaeifam et al. (2023) conducted interviews with 10 command-level experts in Turkey to identify information requirements for situational awareness during fire response.

The survey was organised into four thematic sections:

1. Demographics (user profile) - to capture respondents' background and digital skills.

2. Information type - to assess the perceived importance of different types of information for firefighting operations. In this section, pre-arrival information items were derived from static building data categories in Jones et al. (2005), Li et al. (2014), and Guyo et al. (2023). On-scene information items were adapted from Li et al. (2014), Guyo et al. (2023) and Rezaeifam et al. (2023).

3. Information source - to identify which sources firefighters currently use or would like to have available during operations. In this section, source categories were derived from Li et al. (2014), Rezaeifam et al. (2023), Jones et al. (2005) and NIST (2015).

4. Digital Twin capabilities - to assess firefighters' expectations of Digital Twin deployment (level of support, scale, and preferred formats of information delivery). This section was informed by Li et al. (2014), Jones et al. (2005), NIST (2015) and NFPA 950.

The survey primarily used multiple-choice checklists (e.g., identifying each information source as "In Use", "Desired" or "Not Relevant") and 5-point Likert-type scales (e.g., rating importance from "Not desired" to "Most preferred"). Open-ended questions were included only in the demographic section; all other items were fixed responses. The fixed-response format was adopted for three reasons: to enable direct quantitative comparison with the closest prior study (Li et al., 2014); to produce quantifiable results that allow information items to be prioritised; and to minimise the cognitive and time burden on survey participants.

Data preparation and cleaning

Initial cleaning involved removing non-data fields (e.g., consent confirmations). As the survey did not require the responses to proceed to the next question, all records were reviewed for missing values. A small number of missing responses were found, primarily in the optional "Advanced digital skills" sub-section of demographics. These were excluded from subsequent statistical analysis.

Categorical labels were coded for quantitative analysis (e.g., "Not important" to "Critically important" coded as 1-5). Job titles were harmonised into broader role clusters: strategic level, command level, and frontline level, based on the organisational structure provided in the Fire & Emergency Operations Plan 2023-2028 (Dublin City Council, 2023), considering that all participants were Ireland-based.

Descriptive statistical analysis

For categorical variables (e.g., gender), frequency distribution was calculated to characterise the sample. For the job clusters, the average years of professional

experience were also calculated to illustrate differences between the organisational levels.

Given the small sample size (n=10), the analysis focused on descriptive measures rather than inferential statistics for job clusters. For all sections using Likert-type scales, measures of central tendency (mean and mode) and variability (standard deviation) were calculated for each item. For non-Likert items with three categorical options (e.g., "In Use", "Desired", "Not Relevant"), mode agreement (the proportion of respondents selecting the most frequent category, %) is reported.

Validation

Several quality control strategies were implemented to ensure the validity of the results. First, all questionable or incomplete entries identified during data cleaning were cross-checked against the raw survey responses. Second, quantitative findings (e.g., high importance ratings for specific information types) were compared with established domain knowledge to assess plausibility. In particular, the NIST workshop (Jones et al., 2005) and the NIST Research Roadmap for Smart Fire Fighting (NIST, 2015) were used, as they reflect information needs validated with fire service practitioners. NFPA 950 and 951 provided a normative basis, as both standards address data creation, utilisation, and exchange within the fire service. EuroFSA statistics were used to validate data on vulnerable occupants and high-risk buildings. Finally, the survey by Li et al. (2014) was used as the closest comparable study to determine where the findings align with or contradict earlier results.

Results Analysis and Discussion

Respondent demographics

The respondents were experienced first responders with between 10 and 39 years of professional experience (Table 1). Most of the respondents (70%) were male. The sample is dominated by respondents at strategic and command levels, with only one frontline respondent. However, when classified by organisational branch rather than rank, seven of the ten participants belong to the operations branch directly involved in firefighting (Dublin City Council, 2023).

Table 1: Participants by gender and experience

Job cluster	Gender	Avg. experience
Strategic level	Male = 2 Prefer not to say = 1	35 years
Command level	Female = 2 Male = 4	21 years
Frontline level	Male=1	10 years

Participants reported moderate self-assessed technical skills (mean=3.5, SD=0.8), showing the highest familiarity with data acquisition equipment and the lowest with programming languages (Table 2). All participants indicated at least moderate familiarity with digital technologies, reflected in regular everyday use (mean=3.7, SD=0.9), and working knowledge of Digital Twins (mean=3.5, SD=1.0). However, organisational readiness was perceived as limited, with respondents rating their organisations' digitalisation strategy and past experience with new technology adoption relatively low (mean=2.7, SD=1.0, and mean=1.9, SD=0.8, respectively).

Table 2: Participants' digital familiarity and skills

Statement	Mean*	SD	Mode
Regular digital usage	3.7	0.9	4
Digital Twin familiarity	3.5	1.0	4
Overall technical skills	3.5	0.8	3
Troubleshooting experience	3.1	1.0	3
Organisation's digital strategy	2.7	1.0	2
Organisation's adoption history	1.9	0.8	2
Data acquisition equipment	3.3	0.8	3
GIS	3.0	0.7	3
Modeling	2.9	0.9	3
Databases	3.0	1.1	2
Simulation	2.1	0.6	2
AI/ML	2.0	0.5	2
Programming languages	1.6	1.0	1

*Likert scale: 1=low to 5=high

Information type

Respondents rated information on building occupancy and the presence of hazardous materials as critically important before arriving on scene (mean=4.7, SD=0.5 and mean=4.6, SD=0.5, respectively) as shown in Table 3. Similarly, data on nearby water sources, the location and size of the fire based on sensor alarms, building type (e.g., residential, commercial), technical systems (e.g., HVAC), specialised sensors (e.g., chemical detectors), and evacuation elevators were generally considered "Very" to "Critically important". In contrast, peripheral data, such as information on important objects to be saved, neighbourhood maps and underground utilities, and building parking areas, were rated as less critical ("Slightly" to "Moderately important").

Table 3: Importance of information before arrival at the emergency site

Type of information	Mean*	SD	Mode
Occupancy	4.7	0.5	5
Hazardous materials	4.6	0.5	5
Water sources	4.4	0.8	5
Fire location (sensors)	4.4	0.5	4
Building type	4.3	0.6	4
Technical systems	4.2	0.6	4
Specialized sensors	4.2	0.6	4
Evacuation lifts	4.1	0.7	4
Building layout and site plan	4.0	0.6	4
Access routes	3.7	0.8	4
Owner contacts	3.6	0.7	4
Traffic data	3.4	0.8	4
Security system	3.3	0.9	3
Restricted zones	3.1	0.9	3
Neighbourhood and utilities map	2.9	0.7	3
Priority assets	2.9	0.7	3
Parking area	2.8	1.2	3

*Likert scale: 1=Not important to 5= Critically important

These results are consistent with the findings of previous studies. Li et al. (2014) also identified building occupancy, hazards, and water sources as key information before arriving at the emergency site. However, their

study highlighted traffic data (route information) and priority assets as highly important, while in the present survey, these factors were rated lower ("Moderately" to "Very important"). Similarly, Guyo et al. (2023) reported that hazard locations, parking, and restricted areas within buildings were important to firefighters, although in this survey, such elements were viewed as only "Moderately" important. The importance of building type is further supported by the European Fire Safety Action Plan 2020-2023 (EuroFSA, 2020), which identified priority premises for fire safety improvements as collective housing (37%), high-rise buildings (18%), buildings accommodating less self-reliant individuals (17%), individual housing (13%) and buildings where large groups of people gather (10%).

Once on scene, the respondents identified the location of the occupants, trapped individuals, and warnings of structural collapse as critically important information (mean=5.0, SD=0.0 for both), as shown in Table 4. This information was rated as maximally important across all job clusters. Information on fire growth dynamics (location, size, duration), the number of vulnerable persons (elderly, children and persons with special needs) and the current building condition (e.g., unsafe entry, structural instability) was also rated among the most critical. Overall, nearly 70% of the information items in this section were rated between "Very" and "Critically" important.

Peripheral data, such as the location and condition of deployed or standby units, hazard perimeter, contacts of other emergency agencies, occupancy of nearby buildings, and local weather conditions and forecasts, were rated slightly lower, ranging from "Moderately" to "Very" important, with weather-related information receiving the lowest average score (mean = 3.1, SD = 0.7). Some differences were observed in communication-related items, with strategic level respondents rating contact information for other emergency agencies higher (mean=4.00) than command level (mean=3.50).

Table 4: Importance of information at the emergency site

Type of information	Mean*	SD	Mode
Occupant locations	5.0	0	5
Collapse warnings	5.0	0	5
Vulnerable persons	4.8	0.4	5
Fire dynamics	4.8	0.4	5
Building condition	4.7	0.6	5
Smoke status	4.5	0.5	5
Access openings	4.4	0.7	5
Refuge and assembly areas	4.3	0.6	4
Sprinkler status	4.1	0.7	4
Sprinkler coverage	4.1	0.7	4
Ventilation status	4.1	0.7	4
Hazard zone	3.8	0.6	4
Standing-by responding units	3.8	0.9	4
Other emergency agencies	3.7	0.6	4
Nearby occupancy	3.5	0.7	3
Weather data	3.1	0.7	3

*Likert scale: 1=Not important to 5= Critically important

Previous surveys have reported similar findings. Li et al. (2014) identified occupancy information, fire dynamics (location, growth rate, and spread of the fire), and smoke status as key informational needs. Likewise, Guyo et al. (2023) emphasised the importance of factors such as the building year of construction (as an indicator of structural resilience and risk of collapse) and operational smoke ventilation systems. They also noted the relevance of the occupancy of nearby buildings, which in this survey was rated as "Moderately" to "Very" important. The importance of information on vulnerable occupants is further supported by data from the EuroFSA (2020), which shows that more than half of residential fire fatalities occur among people 65 or older. EuroFSA (2020) also identified key vulnerable groups that need targeted fire safety improvements: elderly people (38%), children (36%), adults with reduced independence (6%), and socially disadvantaged populations (6%).

Information source

For predefined and location-based information sources (Table 5), all respondents reported using and having access to emergency preplans (100% "In use"). There was also strong interest in expanding access to more advanced predefined data, with 90% of respondents expressing a desire for 3D building models and 80% indicating a need for improved on-scene positioning and navigation systems. The occupancy schedules were reported as "In use" by 60% of the respondents. Geoinformation systems, digital maps, and open data sources available via the Internet were each cited by 50% as currently in use.

Table 5: Predefined and location-based information sources

Source of information	Mode*	Mode agreement (%)
Building 3D models	3	90
Positioning and navigation	3	80
GIS	2	50
Occupancy schedules	2	60
Open data	2	50
Emergency plans	2	100

*Multiple-choice: 1=Not Relevant, 2=In Use, 3=Desired

A previous survey by Li et al. (2014) similarly highlighted the widespread use of emergency preplans (approximately 95% of respondents). However, the present study indicates a higher use of occupancy schedules (60% compared to 25% in the survey by Li et al.(2014)). The difference may be explained by the increased digitalisation of building records and wider use of occupancy management systems, with such data now more routinely available to fire services.

The demand for 3D building models and improved positioning and navigation also remains strong, comparable to Li et al. (2014), where 80% and 77% of respondents, respectively, expressed such needs. These findings are further supported by interviews conducted by Chou et al. (2019), which emphasised firefighters' demand for improved navigation and path planning capabilities during operations.

For building systems specifications as an information source (Table 6), the level of agreement on the data currently "In use" was slightly lower. The fire protection and water supply systems were each identified as "In use" by 70% of the respondents, while the electrical system

was reported by 50%. Most of the information sources in this category were classified as "Desired" (i.e., systems respondents would like to have available). These included building automation and control systems, security and access control systems (each 70%), as well as HVAC, lighting, and elevator systems (each 60%).

Table 6: Building systems specifications

Source of information	Mode*	Mode agreement (%)
Building automation and control	3	70
Security and access control	3	70
HVAC	3	60
Lighting	3	60
Elevators	3	60
Electrical	2	50
Water supply	2	70
Fire protection	2	70

*Multiple-choice: 1=Not Relevant, 2=In Use, 3=Desired

These results are consistent with previous studies. Li et al. (2014) similarly reported high usage of fire protection systems and a strong desire for access to HVAC, elevator, security, and lighting systems, with response levels of approximately 50%. Comparable findings were also reported in interviews conducted by Rezaeifam et al. (2023), which highlighted the frequent use of fire protection, water supply and electrical systems in firefighting operations.

For the real-time data sources (Table 7), all items were classified as "Desired" by the respondents, generally showing high levels of agreement. Occupancy and movement tracking were identified as the most desired data sources, with 100% agreement among the respondents. Air quality sensors, water pressure and sprinkler activation data, CCTV and thermal imaging camera feeds, and electrical system monitoring were closely followed with 90% agreement. Real-time data from temperature sensors, HVAC system (e.g., airflow, smoke control), lighting control system, and security system alerts (e.g., door locks, access changes) were also highly desired, each with 80% agreement. Specialised sensor data (e.g., chemical or radiation detection) and fire detection and alarm system data were rated slightly lower, with 80% and 50% agreement, respectively.

Table 7: Real-time data sources

Source of information	Mode*	Mode agreement (%)
Occupancy	3	100
Electrical	3	90
Water pressure and sprinklers	3	90
CCTV	3	90
Lighting	3	80
Air quality	3	90
HVAC	3	80
Temperature	3	80
Security	3	80
Specialized sensors	3	80
Fire alarm	3	50

*Multiple-choice: 1=Not Relevant, 2=In Use, 3=Desired

Strong interest in using real-time data for fire emergency operations is reinforced by domain research, including the Research Roadmap for Smart Fire Fighting (NIST, 2015) and the earlier Workshop to Define Information Needed by Emergency Responders During Building Emergencies (Jones et al., 2005). Both highlight that integrating real-time building data can significantly improve preparedness and situational awareness during incidents. In particular, data from HVAC systems, fire alarms, sprinklers, and occupant sensors were envisioned to be integrated into Smart Firefighting systems.

The cognitive and experiential information sources (Table 8) were identified as highly used by the respondents, indicating strong reliance on personal judgment and experience during operations. On-site observations showed the highest mode agreement (90%), followed by memory and experience from previous incidents (80%), highlighting their central role in decision-making on scene. However, 60% of the respondents identified the need for improvement in live communications and reporting during operations. Memory and experience from past incidents were reported as "In Use" across all clusters (strategic-67%, command-83%, frontline-100%). However, given the small sample and its heavy weighting toward strategic and command roles, these results cannot be generalised.

Table 8: Cognitive and experiential information sources

Source of information	Mode*	Mode agreement (%)
Live communications and reports	3	60
Memory and experience	2	80
On-site observations	2	90

*Multiple-choice: 1=Not Relevant, 2=In Use, 3=Desired

A strong reliance on prior experience and on-site observation aligns with the findings of Li et al. (2014), where 82% of respondents cited memory and experience, and 72% cited observations as their primary information sources. The survey also revealed that 81% of the respondents wanted improvements in live communication and reporting during incidents.

Digital Twin capabilities

Digital Twin functionality in this section can be divided into two clusters: general information management and support for operational activities (Table 9). Among information management functions, respondents identified information collection ("accessing and gathering all required information, including real-time data") as the area where Digital Twins could most effectively support daily operations (mean=4.5, SD=0.5). The prediction of information ("anticipating potential trends, such as fire spread or structural collapse based on real-time data") was rated as the second most important function (mean=4.5, SD=0.7). Among functions supporting operational activities, communication and coordination ("ensuring communication among firefighters, the command centre, and other emergency services") received the highest rating (mean=4.4,

SD=0.7), consistent with the previously expressed need to improve live communications and reporting.

Information integration ("combining data from multiple sources into a unified operational picture") was the fourth most desired function (mean=4.3, SD=0.6). Other key functionalities such as information interpretation ("extracting useful insights from available data"), resource allocation ("optimising the deployment of personnel, equipment, and support services"), information revision ("detecting and correcting data inconsistencies and errors"), post-incident analysis ("recording and analysing emergency response actions to improve future preparedness"), and information representation ("presenting and visualising data in an accessible format"), were all identified as areas where "Significant support" is required. Navigation and evacuation assistance ("providing dynamic routes and safe passage recommendations") was the only function rated slightly lower (mean=3.7, SD=1.0), indicating a perceived need for moderate rather than critical support.

Table 9: Required type of support expected from Digital Twin deployment

Type of support	Mean*	SD	Mode
Information collection	4.6	0.5	5
Information prediction	4.5	0.7	5
Information interpretation	4.3	1.0	5
Information integration	4.3	0.6	4
Information revision	4.2	0.9	4
Information representation	4.1	0.8	5
Communication and coordination	4.4	0.7	5
Resource allocation	4.2	0.9	4
Post-incident analysis	4.1	0.9	4
Navigation and evacuation assistance	3.7	1.0	4

*Likert scale: 1=No support needed to 5= Essential support required

Similarly, the study by Li et al. (2014) identified information collection (87%), prediction (72%), and reasoning (64%) as the most needed types of assistance where firefighters sought support. These preferences were also found to correlate with participants' years of experience and job position.

Regarding the operational scale at which a Digital Twin should function to be most useful to firefighters (Table 10), respondents identified the building level as the most important (mean=4.5, SD=0.7). This was followed by the floor level of the building (mean=4.4, SD=0.7) and the emergency site level (mean=4.2, SD=0.4). The building component level received the lowest rating (mean=3.9, SD=0.7), although it was still considered "Moderately" to "Very important" by the respondents.

Table 10: Digital Twin operating scale

Scale	Mean*	SD	Mode
Building	4.5	0.7	5
Building floor	4.4	0.7	5
Emergency site	4.2	0.4	4
Urban	4.0	0.8	4
Building component	3.9	0.7	4

*Likert scale: 1=Not important to 5= Critically important

Regarding the preferred scale, Li et al. (2014) survey found that the building floor level was most desired (75%), followed by the room level and building level. The component level ranked lowest (27%), which is consistent with the results of our survey. The urban scale was not included in the study by Li et al. (2014).

When evaluating preferred formats for information representation (Table 11), respondents expressed a strong preference for 3D models (mean=4.5, SD=0.8). Other formats, such as 2D graphical representations (e.g., drawings, schematics), augmented reality overlays, colour-coded indicators (e.g., heat maps or highlighted danger zones), and virtual simulations were also rated between "Strongly desired" and "Most preferred."

In contrast, numerical data (e.g., tabular or sensor readings), textual information (e.g., reports or instructions), and haptic feedback (e.g., vibration alerts) were viewed as "Slightly" to "Moderately" desirable. Audio and phonetic cues (e.g., spoken instructions or sound-based alerts) were the least preferred format (mean=2.6, SD=0.8).

Table 11: Preferred formats of information representation

Format	Mean*	SD	Mode
3D models	4.5	0.8	5
2D graphics	4.2	0.7	4
AR overlay	4.1	0.5	4
Virtual simulation	4.0	0.6	4
Colour-coded indicators	4.0	0.6	4
Numeric data	3.2	0.7	4
Text data	3.1	1.3	4
Haptic feedback	3.0	1.1	4
Audio cues	2.6	0.8	3

*Likert scale: 1=Not desired to 5= Most preferred

These findings are also consistent with the earlier research by Li et al. (2014), where graphical representation was identified as the most preferred format (64%), followed by 3D visualisation (50%), virtual reality (38%), and augmented reality (33%). Phonetic representation ranked lowest among the options. The NIST workshop similarly identified a preference for 3D building display as easier to use and interpret (Jones et al., 2005).

The survey did not explicitly address the preferred level of detail of 3D models. However, respondents' preferences for building-, floor-, and emergency-site-level information suggest that geometric resolution should reflect operational need rather than technical capability. In practice, photogrammetry-based 3D models are already used by firefighters for high-risk premises as part of Predetermined Attendance strategies (O'Neill, 2025). Regarding BIM, while it is the most commonly proposed source of 3D building representation in the fire safety literature, the majority of existing buildings predate it and have no such model available.

In summary, survey results show that firefighters consistently rely heavily on available emergency preplans and cognitive sources, such as personal observations and prior experience, during operations. Critical information includes occupancy data (including vulnerable occupants), building hazards and structural integrity, fire

dynamics, and the location of water sources, all of which are considered highly important both from dispatch to arrival and during fire attack. Regarding information visualisation, 3D models and 2D graphic representations of buildings and floor layouts are the most preferred, indicating a strong demand for better spatial awareness. The respondents also expressed a high need for improved general information collection, trend forecasting, and communication support during operations. Real-time data utilisation is highly demanded, particularly in terms of occupancy, water systems, and air quality. However, the survey showed moderate self-rated digital skills and low confidence in organisational digital strategy and prior experience with new technologies. This suggests potential challenges in deploying highly complex digital systems.

Implications and Limitations

The survey results help identify functional and non-functional requirements for system architecture development. A practical next step is to prioritise information needs based on response mode agreement (high, medium, low) and mapping each category to existing or required data sources. For example, occupancy information is already used based on occupancy schedules, while real-time occupancy tracking would require additional sensor data.

Interest in real-time sensing is high, yet many buildings lack the necessary infrastructure. This points out that two scenarios should be considered: (1) buildings without sensors or without real-time data access, where firefighters rely on preincident plans, 3D models, and GIS; and (2) buildings with sensor networks whose data is accessible to firefighters. Another tension appears between the strong demand for real-time information and the low preference for numerical or textual formats. Respondents lean toward visual representations (e.g., 3D models, 2D graphics), suggesting the need for dashboards and data abstractions.

Although building- and floor-level information was rated most useful, the preferred granularity of data remains unclear. Higher resolution data does not necessarily lead to better decisions. Component-level information may overload users and was less preferred, yet structural collapse risk was marked as highly important on the scene.

The reliance on several information sources suggests that federated integration is more realistic than centralisation. Data may stay within native systems where suitable, but should be interoperable through shared schemas and interfaces. In addition to the functional aspects, several non-functional requirements also emerge. Alongside common considerations such as performance, usability, scalability, and security, semantic and technical interoperability will be particularly critical. Firefighting systems must work under uncertainty, shifting resource availability, and dynamic coordination, closer to distributed defence systems than standard enterprise software.

The respondents reported moderate digital skills but low organisational readiness, indicating likely adoption challenges. Incremental deployment is therefore more feasible: new functionalities should extend current workflows and practices rather than replace them. Solutions that require heavy onboarding or major workflow changes may struggle to gain acceptance (Carlson et al., 2025).

This study has several limitations. (1) The survey sample was small and limited to first responders in Ireland, which may reduce the generalisability of the findings to other national contexts. (2) The small sample size and its skew toward strategic and command levels prevented meaningful comparisons across job clusters, and reduced the representativeness of operational firefighter information needs. (3) The survey did not account for building type, treating the operational process as uniform. Information needs may differ for high-risk facilities, although such incidents are considerably less frequent than typical dwelling fires (Dublin City Council, 2024). (4) Preferences for information presentation were assessed through text-based questions without interface mockups, which may have influenced responses; the preferred level of detail and granularity of data representations were also not explored. However, user interface design falls outside the scope of this study. (5) Finally, advanced Digital Twin capabilities such as LLM-based interaction or AI-driven prediction were not explicitly addressed. While functions such as information prediction and interpretation indirectly assume data-driven approaches, these were not explored in depth.

Conclusions

This study examined firefighters' information needs, data usage and expectations for Digital Twin capabilities in fire response. The findings confirm a persistent gap between the available digital technologies and their practical use during incidents. Under conditions of uncertainty and time pressure, decision-making continues to rely primarily on experience and direct observation, despite the recognised value of data-driven support.

The results identify occupancy and structural integrity as the most critical information categories, aligning with the core operational priorities: protecting occupants and ensuring crew safety. While some of this information can be obtained from pre-incident data, its real-time availability and prioritised delivery become critical during active operations. A Digital Twin theoretically can support these needs by linking data collection, analytics, and representation, but its practical implementation requires careful phasing. Not all information should be streamed continuously; prioritizations and timing of data will determine usability in real situations.

The main contribution of this study lies in providing quantitative evidence on firefighter information requirements, enabling explicit prioritisation of information types for system design. In addition, the study delivers a quantitative evaluation of Digital Twin functionality, which provides a basis for defining functional system requirements and aligning system

development with actual operational demand. The findings also update and extend earlier surveys, which are now partially outdated due to the increasing digitalisation of buildings and emergency response systems.

Acknowledgments

The authors express their sincere appreciation to all anonymous participants who took the time to complete the survey. We also thank the Dublin Fire Brigade for their consultation and assistance with the dissemination of the survey. We also thank Bentley Systems Inc. for supporting the collaboration between Kaunas University of Technology and Dublin City University.

References

Almatared, M., Liu, H., Abudayyeh, O., Hakim, O., Sulaiman, M., 2023. Digital-Twin-Based Fire Safety Management Framework for Smart Buildings. Buildings 14. https://doi.org/10.3390/buildings14010004

Carlson, E.J., Manierre, M., Bazzocchi, M.C.F., 2025. "Firefighters Hate Two Things - Change and the Way Things Are" Exploring Firefighters' Perspectives Towards Change. Fire 8, 348. https://doi.org/10.3390/fire8090348

Chen, H., Hou, L., Zhang, G. (Kevin), Moon, S., 2021. Development of BIM, IoT and AR/VR technologies for fire safety and upskilling. Automation in Construction 125, 103631. https://doi.org/10.1016/j.autcon.2021.103631

Chou, J.-S., Cheng, M.-Y., Hsieh, Y.-M., Yang, I.-T., Hsu, H.-T., 2019. Optimal path planning in real time for dynamic building fire rescue operations using wireless sensors and visual guidance. Automation in Construction 99, 1-17. https://doi.org/10.1016/j.autcon.2018.11.020

Dublin City Council, 2023. Dublin Fire Brigade Fire & Emergency Operations Plan 2023-2028. Version 20.02. Dublin: Dublin City Council.

Dublin City Council, 2024. Dublin Fire Brigade annual report 2023. Dublin: Dublin City Council.

European Fire Safety Alliance, 2020. European Fire Safety Action Plan is an Ongoing Success.

Guyo, E.D., Hartmann, T., Snyders, S., 2023. An ontology to represent firefighters data requirements during building fire emergencies. Advanced Engineering Informatics 56, 101992. https://doi.org/10.1016/j.aei.2023.101992

Han, L., Potter, S., Beckett, G., Pringle, G., Welch, S., Koo, S.-H., Wickler, G., Usmani, A., Torero, J.L., Tate, A., 2010. FireGrid: An e-infrastructure for next-generation emergency response support. Journal of Parallel and Distributed Computing 70, 1128-1141. https://doi.org/10.1016/j.jpdc.2010.06.005

Jones, W.W., Davis, W.D., Evans, D.D., Holmberg, D.G., Bushby, S.T., Reed, K.A., 2005. Workshop to Define Information Needed by Emergency Responders During Building Emergencies.

Khan, A.A., Khan, M.A., Leung, K., Huang, X., Luo, M., Usmani, A., 2022. A review of critical fire event library for buildings and safety framework for smart firefighting. International Journal of Disaster Risk Reduction 83, 103412. https://doi.org/10.1016/j.ijdrr.2022.103412

Kim, Y.-J., Kim, H., Ha, B., Kim, W.-T., 2024. Advanced fire emergency management based on potential fire risk assessment with informative digital twins. Automation in Construction 167, 105722. https://doi.org/10.1016/j.autcon.2024.105722

Larsson, A., 2008. A framework for evaluating emergency preparedness plans and response strategies.

Li, N., Yang, Z., Ghahramani, A., Becerik-Gerber, B., Soibelman, L., 2014. Situational awareness for supporting building fire emergency response: Information needs, information sources, and implementation requirements. Fire Safety Journal 63, 17-28. https://doi.org/10.1016/j.firesaf.2013.11.010

National Fire Protection Association (NFPA), 2020. NFPA 950: Standard for Data Development and Exchange for the Fire Service.

National Fire Protection Association (NFPA), 2022. NFPA 951: Guide to Building and Utilizing Digital Information.

National Institute of Standards and Technology (NIST), 2015. Research Roadmap for Smart Fire Fighting.

O'Neill, S., 2025. Dublin Fire Brigade blazes a trail for digital twins in emergency response. Bentley Cities & Campuses.

Rezaeifam, S., Ergen, E., Günaydın, H.M., 2023. Fire emergency response systems information requirements' data model for situational awareness of responders: A goal-directed task analysis. Journal of Building Engineering 63, 105531. https://doi.org/10.1016/j.jobe.2022.105531

Tashakkori, H., Rajabifard, A., Kalantari, M., 2015. A new 3D indoor/outdoor spatial model for indoor emergency response facilitation. Building and Environment 89, 170–182. https://doi.org/10.1016/j.buildenv.2015.02.036

UK Ministry of Housing, Communities and Local Government, 2024. FIRE0508: Injuries Sustained by Firefighters and Firefighter Fatalities.

U.S. Fire Administration (USFA), 2024. Firefighter Fatalities in the United States.

Zhang, T., Wang, Z., Zeng, Y., Wu, X., Huang, X., Xiao, F., 2022. Building Artificial-Intelligence Digital Fire (AID-Fire) system: A real-scale demonstration. Journal of Building Engineering 62, 105363. https://doi.org/10.1016/j.jobe.2022.105363

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

TOWARDS A SEMANTIC ONTOLOGY FOR OCCUPANT-CENTRIC BUILDING CONTROL

Saif Alramahi[1], Mohammad Mayouf[1], Elham Del Zendeh[1], and Khalid Ismail[2]
[1]Department of Architecture and the Built Environment, Birmingham City University, 4 Cardigan Street, Birmingham B4 7BD, United Kingdom
[2]Department of Computer Science, Birmingham City University, 4 Cardigan Street, Birmingham B4 7RQ, United Kingdom

Abstract

Occupant-centric control (OCC) balances energy efficiency and occupant comfort in buildings, yet methodological fragmentation hinders cumulative progress. This paper presents the Occupant-Centric Control Ontology (OCCO), a semantic framework that extends W3C/ETSI vocabularies with constructs for comfort metrics, control strategies, feedback provenance, and performance evaluation. Developed using the Linked Open Terms methodology with competency questions derived from OCC literature, OCCO unifies subjective occupant feedback with objective sensor data. Structural verification via SPARQL confirms query capability, with automated evaluation returning no remaining critical pitfalls and 88.2% FAIR compliance. OCCO provides a semantic foundation for comparable, reproducible OCC evaluation.

Introduction

Occupant-centric control (OCC) embeds real-time user information into building control logic to jointly optimise comfort and energy efficiency (Nagy *et al.*, 2023; Yuan *et al.*, 2024). Despite growing research activity, the field lacks standardised evaluation metrics, and findings remain difficult to compare across implementations (Nagy *et al.*, 2023). Studies employ inconsistent comfort metrics, varied energy baselines, and frequently exclude occupants from evaluation altogether (Shafavi *et al.*, 2020). The consequence is that effective strategies cannot be distinguished from ineffective ones, evidence cannot accumulate into generalisable knowledge, and real-world deployment of adaptive building controls remains limited to isolated case studies.

A key barrier to addressing this fragmentation is the absence of a shared semantic foundation for describing and evaluating OCC systems. Existing ontologies address adjacent concerns (sensing, building devices, occupant feedback) but lack the integration of quantitative and qualitative evaluation metrics within a unified framework. This paper addresses this gap by presenting the Occupant-Centric Control Ontology (OCCO). OCCO extends W3C and ETSI vocabularies with integrated OCC evaluation constructs for comfort-energy assessment, control strategy representation, and feedback provenance, enabling comparable evaluation of how well OCC implementations balance measured environmental conditions with reported occupant experience. The framework is demonstrated through visual comfort in commercial lighting, a domain where the gap between objective standards compliance and subjective occupant satisfaction is particularly apparent.

Background and related work

While OCC lighting has attracted considerable research attention (Haq *et al.*, 2014; Nagy, Yong and Schlueter, 2016; Park *et al.*, 2019; Colaco *et al.*, 2023), the field still faces persistent challenges that hinder cumulative progress (Shafavi *et al.*, 2020). Energy and comfort outcomes receive unequal treatment. For instance, daylight harvesting implementations often measure power consumption rigorously but rely only on Workplane illuminance as a comfort proxy (Yu *et al.*, 2017; Lee, Song and Cho, 2018). Conversely, studies emphasising comfort through glare indices such as Daylight Glare Probability (DGP) or Unified Glare Rating (UGR) sometimes lack well-defined energy baselines for comparison (Motamed, Deschamps and Scartezzini, 2017; Papinutto *et al.*, 2022). Many implementations exclude users from evaluation entirely, relying exclusively on sensor data (Yin and Keoh, 2016; Kumar *et al.*, 2017; Lee *et al.*, 2017), despite evidence that override rates reach 88% for automatic blinds (Reinhart and Voss, 2003) and 78% permanent system disabling due to dissatisfaction (Meerbeek *et al.*, 2014). Energy baselines vary from manual control (Aghemo, Blaso and Pellegrino, 2014), to fixed illuminance setpoints (Kandasamy *et al.*, 2018), or go completely unspecified (Kumar *et al.*, 2017). A reported "40% energy savings" carries different meaning from study to study. Visual comfort metrics range from EN 12464-1:2021 (CEN, 2021) or ISO/CIE 8995-1:2025 (ISO/CIE, 2025) illuminance thresholds, to DGP, UGR, and custom indices (Gunay *et al.*, 2017; Park *et al.*, 2019), with a lack of consensus equivalent to the PMV/PPD models established for thermal comfort (ISO, 2025).

Addressing this gap requires a shared conceptual foundation for describing OCC systems and their outcomes. Semantic ontologies offer a potential solution by providing machine-readable vocabularies with formal relationships enabling consistent data description and cross-system integration. However, existing approaches address only parts of the problem. SOSA/SSN (Janowicz *et al.*, 2019) provides measurement provenance patterns

but lacks comfort semantics. SAREF4BLDG (Poveda-Villalón and García-Castro, 2018) offers device and spatial vocabulary but omits occupant modelling. OFO (Donkers, de Vries and Yang, 2024) captures subjective feedback but excludes control strategies and performance evaluation. Brick (Balaji *et al.*, 2016) provides building metadata without comfort semantics, while the DNAs framework (Hong *et al.*, 2015) formalised occupant behaviour for simulation rather than operational control. Recent holistic performance ontologies (Mahdavi, Taheri and Wolosiuk, 2019; Chung *et al.*, 2023) focus on design-phase assessment or thermal comfort without constructs for integrated OCC evaluation that unify objective measurements with subjective occupant responses. The most closely related effort is WOMO (Bruttini *et al.*, 2022), a linked-data ontology that integrates objective sensor observations with subjective occupant feedback for workplace management. However, WOMO targets facility management decision-making rather than OCC evaluation and does not model control strategies, energy performance, or standards compliance. OCCO addresses these gaps by providing constructs that unify comfort-energy assessment with feedback provenance within a single evaluation framework.

Based on the literature, there are limited semantic efforts that integrate both objective sensor data and subjective occupant feedback for OCC evaluation. OCCO proposes to address this gap through a standards-aligned framework that extends SOSA/SSN, SAREF4BLDG, and OFO with constructs for comfort-energy assessment, control strategy representation, and feedback provenance, demonstrated here through the visual comfort domain.

Methodology

OCCO was developed following the Linked Open Terms (LOT) methodology, an industrial-oriented ontology engineering framework that promotes FAIR (Findable, Accessible, Interoperable, Reusable) principles (Poveda-Villalón *et al.*, 2022). LOT was selected over traditional methodologies such as METHONTOLOGY and NeOn for three reasons: (1) its explicit emphasis on reusing standardised vocabularies rather than creating redundant terms; (2) its iterative, sprint-based development model suited to evolving requirements; and (3) its alignment with Linked Data publication best practices.

Ontology requirements specification

Ontology requirements were formalised using the Ontology Requirements Specification Document (ORSD) template from the LOT methodology (Poveda-Villalón *et al.*, 2022). Competency questions (CQs) help convey ontology requirements, core elements, and their relationships, thereby defining the ontology vocabulary. These questions are used to test requirements satisfiability through axiom checking and answer verification (Ren *et al.*, 2014), and should be evaluated using different use-cases. Ten competency questions (Table 1) were derived from the gaps identified in Section 2. These questions address control automation, preference modelling, performance evaluation, and cross-domain integration of comfort with energy metrics.

Table 1: Competency questions for the developed ontology

ID	Competency Question
CQ-01	What control actions were taken in response to detected occupancy in each building space during working hours?
CQ-02	Which building spaces received visual discomfort feedback, and what were the corresponding glare or illuminance levels?
CQ-03	Which control strategy resulted in the highest combined comfort scores and lowest energy consumption over the evaluation period?
CQ-04	What are the preferred illuminance and CCT levels for occupants performing specific visual tasks (e.g., reading, screen work)?
CQ-05	How does an occupant's preference profile evolve over time in response to feedback events?
CQ-06	Which building spaces have circadian-friendly lighting control strategies applied, based on occupant chronotype?
CQ-07	Which building spaces meet the applicable comfort standard while also receiving negative comfort feedback from occupants?
CQ-08	What proportion of illuminance observations fall within the range recommended by the applicable comfort standard?
CQ-09	Which occupants most frequently override automated controls, and in which building spaces do overrides occur?
CQ-10	Which control strategy achieves the best balance of energy efficiency, occupant satisfaction, and compliance with comfort standards?

Ontology implementation

The LOT methodology structures implementation through four sub-activities: conceptualisation, ontology reuse, encoding, and evaluation (Poveda-Villalón *et al.*, 2022). Conceptualisation transformed the ORSD requirements into a structured ontology model. Domain knowledge was organised into five thematic modules (detailed in Section 4). Visual representation followed the Chowlk notation (Chávez-Feria, García-Castro and Poveda-Villalón, 2022), a UML-based specification for representing classes, properties, and restrictions in ontology diagrams.

Ontology reuse adopted a standards-first approach, prioritising W3C Recommendations and ETSI Technical Specifications over domain-specific vocabularies. Term extraction drew from the competency questions and OCC literature, identifying core concepts (e.g., illuminance, glare, occupant feedback, control action) and their relationships. Candidate ontologies were evaluated against selection criteria including standardisation status, maintenance activity, and semantic coverage. Section 4 details the reused ontologies and their integration rationale.

Encoding translated the conceptual diagrams into OWL 2 DL. Diagrams were exported as XML files and uploaded to the Chowlk web converter, which parses the visual notation and generates OWL code serialised in both Turtle and RDF/XML formats (Chávez-Feria, García-Castro and Poveda-Villalón, 2021). The Turtle file was then imported into Protégé (Musen, 2015), an open-source ontology editor, for manual refinement including axiom adjustments, import statement configuration, and consistency checking with the HermiT reasoner.

Evaluation comprises two distinct phases. Verification assesses structural correctness, confirming the ontology is free from syntactic, modelling, and semantic errors and can answer the specified competency questions. Validation assesses domain accuracy, confirming the ontology captures real-world practice and meets end-user requirements. This phase, comprising domain and ontology expert review, is planned as future work and falls outside the scope of this paper.

Verification process

Synthetic instance data was generated iteratively using a large language model (Claude Sonnet 4.6, Anthropic). For each competency question, the model was prompted with the ontology schema in Turtle format and the specific CQ requirements, producing tailored RDF instances for that test scenario. This approach aligns with emerging research exploring LLM assistance in ontology engineering tasks (Lippolis *et al.*, 2025). Generated instances represented realistic OCC scenarios, including occupancy observations, illuminance measurements, user feedback events, control actuations, and performance outcomes. All instances were manually reviewed for logical and domain correctness.

Each of the ten competency questions was translated into one or more SPARQL 1.1 queries and executed using rdflib 7.6.0 (Python) against the synthetic ABox instances loaded alongside the OCCO TBox. Protégé (Musen, 2015) and the HermiT reasoner were used separately for OWL 2 DL consistency checking. Queries tested core capabilities including temporal filtering, spatial correlation, aggregation, and provenance tracking. Structural quality was additionally assessed using OOPS! (Poveda-Villalón, Gómez-Pérez and Suárez-Figueroa, 2014) and FAIR compliance evaluated using FOOPS! (Garijo, Corcho and Poveda-Villalon, 2021). Detailed results are presented in Section 5.

OCCO framework

OCCO is designed around two guiding principles. First, standards alignment: rather than defining custom vocabulary, OCCO reuses established W3C and ETSI ontologies wherever possible, extending them only where domain-specific concepts are required. This positions the ontology for integration with systems already using these standards. Second, modularity: the ontology is organised into cohesive, loosely coupled modules that can evolve independently as the field matures. The following subsections present the architecture, reused standards, key classes, and competency question coverage.

Architecture overview

OCCO organises its classes into five thematic modules, each addressing a distinct aspect of OCC, demonstrated here through the visual comfort domain. This modular structure supports selective reuse and facilitates alignment with existing domain ontologies. Table 2 summarises the modules, their representative classes, and the standards from which they derive.

The **Physical Context** module describes the built environment: zones, rooms, lamps, and sensing/actuation devices. Classes are drawn primarily from SAREF4BLDG for building topology and SOSA for sensor/actuator abstractions. The **Occupants & Preferences** module captures individual characteristics relevant to visual comfort, including task type (e.g., screen-based work, paper reading) and chronotype (circadian preference). These classes extend the Occupant Feedback Ontology (OFO). The **Feedback & Interaction** module represents both objective sensor observations and subjective occupant feedback, including manual override events. This dual representation directly addresses the exclusion of occupants from evaluation noted in Section 2. The **Control & Automation** module defines control strategies (e.g., daylight harvesting, occupancy-based dimming) and the actuations they produce. Provenance relations from PROV-O link actions to their triggering observations. The **Evaluation & Outcomes** module enables performance assessment against standards such as EN 12464-1:2021. It captures both comfort outcomes (e.g., satisfaction ratings, glare reports) and energy metrics, supporting the balanced comfort-energy evaluation identified as a gap in Section 2.

Table 2: OCCO modular architecture

Module	Representative Classes	Primary Standards
Physical Context	BuildingSpace, Lamp, Sensor, Actuator	SAREF4BLDG, SOSA
Occupants & Preferences	Person, VisualTask, Chronotype, Preference	OFO, OCCO
Feedback & Interaction	Feedback, OverrideEvent, Observation	OFO, SOSA
Control & Automation	ControlStrategy, Controller, Actuation	PROV-O, SOSA
Evaluation & Outcomes	PerformanceOutcome, Standard, Illuminance, GlareIndex	OCCO, QUDT, dcterms

Ontology reuse strategy

OCCO adopts a standards-first reuse strategy, prioritising W3C Recommendations and ETSI Technical Specifications, supplemented by established domain vocabularies where domain-specific coverage is required. The following ontologies are reused:

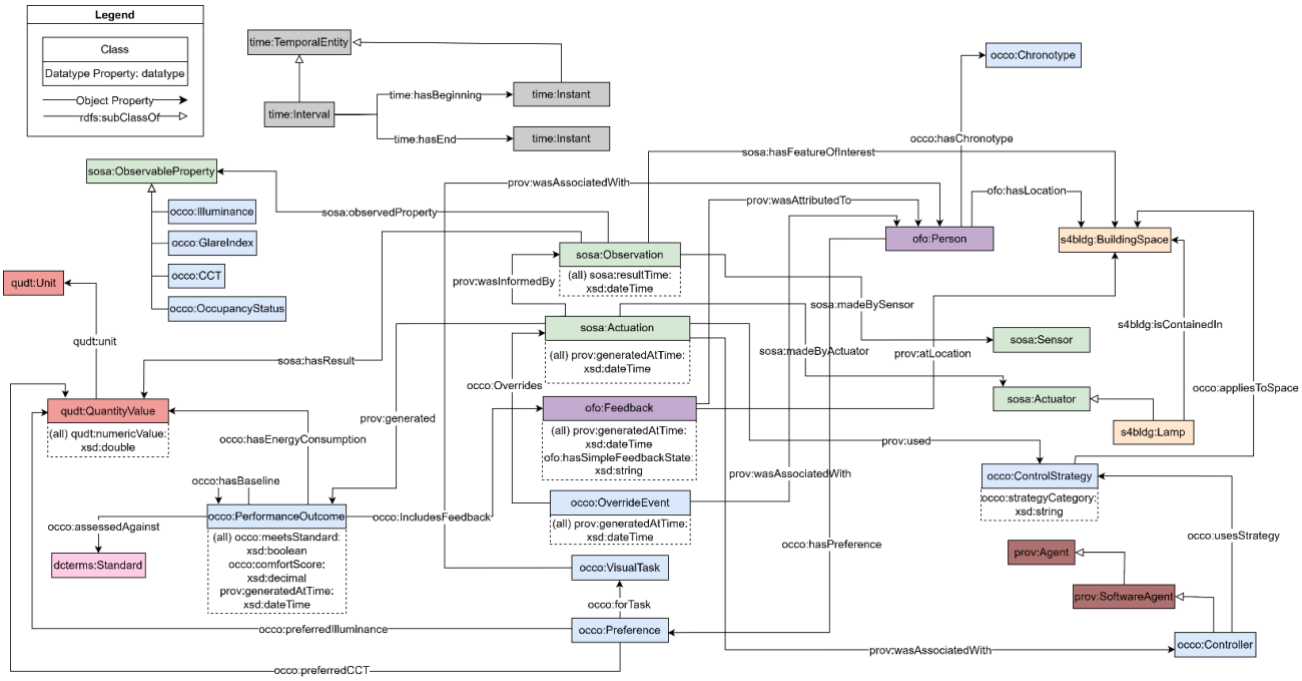


Figure 1: OCCO ontology classes, object properties, and data properties in Chowlk notation

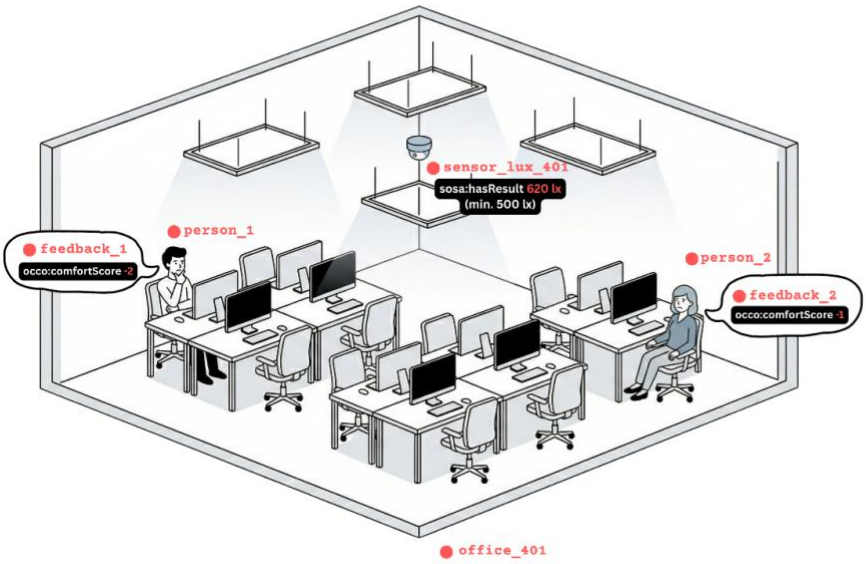
- **SOSA/SSN** (Janowicz *et al.*, 2019): a semantic sensor network ontology for describing sensors, observations, actuators, and actuations. Used in OCCO to model light sensors, occupancy detectors, and dimming actions.
- **SAREF4BLDG** (Poveda-Villalón and García-Castro, 2018): a smart appliances extension for building devices and spaces. Used to represent rooms, zones, and lamps within building topology.
- **OFO** (Donkers, de Vries and Yang, 2024): the Occupant Feedback Ontology for capturing subjective comfort reports and preferences. Used to model occupant feedback, override events, and individual lighting preferences.
- **PROV-O** (Lebo, Sahoo and McGuinness, 2013)**:** a provenance ontology for describing entities, activities, and agents. Used to trace which observations triggered which control actions and attribute feedback to occupants.
- **QUDT** (Hodgson, 2025): a vocabulary for quantities, units, and dimensions. Used to ensure measurement comparability (lux, kWh, UGR, DGP values) across datasets.
- **OWL-Time** (Cox and Little, 2022): a temporal ontology for instants and intervals. Used to represent evaluation periods, schedules, and timestamps for observations and actions.
- **dcterms** (DCMI, 2020): Dublin Core metadata terms. Used to reference lighting and comfort standards (EN 12464-1:2021, ISO/CIE 8995-1:2025) with structured metadata.

Table 3 summarises the namespace prefixes used throughout OCCO. Beyond these reused terms, novel classes in OCCO address concepts absent from existing standards: visual comfort metrics (Illuminance, GlareIndex, CCT), OCC strategies, and performance outcomes integrating subjective feedback with objective measurements.

Table 3: Namespace prefixes and IRIs used in OCCO

Ontology	Prefix	IRI
Occupant-Centric Control Ontology	occo	*https://w3id.org/occo#
Sensor, Observation, Sample, Actuator	sosa	http://www.w3.org/ns/sosa/
SAREF for Buildings	s4bldg	https://saref.etsi.org/saref4bldg/
Occupant Feedback Ontology	ofo	https://w3id.org/ofo#
W3C Time Ontology	time	http://www.w3.org/2006/time#
W3C Provenance Ontology	prov	http://www.w3.org/ns/prov#
Quantities, Units, Dimensions, Types	qudt	http://qudt.org/schema/qudt/
Dublin Core Metadata Terms	dcterms	http://purl.org/dc/terms/

*Full documentation and OWL serialisation are available at w3id.org/occo



Class	Instance
s4bldg:BuildingSpace	office_401
occo:Illuminance	illuminance_401
sosa:Sensor	sensor_lux_401
sosa:Observation	obs_lux_1 sosa:observedProperty illuminance_401; sosa:hasResult 620 lx
ofo:Person	person_1
ofo:Person	person_2
ofo:Feedback	feedback_1 occo:comfortScore -2; prov:atLocation office_401
ofo:Feedback	feedback_2 occo:comfortScore -1; prov:atLocation office_401
occo:Performance Outcome	outcome_401 occo:meetsStandard true (620 lx ≥ 500 lx); occo:assessedAgainst occo:EN_12464_1; occo:includesFeedback feedback_1, feedback_2

Figure 2: Representative OCCO ABox illustrating the scenario addressed by CQ-07

Key classes and relationships

Figure 1 illustrates OCCO's class hierarchy and principal object properties in Chowlk notation (Chávez-Feria, García-Castro and Poveda-Villalón, 2022). Rather than listing all elements, this section highlights relations that demonstrate the ontology's core capability: linking occupant feedback to control actions and enabling unified performance evaluation.

Feedback-Control linkage. When an occupant reports discomfort, the `ofo:Feedback` instance is linked to the occupant via `prov:wasAttributedTo` and to the affected space via `prov:atLocation`. Control actions (`sosa:Actuation`) reference triggering observations through `prov:wasInformedBy`, enabling queries that trace which sensor readings prompted which responses. Manual overrides (`occo:OverrideEvent`) link to the actuation they supersede via `occo:overrides`, preserving the full decision trail.

Performance evaluation. The `occo:PerformanceOutcome` class aggregates both comfort and energy results for a given evaluation period. It connects to subjective data via `occo:includesFeedback` and to objective measurements via `sosa:hasResult`. Compliance with lighting standards is captured through `occo:assessedAgainst`, linking outcomes to `dcterms:Standard` instances representing EN 12464-1:2021 or ISO/CIE 8995-1:2025 standards for example.

Demonstration and verification

Verification was conducted across three complementary methods as outlined in Section 3.3. All evaluation artefacts, including SPARQL queries, result sets, OOPS! and FOOPS! results, are published alongside the ontology documentation at w3id.org/occo.

To illustrate OCCO in operation, Figure 2 presents an ABox excerpt addressing CQ-07: *"Which building spaces meet the applicable comfort standard while also receiving negative comfort* feedback *from occupants?"* The scenario depicts an open-plan office (`office_401`) where a ceiling-mounted lux sensor (`sensor_lux_401`) records an illuminance of 620 lx – compliant with the EN 12464-1:2021 minimum of 500 lx. Two occupants submit `ofo:Feedback` records spatially anchored to the space via `prov:atLocation`, reporting comfort scores of −2 and −1. The resulting `occo:PerformanceOutcome` aggregates both feedback records and asserts `meetsStandard = true` against EN 12464-1:2021. Although the space satisfies the illuminance standard, occupants remain dissatisfied, a discrepancy that quantitative compliance metrics alone would not reveal.

All ten CQs (Table 1) were executed as SPARQL 1.1 queries using rdflib 7.6.0 against the synthetic ABox, returning between 1 and 6 results per query and confirming complete structural coverage. Boundary conditions were verified explicitly across multiple CQs: out-of-hours actuations were correctly excluded (CQ-01), the occupancy-actuation linkage via `prov:wasInformedBy` was confirmed (CQ-01), standards-compliant periods that still generated negative comfort scores were isolated (CQ-07), and zero-override occupants were correctly absent from grouped results (CQ-09).

Automated structural quality evaluation using OOPS! (Poveda-Villalón, Gómez-Pérez and Suárez-Figueroa, 2014) identified one critical pitfall in v0.1.0 (multiple

`rdfs:domain` declarations on `prov:wasAssociatedWith`, resolved in v0.1.1) alongside four minor pitfalls, all addressed prior to publication. FAIRness compliance assessed using FOOPS! (Garijo, Corcho and Poveda-Villalon, 2021), achieved a score of 88.2% (19 of 22 checks passed); remaining failures relate to LOV registry registration and DOI assignment pending acceptance.

Discussion

The ontology proposed in this paper formalises subjective occupant feedback alongside objective sensor measurements, offers constructs for integrated comfort-energy assessment, and provides a shared vocabulary for consistent and reproducible comparisons across OCC implementations. Where existing ontologies (Poveda-Villalón and García-Castro, 2018; Janowicz *et al.*, 2019; Donkers, de Vries and Yang, 2024) each address only part of this landscape, OCCO uniquely combines all four dimensions: visual comfort metrics (illuminance, UGR, DGP, CCT), control strategies, feedback provenance, and performance evaluation within a single, standards-aligned framework.

Whilst former research (Donkers, de Vries and Yang, 2024) provide practical application and value for OCC systems, the holistic integration of objective and subjective metrics presented in the proposed ontology offers wider value for various stakeholders. For instance, researchers gain a shared vocabulary enabling comparable evaluations and meta-analyses across studies. System vendors can implement transparent control logic with standards-aligned interfaces to existing building management systems and digital twins. Facility managers can query comfort-energy trade-offs to support evidence-based and holistic post-occupancy decisions.

Beyond immediate stakeholder applications, OCCO proposes to address a fundamental methodological gap in visual comfort research. Thermal comfort research benefits from established evaluation foundations such as PMV/PPD models codified in ISO 7730 (ISO, 2025), enabling robust cross-study comparison and cumulative knowledge building. Visual comfort research currently lacks an equivalent foundation. By providing standardised constructs for comfort metrics, control strategies, and performance outcomes, OCCO offers a comparable semantic foundation for the lighting domain. The ontology's ability to correlate standards compliance with subjective measures (as demonstrated in CQ-07) enables systematic investigation of cases where normative thresholds fail to predict occupant satisfaction, a recognised but under-studied phenomenon. Furthermore, the explicit representation of override events supports analysis of user intervention patterns, addressing the high override rates documented in the literature. These capabilities position OCCO not merely as a data integration tool, but as an enabler of the holistic, comparable evaluations that the field currently lacks.

Conclusion

This paper presented OCCO, a standards-aligned semantic ontology that integrates visual comfort metrics, control strategies, feedback provenance, and performance evaluation within a unified framework. By extending W3C and ETSI vocabularies alongside established domain ontologies, OCCO provides interoperability with existing building automation systems while offering a semantic foundation for comparable, reproducible OCC evaluation. Structural verification confirmed that all ten competency questions can be answered through SPARQL queries, with automated quality evaluation returning no remaining critical pitfalls (OOPS!) and 88.2% FAIR compliance (FOOPS!).

Future work will address operational challenges within smart buildings and validate the ontology through domain and ontology expert review, as well as applied use cases involving Internet-of-Things (IoT) systems and Digital Twins. The ontology is published as Linked Data with LOT-compliant documentation at the permanent URI w3id.org/occo, supporting community adoption and extension.

References

Aghemo, C., Blaso, L. and Pellegrino, A. (2014) 'Building automation and control systems: A case study to evaluate the energy and environmental performances of a lighting control system in offices', Automation in Construction, 43, pp. 10–22.

Balaji, B. et al. (2016) 'Brick: Towards a unified metadata schema for buildings', in Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments. New York: Association for Computing Machinery, pp. 41–50.

Bruttini, A. et al. (2022) 'Linked data approach for occupant-centric workplace management', in Proceedings of the 22nd International Conference on Construction Applications of Virtual Reality.

CEN (2021) Light and lighting: Lighting of work places. Part 1: Indoor work places. Brussels: CEN.

Chávez-Feria, S., García-Castro, R. and Poveda-Villalón, M. (2021) 'Converting UML-based ontology conceptualizations to OWL with Chowlk', in Verborgh, R. et al. (eds) The Semantic Web: ESWC 2021 Satellite Events. Cham: Springer International Publishing, pp. 44–48.

Chávez-Feria, S., García-Castro, R. and Poveda-Villalón, M. (2022) 'Chowlk: From UML-based ontology conceptualizations to OWL', in Groth, P. et al. (eds) The Semantic Web. Cham: Springer International Publishing, pp. 338–352.

Chung, J. et al. (2023) 'Towards an indoor environmental quality management ontology', in 1st International Workshop on Semantic Web on Constrained Things, SWoCoT 2023, pp. 16–26. Available at: https://ceur-ws.org/Vol-3412/paper2.pdf (Accessed: 15 January 2026).

Colaco, S.G. et al. (2023) ‘A state-of-the-art artificial intelligent techniques in daylighting controller: Models and performance’, Science and Technology for Energy Transition, 78, p. 37.

Cox, S. and Little, C. (2022) Time Ontology in OWL. Available at: https://www.w3.org/TR/owl-time/ (Accessed: 7 January 2026).

DCMI (2020) DCMI Metadata Terms. Available at: https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ (Accessed: 7 January 2026).

Donkers, A., de Vries, B. and Yang, D. (2024) ‘Creating occupant-centered digital twins using the Occupant Feedback Ontology implemented in a smartwatch app’, Semantic Web, 15(2), pp. 259–284.

Garijo, D., Corcho, O. and Poveda-Villalon, M. (2021) ‘FOOPS!: An ontology pitfall scanner for the FAIR principles’, International Semantic Web Conference (ISWC) 2021: Posters, Demos, and Industry Tracks. Available at: http://ceur-ws.org/Vol-2980/paper321.pdf.

Gunay, H.B. et al. (2017) ‘Development and implementation of an adaptive lighting and blinds control algorithm’, Building and Environment, 113, pp. 185–199.

Haq, M.A. ul et al. (2014) ‘A review on lighting control technologies in commercial buildings, their performance and affecting factors’, Renewable and Sustainable Energy Reviews, 33, pp. 268–279.

Hodgson, R. (2025) Quantities, Units, Dimensions and Types (QUDT) Schema. Available at: https://www.qudt.org/doc/DOC_SCHEMA-QUDT.html (Accessed: 7 January 2026).

Hong, T. et al. (2015) ‘An ontology to represent energy-related occupant behavior in buildings. Part I: Introduction to the DNAs framework’, Building and Environment, 92, pp. 764–777.

ISO (2025) ISO 7730:2025: Ergonomics of the thermal environment: Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. Geneva: International Organization for Standardization. Available at: https://www.iso.org/standard/85803.html (Accessed: 14 January 2026).

ISO/CIE (2025) ISO/CIE 8995-1:2025 Light and lighting: Lighting of work places. Part 1: Indoor. Geneva: ISO.

Janowicz, K. et al. (2019) ‘SOSA: A lightweight ontology for sensors, observations, samples, and actuators’, Journal of Web Semantics, 56, pp. 1–10.

Kandasamy, N.K. et al. (2018) ‘Smart lighting system using ANN-IMC for personalized lighting control and daylight harvesting’, Building and Environment, 139, pp. 170–180.

Kumar, A. et al. (2017) ‘Implementation of smart LED lighting and efficient data management system for buildings’, Energy Procedia, 143, pp. 173–178.

Lebo, T., Sahoo, S. and McGuinness, D. (2013) PROV-O: The PROV Ontology. Available at: https://www.w3.org/TR/prov-o/ (Accessed: 7 January 2026).

Lee, W. et al. (2017) ‘A light control technique for energy-efficient buildings in visual comfort constrained environment’, in 2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech), pp. 1–5. Available at: https://ieeexplore.ieee.org/document/8019244 (Accessed: 5 March 2025).

Lee, W., Song, K. and Cho, H. (2018) ‘Visual comfort considered light control methods for energy efficient office buildings: A case study’, Thermal Science, 22(Suppl. 3), pp. 875–885.

Lippolis, A.S. et al. (2025) ‘Large language models assisting ontology evaluation’, arXiv.

Mahdavi, A., Taheri, M. and Wolosiuk, D. (2019) ‘A Building Performance Indicator Ontology: Structure and applications’, in Building Simulation 2019. Rome, Italy, pp. 77–82.

Meerbeek, B. et al. (2014) ‘Impact of blinds usage on energy consumption: Automatic versus manual control’, in Aarts, E. et al. (eds) Ambient Intelligence. Cham: Springer International Publishing, pp. 158–173.

Motamed, A., Deschamps, L. and Scartezzini, J.-L. (2017) ‘On-site monitoring and subjective comfort assessment of a sun shadings and electric lighting controller based on novel High Dynamic Range vision sensors’, Energy and Buildings, 149, pp. 58–72.

Musen, M.A. (2015) ‘The Protégé Project: A look back and a look forward’, AI Matters, 1(4), pp. 4–12.

Nagy, Z. et al. (2023) ‘Ten questions concerning occupant-centric control and operations’, Building and Environment, 242, p. 110518.

Nagy, Z., Yong, F.Y. and Schlueter, A. (2016) ‘Occupant centered lighting control: A user study on balancing comfort, acceptance, and energy consumption’, Energy and Buildings, 126, pp. 310–322.

Papinutto, M. et al. (2022) ‘Saving energy by maximising daylight and minimising the impact on occupants: An automatic lighting system approach’, Energy and Buildings, 268, p. 112176.

Park, J.Y. et al. (2019) ‘LightLearn: An adaptive and occupant centered controller for lighting based on reinforcement learning’, Building and Environment, 147, pp. 397–414.

Poveda-Villalón, M. et al. (2022) ‘LOT: An industrial oriented ontology engineering framework’, Engineering Applications of Artificial Intelligence, 111, p. 104755.

Poveda-Villalón, M. and García-Castro, R. (2018) 'Extending the SAREF ontology for building devices and topology', in Proceedings of the 6th Linked Data in Architecture and Construction Workshop (LDAC 2018), pp. 16–23.

Poveda-Villalón, M., Gómez-Pérez, A. and Suárez-Figueroa, M.C. (2014) 'OOPS! (OntOlogy Pitfall Scanner!): An on-line tool for ontology evaluation', International Journal on Semantic Web and Information Systems, 10(2), pp. 7–34.

Reinhart, C. and Voss, K. (2003) 'Monitoring manual control of electric lighting and blinds', Lighting Research & Technology, 35(3), pp. 243–258.

Ren, Y. et al. (2014) 'Towards competency question-driven ontology authoring', in Presutti, V. et al. (eds) The Semantic Web: Trends and Challenges. Cham: Springer International Publishing, pp. 752–767.

Shafavi, N.S. et al. (2020) 'Occupants visual comfort assessments: A review of field studies and lab experiments', Solar Energy, 208, pp. 249–274.

Yin, X. and Keoh, S.L. (2016) 'Personalized ambience: An integration of learning model and intelligent lighting control', in 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT). Reston, VA: IEEE, pp. 666–671.

Yu, T.-H. et al. (2017) 'An RTP-based dimming control system for visual comfort enhancement and energy optimization', Energy and Buildings, 144, pp. 433–444.

Yuan, Y. et al. (2024) 'A review of current research on occupant-centric control for improving comfort and energy efficiency', Building Simulation, 17(10), pp. 1675–1692.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

ONTOLOGY-BASED SEMANTIC MANAGEMENT OF BRIDGE DRAWINGS

Xiaoli Song[1], Mengyan Peng[1], Steffen Marx[1], and Chongjie Kang[1]
[1]TUD Dresden University of Technology, Dresden, Germany

Abstract

As one of the primary information carriers throughout the bridge lifecycle, drawings contain rich semantic information. Traditional file-based management does not support efficient organization and retrieval. In this work, an ontology-based semantic management method for bridge drawings is proposed, leveraging information extracted by artificial intelligence (AI) models. After processing, AI outputs are transformed into structured semantic entities and relationships defined in the ontology. The resulting knowledge graph enables semantic search and helps engineers trace how structural components appear and evolve across drawings. It enhances the accessibility and reusability of bridge drawings and provides a practical foundation for intelligent asset management.

Introduction

Throughout the lifecycle of bridge engineering projects, a large number of heterogeneous files are generated (Kang et al., 2025). As a key part of this technical documentation, bridge drawings support all major stages, including design, construction, and maintenance. They carry essential engineering knowledge and form the basis for design decisions, construction guidance, and maintenance management.

As bridge projects continue to grow in scale and increase in complexity, a single project can generate a massive volume of drawings (Yang et al., 2022). However, traditional file-system-based drawing management treats each drawing as an independent file and organizes them through folder hierarchies and naming conventions. Although this approach is simple and intuitive, it has clear limitations when applied to large, heterogeneous drawing collections. Drawings from different project stages are sometimes stored in separate locations, even when they depict the same component at different points in time. As a result, engineers cannot efficiently track a component across views and stages, and the drawing content remains fragmented, lacking a systematic and explicit representation.

In recent years, AI has made significant progress in extracting information from engineering drawings (Moreno-García et al., 2018). AI models can identify component outlines, text, and symbols, providing a solid foundation for drawing digitization. However, most current approaches process individual drawings in isolation and cannot structure data across multiple project phases in a systematic manner. This limitation is particularly relevant in bridge engineering, where drawings are continuously produced throughout design, construction, and maintenance stages. Enabling cross-phase component traceability can therefore improve drawing reuse, inspection documentation, and lifecycle-oriented infrastructure management.

An ontology-based semantic management approach for bridge drawings can help address this gap and support the digitalization of bridge engineering. It enables semantic, content-based retrieval without relying on file names or folder paths, thereby improving accessibility. Explicitly modeling relationships enables the system to track the appearance of components across drawings from different project stages, thereby enhancing traceability for inspection and maintenance. Moreover, it can benefit Building Information Modeling (BIM) by supporting the reconstruction of BIM models from drawings, and by enabling model-to-drawing trace-back and collaborative viewing, thereby increasing the reuse value of drawing files.

Overall, this paper introduces a semantic pipeline that organizes bridge drawing information in a unified way. It further provides a new way of storing and organizing drawing content from the perspective of views. A case study is conducted to demonstrate how this pipeline can enhance the retrieval and traceability of drawing information. This work lays a foundation for ontology-based digital management and reuse of bridge engineering file assets.

Related Work

AI-based information extraction from engineering drawings

To support the digitalization of engineering drawings, various information extraction methods have been proposed, which can broadly be divided into traditional image processing approaches and deep learning–based methods. Early studies focused on text extraction and symbol recognition using techniques such as connected component analysis, Hough transforms, and rule-based databases (Fletcher and Kasturi, 2002; Or et al., 2005). These methods were later applied to interpret 2D drawings for 3D model generation, for example, through line detection and connectivity rules (Lewis and Séquin, 1998). In such approaches, geometric primitives extracted from 2D drawings are mapped to 3D elements based on predefined reconstruction rules, which often require strong assumptions about drawing con-

ventions and completeness.

With the advent of deep learning, automatic feature learning has significantly improved the robustness and scalability of drawing interpretation. Recent studies have applied convolutional neural networks and related models for object detection, symbol classification, scene text detection, and OCR in technical drawings (Elyan et al., 2018; Vilgertshofer et al., 2020; Nguyen et al., 2021). These techniques enable the extraction of both geometric features and high-level semantic cues from 2D drawings, which are essential for bridging the gap between 2D representations and 3D models. Consequently, AI-based extraction is increasingly deployed to facilitate downstream applications, such as BIM and IFC-based bridge modeling. In these workflows, detected components, annotations, and spatial relationships serve as critical inputs for semi-automatic or rule-based 3D reconstruction (Lu et al., 2020; Akanbi and Zhang, 2022; Feist et al., 2024).

However, most existing studies focus on extracting information within individual drawings, while the organization and semantic management of extracted results across multiple drawings remain insufficiently addressed.

Semantic Management of Engineering Information

To manage the massive, fragmented data generated by modern projects, the concept of semantic file systems provides a crucial theoretical foundation (Gifford et al., 1991). Rather than treating files merely as "items stored under a rigid path," subsequent semantic file systems incorporated ontologies to model files as explicit objects with semantic metadata, enabling concept-based associative retrieval and reasoning (Ngo et al., 2007; Mashwani and Khusro, 2018). This ontology-driven paradigm is essential for organizing heterogeneous files.

Product Lifecycle Management (PLM) and Product Data Management (PDM) systems are used in mechanical engineering to manage data across the entire lifecycle of standardized, mass-produced items (Ji and Abdoli, 2023). However, they cannot be directly applied to bridge engineering because construction projects are unique, site-specific, and non-standardized, whereas manufacturing typically involves standardized products (Zhang et al., 2024). In the Architecture, Engineering, and Construction (AEC) industry, ontologies are more extensively deployed as a unified semantic layer (Noy et al., 2001; Elshani et al., 2025). Specifically, they are utilized to enhance information integration (Wu et al., 2021) and achieve interoperability across different systems (Hagedorn et al., 2023; Lorvão Antunes et al., 2024). These are particularly beneficial for the growing adoption of infrastructure digital twins (DT) (Zinke et al., 2023; Ramonell et al., 2023).

However, these macro-level approaches often treat engineering drawings as "black-box" attachments. By focusing on physical assets and broad business processes, they overlook the rich geometric and spatial semantics embedded within the drawings. To address this, specialized ontologies have emerged. Early work explored semantic drawing classification (Lu et al., 2008), and recently, Schönfelder and König (2025) proposed an ontology to standardize AI-extracted visual information, successfully introducing a semantic layer for individual drawings.

While such drawing-specific ontologies enhance the queryability of standalone files, they still lack the relational mechanisms required for complex, multi-drawing projects. Specifically, existing research cannot differentiate view types or trace identical components across drawings. To bridge this gap, this paper proposes a semantic management ontology for bridge engineering drawings that supports unified view representation and cross-view component tracking in multi-phase projects.

Methodology

Semantic Processing Pipeline

As shown in Figure 1, the overall pipeline can be divided into three main parts.

First, a set of bridge drawings is processed by trained AI models to extract basic information. The models produce segmented subviews for each drawing and JSON files containing OCR text and other detected elements. These outputs serve as the input for subsequent semantic processing.

Second, a Bridge Engineering Drawing Ontology (BEDO) is constructed. It defines the classes and relationships needed to represent drawing semantics. Where possible, existing domain ontologies are reused to avoid redefining common concepts and to maintain compatibility with established standards. In parallel, a semantic encoding module processes the AI-generated outputs and organizes information such as view types, components, and scales into a better structured encoding JSON file. This file is then used to automatically populate the ABox of the ontology via scripts. On top of these assertions, a set of rules is applied to further enrich the semantics. Since AI models extract unstructured text without understanding project-specific organizational logic, users can define custom rules tailored to specific naming conventions and operational needs. Applying these rules on top of the initial assertions further enriches the semantics and explicitly expands the relational structure of the final knowledge graph.

Finally, in the query stage, the pipeline supports semantic retrieval and visual localization. Users can issue queries based on components, view types, scales, and their relationships, and the system retrieves related drawings and views from the drawing collection folder. For selected results, the corresponding subviews are highlighted directly on the original full drawing to support positioning. In this way, the pipeline turns semi-structured AI outputs into computable knowledge and visualizable query results.

AI-Based Extraction of Drawing Information

2D engineering drawings contain both geometric and non-geometric information, including graphical representations of structural views as well as textual annotations and title blocks. In this work, information extraction follows an

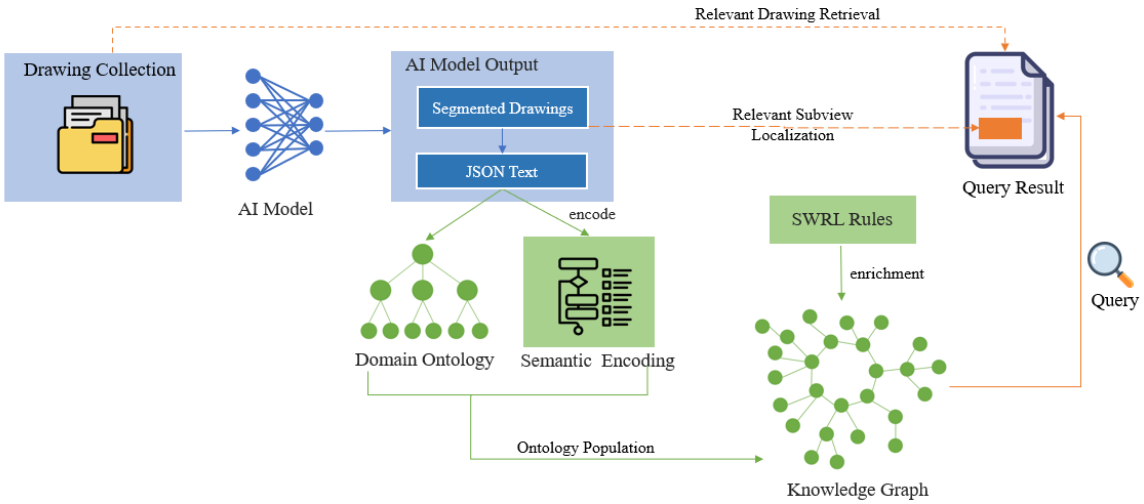


Figure 1: Overall Pipeline: from AI output to ontology-based query and visualization

integrated workflow that combines view localization, text recognition, and semantic post-processing, largely building upon the methodology proposed by Peng et al. (2025). This workflow enables the structured extraction of heterogeneous information from engineering drawings as a foundation for subsequent semantic modeling.

Graphical views in engineering drawings are localized using an AI-based object detection model. In this step, YOLOv7 is applied to identify view regions such as top and side views, providing bounding boxes that describe the spatial extent of each view. These localized view regions establish the geometric reference framework for organizing and associating extracted information. Textual information is obtained through OCR, which detects and recognizes text elements from annotations and title blocks. The OCR outputs include recognized text strings and precise bounding box coordinates, enabling spatial alignment between textual elements and graphical views within the drawings.

To transform the raw outputs of object detection and OCR into semantically meaningful information, post-processing is applied, including title-block metadata extraction, font-size-based view-title identification, and spatial matching between titles and detected views (Peng et al., 2025, 2026). In the current prototype, title-block metadata are extracted using an interactive semi-automatic workflow in which a reference region defined on the first sheet is transferred to subsequent drawings. The results are further filtered using clustering-based analysis and heuristic rules, while uncertain cases are retained for manual verification.

Furthermore, engineering drawings produced by different companies across various project phases exhibit heterogeneous layouts, terminology, varying scales, and geometric distortions from scanning. To handle these inconsistencies, the proposed workflow completely avoids fixed templates and absolute geometric dimensions. Instead, view localization relies on AI-based object detection using normalized bounding-box coordinates, which naturally mitigates the effects of geometric distortion. Similarly, text recognition and view-title identification are driven by relative font-size clustering and spatial relationships rather than predefined title-block positions or fixed pixel thresholds. Finally, explicitly stated scale information is extracted independently and encoded as a semantic attribute. This relative, feature-driven approach ensures that the extraction pipeline remains robust across highly diverse drawing collections.

However, the extracted outputs at this stage still lack a unified relational structure, necessitating further semantic encoding to align with the core ontology concepts.

Semantic Encoding

To transform the semi-structured AI outputs into ontology representations, a semantic encoding module is introduced. Based on the JSON files produced in the information extraction stage, the module identifies and normalizes elements such as view types, components, and scales. The result is an encoding JSON file that is ready for automatic ABox population in the subsequent step.

Figure 2 shows the encoding workflow. To enhance the robustness of the encoding method, the script implements three layers of matching. The matching conditions are progressively relaxed across these layers, so that exact matches are used whenever possible, while fuzzy matches can still be obtained when the available information is incomplete or noisy.

The first layer is exact matching, which uses direct string comparison with word-boundary checks and is suitable for standardized, error-free text. The second layer is substring matching, which relies on substring inclusion to handle compound words (e.g., German compound words). The third layer is fuzzy matching, which compensates for OCR and image segmentation errors, such as missing initial characters or spelling mistakes, by applying a similarity threshold (set to 90% by default).

These three layers are applied on top of the encoding keyword lists, ensuring that semantically relevant terms can still be detected even when the textual inputs deviate from

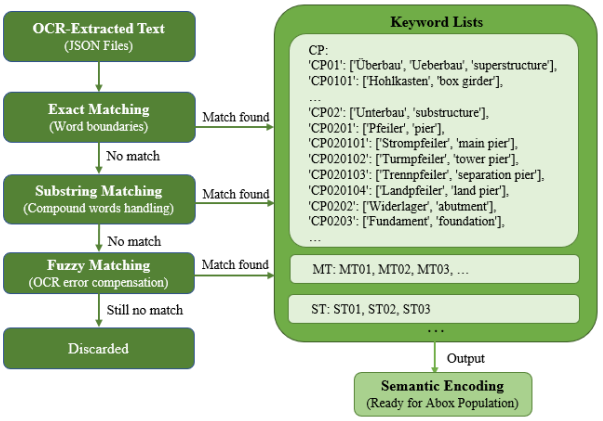


Figure 2: Semantic encoding workflow for OCR-extracted text

their canonical forms. Building on this, the script iterates through seven predefined keyword categories to generate semantic encodings for the extracted text: Bridge Type (BT), Component (CP), Document (DC), View (VW), Material (MT), Lifecycle Stage (ST), and Others (OT).

The keywords used in these categories were developed based on existing domain ontologies reused in BEDO, which are described in detail later in this paper. For each category code, the associated keyword lists include terms in multiple languages, such as German and English, allowing multilingual matching of textual outputs. Figure 2 shows an example encoding keyword list for the Component (CP) category. The design of the encoding keyword lists is extensible. By appending additional language variants to the keywords, the same encoding logic can be directly applied to drawings in new languages. Furthermore, new semantic categories can be added to the keyword lists as AI models become capable of extracting richer information from drawings. In this way, the semantic encoding can be incrementally extended while reusing the same overall pipeline.

Ontology Design

Figure 3 shows the overall structure of the proposed BEDO ontology. The ontology design consists of three main parts: classes, object properties, and data properties.

Classes

To describe engineering drawings, several core classes were designed. The `bedo:Drawing` class, which is a subclass of the `bedo:Document` class, represents technical drawings. `bedo:DrawingView` represents the corresponding views within a drawing. To ensure the defined ontology remains simple and practically applicable, established international standards DIN EN ISO 10209 (2022) and DIN EN ISO 128-3 (2022) were explicitly taken into account when defining the basic nomenclature of drawing views. However, because the definitions in these standards include many view types rarely encountered in the bridge domain, the view classes were selectively filtered and structured to incorporate only those practically utilized in bridge engineering. Additionally, the `bedo:DrawingSeries` class enables higher-level abstraction and organization of drawings when semantic information is more abundant.

The `bedo:Stage` class defines different phases of bridge engineering projects. Throughout a bridge's lifecycle, various types of drawings are produced at different stages, including design drawings, construction drawings, and maintenance drawings. These are represented by the `bedo:Design`, `bedo:Construction`, and `bedo:Maintenance` classes, respectively.

Object Properties

The object properties define the key relationships between classes. The `bedo:hasDrawingView` property links a drawing to its different views, as a single drawing can

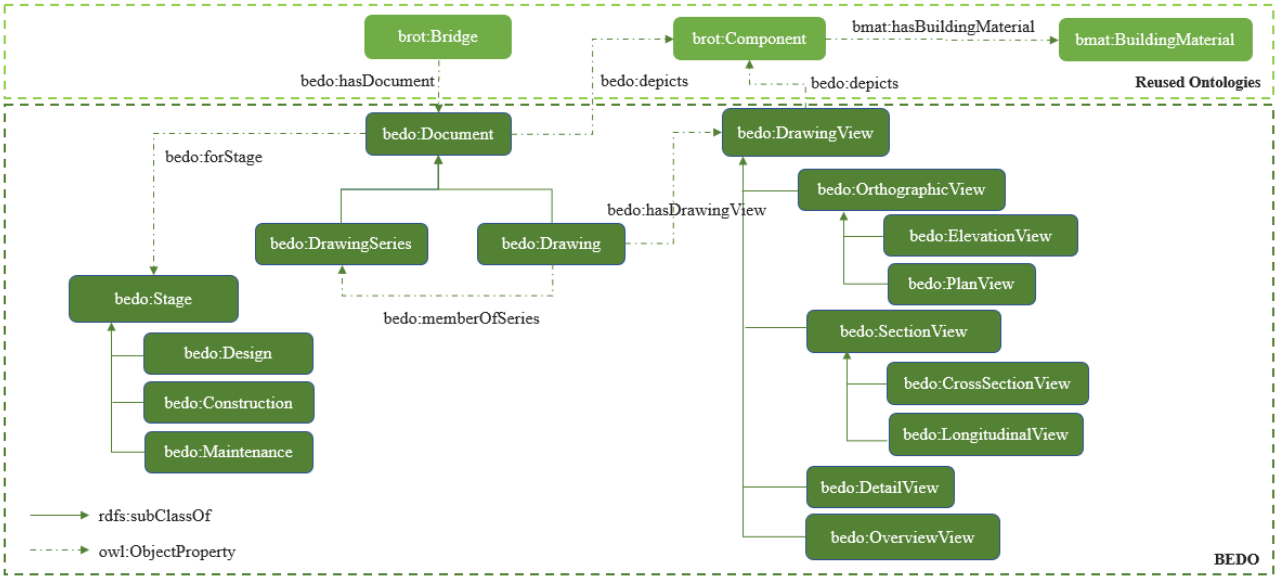


Figure 3: Bridge Engineering Drawing Ontology

contain multiple views. The `bedo:forStage` property describes which engineering stage a bridge document belongs to. The `bedo:memberOfSeries` property indicates that a drawing belongs to a specific drawing series. A single drawing can be included in multiple series based on different classification criteria. The corresponding inverse properties are also defined but are omitted here for brevity.

Data Properties

The data properties primarily capture attribute information from drawing files and processed model outputs. For drawing files, `bedo:drawingFilePath` specifies the file path location within the system, while `bedo:drawingTitle` stores the title from the drawing title block. For subviews generated after AI processing, `bedo:boundingBox` records the bounding box parameters used during AI-based object detection, which can be utilized for reverse view localization during queries. The `bedo:drawingViewTitle` property captures the titles of subviews and thus helps determine the semantic information of each view. Additionally, `bedo:viewFilePath` stores the file path of AI-generated view files. The `bedo:viewJsonFilePath` property records the file path of the OCR output JSON files. Other data properties include `bedo:scale` for scale information and `bedo:seriesType` for classifying drawing series types.

Reusing Existing Bridge Domain Ontologies

To improve interoperability and reusability, some existing bridge domain ontologies were reused. These include the Bridge Topology Ontology (brot), the Bridge Component Ontology (brcomp), and the Bridge Material Ontology (bmat) (Hamdan and Scherer, 2020). While BEDO models drawing- and view-level semantics, it relies on these external ontologies to represent the physical assets depicted in the drawings. Extracted components and materials are mapped directly to classes in `brcomp` and `bmat`, respectively, while topological relationships among components are aligned with `brot`. The reused ontologies and their IRIs are summarized in Table 1.

Table 1: Reused Ontologies in BEDO

Prefix	URI
brot	https://w3id.org/brot#
bmat	https://w3id.org/bmat#
brcomp	https://w3id.org/brcomp#

Application and Results

A case study was conducted using a set of inspection drawings of the Nibelungen Bridge. The Nibelungen Bridge is a historic prestressed concrete bridge in Worms, Germany, constructed in 1953 (Kang et al., 2024).

Ontology Population

The drawing dataset consists of 22 drawing files and was first processed using the model proposed by Peng et al. (2025). Note that the drawings used in this case study are not included in the original training dataset of that model. After semantic encoding and ABox population, a knowledge graph representing this drawing set was generated.

Rule-based Semantic Enrichment

Before being used for downstream tasks, the knowledge graph was further semantically enriched by applying a set of rules. As introduced in the methodology, these rules capture project-specific organizational logic. As shown in Figure 4, the inferred series instances, such as *Drawingseries_003* and *Drawingseries_Hessen*, are automatically linked to their corresponding member drawings, forming a more structured semantic network. Two illustrative examples are as follows:

- Rule 1 (Series by filename): Any drawing whose filename starts with ”003.” is assigned to the drawing series *Drawingseries_003*, where the *seriesType* is set to ”filename”.
- Rule 2 (Series by region): Any drawing whose drawing title contains the word ”Hessen” is assigned to the drawing series *Drawingseries_Hessen*, where the *seriesType* is set to ”region”.

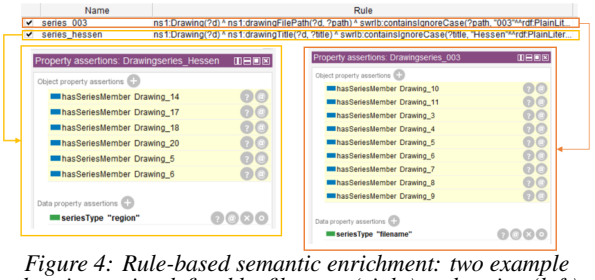


Figure 4: Rule-based semantic enrichment: two example drawing series defined by filename (right) and region (left)

Semantic Retrieval and Visualization

With the knowledge graph, users can perform semantic queries at different levels of granularity. For example, based on the aforementioned enrichment rules, users can directly retrieve all documents belonging to the *Drawingseries_Hessen*. Moreover, the system supports fine-grained, component-level queries, such as requesting all views related to ”pier” components. As shown in Figure 5, when querying for pier components, the system retrieves a list of matching subviews on the left panel, accompanied by their view titles. For each result, its exact location within the original drawing can be further visualized. In this specific example, the red mask effectively highlights the bounding box of the queried pier view.

Figure 6 shows the underlying knowledge graph of the drawing in Figure 5. The graph explicitly represents nine distinct drawing views and their containment relationships. By comparing the two figures, the mapping between visual representations and semantic instances be-

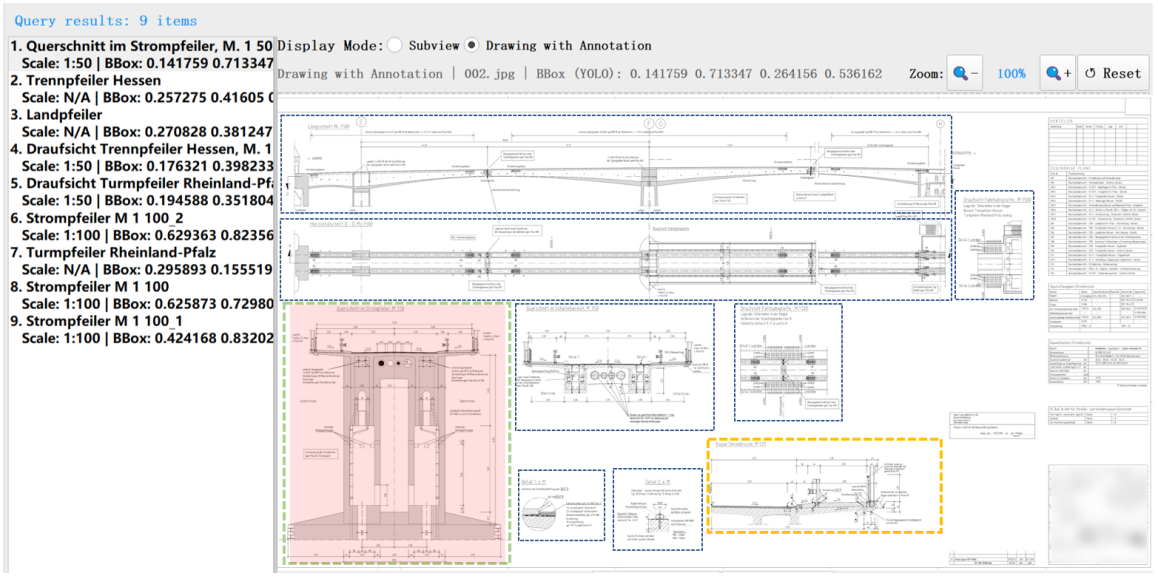


Figure 5: Query and localization of a pier subview within the original drawing

comes transparent. For example, the pier view enclosed in the green dashed box in Figure 5 corresponds to the *DrawingView_23* node. As highlighted by the green dashed box in Figure 6, this node is explicitly marked as a *CrossSectionView* depicting a *Pier* component. Similarly, the view enclosed in the yellow dashed box in the bottom right of Figure 5 maps to *DrawingView_18*, which depicts a *Cap* component. This direct linkage demonstrates how unstructured visual layouts are successfully transformed into a highly organized and queryable semantic network.

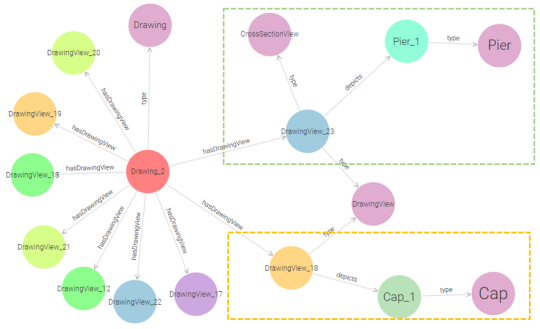


Figure 6: Queried drawing represented in the knowledge graph

Discussion

In the previous section, the proposed pipeline was demonstrated on a set of inspection drawings of the Nibelungen Bridge. The results show that, based on the semantics parsed from subview titles, components can be successfully traced back to their corresponding view locations in the original drawings. As long as the relevant information is explicitly encoded in the drawings, it can be recognized and linked through the ontology, enabling consistent cross-view and cross-stage traceability.

Although the case study uses German bridge drawings, the proposed semantic management approach is not limited to a specific national context. Since the ontology captures generic relationships among drawings, views, components, and project stages rather than country-specific regulations, it can be adapted to drawing collections produced under different standards and documentation conventions. Its applicability is also not constrained by bridge span size. Since BEDO operates at the semantic level of drawings, views, and component occurrences rather than on structural dimensions, it can in principle be applied to bridges with small, medium, or large spans, provided that the relevant information is represented in the drawings. Additionally, the pipeline shows potential for scalability, although this has not yet been validated on a large-scale dataset. Because the AI extraction processes individual drawings independently, the workload can be parallelized to handle much larger numbers of files. Furthermore, the generated semantic data consists of lightweight RDF triples, which modern graph databases can generally manage efficiently at large scale.

However, some limitations remain. In situations where view headings are missing due to overlapping elements, or where titles provide little semantic information on their own (e.g., “Schnitt A - A, M. 1:25”), the current pipeline struggles to assign complete semantics. In these cases, additional context from other parts of the drawing is required to interpret the meaning of the view, which is not yet systematically captured. Consequently, the semantic representation for such views remains incomplete. To further improve the richness of semantic information, additional extraction strategies are needed. These may include exploiting more diverse textual annotations, leveraging layout and graphical context, or incorporating prior domain knowledge to infer missing semantics beyond what is explicitly stated in the view titles. Furthermore, since a manually labeled benchmark dataset for systematically validating metadata extraction is currently unavailable, es-

tablishing one remains an important direction for future work.

Another limitation of the proposed pipeline is that the semantic accuracy of the populated ABox depends on the performance of the upstream AI models used for view detection and OCR. In practical bridge engineering projects, many drawings exist only as scanned documents with heterogeneous quality. Although the pipeline can be applied to such data, low resolution or incomplete scans may reduce extraction accuracy and lead to missing or incorrect semantic assertions. Nevertheless, the ontology-based representation itself remains independent of drawing format once sufficient information has been extracted.

The proposed approach is positioned not as an isolated solution for 2D document processing, but as a complementary mechanism for bridging unstructured legacy drawing data with modern BIM ecosystems. While new infrastructures are increasingly designed natively in 3D, existing assets still rely on vast archives of 2D engineering drawings. To ensure the proposed approach remains highly relevant in a BIM-dominated future, a critical avenue for future work is the direct integration of the proposed ontology with BIM frameworks. This can be achieved by establishing explicit semantic mappings between BEDO and IFC-based bridge models, such as IFCBridge. This integration would enable the transfer and alignment of information between legacy 2D drawings and 3D BIM models. Such a connection would support more comprehensive analyses across multiple data sources and facilitate downstream tasks such as lifecycle management, structural assessment, and digital-twin applications.

Conclusions

This paper presented BEDO, an ontology for representing bridge engineering drawings, together with a pipeline for automatically populating the ontology from the fragmented data produced by AI-based processing of the drawings. By parsing textual information and encoding it as ontology concepts, the approach links drawing content to a formal semantic model. Based on the instantiated ABox assertions, it then enables semantic querying of drawings at the view level. The case study on the Nibelungen Bridge demonstrated that, whenever sufficient semantic cues are explicitly encoded in the drawings, the proposed method can reliably trace component-related views back to their locations in the original documents. At the same time, several limitations remain, such as the incomplete extraction of relationships between views and the dependence on the quality of upstream AI components. Future work will address these limitations and further link BEDO with BIM models to support richer cross-modal analyses and more integrated lifecycle management of bridge infrastructures.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Claude 4.5 and DeepL in order to improve the readability and language of the work. After using these services, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

The authors gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG) within the priority program SPP 2388 “Hundred plus – extending the service life of complex structures through intelligent digitalization” (project number 501804558) and of the Bundesministerium für Verkehr (BMV) within the project “HyBridGen – Hybrid Bridge Generator: AI-based bridge generator with knowledge and experience data and early public participation” (project number 01FV2067A).

References

Akanbi, T. and Zhang, J. (2022). Semi-automated generation of 3d bridge models from 2d pdf bridge drawings. In Construction Research Congress 2022, pages 1347–1354.

DIN EN ISO 10209 (2022). Technische produktdokumentation – vokabular – begriffe für technische zeichnungen, produktdefinition und verwandte dokumentation (ISO 10209:2022). Standard, Deutsches Institut für Normung.

DIN EN ISO 128-3 (2022). Technische produktdokumentation (TPD) – allgemeine grundlagen der darstellung – teil 3: Ansichten, schnitte und schnittansichten (ISO 128-3:2022). Standard, Deutsches Institut für Normung.

Elshani, D., Lombardi, A., Hernandez, D., Staab, S., Fisher, A., and Wortmann, T. (2025). Aec co-design workflow for cross-domain querying and reasoning using semantic web technologies. Automation in Construction, 176:106226.

Elyan, E., Garcia, C. M., and Jayne, C. (2018). Symbols classification in engineering drawings. In 2018 International joint conference on neural networks (IJCNN), pages 1–8. IEEE.

Feist, S., Jacques de Sousa, L., Sanhudo, L., and Poças Martins, J. (2024). Automatic reconstruction of 3d models from 2d drawings: A state-of-the-art review. Eng, 5(2):784–800.

Fletcher, L. A. and Kasturi, R. (2002). A robust algorithm for text string separation from mixed text/graphics images. IEEE transactions on pattern analysis and machine intelligence, 10(6):910–918.

Gifford, D. K., Jouvelot, P., Sheldon, M. A., and O’Toole, J. W. (1991). Semantic file systems. In Proceedings of the thirteenth ACM symposium on Operating systems principles, SOSP91, pages 16–25. ACM.

Hagedorn, P., Liu, L., König, M., Hajdin, R., Blumenfeld, T., Stöckner, M., Billmaier, M., Grossauer, K., and Gavin, K. (2023). Bim-enabled infrastructure asset management using information containers and semantic web. Journal of Computing in Civil Engineering, 37(1).

Hamdan, A.-H. and Scherer, R. J. (2020). Integration of bim-related bridge information in an ontological knowledgebase. In LDAC.

Ji, X. and Abdoli, S. (2023). Challenges and opportunities in product life cycle management in the context of industry 4.0. Procedia CIRP, 119:29–34.

Kang, C., Voigt, C., Eisermann, C., Kerkeni, N., Hegger, J., Hermann, W., Jackmuth, A., Marzahn, G., and Marx, S. (2024). Die nibelungenbrücke als pilotprojekt der digital unterstützten bauwerkserhaltung. Bautechnik, 101(2):76–86.

Kang, C., Walker, M., Bartels, J.-H., Marzahn, G., and Marx, S. (2025). Digital twin technologies for bridge lifecycle management—literature insights and a pilot study on the nibelungen bridge. Results in Engineering, 28:108288.

Lewis, R. and Séquin, C. (1998). Generation of 3d building models from 2d architectural plans. Computer-Aided Design, 30(10):765–779.

Lorvão Antunes, A., Barateiro, J., Marecos, V., Petrović, J., and Cardoso, E. (2024). Ontology-based bim-ams integration in european highways. Intelligent Systems with Applications, 22:200366.

Lu, Q., Chen, L., Li, S., and Pitt, M. (2020). Semi-automatic geometric digital twinning for existing buildings based on images and cad drawings. Automation in Construction, 115:103183.

Lu, T., guan, F., gu, N., and wang, F. (2008). Semantic classification and query of engineering drawings in the shipbuilding industry. International Journal of Production Research, 46(9):2471–2483.

Mashwani, S. R. and Khusro, S. (2018). The design and development of a semantic file system ontology. Engineering, Technology & Applied Science Research, 8(2):2827–2833.

Moreno-García, C. F., Elyan, E., and Jayne, C. (2018). New trends on digitisation of complex engineering drawings. Neural Computing and Applications, 31(6):1695–1712.

Ngo, H. B., Bac, C., Silber-Chaussumier, F., and Le, T. Q. (2007). Towards ontology-based semantic file systems. In 2007 IEEE International Conference on Research, Innovation and Vision for the Future, pages 8–13. IEEE.

Nguyen, M. T., Pham, V. L., Nguyen, C. C., and Nguyen, V. V. (2021). Object detection and text recognition in large-scale technical drawings. Proceedings of the 10th International Conference on Pattern Recognition Applications and Methods (ICPRAM).

Noy, N. F., McGuinness, D. L., et al. (2001). Ontology development 101: A guide to creating your first ontology.

Or, S.-h., Wong, K.-H., Yu, Y.-k., Chang, M. M.-y., and Kong, H. (2005). Highly automatic approach to architectural floorplan image understanding & model generation. Pattern Recognition, pages 25–32.

Peng, M., Cui, A., Kang, C., and Marx, S. (2025). AI-based extraction and management of text and view information from 2D bridge engineering drawings. University of Strathclyde Publishing.

Peng, M., Qian, H., Marx, S., and Kang, C. (2026). Optimizing 2d bridge engineering drawing digitization: A comparative study of text recognition tools and development of lightweight post-recognition structured information extraction methods. Results in Engineering, page 110186.

Ramonell, C., Chacón, R., and Posada, H. (2023). Knowledge graph-based data integration system for digital twins of built assets. Automation in Construction, 156:105109.

Schönfelder, P. and König, M. (2025). Ontology-based reasoning in automatic floor plan analysis. Advanced Engineering Informatics, 68:103761.

Vilgertshofer, S., Stoitchkov, D., Borrmann, A., Menter, A., and Genc, C. (2020). Recognising railway infrastructure elements in videos and drawings using neural networks. Proceedings of the Institution of Civil Engineers-Smart Infrastructure and Construction, 172(1):19–33.

Wu, C., Wu, P., Wang, J., Jiang, R., Chen, M., and Wang, X. (2021). Ontological knowledge base for concrete bridge rehabilitation project management. Automation in Construction, 121:103428.

Yang, J., Xiang, F., Li, R., Zhang, L., Yang, X., Jiang, S., Zhang, H., Wang, D., and Liu, X. (2022). Intelligent bridge management via big data knowledge engineering. Automation in Construction, 135:104118.

Zhang, S., Tang, Y., Zou, Y., Yang, H., Chen, Y., and Liang, J. (2024). Optimization of architectural design and construction with integrated bim and plm methodologies. Scientific Reports, 14(1).

Zinke, T., Reymer, S., Kosse, S., Hagedorn, P., König, M., Wedel, F., Schneider, S., Marx, S., Nieborowski, S., and Windmann, S. (2023). Digital twins for bridges – concept of a modular digital twin based on the linked data approach, pages 1745–1752. CRC Press.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

ENABELING COMPREHENSIVE QUERYING OF ENVIRONMENTAL CONTEXT AND ROAD INFRASTRUCTURE DATA USING A GRAPH-BASED APPROACH

Ina Heise, André Borrmann
Chair of Computing in Civil and Building Engineering, TUM Georg Nemetschek Institute,
Technical University Munich, Germany

Abstract

Enabling predictive maintenance (PM) of road infrastructure by understanding condition developments requires accounting for internal and external influences. However, the corresponding data are highly heterogeneous and distributed, limiting comprehensive large-scale analysis. To bridge this gap, an existing Digital Twin (DT) approach is enhanced with environmental context data. Knowledge graphs enable scalable querying across heterogeneous subsystems, while a federated database addresses data distribution. The framework is validated using real-world data from Germany. Linking historical climate records with bridge condition development demonstrates comprehensive querying. By enabling scalable cross-system data association, the proposed DT framework enhances the practical applicability of existing PM methods.

Introduction

PM is crucial for ensuring the resilience and sustainability of road infrastructure. Its practical implementation is facilitated by informed decision-making, which requires not only a comprehensive understanding of the current state of the system but also an understanding of trends and tendencies in future conditions.

The challenge of comprehensive state representation arises from the system's complexity, driven by its size, heterogeneity, and the various relationships that span multiple scales. Furthermore, when analyzing trends and predicting the development of conditions, considering the system's interactions with its environmental context becomes crucial, as it allows integration of the climate conditions to which the system was exposed. When dealing with complex real-world systems, the DT concept depicts a recognized method. Applying DTs involves using a digital representation of the physical system, along with its various states, to apply computer-based methods. However, the complexity of road infrastructure and its environmental context is also apparent in the existing data, which is distributed across distinct systems managed and updated independently by different institutions and reflects heterogeneity in its underlying data structures. Thus, creating a database suitable for investigating the development of road infrastructure conditions in conjunction with external influences, such as climate, requires associating data with considerable manual effort, which strongly conflicts with the size of the system under consideration. A comprehensive DT that, in a scalable manner, captures road infrastructure alongside its climate exposures and relationships and facilitates automated data association, overcomes this limitation.

Contributions

Studies examining environmental impacts on road infrastructure deterioration primarily rely on a research database that already provides condition and climate data in an integrated manner. While these studies demonstrate the importance of considering environmental context, they neglect the data heterogeneity and distribution prevalent in road infrastructure management practice. Integration of heterogeneous maintenance data, particularly sensor data, is a key topic in research on road infrastructure DTs. Previous work presents frameworks that address the data-integration challenge using graph-based approaches. However, existing approaches focus on individual assets and neglect scalability to entire road infrastructure systems, which is essential for investigating correlations between environmental data and condition development. Other studies on graph-based DT frameworks focus on scalable capture of road infrastructure and its internal relationships, but neglect the system's interactions with the environmental context. Accordingly, a gap remains between the need for large-scale evaluation of road infrastructure maintenance data alongside its environmental contexts and the existing DT approaches that focus either on time-series data integration or scalability. Accordingly, we build on existing approaches for the scalable representation of road infrastructure with its interactions and enhance them by integrating interface relationships with environmental context data, particularly georeferenced time-series climate data. Thereby, not only is the integration of highly heterogeneous data structures at scale accounted for, but also the effects arising from differences in update cycles, spatial structuring approaches, and data distribution across various systems. Furthermore, scalability factoring requires an efficient approach to automatically establish links between time-series data resources and road infrastructure elements during the DT setup. By leveraging characteristics of road infrastructure, we facilitate inference of these relations, thereby significantly reducing the number of required geospatial calculations. The proposed concepts are validated on complete real-world datasets from Ger-

man road infrastructure management and climate recordings. Thus, the ability to handle real-system sizes while acknowledging the data-inherent heterogeneity and distribution is proven.

Related Works

Environmental Context in Infrastructure Deterioration

Considering climatic influences is crucial when investigating the development of road infrastructure conditions. Providing a broad perspective, Sias et al. (2025) conduct a literature review of the effects of climate change on road infrastructure.

Zhang et al. (2025) examine the effects of rising temperatures on asphalt pavement aging. Focusing on more specific mechanisms, Zhou et al. (2020) analyze the effects of freeze-thaw cycles on concrete pavements, while Chen et al. (2023) examines the pavement roughness' dependency on the influences of temperature, moisture, rainfall, and freeze-thaw cycles. Using a data-driven approach, Piryonesi and El-Diraby (2021) integrate temperature ranges, perspiration, and freeze-thaw cycles in their pavement condition prediction. Lamei et al. (2025) investigate climate predictors for bridge component deterioration. It becomes evident that considering the climatic conditions to which a road infrastructure system is, was, and will be exposed is essential in understanding deterioration. This aligns with the Long-Term Pavement Performance (LTPP) program's (Federal Highway Administration, US Department of Transportation (2017), Schwartz et al. (2015)) goal of supporting research on the deterioration of road infrastructure by providing a comprehensive database. The LTPP database contains data on test road sections and bridges in North America and Canada, including condition records and accumulated climate data per test road section. It is apparent that virtually all studies investigating climate effects on deterioration on a larger scale (such as Zhang et al. (2025), Chen et al. (2023), Piryonesi and El-Diraby (2021), Lamei et al. (2025)) build on data from this LTPP database.

Although the LTPP database is a key resource for developing deterioration models, its static storage of aggregated climate data with fixed assignments to road sections introduces limitations. In particular, the lack of flexibility restricts the development of new deterioration models and the application of models that are not aligned with the accumulated metrics provided. In addition, data maintenance and integration of additional data sources are cumbersome, uncertainty quantification is rendered impossible, and adaptation to nonstationary climate conditions is not feasible. Moreover, this centralized data structure does not reflect real-world asset management practices, where road infrastructure and environmental data are typically maintained by separate institutions using distinct systems, each with heterogeneous underlying data structures.

Integrated Infrastructure DTs

Broo and Schooling (2023) highlights DT's capabilities as an effective tool for improving infrastructure systems maintenance through extensive data analysis, but also outlines existing challenges in integrating complex systems and data.

Knowledge graphs (KG) facilitate the integration of heterogeneous datasets by representing elements as nodes connected by edges that capture their relationships as stated in Hogan et al. (2022). Various graph models exist, including the Resource Description Framework (RDF) and Labeled Property Graph (LPG) model. While RDF benefits from unique identifiers and a clearly defined schema through ontologies, the LPG model offers greater flexibility and a more compact structure by allowing properties to be assigned to nodes and edges. While both approaches have their respective advantages and disadvantages, they are, in principle, suitable, as also reflected in existing works.

Hagedorn et al. (2023) presents a graph-based DT framework that integrates heterogeneous maintenance data for roads and bridges, building on a web-based implementation of the Information Container for Linked Document Delivery (ICDD). In Hagedorn et al. (2024), the focus shifts from maintenance to construction, but the ICDD approach is further refined by integrating dynamic sensor data. Although the ICDD approach enables linking RDF graphs with other time-series datasets, thereby facilitating data integration, its reliance on a file-based architecture requires a centralized data storage system. This centralized architecture contrasts with the distributed data management practices, especially when external data sources, such as weather data, are to be included. Furthermore, the largely static structure not only limits adaptability to changing requirements but also increases maintenance effort and reduces scalability in complex, evolving systems.

Li et al. (2024) examine another area of infrastructure systems with their graph-based DT approach for pipes and tunnels. By using LPGs for implementation, the approach benefits from the flexibility and a more compact graph structure. Li et al. (2024) approach involves considering environmental data in the form of measured ground settlements and, by capturing interactions, their effects on the considered structures. However, the approach explicitly stores interaction-describing relationships and operates on a single graph that represents the entire system. Thus, once again, current infrastructure management practices are neglected, just as the system's ability to adapt for integrating additional subsystems, interactions, or data resources.

Ramonell et al. (2023) integrates IoT data with infrastructure IFC models by proposing an LPG-based microservice architecture for asset maintenance. This approach reveals a limitation that also applies to the approaches proposed by Li et al. (2024), Hagedorn et al. (2023), and Hagedorn et al. (2024): In all cases, the integration of sensor data is examined only for individual assets. The scalability of the proposed DT approaches to an entire infrastructure system, including the associated number of assets and the corresponding number of sensors, remains unexamined. Addi-

tionally, when incorporating environmental data, such as climate data, this information often spans a broad area, making it relevant to multiple assets. As a result, adopting existing methods that rely on data integration via explicit component-level links is highly inefficient. In addition, existing approaches also emphasize streaming sensor data. Although this provides a comprehensive view of a component's actual condition, it neglects large-scale evaluation of historical time series, as compared to, e.g., documented condition changes in the context of correlation analysis. Thus, characteristics of road infrastructure systems, such as size, heterogeneity, and the time spans of interest, are not addressed.

Scalable Road Infrastructure DTs

When it comes to the scalable use of knowledge graphs, Purohit et al. (2020) reports improved scalability for LPGs and suggests mapping RDF graphs onto LPGs to combine the efficiency of LPGs with the knowledge provided by the underlying ontologies of RDF graphs.

Thus, existing DT approaches that address the scalable integration of heterogeneous data build upon LPGs. Heise et al. (2025a) introduce a DT-framework for the scalable evaluation of spatial relationships between heterogeneous road infrastructure elements. Heise et al. (2024) extend this approach to capture spatial relationships at multiple scales, and in Heise et al. (2025b), the framework is further enhanced to capture temporal relationships, enabling the investigation of dependencies between condition developments as time-dependent variables. However, while these works enable data analysis that accounts for internal interactions within road infrastructure, external influences, such as climate, remain unaddressed, although existing research underscores their significant impact on deterioration processes.

A gap remains in DT approaches that can also efficiently and scalably integrate long-term observation data at the corresponding frequency, especially since this data naturally lacks infrastructure-specific characteristics and is most efficiently stored in relational (time-series) databases and maintained independently. Thus, we adopt the DT framework from Heise et al. (2025a) for internal relationships in road infrastructure and enhance it by incorporating interface relationships with environmental context data.

Method

The core of capturing the complex relationships and interactions in road infrastructure lies in applying the System-of-Systems (SoS) concept initially introduced by Ackoff (1971). Relationships within road infrastructure and between road infrastructure elements and the environmental context are treated as distinct subsystems, as illustrated in Figure 1 on the left. Elements represented in multiple subsystems serve as connection points that link the subsystems to an overall SoS.

Reflecting this subdivision of the overall system in the architecture of the digital representation not only accounts for the distributed management and storage of data resulting from the actual distribution of responsibilities but also enables the selection of the most suitable database management system (DBMS) type. The right side of Figure 1 depicts the implementation of the SoS concept in a DT framework, incorporating digital representations of road infrastructure and its relationships as *InfraTwin*, and climate data as a distinct *Climate Twin* in *Layer I*. The cross-system evaluation of the subsystems builds on identifying representations of the same elements across subsystems, and implementation is achieved in *Layer II* via a federated database.

Layer I: Subsystems

The *Climate Twin* is assumed to be an existing external system that provides time-series climate data associated with a georeferenced element (typically the weather station where the data were recorded or the area to which the recorded values apply). It is common practice to use relational DBMSs for time series data. As an example for such a system, we consider the *Climate Data Center (CDC)*[1] maintained by the German Meteorological Service (Deutscher Wetterdienst).

The *InfraTwin* comprises further subsystems and builds on the concept Heise et al. (2025a). These subsystems capture both elements of road infrastructure and their conditions, as well as their internal relationships in the *Internal Relationship Graph*. The framework of Heise et al. (2025a) is enhanced by an additional subsystem comprising interface relationships to external elements in the *Interface Relation Graph*. Each *InfraTwin* subsystem is modelled as a knowledge graph and stored in a distinct graph database instance. The modeling of internal relationships within the *InfraTwin* follows the approach presented in Heise and Borrmann (2024) and exploits linear reference elements as a spatial structuring system. Assets and reference elements, stemming from the road network, are represented as nodes. Modeling their relationships implements the principle of absolute linear referencing. To capture the relationships between road infrastructure elements and the environmental context, this spatial structuring approach is leveraged, as it already accounts for road infrastructure-specific characteristics.

Rather than explicitly linking every asset to its environment, the geospatial context elements are mapped onto the linear reference elements as illustrated in Figure 2. Accordingly, it is explicitly modelled which parts of the road network belong to which environmental context. Environmental contexts are represented as areas on a linear reference element with specific start and end points, defined by distance values. Implementing this absolute linear referencing in the graph structure follows Heise and Borrmann (2024) in a simplified manner: environmental data and linear reference elements are represented as nodes, and the relationship between them is captured by connecting edges, as detailed in Figure 2 on the right.

[1] https://cdc.dwd.de/

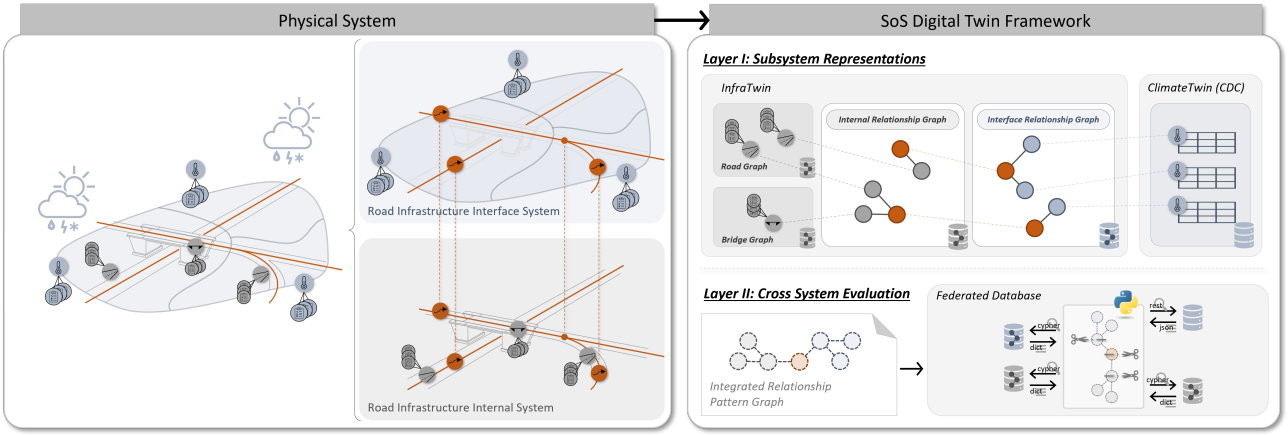


Figure 1: Representing Road Infrastructure with its Environmental Interactions as System-of-Systems Digital Twin Framework

Information on whether a road infrastructure element is located in a specific area can be derived by superimposing its relationships with the common reference element. Therefore, spatial relationships between environmental context and road infrastructure elements can be evaluated with the same precision as those between assets. This includes $n : n$ relationships between areas and assets, which may be important for structures spanning several kilometers, such as noise barrier walls, and their relationships to potentially multiple areas. In summary, the road network serves as the connection point between the two subsystems, which reduces not only the number of cross-system connections but also the number of elements to be represented in the *Interface Relationship Graph*, thereby efficiently reducing its size.

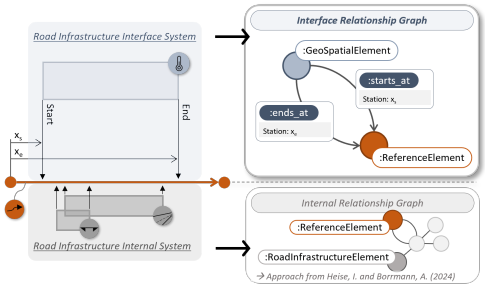


Figure 2: Graph Structures of the Relationship Graphs

Layer II: Comprehensive Evaluation

Applying the SoS principles and the corresponding subdivision into subsystems results in a distribution of information. Thus, elements with their corresponding data and various relationships are spread across several DBMSs. The cross-system relationship can be described as a pattern graph that spans the various subsystem graphs. Thus, identifying elements that exhibit a specific cross-system relationship comes down to a pattern-matching problem, searching for occurrences of this pattern graph in the overall system. The distributed storage of the subsystems turns the problem into a distributed pattern-matching problem, in which each pattern graph fragment corresponds to a single subsystem.

The implementation of distributed pattern matching forms *Layer II* in Figure 1. A federated database connects the distinct DBMSs of the different subsystems without holding any data itself. Instead, for each pattern graph fragment, corresponding queries are sent to the respective subsystem database endpoints. The order (as well as potential parallelization) of this query process is determined by the dependencies defined in the overall pattern graph. Additionally, aligning the query direction with the expected solution set sizes for each subsystem evaluation helps quickly exclude large portions of the system, thereby enhancing overall evaluation efficiency. Implementation of *Layer II* can be achieved through a script, e.g, using Python.

It is worth noting that the query language and the way data is accessed may vary across fragments, as the concept explicitly considers different DBMS types across subsystems. The depiction in *Layer II* of Figure 1 assumes that Neo4j is used as a graph DBMS, and accordingly, interactions involve sending Cypher queries, while the climate time series can be efficiently obtained via a REST API.

Creation of the Interface Relationship Graph

The *Interface Relationship Graph* comprises the relationships between the geospatial environmental context and the road network edges. Since these relationships are not yet present in the data, geospatial calculations are required. Assuming the road network's edge geometries are polylines, this calculation identifies which segments of those polylines fall within specific environmental areas. In principle, this analysis corresponds to geospatial querying already implemented in diverse interfaces and database engines, and it is also applied in existing approaches, such as those presented in the Related Works Section.

Yet, given the specific characteristics of road infrastructure systems, particularly the common sizes of real-world road networks, this poses a fundamental dilemma between geometric analysis accuracy and scalability. In addition, the number of polyline segments of a network edge is often unrelated to its spatial extent. On the contrary, network edges with a small spatial extent, often connecting other main axes, typically have a large number of polyline seg-

ments due to their high curvature. Therefore, considering inherent net characteristics when determining the spatial relationships between the road network and environmental areas, rather than computing them for every road network polyline segment, appears reasonable.

The proposed approach builds on the assumptions that (1) the geometry of the road network is significantly finer-grained than its environmental context areas. Thus, one network edge intersects at most one boundary between environment areas. And (2) only qualitative relationships between the road network and the environmental context areas are to be evaluated, and a tolerance of the geometric calculations of several meters is acceptable.

The approach consists of a two-step process: Firstly, the geospatial relationship between the net nodes and the areas is computed. Based on this information, the network edges themselves are considered second. During the analysis of the network edges, two cases are distinguished, aiming to use inference as far as possible rather than a geometric analysis. Figure 3 illustrates both cases.

In *Case I*, relationships between net edges and areas can be completely inferred by assuming that if two net nodes V_1 and V_2 are within an area A, this also applies to the network edge (and correspondingly the reference element) E connecting V_1 and V_2. Accordingly, the spatial relation required to create the *Interface Relation Graph* is determined using Equation 1. As the environmental area spans the entire reference element, the start station is 0, and the end station is the length of the reference element, respectively.

$$\frac{\mathrm{in}(V_1, A) \quad \mathrm{in}(V_2, A) \quad \mathrm{starts}(V_1, E) \quad \mathrm{ends}(V_2, E)}{\mathrm{in}(E, A)} \tag{1}$$

Case II covers V_1 and V_2 contained in different areas A_1 and A_2 and therefore involves the reference element intersecting a boundary. To localize the boundary P_{Bound} on the reference element, the individual points of the polyline $P_1, ..., P_n$ are checked for their containment in A_1. P_{Bound} corresponds to the first point identified outside A_1, and investigations stop if P_{Bound} is identified. The P_{Bound} station together with the length of the reference element provides the information to model the relationship between the two environmental areas and the reference element, as illustrated in Figure 3.

The result is the *Interface Relation Graph*, which captures the spatial relationships between environmental context areas and linear reference elements, thereby facilitating the comprehensive evaluation of climate and asset data.

Prototypical Real-World Implementation

To validate the practical applicability of the proposed framework, it is applied to a real road infrastructure system. Real-world data from German infrastructure management and publicly available climate data, both structured differently and originating from distinct sources, are used.

Setting Up the DT-Framework

The *InfraTwin* integrates data on bridges and the German road network, while the *ClimateTwin* is represented by the CDC. Bridge data in Germany is maintained in SIB-Bauwerke[2], a proprietary asset management system. SIB-Bauwerke builds on a relational database and supports only file-based data exchange. For the case study, we used a copy of the database for all Bavarian (a state of Germany) assets provided to us as part of a research project. The German road network is centrally managed and publicly provided by the Federal Highway Administration (Bundesanstalt für Straßenwesen (BASt)). For the case study, the data were obtained through the WFS interface[3]. The CDC provides several interfaces, including a REST API and an open file server. Since the REST API interface is still under construction and therefore does not yet provide access to the full data set, the HTTP-based open file server[4] was used, despite the resulting limitations in filtering options. The prototypical implementation of the concept employs Labeled Property Graphs (LPG) and Neo4j, a widely used graph DBMS, for their storage. The federated database is implemented in a Python script.

Internal Relationship Graphs Creation

To facilitate integration of the different legacy data sources, graph representations of the bridge data and the road network were created and stored in distinct graph database instances. Capturing the internal relationships follows the approach presented in Heise et al. (2025a). This involves creating the *Internal Relationship Graph*, which captures the spatial relationships between bridges and reference elements (as referred to in Heise et al. (2025a) as the *InfraNetRelation Graph*). Furthermore, graphs representing the net (*Net Graph*) and the bridge condition recordings (*Bridge Graph*), each stored in a distinct graph database, were generated.

Interface Relationship Graph Creation

In this case study, the *Net Graph* serves as a basis for creating the *Interface Relationship Graph*. According to the concept presented earlier, this involves assigning net nodes to their environmental contexts first. Since the climate data in the CDC is stored associated with weather stations possessing point coordinates, calculating the required information yielded a Voronoi-like decomposition via Euclidean nearest-neighbor assignment. To ensure computational efficiency, the distance calculation was performed using a vectorized approach. For each net node, distances to all weather stations were computed simultaneously, and the minimum value was selected, which avoided the construction of a full distance matrix. Implementation was achieved through a Python script using the NumPy library. The computed assignments of net nodes and weather sta-

[2] https://sib-bauwerke.de/

[3] https://inspire.bast.de/geoserver/bisstra.strasse/wfs?service=WFS

[4] https://opendata.dwd.de/

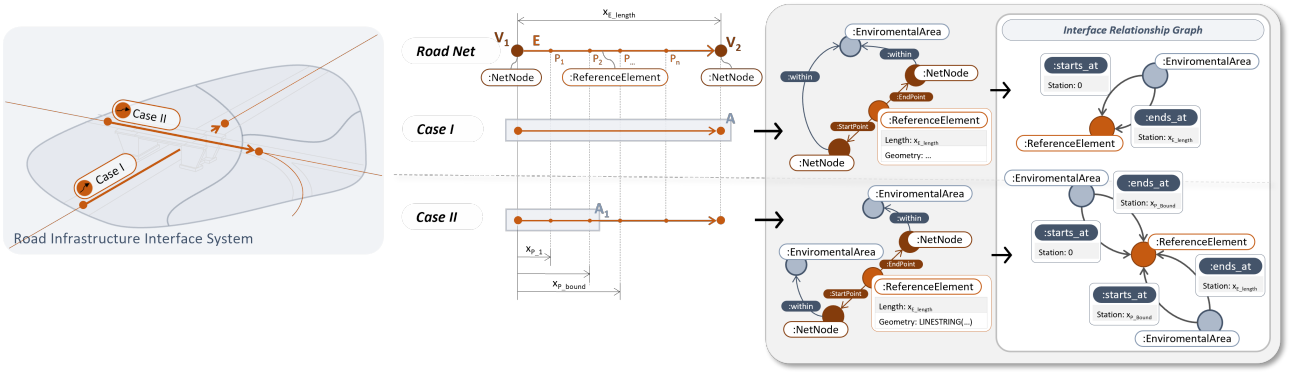


Figure 3: Interface Relationship Graph Creation

tions are fed into the *Net Graph*.

Based on this, the interface relationships between reference elements and weather stations satisfying the conditions of *Case I* could be inferred. The inference is implemented using a CYPHER query, which first extracts the corresponding reference element-weather station pairs (along with the length of the reference element and the respective IDs). Based on the created dictionary, the corresponding nodes and relations in the *Interface Relationship Graph* are generated.

The remaining reference elements fall under *Case II*. To determine x_{P_bound}, a CYPHER query directed to the *Net Graph* identifies the corresponding reference elements with associated weather stations linked to its start and end point. The IDs, along with their geometry (stored as WKT), are given back. The calculation of x_{P_bound} is then implemented in Python using the NumPy and Shapely libraries. The calculation results facilitate the enhancement of the *Interface Relationship Graph* with the remaining relationships between weather stations and reference elements.

The Example Use Case

Associating bridge conditions with frost-thaw cycles derived from daily temperature recordings serves as an example use case. The bridge conditions are contained in the *Bridge Graph* linked to the respective asset. Temperature recordings are provided in different resolutions in the CDC. Automatic association is facilitated by comprehensively evaluating the *Interface Relationships Graph* and *Internal Relationship Graph*, as illustrated in Figure 4. Firstly, the *Internal Relationship Graph* provides all Bridges in the dataset, along with their localization on a reference element. This information is used to subsequently query the *Interface Relationship Graph*. Using the reference elements' IDs, the weather station corresponding to each bridge is determined. The resulting dictionary contains bridge IDs assigned to weather station IDs.

For further querying, the capabilities of the provided interfaces were taken into account. If fetching temperature records for a specific time interval is possible, e.g., via a REST API, continuing the sequential process in which the condition ratings of the bridges were queried first, followed by the temperature data for the covered intervals, is beneficial. However, the CDC interface provides only a single text file containing all historical temperature data for each weather station, prohibiting the preselection of climate data based on bridge data availability. Accordingly, parallel data retrieval was implemented. The first subprocess extends the dictionary of weather station IDs and bridge IDs with a list of available condition grade recordings per bridge by interacting with the *Bridge Graph*. For each weather station ID, historical soil temperature records are retrieved from the CDC. Within the CDC, the ZIP archives corresponding to a specific weather station could be automatically identified and downloaded (via an HTTP request) using the employed naming convention. After downloading the ZIP archives, they are opened in memory, and the text file containing the actual temperature data is extracted and stored, using the corresponding weather station's ID as the file name.

Finally, investigating temporal relationships by implementing the approach presented in Heise et al. (2025b) enables associating frost-thaw cycles with bridge condition recordings, thereby facilitating further data analysis.

Results

The prototypical implementation demonstrated the applicability of the proposed framework to real-world data sets from Germany. The CDC dataset comprised data from 5.311 weather stations. The road network contains 36.565 network edges (and respectively reference elements). The relationship of 87% (31.916) of these reference elements to weather stations could be completely inferred as they fell under *Case I* (Figure 3). Only the remaining 4.649 reference elements required detailed analysis.

The *Internal Relationship Graph* contains 31.309 asset representations. Since we had access to asset data only from a part of Germany (Bavaria), but could use the road network for the entire country, this resulted in fewer assets as reference elements. Nevertheless, the reduction in explicitly modeled relationships by using the net as a connecting element is evident in the average number of assets per reference element, which is 3.3 and is expected to be even higher if road sections are to be integrated.

Implementing automated association of bridge condition

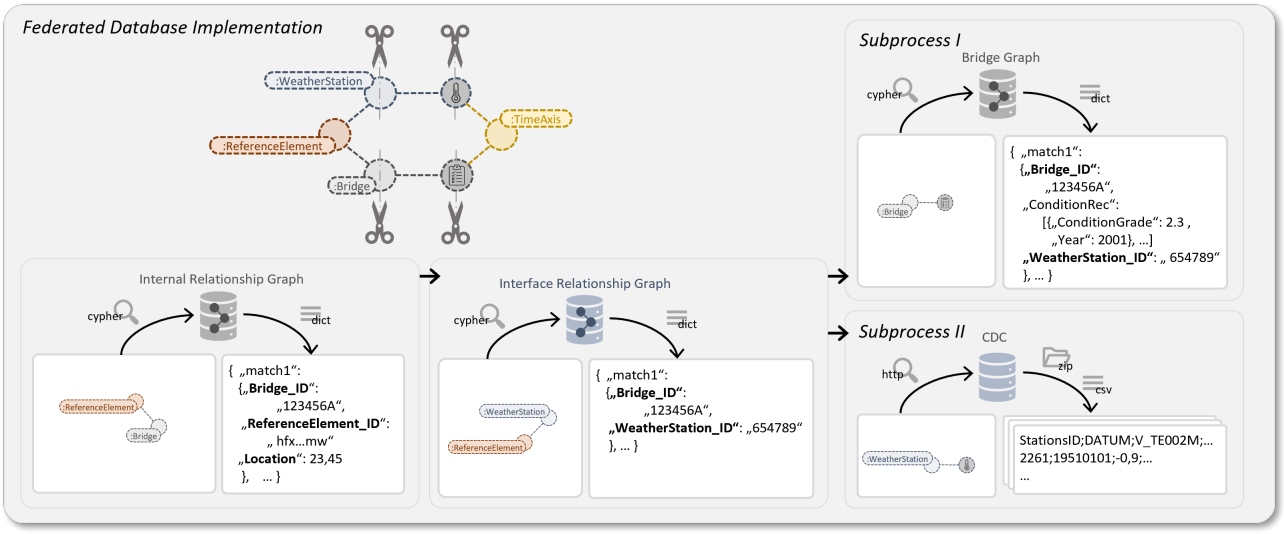


Figure 4: Prototypical Implementation of Layer (II): Federated Database for Comprehensive Evaluation

data with temperature recordings was successful. However, evaluating the temporal relationships proved a bottleneck due to the large number of temperature recordings.

Discussion

The prototypical implementation demonstrates the proposed concept's ability to comprehensively evaluate environmental context and road infrastructure data at a large scale, using temperature and bridge condition records. By considering complete real-world datasets, not only are realistic system sizes accounted for, but also the inherent distributed storage and data maintenance, as well as the significant heterogeneity of data structures involved. This paves the way for applying existing works on multifactorial deterioration models, which currently build on the central, homogenised LTPP data set, to a broader context, including infrastructure management systems currently in use.

By using the road network as a connection point between the two subsystems, a spatial structuring approach accounting for infrastructure-specific characteristics is employed. The approach's effectiveness results from reducing explicitly modeled relationships between road infrastructure and external elements. This minimizes not only the size of the *Interface Relationship Graph*, thereby improving performance during querying and creation, but also the number of cross-system connections that must be considered and maintained.

In addition, mapping environmental context areas to the reference elements enhances the accuracy of spatial relationship evaluation by accounting for the large longitudinal extent of the infrastructure assets. However, due to the very large difference in granularity when mapping climate data to the German road network, only 13% of the reference elements even cross a boundary between two environmental context areas. The effect can therefore be better exploited by examining relationships with environmental data of finer resolution, such as soil data.

The approach faces limitations in ensuring consistency across several subsystems. In particular, this affects changes to the road network or the available weather stations, which require updating the *Interface Relationship Graph* to ensure complete integration of the available data sources. Provided changes are known, the elements' IDs enable efficient updates. However, as the data sources involved are maintained completely independently, this knowledge most likely does not exist. In this case, there is a lack of efficient update strategies that prevent the entire subsystem graph from having to be checked for updates or completely regenerated. Although complete regeneration of the *Interface Relationship Graph* is not the optimal updating strategy, it still benefits from minimizing connection points between subsystems by using the road network instead of individual assets, and from deploying inference to avoid unnecessary geospatial calculations.

Furthermore, the case study revealed the limited effectiveness of the approach of Heise et al. (2025b) for analyzing temporal relationships when the data exhibit a very high degree of heterogeneity in update cycles. Integrating the concept of reference intervals proposed by Allen (1983) appears worth investigating.

Conclusion

The paper presents a comprehensive DT framework that captures road infrastructure and its environmental context. The integration is facilitated by applying SoS principles in the digital representation. A knowledge-graph-based approach addresses scalable semantic integration by capturing contextual relationships and enabling their flexible evaluation despite their inherent heterogeneity. The comprehensive evaluation, accounting for data distribution across heterogeneous DBMS (types), is implemented as a federated database. The applicability of the concept is proven by a prototypical implementation using complete real-world datasets from German infrastructure management and the German meteorological service (Deutscher Wetterdienst, DWD). The achieved data association en-

ables investigating dependencies between condition development and temperature fluctuations. Thus, the presented approach paves the way for PM in infrastructure by leveraging existing data to better understand deterioration.

Acknowledgements

The presented research has been funded by the TUM Georg Nemetschek Institute - AI for the Built World.

References

Ackoff, R. L. (1971). Towards a System of Systems Concepts. Management Science, 17:661–671.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. Communications of the ACM, 26:832–843.

Broo, D. G. and Schooling, J. (2023). Digital twins in infrastructure: definitions, current practices, challenges and strategies. International Journal of Construction Management, 23:1254–1263.

Chen, C., Deng, Y., Li, M., and Shi, X. (2023). Investigation of key climatic factors affecting asphalt pavement roughness in different climate regions. Transportation Research Part D: Transport and Environment, 122:103877.

Federal Highway Administration, US Department of Transportation (2017). The Long-Term Pavement Performance Program. Technical report.

Hagedorn, P., Liu, L., König, M., Hajdin, R., Blumenfeld, T., Stöckner, M., Billmaier, M., Grossauer, K., and Gavin, K. (2023). Bim-enabled infrastructure asset management using information containers and semantic web. Journal of Computing in Civil Engineering, 37.

Hagedorn, P., Petrova, E., and König, M. (2024). A system-of-systems approach for deploying containerized construction digital twins using linked data. Proceedings of the 20th conference of the International Society for Computing in Civil and Building Engineering (ICCCBE2024).

Heise, I. and Borrmann, A. (2024). infraspatialOT: An ontology for the representation of spatial relationships in road infrastructure. In LDAC 2024 - Linked Data in Architecture and Construction.

Heise, I., Esser, S., and Borrmann, A. (2025a). A graph-based systems-of-systems architecture enabling multi-scale Digital Twins for maintaining road infrastructure. Advanced Engineering Informatics, 68:103649.

Heise, I., Esser, S., and Borrmann, A. (2025b). Enabling Spatial and Temporal Correlation of Heterogeneous Infrastructure Data Using Knowledge Graphs. Proceedings of the 2025 European Conference of Computing in Construction (2025 EC^3).

Heise, I., Göbels, A., Borrmann, A., and Beetz, J. (2024). Enabling Comprehensive Querying of Road and Civil Structure Data using Graph-based Methods. In Proceedings of the 41st International Conference of CIB W78.

Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., and Zimmermann, A. (2022). Knowledge Graphs. ACM Computing Surveys, 54:1–37.

Lamei, T., Abdel-Rahim, A., and Ibrahim, A. (2025). Development of performance-based deterioration models for bridge components. Innovative Infrastructure Solutions, 10:230.

Li, T., Rui, Y., Zhu, H., Lu, L., and Li, X. (2024). Comprehensive digital twin for infrastructure: A novel ontology and graph-based modelling paradigm. Advanced Engineering Informatics, 62:102747.

Piryonesi, S. M. and El-Diraby, T. (2021). Climate change impact on infrastructure: A machine learning solution for predicting pavement condition index. Construction and Building Materials, 306:124905.

Purohit, S., Van, N., and Chin, G. (2020). Semantic Property Graph for Scalable Knowledge Graph Analytics. In 2020 IEEE International Conference on Big Data (Big Data), pages 2672–2677. IEEE.

Ramonell, C., Chacón, R., and Posada, H. (2023). Knowledge graph-based data integration system for digital twins of built assets. Automation in Construction, 156.

Schwartz, C., Elkins, G., Li, R., Visintine, B. A., Forman, B., Rada, G. R., and Groeger, J. L. (2015). Evaluation of LTPP Climatic Data for Use In Mechanistic-Empirical Pavement Design Guide(MEPDG) Calibration and other Pavement Analysis. Technical report.

Sias, J. E., Dave, E. V., Underwood, B. S., Bowers, B. F., Harvey, J. T., Henning, T. F. P., Tighe, S. L., Jacobs, J. M., Pregnolato, M., Qiao, Y., Mecray, E., Golalipour, A., Chamorro, A., and Hendrick, P. (2025). Climate change impacts on roadways. Nature Reviews Earth & Environment, 6:555–573.

Zhang, R., Sun, L., Qiao, Y., Sias, J. E., and Dave, E. V. (2025). Multidimensional comparative analysis of future climate change impacts on pavement infrastructure aging. Transportation Research Part D: Transport and Environment, 142:104702.

Zhou, C., Lan, G., Cao, P., Tang, C., Cao, Q., Xu, Y., and Feng, D. (2020). Impact of freeze-thaw environment on concrete materials in two-lift concrete pavement. Construction and Building Materials, 262:120070.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

GREENBIMXELS: A DATA SPACE UNLOCKING ENVIRONMENTAL VALUE FOR PUBLIC FACILITIES IN THE CONSTRUCTION SECTOR

Gordo-Gregorio, Paula[1]; Coloma Picó, Eloi[1]; Martínez Ibañez, Manel[1]; Macarulla Martí, Marcel[1]; Angulo, Cecilio[1]
[1]Polytechnic University of Catalonia, Barcelona,Spain

Abstract

Public administration buildings generate vast operational data, which is rarely used for environmental purposes. GreenBIMxels, a research project led by the Polytechnic University of Catalonia, explores how Data Spaces support environmental assessments in public facilities.

This paper defines environmental services for BIM models within a Data Space named UPCxels. The study addresses the interests of public administrations and private-sector stakeholders, examining how Data Spaces can respond to their needs. Three use cases are designed by combining participatory workshops and a questionnaire survey. The use cases include energy-efficiency, water-consumption optimization and carbon-footprint assessment, each operationalized through KPI sets and BIM/IFC-based properties.

Introduction

Building Information Modeling (BIM) is increasingly used to manage building and infrastructure assets by means of shared digital representations that support decision-making across design, construction and facility management throughout the asset lifecycle. By consolidating geometric, technical and operational information into a single model, BIM improves coordination among stakeholders and provides a data-rich foundation for operation and maintenance activities (Gordo-Gregorio et al., 2025). Despite this potential, BIM is still frequently applied in a construction-centric way, which constrains the systematic reuse of information for environmental performance monitoring or long-term portfolio planning, particularly within public administrations (Akbari et al., 2024; Kivits & Furneaux, 2013).

In parallel, the European data strategy and recent work on Data Spaces propose federated infrastructures where data holders remain sovereign over their assets while enabling controlled sharing and cross-organizational analytics and demand alignment with the international data-space principles (the Data Spaces Support Centre (DSSC), Gaia-X, and the International Data Spaces Association (IDSA)) (Enes, 2025; Valk & Ryan, 2025). Sectoral Data Spaces define common governance, trust and interoperability frameworks so that heterogeneous datasets can be securely combined and analyzed without centralizing sensitive information (Kovach et al., 2026). For the built environment, this paradigm is particularly relevant, because public owners, operators and service providers need to cooperate around services such as energy demand, while respecting contractual constraints, privacy and institutional autonomy (Janev et al., 2021).

At the same time, research in BIM-based sustainability assessment shows the importance of harmonized, object-level data structures to integrate building models with environmental information such as life-cycle assessment (LCA) datasets and operational performance indicators (Klumbyte et al., 2023). Several authors have proposed BIM-centered workflows for energy analysis or water-use evaluation, but these approaches often rely on project-specific schemas, ad hoc integrations or proprietary platforms, which hampers scalability (González et al., 2021; Sameer & Bringezu, 2021). There is therefore a need for an open semantic framework that connects BIM, standardized data dictionaries and environmental datasets within a data-space architecture, so that AI-based services can perform analytics on public facilities.

To demonstrate the practical impact of this approach, GreenBIMxels develops three use cases focused on environmental performance of public facilities (energy, water and carbon footprint). For each case, structured data from BIM models of public facilities are combined with external data bases and IoT data within the Data Space, enabling advanced analytics and AI-based services without exposing raw operational data outside their institutional boundaries. These use cases are defined in this paper to evaluate how a BIM-based Data Space can unlock environmental value from existing BIM datasets of public facilities, support evidence-based decision-making, add value and reduce fragmentation between administrations in the management of building and infrastructure assets.

Methodology

The three use cases were developed with companies and public administrations, within the Catalan context, through a structured, multi-stage Co-Creation Method that combined participatory workshops and a formalized questionnaire-based survey targeting future data consumers. This approach ensured that the selected use cases were not only technically feasible within the GreenBIMxels data-space architecture but also aligned with real market needs, operational practices, and governance expectations. The process started with a series

of dedicated workshops with data providers, aimed at identifying priority environmental challenges, and feasible analytical services that could be realistically deployed within a data-space context. The development of these collaborative workshops is aligned with the guidelines of the DSSC (Data Spaces Support Centre), which define the Co-Creation Method with the aim of ensuring that the evolution process of any Data Space is carried out in an appropriate and structured manner. This process, as shown in the following figure, consists of three stages: the exploratory phase, the preparation phase and the implementation phase (Fauth et al., 2024; Weiss et al., 2025). This paper includes the exploratory and preparatory stages.

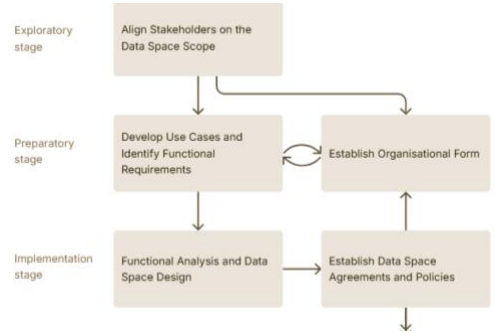


Figure 1 Workflow of the Co-Creation Method. Source: DSSC

A total of ten public administrations participated, including municipalities, regional government departments, public institutes, and public bodies responsible for infrastructure and transport, with technical staff drawn from energy services (end users with in-depth knowledge of operational processes), maintenance representatives (end users with process expertise), and organizational managers and BIM specialists acting as data providers or decision-makers. During these workshops, participants discussed current practices related to energy management, water consumption monitoring, and carbon-footprint assessment, as well as limitations in accessing, sharing, and reusing BIM and operational data across organizational boundaries. The workshops combined exploratory and convergence-oriented activities. First, they gathered information on environmental data services by examining the types of BIM models and IFC datasets currently available, the environmental functionalities that data providers aim to enable through the Data Space, and the key performance indicators (KPIs) used or anticipated for decision-making. Building on these discussions, the workshops converged on three representative, high-impact domains (energy efficiency, water consumption optimization, and life-cycle carbon-footprint assessment) which are common across multiple public administrations and closely aligned with policy-driven sustainability objectives. In parallel, the workshops focused on defining the implementation process, identifying the main barriers that administrations expect to face during the project, and outlining the required technological approach for effective deployment. To complement the qualitative insights with quantitative rigor, workshop contributions from public administrations were analyzed using co-occurrence network analysis and text mining techniques implemented in the R programming language (Amancio et al., 2025).

Furthermore, a structured questionnaire survey was subsequently distributed to prospective data consumers, including architecture and engineering firms, technical consultancies, software providers, and sustainability-oriented service companies. The questionnaire was designed to capture a detailed profile of data consumers, their typical data usage patterns, integration challenges, and required update frequencies. In addition, the survey explored perceived value and priority of different environmental use cases, expectations regarding AI-based analytical services, preferred levels of data processing, and requirements related to data governance, quality and security. This method enables a thorough assessment by integrating multiple data sources, thereby strengthening the reliability of results in applied research contexts (Mayo-Wilson et al., 2018). Consumer surveys provide structured quantitative and qualitative feedback on perceived outcomes, while workshop Post-it notes offer spontaneous, context-rich reflections from providers directly involved in delivery.

The results of the workshops and surveys were jointly analyzed to define the scope, data requirements and KPIs. For each case study, a clear mapping was established between the proposed analytical services, the required data properties and IFC-based data elements, and the KPIs used to assess environmental performance and impact. As part of the preparatory stage, the definition of indicators was carried out by translating the co-created use cases into measurable parameters. This process aimed to operationalize the needs and analytical requirements identified during the workshops into a structured set of indicators suitable for implementation within the data-space architecture.

The development of these indicators followed an iterative approach. First, candidate parameters were derived from workshop discussions, reflecting their data needs and decision-making criteria. Second, these parameters were refined through two complementary activities, an analysis of the digital tools and information systems currently used to ensure alignment with existing data availability and formats and validation through expert consultations and literature review to assess completeness and relevance.

During this process, some parameters initially identified were discarded due to their limited contribution to the proposed services, while additional parameters emerged as necessary to ensure analytical feasibility and robustness. The final set of parameters was then structured into KPIs and mapped to BIMxel properties and IFC-based data elements, forming the basis for the KPI tables presented in the Results section.

As a result, the methodology ensured coherence between stakeholder needs, available and shareable data, and the analytical capabilities supported by the GreenBIMxels infrastructure, thereby increasing both technical robustness and adoption potential.

Results and discussion

The results presented in this section derive from the combined analysis of two complementary empirical sources, a series of structured workshops with public administrations and data providers, and a questionnaire-based survey addressed to prospective data consumers from the construction, engineering, digitalization, and sustainability sectors. This approach simultaneously provided feedback on the project implementation and supported the definition of the three use cases, including their associated KPIs.

On the questionnaire survey most respondents operate as direct data consumers but also express a strong interest in developing data-based services or advisory products, indicating that data consumption and value creation are closely linked in their business models. From a digital readiness perspective, 87% of respondents reported a medium to high level of maturity in managing BIM/IFC data. This finding is consistent with the workshop results, where public administrations demonstrated the availability of IFC models, monitoring systems, and sectorial platforms (such as BMS, SCADA or energy monitoring tools), albeit in a fragmented and heterogeneous manner. Together, these results suggest that both providers and consumers possess a sufficient baseline of digital capabilities to engage with a data-space-based solution, while still facing significant integration challenges.

Project implementation

The findings reflect consumers' perspectives as captured through questionnaire responses and providers' viewpoints as emerging from the workshop discussions. The questionnaire survey was completed by 8 organizations out of 15 invited (response rate ≈ 53%). The questionnaire sample could be acknowledged as modest in absolute terms, yet it can still be considered adequate for the exploratory and descriptive aims of this study. In survey methodology, sample adequacy is not determined solely by size but by its ability to capture the relevant heterogeneity of the target population (Erdogan & Baker, 2002; Fosnacht et al., 2017). From this perspective, the present sample covers micro, small, and medium-sized organizations involved in different stages of the building and infrastructure life cycle, which enhances its structural similarity to the population of interest and supports its analytical representativeness. Post-it notes collected during co-creation workshops were analyzed to identify emerging themes and stakeholder concerns.

As a result, a co-occurrence network (Figure 2) was built based on the content of the post-it notes, which were generated in response to two guiding questions addressing the needs and challenges related to the sustainable management and equipment maintenance of their building assets. Each post-it note was systematically coded to extract key terms. The network was then developed by mapping the frequency with which these terms co-occurred across responses, enabling the identification of the following thematic clusters:

- Need for research: Participants highlighted gaps in applied knowledge on data-space integration for BIM, calling for more pilot studies and evidence on ROI.
- Need for interoperability: Frequent mentions of heterogeneous formats, proprietary tools, and semantic standardization.
- Data and models availability/sharing from administrations: This emerged as the dominant cluster, reflecting frustrations with fragmented access, outdated datasets, and bureaucratic silos.
- Lack of human resources: Capacity constraints, including skilled BIM personnel and training, were recurrent, limiting sustained adoption.

The diagram's visual prominence of the data-sharing cluster signals broader sector challenges like platform fragmentation. Both data consumers and public data providers highlighted that access to operational consumption data (energy and water) and geospatial information remains uneven and often constrained by organizational or technical barriers.

A key convergent result concerns the main barriers to effective data use (Figure 3). Respondents gave greatest importance to practical and structural barriers, especially the cost or time required to obtain data, along with interoperability and incomplete data. In contrast, lack of knowledge was considered less important. This perception mirrors the workshop findings, where public administrations repeatedly emphasize platform fragmentation, dependency on proprietary systems, and difficulties in maintaining up-to-date and consistent datasets across multiple tools and departments.

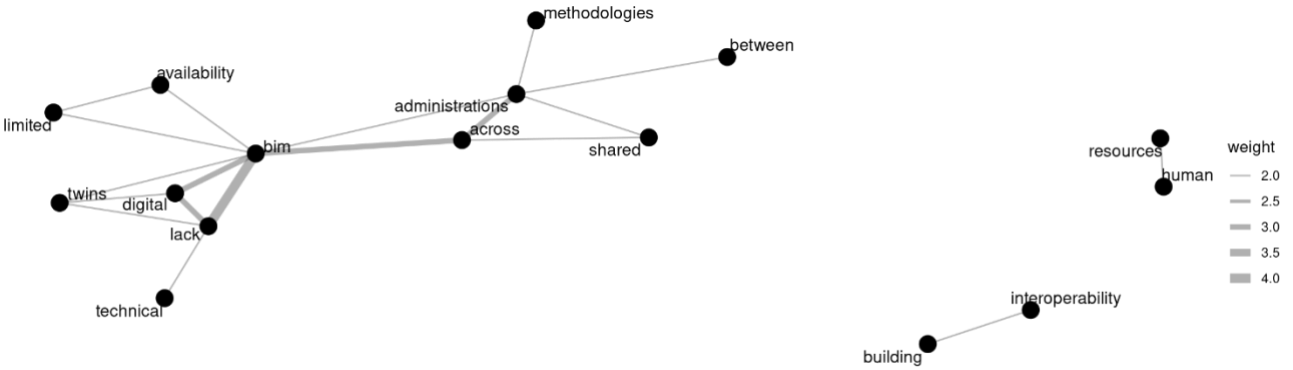


Figure 2 Co-occurrence network of implementation barriers identified in the public administration workshops

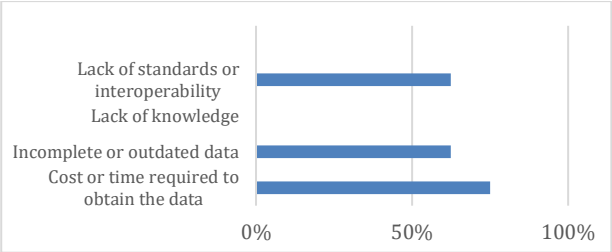


Figure 3 Main difficulties in integrating facility management information into data-consumers workflows (n=8 Companies).

Regarding adoption, consumers emphasized that technical feasibility alone is insufficient, uncertainty about return on investment remains the main barrier (Figure 4). Expected benefits are framed primarily in terms of market differentiation and access to new business opportunities, rather than purely internal efficiency gains (Figure 5).

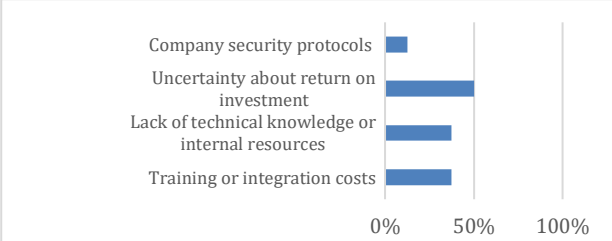


Figure 4 Identified barriers by data-consumers on connecting to the Data Space (n=8 Companies).

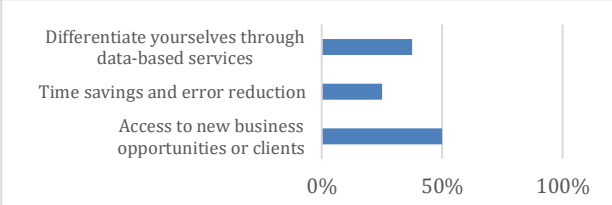


Figure 5 Identified benefits by data-consumers on connecting to the Data Space (n=8 Companies).

Trust-related aspects emerged as critical enablers of participation in the GreenBIMxels Data Space. Data consumers identified confidentiality, secure access, and ease of integration as essential guarantees, with data traceability and clear licensing conditions also playing an important role (Figure 6). These expectations align closely with the governance concerns expressed in the workshops, where compliance with European data-space principles and security requirements was considered a prerequisite for institutional adoption.

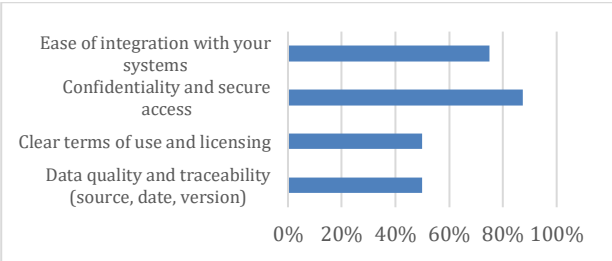


Figure 6 Key guarantees required for data consumers to connect to the Data Space (n=8 Companies).

These results underscore the importance of demonstrating concrete, replicable value through the selected use cases.

Definition of use cases

The network graph in Figure 8 visualizes the thematic clustering of post-it notes from workshops, where node size reflects frequency of mentions, colors denote domains (orange: energy, purple: water, green: carbon, gray/black: transversal), and edges indicate interconnections that directly informed the definition of GreenBIMxels use cases and KPIs.

Gray central nodes dominate the core including "data", "comparison", "monitoring", "consumptions", "IFC", forming a highly repetitive and interconnected hub that underscores these topics as transversal KPIs essential to all three use cases. Their prominence and repetition signal that reliable, standardized access to BIM/IFC models and operational consumption metrics constitutes the primary analytical foundation, enabling portfolio-scale visualization, anomaly detection, and benchmarking across energy, water, and carbon domains.

Energy emerges as the most prominent use case, with the largest orange cluster showing high frequency and dense connections to gray transversals, indicating it as the top stakeholder priority for KPIs like power curves, efficiency patterns, and predictive interventions. This priority is also reflected among consumers in the questionnaire survey (Figure 7). Strong orange-green linkages reveal a tight relationship between energy and carbon footprint, particularly for renovation decisions, material choices, and lifecycle optimizations, positioning combined energy-carbon KPIs (such as CO2eq reductions via efficiency upgrades) as high-impact composites. Water stands apart with fewer blue nodes, mostly peripheral and linked to monitoring, highlighting leak detection and usage patterns (liters) as a more niche but still a relevant use case with lower overall emphasis. The graph's density confirms stakeholder consensus on actionable, data-space-enabled indicators rather than exploratory ones.

With respect to consumers, all proposed functionalities were considered important by the respondents. However, automatic simulations and predictions stood out as the most demanded feature (Figure 7).

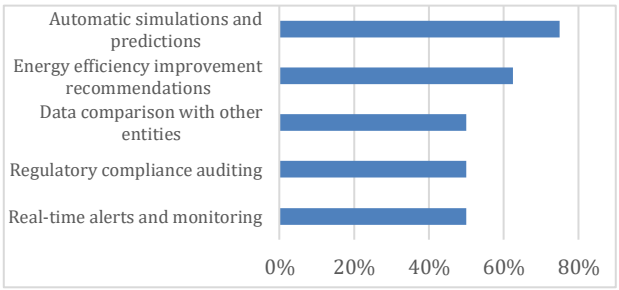


Figure 7 Identified useful domains by data-consumers on connecting to a Data Space (n=8 Companies).

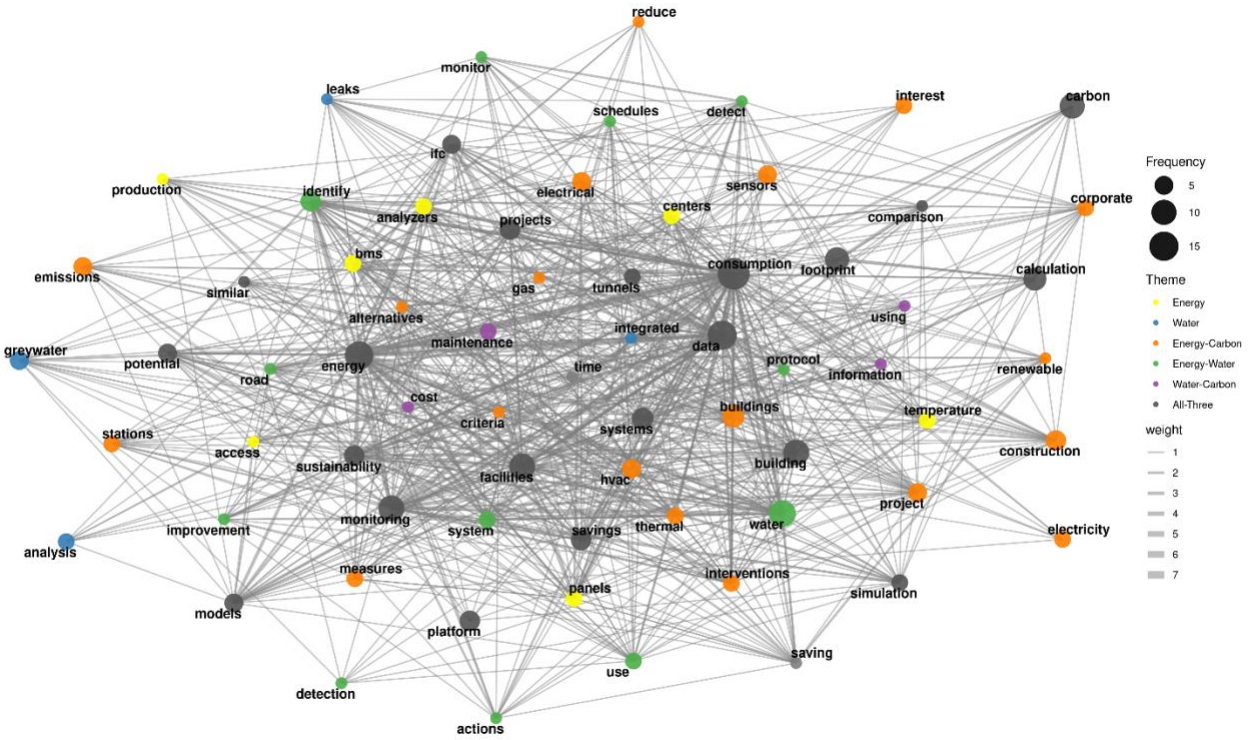


Figure 8 Co-occurrence network of the use cases requirements identified in the public administration workshops

Figure 9 shows that respondents mainly aim to use the Data Space for decision-making in design or planning, followed by developing digital products or services and only a few respondents are uncertain about which data to consume, indicating generally clear expectations.

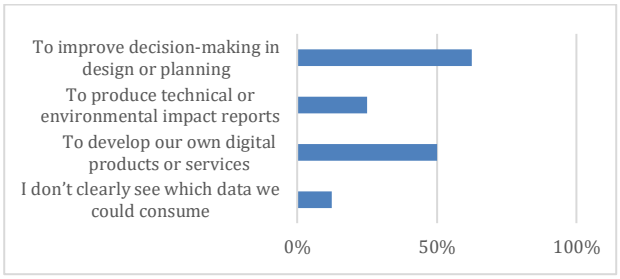


Figure 9 Identified data use by data-consumers on connecting to the Data Space (n=8 Companies).

The results show strong alignment between workshop insights and user expectations. Transversal data elements such as monitoring, comparison and predictions, form the core foundation across all use cases, highlighting the importance of standardized, accessible and interoperable data. Energy emerges as the main priority, closely linked to carbon, while water plays a more secondary role. From the consumer side, the proposed functionalities are well received, with a clear preference for simulations and predictive analytics. Overall, the findings confirm that the Data Space approach is relevant and valued, particularly for decision-making and advanced, actionable insights.

KPIs Definition

The KPI tables presented in this section are the result of the properties-definition process described in the preparatory stage of the methodology. The defined KPIs are directly linked to BIMxel properties that are systematically derived from IFC databases, external reference sources (such as Datadis for consumptions or AEMET for temperatures), and operational data obtained from IoT systems. Detailed KPI tables for each service are provided in Appendix 1[1].

The correlation network built from the workshop post-it notes showed that stakeholders perceive standardized access to BIM/IFC models, asset identification, geometry, space–system mapping and basic consumption metrics as prerequisites for any environmental analysis. The transversal data properties repeated across the KPI tables (IfcBuilding, IfcSpace, asset codes, consumption and anomaly properties) operationalize those clusters by turning the most frequent and central post-it notes topics into reusable, object-level data requirements shared by all three services. Based on the needs from the participants and the existing literature review the KPIs and their properties have been defined as shown in Appendix 1 (Table 1: Energy KPIs, Table 2: Water KPIs and Table 3: Carbon Footprint KPIs).

The correlation graph showed energy as the dominant cluster, the most interconnected, highlighting stakeholder priority for immediate value in monitoring, anomaly detection, and benchmarking. This aligns with

[1] Detailed KPI tables are provided in Appendix 1 and are available in: https://doi.org/10.5281/zenodo.19912259

questionnaire findings favoring AI-driven energy services (simulations, predictions, recommendations). Consequently, the energy KPI (Appendix 1: Table 1) emphasizes portfolio-level and pattern-oriented indicators (including total and seasonal consumption, normalized intensities, PV share, number, frequency and duration of anomalies, and seasonal HVAC efficiency) supported by the transversal data properties needed for pattern analysis (stable IDs, geometry, use, occupancy, meter–asset mapping), directly reflecting the central topics in the correlation network.

Regarding water, surveys and workshop notes pinpoint specific barriers such as uneven access to operational data across municipalities, high costs and time delays for data acquisition, and incomplete datasets lacking historical depth. Therefore, administrations prioritize short-term services like leak detection, consumption anomaly alerts, and basic usage pattern analysis using existing smart meters and sectorization data. This is reflected in the water KPI (Appendix 1: Table 2).which focuses on seasonal consumption, consumption by use type, irrigation and external uses, regenerated water ratio, anomaly counts and resolution times, normalized intensities, monitoring coverage and tank capacity, all grounded on a minimal but coherent set of BIMxel properties (space/use classification, meter linkage, reuse flags, etc.) mirroring the main nodes in the correlation graph.

For carbon footprint, the network shows strong energy-material linkages around renovation and lifecycle impacts, framing it as a composite of operational and embodied decisions rather than standalone. Workshops and surveys reinforce this through emphasis on regulatory compliance, portfolio-scale assessment, and urban context factors like density and street canyons. Accordingly, the carbon KPI (Appendix 1: Table 3). extends the energy and material KPIs to include total and operational CO2eq reductions, PV-related avoided emissions, urbanization interventions, embodied carbon by material, share of low-carbon and circular materials and mobility-related emissions, all parameterized with the same transversal building and consumption data plus additional indicators correspond to the green and transversal nodes in the network. For carbon footprint services, regulatory and taxonomy compliance checks (aligning with frameworks like EPBD, GENERCAT 2023-2027, EU Taxonomy, and DNSH criteria) did appear in stakeholder discussions but will require location-specific development. Thus, they are excluded from the main required properties, though they will be incorporated during implementation by mapping operational/embodied carbon KPIs, and metadata (such as YearOfConstruction, Phase, PrincipalUse) against local regulatory thresholds for automated eligibility screening and portfolio compliance reports.

The three KPI tables specify only those BIMxel properties that can be consistently exposed through a shared object-level layer (IFC entities, classification codes, environmental properties, governance attributes), making it possible to deliver the requested AI-enabled services (anomaly detection, benchmarking, predictive analysis, renovation-support KPIs) through a service-layer architecture where raw BIM and time-series data remain with providers. In this way, the KPIs emerge as a direct translation of the most central and recurrent patterns in the correlation network.

Services definition

Consequently, the AI-based services defined in this work focus on three main families, each leveraging the co-created KPIs and BIMxel properties to deliver actionable environmental analytics within the Data Space.

- Anomaly detection and monitoring services automatically identify abnormal energy and water consumption patterns at both asset and portfolio levels. These services process time-stamped meter data linked to BIM elements (via AssetManagementCode, IfcSpace mappings and Subdiscipline flags) to compute anomaly counts, mean durations and deviation scores against historical baselines or seasonal patterns, enabling early fault detection and prioritized maintenance interventions (Zhang et al., 2023).
- Benchmarking and portfolio comparison services normalize KPIs by floor area (GrossFloorArea), user occupancy or functional use across energy, water and carbon domains. By comparing normalized energy intensity (kWh/m²/year), water intensity (m³/m²/year) or total CO_2eq reductions between comparable public facilities, these services identify efficiency outliers, best practices and intervention priorities at scale (Martínez-Rocamora et al., 2021).
- Predictive performance analytics services forecast future consumption trends, efficiency degradation and carbon trajectories. These integrate KPI time series with BIM-derived context (climate via TerritorialScope/LatitudeLongitude, occupancy patterns, envelope U-values, material properties) and external baselines using time-series models or multivariate learning to simulate intervention impacts and risk profiles (Billanes et al., 2025; Cheng et al., 2020).

These AI-based services collectively address key gaps identified in the workshops (the need for streamlined access to standardized BIM/IFC data, reliable anomaly alerts, cross-asset benchmarking and actionable compliance insights) while respecting data-sovereignty principles through federated processing within the GreenBIMxels Data Space. By operationalizing stakeholder-defined KPIs via BIMxel properties, they enable public administrations to transition from fragmented, project-centric environmental monitoring to scalable, portfolio-wide decision support that aligns with European data-space architectures.

At the current stage of the project, GreenBIMxels is in the implementation phase, where the proposed approach is being developed and tested through pilot deployments. These pilots are being conducted using real-world datasets, BIM models, and operational information from

public buildings, primarily from the Universitat Politècnica de Catalunya and a collaborating municipality. This allows the validation of the approach under realistic conditions, including heterogeneous data availability and existing system constraints.

The three use cases are being implemented through two main development streams. Energy and water management services are being developed within a shared analytical environment, as both rely on similar data sources (such as consumption data from monitoring systems) and analytical logic, particularly anomaly detection and performance monitoring. In parallel, the carbon footprint use case is being developed in a separate tool, currently focused on material-based analysis and comparison, reflecting its different data requirements and methodological approach.

For all three use cases, an initial Minimum Viable Product (MVP) has been defined and is currently under development and testing. This MVP is based on the KPI framework and BIMxel properties defined in this study. However, at this stage, only a subset of the defined properties is being actively used, corresponding to the functionalities implemented in the MVP. The full set of KPIs and properties represents the target model to be progressively integrated as the implementation phase advances. The pilot implementations have already highlighted several practical issues. First, a substantial share of the required BIMxel property values is missing or incomplete in the source BIM models, confirming the relevance of the harmonized property set. Second, energy and water consumption data are currently distributed across heterogeneous platforms (including DEXMA, Datadis and other provider-specific systems), which has motivated the ongoing development of dedicated connectors to federate these sources within the Data Space. Third, temporal and semantic gaps in consumption records have emerged as recurrent, reinforcing the need for analytical services capable of handling incomplete, noisy or inconsistent data. Finally, to address interoperability challenges, a tool for the automated conversion of IFC parameters into BIMxel properties has been developed and integrated into the workflow, reducing manual effort and improving consistency in the preparation of KPI-ready datasets.

Limitations include the focused size and geographical scope of the empirical sample, limiting behavioral insights to the Catalonian context. The paper also concentrates on the exploratory and preparatory stages of data-space implementation, so future work should examine the long-term performance and maintenance effort of the proposed services in operational deployments, thereby improving and complementing the properties set in subsequent steps.

Conclusions

The study concludes that a BIM-based Data Space can effectively unlock underused environmental information from public facility portfolios, enabling administrations and service providers to implement AI-based services for energy, water and carbon management while maintaining data sovereignty. It demonstrates that co-created KPIs, encoded as BIMxel properties and systematically derived from IFC models, external databases and operational data, provide a robust semantic backbone for anomaly detection, benchmarking and predictive analytics. The three use cases prove to be strongly aligned with stakeholder priorities, with energy-related services emerging as the dominant entry point. Results from workshops and surveys indicate that both public administrations and data consumers possess sufficient digital maturity to engage with data-space-based solutions but require harmonized data structures, reduced integration friction and clear governance guarantee to translate this potential into operational value.

This paper confirms the dual-sided value proposition, public data providers in Catalonia gain scalable portfolio insights without data centralization, while engineering firms and sustainability consultants access validated KPIs and analytics foundations to build commercial services. By implementing a service-layer architecture that processes data locally and exposes aggregated indicators via harmonized APIs, GreenBIMxels unlocks environmental value from existing BIM models of public facilities and demonstrates stakeholder's interest in data-space adoption.

Acknowledgments

We thank the public administrations and private stakeholders for their participation, and the Spanish Ministry for Digital Transformation and the Civil Service for funding through the 2024 Call for Demonstrators and Use Cases of Sectoral Data Spaces.

References

Akbari, S., Sheikhkhoshkar, M., Pour Rahimian, F., El Haouzi, H. B., Najafi, M., & Talebi, S. (2024). Sustainability and building information modelling : Integration, research gaps, and future directions. *Automation in Construction*, *163*, 105420. https://doi.org/10.1016/j.autcon.2024.105420

Amancio, D. R., Machicao, J., & Quispe, L. V. C. (2025). Leveraging word embeddings to enhance co-occurrence networks : A statistical analysis. *PLOS ONE*, *20*(7), e0327421. https://doi.org/10.1371/journal.pone.0327421

Billanes, J. D., Ma, Z. G., & Jørgensen, B. N. (2025). Data-Driven Technologies for Energy Optimization in Smart Buildings : A Scoping Review. *Energies*, *18*(2), 290. https://doi.org/10.3390/en18020290

Cheng, J. C. P., Chen, W., Chen, K., & Wang, Q. (2020). Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. *Automation in Construction*, *112*, 103087. https://doi.org/10.1016/j.autcon.2020.103087

Enes, G. (2025). The Common European Data Space (s). A Governance Model "the European Way". *A Digital*

Europe for Citizens: Data Governance, Data Markets, Data Services, 35.

Erdogan, B. Z., & Baker, M. J. (2002). Increasing Mail Survey Response Rates from an Industrial Population : A Cost-effectiveness Analysis of Four Follow-up Techniques. *Industrial Marketing Management*, *31*(1), 65-73. https://doi.org/10.1016/S0019-8501(00)00117-6

Fauth, J., Strnadl, C. F., Brenner, M., Heitzhausen, H., & di Pascale, M. (2024). *A Digitalization Framework for Ecosystem-Based* Servitization *: The Use Case of the Building Permit Process* (SSRN Scholarly Paper No. 4827733). Social Science Research Network. https://doi.org/10.2139/ssrn.4827733

Fosnacht, K., Sarraf, S., Howe, E., & Peck, L. K. (2017). How Important are High Response Rates for College Surveys? *The Review of Higher Education*, *40*(2), 245-265.

González, J., Soares, C. A. P., Najjar, M., & Haddad, A. N. (2021). BIM and BEM Methodologies Integration in Energy-Efficient Buildings Using Experimental Design. *Buildings*, *11*(10), Article 10. https://doi.org/10.3390/buildings11100491

Gordo-Gregorio, P., Alavi, H., & Forcada, N. (2025). Decoding BIM Challenges in Facility Management Areas : A Stakeholders' Perspective. *Buildings*, *15*(5), 811. https://doi.org/10.3390/buildings15050811

Hamilton, A. B., & Finley, E. P. (2019). Qualitative methods in implementation research : An introduction. *Psychiatry Research*, *280*, 112516. https://doi.org/10.1016/j.psychres.2019.112516

Janev, V., Vidal, M.-E., Endris, K., & Pujic, D. (2021). Managing Knowledge in Energy Data Spaces. *Companion Proceedings of the Web Conference 2021*, 7-15. https://doi.org/10.1145/3442442.3453541

Kivits, R. A., & Furneaux, C. (2013). BIM : Enabling Sustainability and Asset Management through Knowledge Management. *The Scientific World Journal*, *2013*, 983721. https://doi.org/10.1155/2013/983721

Klumbyte, E., Georgali, P.-Z., Spudys, P., Giama, E., Morkunaite, L., Pupeikis, D., Jurelionis, A., & Fokaides, P. (2023). Enhancing whole building life cycle assessment through building information modelling : Principles and best practices. *Energy and Buildings*, *296*, 113401. https://doi.org/10.1016/j.enbuild.2023.113401

Kovach, A., Montalvillo, L., Lanza, J., Sotres, P., & Urbieta, A. (2026). Understanding Data Spaces : A Systematic Mapping Study of foundations, technical building blocks, and sectoral adoption. *Computer Science Review*, *59*, 100819. https://doi.org/10.1016/j.cosrev.2025.100819

Martínez-Rocamora, A., Rivera-Gómez, C., Galán-Marín, C., & Marrero, M. (2021). Environmental benchmarking of building typologies through BIM-based combinatorial case studies. *Automation in Construction*, *132*, 103980. https://doi.org/10.1016/j.autcon.2021.103980

Mayo-Wilson, E., Li, T., Fusco, N., Dickersin, K., & Investigators, for the M. (2018). Practical guidance for using multiple data sources in systematic reviews and meta-analyses (with examples from the MUDS study). *Research Synthesis Methods*, *9*(1), 2-12. https://doi.org/10.1002/jrsm.1277

Sameer, H., & Bringezu, S. (2021). Building information modelling application of material, water, and climate footprint analysis. *Building Research & Information*, *49*(6), 593-612. https://doi.org/10.1080/09613218.2020.1864266

Valk, O. M. C. van der, & Ryan, M. (2025). Data for the common good in the common European Data Space. *Data & Policy*, *7*, e32. https://doi.org/10.1017/dap.2025.5

Weiss, W., Russegger, S., Kozic, D., Cossa, A., Huber, A., Hofer, T., Draxler, S., Schützenhofer, E., & Schoiswohl, P. (2025). A Data Space for Accessible Multimodal Transport : Enabling Inclusive Mobility. *URBAN INNOVATION: TO BOLDLY GO WHERE NO CITIES HAVE GONE BEFORE. Medium Sized Cities and Towns as a Major Arena of Global Urbanisation. Proceedings of REAL CORP 2025, 30th Intl. Conference on Urban Development, Regional Planning and Information Society*, 53-64.

Zhang, Z., Chen, Y., Wang, H., Fu, Q., Chen, J., & Lu, Y. (2023). Anomaly detection method for building energy consumption in multivariate time series based on graph attention mechanism. *PLOS ONE*, *18*(6), e0286770. https://doi.org/10.1371/journal.pone.0286770

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

DEEP LEARNING–BASED POINT CLOUD SEGMENTATION AND RELATIONSHIP GRAPH CONSTRUCTION FOR HYDROPOWER DAM DIGITAL TWINS

Alwyn Mathew, Soheila Kookalani, and Ioannis Brilakis
University of Cambridge, England, United Kingdom

Abstract

This paper proposes a deep learning pipeline for generating semantic meshes and relationship graphs from intensity-only laser point clouds of hydropower plants. It segments point clouds into object-level instances using supervised models and derives spatial relationships, including proximity and containment, between components. These are refined into a structured graph, providing a lightweight representation for digital twin applications. Replacing rule-based methods with learning-based segmentation enables automated graph construction for dam infrastructure despite irregular geometries, limited features, and the lack of standardised schemas.

Introduction

Hydropower dams are long-lived, safety-critical assets that underpin electricity supply, water storage, flood control, and regional resilience. Their operation and maintenance increasingly depend on reliable "as-is" information about structural components as well as embedded mechanical and hydraulic equipment. Digital twin (DT) approaches have therefore gained traction to link physical assets to continuously usable digital representations that support inspection planning, performance assessment, anomaly detection, and scenario analysis. Recent large-scale evidence also shows that the DT concept is used inconsistently across the built environment, motivating clearer operational definitions and explicit data requirements for DT implementations as reported by Abdelrahman et al. (2025); Kookalani et al. (2026).

A practical DT for civil infrastructure is not only a geometric model, but an integrated information system that connects geometry, semantics, and relationships to enable querying, automation, and analytics. Construction DT research emphasizes that robust DT pipelines require systematic data capture, semantic structuring, and interoperability-aware information integration across lifecycle workflows as described by Tuhaise et al. (2023).

In this context, terrestrial laser scanning and photogrammetry-derived point clouds have become key modalities for capturing high-fidelity "as-is" geometry. However, translating raw point clouds into usable DT-ready representations remains challenging due to occlusions, clutter, irregular shapes, scale, and the need for detailed semantic labelling as outlined by Yue et al. (2024).

Most scan to building information modelling (BIM) and point-cloud-to-semantic-model workflows have been developed around buildings and industrial facilities, where component taxonomies are mature and standardized schema are relatively well supported as reported in Alavi et al. (2024). For dam infrastructure, the semantic breadth of elements and the heterogeneity between structural and electromechanical subsystems create gaps in standardization and representation. Commonly used exchange standards were originally designed for buildings and do not fully cover dam-specific components and hierarchies, which limits direct "BIM-first" relationship modelling for DT creation. This mismatch motivates DT representations that can be inferred directly from sensed data, especially when a complete parametric BIM is infeasible or uneconomical to create.

A second practical limitation is that many real-world infrastructure point clouds, especially open datasets, do not include rich appearance cues (RGB), and instead provide geometry plus intensity only. This reduces the discriminative signal available to segmentation algorithms and amplifies the need for methods that can learn robust geometric and reflectance patterns from limited features, while also handling the severe class imbalance typical in infrastructure scenes. Domain gaps between synthetic data and real scans further complicate generalization according to Hu et al. (2024).

Finally, DT utility depends heavily on relationships. Even if components are segmented accurately, downstream tasks such as navigation, hierarchical organization, adjacency reasoning, and structured asset inventories require an explicit relationship graph. Prior work in the built environment has shown the value of enriching geometric reconstructions into graph-structured semantic representations, enabling machine-readable links between objects beyond pure geometry as noted by Werbrouck et al. (2020).

To address these challenges, this paper proposes an end-to-end, learning-based framework for deriving digital twin–ready representations directly from intensity-only dam point clouds. The method is evaluated on the Platanovrisi Dam dataset, covering both exterior dam surfaces and interior hydropower plant spaces. The main contributions of this work are summarised as follows:

1. Learning-based segmentation for large-scale dam point clouds: We develop a deep learning pipeline



Figure 1: Platanovrisi dataset: HPP ground level (left) and turbine floor (right).

tailored for non-planar, complex dam surfaces that enables semantic labeling of intensity-only point clouds.

2. Semantic mesh generation and instance relationship extraction: The pipeline converts segmented point clouds into structured semantic meshes and identifies spatial relationships between individual components.
3. Relationship graph construction and connectivity modeling: Semantic predictions are transformed into a connected graph representation that captures adjacency and containment relationships among dam components.
4. End-to-end digital twin workflow from point cloud to UML (Unified Modeling Language): The connected graph can be exported into a generic format and used to generate UML class diagrams, providing a practical scaffold for digital twin modeling and downstream system integration.

Background

DTs in the built environment are commonly positioned as integrated digital representations that can be updated and used to support decision-making across the asset lifecycle. Large-scale bibliometric analyses highlight that the term digital twin is often used ambiguously in built-environment literature, but converges on core characteristics such as representational fidelity, purposeful synchronization and update mechanisms, and actionable use in analysis and operations as described by Abdelrahman et al. (2025). Construction-focused DT reviews in Tuhaise et al. (2023) further emphasize that DT value depends on connecting sensing, data integration, and analytics into an end-to-end workflow, rather than treating the DT as a standalone 3D model. For civil infrastructure, DT research is expanding beyond buildings into bridges, tunnels, and dams, where monitoring and risk assessment are central. For example, recent work in Tian et al. (2025) demonstrates DT approaches for high arch dams by fusing physics-based models with learning-based methods and uncertainty quantification to support real-time state estimation and failure risk assessment. These efforts underscore two recurring bottlenecks for infrastructure DTs: (1) obtaining up-to-date, spatially consistent as-is representations, and (2) structuring heterogeneous information into computable forms that support reasoning and automation. A key implication is that DT-ready representations must encode more than geometry. They must support asset semantics and connectivity or hierarchy, enabling queries such as "which elements belong to the turbine floor?", "what is adjacent to the intake tower?", or "which equipment is connected to which pipes?". Knowledge graph approaches in Tuhaise et al. (2023) explicitly target this integration challenge by using graph-based models to unify multi-source information for built assets, thereby improving cross-system interoperability and downstream analytics.

Laser scanning produces dense 3D point clouds that capture detailed surface geometry, making them attractive for modeling existing assets and enabling downstream DT workflows. However, scan-to-BIM pipelines typically require preprocessing, semantic segmentation and instance grouping, and object reconstruction and relationship modelling. Recent work of Patil and Kalantari (2025) emphasizes how segmentation accuracy can directly influence reconstruction outcomes and highlights practical hybrid workflows that combine deep learning segmentation with geometric fitting for downstream modelling. Across the Architecture, Engineering, and Construction (AEC) domain, deep learning has become the dominant approach for point cloud semantic segmentation because it reduces reliance on handcrafted geometric features and improves robustness across complex scenes as shown by Liu et al. (2025). A recent review in Yue et al. (2024) summarizes how point-cloud deep learning has been increasingly adopted for construction tasks, while also documenting persistent challenges: data scarcity, domain shift, scale variation, computational cost, and class imbalance, issues that are amplified in large infrastructure environments. Seminal architectures such as PointNet and PointNet++ as in Qi et al. (2017a) established the foundational concept of learning directly from unordered point sets, with hierarchical variants capturing local-to-global context for improved segmentation in complex geometries. In the construction and infrastructure domain, prior work reported in Yin et al. (2021) has demonstrated the effectiveness of deep learning-based segmentation for industrial scenes and transportation assets, including ResPointNet++ for industrial point clouds and hierarchical graph-based models for bridge component segmentation. Nevertheless, dam environments introduce additional complexity:

- Feature limitations: Many datasets are intensity-only (no color information), requiring models to separate classes using geometry and reflectance patterns rather than appearance cues.
- Heterogeneous scale and geometry: Dams mix large continuous surfaces with small mechanical compo-

nents and dense equipment layouts, increasing confusion between visually similar elements.

- Domain shift and annotation scarcity: Methods trained on synthetic BIM-derived point clouds often underperform on real scans due to the domain gap. Domain adaptation methods such as DawNet explicitly address this for BIM-to-scan segmentation Hu et al. (2024).
- Class imbalance: Infrastructure point clouds typically have extreme imbalance, motivating loss reweighting, focal losses, and sampling strategies, as also discussed in the proposed pipeline.

Even with accurate semantic segmentation, DT value is limited if the output remains an isolated labelled point cloud. Relationship modelling provides the scaffolding for higher-level reasoning, data integration, and automation. In conventional BIM workflows, relationships are typically defined through schema-specific hierarchies and predefined object types. However, infrastructure DTs often face schema gaps for dams. Standardized BIM schemas may not fully capture dam-specific components and their relationships, motivating alternative graph-first representations inferred directly from data. Two complementary directions in the literature motivate relationship graph construction: (1) Semantic enrichment / scan-to-graph: The scan-to-graph concept integrates scan-to-BIM ideas with semantic network by representing objects and their relationships as a graph, enabling richer queries and greater interoperability than geometry alone as presented in Werbrouck et al. (2020). This approach is particularly relevant when full parametric BIM creation is not feasible, but a computable representation is still required. (2) Connectivity-aware modelling from incomplete data: Infrastructure scans are commonly incomplete due to occlusions and access constraints. Methods that exploit connectivity relationships to improve as-built modelling from incomplete point clouds demonstrate the importance of explicitly reasoning about adjacency and inter-element connections, rather than treating components independently as shown by Kim and Kim (2021). Graph-based asset as outlined by Tuhaise et al. (2023) representations also align naturally with knowledge graph integration approaches for DTs, where nodes or edges support linking geometry-derived entities to metadata, sensor streams, inspection records, and analytical outputs. For dam DTs, this suggests a pragmatic pathway: infer lightweight spatial relations from point cloud instances, progressively enrich those graphs with domain semantics as schema mature and expert knowledge becomes available.

Methodology

Constructing a geometric model of a dam involves several key stages. The point cloud is first segmented into individual components, such as walls, gates, and turbines, breaking down the complex structure into manageable elements. Spatial proximity analysis is then applied to these components to infer their interactions, resulting in a relationship graph that captures spatial and semantic connections. Surface reconstruction techniques are subsequently employed on the segmented point clouds to generate continuous meshes, producing a detailed and coherent geometric representation of the dam.

Semantic mesh

The proposed segmentation network follows the PointNet architecture according to Qi et al. (2017a) for point-wise semantic labeling. It begins with an input transformation module (T-Net) that learns a 3×3 affine matrix to align the raw point cloud, thereby improving invariance to rigid geometric transformations. Feature extraction is performed through a sequence of shared multilayer perceptrons implemented as one-dimensional convolution layers, progressively expanding the channel dimension from the input to 64, 64, 64, 128, and finally 1024. A second transformation network is applied after the early convolution layers to learn a 64×64 feature alignment matrix, enabling canonicalization in the learned feature space. Global context is captured by max pooling across all points to produce a 1024-dimensional global feature vector. For semantic segmentation, the global feature is concatenated with per-point local features to form a fused representation for each point. This combined feature is processed by a segmentation head composed of additional shared multilayer perceptrons with channel dimensions 1088, 512, 256, and 128, followed by a final classification layer that outputs per-point semantic labels for the target number of classes.

Several modifications are introduced relative to the original PointNet to improve robustness and representational power. The network incorporates flexible normalization and dropout strategies, which enhance stability and reduce overfitting, and it can process additional input features beyond spatial coordinates, such as intensity or surface normals, enabling richer point representations. Transformation modules are refined to better align features across points while supporting these enhancements, ensuring more effective learning of both local and global geometric structures. Furthermore, a hierarchical variant inspired by PointNet++ as presented in Qi et al. (2017b) is included to capture local geometric patterns at multiple scales. This extension employs set abstraction and feature propagation modules to encode multi-scale contextual information, significantly improving the network's ability to model complex spatial structures in large-scale infrastructure environments. Strategies for handling class imbalance and tile-based oversampling further enhance segmentation of underrepresented classes, and these are described in detail in the Discussion section. The segmentation method has shown improved accurately as shows in Figure 2 and Table 1.

Surface reconstruction is essential for converting segmented point clouds into continuous, high-fidelity geometric models suitable for digital twin applications. In this study, surface normals computed during segmentation provide local geometric context, guiding the reconstruc-

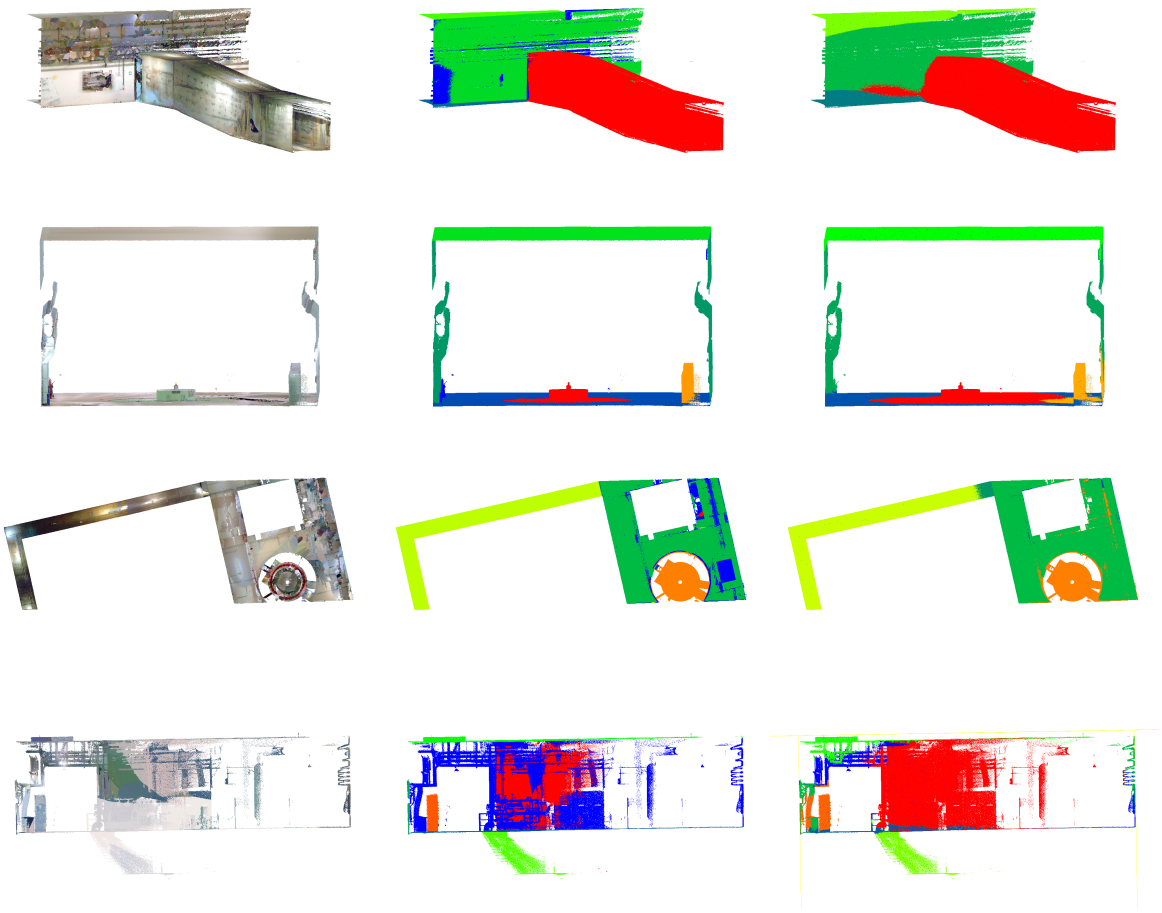


Figure 2: Platanovrisi dataset: color pointcloud (left), ground truth segmentation (middle) and predicted segmentation (right).

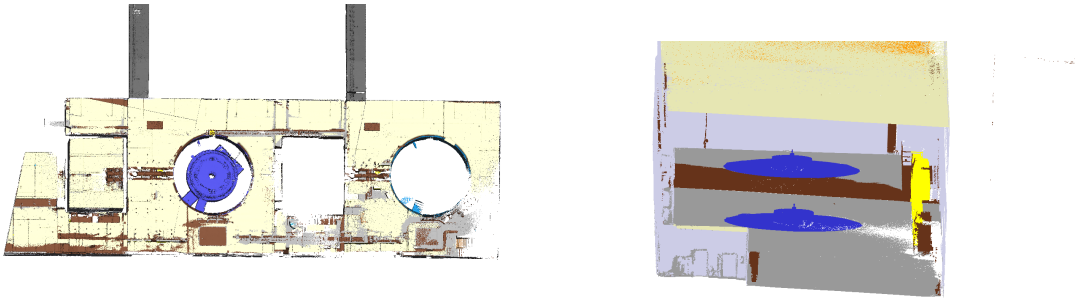


Figure 3: Semantic mesh generated from sematic pointcloud of the Platanovrisi dataset.

tion algorithm to accurately capture the orientation and topology of each component. We employ the Ball Pivoting Algorithm (BPA) Bernardini et al. (1999) to generate triangular meshes, where a virtual ball rolls and pivots across points to form triangles, adapting its radius to handle both smooth and intricate surfaces. This approach is particularly effective for hydropower plants, which combine curved, flat, and angular geometries such as turbine blades, pipes, and structural reinforcements. Challenges such as noise, outliers, occlusions, and irregular geometries are addressed through careful parameter tuning and post-processing techniques, including smoothing, hole filling, and mesh decimation. The resulting semantic meshes provide a coherent and accurate representation that serves as the foundation for visualization, simulation, and downstream tasks, integrating seamlessly with the segmentation and relationship graph pipeline to model complex hydropower infrastructure as shown in Figure 3.

Instance relationships

In traditional BIM workflows, relationship graphs are usually derived from well-defined object hierarchies and standardized schemas. However, for dam infrastructure, the creation of complete BIM models is limited because commonly used exchange formats, such as Industry Foundation

Classes (IFC), were originally developed around building and industrial facility taxonomies Alavi et al. (2024); Zhang et al. (2024) and do not yet provide comprehensive coverage of the heterogeneous structural and electromechanical component hierarchies typical of dam infrastructure.

To address this gap, this study proposes constructing a relationship graph directly from segmented point cloud data. Unlike conventional BIM graphs, this representation is lightweight and inferred, relying on spatial and semantic cues rather than predefined schemas. The graph is generated by establishing "connected" relationships based on spatial adjacency between segmented components. Each node represents a single component, while edges indicate spatial or semantic associations.

A k-dimensional tree (KD-Tree) is employed to efficiently index points and perform radius based searches within the point cloud. For each pair of segments, if any point from one segment lies within a radius of 0.5 m of a point in the other, an adjacency edge is created between their corresponding graph nodes. The KD-Tree is particularly effective in handling the geometric complexity of hydropower plants, which include irregularly shaped components such as turbines, generators, pipes, and reservoirs. Manual identification of such relationships is impractical due to this complexity, making automated inference highly valuable.

Beyond adjacency, components are also associated with higher-level structural abstractions, such as the dam body, ground floor, or turbine floor. If a component is spatially contained within one of these broader regions, a hierarchical "has" relationship is inferred (e.g., the ground floor has a pipe). This introduces an organized semantic structure, though it is limited to spatial and containment logic and does not capture functional or operational dependencies. Capturing these deeper relationships would require domain-specific schemas and expert knowledge, which are currently unavailable for dam infrastructure.

The basic relationships extracted from the point cloud are exported to a generic format, such as puml. This puml file serves as the input to PlantUML, to automatically generate a draft UML diagram. While the generated UML is not complete, it provides an initial framework that can be manually refined, extended, and annotated to include more detailed relationships. Figure 4 shows a preliminary example of the resulting graph, illustrating inferred connectivity and hierarchical groupings among segmented components.

Table 1: Segmentation evaluation result on test split of Platanovrisi dataset

Method	Precision	Recall	F1	IoU
RANSAC	0.5-0.6	0.4-0.5	2 0.5	0.40.4-0.6
Our	0.6813	0.9271	0.7393	0.6253

Dataset

The Platanovrisi Dam is a concrete facility located roughly 7.5 km southeast of the Thesauros Dam in northeastern Greece, has been operational since 1999 (Figure 1). It forms a reservoir capable of storing around 90 million cubic meters of water. Together with the 175-meter-tall earthen Thesauros Dam, among the tallest in Europe, Platanovrisi functions as part of a coordinated hydropower system that maximizes efficiency through water reuse and nighttime pumping between the reservoirs. The combined infrastructure supports a total generation capacity of 500 MW under a Greek-Bulgarian water management agreement. The dataset captures key structural elements of the dam, including the spillway, auxiliary spillway, downstream face, toe still basin, crest, and an intake tower. Interior scans encompass two levels of the hydropower plant: the ground-floor turbine gallery, providing maintenance access, and the lower turbine floor, which houses critical operational components such as the turbine well, piping, tanks, pumps, and control panels.

The dataset comprises approximately 369 million valid points distributed across 33 spatial tiles and annotated with 10 semantic classes: floor, wall, ceiling, stairs, crane, tunnels, electric distributor, turbine well, penstock, and water tank. The class distribution is severely imbalanced, as shown in Table 2, with wall and tunnels together accounting for over 55% of all points while minority classes such as crane and water tank each represent less than 1%. The dataset was partitioned into training, validation, and test splits using an 8:1:1 ratio at the tile level to prevent spatial leakage between splits.

Table 2: Distribution of dominant classes in the Platanovrisi dataset

Classes	#points (in M)	Percentage
Wall	105	28.8
Tunnels	98	27
Floor	73	20.2
Ceiling	49	13.4
Penstock	13	3.7
Electric distributor	9	2.6
Stairs	7	2
Turbine well	3	1.1
Crane	2	0.6
Watertank	2	0.6

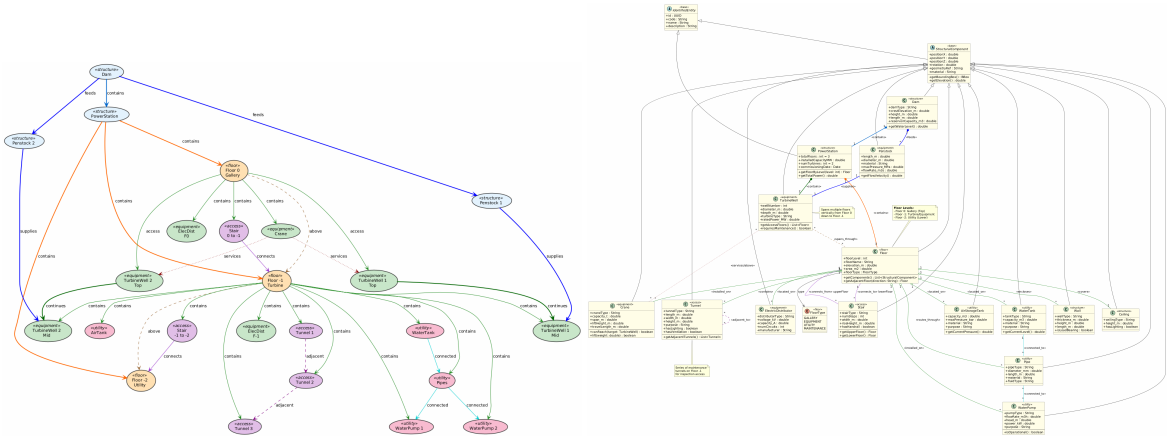


Figure 4: Relationship graph extracted from semantic output (left) and UML from relationship graph (right).

Discussion

The integration of semantic mesh generation and relationship graph construction relies critically on accurate deep learning–based segmentation, which forms the foundation of the proposed dam digital twin pipeline. Unlike conventional building structures, dams exhibit highly irregular geometries such as curved spillways, inclined faces, and embedded tunnels that challenge both traditional scan-to-BIM workflows and primitive-fitting approaches such as RANSAC as presented in Mathew et al. (2025) as seen in segmentation evaluation Table 1. Learning-based segmentation is therefore essential for capturing these complex, non-planar surfaces and heterogeneous mechanical components directly from point cloud data.

These geometric difficulties, together with scarce public datasets and inconsistent component definitions across projects, further complicate standardization and modeling. Additional obstacles include incomplete scans, missing hierarchical information, and limited prior knowledge of functional interactions between components. By leveraging deep neural networks to produce semantic predictions that drive surface reconstruction and relationship inference, the proposed pipeline generates compact yet structured representations that can support spatial queries and component-level data integration. This output provides an initial foundation for digital twin applications tailored to dam infrastructure, where the geometric and semantic complexity poses greater challenges than in typical building environments.

Semantic segmentation of large-scale infrastructure point clouds, however, often suffers from severe class imbalance as shown in Table 2. Structural components such as walls and tunnels dominate the point clouds, while smaller elements such as pipes, turbines, or auxiliary equipment contain far fewer points. To mitigate this issue, two complementary strategies were employed during training: loss reweighting and tile-based oversampling. Class loss balancing adjusts the contribution of each category to the training objective in proportion to its prevalence, penalizing misclassifications of underrepresented classes more heavily. Various weighting schemes can be applied, including balanced weighting based on inverse class frequency, simple inverse-frequency weighting, and normalized custom weights. In addition, focal loss is used to down-weight easy examples and focus optimization on hard-to-classify points, while label-smoothed negative log-likelihood further improves generalization by preventing overconfident predictions.

At the data level, tile-based oversampling ensures rare classes are sufficiently represented during training. Point clouds are divided into spatial tiles, and tiles containing rare classes are sampled more frequently based on the rarity of the least common component within the tile, with a maximum oversampling limit to prevent excessive duplication. This increases exposure to underrepresented geometries without artificially inflating individual points. Together, loss reweighting and tile oversampling act at complementary stages of the pipeline, aiming to improve segmentation accuracy for minority classes in complex hydropower environments. Training robust models for dam infrastructure remains challenging due to the limited availability of high-quality, publicly accessible datasets. The Platanovrisi dataset used in this study, while adhering to high-quality standards, still suffers from incomplete scans and inconsistent color mapping. As a result, the model was trained exclusively on laser intensity data, avoiding reliance on potentially unreliable color information to ensure robust generalization. A key limitation of the current study is that the pipeline has been evaluated on a single site (Platanovrisi Dam). While the architectural and geometric diversity within this dataset provides a meaningful testbed, the generalizability of the trained model to other dam structures with different typologies, materials, or scanning conditions has not yet been assessed. Evaluating transferability across multiple dam sites and identifying domain adaptation needs are planned for the follow up journal extension.

The segmentation model receives 4 input channels per point (XYZ coordinates and laser intensity) and was trained with a batch size of 5 using the Adam optimizer with a learning rate of 0.0008, weight decay of 1e-4, and

cosine annealing with a 5 epoch warmup to a minimum learning rate of 1e-6. Group normalization was used for stability with small batches, and gradient clipping with a maximum norm of 1.0 was applied. Training ran for up to 150 epochs with early stopping (patience of 15 epochs). To address the severe class imbalance, balanced inverse frequency weighting was combined with focal loss (alpha = 0.75, gamma = 2.0). As reported in Table 1, the model segments dominant structural categories such as floor, wall, and ceiling with reasonable accuracy. However, minority classes including crane, turbine well, and water tank remain challenging due to their extremely low representation in the dataset (each below 1% of total points, as shown in Table 2). This reflects the inherent difficulty of dam point cloud segmentation compared to building environments, where class distributions tend to be more balanced and component geometries more standardized.

The quality of the relationship graph depends directly on the accuracy of the upstream segmentation. In the proposed pipeline, adjacency (''connected") edges are inferred using a KD-Tree radius search: for each segment pair, if any point from one segment falls within a radius of 0.5 m of a point in the other, an adjacency edge is created between their corresponding nodes. Containment (''has") relationships are assigned when a component lies spatially within a higher level region such as the ground floor or turbine floor. Segmentation errors, particularly oversegmentation of large surfaces such as walls or confusion between adjacent classes, can introduce spurious edges or miss valid connections. For instance, if a wall segment is split into multiple fragments, each fragment generates its own node and edges, inflating the graph with redundant adjacency links. Conversely, if a small component such as a pipe is absorbed into a neighbouring wall segment, the corresponding node and its relationships are lost entirely. Since no ground truth relationship graph currently exists for the Platanovrisi dataset, quantitative evaluation of graph quality (e.g., edge level precision and recall) is not feasible at this stage. Constructing such a reference through expert annotation and conducting sensitivity analysis of the distance threshold are planned for the follow up journal paper.

Conclusion

This paper presented an end-to-end pipeline for constructing relationship graphs from intensity-only laser point clouds of hydropower dams by integrating deep learning based segmentation, surface mesh reconstruction, and instance-level relationship inference. Evaluated on the Platanovrisi Dam dataset across exterior structures and interior hydropower plant spaces, the framework demonstrates the feasibility of producing structured geometric and relational models directly from laser scans without relying on complete parametric BIM schemas. Dam infrastructure poses greater challenges than typical building environments due to irregular geometries, heterogeneous component types, the absence of standardized schemas, and severe class imbalance in point cloud data. Remaining challenges include limited training data, poor performance on minority classes, the lack of a ground truth relationship graph for quantitative evaluation, and the current restriction to spatial relationships. Future work includes multi-site evaluation, construction of an expert annotated reference graph, improved handling of class imbalance, and richer domain informed semantics toward scalable digital twins for hydropower infrastructure.

Acknowledgments

This work is supported by the D-HYDROFLEX project under the European Union's Horizon Programme (Grant Agreement No. 101122357), BUILDSPACE project under the Horizon Europe (Grant Agreement No. 101082575) and M-TABS under Innovate UK (Grant Agreement No. 10132366).

References

Abdelrahman, M., Macatulad, E., Lei, B., Quintana, M., Miller, C., and Biljecki, F. (2025). What is a digital twin anyway? deriving the definition for the built environment from over 15,000 scientific publications. Building and Environment, 274:112748.

Alavi, H., Kookalani, S., Rahimian, F., and Forcada, N. (2024). Integrated Building Intelligence. Springer.

Bernardini, F., Mittleman, J., Rushmeier, H., Silva, C., and Taubin, G. (1999). The ball-pivoting algorithm for surface reconstruction. IEEE Transactions on Visualization and Computer Graphics, 5(4):349–359.

Hu, D., Gan, V. J., and Zhai, R. (2024). Automated bim-to-scan point cloud semantic segmentation using a domain adaptation network with hybrid attention and whitening (dawnet). Automation in Construction, 164:105473.

Kim, H. and Kim, C. (2021). 3d as-built modeling from incomplete point clouds using connectivity relations. Automation in Construction, 130:103855.

Kookalani, S., Green, S., Luo, P., Alavi, H., Parn, E., Sun, Z., and Brilakis, I. (2026). Mapping digital twin applications in infrastructure and the built environment across research types, methods, sectors, phases, and scales. Automation in Construction, 182:106778.

Liu, H., Su, H., Kookalani, S., Cao, M., Lei, Y., and Wu, W. (2025). Attention-enhanced spatiotemporal deep learning-based automatic warning model with uncertainty estimation in dam safety. Structural Health Monitoring, page 14759217251358537.

Mathew, A., Alavi, H., Wang, G., Su, Y., Che, W., and Brilakis, I. (2025). Geometric model of hydropower

plants: a foundational step towards creating a hydropower digital twin. In EC3 Conference 2025, volume 6, pages 0–0. European Council on Computing in Construction.

Patil, J. and Kalantari, M. (2025). Automatic scan-to-bim—the impact of semantic segmentation accuracy. Buildings, 15(7):1126.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017a). Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 652–660.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017b). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. arXiv preprint arXiv:1706.02413.

Tian, J., Chen, C., Zhang, L., Chen, J., Huang, H., and Zhang, P. (2025). A digital twin method for real-time analysis of structural deformation and failure for high arch dams. Engineering Structures, 345:121426.

Tuhaise, V. V., Tah, J. H. M., and Abanda, F. H. (2023). Technologies for digital twin applications in construction. Automation in Construction, 152:104931.

Werbrouck, J., Pauwels, P., Bonduel, M., Beetz, J., and Bekers, W. (2020). Scan-to-graph: Semantic enrichment of existing building geometry. Automation in Construction, 119:103286.

Yin, C., Wang, B., Gan, V. J., Wang, M., and Cheng, J. C. (2021). Automated semantic segmentation of industrial point clouds using respointnet++. Automation in Construction, 130:103874.

Yue, H., Wang, Q., Zhao, H., Zeng, N., and Tan, Y. (2024). Deep learning applications for point clouds in the construction industry. Automation in Construction, 168:105769.

Zhang, S., Zhang, S., Wang, C., Zhu, G., Liu, H., and Wang, X. (2024). Extended ifc-based information exchange for construction management of roller-compacted concrete dam. Automation in Construction, 163:105427.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DATA QUALITY CONTROL FOR OPENBIM-BASED DIGITAL TWINS IN RETROFIT: A CASE STUDY ON IFC VALIDATION AND INFORMATION REQUIREMENTS

Sarah Mahbod[1], Atefeh Hesarbani[1], Erik Poirier[2], Farzad Jalaei[3], Arash Hosseini[3], and Mazdak Nik-Bakht[1]
[1]Concordia University, Montréal, Canada
[2]École de technologie supérieure, Montréal, Canada
[3]National Research Council Canada, Ottawa, Canada

Abstract

Building Information Modelling provides the geometric and semantic foundation for Digital Twins (DT), but its effectiveness depends on the availability of use-case-specific information. In OpenBIM workflows, Industry Foundation Classes (IFC) models often contain missing or incomplete information, limiting reliable integration with DT platforms. This paper presents an envelope retrofit-oriented data quality control using Information Delivery Specifications (IDS). Minimum requirements were formalized into IDS and validated against an IFC model. Results show that while geometry was generally complete, critical semantics, particularly material and thermal properties, were frequently missing. IDS-based validation enables early detection of data gaps and improves IFC data readiness.

Introduction

The built environment is a major contributor to global energy use and CO_2 emissions (International Energy Agency, 2023), making existing building stock an important target for retrofit interventions aimed at reducing energy use, lowering emissions, and extending operational lifespans.

Digital technologies, particularly Building Information Modelling (BIM) and Digital Twin (DT), offer significant potential to support and optimize retrofit processes (D'Angelo et al., 2022). However, the effectiveness of these technologies is often constrained by persistent interoperability and data exchange challenges.

Building retrofitting aims to enhance energy efficiency (Ma et al., 2025), occupant comfort (Hosamo et al., 2023a), and environmental performance. Traditional retrofit approaches rely primarily on static building properties, historical weather data, and simplified energy models (Feng et al., 2022). Such approaches often fail to capture the true complexity of existing buildings, leading to suboptimal design decisions, unforeseen implementation challenges, and performance outcomes that deviate from predictions, commonly referred to as the energy performance gap (Cuerda et al., 2020).

DTs offer a promising alternative by enabling data-driven retrofit decision-making through continuous monitoring and real-time performance feedback (Ba et al., 2025). By integrating live data streams, DTs can support optimized energy use and improved occupant comfort (Hosamo et al., 2023b). A DT can be defined as a digital replica of a physical asset that integrates data from Internet of Things (IoT) devices with analytics and computational models to represent the state and behavior of the asset across its lifecycle (Agostinelli et al., 2021). DT ecosystems typically consist of a physical asset, its virtual counterpart, and a bidirectional flow of information between them (Hadjidemetriou et al., 2023). Within this context, BIM provides the geometric and semantic foundation of the virtual model, while IoT technologies enable dynamic data acquisition, making BIM and IoT core enablers for DT development in the built environment (Li et al., 2023). BIM models represent the physical and functional characteristics of buildings through both geometric and non-geometric information, providing a structured baseline for simulation, analysis, and decision support (Azhar et al., 2020).

Despite this conceptual synergy, integrating BIM models into DT workflows remains technically challenging due to interoperability limitations (Muchla et al., 2025). Interoperability refers to the ability to exchange and reuse data across heterogeneous software platforms throughout the building lifecycle while minimizing information loss and manual intervention (Agostinelli et al., 2021). In practice, BIM platforms, energy simulation tools, IoT systems, and DT environments often rely on proprietary data structures and inconsistent schemas (Piras et al., 2025). During BIM data exchange, critical semantic information such as thermal properties, material definitions, and equipment specifications may be omitted, misclassified, or improperly mapped due to incomplete information requirements, variations in modelling practices, and IFC export configurations, particularly in workflows using open standards such as IFC and gbXML (Bastos Porsani et al., 2021; Li et al., 2023). These issues hinder the automated linkage of BIM-based static data with dynamic data sources, such as Building Automation Systems (BAS), Building Management Systems (BMS), and IoT sensors, resulting in manual data correction, extended workflows, and unreliable analytics.

While interoperability challenges in BIM–DT integration are well recognized, existing studies predominantly focus on conceptual frameworks or platform-specific implementations. Limited attention has been given to the explicit definition and automated validation of retrofit-oriented information requirements in IFC models prior to DT deployment. As a result, data quality issues are often

identified late in the workflow, reducing scalability and limiting the effectiveness of DT applications for retrofit projects. Accordingly, this study is guided by the following research question: How can minimum, use-case-specific information requirements for envelope retrofit be explicitly defined and automatically validated in IFC models to support DT integration? This paper addresses this gap by proposing a structured approach for defining and validating minimum, use-case-specific information requirements and demonstrates how IDS can be used not only as a checking tool but as a mechanism for improving upstream data definition in envelope retrofit-oriented workflows. The main contributions of this study are:
(1) the definition of a minimum, use-case-specific set of envelope retrofit information requirements, (2) their formalization into machine-interpretable IDS specifications, and
(3) the demonstration of an automated validation workflow to assess IFC data readiness for DT integration.

In this research, the term retrofit refers specifically to building envelope retrofit, focusing on elements that govern heat exchange with the external environment. The proposed approach is demonstrated through a case study on an IFC model. The objective is to improve IFC data readiness and support scalable, openBIM-compliant DT applications for retrofit projects.

Methodology

This study applies a case-study approach to assess IFC data readiness for envelope retrofit-oriented DT integration through requirement definition and IDS-based validation. The methodological workflow consists of five main steps: (1) identification of retrofit-oriented information requirements, (2) formalization of these requirements using IDS, (3) preparation and export of the BIM model to IFC format, (4) validation of IFC models against the defined IDS rules, and (5) analysis of validation results to identify data quality issues.

Figure illustrates the validation workflow used in this study. This workflow represents one stage within a broader process that begins with use-case-based requirement definition, followed by validation (this research), and subsequent model refinement and DT integration in future work.

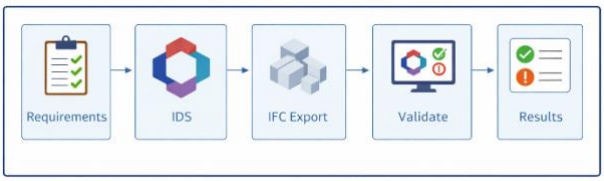


Figure 1: IDS-based validation workflow

The methodology defines information requirements rather than as a general measure of BIM completeness. For the use-case addressed in this study, the minimum set was limited to the static envelope information needed for initial retrofit assessment and subsequent DT use. Requirement selection was based on the information needed for envelope performance evaluation. Although additional information may be available in external sources such as product and material databases or environmental product declarations, these were considered outside the scope of this study. The scope of the IDS was therefore limited to information that could be reliably represented in IFC and automatically validated.

Retrofit-oriented information requirements were identified based on the literature insights, common inputs required for building energy analysis, and domain knowledge relevant to envelope retrofit. The scope was intentionally limited to external envelope elements, including walls, roofs, slabs, windows, and doors, which are central to envelope retrofit assessment. The resulting requirements focus on the presence and correct assignment of envelope-related information, including element identification, material assignment, and the availability of thermal performance properties. Together, these requirements represent a core but non-exhaustive subset of information relevant to the selected use-case. The identified requirements were formalized using buildingSMART IDS (buildingSMART International, 2024), enabling machine-interpretable definition of expected IFC entities, properties, and quantities. In this study, IDS rules were designed to verify information availability and completeness without prescribing how elements must be modelled or requiring specific property values, allowing flexibility across BIM authoring practices. A prototype validation workflow was implemented by exporting a BIM model of an existing building to IFC format and validating it using an IDS-compatible tool.

Case Study and Validation Workflow

Case Study Description

The case study is based on the Future Buildings Laboratory (FBL), a research facility located at Concordia University in Montreal, Quebec. The FBL functions as a living laboratory supporting multidisciplinary research and education in sustainable buildings and infrastructure. The facility provides a controlled environment for testing building technologies and monitoring performance, making it a suitable case study for investigating data readiness in retrofit-oriented DT workflows.

The building is a single-storey structure with approximate plan dimensions of 14.7 m × 7.5 m and an overall height of about 8 m. From a thermal perspective, the facility operates primarily as a single thermal zone; however, it is internally subdivided into four dedicated test cells used for independent experimental studies, along with two mechanical cells. The building is equipped with a compact set of building systems, including an outdoor heat pump unit paired with an indoor air handling unit for heating and cooling, as well as an electric domestic hot water system. These characteristics make the FBL representative of small-scale experimental buildings commonly used in retrofit research and DT development.

BIM Modelling and IFC Export

A three-dimensional BIM model of the FBL was developed using Autodesk Revit to represent the actual building geometry and associated static building information. The model focuses on architectural and envelope components required for retrofit assessment and DT initialization. The BIM model was subsequently exported to the IFC format using IFC 4 and the Reference view MVD to enable open and tool-independent data exchange.

The IFC export was performed without manual post-processing or correction before validation. This approach ensures that the validation results reflect the completeness and correct assignment of information as produced by the BIM authoring and export process, rather than improvements introduced through manual intervention. The exported IFC model served as the sole input for IDS-based validation.

IDS Definition and Scope

To develop the IDS, usBIM.IDSeditor (ACCA software, 2026) was used due to its native support for the buildingSMART IDS standard and its schema-aware environment for defining requirements without manual XML editing. The tool was selected because it is dedicated to requirement authoring rather than model checking, is independent of any specific IFC model, and produces portable IDS definitions accompanied by human-readable documentation. This separation allows IDS validation to be performed independently using an IDS-compatible tool.

Retrofit-Oriented Information Requirements

This section details how retrofit-oriented requirements were selected. The requirements target static building information and should therefore be understood as a minimum information specification for this use-case, rather than as a complete definition of all information that could be relevant to building lifecycle analysis.

These requirements were derived from the information typically needed for envelope performance evaluation. In particular, the selection reflects inputs commonly required by building energy simulation workflows, including geometry and zoning context, material and construction properties, and thermal performance parameters.

The IDS requirements were defined as a core subset that is both (i) retrofit-relevant and (ii) practically checkable in IFC exports across tools. For opaque elements (external walls, roofs, and slabs), requirements emphasize element identification, material assignment, and thermal transmittance properties. For openings (windows and doors), the focus is on thermal transmittance, with quantity information treated as optional.

In the IDS structure, each requirement is defined within a Specification that includes an Applicability definition and a set of Requirements, enabling the translation of these information needs into explicit, machine-interpretable validation rules.

Five envelope element categories were selected as the validation scope of the IDS, and requirements were defined for applicable instances of these categories using standard IFC entity types, property sets, and quantity take-off sets. The detailed requirements are summarized in Table 1.

Table 1: Summary of IDS requirements for envelope elements

Element Type	IFC Entity	Properties
External Walls	IfcWall	Name, IsExternal, LoadBearing, ThermalTransmittance, NetSideArea, NetVolume, Material
Roofs	IfcRoof	IsExternal, ThermalTransmittance, GrossArea, Material
Slabs/Floors	IfcSlab	IsExternal, ThermalTransmittance, Material
Windows	IfcWindow	IsExternal, U-value
Doors	fcDoor	IsExternal, ThermalTransmittance

For external walls (Figure), the IDS defines the most extensive set of checks, reflecting the central role of walls in envelope heat transfer. The specification includes identification and applicability conditions to define which elements are evaluated, together with requirements for thermal, material, and quantity-related properties; however, classification and fire rating remain optional to avoid over-constraining IFC exports.

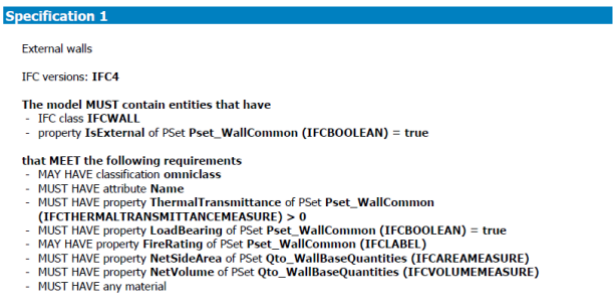


Figure 2: External walls specification

For roofs, the IDS specifies external applicability together with thermal, material, and gross area information. For

slabs and floors, the IDS includes external applicability, material assignment, and thermal transmittance properties, without additional quantity requirements.

For windows, the IDS requires external applicability and thermal transmittance, while area-related quantities are optional. For doors, the external applicability and thermal transmittance were included, without additional quantity requirements.

Requirements related to dynamic operation (e.g., real-time sensor data, control setpoints, or occupancy schedules), performance targets (e.g., prescribed U-value thresholds or regulatory compliance limits), proprietary or tool-specific parameters (e.g., authoring-tool custom fields), detailed construction layering (e.g., full multi-layer wall build-ups with material thicknesses), and non-retrofit lifecycle data (e.g., construction sequencing or maintenance scheduling) were excluded, as they cannot be consistently expressed, validated, or verified as static information within IFC-based IDS workflows.

In addition, some information that is typically maintained in external sources rather than within the IFC model was excluded. This includes manufacturer-specific product data, detailed material properties, and environmental declaration information such as embodied carbon indicators. While such information may be important for broader sustainability or lifecycle analyses, it falls outside the scope of this study, which focuses on validating the minimum IFC-based information required for retrofit-oriented DT initialization. The resulting set supports early, automated IFC data quality assessment for retrofit-oriented DT integration without defining a comprehensive modelling specification.

Validation Workflow

The validation workflow evaluates IFC data readiness through automated, standards-based checks. The IFC4 model was validated against the defined IDS specifications using BIMcollabZoom (BIMcollab, 2026.), as shown in Figure , which assessed each applicable element and generated element-level pass/fail results. For non-compliant elements, the output included requirement-specific messages indicating missing or incomplete information, such as absent thermal properties, missing material assignments, or missing quantity data.

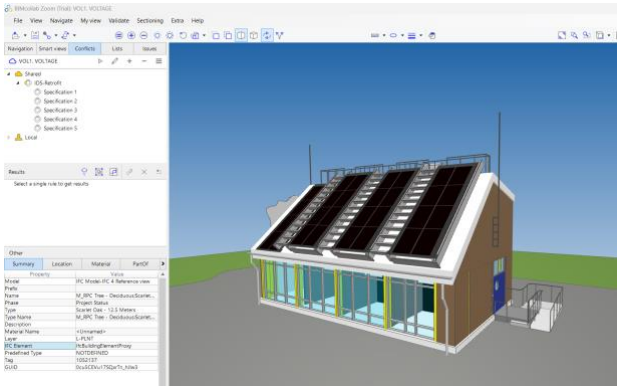


Figure 3: Case study 3D model in BIMcollab

The validation results were then analyzed to identify recurring data gaps across envelope elements and to assess overall IFC data readiness. Non-compliant elements can be traced and corrected in the BIM authoring environment, followed by re-export and re-validation; however, this corrective loop is outside the scope of this study.

Result Analysis and Discussion

Results Analysis

The validation assessed compliance with the retrofit-oriented information requirements defined for envelope elements and showed that a substantial number of applicable elements failed one or more IDS checks. In addition to the pass/fail indicators, the BIMcollabZoom provided detailed Property–Value information for each non-compliant element. These fields are essential for interpreting why elements failed IDS requirements and for distinguishing between geometric completeness and semantic deficiencies. In several cases, elements included correct geometry, dimensions, and basic identifiers, while required attributes such as material assignments, thermal transmittance values, or external applicability flags were undefined or absent. These failures were primarily associated with missing thermal properties and material assignments. This indicates that data gaps are concentrated in the semantic information required for envelope retrofit assessment rather than in geometric representation.

For walls (Figure), the validation output shows repeated failures associated with undefined material assignments and missing thermal transmittance properties, despite the presence of geometric representations and basic element identifiers. It should be noted that the figure primarily shows internal walls. Their inclusion in the validation results indicates that the external applicability condition (e.g., IsExternal) was missing or not correctly assigned, causing these elements to be evaluated against IDS requirements intended for external walls. This highlights the importance of correctly defining applicability attributes in addition to providing required semantic information.

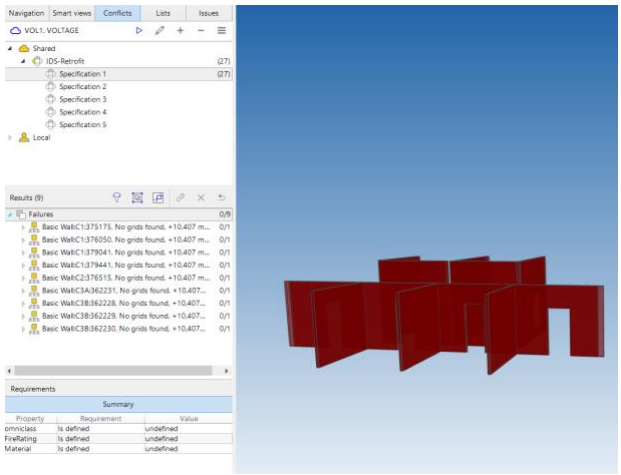


Figure 4: Validation of External Walls

For slabs and floors (Figure), although geometries were present and correctly represented as IfcSlab entities, the absence of material assignments resulted in systematic non-compliance.

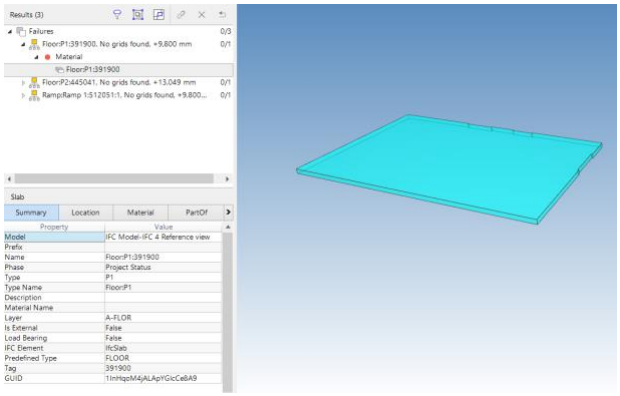


Figure 5: Validation of Slabs/Floors

For windows (Figure), the validation results show that while window geometries and basic dimensional attributes were available, the absence of thermal transmittance properties, particularly U-values prevented compliance with the retrofit-oriented requirements. This suggests that thermal properties were either not defined at the authoring stage or not reliably transferred during IFC export

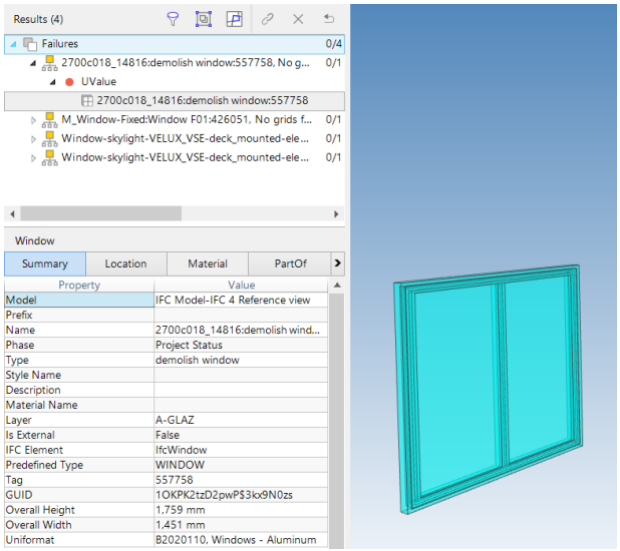


Figure 6: Validation of Windows

For doors (Figure), validation failures were associated with missing thermal transmittance properties and missing or incorrect external applicability information. Even when doors were correctly classified as external and included additional attributes such as fire rating, the lack of thermal performance information resulted in non-compliance with the IDS requirements.

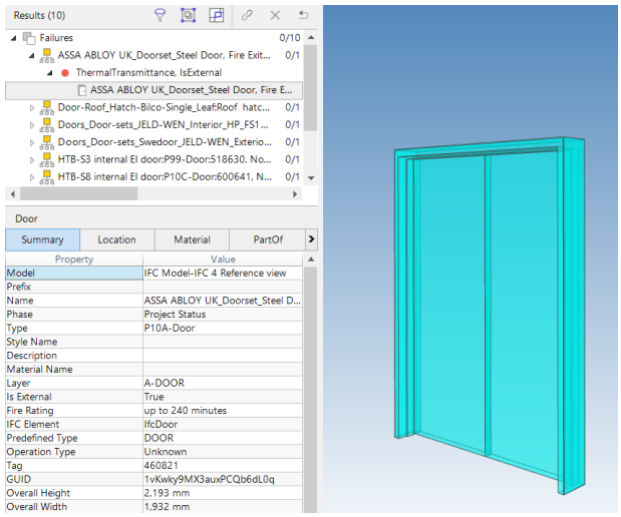


Figure 7: Validation of Doors

Discussion

The validation outcomes highlight a fundamental challenge in using IFC models as a foundation for retrofit-oriented DT workflows: geometric completeness alone does not guarantee data readiness. Although the BIM model provided visually complete representations of envelope elements, the frequent absence of material and thermal properties significantly limits its suitability for retrofit assessment and DT initialization. The findings should be interpreted within the scope of envelope retrofit, as the requirement set does not address system-level or operational retrofit interventions.

The results indicate that IFC data readiness is not solely a technical issue, but also a methodological one related to how information requirements are defined and communicated prior to modelling. In the absence of explicit, use-case-driven requirements, critical attributes such as thermal transmittance and material definitions remain optional, leading to systematic data gaps across the model. This highlights the importance of linking requirement definition to model authoring and export processes rather than relying on implicit assumptions about data availability.

The study also clarifies the role of IDS within the broader process of preparing BIM data for DT use. In this work, IDS was used to validate the minimum set of IFC-embedded information required for the selected use-case. This does not imply that all relevant information must reside within the IFC model. In practice, richer data such as manufacturer-specific product information, material databases, or environmental product declarations may be integrated after the baseline IFC model has been validated. IDS-based validation should therefore be understood as an early quality-control step within a broader workflow that includes model preparation, enrichment, and DT integration. The IDS-based validation approach proved effective in making data deficiencies explicit and traceable. By encoding requirements in a machine-interpretable form, the validation process clearly distinguished between

geometric completeness and semantic readiness, identifying missing or incomplete information at the element level. This supports early detection of data gaps before DT integration, where manual correction would otherwise be required.

Despite its advantages, IDS-based validation has several limitations in the context of envelope retrofit-oriented DT workflows. IDS supports verification of the presence and correct assignment of information, but does not assess the accuracy or physical validity of property values. In addition, IDS operates on static IFC data and does not capture dynamic information such as sensor data or operational conditions. Its effectiveness also depends on how models are authored and exported, meaning that variations in modelling practices and export configurations can influence validation outcomes. IDS should therefore be considered a necessary but not sufficient mechanism for ensuring data readiness.

Overall, the results demonstrate that explicitly defined retrofit-oriented information requirements, when validated through IDS, improve transparency regarding IFC data readiness. This enables earlier and more reliable data quality control and provides a practical mechanism for strengthening interoperability between BIM models and DT workflows.

Beyond demonstrating current IDS usage, the study contributes a structured approach for defining and validating minimum information requirements tailored to a specific envelope retrofit use-case.

Conclusion

This study presented an envelope retrofit-oriented approach to assessing IFC data readiness for DT integration using Information Delivery Specifications (IDS). By explicitly defining use-case-specific information requirements and validating them through an automated, standards-based workflow, the study demonstrated how IDS can be used to identify missing and incomplete information in IFC models before DT deployment.

The case study showed that, while the IFC model provided complete geometric representations of envelope elements, essential semantic information, particularly material definitions and thermal performance properties, was frequently absent or only partially defined. Validation results highlight the limitations of relying on visual completeness alone when evaluating IFC models for retrofit-oriented DT workflows.

The findings further indicate that IDS-based validation enables early and repeatable data quality control in a tool-independent manner. They also show that the primary barrier to IFC data readiness is not the data format itself, but the lack of explicitly defined information requirements. By formalizing these requirements in a machine-interpretable form, IDS supports systematic detection of data gaps and facilitates targeted model improvement, reducing the need for manual data correction during later stages of DT integration. This approach provides a practical basis for improving IFC data readiness and supporting DT integration.

This study has limitations: the requirement set focuses on a core subset of retrofit-relevant information rather than a comprehensive specification, and validation was conducted on a single case-study building. While the workflow supports iterative correction of detected issues through model updates and re-validation, the implementation of model corrections and subsequent validation cycles is outside the scope of this study, which focuses on requirement definition and initial data quality assessment. Nevertheless, the proposed workflow is transferable and provides a solid foundation for extending IDS-based validation to broader retrofit scenarios. Beyond demonstrating a practical application, the study provides analytical insights into common patterns of information deficiency in IFC models and highlights the role of explicit requirement definition in improving data readiness.

Future work will expand the requirement scope to additional building systems, integrate semantic linking through the buildingSMART Data Dictionary(bSDD), and explore the combined use of IDS and Information Delivery Manuals (IDM) to better define roles and responsibilities in retrofit-oriented DT workflows.

Acknowledgments

The authors thank Concordia University for access to the Future Buildings Laboratory used as the case study in this research. The authors acknowledge the use of AI-assisted tools to support the drafting and refinement of this manuscript.

References

ACCA software. (2026). usBIM.IDSeditor (Version 2.3.20.0) [Computer software]. https://www.usbim.com

Agostinelli, S., Cumo, F., Guidi, G., & Tomazzoli, C. (2021). Cyber-Physical Systems Improving Building Energy Management: Digital Twin and Artificial Intelligence. Energies, 14(8). https://doi.org/10.3390/en14082338

Azhar, S., Khalfan, M., & Maqsood, T. (2020). Building Information Modeling (BIM): Now and beyond. The Australasian Journal of Construction Economics and Building, 12(4), 15–28. https://doi.org/10.3316/informit.013120167780649

Ba, L., Tangour, F., Abbassi, I. E., & Absi, R. (2025). Analysis of Digital Twin Applications in Energy Efficiency: A Systematic Review. Sustainability, 17(8). https://doi.org/10.3390/su17083560

Bastos Porsani, G. B., de Lersundi, K. D. V., Sánchez-Ostiz Gutiérrez, A. S.-O., & Fernández-Bandera, C. F. (2021). Interoperability between building information modelling (Bim) and building energy model (bem). Applied Sciences (Switzerland), 11(5), 1–20. Scopus. https://doi.org/10.3390/app11052167

BIMcollab. (2026). BIMcollab Zoom [Computer software].

buildingSMART International. (2024). Information Delivery Specification (IDS). buildingSMART International. https://www.buildingsmart.org/standards/bsi-standards/information-delivery-specification-ids/

Cuerda, E., Guerra-Santin, O., Sendra, J. J., & Neila, Fco. J. (2020). Understanding the performance gap in energy retrofitting: Measured input data for adjusting building simulation models. Energy and Buildings, 209, 109688. https://doi.org/10.1016/j.enbuild.2019.109688

D'Angelo, L., Hajdukiewicz, M., Seri, F., & Keane, M. M. (2022). A novel BIM-based process workflow for building retrofit. Journal of Building Engineering, 50, 104163. https://doi.org/10.1016/j.jobe.2022.104163

Feng, K., Lu, W., Wang, Y., & Man, Q. (2022). Energy-Efficient Retrofitting under Incomplete Information: A Data-Driven Approach and Empirical Study of Sweden. Buildings, 12(8). https://doi.org/10.3390/buildings12081244

Hadjidemetriou, L., Stylianidis, N., Englezos, D., Papadopoulos, P., Eliades, D., Timotheou, S., Polycarpou, M. M., & Panayiotou, C. (2023). A digital twin architecture for real-time and offline high granularity analysis in smart buildings. Sustainable Cities and Society, 98, 104795. https://doi.org/10.1016/j.scs.2023.104795

Hosamo, H. H., Nielsen, H. K., Kraniotis, D., Svennevig, P. R., & Svidt, K. (2023a). Digital Twin framework for automated fault source detection and prediction for comfort performance evaluation of existing non-residential Norwegian buildings. Energy and Buildings, 281, 112732. https://doi.org/10.1016/j.enbuild.2022.112732

Hosamo, H. H., Nielsen, H. K., Kraniotis, D., Svennevig, P. R., & Svidt, K. (2023b). Improving building occupant comfort through a digital twin approach: A Bayesian network model and predictive maintenance method. Energy and Buildings, 288, 112992. https://doi.org/10.1016/j.enbuild.2023.112992

International Energy Agency. (2023). Buildings. International Energy Agency. https://www.iea.org/topics/buildings

Li, C., Lu, P., Zhu, W., Zhu, H., & Zhang, X. (2023). Intelligent Monitoring Platform and Application for Building Energy Using Information Based on Digital Twin. Energies, 16(19), 6839. https://doi.org/10.3390/en16196839

Ma, T., Xiao, F., Zhang, C., Zhang, J., Zhang, H., Xu, K., & Luo, X. (2025). Digital twin for 3D interactive building operations: Integrating BIM, IoT-enabled building automation systems, AI, and mixed reality. Automation in Construction, 176, 106277. https://doi.org/10.1016/j.autcon.2025.106277

Muchla, A., Kurcjusz, M., Sutkowska, M., Burgos-Bayo, R., Koda, E., & Stefańska, A. (2025). The Use of BIM Models and Drone Flyover Data in Building Energy Efficiency Analysis. Energies, 18(13). Scopus. https://doi.org/10.3390/en18133225

Piras, G., Agostinelli, S., & Muzi, F. (2025). Smart Buildings and Digital Twin to Monitoring the Efficiency and Wellness of Working Environments: A Case Study on IoT Integration and Data-Driven Management. Applied Sciences, 15(9). https://doi.org/10.3390/app15094939

Appendix A

Specification 1

External walls

IFC versions: IFC4

The model MUST contain entities that have

- IFC class IFCWALL
- property IsExternal of PSet Pset_WallCommon (IFCBOOLEAN) = true

that MEET the following requirements

- MAY HAVE classification omniclass
- MUST HAVE attribute Name
- MUST HAVE property ThermalTransmittance of PSet Pset_WallCommon (IFCTHERMALTRANSMITTANCEMEASURE) > 0
- MUST HAVE property LoadBearing of PSet Pset_WallCommon (IFCBOOLEAN) = true
- MAY HAVE property FireRating of PSet Pset_WallCommon (IFCLABEL)
- MUST HAVE property NetSideArea of PSet Qto_WallBaseQuantities (IFCAREAMEASURE)
- MUST HAVE property NetVolume of PSet Qto_WallBaseQuantities (IFCVOLUMEMEASURE)
- MUST HAVE any material

Specification 2

Roofs

IFC versions: IFC4

The model MUST contain entities that have

- IFC class IFCROOF
- attribute ObjectType = ROOF

that MEET the following requirements

- MUST HAVE property ThermalTransmittance of PSet Pset_RoofCommon (IFCTHERMALTRANSMITTANCEMEASURE) > 0
- MUST HAVE property IsExternal of PSet Pset_RoofCommon (IFCBOOLEAN) = true
- MUST HAVE property GrossArea of PSet Qto_RoofBaseQuantities (IFCAREAMEASURE)
- MUST HAVE any material

Specification 3

Slabs/Floors

IFC versions: IFC4

The model MUST contain entities that have

- IFC class IFCSLAB
- property IsExternal of PSet Pset_SlabCommon (IFCBOOLEAN) = true

that MEET the following requirements

- MUST HAVE property ThermalTransmittance of PSet Pset_SlabCommon (IFCTHERMALTRANSMITTANCEMEASURE) > 0
- MUST HAVE property IsExternal of PSet Pset_SlabCommon (IFCBOOLEAN) = true
- MUST HAVE any material

Specification 4

Windows

IFC versions: IFC4

The model MUST contain entities that have

- IFC class IFCWINDOW

that MEET the following requirements

- MUST HAVE property UValue of PSet Pset_WindowCommon (IFCLABEL) > 0
- MUST HAVE property IsExternal of PSet Pset_WindowCommon (IFCBOOLEAN) = true
- MAY HAVE property Area of PSet Qto_WindowBaseQuantities (IFCAREAMEASURE)

Specification 5

Doors

IFC versions: IFC4

The model MUST contain entities that have

- IFC class IFCDOOR

that MEET the following requirements

- MUST HAVE property ThermalTransmittance of PSet Pset_DoorCommon (IFCTHERMALTRANSMITTANCEMEASURE) > 0
- MUST HAVE property IsExternal of PSet Pset_DoorCommon (IFCBOOLEAN) = true

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

INTEROPERABILITY IMPERATIVE: A BUILT ENVIRONMENT MATURITY-BARRIER ANALYSIS FOR DIGITAL TWIN CONVERGENCE WITH BIM AND IOT

Ubongabasi Inyang, Bamidele Adebisi, and Weizhuo Wang
Department of Engineering, Manchester Metropolitan University, Manchester, UK

Abstract

The built environment faces pressure to improve efficiency and sustainability throughout its lifecycle. Digital twins (DT), leveraging BIM, IoT sensor data, and AI, offer significant potential for proactive and intelligent asset management. However, deployment is hindered by the Interoperability Imperative—the need for heterogeneous systems to exchange and interpret data seamlessly. This paper develops a Maturity–Barrier Logic that frames DT evolution as constrained by interoperability-related barriers. It identifies a persistent implementation gap, with applications stagnating at the Digital Shadow and failing to achieve autonomous Cognitive Twins. A diagnostic framework is proposed to support the transition toward federated, system-of-systems DT.

Introduction

The Architecture, Engineering, Construction, and Operations (AECO) industry is undergoing a significant transformation driven by the digitalisation and integration of cyber-physical systems (CPS) (Boje et al., 2020). Central to this transition is the convergence of Building Information Modelling (BIM), the Internet of Things (IoT), and DT technologies, enabling a shift from static representations to dynamic, data-driven systems capable of real-time monitoring, simulation, and predictive analysis. Despite this potential, the realisation of fully functional DTs remains constrained by the interoperability imperative for heterogeneous systems to exchange, interpret, and utilise data without ambiguity. In this study, a Digital Twin is defined as a virtual representation of a physical asset that is continuously updated through automated, bi-directional data exchange, enabling real-time monitoring, prediction, and optimisation (Rasheed et al., 2020; Grieves and Vickers, 2017). This emphasises the distinction between true DTs and lower levels of digital representation, where limited data integration restricts system functionality (Kritzinger et al., 2018). As DT evolves, Artificial Intelligence (AI) is increasingly playing a critical role in enabling advanced capabilities. While BIM provides semantic structure and IoT enables real-time sensing, AI functions as the computational engine that supports pattern recognition, predictive analytics, and autonomous decision-making. This integration enables the emergence of Cognitive Twins at higher maturity levels that are capable of adaptive and self-learning behaviour (Boje et al., 2020; Madni et al., 2019; Yitmen et al., 2021). These capabilities also further intensify interoperability requirements, particularly in relation to semantic integration, data governance, and trust in algorithmic outputs. Despite extensive research on DT technologies and their associated implementation barriers, a critical limitation persists as existing studies either focus on identifying barriers ranging from technical and data-related challenges to organisational constraints (Opoku et al., 2023; Kineber et al., 2023) or on defining Digital Twin Maturity Models (DTMM) that describe system evolution (Deng et al., 2021; Liu et al., 2024). To address this gap, this paper develops a structured analysis that links interoperability barriers to distinct stages of DT maturity. Several DTMMs have been proposed and a referenced classification by Kritzinger et al. (2018) distinguishes between Digital Models, Digital Shadows, and DTs based on the level of data integration and automation. Studies highlight limitations within existing maturity models and Liu et al. (2024) noted that many DTMMs lack clarity in defining capability progression and fail to provide continuous and measurable development pathways thus resulting in models that effectively describe idealised end states. Interoperability is consistently identified as the most critical technical constraint (Naderi and Shojaei, 2023), and challenges arise at both syntactic and semantic levels, reflecting difficulties in data exchange formats as well as inconsistencies in meaning and interpretation across systems. A key issue is the integration of BIM data, typically structured in Industry Foundation Classes(IFC) schemas, with IoT-generated time-series data, which operates on fundamentally different data models (Boje et al., 2020). Machine-readable knowledge representation and integration such as semantic web technologies and ontologies, such as Brick Schema, SOSA/SSN (Sensor, Observation, Sample, and Actuator/Semantic Sensor Network), and ifcOWL(Web Ontology Language representation of the IFC schema), have been proposed to address these challenges (Petrova and Pauwels, 2020; Farghaly et al., 2023), but their adoption remains limited due to implementation complexity and the absence of unified industry standards. In addition, technical, organisational, and governance barriers further impede DT adoption. Concerns related to data security, privacy,

and ownership also complicate collaboration across stakeholders, limiting trust and restricting data sharing (AlBalkhy et al., 2024; Kineber et al., 2023).

Further analysis reveals a fundamental disconnect between two dominant research streams: the study of DTMM and the identification of implementation barriers. While maturity models define the evolutionary trajectory of DT systems (Deng et al., 2021; Liu et al., 2024), and barrier-focused studies classify implementation challenges (Opoku et al., 2023), there is a lack of integration between these perspectives. Existing research does not systematically examine how particular barriers constrain progress between maturity stages. For example, early-stage DT development is often limited by data acquisition and sensor integration challenges, whereas advanced stages require sophisticated semantic interoperability, data governance, and trust in AI-driven decision-making (AlBalkhy et al., 2024; Boje et al., 2020). This lack of alignment results in a practical limitations where stakeholders are unable to identify which barriers must be addressed to achieve specific maturity transitions, as current frameworks do not provide actionable guidance for implementation or strategic planning (Liu et al., 2024). By synthesising technical, semantic, and organisational challenges identified in the literature, this study proposes a Maturity–Barrier Logic that maps barriers, interoperability challenges to DT maturity stages to address this gap and enable a more systematic understanding of transition constraints and supporting more informed decision-making in DT implementation in the built environment. To guide this investigation, the study addresses the following research questions:

- How do interoperability barriers manifest across different DT maturity stages?
- What specific barriers inhibit progression between these maturity stages?
- How can these barriers be mapped to support DT diagnostic framework development?

Methodology

This study adopts a conceptual synthesis approach to systematically identify and analyse interoperability-related barriers and their relationship to DT maturity stages in the built environment. Guided by PRISMA principles, the review was conducted using Scopus as the primary database. Searches, restricted to English-language publications from 2015 onward, using this terms : Core Concept(Federated DTs); Domain & Context (Built Environment, Construction Industry, Asset Lifecycle, Facility Management (O&M)); Tech & Systems (BIM, IoT, Sensors, Cyber-Physical Systems, Data/System Integration); Interoperability (Semantic & Organisational Standards, IFC, Asset Administration Shell, Ontologies); Performance (Predictive/Proactive Maintenance, Energy Optimisation, IEQ, Occupant Comfort); Governance (Data/DT Governance, Ecosystems, Policy, Stakeholder & Tech Readiness); Implementation (Adoption Barriers (Org/Process), Deployment Challenges, Transformation); Maturity (Digital Twin Maturity Models, Technology & Digital Maturity Levels); Methodology (Systematic Literature Reviews, Bibliometric Analysis). To capture interoperability challenges across the full technological spectrum, keywords included not only “digital twin” but also associated enablers such as BIM and IoT. The search process yielded 1,089 records. Following relevance screening, 180 studies were retained. The analysis covered both technical barriers (data integration, semantic alignment, standards fragmentation, legacy systems) and organisational issues (governance, stakeholder readiness, and workflow alignment). The outcome is a structured Maturity–Barrier Mapping that aligns interoperability challenges with DT maturity progression.

Barrier Taxonomy and Distribution

The systematic analysis of the provided literature reveals that the convergence of BIM, IoT, and DT is hindered by a complex matrix of technical and organisational barriers. These barriers confirm the Interoperability Imperative, in which the primary friction lies not merely in the availability of technology but in the capacity to integrate heterogeneous systems and govern them effectively. The Barrier-to-Maturity Mapping Logic was developed by extracting interoperability-related barriers and grouping them into categories and then mapping them to DT maturity stages. Table 1 illustrates the delineation of these challenges, categorised by their impact on the DT.

Interoperability Barriers:

- **Data Capture and Connectivity:** A fundamental technical barrier is the difficulty in establishing real-time synchronisation between the physical asset and the virtual counterpart (Naderi and Shojaei, 2023).
- **Semantic Alignment:** A critical limitation identified is the semantic gap between BIM that relies on static, hierarchical schemas (e.g., IFC), while IoT generates dynamic, time-series data.
- **Visualisation and Integration:** The integration of diverse software tools and the diversity in source systems create fragmented data silos.

Governance and Organisational Barriers:

- **Organisational Inertia:** This poses challenges that are as great as, or greater than, technical limitations. Overcoming the organisational inertia inhibiting DT adoption requires a shift from voluntary adoption to mandated governance frameworks that support a System-of-Systems approach.
- **Fragmentation and Silos:** The industry is characterised by fragmented supply chains and silos between departments, which obstruct the holistic data sharing required for DTs (AlBalkhy et al., 2024;

Table 1: Taxonomy of Barriers to DT-BIM-IoT Convergence

Category	Barrier Sub-Type	Description & Key Citations
Technical/ Semantic	Semantic Interoperability	Inability of systems to interpret data across BIM (static) and IoT (dynamic) domains; lack of unified ontologies.
	Data Integration/Fusion	Difficulty fusing heterogeneous data formats (e.g., geometric vs. time-series); diversity in source systems.
	Connectivity/Latency	Real-time data transmission, bandwidth, and latency issues that disrupt the physical–virtual link.
	Data Quality	Uncertainties with data quality and reliability, including noisy sensor data and lack of accuracy.
Organisational Governance	Readiness & Skills	Skilled professionals shortage, low level of knowledge, resistance to change, and lack of training.
	Structural/Process	Fragmentation of the industry; silos between departments.
	Economic/Financial	High investment costs, uncertain ROI, and lack of financial incentives.
	Trust & Security	Lack of trust in data security, privacy concerns, and unclear data ownership/intellectual property rights.

Opoku et al., 2023).

- **Readiness and Skills:** The pervasive low level of knowledge and lack of competence regarding DT technologies among stakeholders contributes to resistance to change and a preference for traditional workflows (Kineber et al., 2023; Opoku et al., 2023).
- **Cost and Value:** High initial investment costs (CapEx) and an unclear value proposition or unproven Return on Investment (ROI) act as significant deterrents, particularly for Small and Medium Enterprises (SMEs) (AlBalkhy et al., 2024; Kineber et al., 2023).

Exemplars and Benchmarks:

The literature frequently cites specific frameworks as benchmarks for guiding DT implementation:

- **The Gemini Principles:** Proposed by the Centre for Digital Built Britain (CDBB), Purpose, Trust, Function principles are cited as a foundational ethical and functional value-based framework that serves as a benchmark for national-level DT strategy (Al-Sehrawy et al., 2021; Bolton et al., 2018; Chen et al., 2021). It emphasises that a DT must have a clear purpose, must be trustworthy, and must function effectively. Chen et al. (2021) utilised these principles to develop a maturity model for asset management, validating that alignment with these principles correlates with higher DT maturity.
- **Incremental Digital Twin (iDTw) Framework:** A framework, proposed by Calvetti et al. (2025), that breaks down DT implementation into granular levels (Static, Detailed, Sensored, Responsive, Adaptive, Intelligent). It is particularly effective for smart enterprises and construction logistics, offering a phased deployment strategy that lowers the barrier to entry for organisations daunted by the complexity of complete autonomy.
- **BIM-IoT Data Integration (BIM-IoTDI) Architecture:** Developed by Eneyew et al. (2022), this framework explicitly addresses the semantic interoperability gap in smart buildings. By utilising an ontology-based query mediation method, it enables the integration of static BIM data and dynamic IoT data without loss of semantic meaning. It is critical for supporting proactive FM tasks such as predictive maintenance and energy management.
- **CITYSTEPS Maturity Model:** This model assesses governance and technological readiness in urban DTs. It highlights that while many projects achieve visualisation, few progress to real-time decision-making or autonomous stages due to governance and interoperability failures (Masoumi et al., 2023).
- **ISO Standards:** References are made to ISO 19650 for information management and ISO 23247 for automation systems as emerging standards attempting to regulate this space (Calvetti et al., 2025; Palihakkara et al., 2025).

Mapping Maturity Levels and Interoperability Gaps

The analysis identifies a consensus on multi-stage evolutionary pathways for DTs, as presented in Table 2. While terminologies vary, the progression moves from static representation to autonomous interaction. Crucially, specific interoperability gaps act as limitations, preventing progression to higher maturity levels. The strategic roadmap synthesises the technical, semantic, and organisational requirements necessary to navigate the convergence to address the Interoperability Imperative by structuring deployment into maturity phases to overcome specific barrier limitations. To navigate the Interoperability Imperative, stakeholders should adopt a

Table 2: Maturity Mapping and Levels Gaps

Maturity Level	Characteristics and Limitation
Level 1: Descriptive (Digital Model)	Static 3D representation (e.g., BIM) with manual data exchange. **Limitation:** Digitisation Gap. The transition to the next level is blocked by the inability to digitise physical assets and processes effectively.
Level 2: Informative (Digital Shadow)	One-way automatic data flow from physical to digital (monitoring)- real-time visualisation. **Limitation:** Connectivity Gap. Progress is restricted by the lack of robust IoT infrastructure and reliable data transmission protocols.
Level 3: Predictive (Adaptive Twin)	Bi-directional data flow or advanced simulation capabilities. Uses historical data and AI to predict future states. **Limitation:** Semantic Gap. Lack of semantic interoperability (ontologies) prevents the DT from contextualising IoT readings within BIM geometry.
Level 4: Comprehensive/ Intelligent (Cognitive Twin)	Autonomous decision-making, self-learning, and prescriptive analytics. The DT can act on the physical world. **Limitation:** Trust and Governance Gap. Stakeholders lack trust in black box AI algorithms and fear security breaches.
Level 5: Federated (System-of-Systems)	Interconnected DTs interacting to optimise broader systems. **Limitation:** Standardisation Gap. The lack of universal data standards prevents disparate DTs from interacting.

phased approach, as illustrated in Table 3 case studies for overcoming distinct interoperability challenges at each phase to unlock subsequent maturity levels. For Foundational (Static Twin) where the asset exists as a high-fidelity digital model, the primary hurdle data quality and digitisation is the static nature of building data and the lack of accurate as-built geometric information (Opoku et al., 2023). Without resolving the digitisation gap, the twin remains a passive repository. For digital shadow, the one-way automated data flow from physical sensors to the digital platform affects connectivity and integration. Progression requires overcoming the Connectivity Gap by integrating IoT sensors with legacy systems and ensuring low-latency data transmission (Kang and Mo, 2024; Sepasgozar, 2021). This phase establishes the digital shadow. In Autonomous Cognitive Twin where bi-directional data flow with AI helping drive autonomous decision-making, Semantic Alignment Interoperability is required to move from a shadow to a true twin capable of prediction and control. This involves mapping heterogeneous data streams to unified ontologies (e.g., aligning IFC with SSN/SOSA) to enable automated reasoning and simulation (Boje et al., 2020; Huang and Hsieh, 2026). On the Federated System-of-Systems, trust is the key anchor.

Bridging the Gap: From Connectivity to Cognition

The progression through the roadmap is not linear but rather stepwise; specific barriers serve as limitations that must be overcome to advance.

- **Foundational to Integrative (The Connectivity Limitations):** To move from Level 1 (Static) to Level 2/3 (Dynamic), practitioners must resolve data and/or integration barriers. Specifically, the physical-digital disconnect must be bridged by retrofitting assets with IoT sensors and establishing reliable middleware (e.g., Node-RED) to handle the velocity of real-time data (Biagini et al., 2024; Naderi and Shojaei, 2023).
- **Integrative to Autonomous (The Semantic Limitations):** To move from Level 3 (Monitoring) to Level 4/5 (Prediction/Autonomy), the Semantic/Standards Barrier is the primary obstacle. A DT cannot predict if it does not understand the relationship between a temperature reading and the volume of a specific room. This requires resolving semantic silos by mapping BIM data to ontologies (e.g., Brick, SSN), moving beyond simple data transfer to knowledge graph integration (Boje et al., 2020; Eneyew et al., 2022).
- **Autonomous to Federated (The Trust Limitations):** To achieve a true System-of-Systems (Federated DT), the organisational governance barrier must be resolved. Technical connectivity is insufficient if stakeholders do not trust the data security or the AI's decision-making logic. Overcoming this requires adherence to governance frameworks as these legal structures are essential for resolving data sovereignty issues, enabling stakeholders to share proprietary data within federated ecosystems without compromising control, thereby meeting the governance requirements for Level 5 (AlBalkhy et al., 2024; Jørgensen, 2025). Policy-makers must establish a regulatory environment that incentivises interoperability through three core actions. First, public procurement must mandate open standards to ensure that data remain Findable,

Table 3: Maturity Mapping and Interoperability Gaps: Empirical Case Studies and Barrier Resolution

Maturity Phase	Empirical Case Study Evidence
Phase 1: Foundational (Descriptive Twin)	**The Digitisation Gap:** Overcomes data/integration barriers (AlBalkhy et al., 2024). **Case Study:** Northumbria University City Campus (UK) (Boje et al., 2020; Lu et al., 2020) **Digitisation Barrier:** Case study proves that relying on fragmented legacy documents restricts facility management to manual updates. It demonstrates that establishing a unified, standardised 3D semantic model (ISO 19650) is the foundational requirement to break non-machine-readable data
Phase 2: Integrative (Responsive Twin)	**The Connectivity Gap:** Overcomes established data connectivity barriers. (Boje et al., 2020; Jahangir et al., 2024). **Case Study:** University of Florence & UNF Engineering Building (Biagini et al., 2024; Kang and Mo, 2024) **Connectivity Barrier:** The Florence case proves that a middleware layer (Node-RED) is required to overcome protocol mismatches, while the UNF case shows that custom IFC Property Sets (pSets) anchor dynamic data feeds to static BIM elements, enabling a real-time Digital Shadow.
Phase 3: Autonomous (Cognitive Twin)	**The Semantic Gap:** Overcomes AI reasoning barriers (Boje et al., 2020). **Case Study:** Institute for Manufacturing (West Cambridge, UK) (Eneyew et al., 2022) **Semantic Barrier:** Proves an IFCxBrick pipeline is required for semantic interoperability. Mapping IFC to Brick via RDF/OWL achieves machine-understandable context, making automated tasks like Fault Detection and Diagnosis (FDD) feasible.
Phase 4: Federated (System-of-Systems)	**The Trust Gap:** Overcomes multi-owner data barriers (AlBalkhy et al., 2024). **Case Study:** Smart Dublin / Urban DT (Jørgensen, 2025; Masoumi et al., 2023) **Trust & Governance Barrier:** Proves that formal ''Data Spaces'' aligned with the Data Governance Act (DGA) are required for trusted exchange. It demonstrates that technical interoperability alone is insufficient without legally compliant open-data frameworks.

Accessible, Interoperable, and Reusable (FAIR), thereby preventing vendor lock-in (Klar et al., 2024). Second, the adoption of the Smart Readiness Indicator (SRI) should be promoted as a regulatory benchmark for assessing a building's capacity to adapt to needs and demands, serving as a key performance indicator of DT effectiveness. Finally, governance must align with the Data Governance Act (DGA) and Data Act (DA) to establish secure Data Spaces (Jørgensen, 2025).

Digital Shadow to Cognitive Twin Transition:

The most significant limitation identified in the transition from Level 2 (Digital Shadow) to Level 3/4 (Cognitive Twin) is Semantic Interoperability. To resolve this semantic gap and evolve the DT from a passive visualisation tool into a Cognitive Twin, a Diagnostic Implementation Guide (Figure 1) was developed. By synthesising theoretical concepts, maturity models, and empirical case studies from existing literature, this guide was structured to address specific technical constraints. The involved transforming disparate technical data into a cohesive, step-by-step workflow designed to transition a DT from a stalled state to a cognitive capability. Consequently, Figure 1 serves as a diagnostic framework outlining the technical protocols required to overcome this barrier—a critical distinction, as practitioners frequently conflate mere data connectivity (the transmission of bytes) with true data integration (the synthesis of meaning).

Discussion and Implications

DTs as Socio-Technical Systems: Beyond Technical Maturity: As illustrated in Figure 2, DT progression extends beyond static 3D models to real-time monitoring, predictive analytics, and autonomous decision-making, rather than being solely determined by technological capability (Sacks et al., 2020; Yitmen et al., 2021). Instead, it is inherently socio-technical, requiring alignment between technical infrastructure and organisational practices. The transition across maturity levels is systematically constrained by interoperability-related barriers—digitisation, connectivity, semantic, and governance gaps—which reflect both technical limitations and stakeholder misalignment. The evolution to Cognitive and Federated Twins is not simply a matter of deploying more advanced technologies, but of resolving the human, organisational, and institutional constraints embedded in each transition stage. To bridge and resolve the semantic gap between connectivity and cognition, the diagnostic guide in Figure 1 is useful to aid the transformation of the DTs, a prerequisite for envisioned predictive maintenance and energy optimisation (Boje et al., 2020; Liu et al., 2024).

Stakeholder-Specific Implications: Operationalising the Framework: The Maturity–Barrier Logic provides a structured lens for operationalising stakeholder

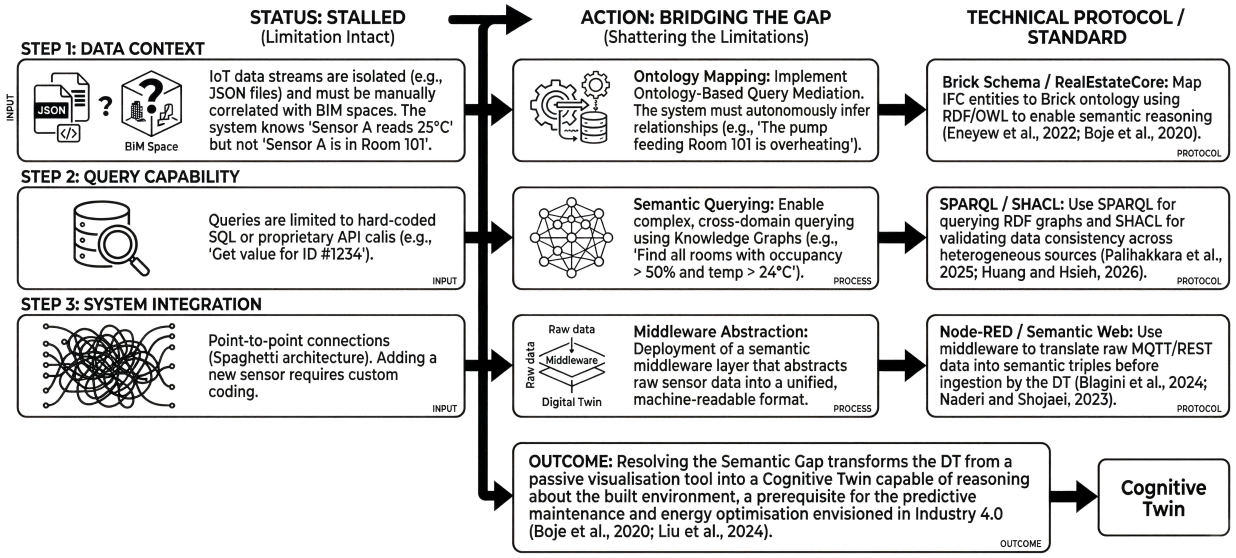


Figure 1: Diagnostic Implementation Guide for Overcoming the Semantic Barrier

responsibilities. Each blockade identified in Figure 2 corresponds to specific actions required by key actors:

- **Asset Owners:** To overcome the barriers at early and intermediate maturity stages of unclear value proposition, asset owners must transition from passive recipients of static BIM deliverables to active drivers of digital continuity. This requires embedding interoperability and semantic data requirements into procurement contracts, ensuring that deliverables support lifecycle integration rather than project completion alone.
- **Technology Providers:** Providers reliance on proprietary systems and closed data architectures inhibits cross-platform integration. To enable progression, providers must adopt open, modular system architectures, supported by middleware solutions (e.g., Node-RED) and semantic technologies (RDF/OWL). Mapping BIM standards such as IFC to ontologies like Brick Schema or RealEstateCore enables machine-readable context, facilitating advanced analytics and automated reasoning capabilities.
- **Policy-Makers and Industry Bodies:** The transition toward higher maturity levels (Levels 4–5) is constrained by trust and governance barriers, where data sharing is limited by legal, ethical, and commercial concerns. Policy-makers need to establish regulatory frameworks and data governance mechanisms to promote secure and standardised data exchange.

Practical and Research Outlook

The proposed Maturity–Barrier Logic serves not only as an analytical model but as a practical diagnostic tool for guiding DT implementation:

- **Industry Practitioners:** Organisations should utilise the framework to assess their current maturity level and identify the specific barriers preventing progression. For firms operating at the Digital Shadow stage, the priority should shift from additional sensor deployment to semantic data integration and architecture development. Implementing ontology-based pipelines (e.g., IFC-to-Brick) and establishing interoperable data environments are critical next steps.
- **Academic Community:** Future research must move beyond conceptual definitions toward empirical validation and implementation strategies. Key priorities should include quantitative evaluation of DT ROI across lifecycle stages, development of automated ontology mapping techniques and investigation of socio-technical governance models addressing data sovereignty, security, and ethics.

Conclusions

This study addressed the challenge of fragmented DT implementation in the AECO industry by developing a structured Maturity–Barrier Logic that links DT evolution to interoperability constraints. A systematic analysis of existing literature, technical standards, and industry practices was conducted to examine the convergence of BIM, IOT, and DT technologies. Through this analysis, DT maturity stages were explicitly mapped against four critical transition barriers—Digitisation, Connectivity, Semantic, and Trust/Governance gaps—thereby reframing DT progression as a sequence of constrained transitions rather than a purely linear technological evolution. The findings reveal a significant disparity between theoretical DT maturity models and real-world implementation. While existing frameworks suggest a clear pathway toward autonomous and federated systems, most practical applications remain constrained at Level 2 (Digital Shadow). This stagnation is primarily driven by a semantic limitation where systems achieve syntactic connectivity but lack the ontology-based interoperability

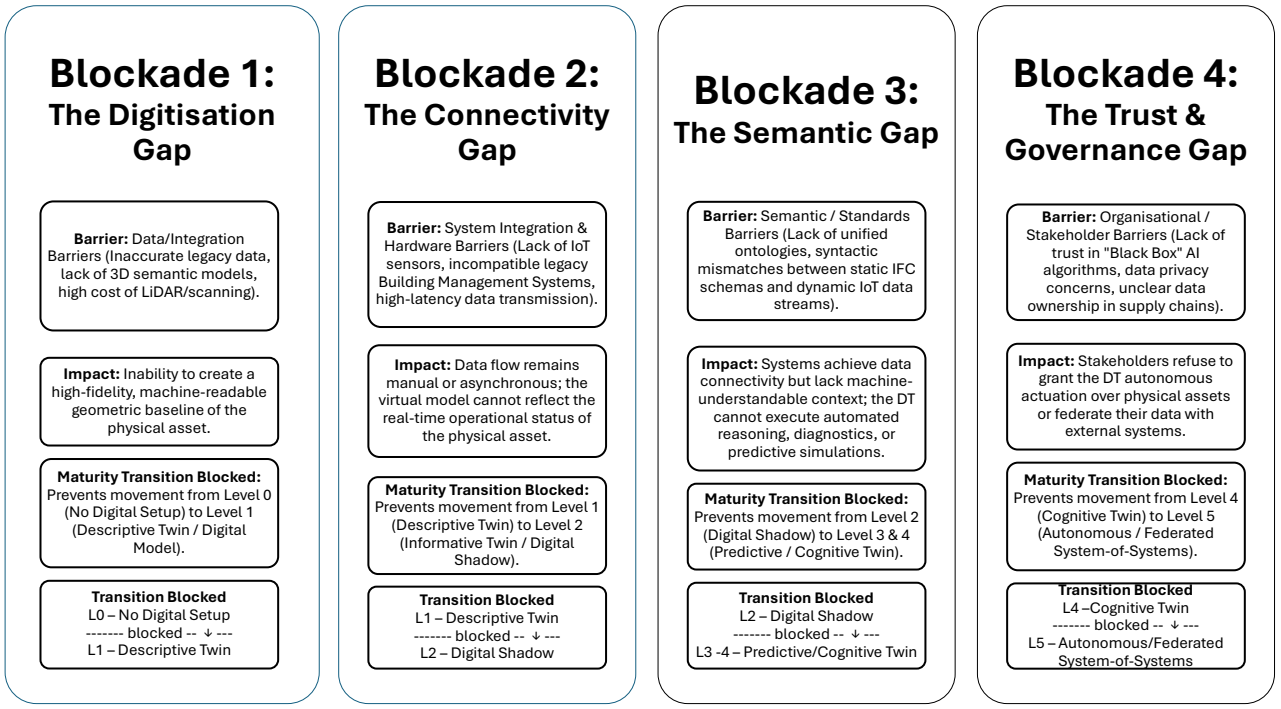


Figure 2: Barrier → Impact → Maturity Transition Diagram

required for machine understanding and AI-driven reasoning. Furthermore, progression to higher maturity levels is inhibited by a trust and governance gap, where organisational, legal, and ethical concerns surrounding data sharing and algorithmic transparency limit the adoption of fully autonomous and federated DT systems. These findings establish interoperability not as a secondary technical issue, but as the central enabling condition for DT maturity. Advancing DTs is less about deploying additional sensing technologies and more about enabling machine-readable, semantically aligned, and governable data ecosystems. Without addressing these barriers, the industry risks remaining confined to isolated, low-value digital implementations. This paper presents diagnostic and decision-support logic for practitioners and researchers alike. This review repositions the future of DTs as an interoperability-driven transformation, where the successful alignment of technology, data, and stakeholders determines the pathway from isolated digital models to fully integrated, intelligent DTs. Achieving this vision requires systematic resolution of the interoperability barriers identified in this study. By addressing these constraints, the industry can transition from fragmented digital implementations to cohesive, intelligent ecosystems that support global sustainability objectives, including Sustainable Development Goal 11.

References

Al-Sehrawy, R., Kumar, B., and Watson, R. (2021). A digital twin uses classification system for urban planning & city infrastructure management. Journal of Information Technology in Construction (ITcon), 26:832–862.

AlBalkhy, W., Karmaoui, D., Ducoulombier, L., Lafhaj, Z., and Linner, T. (2024). Digital twins in the built environment: Definition, applications, and challenges. Automation in Construction, 162:105368.

Biagini, C., Bongini, A., and Marzi, L. (2024). From bim to digital twin: Iot data integration in asset management platform. Journal of Information Technology in Construction (ITcon), 29:1103–1127.

Boje, C., Guerriero, A., Kubicki, S., and Rezgui, Y. (2020). Towards a semantic construction digital twin: Directions for future research. Automation in Construction, 114:103179.

Bolton, A., Butler, L., Dabson, I., Enzer, M., Evans, M., Fenemore, T., Harradence, F., Keaney, E., Kemp, A., and Luck, A. (2018). The Gemini Principles. Centre for Digital Built Britain, Cambridge.

Calvetti, D., Mêda, P., Hjelseth, E., and de Sousa, H. (2025). Incremental digital twin framework: A design science research approach for practical deployment. Automation in Construction, 170:105954.

Chen, L., Xie, X., Lu, Q., Parlikad, A., Pitt, M., and Yang, J. (2021). Gemini principles-based digital twin maturity model for asset management. Sustainability, 13(15):8224.

Deng, M., Menassa, C., and Kamat, V. (2021). From bim to digital twins: A systematic review of the evolution of intelligent building representations in the aec-fm industry. Journal of Information Technology in Construction (ITcon), 26:58–83.

Eneyew, D., Capretz, M., and Bitsuamlak, G. (2022). Toward smart-building digital twins: Bim and iot data integration. IEEE Access, 10:130487–130506.

Farghaly, K., Soman, R. K., and Zhou, S. A. (2023). The evolution of ontology in aec: A two-decade synthesis, application domains, and future directions. Journal of Industrial Information Integration, 36:100519.

Grieves, M. and Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In Transdisciplinary Perspectives on Complex Systems, pages 85–113. Springer, Cham.

Huang, C.-P. and Hsieh, S.-H. (2026). Ontology-driven automation for bim-fm data integration using neo4j, python, and workflow platforms. Advanced Engineering Informatics, 71:104240.

Jahangir, M., Schultz, C., and Kamari, A. (2024). A review of drivers and barriers of digital twin adoption in building project development processes. Journal of Information Technology in Construction (ITcon), 29:141–178.

Jørgensen, B. N. (2025). Digital twins under eu law: a unified compliance framework across smart cities, industry, transportation, and energy systems. Electronics, 14(24):4881.

Kang, T. and Mo, Y. (2024). A comprehensive digital twin framework for building environment monitoring with emphasis on real-time data connectivity and predictability. Developments in the Built Environment, 17:100309.

Kineber, A., Singh, A., Fazeli, A., Mohandes, S., Cheung, C., Arashpour, M., Ejohwomu, O., and Zayed, T. (2023). Modelling the relationship between digital twins implementation barriers and sustainability pillars: Insights from building and construction sector. Sustainable Cities and Society, 99.

Klar, R., Arvidsson, N., and Angelakis, V. (2024). Digital twins' maturity: The need for interoperability. IEEE Systems Journal, 18:713–724.

Kritzinger, W., Karner, M., Traar, G., Henjes, J., and Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, 51(11):1016–1022.

Liu, Y., Feng, J., Lu, J., and Zhou, S. (2024). A review of digital twin capabilities, technologies, and applications based on the maturity model. Advanced Engineering Informatics, 62:102592.

Lu, Q., Parlikad, A., Woodall, P., Don Ranasinghe, G., Xie, X., Liang, Z., Konstantinou, E., Heaton, J., and Schooling, J. (2020). Developing a digital twin at building and city levels: Case study of west cambridge campus. Journal of Management in Engineering, 36(3):05020004.

Madni, A., Madni, C., and Lucero, S. (2019). Leveraging digital twin technology in model-based systems engineering. Systems, 7(1):7.

Masoumi, H., Shirowzhan, S., Eskandarpour, P., and Pettit, C. (2023). City digital twins: Their maturity level and differentiation from 3d city models. Big Earth Data, 7(1):1–36.

Naderi, H. and Shojaei, A. (2023). Digital twinning of civil infrastructures: Current state of model architectures, interoperability solutions, and future prospects. Automation in Construction, 149:104785.

Opoku, D. G. J., Perera, S., Osei-Kyei, R., Rashidi, M., Bamdad, K., and Famakinwa, T. (2023). Barriers to the adoption of digital twin in the construction industry: a literature review. In Informatics, volume 10, page 14. MDPI.

Palihakkara, A. et al. (2025). Toward cloud-based scalable and interoperable digital twins: Review of the state of the art. Journal of Information Technology in Construction (ITcon), 30:1209.

Petrova, E. and Pauwels, P. (2020). Semantic enrichment of association rules discovered in operational building data. In Proceedings of the 37th CIB W78 Information Technology for Construction Conference (CIB W78), pages 308–326, São Paulo, Brazil.

Rasheed, A., San, O., and Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. IEEE access, 8:202438.

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S., and Girolami, M. (2020). Construction with digital twin information systems. Data-Centric Engineering, 1:e14.

Sepasgozar, S. (2021). Differentiating digital twin from digital shadow: Elucidating a paradigm shift to expedite a smart, sustainable built environment. Buildings, 11(4):151.

Yitmen, I., Alizadehsalehi, S., Akıner, and Akıner, M. E. (2021). An adapted model of cognitive digital twins for building lifecycle management. Applied Sciences, 11(9):4276.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

OPEN STANDARDS FOR DATA INTEGRATION AND VISUALIZATION ACROSS THE PROJECT LIFECYCLE: CHALLENGES AND OPPORTUNITIES OF IFC5

Jacopo Cassandro[1], Alessandro Bruttini[2], Andrea Bongini[2], Lukas Guntermann[3], and Markus König[3]
[1]Department of Architecture, Built Environment, and Construction Engineering (DABC), Politecnico di Milano, Milan, Italy
[2]Department of Architecture (DIDA), University of Florence, Florence, Italy
[3]Chair of Computing in Engineering, Ruhr University Bochum, Bochum, Germany

Abstract

Industry Foundation Classes (IFC) enable interoperability across the AECO sector, yet IFC2x3 and IFC4 remain constrained by monolithic structures, limited external data integration, and weak support for real-time and web-based workflows. This paper applies comparative schema analysis and domain-based evaluation across four lifecycle phases, design, construction, operation, and deconstruction, to critically assess these limitations. Emerging IFC5 concepts are examined against identified requirements, highlighting both improvements and unsolved challenges, including governance complexity and implementation maturity. Results indicate increased flexibility and potential, alongside new dependencies and trade-offs. In conclusion, a forward-looking assessment and early-adoption priorities for IFC5-enabled data-centric workflows are provided.

Introduction

The architecture, engineering, construction, and operation (AECO) sector increasingly relies on Building Information Modeling (BIM) to meet growing demands for coordination, efficiency, and information exchange. Initially used for geometric representation, BIM has expanded to support model-based collaboration and a wide range of domain-specific use cases across the project lifecycle (Wang and Chen, 2023). However, to deliver its full value, such collaboration requires the continuous exchange of structured data among stakeholders (Shelden et al., 2020). Addressing this need, the Industry Foundation Classes (IFC) standard, developed by buildingSMART International (bSI), serves as an open and vendor-neutral data schema to support interoperability. It provides a standardized object-based digital description through entities, attributes, and relationships. This includes the geometry of elements as well as information related to processes and operations from design and construction, through safety, operational, and maintenance, to deconstruction (Li et al., 2025).

As industry requirements evolve, IFC technical committees extend the standard with new domain-specific definitions (Laakso and Kiviniemi, 2012), but formal adoption into official releases depends on lengthy ISO-compliant review processes. Consequently, practitioners and researchers frequently used custom solutions and dedicated IFC extensions, highlighting the need for more flexible data structures and interoperable information exchange (Yu et al., 2023).

With the construction industry becoming increasingly data-intensive and reliant on dynamic, multi-source digital workflows, the limitations of existing IFC versions (IFC2x3 and IFC4) have become more evident. These versions lack native support for real-time data exchange, integration with external data sources, advanced analytics, and modern visualization tools. To address these gaps, bSI started the development of IFC5 as a next-generation schema (Berlo, van et al., 2021).

Currently under development, IFC5 adopts modular and extensible architecture and supports multiple interoperable formats, including XML Schema Definition (XSD), Web Ontology Language (OWL), and JavaScript Object Notation (JSON). Moreover, its integration with visualization technologies, such as Universal Scene Description (USD), combines semantic content with real-time, interactive visualizations.

This paper examines current challenges and future opportunities of IFC5 for heterogeneous data integration and visualization across the project lifecycle. Four lifecycle phases and representative use cases are considered: design (cost and time management), construction site health and safety (H&S) management, operation and maintenance (O&M), and circular deconstruction. Based on a literature review of IFC-based approaches for integrating BIM data, a set of functional requirements that the schema should support in each domain, with emphasis on data representation, exchange, and integration, was identified. These requirements ensure a comparative schema analysis of IFC2x3 and IFC4, highlighting recurring gaps and challenges across phases. The analysis subsequently examines emerging IFC5 concepts to discuss how far they address these limitations and to what extent they can support more flexible and interoperable data environments. Overall, the paper analyses the ongoing evolution of IFC and discusses early adoption, as well as the practical implications of integrating IFC5 into data-centric AECO workflows.

Research questions and methodology

To assess the current limitations and future potential of IFC for data-centric integration and visualization across four representative lifecycle use cases, the following research questions (RQs) were identified:

- RQ1: *Which limitations emerge for data representation, exchange, integration, and visualization in the IFC2x3 and IFC4 versions across key lifecycle phases of the built asset?*
- RQ2: *To what extent do emerging IFC5 concepts address these gaps, and what trade-offs and implications emerge for implementation across the same lifecycle phases?*

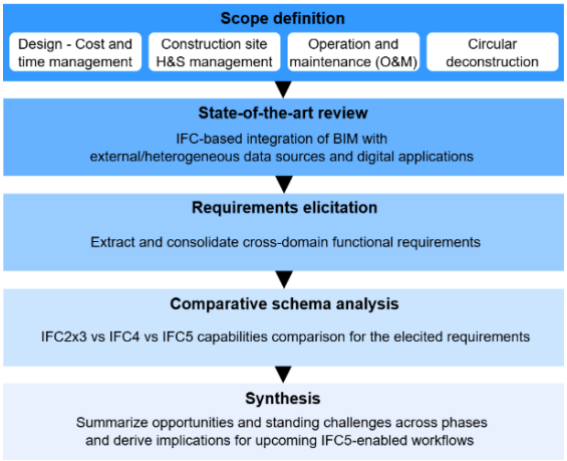


Figure 1: Research methodology

Building on these RQs, the study follows the workflow reported in Figure 1. First, the scope is defined by selecting four lifecycle phases and representative use cases: cost and time management, construction site H&S management, O&M, and circular deconstruction. A state-of-the-art review is then conducted, focusing on IFC-based approaches that integrate data sources and digital applications in these domains. Building on the literature, both specific and cross-domain functional requirements are derived and are intended to capture recurring patterns across lifecycle phases and use cases. These are subsequently used as evaluation criteria in a comparative assessment of IFC versions' capabilities, to determine which address the identified needs and to highlight gaps, trade-offs, and implications for upcoming implementations. Finally, findings are synthesized to summarize key challenges and opportunities.

Background

Since its early releases, IFC has been developed as a vendor-neutral schema to support information exchange across the AECO sector. Its evolution since the late 1990s reflects growing industry requirements alongside data modeling constraints (Laakso and Kiviniemi, 2012).

IFC2x3, standardized as ISO/PAS 16739:2005, remains the most widely implemented version. Based on a monolithic STEP/EXPRESS architecture optimized for file-based exchange, it reliably represents geometry and core building semantics. However, limited extensibility, weak integration of external data, and reduced support for process-oriented information constrain its suitability for data and real-time workflows (Jia et al., 2024).

IFC4 ADD2 TC1, standardized as ISO 16739-1:2018, following a first release in 2013, introduced incremental refinements over IFC2x3, including improved geometric representations, clearer and more extensive property sets (PSets), stronger semantic consistency, and closer alignment with Model View Definitions (MVDs). Complementary Information Delivery Specifications (IDS) and buildingSMART Data Dictionary (bSDD), further support the validation and definition of information requirements. IFC4 also expanded process-related modeling and domain coverage, though it preserved the underlying monolithic exchange paradigm. Subsequent revisions culminated in IFC4.3, released in 2024 as ISO 16739-1:2024, extending the schema to infrastructure assets with the same architectural foundations. As a result, despite a broader scope and richer semantics, persistent structural constraints remain, particularly regarding extensibility, modular evolution, and alignment with web and cloud-native environments (Yu et al., 2023). To address these constraints, IFC5 is under development as a step-change toward a modular, decomposed architecture supporting multiple serializations (e.g., XSD, OWL, JSON). This shift aims to improve interoperability with digital ecosystems and real-time workflows, while introducing new challenges related to governance, migration, and tooling maturity. In parallel, emerging OpenUSD-based pipelines illustrate how IFC semantics can be integrated with high-performance visualization frameworks to support digital twins and interactive applications (Ren et al., 2025).

The following subsections review IFC-based implementations across the selected lifecycle phases, summarizing the state of the art and informing of the functional requirements used in the comparative analysis.

Cost and time management – Design phase

During the design phase, BIM has been increasingly adopted to support cost estimation and scheduling, enabling earlier and more informed decision-making. In this context, 4D and 5D BIM, which integrate time and cost information with 3D geometry, have been widely investigated to improve transparency and control during design (Yang et al., 2022). Prior research has shown that model-based quantity take-off, automated cost calculation, and links to external cost databases can improve accuracy and consistency compared to document-based workflows (Sacks et al., 2018).

Recent studies have revisited the role of IFC in cost estimation and scheduling, demonstrating that cost and time information can be explicitly modeled using entities such as *IfcCostItem*, *IfcCostSchedule*, *IfcTask*, and *IfcWorkSchedule*. For example, Cassandro et al. (2024) propose an IFC-based approach linking cost items directly to geometric elements, enabling semi-automated cost calculation and validation, while comparative results indicate reductions in manual coding and errors relative to traditional 5D BIM software. Similarly, Fürstenberg et al. (2021) show that IFC4-based process models can support data exchange across stakeholders and software.

Despite these advances, cost and scheduling information are still rarely managed directly within IFC in practice. Instead, they are typically handled in external tools using

proprietary formats (Borrmann et al., 2018). Links between geometry, cost, and time are commonly established through key–value mechanisms (e.g., identifiers, classification systems, or WBS codes), rather than through semantically explicit relationships (Boje et al., 2020). Consequently, current IFC-based workflows support data exchange rather than full data integration, leaving cost and time data external to the model and highlighting persistent challenges in modeling effort, tool support, and workflow integration.

Site and safety management – Construction phase

During construction, processes become more dynamic and seamless information exchange becomes crucial. Construction site H&S management is shaped by spatiotemporal complexity, as risk emerges from changing conditions, temporary works, and interactions among workers, equipment, and partially completed assets. Therefore, effective digital support requires situational awareness of what is being built, the sequencing and spatial context of activities, and the evolution of hazards and mitigation measures as the site layout changes (Zhang et al., 2024). IFC-based workflows address these needs by offering a shared, vendor-neutral backbone that anchors H&S information to a consistent spatial and semantic reference (Hu et al., 2024). Safety relevant attributes can be linked to elements and spaces via PSets and classifications, while process constructs (e.g., tasks, schedules) connect activities to affected components, enabling 4D safety reasoning and automated or semi-automated checks (e.g., fall risks) and visual outputs such as hazard maps (Hong and Teizer, 2025). Temporary works and site logistics remain essential yet often under-specified, so many approaches rely on modeling conventions, dedicated libraries, or external datasets linked through IFC identifiers to maintain consistency across revisions (Getuli et al., 2025). Given the document-intensive nature of H&S, exchanges must preserve traceability (e.g., inspection reports or photos) and support heterogeneous, frequently updated data streams (e.g., sensor observations) through linked services or databases rather than embedding volatile data in static files (Amo Larbi et al., 2024). Issue communication and visualization-ready delivery (e.g., dashboards) are also central, with growing attention to AR-assisted compliance checks (Corneli et al., 2025) and integrated VR/AR safety training (Getuli et al., 2020).

Operation and Maintenance – Operational phase

Effective management decisions during the operational phase require timely, accurate information on building components, systems, and their performance over time. IFC-based methodologies have been investigated to support the transition from construction to O&M. Key IFC entities, including MEP element classes, O&M-related PSets, and space-related constructs, provide the semantic linking of physical assets with maintenance data, performance indicators, and lifecycle cost data (Patacas et al., 2016). Several extensions to IFC4 have been proposed to improve the representation of maintenance activities, inspection, and failure modes. In parallel, COBie has become a widely adopted standard for structuring O&M data at handover; however, its implementation continues to face challenges related to data consistency and tool interoperability (Oliveira et al., 2024).

Despite these efforts, interoperability and data quality problems remain significant. O&M systems (such as CAFM and CMMS) rely on proprietary schemas that are incompatible with IFC, often requiring custom integrations that increase complexity and error risk (Farghaly et al., 2018). Operational data, including sensor measurements and work orders, are typically stored in external databases and only referenced in BIM models. Existing PSets frequently fail to meet O&M needs, leading to custom extensions and data loss during data exchanges (Tomczak et al., 2022). Additional barriers include inaccurate as-built models, generic O&M MVD, and organizational limitations such as insufficient BIM expertise and high implementation costs (Otranto et al., 2025). Emerging digital twin approaches aim to address these limitations by integrating real-time Internet of Things (IoT) sensor data with IFC, enabling continuous monitoring and predictive maintenance (Biagini et al., 2024). However, challenges related to data governance, interoperability, and standardized workflows limit large-scale implementation.

Circularity and reuse – Deconstruction phase

Robust reuse outcomes require reliable data: accurate as-built models and maintenance records enable assessment of reuse potential. Requirements for circularity, reuse, and deconstruction are inconsistently defined. Life Cycle Assessment (LCA) is central, but in IFC2x3/IFC4, its integration typically relies on custom PSets (Guntermann et al., 2025). Diverse metrics (e.g., global warming potential, service life, disassemblability) complicate integration. Harmonization efforts, as in the bSDD, are critical for a design for deconstruction (DfD) and for planning with reused components in a closed-loop circular economy. Kim, S. and Kim, S.-A. (2023) developed a graph-based DfD tool for steel construction using *IfcRelConnectsElements,* and Allam and Nik-Bakht (2025) combined knowledge graphs with IFC to automate deconstruction planning, deriving XML-based schedules with 4D simulations from PSets and IFC classes. Representation and extraction of circular information remain challenging. Tomczak et al. (2024) reported difficulties documenting disassembly instructions and element connections in IFC. While *IfcStructuralPointConnection* captures connection types for structural elements, non-structural components (e.g., windows) are more problematic. Mantelatto et al. (2026) used a Python routine based on IfcOpenShell to extract interdependence between structural elements.

IDS validates IFC information requirements. For circularity, limitations stem from IDS v1.0, not IFC: requirements address only individual entities (no cross-entity generalization), and abstract IFC classes cannot be referenced – since *IfcPort* is abstract, non-structural elements cannot be validated this way (Guntermann et al., 2025; Tomczak et al., 2024). Reuse Marketplaces

(Hagedorn et al., 2025) impose further requirements: component separability is critical; workflows require extracting and publishing component-level IFC files, raising feasibility questions around reliable extraction/integration and ID management, especially the fate of GUIDs when components are cut and whether parent IDs can generate child IDs.

Data requirements for IFC standards

This section consolidates five shared, cross-domain requirements for data integration and visualization, identified through analysis of the four lifecycle use cases. Although their relative importance varies across different stages, all requirements recur in multiple domains and therefore constitute shared evaluation criteria for the comparative analysis of IFC2x3, IFC4, and IFC5 concepts presented in the following sections.

Data model extensibility

To support evolving lifecycle needs, a schema is required that can be extended with additional concepts, attributes, and relationships while remaining stable across versions. This is essential for integrating heterogeneous data, such as cost structures, safety rules, operational and deconstruction data, without compromising interoperability, since limited extensibility often leads to proprietary extensions or reliance on external data, weakening standard-based integration.

Integration with external datasets

The capability to reference, align, and synchronize non-IFC data sources with IFC-based models is critical for integrating cost databases, construction schedules, sensor streams, regulatory documents, maintenance records, and environmental data. Beyond simple identifiers or file references, it requires semantically explicit links, version awareness, and support for heterogeneous data models, since authoritative information often resides outside BIM authoring environments across all lifecycle phases.

Time-aware and real-time application support

Representing and managing time-dependent states, events, and changes is essential to support dynamic workflows such as construction sequencing, safety monitoring, asset condition tracking, and lifecycle performance analysis. Time-aware data support includes explicit temporal relationships, event modelling, and the capacity to accommodate incremental updates and streaming observations. Without these capabilities, IFC models remain difficult to use in real-time applications.

Web-friendly data exchange

Suitability for modern web-based and cloud-native toolchains depends on efficient exchange and integration mechanisms, including lightweight serializations, API-based access, partial data retrieval, and reduced parsing overhead. These capabilities are increasingly important for collaborative platforms, digital twins, and distributed applications, where file-based, monolithic transfers introduce performance and scalability bottlenecks.

Performance-oriented geometry representation

Efficient use in high-performance visualization and rendering pipelines – such as real-time viewers, simulation engines, and immersive environments – depends on geometry representations that are performant to parse, stream, and render. While semantic richness remains essential, excessive geometric complexity or tightly coupled semantics can hinder interactive performance. This need is especially relevant for large-scale models, for advanced construction monitoring and XR-enabled interactive applications across the lifecycle.

Comparative analysis of IFC capabilities

This section qualitatively evaluates how IFC2x3, IFC4, and IFC5 support the identified requirements. Using the four use cases as reference scenarios, it focuses on structural capabilities rather than tool implementations.

Data model extensibility

IFC2x3 exhibits limited extensibility due to its monolithic schema and rigid inheritance structure. Concepts beyond the design phase are therefore typically introduced through custom PSets and informal conventions rather than controlled domain evolution, leading to inconsistent semantics and weak validation. In cost and time management, entities such as *IfcCostItem* and *IfcTask* exist but offer limited expressiveness, often prompting ad-hoc extensions. In construction H&S, safety concepts (e.g., hazards) are not first-class entities and are commonly represented via custom PSets and linked documents, reducing interoperability. Similarly, in O&M, *IfcAsset* provides only a basic anchor and lacks the operational structures expected by CAFM/CMMS systems, while deconstruction workflows rely almost entirely on project-specific extensions due to minimal native support for circularity and LCA information.

IFC4 expands semantic coverage and improves conceptual clarity, reducing the need for customization in some areas, while largely preserving the same extensibility mechanisms based on PSets, classifications, and external links. As a result, fragmentation persists in domains that require structured, evolving semantics. Refinements in cost and time entities, standardized risk-related properties for H&S, and the introduction of entities for O&M represent incremental progress. Still, they do not fully support complex workflows such as work-order hierarchies or lifecycle-oriented extensions. Deconstruction-related information still depends heavily on custom PSets and external vocabularies (e.g., bSDD).

IFC5 explicitly addresses these limitations through a modular architecture that separates stable core concepts from domain-specific modules, enabling parallel and governed semantic evolution across lifecycle phases. While this improves the feasibility of dedicated extensions, it also introduces coordination challenges related to versioning and cross-module consistency. Moreover, while increased extensibility facilitates customization, it must be carefully managed to avoid a proliferation of project-specific data that could compromise interoperability.

A comparative overview of IFC extensibility across lifecycle phases is provided in Figure 2.

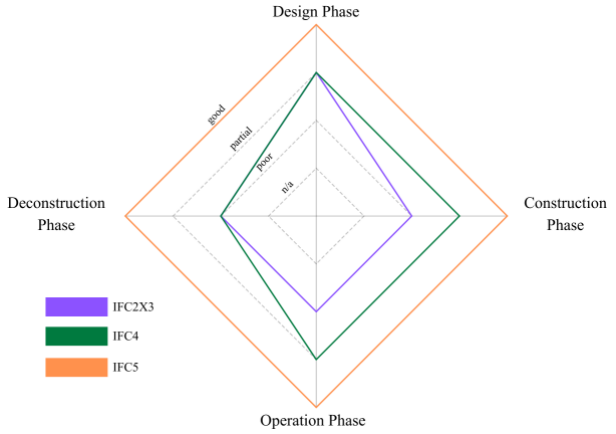


Figure 2: Data model extensibility chart

across data sources. The IFC integration with external datasets across lifecycle phases is provided in Figure 3.

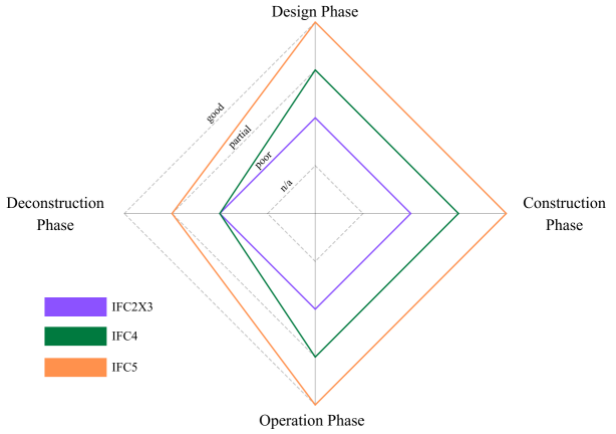


Figure 3: Integration with external datasets chart

Integration with external datasets

In IFC2x3, integration with external datasets is largely implicit and identifier-based, relying on GUID matching, classification codes, and file references, with limited semantic alignment. In cost estimation, this leads to static connections of costs, schedules, and physical elements, pushing 4D/5D workflows to external systems without mechanisms for consistency checking. In construction H&S, safety documentation and site observations remain outside the model and are merely referenced, limiting traceability, automated compliance checks, and reliable synchronization. Similarly, in O&M, basic reference constructs provide insufficient structure to integrate CMMS work orders, IoT streams, or maintenance data. Deconstruction workflows face comparable limitations, as native support for integration with Geographic Information Systems (GIS) and Environmental Product Declaration (EPD) datasets is largely absent.

IFC4 improves reference mechanisms and clarifies process-related semantics, enabling more explicit associations between IFC and external datasets such as cost databases, schedules, H&S checklists, and O&M resources. While this reduces ambiguity compared with IFC2x3, data typically remains external to the IFC model, and integration continues to rely on GUIDs. As a result, bidirectional synchronization with frequently updated sources, such as sensors, KPIs, or H&S issue lifecycles, remains weak and project-specific. Deconstruction scenarios still depend on external tools and additional semantic layers for GIS and EPD integration.

IFC5 explicitly targets heterogeneous integration through decoupled representations and web- and Linked-Data-friendly encodings (e.g., JSON/OWL), improving semantic interoperability with external services across domains. Alignment with OWL/RDF and Linked Data principles can support richer, machine-interpretable integration with ontologies (e.g., CI/SOSA/SSN), compliance repositories, CMMS, and issue-tracking systems. However, effective interoperability will ultimately depend on standardized mappings, governance frameworks, and robust identity and version management

Time-aware and real-time application support

In IFC2x3, support for time-aware and real-time applications is minimal and largely planning-oriented, with no effective constructions for dynamic events or state changes. As a result, cost and time management updates (e.g., schedule changes or actual versus planned values) are typically handled outside IFC, and 4D/5D workflows rely on external tools. In construction H&S, time-stamped safety events and evolving site conditions cannot be natively represented, limiting IFC's role to a static geometric reference. In O&M, *IfcTimeSeries* allows storage of time-dependent values but does not support event-driven transitions or maintenance histories, constraining native performance tracking. Circular planning workflows are similarly unsupported, as IFC2x3 provides no mechanisms for real-time monitoring or incremental change management.

IFC4 enhances scheduling semantics through richer process entities and relationships, enabling more explicit representation of planned timelines and dependencies. This supports improved cost and time planning and structured scheduling of H&S activities, such as inspections. However, the model remains essentially static: event-driven updates, state transitions, and continuous data flows are not natively supported. In O&M, although device entities are introduced, real-time data feeds and dynamic updates still depend on external platforms, reducing IFC4 to a reference role. Limitations also persist in circular planning, where incremental model changes and real-time visualization of evolving variants are not directly supported.

IFC5 introduces a conceptual shift toward time-aware and event-based representations, aligning more closely with digital-twin-oriented workflows. This could enable tighter coupling between schedules and changing quantities or costs, explicit modeling of temporal validity and evolving risk states in construction H&S, and dynamic asset states with predictive triggers in O&M. For deconstruction planning, event-based and incremental-change concepts could support more responsive variant tracking. However, their practical impact will depend on

specification choices, serialization, querying, synchronization, and performance in time-critical applications. Figure 4 summarizes the IFC time-aware and real-time application support across lifecycle phases.

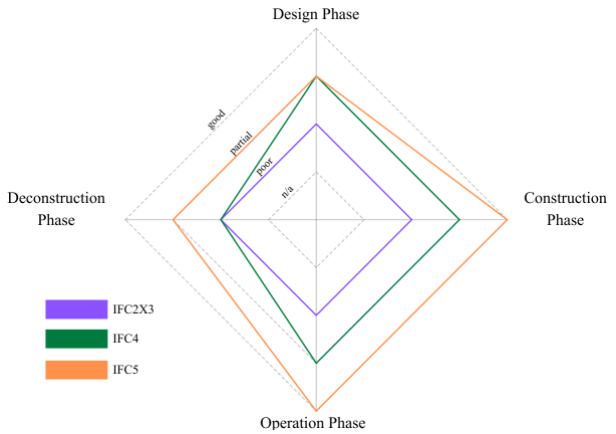


Figure 4: Time-aware and real-time application support chart

components and attributes. The trade-off is greater architectural complexity, requiring standardized interfaces and strong governance across distributed services. The outcomes are summarized in Figure 5.



Figure 5: Web-friendly data exchange chart

Web-friendly data exchange

IFC2x3 is fundamentally based on STEP and file-oriented exchange, resulting in large, monolithic datasets that are poorly suited to web-based workflows. This limits scalable collaboration in cost and time management, where querying, incremental updates, and API-driven interactions are weakly supported, and 4D/5D coordination typically relies on external tools and full-file transfers. In construction H&S, frequent dissemination of updated information requires repeated exchange of complete IFC files and linked documents, creating overhead for mobile and field use and weakening traceability. In O&M, routine cloud or mobile queries require full-file parsing, as spatial containment and relationships cannot be selectively accessed, making web-scale operations inefficient. Similar limitations affect deconstruction scenarios involving distributed stakeholders and reuse-oriented platforms.

IFC4 improves schema consistency and clarifies relationships but does not alter the underlying exchange paradigm. Web-based use cases, therefore, still depend on full file transfers or middleware layers. As a result, just-in-time access to relevant model subsets, such as elements affected by a task, safety issue, or cost update, remains difficult without external services. In O&M, despite clearer spatial constructs, the lack of native API patterns and partial extraction mechanisms continues to constrain cloud dashboards and mobile applications. Deconstruction workflows likewise remain poorly supported by monolithic exchange models.

IFC5 explicitly moves toward web-native exchange through lightweight, language-independent encodings that enable API-based access, partial data exchange, and incremental updates aligned with cloud-native toolchains. These capabilities support specific queries and updates in cost/time, fast publication of H&S changes, and REST-like access to assets and spaces for O&M applications. These capabilities suit distributed reuse and urban-mining workflows needing selective extraction and publication of

Performance-oriented geometry representation

IFC2x3 geometry prioritizes semantic completeness over real-time performance, which limits its suitability for interactive and visualization-intensive applications. In cost and time workflows, complex boundary representations and detailed solids often degrade performance in large models. In construction H&S, these limitations are particularly evident in AR/VR-based coordination and safety training, where low latency and stable frame rates are essential, requiring external conversion and simplification. In O&M, detailed geometric representations are excessive for dashboards, space management, and mobile wayfinding, while standardized lightweight proxies are not natively defined. Similar constraints hinder interactive visualization in deconstruction and reuse-oriented scenarios.

IFC4 improves geometric robustness and introduces tessellated representations, which better align with real-time rendering pipelines. However, performance-oriented geometry remains a secondary concern. Visualization workflows across cost/time analysis, H&S AR/VR applications, and O&M digital twins still depend on external optimization techniques or conversion to alternative formats. A key limitation is the tight coupling between semantics and geometry within product shape structures, which restricts selective extraction of lightweight, visualization-only subsets for streaming, incremental loading, and mobile or immersive use. Circular planning workflows face comparable constraints.

IFC5 explores a clearer separation between authoritative semantics and visualization-optimized geometry, with conceptual alignment to high-performance scene frameworks such as USD. This approach can support responsive model exploration in cost/time management, multiple levels of geometry tailored to AR/VR safety coordination, and simplified operational views for O&M applications consumed directly by web or AR tools. For circular design, it enables interactive evaluation of reuse scenarios. The main challenge lies in preserving semantic-

geometric coherence and traceability, ensuring that optimized visual assets remain reliably linked to authoritative IFC entities for auditing and regulatory purposes. A comparative overview of outcomes is provided in Figure 6.

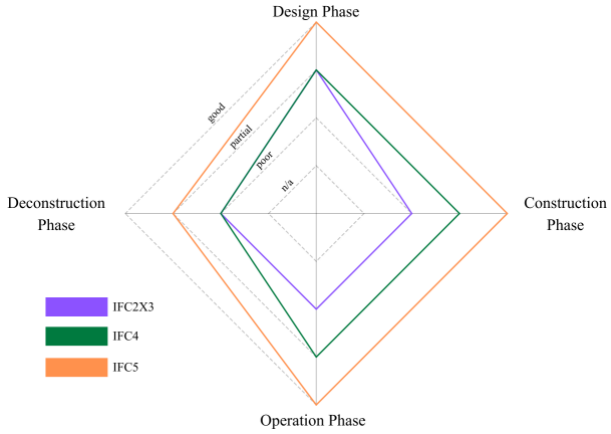


Figure 6: Performance-oriented geometry representation chart

Conclusion

This paper evaluates the current state of IFC2x3 and IFC4 and examines the future potential of IFC5 for data-centric integration across the building lifecycle. The comparative analysis highlights that IFC2x3 and IFC4 remain largely constrained by a monolithic, file-based exchange paradigm. Although IFC4 introduces notable semantic and geometric improvements, persistent limitations affect all lifecycle phases, particularly in operation and deconstruction. These include limited extensibility beyond geometry-centric models, reliance on identifier-based links to external data sources, and insufficient support for time-aware and real-time workflows.

The analysis of IFC5 indicates a strategic shift toward modular, web-aligned architectures, supporting multiple encodings and improved integration with contemporary toolchains. However, these advances redistribute rather than eliminate complexity, introducing new challenges related to governance, tooling maturity, identity management, and semantic consistency across distributed representations. Moreover, while increased extensibility facilitates customization, it must be carefully managed to avoid a proliferation of project-specific data that could undermine interoperability. Several critical capabilities also remain dependent on implementation choices.

The main limitations of the study comprise the lack of quantitative evaluation and empirical validation, as well as limited consideration of context-specific regulatory aspects and management implications. Future work will address these issues through quantitative metrics, real-world validation, and further investigation of governance and coordination aspects.

In conclusion, while IFC4 represents an essential evolution beyond IFC2x3, it does not fully address interoperability constraints. IFC5 offers a promising reorientation toward integration-oriented workflows, whose success will depend on robust governance frameworks, mature tooling, and standardized integration patterns beyond traditional file-based exchange.

References

Amo Larbi, J., Tang, L., Amo Larbi, R., Asante Abankwa, D. and Darko Danquah, R. (2024) Developing an Integrated Digital Delivery Framework and Workflow Guideline for Construction Safety Management in a Project Delivery System, doi:10.1016/j.ssci.2024.106486.

Allam, A. and Nik-Bakht, M. (2025) Knowledge Graph-Based Customer-Centric Deconstruction Assessment Model. vol. 6. European Council on Computing in Construction, pp. 0–0. doi:10.35490/EC3.2025.234.

Berlo, L. van, Krijnen, T., Tauscher, H., Liebich, T., Kranenburg, A. van, and Paasiala, P. (2021) Future of the Industry Foundation Classes: Towards IFC 5.

Biagini, C., Bongini, A. and Marzi, L. (2024) From BIM to Digital Twin. IOT Data Integration in Asset Management Platform. Journal of Information Technology in Construction (ITcon), 29 (49) doi:10.36680/j.itcon.2024.049.

Boje, C., Guerriero, A., Kubicki, S. and Rezgui, Y. (2020) Towards a Semantic Construction Digital Twin: Directions for Future Research. Automation in Construction. doi:10.1016/j.autcon.2020.103179.

Borrmann, A., König, M., Koch, C. and Beetz, J. (2018) Building Information Modeling: Why? What? How?: Technology Foundations and Industry Practice. doi:10.1007/978-3-319-92862-3_1.

Cassandro, J., Mirarchi, C., Zanchetta, C. and Pavan, A. (2024) Enhancing Accuracy in Cost Estimation: Structured Cost Data Integration and Model Validation. Journal of Information Technology in Construction (ITcon), 29 (58) December, pp. 1293–1325. doi:10.36680/j.itcon.2024.058.

Corneli, A., Carbonari, A., Spegni, F., Pieroni, T. and Naticchia, B. (2025) On-Site Safety Inspections Through Marker-Less Augmented Reality and Blockchain Notarization of BIM-Based Processes. Buildings. doi:10.3390/buildings15132318.

Farghaly, K., Abanda, F. H., Vidalakis, C. and Wood, G. (2018) Taxonomy for BIM and Asset Management Semantic Interoperability. Journal of Management in Engineering. doi:10.1061/(ASCE)ME.1943 - 5479.0000610.

Fürstenberg, D., Wikström, L., Hjelseth, E. and Lædre, O. (2021) Enabling Automation of BIM-Based Cost Estimation by Semantic Web Technology. October.

Getuli, V., Bruttini, A. and Rahimian, F. (2025) Parametric Design Methodology for Developing BIM Object Libraries in Construction Site Modeling. Automation in Construction, 170 February, p. 105897. doi:10.1016/j.autcon.2024.105897.

Getuli, V., Capone, P., Bruttini, A. and Isaac, S. (2020) BIM-Based Immersive Virtual Reality for Construction Workspace Planning: A Safety-Oriented Approach. Automation in Construction, 114 June, p. 103160. doi:10.1016/j.autcon.2020.103160.

Guntermann, L., Hagedorn, P. and König, M. (2025) Enhancing IFC Models for Circular Design: Validation of Required Information through an IDS Framework. International Symposium on Automation and Robotics in Construction (ISARC) Proceedings, 2025, Montreal, Canada, doi:10.22260/ISARC2025/0093.

Hagedorn, P., Donkers, A., Petrova, E. and König, M. (2025) Decentralised Data Exchange in Construction Data Spaces Using Information Containers. vol. 6. European Council on Computing in Construction. doi:10.35490/EC3.2025.301.

Hong, K. and Teizer, J. (2025) Digital Construction Site Layout Planning and Real-Time Trajectory Analysis for Proactive Safety Monitoring and Control of Struck-by Hazards. Automation in Construction, September. doi:10.1016/j.autcon.2025.106353.

Hu, R., Chen, K., Jiang, W. and Luo, H. (2024) IFC Data Extension for Real-Time Safety Monitoring of Automated Construction in High-Rise Building Projects. Automation in Construction, 162 June. doi:10.1016/j.autcon.2024.105408.

Jia, J., Ma, H. and Zhang, Z. (2024) Integration of Industry Foundation Classes and Ontology: Data, Applications, Modes, Challenges, and Opportunities. Buildings. doi:10.3390/buildings14040911.

Kim, S. and Kim, S.-A. (2023) A Design Support Tool Based on Building Information Modeling for Design for Deconstruction: A Graph-Based Deconstructability Assessment Approach. Journal of Cleaner Production. doi:10.1016/j.jclepro.2022.135343.

Laakso, M. and Kiviniemi, A. (2012) The IFC standard - a review of history, development, and standardization. Journal of Information Technology in Construction.

Li, Y., Zhao, Q., Yang, M., Ma, Z. and Hei, X. (2025) Advancements and Applications of Industry Foundation Classes Standards in Engineering: A Comprehensive Review. Buildings, 15 (16) August. doi:10.3390/buildings15162927.

Mantelatto, A., Njuguna, M., Kuzminykh, A., Parente, M., Granja, J. and Azenha, M. (2026) Circularity of Construction Materials: Precedence Between Elements Based on Digital Models for Demolition. In: Daoud, A. and Balázs, G. L. ed., Concrete and Circular Economy, 2026. doi:10.1007/978-3-032-06802-6_27.

Oliveira, A., Granja, J., Bolpagni, M., Motamedi, A. and Azenha, M. (2024) Development of Standard-Based Information Requirements for the Facility Management of a Canteen. Journal of Information Technology in Construction (ITcon). doi:10.36680/j.itcon.2024.014.

Otranto, R. B., Junior, G. M. and Pellanda, P. C. (2025) BIM-FM Integration through openBIM: Solutions for Interoperability towards Efficient Operations. Journal of Information Technology in Construction (ITcon). doi:10.36680/j.itcon.2025.012.

Patacas, J., Dawood, N., Greenwood, D. and Kassem, M. (2016) Supporting Building Owners and Facility Managers in the Validation and Visualisation of Asset Information Models (AIM) through Open Standards and Open Technologies. Journal of Information Technology in Construction, 21 November.

Ren, G., Huang, C. and Su, T. (2025) From Building Deliverables to Open Scene Description: A Pipeline for Lifecycle 3D Interoperability. Buildings, 15 (24) December. doi:10.3390/buildings15244503

Sacks, R., Eaastman, C., Lee, G. and Teicholz, P. (2018) BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors. New Jersey.

Shelden, D., Pauwels, P., Pishdad-Bozorgi, P. and Tang, S. (2020) Data Standards and Data Exchange for Construction 4.0. In: Sawhney, A., Riley, M. and Irizarry, J. ed., Construction 4.0. Taylor and Francis Ltd.. doi:10.1201/9780429398100-12.

Tomczak, A., Benghi, C., Berlo, L. van and Hjelseth, E. (2024) Requiring Circularity Data in BIM With Information Delivery Specification. Journal of Circular Economy. doi:10.55845/REJY5239.

Tomczak, A., Berlo, L. v, Krijnen, T., Borrmann, A. and Bolpagni, M. (2022) A Review of Methods to Specify Information Requirements in Digital Construction Projects. IOP Conference Series: Earth and Environmental Science, 1101 (9) November. doi:10.1088/1755-1315/1101/9/092024.

Wang, T. and Chen, H.-M. (2023) Integration of Building Information Modeling and Project Management in Construction Project Life Cycle. Automation in Construction, 150 June, p. 104832. doi:10.1016/j.autcon.2023.104832.

Yang, S.-W., Moon, S.-W., Jang, H., Choo, S. and Kim, S.-A. (2022) Parametric Method and Building Information Modeling-Based Cost Estimation Model for Construction Cost Prediction in Architectural Planning. Applied Sciences, 12 (19) September. doi:10.3390/app12199553.

Yu, Y., Kim, S., Jeon, H. and Koo, B. (2023) A Systematic Review of the Trends and Advances in IFC Schema Extensions for BIM Interoperability. Applied Sciences. doi:10.3390/app132312560.

Zhang, Z., Guo, B. H. W., Chang-Richards, A., Feng, Z., Jin, R., Zou, Y. and Miang Goh, Y. (2024) Digital Technology Enhanced Situation Awareness for Construction Safety: Systematic Review and Future Research Directions. August. doi:10.1016/j.ssci.2023.106280.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

SYMBOL RECOGNITION IN PIPING AND INSTRUMENTATION DIAGRAMS VIA MASK-AWARE TEMPLATE MATCHING WITH SPATIAL SIMILARITY

Juho Han[1], Miyoung Uhm[1], Heejun Youn[1], H. David Jeong[2], and Ghang Lee[1,3]
[1]Yonsei University, Seoul, Korea, Republic of South Korea
[2]Texas A&M University, College Station, TX, USA
[3]Technical University of Munich, Munich, Germany

Abstract

Automated symbol recognition in Piping and Instrumentation Diagrams (P&IDs) is challenging due to varying standards and complex structures where text, lines, and symbols are intermingled. This study proposes a non-learning-based framework applying extended template matching using the symbol legend. The framework extracts the Region of Interest (RoI) masks via the Convex Hull algorithm and performs matching using Intersection over Union (IoU) as a similarity metric to reflect spatial similarity. Subsequently, Intersection over Minimum (IoM)-based Non-Maximum Suppression (NMS) is introduced to suppress duplicate detections. Experiments on real-world industrial P&IDs achieved an F1 Score of 0.992, demonstrating its applicability to P&ID digitization.

Introduction

Piping and Instrumentation Diagrams (P&IDs) are fundamental engineering drawings in the plant industry, illustrating equipment, piping, and control systems along with their functional relationships. Serving as essential reference materials throughout the project lifecycle from design to operation, P&IDs utilize standardized symbols to represent critical information regarding process flow, control logic, and safety mechanisms (Kim et al., 2025).

P&ID reviews are conducted iteratively to ensure design accuracy, safety, and consistency. Although P&IDs are generated using Computer-Aided Design (CAD) tools where symbols are managed as blocks, design attributes are often not fully embedded within the drawing. Consequently, detailed specifications are managed in external documents rather than being intrinsically linked to the P&ID, creating challenges in reviewing and linking data (Kang et al., 2019; Koziolek and Koziolek, 2024).

In this environment, digitizing P&IDs for integration into downstream systems requires the preliminary process of manually identifying instruments, equipment, piping elements, and their interconnections. For instance, symbol label information must be matched with external documents such as equipment lists and datasheets. Most of these tasks rely on manual work by skilled engineers, rendering the process time-consuming and labor-intensive. Furthermore, this approach often leads to errors, omissions, and data inconsistencies, highlighting the necessity for automated symbol recognition (Li et al., 2025; Shteriyanov et al., 2025a).

However, the automation of P&ID symbol recognition faces several challenges. First, when CAD drawings are distributed as static images, block-based structures are lost, leaving symbols as a collection of simple geometric primitives without object-level semantics (Su et al., 2024). Second, as illustrated in Fig. 1, symbol geometries and labeling conventions vary significantly depending on the project, client, and design standards. General rule-based approaches or models trained on fixed datasets often fail to generalize across these diverse project-specific variations (Gupta et al., 2024).

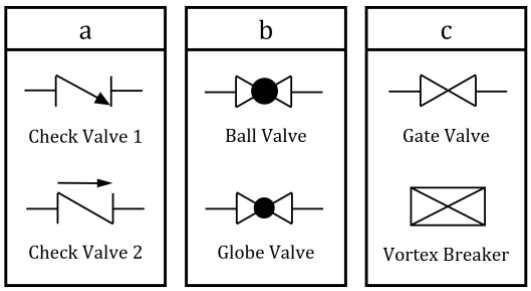


Figure 1: Examples of variations in P&ID symbol and labeling conventions: a) Different symbol geometries with identical labels; b) Identical symbol geometries with different labels; c) Similar symbol geometries with different labels.

To address this challenge, this study exploits symbol legend included in P&IDs. The symbol legend systematically defines the symbols, names, and codes used in a specific project, clearly presenting the project-specific symbol standards applied throughout the drawing set. In this study, symbols defined in the legend are constructed as symbol templates to serve as references for automated recognition. Based on this, we propose a template matching approach utilizing the Intersection over Union (IoU) metric.

First, symbol templates are constructed through preprocessing, and the Region of Interest (RoI) for the symbol is extracted using the Convex Hull algorithm. Subsequently, the symbol template is overlaid on the P&ID using a sliding window method to systematically scan the drawing. The similarity between the symbol template and the symbol candidate area (P&ID patch) is quantitatively evaluated using the IoU metric. If the similarity exceeds a predefined threshold, the instance is stored as a symbol candidate. Thereafter, to resolve the issue of duplicate detections, an Intersection over

Minimum (IoM)-based Non-Maximum Suppression (NMS) technique is applied to ensure a single, unambiguous detection result for each symbol instance.

This approach enables accurate symbol recognition in image-format drawings without requiring additional training data, while simultaneously allowing for the reflection of project-specific symbol standards. The main contributions of this study are summarized as follows:

- An IoU-based template matching framework is proposed to evaluate spatial similarity instead of pixel intensity, demonstrating superior performance in binary P&ID environments.
- The Convex Hull algorithm is utilized to extract precise RoI masks, thereby ensuring matching reliability against background noise.
- An IoM-based NMS is adopted to strictly and effectively filter duplicate detections in symbol-dense areas.

Background

Driven by advancements in AI, research on automated P&ID symbol recognition has primarily focused on approaches utilizing deep learning models and Visual Language Models (VLMs). These studies aim to automatically extract equipment, piping, and instrumentation information from P&IDs and convert them into structured digital data. However, existing studies have limitations, such as the prohibitive costs of constructing massive training datasets, high computational resource consumption, and compromised positional precision. To overcome these issues, this study proposes an improved template matching technique utilizing a RoI-based mask-aware matching strategy and the IoU spatial similarity metric.

Deep Learning Model for Symbol Recognition

Deep learning-based approaches offer the advantage of simultaneously predicting the location and class of symbols within an image, enabling robust recognition even in complex P&ID environments where numerous symbols coexist. For instance, Gada. (2021) applied transfer learning by fine-tuning various detection models pre-trained on MS-COCO dataset, such as Single-Shot Detection Network, Region-based Fully Convolutional Networks, and Faster Region-based Convolutional Neural Networks (Faster R-CNN), on a self-constructed P&ID dataset. Experimental results showed that Faster R-CNN achieved the best performance with an F1 Score of 66.1%. Additionally, Kim et al. (2021) generated a P&ID symbol dataset using the SmartPlant P&ID program and adopted RetinaNet, which leverages Focal Loss, significantly improving the detection rate for small or overlapping objects and achieving an F1 Score of 96.2%. In a related effort, Gajbhiye et al. (2023) constructed a dataset by segmenting P&IDs into square images of various sizes and labeling them,and performed symbol recognition using You Look Only Once (YOLO) v5 model trained on this dataset, achieving an F1 Score of 93.0%. Furthermore, Su et al. (2024) incorporated a Squeeze-and-Excitation attention mechanism into the YOLOv5 model to enhance feature representation, achieving an F1 Score of 93.9% through this approach.

Despite the strong representational capabilities of these models, deep learning-based approaches require extensive labeled datasets and iterative training cycles to achieve satisfactory performance. As illustrated in Fig. 2, building a symbol detection model necessitates an iterative process of dataset construction, model training, model evaluation, and model retraining, which imposes significant time and resource burdens in real-world industrial environments. This dependency limits their applicability in environments where design standards change frequently, and new symbols are continuously introduced.

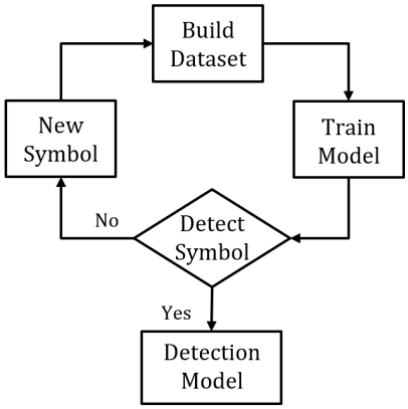


Figure 2: Building process of a deep learning model for symbol detection.

Visual Language Model for Symbol Recognition

Recently, P&ID recognition approaches based on VLMs have been proposed. Since VLMs can process images and text simultaneously, they can interpret symbols and text labels within P&IDs contextually. For example, Shteriyanov et al. (2025b) established a hybrid pipeline that interprets various formats of symbol legend without training through multi-modal prompting of VLMs and combines this with OCR-based drawing recognition results, recording 100% accuracy in legend classification and information extraction. Furthermore, Ghosh (2025) proposed a framework that utilizes VLMs as an automated quality evaluator for object detection results to verify the outputs of existing object detectors and identify missed objects. To overcome the limitations in the spatial reasoning capabilities of VLMs, the study combined visual, text, and geometric grounding, achieving an F1 Score of 87.6% in object identification.

While these studies demonstrate the potential of VLMs for P&ID interpretation, whether such results hold under the challenging visual conditions commonly encountered in real engineering drawings has not been sufficiently validated. In this regard, Shteriyanov et al. (2025c) introduced BlueprintSymVL, a benchmark for evaluating VLM symbol recognition. This benchmark is designed to isolate and test visual challenges inherent to the blueprint domain, such as symbol density, occlusion, and visual

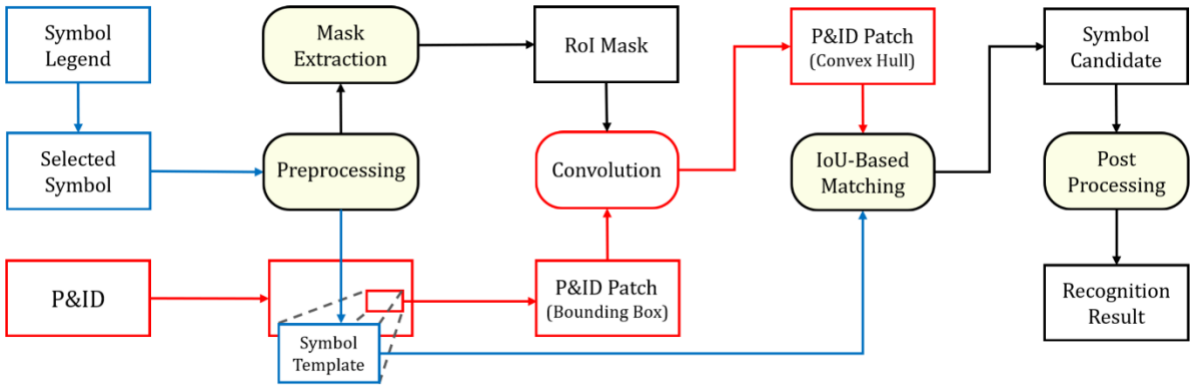


Figure 3: Flowchart of the proposed method. Blue and red paths represent symbol template generation and P&ID patch refinement, integrated for IoU-based symbol recognition.

similarity, and evaluates the visual generalization capability of models through a one-shot visual in-context querying strategy. An evaluation of four VLMs (GPT-4o, Gemini 2.5 Pro, InternVL 2.5 78B, and Qwen 2.5 VL 72B) showed that while the rate of simultaneously matching both symbol counts and labels reached up to 86% in simple scenarios, it dropped sharply to below 32% in dense conditions and below 38% in the presence of visually similar symbols.

In conclusion, VLM-based approaches can alleviate data scarcity as they handle diverse symbols without additional training. Nevertheless, these methods incur high computational costs due to image-based inputs, and recognition quality degrades substantially under such domain-specific challenges. Additionally, specifying the positional coordinates of symbols on P&IDs remains often infeasible with current VLM-based approaches.

Template Matching for Symbol Recognition

Before the emergence of deep learning model and VLMs, the most prevalent approach for symbol recognition was template matching (Kang et al., 2019). This method quantitatively calculates the similarity between a predefined template image and an input drawing to search for the location with the highest value. To quantify this degree of matching, various mathematical formulations based on pixel intensity distributions are employed as similarity metrics.

Representative similarity metrics include the Normalized Correlation Coefficient (TM-CCOEFF-NORMED), Normalized Cross Correlation (TM-CCORR-NORMED), and Normalized Squared Difference (TM-SQDIFF-NORMED) provided by the OpenCV library (OpenCV: Template Matching).

Specifically, the Normalized Correlation Coefficient normalizes the sum of the products of values after subtracting the mean from both the template and the image patch to a range of -1 to 1; a value closer to 1 indicates a perfect match, while a value closer to -1 implies an inverted match. Normalized Cross-Correlation normalizes the sum of the direct products of pixel values between the template and the image patch to a range of 0 to 1, where a value closer to 1 signifies a perfect match. Normalized Squared Difference normalizes the sum of squared differences between the pixel values of the template and the image patch to a range of 0 to 1, with a value closer to 0 indicating a perfect match.

Since these metrics evaluate similarity based on the statistical distribution or correlation of pixel intensities, they demonstrate robust matching performance in typical natural image environments characterized by illumination variations or noise. However, in P&ID environments characterized by binary properties (black and white pixels), these intensity-based metrics may not be intuitive. For instance, TM-CCOEFF-NORMED calculates similarity within a range of -1 to 1 based on the mean and deviation of pixel values. In binary images, there is a risk that such statistical operations may lead to an over-estimation of the score due to the dominance of matching background pixels, rather than directly representing the morphological similarity of the actual object.

Therefore, considering the binary characteristics of P&IDs, this study adopts IoU as the core metric for similarity measurement. IoU calculates the ratio of the intersection to the union between two areas, intuitively representing the degree of pixel overlap based on spatial occupancy rather than pixel values. Furthermore, for the purpose of this study, which requires verifying how accurately the P&ID patch refined by RoI aligns with the symbol template, the IoU-based approach, which imposes a strict penalty for morphological mismatch, is judged to be more suitable.

Methodology

Fig. 3 illustrates the overall workflow of the framework proposed in this study. The entire process consists of four stages: 1) Preprocessing and Symbol Template Construction, 2) RoI Mask Extraction, 3) IoU-Based Template Matching, and 4) IoM-Based NMS Postprocessing.

Preprocessing and Symbol Template Construction

To achieve symbol recognition in P&IDs using a template matching approach, it is essential to clearly define the target symbols and prepare standardized templates. In this study, to accurately reflect the actual symbol system used in the project, we referenced the symbol legend included in the P&IDs. Target symbols were selected based on their

frequency of occurrence and structural simplicity, and each symbol was extracted in an image format.

Since symbols in actual P&IDs do not maintain a form perfectly identical to the selected symbols, rotation, flipping, and scale adjustment were performed on the symbols prior to matching. However, certain symbols retain the same shape even after flipping or rotation. Since such cases reduce computational efficiency, rotation and flipping were performed first, followed by the removal of redundant variations.

Subsequently, scale adjustment was applied to prevent matching failures caused by minor size discrepancies. The symbol images processed in this manner were binarized to reduce background effects and ensure consistent similarity evaluation in IoU-based matching. Furthermore, optimal symbol templates for matching were constructed by applying a Bounding Box algorithm that computes the minimum enclosing rectangle of the symbol object to eliminate unnecessary margins. This series of preprocessing steps establishes a foundation for handling morphological variations in symbols without relying on learning-based approaches.

RoI Mask Extraction

Additionally, in this study, the Convex Hull algorithm was introduced to extract RoI mask to enhance the quality of template matching. The Convex Hull algorithm is a technique that computes the smallest convex polygon that tightly encloses all line segments and geometric shapes. By utilizing this, it is possible to define an optimal region that encompasses the entire symbol without being affected by complex internal structures or voids. As shown in Fig. 4, the extracted RoI is converted into a binary mask where the region of interest is assigned a value of 1 and the background is assigned a value of 0, enabling its direct utilization in the subsequent IoU-based template matching process.

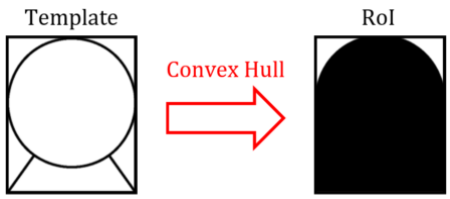


Figure 4: Example of RoI mask extraction results for Symbol Templates using the Convex Hull algorithm.

IoU-Based Template Matching

This study applies a template matching approach utilizing IoU as a similarity metric to evaluate the similarity between the preprocessed symbol templates and P&ID patches. Template matching is performed using a sliding window strategy, where each symbol template is moved with a fixed stride to overlap with the entire P&ID, and the IoU value between the symbol template and the underlying P&ID patch is calculated at each position.

The IoU metric is calculated as the ratio of the intersection area to the union area between the symbol template and the P&ID patch. This formulation enables a quantitative evaluation of pixel-level spatial overlap and ensures that higher similarity scores are assigned when the symbol position, scale, and shape are well-aligned. The IoU is defined as Eq. (1), where T denotes the symbol template and P represents the P&ID patch.

$$IoU = \frac{area(T \cap P)}{area(T) + area(P) - area(T \cap P)} \quad (1)$$

The calculated value of the IoU metric varies significantly depending on the margins of the symbol template. As the margins in the symbol template increase, more noise from the P&ID patch is included, leading to lower similarity values and, consequently, matching failure. Therefore, in this study, matching was performed through a more refined IoU calculation utilizing the RoI mask obtained in the previous step.

Specifically, the IoU between the symbol template and the P&ID patch is calculated by dividing the area of binary pixel intersection of the two regions by the area of binary pixel union of the two regions. Here, the area of union was calculated by summing the areas of individual regions and subtracting the area of intersection between the two regions. Specifically, the pixel area of the symbol template is calculated by summing all pixels constituting the symbol. The pixel area of the P&ID patch is calculated by summing the values obtained by convolution using the RoI mask as a kernel. Similarly, the intersection of the two regions can be calculated by summing the values obtained by convolving the result of applying the RoI mask to the symbol template and the P&ID patch. If the IoU value calculated through this process exceeds a predefined threshold, it is determined as a match and stored as a symbol candidate.

IoM-Based NMS Postprocessing

As a result of template matching, multiple duplicate symbol candidates are generated for a single actual symbol location. If such duplicate detections are not removed, multiple bounding boxes corresponding to the same symbol remain, thereby degrading the reliability of the recognition results. In the general object detection domain, this issue is commonly addressed using IoU-based NMS. However, since conventional IoU-based NMS relies on the union area of two bounding boxes to evaluate overlap, a problem arises where the IoU value is calculated to be low, resulting in failure to remove duplicates despite actual overlap, particularly when the difference in scale between two candidates is significant.

To address this limitation, this study applies an IoM-based NMS method. IoM is defined as the ratio of the intersection area between two candidate regions to the smaller of their areas. This definition allows for a high degree of redundancy to be assigned even when a smaller candidate is fully or partially contained within a larger one. This characteristic is particularly suitable for the P&ID environment, where physical overlap between symbols is rare, but inclusion relationships between candidates of different scales occur frequently. The IoM is calculated as Eq. (2), where A and B denote the

Bounding Boxes of two distinct Symbol Candidates, respectively.

$$IoM = \frac{area(A \cap B)}{\min\big(area(A),\ area(B)\big)} \quad (2)$$

Candidates are first sorted in descending order based on their IoU-based similarity scores, and the candidate with the highest similarity is selected. Subsequently, other candidates whose IoM values with the selected candidate exceed a predefined threshold are considered duplicates and removed. By repeating this process, a single final detection result can be derived for each symbol location.

Experiments

Experimental Setup

To evaluate the performance of the proposed RoI-based IoU template matching and IoM-based NMS approaches, this section details the template construction, preprocessing parameter settings, and experimental data composition. As shown in Fig. 5, a total of 16 basic symbol types were extracted from the symbol legend to construct the template. Template dimensions vary by symbol type, ranging from 150 × 80 pixels to 510 × 510 pixels. Additionally, rotation and flipping were applied to the basic templates, and duplicates were removed to produce 66 symbol templates. Scale transformations were then applied, resulting in a total of 198 symbol templates. The parameters used for preprocessing are summarized in Table 1. All experiments were conducted on a Windows-based local PC equipped with an Intel i7-11700KF CPU (3.60 GHz) and 64 GB of RAM, utilizing a total of 8 worker threads for matching computations.

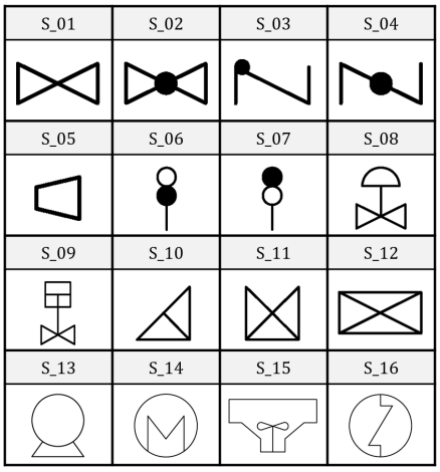


Figure 5: A total of 16 constructed Symbol Templates.

Table 1: Parameters used for preprocessing.

Parameters	Values
Rotataion	[0°, 90°, 180°, 270°]
Flippling	[True, Flase]
Scaling	[0.95, 1.0, 1.05]

Evaluation Dataset

To evaluate the practical applicability of the proposed method in industrial environments, four real-world plant P&IDs were utilized. These drawings were obtained from actual plant design projects and possess a high resolution of 19,867 × 14,034 pixels. Fig. 6 presents a representative example of the P&IDs used in this experiment, effectively illustrating the typical characteristics of real-world design drawings where various symbols, text, and fine connecting lines are intricately interwoven. Furthermore, the coexistence of symbols varying significantly in size and shape ranging from large equipment to small valves and instruments makes this dataset highly suitable for validating the robustness of the proposed method. In particular, small symbols such as valves and instruments often share similar shapes and tend to be densely clustered together, posing additional challenges for accurate recognition. These characteristics of the dataset play a crucial role in demonstrating the practical efficiency of the proposed method within real-world plant P&ID environments. The statistics of symbols contained in each drawing are summarized in Table 2.

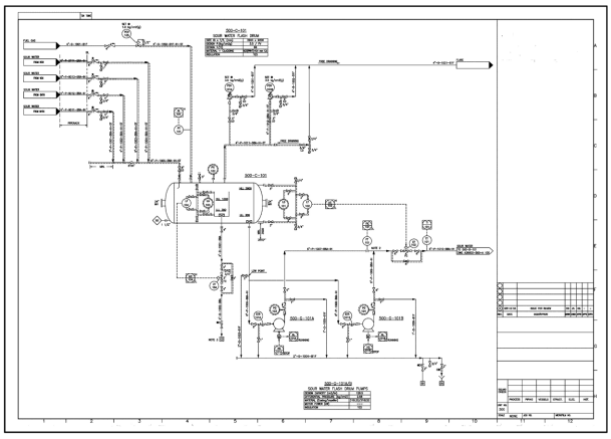


Figure 6: Example of P&ID (Drawing 1).

Tabel 2. Statistics of symbols contained in real-world plant P&IDs.

Symbol	s01	s02	s03	s04	s05	s06	s07	s08	s09	s10	s11	s12	s13	s14	s15	s16	**Total**
P&ID 1	61	3	7	0	14	1	6	3	0	2	0	2	2	0	0	0	**101**
P&ID 2	35	1	4	0	4	4	2	2	1	0	1	0	2	0	0	0	**56**
P&ID 3	23	1	2	2	6	1	2	1	0	1	0	0	2	1	1	2	**45**
P&ID 4	87	8	3	0	32	1	4	7	1	4	0	1	2	0	1	0	**151**
Total	**206**	**13**	**16**	**2**	**56**	**7**	**13**	**13**	**2**	**7**	**1**	**3**	**8**	**1**	**2**	**2**	**353**

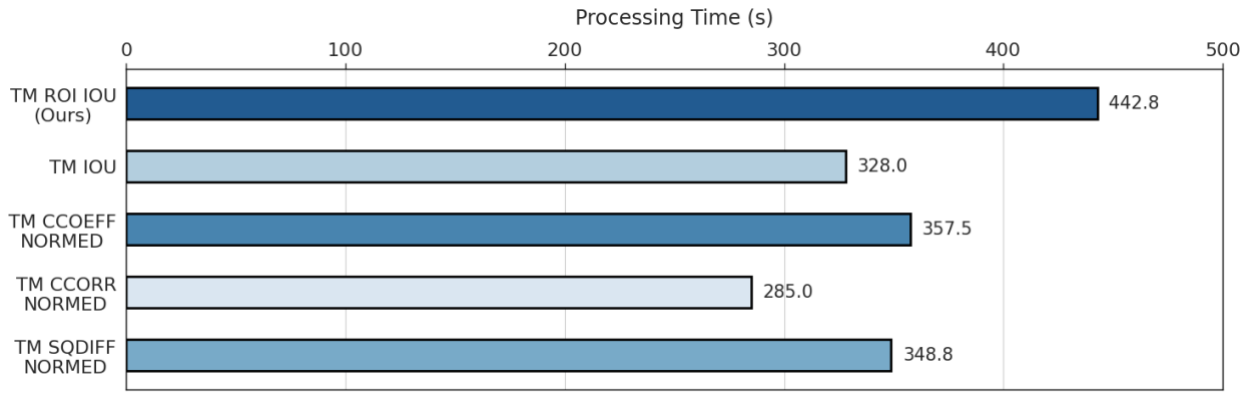


Figure 7: Processing time of each algorithm for a single P&ID.

Evaluation Metrics

In this study, to quantitatively evaluate the matching performance of the proposed symbol recognition framework, the F1 score, which is the harmonic mean of Precision and Recall, was selected as the primary evaluation metric.

First, the validity of symbol recognition is classified based on the correspondence between the predicted symbol class and the Ground Truth. Cases where the predicted symbol class matches the ground truth are considered True Positives (TP). Conversely, cases where the predicted class differs from the actual class, or where a background region containing no symbol is misidentified as a symbol, are defined as False Positives (FP). Cases where an existing symbol is not detected are classified as False Negatives (FN).

In environments such as the dataset used in this experiment, where severe class imbalance exists or where the risks of False Positives and False Negatives coexist due to complex backgrounds, it is difficult to represent the comprehensive performance of the model using only single metrics like Precision or Recall. Therefore, this study adopted the F1 score as the definitive performance benchmark to objectively assess the model's robustness by balancing these two factors. The definition of the F1 score is as shown in Eq. (3).

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP + FP + FN} \quad (3)$$

Experimental Results

This experiment conducted a comparative analysis of algorithm speed and matching quality among the conventional template matching methods provided by OpenCV, the standard IoU matching method (TM-IOU), and the RoI-based IoU matching method (TM-ROI-IOU).

First, the processing times of each algorithm were compared. Fig. 7 compares the time required to apply 198 preprocessed templates to a single P&ID. As shown in the graph, TM-CCORR-NORMED was the fastest, recording 328.03 seconds. In contrast, the TM-ROI-IOU proposed in this study was the slowest at 442.78 seconds, taking approximately 35% longer than TM-CCORR-NORMED. This increase in computational cost is primarily attributed to the additional pixel-level convolution operations required to apply RoI masks during the matching process.

Prior to evaluating matching quality, score interpretations were standardized to ensure a fair comparison across algorithms employing different similarity metrics. Specifically, for TM-SQDIFF-NORMED, where a value of 0 indicates a perfect match, the matching criterion was inverted (1 - value) to align with other metrics where higher values denote higher similarity. Under this standardized framework, the F1 score was assessed by increasing the matching threshold from 0.50 to 0.95 in increments of 0.05, as summarized in Table 3. Consequently, the TM-ROI-IOU demonstrated superior performance, achieving a peak F1 score of 0.992 at a threshold of 0.85.

Tabel 3. F1-Score of each algorithm according to threshold changes.

Threshold	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95
TM ROI IOU (Ours)	0.443	0.743	0.850	0.961	0.978	0.984	0.984	**0.992**	0.907	0.393
TM IOU	0.838	0.864	0.871	0.974	0.978	**0.980**	0.971	0.957	0.831	0.308
TM CCOEFF NORMED	0.358	0.373	0.410	0.503	0.934	0.978	**0.980**	0.971	0.949	0.572
TM CCORR NORMED	0.348	0.372	0.379	0.403	0.494	0.899	0.961	**0.977**	0.964	0.788
TM SQDIFF NORMED	0.868	0.936	0.977	0.977	**0.977**	0.976	0.961	0.946	0.788	0.295

Discussion

Experimental results showed that the proposed template matching framework entails a higher computational cost than existing methods. This is attributed to the inclusion of the RoI mask extraction process to define valid regions within P&ID patches, as well as the subsequent convolution operations required throughout the matching process. Nevertheless, it achieves superior matching performance in terms of symbol recognition quality and reliability. Furthermore, the proposed method maintained stable performance even in real-world plant P&ID environments characterized by high resolution, a wide range of symbol sizes, dense arrangements, and complex piping structures. Notably, the improved results of RoI-based IoU matching over the non-RoI approach confirm that eliminating margins and minimizing unnecessary background calculations effectively mitigates missed detections caused by background noise. Fig. 8 illustrates examples of correctly recognized symbols in complex environments using the RoI-based IoU matching method.

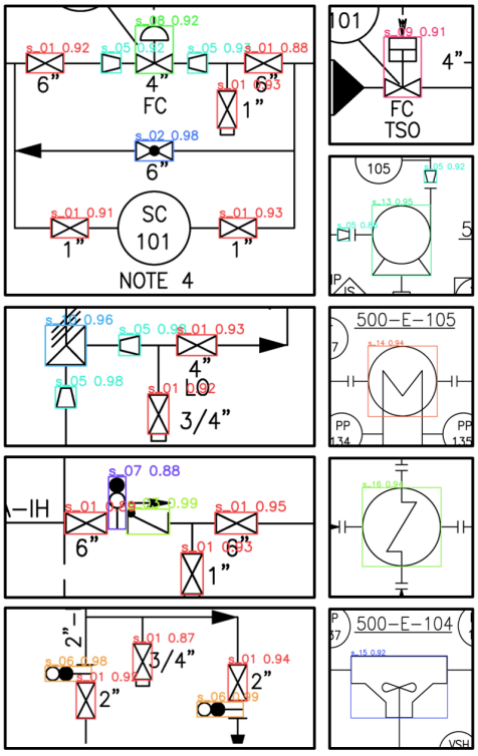


Figure 8: Examples of successful symbol recognition.

Although the proposed method successfully recognized the majority of symbols, limitations resulting in false positives and false negatives were observed for a minority of symbols. These failure cases can be categorized into three main types, as shown in Fig. 9.

First, there were cases where existing symbols were missed because their calculated similarity scores failed to exceed the threshold. For example, in the case of s_08, the calculated similarity fluctuated around the set threshold (0.85), failing to provide consistent matching results. Second, detection failures occurred when symbol deformation exceeded the scope of transformations covered in this study, such as rotation, flipping, or scaling. For instance, s_12 was situated within the boundaries of a specific equipment unit, causing its aspect ratio to change in accordance with the equipment's shape. Such geometric distortions could not be compensated for by the preprocessing steps of rotation or scale adjustment, rendering detection impossible. Third, false recognitions arose from inclusion relationships between symbols, where a larger symbol was misidentified as a smaller one. For example, s_09 was misclassified as s_01, and s_11 as s_10. This error is attributed to the connecting lines and piping attached to the larger symbol being included in the similarity calculation, which resulted in the smaller symbol (or a sub-part) yielding a higher similarity score than the actual larger symbol.

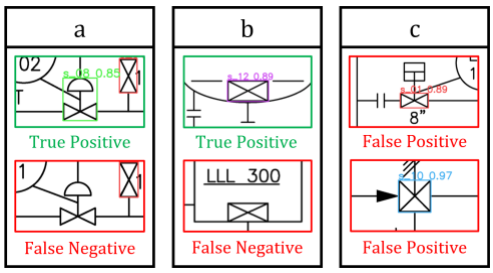


Figure 9: Examples of failed symbol recognition: a) Missed detection due to high threshold; b) Missed detection due to severe symbol deformation; c) Incorrect recognition of symbols in inclusion relationships.

Conclusion

This study addressed the industrial challenge of automated symbol recognition as a key method for reviewing and digitizing image-based P&IDs. To overcome the issues of semantic loss due to the conversion of P&IDs into image formats and the project-specific variations in symbol shapes and meanings, an extended template matching framework based on the symbol legend was proposed. The proposed method enables symbol recognition without the need for dataset construction or model training by directly leveraging the symbol legend defined in each project. Specifically, the detection process was refined by extracting RoI masks using the Convex Hull algorithm, and IoU-based matching with IoM-based NMS was introduced to provide a quantitatively interpretable and reliable recognition mechanism.

Comparative experiments demonstrated that the proposed framework is more effective than conventional template matching methods in real-world P&ID environments characterized by densely arranged symbols with similar shapes and complex piping structures. These results are significant in that they demonstrate that the proposed approach can serve as a robust symbol recognition technique for P&ID digitization in the field of engineering drawing analysis, which often suffers from data scarcity.

Despite these contributions, this study has several limitations. First, constructing symbol templates from the symbol legend involves labor-intensive manual effort, and symbols not included in the legend cannot be

recognized. Second, the generalization of the method is limited because rotation, flip and scale transformations are applied only within predefined parameters. Third, the system is sensitive to hyperparameter settings; if an appropriate matching threshold is not established, the system may fail to detect target symbols. Fourth, the sliding window approach incurs high computational costs, making real-time processing difficult. Finally, the reliability of the F1 Score evaluation for specific symbol classes is limited due to severe class imbalance in the P&ID dataset used for verification.

Future research will focus on addressing the limitations identified in the experiments. First, false recognitions arising from inclusion relationships are to be resolved by prioritizing labeling for larger symbols during the preprocessing stage. Additionally, more consistent and robust matching will be enabled by incorporating geometric similarity metrics for symbol shapes. Moreover, computational efficiency will be improved by exploring algorithm optimization and GPU-based parallel processing. Furthermore, an automated symbol template construction model will be developed from the symbol legend to alleviate the burden of manual effort. Finally, the reliability of results will be enhanced by applying this framework to a wider variety of industrial P&IDs.

Through these efforts, the proposed framework is expected to extend beyond symbol recognition itself and support practical engineering tasks in the plant industry, without requiring extensive training data or domain-specific model retraining. For instance, the recognized symbol types and their spatial arrangements can be cross-referenced with design rules to detect violations or linked with equipment datasheets to support P&ID quality reviews. Furthermore, comparing recognition results between revisions can enable systematic management of design change histories. Ultimately, this study is anticipated to contribute to accelerating the digital transformation of the plant industry by bridging the gap between legacy engineering documents and intelligent information systems.

Acknowledgments

This work was supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (No. RS-2024-00407028) and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. RS-2025-25432239).

References

Gada, M. (2021) Object detection for P&ID images using various deep learning techniques. In: 2021 International Conference on Computer Communication and Informatics (ICCCI). pp.1–5. IEEE.

Gajbhiye, S.M., Bhamre, S.R., Tadepalli, L.T., Pillai, M.R. & Uplaonkar, D. (2023) Advancing P&ID digitization with YOLOv5. In: 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS). pp.1–6. IEEE.

Ghosh, S. (2025) Visual Language Model as a Judge for Object Detection in Industrial Diagrams. arXiv preprint arXiv:2510.03376.

Gupta, M., Wei, C., & Czerniawski, T. (2024). Semi-supervised symbol detection for piping and instrumentation drawings. Automation in Construction, 159, 105260.

Kang, S. O., Lee, E. B., & Baek, H. K. (2019). A digitization and conversion tool for imaged drawings to intelligent piping and instrumentation diagrams (P&ID). Energies, 12(13), 2593.

Kim, H., Lee, W., Kim, M., Moon, Y., Lee, T., Cho, M., & Mun, D. (2021). Deep-learning-based recognition of symbols and texts at an industrially applicable level from images of high-density piping and instrumentation diagrams. Expert Systems with Applications, 183, 115337.

Kim, J. B., Moon, Y., Han, S. T., & Mun, D. (2025). Automated inspection of P&ID object recognition using deep learning. Scientific reports, 15(1), 39031.

Koziolek, H. & Koziolek, A. (2024) LLM-based control code generation using image recognition. In: Proceedings of the 1st International Workshop on Large Language Models for Code. pp.38–45.

Li, Z., Zhao, J., Zhou, H. & Hu, X. (2025) A novel symbol recognition framework for digitization of piping and instrumentation diagrams. In: The 35th European Symposium on Computer Aided Process Engineering. Ghent, Belgium. pp.1800–1805.

OpenCV (n.d.) Template matching. [Online] Available from: https://docs.opencv.org/4.x/d4/dc6/tutorial_py_template_matching.html [Accessed 29 January 2026].

Shteriyanov, V., Dzhusupova, R., Bosch, J. & Olsson, H.H. (2025a) Automating the expansion of instrument typicals in piping and instrumentation diagrams (P&IDs). In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 39, No. 28, pp.28885–28891.

Shteriyanov, V., Dzhusupova, R., Bosch, J. & Olsson, H.H. (2025b) Enhancing OCR-based engineering diagram analysis by integrating diverse external legends with VLMs. Journal of Software: Evolution and Process, 37(12), e70072.

Shteriyanov, V., Dzhusupova, R., Bosch, J. & Olsson, H.H. (2025c) BlueprintSymVL: A discriminative benchmark for VLM symbol recognition in engineering blueprints. Results in Engineering, 108171.

Su, G., Zhao, S., Li, T., Liu, S., Li, Y., Zhao, G. & Li, Z. (2024) Image format pipeline and instrument diagram recognition method based on deep learning. Biomimetic Intelligence and Robotics, 4(1), 100142.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A COMPARATIVE ANALYSIS OF IoT DATA PIPELINES FOR REAL-TIME MIXED REALITY VISUALISATION IN UNITY-BASED DIGITAL TWINS

Muhammad Shahzad[1], Joseph H. M. Tah[1], Avar Almukhtar[1], and Muhammad Younas[1]
[1]Oxford Brookes University, Oxford, United Kingdom

Abstract

Digital Twins increasingly rely on real-time IoT data, yet the end-to-end architectural implications of alternative integration approaches remain underexplored. This paper presents a controlled comparative analysis of three IoT data pipelines connecting The Things Network (TTN) to a Unity-based Digital Twin for mixed reality visualisation: cloud-centric pipelines using Amazon Web Services and Microsoft Azure, and a lightweight publish–subscribe pipeline based on direct MQTT delivery. By isolating the data integration layer, a benchmarking approach evaluates data transmission behaviour, persistence, integration complexity, scalability, and cost. The findings provide practical architectural insight into how pipeline design choices shape Digital Twin behaviour.

Introduction

Digital Twins are increasingly adopted in the built environment to integrate physical assets, sensing systems, and digital representations in support of monitoring, operation, and maintenance activities. As originally articulated by Grieves and Vickers (2016) and subsequently extended to construction and infrastructure contexts by Boje et al. (2020), Digital Twins enable continuous coupling between physical systems and their digital counterparts. By linking real-world measurements with digital models, they provide situational awareness. They also support informed decision-making across the lifecycle of buildings and infrastructure (Boschert and Rosen, 2016; Lu et al., 2020; Shahzad et al., 2025).

A fundamental requirement for operational Digital Twins is the reliable and continuous integration of real-time data from distributed sensors. Internet of Things (IoT) and related sensor technologies have widely deployed in buildings to capture environmental and operational parameters relevant to performance assessment and maintenance (El-Ammari and Hammad, 2019; Al-Nasser et al., 2024). At the network layer, Low-Power Wide-Area Network (LPWAN) technologies such as LoRaWAN enable scalable sensor deployments with low energy consumption (Centenaro et al., 2016; LoRa Alliance, 2025). However, as emphasised by Wang et al. (2013) and Tao et al. (2018), sensor deployment alone does not constitute a functional Digital Twin. Real-time data must be transmitted, processed, and delivered through an integration pipeline that is sufficiently robust, scalable, and responsive for operational use.

Immersive Digital Twin applications are increasingly reported in construction and facility management research. However, existing studies largely foreground application-level outcomes such as training, inspection, or collaboration, while treating the underlying IoT data pipeline as an implementation detail (Palmarini et al., 2018). In practice, IoT-enabled Digital Twin systems are commonly realised through either cloud-centric architectures, which provide managed ingestion, storage, and dashboard services, or lightweight publish–subscribe approaches that prioritise low-latency data delivery. Cloud platforms such as Amazon Web Services (AWS) and Microsoft Azure are frequently adopted for Digital Twin deployments due to their scalability and managed services (Lu et al., 2020; Al-Nasser et al., 2024), while messaging protocols such as MQTT are widely used for real-time IoT streaming (Thangavel et al., 2014).

At the platform level, several studies have presented comparative analyses of major cloud providers, most notably AWS and Azure, focusing on pricing models, security mechanisms, quality-of-service attributes, and service capabilities (Kotas et al., 2018; Eisa et al., 2020; Palumbo et al., 2021; Hameed et al., 2025). While these contributions provide valuable insight into cloud platform characteristics, they are typically conducted from a general cloud computing or enterprise systems perspective. As noted by Palumbo et al. (2021) and Eisa et al. (2020), such comparisons often rely on abstract performance metrics or simulation-based evaluation and do not examine how alternative cloud architectures behave when embedded within end-to-end Digital Twin data pipelines. Moreover, controlled comparisons that trace IoT data flow from sensing infrastructure to immersive visualisation environments under identical conditions remain limited.

This paper addresses this gap through a controlled comparative analysis of three end-to-end IoT data pipeline architectures that deliver live sensor readings from The Things Network (TTN) to a Unity-based Digital Twin. TTN is an open, LoRaWAN-based IoT network providing device management and standardised data forwarding services (TTN, 2023). The evaluated architectures comprise: (i) cloud-centric pipelines implemented using AWS and Microsoft Azure, providing

managed ingestion, serverless processing, and persistent storage (AWS, 2024a, Microsoft, 2024); and (ii) a lightweight publish–subscribe integration path in which Unity subscribes directly to sensor topics via an MQTT broker, enabling low-latency data delivery.

The conceptual scope of the study is illustrated in Figure 1, which positions the data pipeline layer as an intermediary between the IoT sensing infrastructure and the Digital Twin visualisation environment. Within this framework, the case study environment, TTN configuration, and Unity-based Digital Twin are held constant, and only the data pipeline architecture between TTN and Unity is varied. Mixed Reality (MR) and Virtual Reality (VR) are employed as representative high-demand real-time visualisation contexts such as live walkthroughs, remote asset inspection, and collaborative review where low latency, continuous updates, and immediate user feedback are critical for effective decision-making rather than as the primary objects of evaluation.

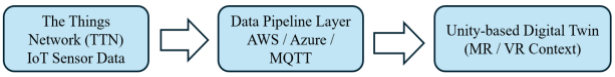


Figure 1: Conceptual Scope of the Study

The contribution of this work is threefold. First, it documents the implementation of three TTN-to-Unity data pipelines under identical case study conditions, enabling controlled architectural comparison. Second, it applies a structured benchmarking approach to evaluate IoT data pipelines across data transmission behaviour, platform capabilities, data persistence, Unity integration complexity, and architectural maintainability. Third, it synthesises the observed trade-offs to support evidence-informed selection of IoT data pipeline architectures for real-time Digital Twin visualisation in the built environment.

Case Study and Experimental Context

Case Study Overview

The study is based on an operational building used for teaching and research activities, instrumented with six IoT sensors for continuous environmental monitoring. The building provides a realistic yet controlled context for evaluating alternative IoT data pipeline architectures under normal operational conditions, without introducing domain-specific application bias.

Environmental monitoring was performed using Elsys CO_2 sensors deployed exclusively within the shared office space of the case study building. The sensors measure temperature, relative humidity, and CO_2 concentration. These are widely accepted indicators of indoor environmental quality in office environments and representative of sensing modalities commonly used in building monitoring and facilities management. The sensors operate on a low-power basis and are suitable for continuous, long-term deployment. Figure 2 illustrates the spatial distribution of the six deployed sensors within the shared office space, which provides the physical context for mapping measured conditions to corresponding locations in the Unity-based Digital Twin.

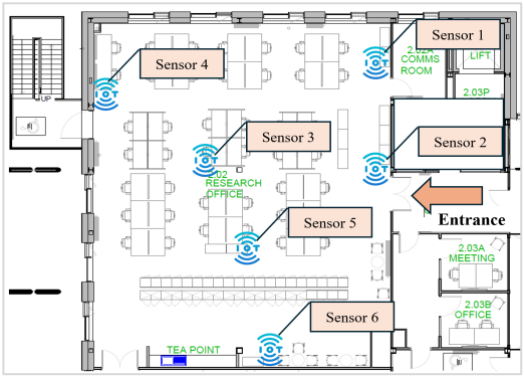


Figure 2 – Spatial deployment of IoT sensors within the case study building

All sensors communicated using a Low-Power Wide-Area Network (LPWAN) based on the LoRaWAN protocol. Sensor uplinks were received and managed by The Things Network (TTN), which served as the common IoT network layer across all experimental configurations. TTN was selected due to its widespread adoption, support for LoRaWAN devices, and ability to forward sensor data through standardised integration mechanisms. Importantly, TTN was already operational within the case study environment, ensuring a stable and continuously functioning network infrastructure and avoiding additional variability associated with network setup.

At the visualisation layer, sensor data were integrated into a Unity-based Digital Twin representing the monitored building. Sensor readings were mapped to their corresponding spatial locations within a three-dimensional model and updated dynamically in response to incoming data. MR and VR devices were used as representative high-demand visualisation contexts, where lag-free interaction and continuous state synchronisation are essential, while the visualisation logic and interaction mechanisms were kept constant throughout the study.

Controlled Comparison Strategy

To enable controlled architectural comparison, the physical sensing infrastructure, including sensor types, deployment locations, payload structure, and data transmission frequency, was held constant across all experiments. The sensor data transmission interval was configured at 10 minutes for all devices, ensuring consistent temporal resolution across all pipeline evaluations. The TTN configuration remained unchanged, ensuring identical data semantics and network behaviour for each pipeline.

The Unity-based Digital Twin, including the 3D model, data mapping logic, and update routines, was also held constant across all experimental scenarios. As shown in Figure 3, this setup fixes both the upstream sensing and networking layer and the downstream visualisation component. The data integration layer is isolated as the sole variable element.

All pipelines were implemented and executed in the same computing environment, using identical hardware and software configuration, to ensure that observed differences were not influenced by variations in computational resources or execution conditions.

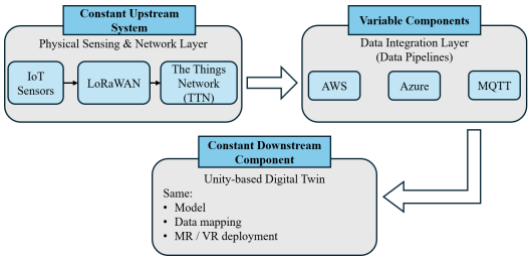


Figure 3 – Controlled Experimental Setup

Within this controlled configuration, three data pipeline architectures were implemented and evaluated: cloud-centric pipelines using AWS and Azure, and a lightweight publish–subscribe pipeline using an MQTT broker. This design ensures that observed differences in system behaviour, complexity, and capability can be attributed directly to the data pipeline architecture rather than to sensing, networking, or visualisation factors.

IoT Data Pipeline Architectures

This section describes the IoT data pipeline architectures implemented and evaluated in the study. All architectures connect the same upstream IoT network to the same Unity-based Digital Twin, enabling controlled comparison of alternative data integration approaches. A common baseline is provided by The Things Network (TTN), followed by three downstream pipelines that differ only in how sensor data are processed, stored, and delivered.

Common Baseline: TTN Data Flow

All sensor data originate from distributed IoT devices deployed within the case study building and are transmitted using the LoRaWAN protocol. Uplink messages are received and managed by The Things Network (TTN), which serves as the common IoT network layer across all experimental configurations.

Sensor uplinks are encoded as structured JSON payloads containing timestamped measurements and device identifiers. Payload structure and transmission frequency remain identical across all pipelines. TTN functions exclusively as a data forwarding layer, providing webhook- and MQTT-based interfaces to downstream systems without performing data processing or persistence. This baseline data flow configuration ensures that observed differences arise from downstream pipeline architectures rather than sensing or network variability.

Cloud-Centric Pipelines: TTN–AWS–Unity and TTN–Azure–Unity

The two TTN–AWS–Unity and TTN–Azure–Unity pipelines follow a cloud-centric architectural pattern designed to support persistence and structured data access. Sensor data are forwarded from TTN via webhooks to cloud ingestion endpoints, triggering serverless processing functions that perform lightweight payload parsing and validation.

Processed data are stored in managed cloud databases, enabling retention of both recent and historical sensor readings. An API layer exposes stored data to the Unity-based Digital Twin via HTTP-based requests. Unity retrieves updated sensor data at a fixed polling interval of 10 seconds, ensuring regular synchronisation between the cloud database and the visualisation environment. This pull-based access model decouples ingestion from consumption. However, it introduces multiple intermediary components between the IoT network and the visualisation layer.

Architecturally, the AWS- and Azure-based pipelines exhibit functional parity. Differences observed between them relate primarily to platform-specific tooling and configuration workflows rather than fundamental architectural design.

Streaming-Based Pipeline: TTN–MQTT Broker–Unity

The TTN–MQTT broker–Unity pipeline represents a lightweight, streaming-based architecture based on a publish–subscribe communication model. Sensor data are forwarded from TTN to an MQTT broker, from which the Unity application subscribes directly to relevant topics.

Data delivery is push-based and occurs as messages are published, enabling near real-time updates. Unlike the cloud-centric pipelines, this architecture introduces no intermediary processing or inherent data persistence. Sensor readings are therefore transient unless additional storage is explicitly introduced. This design prioritises immediacy and simplicity while shifting responsibility for data handling and temporal context to the consuming application.

Architectural Abstraction

Figure 4 presents an abstracted comparison of the two dominant pipeline paradigms evaluated: cloud-centric and streaming-based pipelines.

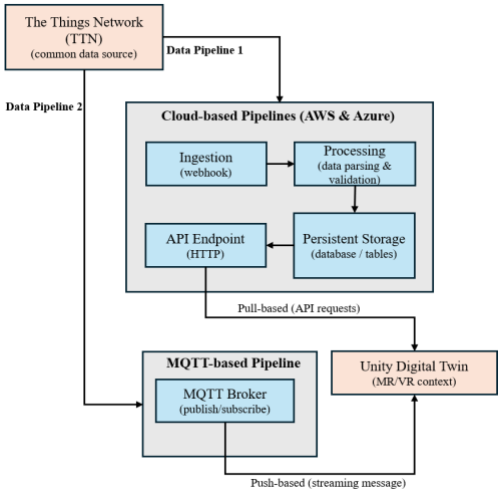


Figure 4 - Abstracted IoT data pipeline architectures connecting TTN to a Unity-based Digital Twin

The figure highlights differences in component composition, data flow, and data persistence while omitting platform-specific services. This abstraction provides architectural context for the benchmark evaluation presented in the following section.

Benchmarking Approach

The comparative evaluation in this paper adopts a structured benchmarking approach to assess alternative IoT data pipeline architectures supporting real-time Digital Twin visualisation. Consistent with established cloud and systems benchmarking practice, complex platforms are evaluated using common architectural and operational criteria rather than vendor-specific performance ranking (Kotas et al., 2018; Eisa et al., 2020). Accordingly, this study focuses on architectural behaviour and practical integration characteristics observed during implementation and deployment.

Benchmark Design Rationale

The benchmarking criteria were developed according to three principles: platform agnosticism, architectural relevance, and implementation grounding. First, the criteria are platform-agnostic, focusing on architectural and functional properties that can be compared across cloud-managed and publish–subscribe approaches rather than vendor-specific services. This aligns with established cloud benchmarking literature, where heterogeneous platforms are normalised into comparable capability and quality dimensions (Kotas et al., 2018; Eisa et al., 2020), and with more recent work emphasising architectural capability over provider-specific implementation detail (Hameed et al., 2025). Palumbo et al. (2021) further demonstrate the importance of vendor-neutral analysis when assessing end-to-end latency behaviour.

Second, the criteria are architecture-oriented, emphasising data flow, component responsibilities, and coupling between ingestion, persistence, and application access layers. This reflects evidence that system-level properties such as latency, responsiveness, and integration overhead emerge primarily from architectural composition rather than from individual services in isolation (Palumbo et al., 2021).

Third, the criteria are deployment-informed, grounded in deploying and troubleshooting multiple TTN-to-Unity pipelines in a real building context. While protocol-level studies quantify MQTT performance under controlled conditions (Thangavel et al., 2014; Ansyah et al., 2023), they do not capture the integration and operational challenges encountered in end-to-end Digital Twin deployments. Incorporating factors such as configuration effort, tooling complexity, and operational usability therefore extends existing benchmarking work toward application-facing Digital Twin systems.

Each pipeline was implemented and tested repeatedly under identical conditions. Benchmark assessments were based on direct observation of system behaviour during deployment, configuration, and runtime operation. Comparative evaluation was conducted using consistent criteria across all pipelines, ensuring that observed differences reflect architectural characteristics rather than experimental variability.

In this study, latency is defined as the elapsed time between sensor data transmission from TTN and its availability within the Unity-based Digital Twin. Based on repeated implementation observations, the MQTT based pipeline exhibited near real-time updates (typically within 1-2 seconds), whereas cloud-based pipelines introduced additional delay (approximately 5-8 seconds) due to ingestion, storage, and API retrieval stages.

Benchmark Categories

Based on these principles, including platform agnosticism, architectural relevance, and implementation grounding, six benchmark categories were defined.

Data Transmission Characteristics assess setup effort, communication mechanisms, and latency, including the distinction between push-based streaming and pull-based API access.

Platform Capabilities evaluate native support for data inspection, monitoring, and debugging.

Data Storage and Representation address persistence, ease of querying, and built-in visualisation or analytics.

Unity Integration Complexity captures software development kit (SDK) availability, programming effort, and runtime debugging effort.

Architectural Complexity and Maintainability consider component count, scalability, and suitability for sustained deployment.

Cost and Resource Implications capture licensing constraints, tooling requirements, and access to platform features relevant to academic and early-stage deployments.

Together, these benchmarks provide a concise basis for comparing cloud-based and streaming-based IoT data pipeline architectures under controlled conditions. Their application to the AWS-, Azure-, and MQTT broker–based pipelines is presented in the following section.

Comparative Analysis and Results

This section presents a comparative analysis of the three IoT data pipeline architectures evaluated in the study: TTN–AWS–Unity, TTN–Azure–Unity, and TTN–MQTT broker–Unity. The analysis applies the benchmarking approach defined in the preceding section and focuses on architectural behaviour observed during implementation and deployment. Results are presented using high-level comparison tables supported by restrained qualitative interpretation. The objective is not to rank platforms, but to identify architectural trade-offs relevant to real-time Digital Twin visualisation.

Comparison of Cloud-Based Pipelines: AWS vs Azure

Both the TTN–AWS–Unity and TTN–Azure–Unity pipelines follow a cloud-centric architectural pattern, employing webhook-based ingestion, serverless processing, managed storage, and API-mediated data

access. Both architectures support persistent data storage, structured querying, and controlled access from the Unity-based Digital Twin.

Despite the architectural parity, differences were observed in integration effort, workflow transparency, tooling requirements, and practical usability, particularly under student-tier account constraints. Table 1 summarises these differences across key architectural and implementation-oriented criteria.

Table 1. High-level comparison of AWS and Azure cloud-based IoT pipelines for digital twin data integration

Criterion	AWS	Azure
Overall architectural pattern	Cloud-centric, serverless	Cloud-centric, serverless
Data ingestion from TTN	Webhook-based	Webhook-based
Serverless processing	AWS Lambda	Azure Functions
Data persistence	Managed database (table-based)	Managed storage / tables
Querying capability (ease of data access and retrieval)	Strong, straightforward querying	Strong but more workflow-dependent
Data export (CSV/Excel)	Direct export from tables	Indirect; requires queries or external tools
Built-in dashboards (student tier)	Limited without paid services	Available for basic charts
Unity integration effort	Moderate	High
Workflow transparency	Relatively clear and linear	Fragmented, multi-step
Tooling requirements	Mostly web-based	Web + desktop tools (e.g. Storage Explorer)
Maintainability (ease of modifications)	High	Medium
Learning curve	Moderate	Steep
User Interface Usability	Relatively Intuitive	Complex

Table 1 compares the AWS- and Azure-based pipelines for digital twin data integration, focusing on how each cloud platform supports real-time sensor data delivery, persistence, and access within a Unity-based Digital Twin. The AWS-based pipeline exhibited a clearer and more linear workflow, enabling easier data access, modification, and maintenance during development. The Azure-based pipeline achieved comparable functional outcomes but required more extensive configuration, additional tooling, and multi-step workflows, resulting in higher development and debugging overhead. These differences reflect platform-level implementation characteristics rather than fundamental architectural distinctions.

Cloud-Based Pipelines versus MQTT broker-based Streaming

While AWS and Azure are architecturally comparable, the TTN–MQTT broker–Unity pipeline represents a fundamentally different integration paradigm. Rather than relying on cloud-managed processing and storage, this pipeline adopts a lightweight publish–subscribe model in which sensor data are streamed directly from TTN to the Unity application via an MQTT broker.

From a data transmission perspective, the streaming-based pipeline demonstrated lower end-to-end latency and simpler configuration, as it does not introduce serverless processing, databases, or API gateways. However, this architectural simplicity comes at the cost of data persistence, platform-level analytics, and historical access, which are inherent features of cloud-centric pipelines.

Table 2 – Comparison of cloud-based pipelines and MQTT broker-based streaming for Digital Twin data integration

Criterion	AWS/Azure	MQTT
Architectural paradigm	Centralised, cloud-centric	Decentralised, streaming
Data delivery model	Pull-based (API requests)	Push-based (publish–subscribe)
End-to-end latency	Medium (5-8 seconds)	Low (1-2 seconds)
Data persistence	Native and integral	Not provided
Historical data access	Supported	Not supported
Built-in analytics and dashboards	Available (tier-dependent)	Not available
Unity integration complexity	Medium to High	Low
Number of system components	High	Low
Setup and configuration effort	High	Low
Cost sensitivity (student tier)	High	Low
Suitability for live visualisation	Appropriate	Appropriate
Suitability for long-term monitoring	Appropriate	Limited
Data latency	Medium	Low

Table 2 summarises the key architectural differences as they affect real-time Digital Twin data integration, contrasting cloud-based pipelines with the MQTT broker–based streaming approach. The analysis indicates that streaming-based integration is fairly effective for immediate, real-time data delivery in live Digital Twin visualisation. In contrast, cloud-based pipelines provide stronger support for long-term monitoring, historical analysis, and system observability, but with higher integration overhead.

Architectural Trade-off Observations

The comparative analysis reveals several recurring architectural trade-offs across the evaluated pipelines.

Cloud richness versus simplicity – Cloud-centric pipelines provide a rich ecosystem of managed services, including ingestion, storage, monitoring, and analytics, supporting structured data management and operational visibility (AWS, 2024b). However, these capabilities require coordination of multiple components. In contrast, the MQTT broker–based pipeline minimises architectural footprint and setup effort but offers limited platform-level tooling.

Persistence versus immediacy – Persistent storage is a defining characteristic of cloud-based pipelines, enabling historical access and retrospective analysis (Palmarini et al., 2018; Qiu et al., 2025). By contrast, streaming-based pipelines prioritise immediacy by delivering data directly to the consuming application, supporting near real-time updates but offering no inherent data retention (Thangavel et al., 2014).

Integration overhead versus architectural responsibility – Cloud-based pipelines introduce higher integration overhead due to serverless processing, databases, and API layers, but centralise responsibilities such as data validation, persistence, and access control. Streaming-based pipelines reduce backend complexity but shift responsibility for data handling, temporal context, and robustness to the consuming application (Al-Nasser et al., 2024).

Summary of Findings

The comparative analysis demonstrates that no single IoT data pipeline architecture is universally optimal. Cloud-centric pipelines offer robustness, persistence, and extensibility, making them suitable for long-term, data-rich Digital Twins. Streaming-based integration prioritises low latency and simplicity, supporting responsive real-time visualisation with minimal backend overhead.

These findings reinforce that data pipeline selection should be guided by architectural alignment with application requirements, resource constraints, and system lifecycle considerations, rather than by platform capability alone. The implications of these architectural trade-offs for real-time Digital Twin visualisation are discussed in the following section.

Implications and Discussion

The comparative analysis demonstrates that architectural decisions at the IoT data pipeline level have direct and measurable implications for how effectively Digital Twins support real-time visualisation. While prior studies have examined cloud platform selection using quality-of-service metrics, simulation models, or abstract optimisation criteria (e.g. Kotas et al., 2018; Eisa et al., 2020; Hameed et al., 2025), such analyses are typically decoupled from operational Digital Twin deployments. By contrast, this study evaluates data pipeline behaviour under controlled but realistic conditions. It reveals how architectural composition shapes update continuity, data availability, and development and maintenance effort in real-time Digital Twin systems.

Latency, Update Continuity, and Immediacy

Latency is a critical determinant of real-time Digital Twin fidelity, as temporal misalignment between physical processes and their digital representations can undermine user confidence and operational reliability. As Wang et al. (2013) and Palmarini et al. (2018) have argued, delayed or inconsistent updates reduce the effectiveness of immersive and monitoring-oriented Digital Twins. In MR and VR environments, such latency can manifest as perceptible delays between user interaction and system response, disrupting the sense of immersion and reducing the reliability of real-time feedback. This is particularly critical in applications such as live walkthroughs and remote monitoring, where users rely on immediate visual updates to interpret environmental conditions and support timely decision-making.

The results indicate that publish–subscribe streaming architectures are architecturally conducive to lower end-to-end latency than cloud-centric pipelines, primarily due to the absence of intermediary storage, querying, and API retrieval stages. This design supports more continuous and immediate visual updates within Digital Twin. In contrast, cloud-based pipelines introduce additional processing layers associated with ingestion, persistence, and pull-based access, which can lead to discretised update behaviour. These observations suggest that low-latency streaming architectures are particularly well suited to live monitoring and event-driven Digital Twin applications, while recognising that such latency advantages arise from architectural simplification rather than from quantified performance measurements.

Data Persistence and Long-Term Use

Persistent data storage is a defining characteristic of cloud-centric IoT data pipelines and is widely recognised as essential for diagnostics, performance assessment, and maintenance planning in the built environment. Palmarini et al. (2018) and Qiu et al. (2025) emphasise that historical data enables Digital Twins to move beyond real-time state representation toward analytical and decision-support functions.

Consistent with this view, the analysis shows that cloud-based pipelines inherently support structured data retention and querying, enabling historical inspection and

aggregated representations within or alongside the Digital Twin. Streaming-based architectures, by contrast, prioritise transient data delivery and do not retain historical information unless supplemented by additional storage components. While this does not affect real-time update continuity, it constrains retrospective analysis, highlighting persistence and immediacy as complementary but not always compatible architectural objectives.

Integration Complexity and Architectural Responsibility

Integration complexity directly affects the feasibility and sustainability of Digital Twin implementations. Recent studies by Al-Nasser et al. (2024) and Hadjidemetriou et al. (2023) emphasise that Digital Twin viability depends as much on integration effort as on conceptual design.

The benchmark results indicate that cloud-centric pipelines introduce higher integration overhead due to the configuration of serverless processing, databases, and API layers. This complexity provides flexibility, access control, and extensibility, but increases development and maintenance effort. Streaming-based pipelines reduce backend complexity by enabling direct data delivery to the visualisation layer, but shift responsibility for data validation, temporal context, and robustness to the consuming application. This redistribution of responsibility clarifies why MQTT broker–based pipelines are not inferior, but rather represent a different architectural choice aligned with immediacy and simplicity.

Design Guidance and Context-Aware Architectures

The observed trade-offs confirm that no single data pipeline architecture is universally optimal. Cloud-based pipelines are most appropriate for Digital Twins intended as long-term, data-rich systems requiring persistence, analytics, and enterprise integration. MQTT broker–based streaming is advantageous where low latency, simplicity, and rapid deployment are prioritised.

Consistent with architectural Digital Twin literature (Wang et al., 2013; Palmarini et al., 2018), these findings also point toward the value of hybrid architectures. By combining streaming mechanisms for immediate data delivery with cloud-based services for persistence and analytics, hybrid pipelines can balance responsiveness with long-term data value. Rather than competing alternatives, cloud-centric and streaming-based pipelines should therefore be understood as complementary components within a broader Digital Twin ecosystem.

Overall, the results reinforce that IoT data pipeline selection should be guided by intended Digital Twin functionality, lifecycle requirements, and operational context. By treating immersive visualisation as a demanding application context rather than the primary object of evaluation, this study highlights how backend architectural choices shape the robustness, deployability, and real-time behaviour of Digital Twin systems in the built environment.

Conclusion

This paper presented a comparative architectural analysis of IoT data pipeline approaches supporting real-time Digital Twin visualisation in the built environment. Using a controlled case study, three end-to-end pipelines; TTN–AWS–Unity, TTN–Azure–Unity, and TTN–MQTT broker–Unity were implemented and evaluated under identical sensing, visualisation, and mixed reality deployment conditions. The data integration layer was isolated as the primary variable.

The findings demonstrate that data pipeline architecture significantly influences Digital Twin responsiveness, complexity, and functional scope, particularly in high-demand visualisation contexts such as mixed reality. Cloud-centric pipelines support persistence, scalability, and long-term data management, making them suitable for sustained and data-rich Digital Twin deployments, albeit with higher integration and operational overhead. In contrast, streaming-based pipelines prioritise architectural simplicity and immediacy, supporting responsive real-time visualisation but offering limited inherent support for historical data handling.

Beyond platform-specific performance claims, the study provides architectural insight into how data integration design choices shape the practical behaviour of real-time Digital Twins when deployed in immersive and interactive environments. Future research may extend this work through hybrid pipeline architectures that combine low-latency streaming with cloud-based persistence and analytics, as well as through quantitative latency evaluation and deployment at larger or multi-building scales.

Overall, the findings confirm that no single data pipeline architecture is universally optimal. Instead, effective pipeline selection should be guided by the intended role, lifecycle stage, and interaction demands of the Digital Twin, underscoring the importance of aligning backend data architectures with application requirements rather than adopting a one-size-fits-all integration strategy.

References

Al-Nasser, H., Ahmad, M.E., Peralta, A., Patricia, C.G.C., Al-Zuriqat, T. Dragos, K. and Smarsly, K. (2024). Digital twin architectures in civil engineering: A systematic literature review, conference, Fachtagung Baustatik – Baupraxis: Hamburg, Germany.

Ansyah, A.S.S., Arifin, M., Alfan, M.B., Suriawan, M.V., Farhansyah, N.H., Shiddiqi, A.M. and Studiawan. H. (2023) MQTT Broker Performance Comparison between AWS, Microsoft Azure and Google Cloud Platform, 2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC), Mysore, India, 2023, pp. 1-6, doi: 10.1109/ICRTEC56977.2023.10111870.

AWS (2024a) Guidance for Digital Twin Framework on AWS. Available at: https://aws.amazon.com/solutions/guidance/digital-twin-framework-on-aws/ (Accessed: 13 January 2026).

AWS (2024b) Building your digital twin solution using the Digital Twin Framework on AWS. Available at: https://aws.amazon.com/blogs/hpc/building-your-digital-twin-solution-using-the-digital-twin-framework-on-aws/ (Accessed: 16 January 2026).

Boje, C., Guerriero, A., Kubicki, S. and Rezgui, Y. (2020) 'Towards a semantic Construction Digital Twin: Directions for future research', Automation in Construction, 114, 103179.

Boschert, S. and Rosen, R. (2016) 'Digital Twin—The simulation aspect', in Mechatronic Futures. Cham: Springer, pp. 59–74.

Centenaro, M., Vangelista, L., Zanella, A. and Zorzi, M. (2016) 'Long-range communications in unlicensed bands: The rising stars in the IoT and smart city scenarios', IEEE Wireless Communications, 23(5), pp. 60–67.

El Ammari, K. and Hammad, A. (2019) Remote interactive collaboration in facilities management using BIM-based mixed reality, Automation in Construction ,107(2019), 102940,

Eisa, M., Younas, M., Basu, K. and Awan, I. (2020) Modelling and Simulation of QoS-Aware Service Selection in Cloud Computing, Simulation Modelling Practice and Theory, 103, 102108.

Grieves, M. and Vickers, J. (2016) 'Digital Twin: Mitigating unpredictable, undesirable emergent behavior in complex systems', in Transdisciplinary Perspectives on Complex Systems. Cham: Springer, pp. 85–113.

Hadjidemetriou, L., Stylianidis, N., Englezos, D., Papadopoulos, P., Eliades, D., Timotheou, S., Polycarpou, M.M. and Panayiotou, C. (2023) A digital twin architecture for real-time and offline high granularity analysis in smart buildings, Sustainable Cities and Society, 98, 104795,

Hameed, S., Bukhari, S.H.R., Rizvi, S.U.E.A., Tahir, M. and Aadil, F. (2025) A Comparative Analysis on the Services Offered by Amazon Web Services and Microsoft Azure, Concurrency and Computation: Practice and Experience, WILEY, pp. 1-19.

Kotas, C., Naughton, T. and Imam, N. (2018) "A Comparison of Amazon Web Services andMicrosoft Azure Cloud Platforms for High Performance Computing," in 2018 IEEE International Conference on Consumer Electronics (ICCE) (IEEE, 2018), pp. 1–4.

LoRa Alliance (2025) What is LoRaWAN? Available at: https://lora-alliance.org/about-lorawan/ (Accessed: 12 January 2026).

Lu, Q., Parlikad, A., Woodall, P., Ranasinghe, G. D., Xie, X., Liang, Z., Konstantinou, E., Heaton, J., & Schooling, J. (2020). Developing a dynamic digital twin at building and city levels: A case study of the West Cambridge campus.

Microsoft (2024) Microsoft Azure – Cloud Computing Services. Available at: https://azure.microsoft.com (Accessed: 13 January 2026).

Palmarini, R., Erkoyuncu, J.A., Roy, R. and Torabmostaedi, H. (2018) 'A systematic review of augmented reality applications in maintenance', Robotics and Computer-Integrated Manufacturing, 49, pp. 215–228.

Palumbo, F., Aceto, G., Botta, A., Ciuonzo, D., Persico, V. and Pescape, A. (2021) 'Characterization and Analysis of Cloud-To-User Latency: The Case ofAzure and Aws', Computer Networks, 184 (2021): 107693.

Qiu, S., Zaheer, Q., Ali, F., Wajid, S., Chen, H. Ai, C. and Wang, J. (2025). Exploring the impact of digital twin technology in infrastructure management: a comprehensive review. Journal of Civil Engineering and Management, 31, pp. 395-417.

Shahzad, M., Joe, H.M. Tah, Younas, M. and AlMukhtar, A. (2025) Technologies and techniques in digital twins for real-time data visualisation in building maintenance: A state-of-the-art review', Journal of Infrastructure Intelligence and Resilience, 4(4), 100185.

Tao, F., Zhang, H., Liu, A. and Nee, A.Y.C. (2018) 'Digital twin in industry: State-of-the-art', IEEE Transactions on Industrial Informatics, 15(4), pp. 2405–2415.

Thangavel, D., Ma, X., Valera, A., Tan, H.-X. and Tan, C.K.-Y. (2014) 'Performance evaluation of MQTT and CoAP via a common middleware', 21-24 April 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP). IEEE, pp. 1–6.

The Things Network (TTN) (2023) The Things Network Documentation. Available at: https://www.thethingsnetwork.org/docs/ (Accessed: 15 January 2026).

Wang, X., Love, P. E. D., Kim, M. J., Park, C.-S., Sing, C.-P., & Hou, L. (2013). A conceptual framework for integrating building information modeling with augmented reality. Automation in Construction, 34, pp.37- 44.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

PRESENTING THE LOINxml SCHEMA: DISCUSSING DESIGN DECISIONS AND INTENDED APPLICATION

Martina Mellenthin Filardo[1,2], Paul Debus[1], Sven Zentgraf[3], Štefan Jaud[4], Christian Clemen[5], Ghang Lee[6], Jakob Beetz[7] and André Borrmann[8]
[1]Bauhaus-Universität Weimar, Weimar, Germany
[2]Erfurt University of Applied Sciences, Erfurt, Germany
[3]Ruhr-Universität Bochum, Bochum, Germany
[4]Jaud IT GmbH, Moorenweis, Germany
[5]Universtiy of Applied Sciences Dresden, Dresden, Germany
[6]Yonsei University, Seoul, South Korea
[7]RWTH Aachen University, Aachen, Germany
[8]Technical University of Munich, Munich, Germany

Abstract

The Level of Information Need introduced in ISO 19650 gained a standardized machine-readable data model with the LOINxml (ISO 7817-3). This paper presents the normative context of this standard with a short outline of the specification process of the LOINxml data model. Through systematic analysis of schema elements and design decisions, this work demonstrates how LOINxml enables structured, transaction-based requirement specifications supporting alphanumerical, documentation, and geometrical information across project lifecycles. The contribution establishes LOINxml as foundational infrastructure for digital information requirements while identifying threshold conditions and future research directions essential for industry implementation.

Introduction

Precise information requirements are critical for Building Information Modeling (BIM) information delivery and conformant project deliverables in the architecture, engineering, construction, and operation (AECO) industry (Sacks et al., 2025). Machine-readable representation of these requirements enables not only vendor-neutral definition, but also more precise requirements and higher quality of project deliverables. A framework for such requirement definition has been established in the Level of Information Need (LOIN) standards. While fundamentals have been specified in part 1 of ISO 7817, application guidelines are presented in part 2 (ISO/TS 7817-2, 2025), and a corresponding data model is defined in part 3. (ISO 7817-3, 2025). Of these three interrelated standards, this paper discusses the design decisions and intended applications of the ISO 7817-3 LOINxml schema.

This paper documents the context and design decisions of the data model presented in ISO 7817-3 (2025) to promote practitioner adoption. The context in which the data model is embedded is explained in the normative setting. Subsequently, the development process is explained. The data model published in ISO 7817-3 is then presented and discussed. The paper finishes with the conclusions and outlook.

Normative setting

The Level of Information Need (LOIN) concept was introduced in ISO 19650-1 (2018). However, the concept had to be elaborated into precise normative specifications, which has been done in the ISO 7817 series, where fundamentals and principles are described in ISO 7817-1 (2024), guidelines and application in ISO/TS 7817-2 (2025), and an XML-based data model in ISO 7817-3 (2025). ISO 7817-3, specifying a schema for the Level of Information Need, aligns closely with ISO 23387 (2025), which establishes a data template structure for construction objects based on the methodology for describing, creating, and maintaining properties in interconnected data dictionaries (ISO 23386, 2020) described in ISO standard 12006-3. This approach provides the concepts and relationships required to describe information about building objects (ISO 23387, 2025).

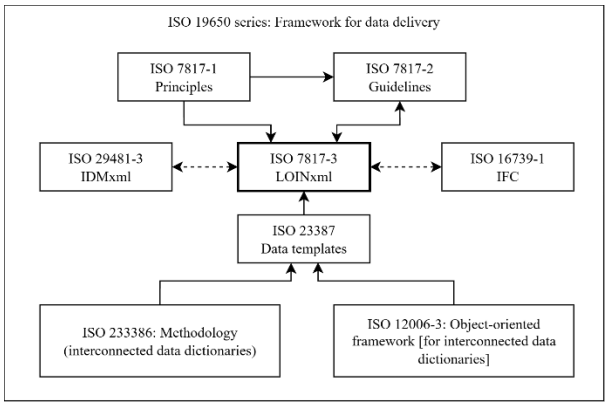


Figure 1: Normative context of ISO 7817-3.

In this context, a data template is a subset of the model ISO 23387 extends the ISO 23386 methodology, emphasizing template reusability and both standards harmonize with ISO 12006-3 (2023), thus enabling an object-oriented approach – a strong argument for adoption of the data templates elements in the LOINxml data model. This approach was implemented by importing

XML schema elements provided by ISO 23387 into the LOINxml schema, which represents the strongest form of integration. ISO 7817-3 coexists with ISO 29481-3 (2022) on the Information Delivery Manual (IDM). While both standards provide machine-readable XML structures for information requirements, IDM focuses on information requirements exchanged within specific use cases, whereas LOIN focuses on the data model for the level of information need. While being agnostic to specific data models for actual model contents, the LOINxml is compatible with the Industry Foundation Classes (IFC), among other formats and breakdown structures.

A strong principle implemented by LOINxml is the use of references to information provided in other data schema and resources. This may include references to data dictionaries, such as bSDD, for specifying ObjectTypes and Properties, or references to IDMxml data for referencing project milestones. The references are expressed by International Resource Identifiers (IRI) to allow easy lookup and retrieval, while providing a maximum of flexibility. The context of ISO standards relevant to the LOINxml data model are outlined in Figure 1. The LOINxml data schema can be used for file-based exchange (including multiple files referencing each other) but can also act as a blueprint for defining database schemas and providing access through REST (Representational State Transfer) APIs (Application Programming Interfaces).

Methodology

The development of the schema for ISO 7817-3 was conducted within the framework of standardization work (Vienna Agreement, 1991) for cooperation between the European Committee for Standardization (CEN) and the International Organization for Standardization (ISO) (European Committee for Standardization and International Organization for Standardization, 2016). Delegated experts from all interested countries participate in standard development. In the case of ISO 7817-3, experts came from Belgium, Canada, Czech Republic, France, Germany, Norway, South Korea, and the UK. Meetings were conducted virtually. Over the course of six years of development work, the meeting frequency varied between biweekly and approximately monthly (accounting for breaks and holidays). Figure 3 outlines the process. Other works aided in the development of this standard, such as Jeon et al. (2021); Jaud, Gregory and Borrmann (2021); Bolpagni et al. (2022); Abualdenien and Borrmann (2022); Son et al. (2022); Mellenthin Filardo et al. (2024a); Mellenthin Filardo et al. (2024b); Akbas et al. (2025); Zentgraf et al. (2025) among others.

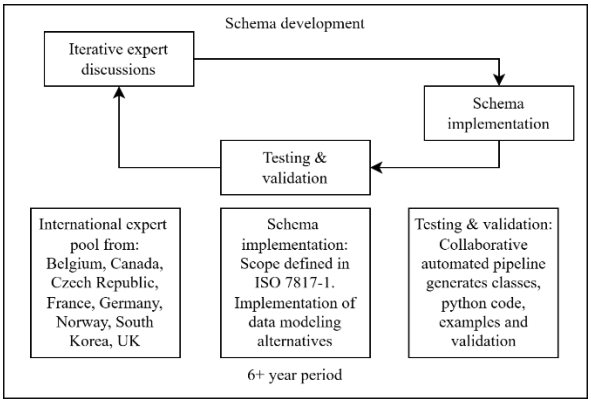


Figure 2: Overview of the schema development procedure

The schema scope derives from ISO 7817-1. Guidelines in ISO/TS 7817-2, developed parallel to ISO 7817-3, informed LOIN elaborations subsequently discussed, tested, and eventually incorporated into LOINxml. Substantive congruence between ISO 23387 and ISO 7817-3 necessitated parallel mechanism implementation. Collaboration through integration avoided duplicate work and contradictory approaches. Individual elements and mechanisms were iteratively discussed among experts.

Discussion material alternated between UML diagrams, XSD files, and XML examples using a collaborative repository. Automated actions generated classes from XSD files using xsData libraries. Python scripts produced and validated XML examples against the schema. XSD and example generation proceeded collaboratively through expert-proposed improvements across multiple iterations. Both XSD and XML files underwent expert review. Drafts circulated through designated internal committees until the standard entered its final phase.

Presenting the LOINxml

The root element of the LOINxml data model is the *LevelOfInformationNeed* element, which acts as a container for the *Specification* element(s), as shown in Figure 2. The *Specification* aggregates the elements *Prerequisites* (mandatory), for context information, *GeoReferencing* (optional), for coordinate system information, and *SpecificationPerObjectType* (optional, zero to many), for specifying requirements.

The *Prerequisites* element (Figure 4) establishes the transaction framework by specifying the purpose, information delivery milestone, and actors of the

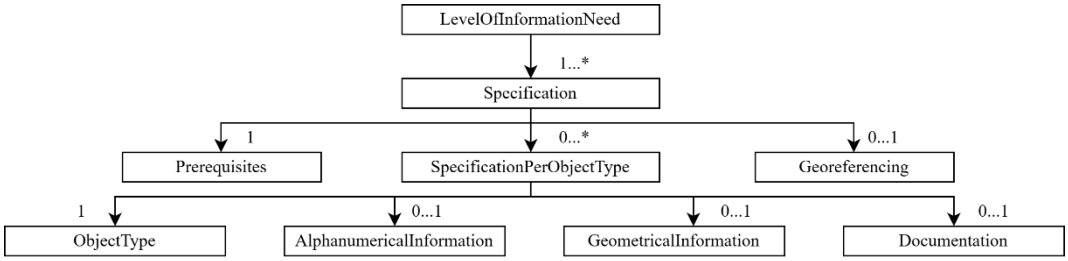


Figure 3: Simplified overview of the LOINxml data model. A detailed version of the schema can be found in ISO 7817-3.

LOINxml. The LOINxml is understood as a transaction limited to a receiving and a providing actor.

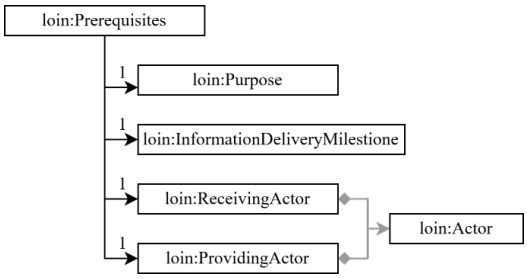


Figure 4: Structure of the Prerequisites element

The *Purpose* element provides the rationale for information need and includes name, optional definition, references, description, language, region, and dictionary links. The *InformationDeliveryMilestone* specifies delivery timing through name, optional description, references, and date/time. The *ProvidingActor* identifies the information provider through role, optional name, affiliation, description, and email. The *ReceivingActor* identifies the information recipient using the same structure as *ProvidingActor*. Varying actors or purposes require separate *Specification* instances rather than consolidated prerequisites. Multiple specifications with different prerequisites can coexist within one document.

The *GeoReferencing* element, shown in Figure 5, specifies the coordinate reference system and transformation parameters from model to coordinate reference system (CRS) coordinates. It comprises the elements *CoordinateReferenceSystem*, defining CRS type and datum, and *ModelCoordinateSystem*, establishing model origin and orientation within the CRS where multiple instances are possible.

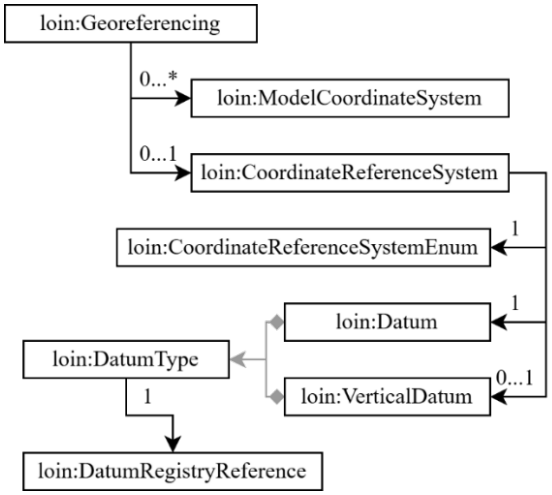


Figure 5: Georeferencing element structure

The *CoordinateReferenceSystem* defines the reference system through Type specifying CRS type as *ProjectedCRS*, *EngineeringCRS*, *GeographicCRS*, or *NotRequired*, *Datum* providing required horizontal datum, and *VerticalDatum* offering optional vertical datum. Each datum includes Name such as "EPSG:25832" or WKT2 string and Type as *DatumRegistryReference*. The *DatumRegistryReference* enables flexible geodetic registry specification per ISO 19127:2019 through *Name*, optional *Description*, and optional *RegistryReference* provided through a Unified Resource Identifier (URI).

The *ModelCoordinateSystem* element defines model positioning within CRS using generic coordinate names. *IsProjected* indicates whether the model inherits CRS projection distortion. *FirstCoordinate* specifies the first horizontal coordinate, typically Easting or Latitude. *SecondCoordinate* defines the second horizontal coordinate, typically Northing or Longitude. *Height* establishes the vertical coordinate. *XAxisAbscissa* and *XAxisOrdinate* provide optional model X-axis direction components. *UnitScale* offers optional model-to-CRS unit conversion. *HorizontalScale* provides optional horizontal distance scale factor.

Avoiding to “reinvent the wheel”, the LOINxml schema uses several elements from the data templates schema (namespace *dt*) defined in ISO 23387 (2025), in particular when it comes to the definition of properties, data types, and units. The schema employs Globally Unique Identifiers (GUIDs) and *ReferenceType* for element referencing within documents and across external resources. The schema definition of the used reference mechanisms is shown in Figure 6.

```
<xs:complexType name="ReferenceType">
    <xs:attribute ref="referenceURI" use="optional" />
    <xs:attribute ref="GUID" use="optional" />
</xs:complexType>
<xs:attribute name="referenceURI" type="xs:anyURI">
    <xs:annotation>
        <xs:documentation>
            Defines a referenceURI to identify a concept
            using URI
        </xs:documentation>
    </xs:annotation>
</xs:attribute>
<xs:attribute name="GUID" type="GuidType">
    <xs:annotation>
        <xs:documentation>
            Defines the GUID of a concept
        </xs:documentation>
    </xs:annotation>
</xs:attribute>
<xs:simpleType name="GuidType">
    <xs:annotation>
        <xs:documentation>
            The representation of a GUID, generally the id
            of an element.
        </xs:documentation>
    </xs:annotation>
    <xs:restriction base="xs:string">
        <xs:pattern value="[a-fA-F0-9]{8}-[a-fA-F0-9]{4}-
                           [a-fA-F0-9]{4}-[afA-
                           F0-9]{4}-[a-fA-F0-9]{12}" />
    </xs:restriction>
</xs:simpleType>
```

Figure 6: Schema definition of the elements from ISO 23387 used for linking in ISO 7817-3.

GUIDs provide unique element identification serving two purposes: establishing element identity and enabling internal cross-references. Most elements support the *dt:GUID* attribute for same-document referencing.

The *dt:ReferenceType* creates references via two attributes. The *dt:GUID* attribute references elements within the same document. The *dt:referenceURI* attribute references external resources. When both attributes are present, the reference points to a specific element identified by GUID within an external resource identified by URI. Referencing guidelines recommend using

dt:GUID alone for same-file references, *dt:referenceURI* alone for external resources, and both attributes for specific elements within external resources.

The *SpecificationPerObjectType* element defines information requirements for specific object types, such as walls, doors, or bridge abutments, binding object definitions to three requirement aspects: alphanumerical information, documentation, and geometrical information. Each Specification may contain multiple *SpecificationPerObjectType* elements, with one per deliverable object type. The structure comprises *ObjectType* defining the target object type as a required element, *AlphanumericalInformation* specifying required properties, quantities as optional, *Documentation* defining required documents as optional, and *GeometricalInformation* specifying geometrical requirements as optional.

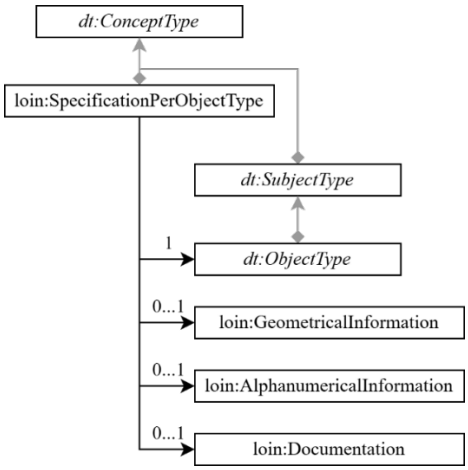


Figure 7: Structure of the SpecificationPerObjectType element

Since *loin:SpecificationPerObjectType* extends *dt:ConceptType* from ISO 23387, it includes Name and Definition elements separate from the *ObjectType* it addresses. Figure 7 shows the *dt:ConceptType* element and how it is incorporated in the LOINxml.

The *ObjectType* element identifies requirement targets using *dt:ObjectTypeType* from ISO 23387. Each object type requires a *dt:GUID* for internal referencing. Two main definition approaches exist: inline definition with Name, Definition, and metadata, or external reference via *dt:ReferenceDocumentRef* linking to libraries or other external resources. Multiple language variants support international projects as specified in ISO 23387. Version tracking occurs through *dt:MajorVersion*, *dt:MinorVersion*, and *dt:Status* as defined in ISO 23387. External references enable linkage to established classifications such as IFC or UniClass while maintaining internal GUIDs and optional local metadata.

A reduced instance example of a specification per object type element with emphasis on an object type definition using CityGML is shown in Figure 8. The *dt:ObjectType* is a child of the *dt:SubjectType* element, thus covering all the metadata of the *dt:ConceptType* element and allowing references defined in the subject type (reference for '*hasParts*' and '*isSubtypeOf*'). By aggregating specific information and associating it with a specific object type, the specification per object type container acts as a 'has' relation inherently (XML does not support relations in that sense).

```
<SpecificationPerObjectType dt:GUID="2c34312at-6h81-4907-
827fdc5eb9b2f28d" dateOfCreation="2026-01-07T10:30:00">
    <dt:Name language="en">CityGML abstract relief component
        properties</dt:Name>
    <dt:Definition language="en">CityGML abstract relief
        component to aggregate inheritance properties for terrain
        elements</dt:Definition>
    <ObjectType dt:about="https://schemas.opengis.net/citygml/relief/
1.0/relief.xsd" dt:GUID="ab294g80-2c6c-6f13-2adf-
46ed115369e9" dateOfCreation="2026-01-07T10:30:00">
        <dt:Name language="en">CityGML abstract relief
            component</dt:Name>
        <dt:Definition language="en">CityGML abstractrelief
            component</dt:Definition>
        <dt:Description language="en">abstract component to
            aggregate inheritance attributes</dt:Description>
        <dt:LanguageOfCreator>en-US</dt:LanguageOfCreator>
        <dt:CountryOfOrigin>Germany</dt:CountryOfOrigin>
        <dt:MajorVersion>3</dt:MajorVersion>
        <dt:MinorVersion>2</dt:MinorVersion>
        <dt:Status>Released</dt:Status>
    </ObjectType>
</SpecificationPerObjectType>
```

Figure 8: XML instance example of an object type definition for a CityGML element

The *AlphanumericalInformation* element specifies required properties through several components. Its structure is presented in Figure 9.

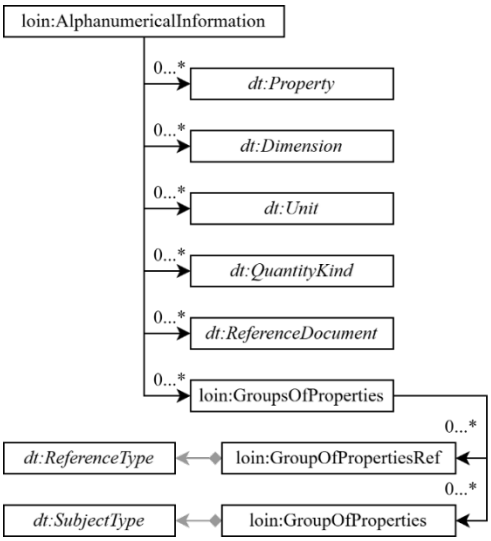


Figure 9: Structure of the AlphanumericalInformation element

Property defines individual property requirements either inline or referenced, with data types, constraints, and measurement references. *QuantityKind* provides quantity type definitions such as Length, Mass, or Pressure for inline properties or by referencing *Dimension*. *Dimension* expresses physical dimensionality (i.e., $L^1M^0T^0$) as SI base unit exponents, supporting inline units and quantity kinds. *Unit* specifies measurement units with conversion information including Scale, Base, Coefficient, and Offset. *DataType* specifying constraints. External definitions via *dt:referenceURI* on reference elements can point to standards like the Quantity, Unit, Dimension, and Type (QUDT) ontology (Hodgson, 2024).

Figure 10 shows an XML example of a self-contained LOINxml definition (inline in one file) while Figure 11 demonstrates the definition in separate files (e.g., one LOINxml and a separate Library with referenceable property definitions).

```
<AlphanumericalInformation dt:GUID="94eb4cfa-46a4-4ea6-bab5-
c6711b1ae35a">
    <Property dt:GUID="8ef0be47-929f-428a-9016-ec1e1f1f7616" dateOfCreation="2026-01-07T10:30:00">
        <dt:Name language="en">relativeToTerrain</dt:Name>
        <dt:Definition language="en">Property to filter elements without
            checking geometry</dt:Definition>
        <dt:DataType name="STRING">
            <dt:PossibleValues>
                <dt:ValueList language="en">
                    <dt:Value order="1">entirelyAboveTerrain </dt:Value>
                    <dt:Value order="2">substantiallyBelowTerrain
                    </dt:Value>
                </dt:ValueList>
            </dt:PossibleValues>
        </dt:DataType>
    </Property>
    <dt:Property dt:GUID="f1090232-9199-22ab-23gh-4a88b6d6b86d" dateOfCreation= "2026-01-07T10:30:00">
        <dt:Name language="en"> FrostResistanceClasses
        </dt:Name>
        <dt:Name language="de">Frostschutzklassen</dt:Name>
        <dt:Definition language="en">Exposure classes related to freezethaw
            cycles according to DIN 1054</dt:Definition>
        <dt:LanguageOfCreator>en-US</dt:LanguageOfCreator>
        <dt:DataType name="STRING">
            <dt:PossibleValues>
                <dt:ValueList language="en">
                    <dt:Value order="1">XF1</dt:Value>
                    <dt:Value order="2">XF2</dt:Value>
                    <dt:Value order="3">XF3</dt:Value>
            </dt:PossibleValues>
        </dt:DataType>
    </dt:Property>
</AlphanumericalInformation>
```

Figure 10: XML instance example of an inline (self-contained) property definition for the specified CityGML object type

```
  <AlphanumericalInformation dt:GUID="82cd6dfc-24a7-1ed9-cda8-a6727a1be28a">
    <GroupOfProperties dt:GUID="b2120e0b-47b1-4f66-b0f7-70a9e0c559ce" dateOfCreation="2026-01-07T10:30:00">
        <dt:Name language="en">Structural Analysis Properties</dt:Name>
        <dt:Definition language="en">Collection of exemplary properties
            <dt:LanguageOfCreator>en-US</dt:LanguageOfCreator>
            <dt:HasPropertyRef dt:referenceURI="file:structural_properties_library.xml" dt:GUID="55492688-2348-4986-85b3-b1a3fb68b31b" />
            <dt:HasPropertyRef dt:referenceURI="file:structural properties library.xml" dt:GUID="f9660631-7197-41cc-95bc-8a98a5d5b89c" />
    </GroupOfProperties>
    </GroupsOfProperties>
</AlphanumericalInformation>
and

<dt:Property dt:about="https://ifc43-docs.standards.buildingsmart.org/IFC/RELEASE/IFC4x3/HTML/lexical/IfcPropertySingleValue.htm" dt:GUID="55492688-2348-4986-85b3-b1a3fb68b31b" dateOfCreation="2024-01-10T09:00:00">
    <dt:Name language="en">LoadBearingCapacity</dt:Name>
    <dt:Name language="de">Tragfähigkeit</dt:Name>
    <dt:Definition language="en">Maximum axial load the column can support
        under design conditions</dt:Definition>
    <dt:Description language="en">Ultimate load capacity considering safety
        factors and material properties</dt:Description>
    <dt:LanguageOfCreator>en-US</dt:LanguageOfCreator>
    <dt:CountryOfOrigin>Germany</dt:CountryOfOrigin>
    <dt:DataType name="REAL">
        <dt:MinInclusive value="100.0" />
        <dt:MaxInclusive value="50000.0" />
        <dt:DataFormat value="[0-9]+\.[0-9]{1,2}" />
    </dt:DataType>
    <dt:Symbol>F Rd</dt:Symbol>
    <dt:DimensionRef dt:GUID="63f6299d-6ef4-4371-947f-4d3cef25fcdb" />
    <dt:UnitRef dt:GUID="e3bf386d-6baa-473e-aa8b-d95b20dacelb" />
    <dt:QuantityKindRef dt:GUID="43daf97c-ee29-4800-b010-c5f2ab4f05be" />
</dt:Property>
```

Figure 11: XML instance example outlining the linking mechanism (yellow highlight) of properties in separate files

GroupsOfProperties contains collections of related properties through inline *GroupOfProperties* or external *GroupOfPropertiesRef* elements. *ReferenceDocument* provides optional links to external dictionaries or standards.

The *GeometricalInformation* element defines geometrical requirements through five aspects, as shown in Figure 12.

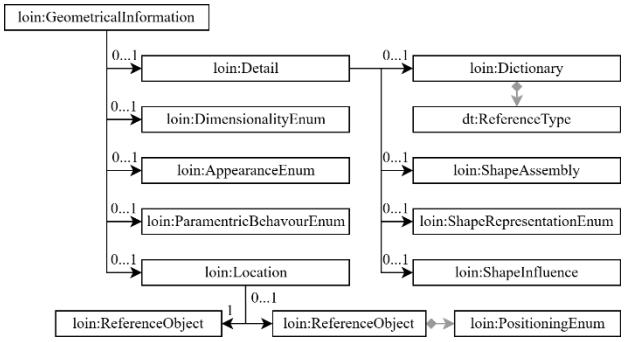


Figure 12: Structure of the GeometricalInformation element

The *Detail* aspect specifies geometry detailing information through a *Dictionary*, providing optional external definition references, *ShapeAssembly*, describing object assembly per ISO/TS 7817-2, *ShapeRepresentation* defining shape representation method per ISO/TS 7817-2, and *ShapeInfluence* establishing geometric feature requirements per ISO/TS 7817-2. The *Dimensionality* aspect describes dimensional representation as *0D, 1D, 2D, 3D*, or *NotRequired* per ISO/TS 7817-2.

The *Appearance* aspect specifies visual appearance requirements per ISO/TS 7817-2. The *ParametricBehaviour* aspect defines parametric behavior specification per ISO/TS 7817-2. The *Location* aspect establishes positioning method as *Absolute, Relative*, or *NotDefined* per ISO/TS 7817-2, where *RelativeOrAbsolute* specifies the positioning method as required and *ReferenceObject* provides the reference for relative positioning through GUID, filter, XML path, or natural language as optional.

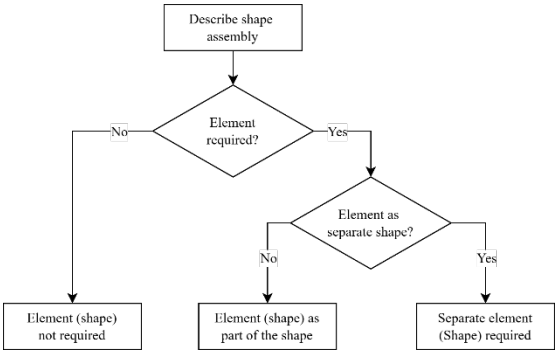


Figure 13: Simplified decision tree showing shape composition

The *Placeholder* attribute indicates placeholder-only geometry. *ShapeInfluence* specifies geometric feature representation with enumerated values according to the decision trees presented in ISO/TS 7817-2. A simplified version of the decision tree logic is outlined in Figure 13. The resulting enumerations (*NotRequired, No[Element], [Element]AsPartOfShape,* or *SeparateShapes*), and optional *ThresholdDimension* for size thresholds.

An XML example is given in Figure 14 to show application of the geometrical information.

```
<GeometricalInformation dt:GUID="5a807595-d9a9-41ce-868d-3912fa16b19c" placeholder="false">
    <Detail>
```

```
4          <Dictionary dt:referenceURI=
         "https://www.buildingsmart.co.at/glossarentries/
5        level-of-detail-200-lod-200/" />
6          <ShapeAssembly> SingleObjectSingularShape
           </ShapeAssembly>
7          <ShapeRepresentation> SingleBoundingPrimitive
                  </Shape Representation>
8              <ShapeInfluence>
9                  <InsideGeometry>
         InsideGeometryAsPartOfShape </Inside Geometry>
10                     <Connections> ConnectionsAsPartOfShape
                       </Connections>
11                     <Openings>SeparateShapes</Openings>
12                     <OperatingAndClearanceZones>
         SeparateShapes </Operating AndClearanceZones>
13
         <Features>FeaturesAsPartOfShape</Features>
14             </ShapeInfluence>
15     </Detail>
16     <Dimensionality>3D</Dimensionality>
17     <Appearance>SingularMaterial</Appearance>
18     <ParametricBehaviour>Requested</ParametricBehaviour>
19     <Location>
20         <RelativeOrAbsolute>Relative</RelativeOrAbsolute>
21         <ReferenceObject>Relative to the bridge
22             bearing</ReferenceObject>
23     </Location>
24 </GeometricalInformation>
```

Figure 14: XML instance example showing a geometrical information definition

The *Documentation* element specifies required documents via one or more *Document* elements. Each *Document* defines name as the document identifier, description specifying content as optional, *ReferenceDocument* providing external specification references as optional, *Format* establishing required file format with name and version, and optional classification attributes including type, form, and content.

Discussion of design decisions

The LOINxml data model prioritized reusing established principles and respecting existing frameworks. A major consideration was the breakdown structure foundation.

While IFC is widely established in AECO industry, not all requirements can be defined within the IFC data schema due to receiving actors' limited IFC knowledge or requirement nature. Projects requiring alternative breakdown structures are not uncommon. To enable broad application, the ISO 7817-3 (2025) data model remains agnostic, permitting any breakdown structure. The design supports multiple breakdown structures, enabling requirements spanning various disciplines. Though this decision introduces known challenges, the primary objective was allowing requirement definition as receiving actors deem necessary. Translating formalized requirements into IFC or other formats constitutes a subsequent step.

LOIN specifications require prerequisites elements while georeferencing and specification per object type remain optional. This ensures LOINxml specifications always include contextual information. The transaction-based model restricting specifications to one receiving and one providing actor was deliberate. Complex actor relationships such as sub-contractors or multiple providing actors were considered. However, such transactions vary either in purpose, warranting separate specifications, or in requirements themselves as with sub-contractors or separate providing actors delivering partial models. Given LOINxml's extensive linking options for alphanumerical information and flexible property structuring, the two-actor solution balances clearly framed requirements with lightweight definition.

The *GeoReferencing* structure separates CRS definition from model positioning through distinct *CoordinateReferenceSystem* and *ModelCoordinateSystem* elements. This separation acknowledges that multiple model coordinate systems may exist within a single CRS, supporting scenarios where different project components use different local origins or orientations within a shared global reference frame. The use of generic coordinate names (*FirstCoordinate, SecondCoordinate, Height*) rather than CRS-specific terminology (Easting, Northing, Latitude, Longitude) enables the schema to accommodate projected, geographic, and engineering coordinate systems uniformly. The *IsProjected* parameter addresses the fundamental distinction between models that inherit CRS projection distortion and those maintaining true-to-reality geometry, reflecting practical considerations in infrastructure versus building modeling workflows. The optional transformation parameters (*UnitScale, HorizontalScale, XAxisAbscissa, XAxisOrdinate*) provide flexibility for complex coordinate transformations but increase specification complexity and potential for misapplication.

The integration of multilingual capabilities through ISO 23387's *dt:ConceptType* enables specifications to function across language boundaries without translation or duplication. This design decision reflects the international nature of construction projects and standards development but introduces complexity in specification authorship and maintenance where multiple language variants must remain synchronized. The inclusion of versioning metadata (*MajorVersion, MinorVersion, Status*) at the object type level enables tracking of definition evolution over time, supporting change management and specification maturity assessment. However, the schema does not enforce versioning constraints or define version compatibility rules, leaving version management protocols to implementation conventions rather than schema-level enforcement.

The distinction between *ObjectType* definition and *SpecificationPerObjectType* implements a separation of concerns where object identity remains independent of requirement specification. This architecture permits the same object type to appear in multiple specifications with varying requirements while maintaining consistent identity through GUIDs. The extension of *dt:ConceptType* from ISO 23387 enables the specification itself to carry metadata distinct from the object it addresses. This dual-layer approach supports requirement versioning and modification without altering object type definitions, facilitating reuse across different project contexts and specification documents.

The schema accommodates both inline definition and external referencing across multiple element types, reflecting a fundamental tension between self-contained specifications and distributed definition management. Inline definitions create complete, autonomous

documents requiring no external dependencies but introducing redundancy when similar definitions recur across specifications. External referencing through *dt:referenceURI* reduces redundancy and enables centralized maintenance of definitions but creates dependencies on external resources whose availability and versioning must be managed. The property definition chain exemplifies this design where inline properties require complete definition of supporting *Units, Dimensions,* and *QuantityKinds*, while externally referenced properties inherit these definitions from their source dictionaries. This flexibility accommodates varying organizational capabilities and project requirements but places responsibility for consistency management on specification authors.

The decomposition of requirements into alphanumerical, documentation, and geometrical aspects reflects established information modeling paradigms where object attributes, associated documents, and geometric representations constitute distinct information categories. This separation enables independent specification of each aspect's requirements, supporting scenarios where different aspects require different levels of detail (for lack of a better word) or where responsibility for different information types resides with different project participants. The optional nature of all three aspects acknowledges that not all specifications require all information types. However, this structure assumes these three categories comprehensively cover information needs, potentially constraining representation of emerging requirement types such as behavioral specifications or process-related information.

The universal application of GUIDs for element identity establishes a stable referencing mechanism independent of element names, positions, or hierarchical relationships. This approach supports requirement traceability and change management by providing persistent identifiers that remain valid across document modifications. The distinction between *dt:GUID* for internal references and *dt:referenceURI* for external references, with their combination enabling references to specific elements within external resources, implements a two-tier resolution strategy. Internal resolution operates within document scope without external dependencies, while external resolution requires access to referenced resources. The schema's guidance to prefer GUID-only for internal references and URI-only for external resources suggests optimization for reference resolution performance and reliability, though the combined approach enables precise external element targeting when necessary.

The five-aspect geometric requirements structure (*Detail, Dimensionality, Appearance, ParametricBehaviour, Location*) aligns with ISO/TS 7817-2 guidelines for geometric specification. This decomposition treats geometric information as multidimensional rather than monolithic, acknowledging that different project phases or use cases may require different geometric characteristics. The *ShapeInfluence* mechanism within *Detail* enables fine-grained specification of individual geometric features with threshold-based rules, supporting complexity management where minor features may be omitted while significant features require explicit modeling. However, the reliance on enumerated values for each aspect constrains expressiveness to predefined options, potentially limiting specification of novel geometric requirements not anticipated in the enumeration design.

The practical implementation of the proposed schema necessitates a shift from purely geometric modeling toward robust semantic integrity, requiring rigorous alignment between proprietary parametric environments and standardized data structures. This integration typically demands the development of bespoke middleware or API-driven scripts to facilitate the mapping of local attributes to schema-specific fields, thereby ensuring that data fidelity is maintained throughout the project lifecycle. Furthermore, the operationalization of the schema is contingent upon the precise calibration of the Level of Information Need, which prevents computational overhead by restricting data density to the specific requirements of defined project milestones. In practice, the schema finds utility i.e. in the automation of quantity takeoffs and the enhancement of multi-disciplinary coordination through automated data validation; this allows for the identification of "data gaps" where physical components lack the requisite metadata for facility management.

Conclusions

The schema represents a significant advancement in formalizing information requirements specification across different dimensions of detail within the AEC industry, providing technical mechanisms for precise, machine-readable requirement definition. The schema's utility depends not primarily on its technical merits but on ecosystem development including software tool support, reference resource availability, practitioner capability building, and governance framework establishment. These ecosystem elements remain underdeveloped, suggesting near-term adoption will be limited to advanced organizations and pilot projects rather than achieving broad industry penetration. The schema provides necessary technical infrastructure for digital information delivery but proves insufficient without surrounding social and technical systems.

Future success depends on recognizing schema implementation as a socio-technical challenge requiring coordinated development of technical, organizational, and human capabilities rather than purely a standards adoption problem.

Acknowledgments

The authors would like to thank ISO, CEN and national standardization bodies and collaborators for the support throughout the development of the mentioned standards, especially CENT/TC 442/WG 2 and JWG ISO/TC 59/SC 1

References

Abualdenien, J. and Borrmann, A. (2022). Levels of detail, development, definition, and information need: a critical literature review. Journal of Information Technology in Construction, 27:363–392.

Akbas, E., Bolpagni, M., Borrmann, A., Boeykens, S., Mellenthin Filardo, M., Liu, L., and Beetz, J. (2025). A holistic approach to information requirements: Integration of level of information need and information delivery specification. J. Inf. Technol. Constr., 30:731–744.

Bolpagni, M., Bosché, F., de Boissieu, A., Akbarieh, A., Shaw, C., Mêda, P., Puust, R., Medineckiene, M., Popov, V., and Sacks, R. (2022). An explorative analysis of European standards on building information modelling. Proceedings of the 2022 European Conference on Computing in Construction, volume 3 of Computing in Construction.

European Committee for Standardization and International Organization for Standardization (2016). Guidelines for the implementation of the Agreement on Technical Cooperation between ISO and CEN. ISO/CEN, 7 edition.

Hodgson, R. (2024). Quantities, Units, Dimensions and Types (QUDT) Schema - Version 2.1.35. last accessed: 14.01.2026.

ISO 12006-3 (2023). Building construction— Organization of information about construction works: Part 3: Framework for object-oriented information. International Organization for Standardization, Geneva, CH, 1 edition.

ISO 19650-1 (2018). Organization of information about construction works - Information management using building information modelling. Part 1: Concepts and principles. International Organization for Standardization, Geneva, CH, 1 edition.

ISO 23386 (2020). Building information modelling and other digital processes used in construction: Methodology to describe, author and maintain properties in interconnected data dictionaries. International Organization for Standardization, Geneva, CH, 1 edition.

ISO 23387 (2025). Building information modelling (BIM) — Data templates for construction objects used in the life cycle of built assets: Concepts and principles. International Organization for Standardization, Geneva, CH, 1 edition.

ISO 29481-3 (2022). Building information models — Information delivery manual: Part 3: Data schema. International Organization for Standardization, Geneva, CH, 1 edition.

ISO 7817-1 (2024). Building information modelling—Level of information need: Part 1: Concepts and principles. International Organization for Standardization, Geneva, CH, 1 edition.

ISO/TS 7817-2 (2025). Building information modelling—Level of information need: Part 2: Guidelines for Application. International Organization for Standardization, Geneva, CH, 1 edition.

ISO 7817-3 (2025). Building information modelling—Level of information need: Part 3: Data model and schema. International Organization for Standardization, Geneva, CH, 1 edition.

Jaud, Š., Gregory, L. J., Borrmann, A. (2021). Georeferencing. T. H. Kolbe, R. Bill, A. Donaubauer (eds), Geoinformationssysteme 2021: Beiträge zur 8. Münchener GI-Runde, Runder Tisch GIS e.V., Munich, Germany, 49-50.

Jeon, K., Lee, G., Kang, S., Roh, H., Jung, J., Lee, K. and Baldwin, M. (2021). A relational framework for smart information delivery manual (IDM) specifications. Advanced Engineering Informatics, vol. 49, p. 101319.

Mellenthin Filardo, M., Liu, L., Hagedorn, P., Zentgraf, S., Melzner, J., and König, M. (2024a). A standard-based ontology network for information requirements in digital construction projects. Proceedings of the 12th Linked Data in Architecture and Construction Workshop (LDAC 2024).

Mellenthin Filardo, M., Debus, P., Melzner, J., Bargstädt, H.-J. (2024b). Enhancing Information Requirement Compliance Through Automated Integration in a Modelling Environment. Proceedings of the 31st International Workshop on Intelligent Computing in Engineering (EG-ICE), Vigo, Spain.

Sacks, R., Lee, G., Burdi, L., and Bolpagni, M. (2025). BIM handbook: A guide to building information modeling for owners, designers, engineers, contractors, and facility managers. John Wiley & Sons, Hoboken, New Jersey, 4 edition.

Son, S., Lee, G., Jung, J., Kim, J., and Jeon, K. (2022). Automated generation of a model view definition from an information delivery manual using idmXSD and buildingSMART data dictionary. Advanced Engineering Informatics, 54:101731.

Zentgraf, S., Mellenthin Filardo, M., Liu L., Hagedorn, P., Melzner J., König, M. (2025). Update of the standard-based ontology network for information requirements in digital construction projects, LDAC 2025: 13th Linked Data in Architecture and Construction Workshop, July 09-11, 2025, Porto, Portugal.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

Structuring Knowledge for AI-Driven Schedule Management in Digital Construction Twins

Tomas Lay Herrera[1], Matthieu Gillet[2], Mehrzad Shahinmoghadam[3] and Simon Bee[4]
[1]Autodesk GmbH, Germany
[2]Autodesk France, France
[3]Autodesk Research, Canada
[4]Autodesk Ltd., United Kingdom

Abstract

Construction projects, which are prone to delays and budget overruns, struggle to transform fragmented data streams into coherent intelligence. We propose a novel approach to monitoring that combines onsite data capture, agentic workflows and knowledge graph architecture to provide a foundation for evidence-based progress tracking. This approach is built on an ontology for structured schedule management that unifies concepts from domain-specific ontologies and Autodesk Construction Cloud. Autonomous AI-driven agents enable knowledge evolution by interacting with a knowledge graph. This lays the foundation for a living digital twin that adapts as new data emerges, supporting consistency checking and cross-system integration.

Introduction

Construction project execution involves large volumes of heterogeneous data, including schedules, building information models, and field reports, yet this information is rarely transformed into timely and coherent project intelligence. This gap continues to contribute to schedule delays and cost overruns, particularly during the execution phase where coordination and interpretation of evolving site conditions are critical (Gómez-Cabrera et al., 2023). In this context, effective schedule management depends on continuously reconciling plan and actual status, identifying emerging deviations, and relating observed site conditions to the activities and building elements they affect. Digital twins have been proposed as a means to address these challenges by maintaining synchronized digital representations of physical assets and construction processes. In construction, digital twin information systems have been framed as closed-loop mechanisms that explicitly represent and compare project intent and project status to support production control and decision making (Sacks et al., 2020). Central to this perspective is the interpretation of observations in relation to planned work. Prior studies show that progress understanding and predictive analyses become possible when site observations are grounded in a formalized project representation capturing spatial and temporal structure (Lilis et al., 2017). In practice, however, sustaining a "living" construction digital twin remains difficult, as information is heterogeneous, incomplete, and distributed across tools and stakeholders, complicating the maintenance of a coherent and queryable project state. Recent work has increasingly explored semantic and ontology-based approaches to structure construction information for digital twin applications, highlighting the role of explicit representations in integrating planning, monitoring, and derived knowledge across systems. In parallel, knowledge-graph-based approaches demonstrate how BIM data, schedules, and site information can be fused into machine-interpretable representations supporting analysis and downstream services (Katsigarakis et al., 2022). Despite these advances, many approaches either remain primarily representational without clear operational workflows, or focus on analytics without maintaining an explicit, inspectable project state over time. Building on these directions, this paper proposes an ontology-grounded, agentic framework for AI-driven schedule management in digital construction twins. The approach introduces the Intelligent Construction (INTC) ontology, which provides a unified semantic model integrating building topology, construction planning, on-site observations, and construction management concepts drawn from BOT, Digital Twin Construction (DTC), and Autodesk Construction Cloud (ACC). The ontology governs a construction knowledge graph that serves as persistent project state and supports explicit representation of observations, evidence, and state assertions. Over this shared representation, we define an agentic knowledge workflow in which specialized agents ingest project data, align observations with planned activities and building elements, and derive evidence-backed progress and deviation assessments. The framework is evaluated through an observation-to-insight pipeline instantiated in a controlled construction setting and validated using competency-question-driven queries over the resulting knowledge graph.

Related Work

Construction Site Monitoring

Construction site monitoring research has largely focused on two complementary problems: capturing as-built conditions and interpreting those conditions relative to planned scope and sequence. Prior work shows that integrating multiple monitoring modalities with BIM improves the completeness and reliability of project status acquisition compared to any single technology alone (Hasan

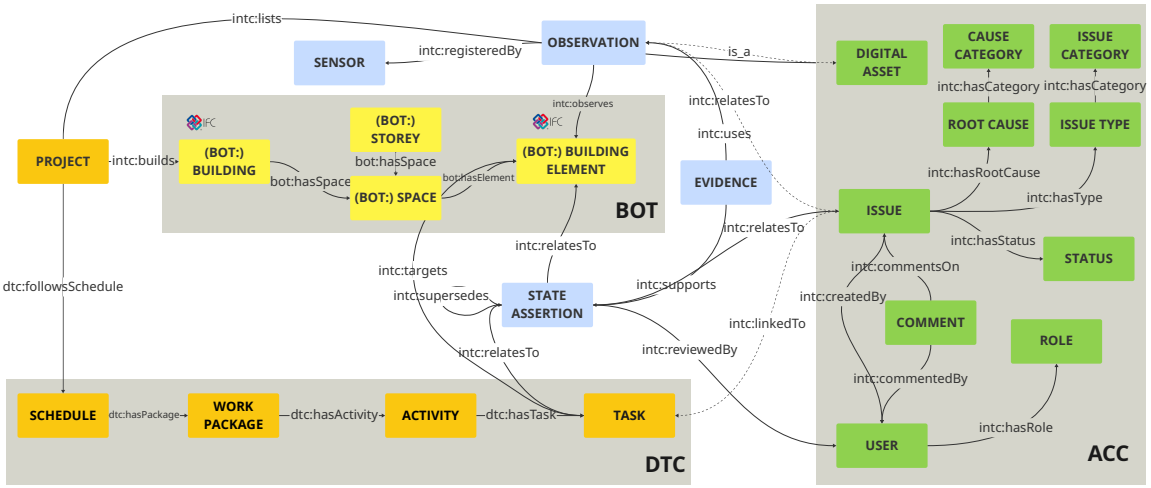


Figure 1: General overview of the INTC ontology

and Sacks, 2023). However, the central challenge remains not sensing per se, but interpreting heterogeneous observations in relation to formal project representations. Vision-based progress monitoring has received particular attention due to the practical availability of site imagery and time-lapse feeds. While recent systems have progressed beyond coarse, binary reporting toward richer state descriptions, they continue to face challenges related to occlusion, viewpoint variability, and robust mapping from detected signals to the project's formal objects and activities (Pal et al., 2023). As a result, visual signals alone often fail to produce reliable, attributable progress assessments. More recent work has explored lifting monitoring outputs into structured project knowledge rather than treating them as disconnected detections. Knowledge-graph based approaches demonstrate how BIM data, schedules, and site observations can be fused into machine-interpretable representations that support analysis and downstream services (Katsigarakis et al., 2022; Pfitzner et al., 2024). However, many such approaches either remain primarily representational, without clear operational workflows, or focus on analytics without maintaining an explicit, inspectable project state over time. This gap motivates our focus on persistent, evidence-linked representations that support plan–vs–actual reconciliation and traceable progress interpretation.

Ontologies for Construction

Ontologies provide a standardized machine-readable terminology to describe complex data and processes present in the construction process, while also enabling semantic interoperability across software capabilities involved in a construction project. Construction ontologies are an active research area, with multiple proposals that go from mid-level ontologies, such as the Digital Construction Ontology Suite (DiCon) (Zheng et al., 2021) or the Digital Twin Construction Ontology (DTC) (Schlenger et al., 2025), which attempt to cover all relevant aspects in the construction process, to domain-specific ontologies for scheduling, BIM data management, compliance and real-time site observations. Getuli (2020) presents a comprehensive study on knowledge modeling in construction planning, while Wu and Ma (2023) automatically generate construction schedules by combining constraint rules and genetic algorithms and Avogaro et al. (2024) proposed an IFC-based application for process scheduling. In the BIM data management field, Balaji et al. (2016) presented a unified schema for building elements, while Pauwels and Terkaj (2016) proposed ifcOWL, an attempt to convert IFC EXPRESS schemas into a usable ontology. In an attempt to address the practicality issues of ifcOWL, Rasmussen et al. (2020) published BOT, a shallower ontology focused on the topological relationships between building spaces. Ontologies such as OntoBPR from Hagedorn et al. (2025), or the Architecture, Engineering, Construction Compliance Checking and Permitting Ontology (AEC3PO) from Dridi et al. (2024) allow the modeling of compliance and permitting process using semantic web tools. Sensor-related ontologies like the Semantic Sensor Network Ontology (SSN) from Haller et al. (2018) or the lightweight ontology for sensors, observations, samples, and actuators (SOSA) from Janowicz et al. (2019) define a comprehensive vocabulary that enables the modeling of site monitoring activities.

AI-driven Agentic Frameworks

Recent work has increasingly framed large language models as agents, systems that interleave reasoning with tool use and state updates to support structured decision-making. ReAct (Yao et al., 2022) formalizes this paradigm by combining reasoning and acting in multi-step workflows, improving robustness and interpretability. While subsequent frameworks extend these ideas with modular orchestration, memory, and tool connectivity, most agent stacks remain domain-general. In construction, effective agent behavior requires grounding actions and inferences in a formal domain model and a continuously maintained project state, motivating our ontology- and knowledge-graph-centered approach.

Proposed Approach

The INTC ontology

Knowledge management in our approach is centered on a construction knowledge graph whose topology is governed by the Intelligent Construction (INTC) ontology. The INTC ontology acts as a connective tissue for four main areas of interest in construction: building information model (BIM), construction scheduling, on-site observations and construction management, the last one covered by Autodesk Construction Services (ACS), linking all these fields in a simple yet robust way. See Figure 1 for an overview of the INTC vocabulary and check (GitHub, 2026) for a version of the ontology, serialized in Turtle format. The BIM-related terminology is built upon an integration of the BOT and ifcOWL ontologies, to increase the coverage of building elements while considering IFC as a common input format for BIM data. The integration with the DTC ontology provides the required scheduling terms, while a one-to-one mapping with ACS concepts allows the definition of construction management concepts at the ontological level. The vocabulary for on-site observations is kept simple, focused less on sensor-related aspects and more on the modeling of insights derived from captured sensor data (see Figure 1 for an overview of the INTC vocabulary). The integration with ACS and the insights-driven on-site monitoring are the main reasons for the design on the INTC ontology.

INTC – ACS connection. ACS provides a comprehensive set of tools to interact with construction data stored in the Autodesk Construction Cloud (ACC). The ACC data model is intrinsically linked to the ACS services; however, there is no ontology to properly describe it. The INTC fills that gap, bringing ACC concepts like roles, users, and issues, among others, and adding semantic relationships between them. In order to avoid replicating the ACC data in the INTC knowledge graph, all entities related to ACC define an URN parameter that links to the corresponding resource for retrieval.

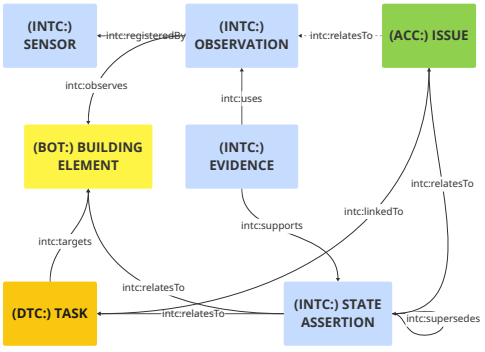


Figure 2: Observations, evidences, and state assertions in the INTC ontology

On-site observations in INTC. The sensor-related module of the INTC ontology aims to describe insights from on-site observations and how they connect with building elements, schedule activities, and potential ACC issues (see Figure 2). An important concept is the "Observation", captured by a "Sensor" (a camera, a microphone or any other sensing device) and used by a "Evidence". An "Evidence" leverages multiple "Observation" to support a "State Assertion" with a certain level of confidence. "State Assertion" is another central concept in this module and can be seen as a sort of update on the building element and the corresponding task in the schedule that targets it. An assertion might also eventually be related to an ACC issue, thus connecting the on-site monitoring with the ACC world. In summary, the core idea behind this design is that the progress of a task, associated with a building element, can be tracked through the associated state assertions, which are supported by evidence built upon observations. The "Ontology Validation" chapter describes a real-life scenario in which all these concepts are leveraged.

Agentic Knowledge Workflow

To operationalize AI-driven schedule management within a digital construction twin, we define an agentic knowledge workflow in which specialized agents operate on a shared construction knowledge graph governed by the INTC ontology (see Figure 3). The knowledge graph serves as the persistent project state and as the main coordination structure: agents record extracted information as typed entities and relations with provenance, timestamps, and confidence. Coordination is event-driven rather than purely sequential, so updates to graph state can activate subsequent processing steps while preserving an auditable record of how project understanding develops over time. The agentic workflow is decomposed into the following roles:

ACC Ingestion Agents (read-only). Retrieve authoritative project artifacts from Autodesk Construction Cloud (ACC) via Autodesk Platform Services (APS) and represent BIM/model artifacts, schedule structures, and documentation/specifications as INTC-aligned entities.

Semantic Linking Agent. Establishes and maintains the links between building elements, tasks, and relevant documentation/specifications required for monitoring, while preserving the provenance of these links.

Visual Observation Agent. Processes site images, registers each image as a DigitalAsset, and creates structured Observation entities linked to the most likely building element and task, while retaining confidence for perception and localization.

Progress and Deviation Agent. Uses newly added or updated observations to build Evidence supporting State-Assertion entities describing task progress and plan vs actual deviation. These outputs should be interpreted as evidence-based assessments rather than as ground truth. Conflicting evidence or assertions may coexist in the graph with their provenance and confidence, allowing later evidence or human review to refine the effective project state without removing prior interpretations.

ACC Sync Agent. Propagates selected progress updates back to ACC via APS. Synchronization is gated: low-confidence or conflicting interpretations are held for hu-

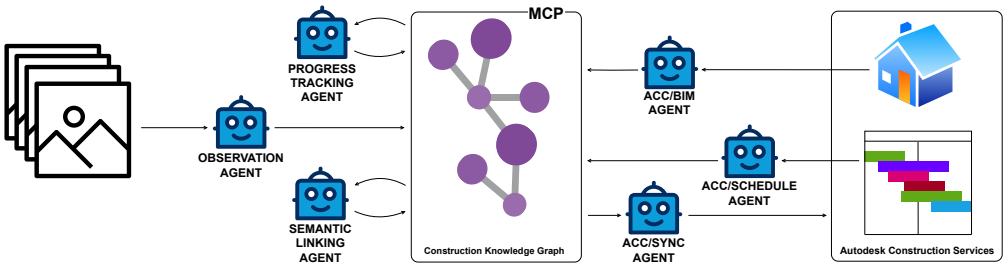


Figure 3: Simple, single-purpose agents operating on the construction graph

man review before write-back.

Overall, ACC provides the authoritative project context, while the knowledge graph provides the semantic structure through which site imagery is connected to observations, evidence, and state assertions for progress tracking and deviation assessment.

Observation-to-Insight Pipeline

The observation-to-insight pipeline converts raw observations into evidence-grounded project knowledge that can be compared against the plan to support activity-level progress monitoring and issue surfacing. The pipeline assumes an initialized project graph (from ACC ingestion and semantic linking) containing Task (planned activities) and BuildingElement entities, plus links to relevant requirements (e.g., execution specifications). This aligns with the DTC architecture that emphasize explicitly managing planning information, monitoring inputs, and derived knowledge in a unified structure (Schlenger et al., 2025).

Figure 4 instantiates the pipeline in the context of image-based progress monitoring as follows: For each new site image, the system (i) registers the file as a DigitalAsset (URI/URN, timestamp, capture metadata), (ii) applies a computer vision toolchain to extract candidate signals (e.g., detected components, stage indicators, potential defects), consistent with vision-based progress monitoring trends in construction (Pal et al., 2023). Next, (iii) it instantiates one or more Observation entities that are explicitly linked to the originating DigitalAsset and grounded to the most likely BuildingElement and Task (preserving uncertainty via confidence scores and, if needed, multiple candidate links). This "data → structured knowledge → metrics" flow is consistent with approaches that lift continuous image capture and object detection outputs into a knowledge-graph representation for construction process analysis (Pfitzner et al., 2024). Finally, with this pipeline in place, (iv) we can consolidate observations over time to maintain a current plan-vs-actual view at the activity level and emits actionable exceptions as Issue entities (e.g., suspected non-compliance or defect), each traceable through an Issue → Observation → DigitalAsset evidence chain. Where appropriate, issues can be exported back to ACC (e.g., as ACC Issues) under guarded policies (confidence thresholds and/or human approval).

While the individual stages of this pipeline (image capture, visual interpretation and progress reporting) are well represented in prior work, the central challenge lies in the alignment step: reliably mapping site observations into the project's object-activity model in a way that remains inspectable over time. In many existing systems, this alignment is either implicit (hard-coded mappings from detections to BIM objects, or transient (performed ad hoc for a specific metric or dashboard), which limits both extensibility and auditability. As a result, downstream "insights" often collapse heterogeneous evidence into a single status signal, obscuring how conclusions were reached and making them difficult to contest or revise as new information arrives.

The pipeline described here treats alignment as a persistent knowledge modelling operation. Observations are explicitly instantiated as first-class entities that mediate between raw digital artefacts and higher-level project claims. This allows multiple, potentially competing interpretations of the same evidence to coexist, each grounded in the project graph with explicit confidence and provenance, rather than forcing early commitment to a single explanation.

Experiments

Building on the pipeline logic introduced above (see Figure 4), we conducted a set of experiments focused on progress tracking using time-indexed site imagery. We assess the effectiveness of the ontology in: (i) grounding visually observable state changes in a knowledge graph instance, and (ii) maintaining an evidence-linked and traceable representation of construction state updates. It should be noted that our evaluation is intentionally scoped to the pipeline's semantic backbone rather than the end-to-end performance of the agentic workflow.

Experimental Setup

Our experiments are based on the ongoing construction of a real-scale mock-up room conducted in one of the Autodesk Technology Center in Boston. Figure 5 shows the 3D BIM representation of the room. The workshop floor is instrumented with multiple fixed cameras positioned to capture time-lapse sequences of construction activities from different angles.

Our setup takes Neo4j, a graph database following the Labelled Property Graph (LPG) as the storage solution for the construction knowledge graph. The INTC ontology is represented in RDF and serialized using the Turtle syntax, and

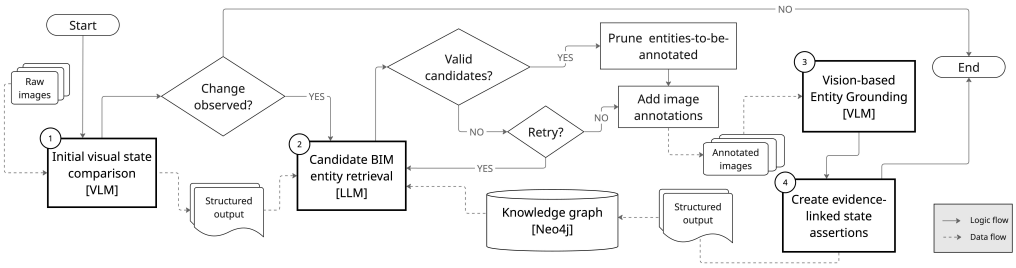


Figure 4: Logic and data flows of the observation-to-insight pipeline in the context of image-based progress monitoring; Steps 1–3 represent agentic enrichment of ingested imagery, while Step 4 establishes evidence and writes state assertions to the knowledge graph



Figure 5: 3D model of the mock-up room

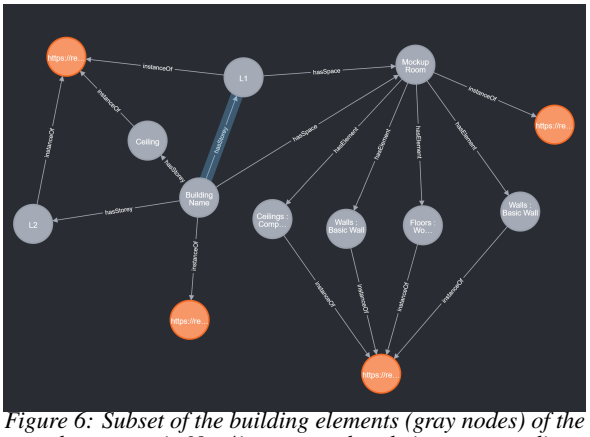


Figure 6: Subset of the building elements (gray nodes) of the mock-up room in Neo4j, connected to their corresponding ontology classes (orange nodes)

we leverage Neo4j's Neosemantics package to import it into the graph, forming the ''TBox'' (vocabulary) layer in the database. In order to link the instance data (''ABox'') with the vocabulary, we label the instance nodes with a single ''Instance'' label and link them with the corresponding class in the ''TBox'' with an ''instanceOf'' predicate, and ensure that all created relationships comply with the restrictions defined in the vocabulary, thereby leveraging ABox consistency while preserving key RDF semantics within an LPG environment (see Figure 6).

Figure 7 demonstrates instantiation of the main INTC ontology constructs, where the main subject of monitoring is the progress (stateType=''PercentComplete'') of one of the wall elements. An overview of instantiating the ontology artiacts for imagery observations is as follows:

Incremental ingestion: The camera captures from the the monitoring interval are represented as Observation instances (with stable external storage identifiers). Note that other sources of information in different formats, such as the site inspection report (text file), can be also represented as observation nodes (see top-left of Figure 7).

Agentic enrichment: The frames from the start and end of the current interval are fed to the image enrichment pipeline (see Figure 4), where a Vision Language Model (VLM) first compares the images to detect visible state changes. Then, the ACC/BIM agent (LLM-based) uses the generated structured change descriptions (in JSON format) and queries the BuildingElement nodes within the Neo4j instance to retrieve a set of likely BIM entity candidates that are set to be visible within the camera's field of view. Finally, it uses annotated overlays (pre-computed) to match each change to the correct BIM GUID by re-invoking the VLM for entity grounding.

Evidence establishment: The enriched observation artifacts become materialized as EvidenceItem nodes. Maintaining this explicit evidence layer is critical because multiple observations may support/contradict a single assertion (derived conclusion), and a single observation may be re-used to support multiple conclusions (e.g., element-level status and task-level completion). Evidence units can also carry indicators and references to intermediate outputs (e.g., confidence, image segmentation masks generated throughout enrichment steps).

State assertion: We use StateAssertion nodes to represent time-scoped and evidence-backed claim about the evolving condition of a subject entity. Each assertion node is associated with a set of key properties, e.g., StateType label (here ''PercentComplete''), state value (here percentage float), validity time (instant or interval), and provenance metadata (e.g., pipeline version). In the scenario depicted in Figure 7, we can see how multiple evidence units (originating from different sources at different times) can support a new assertion about the task's completion and supersede an existing one, while preserving the auditability chain from drawn conclusions back to raw observations.

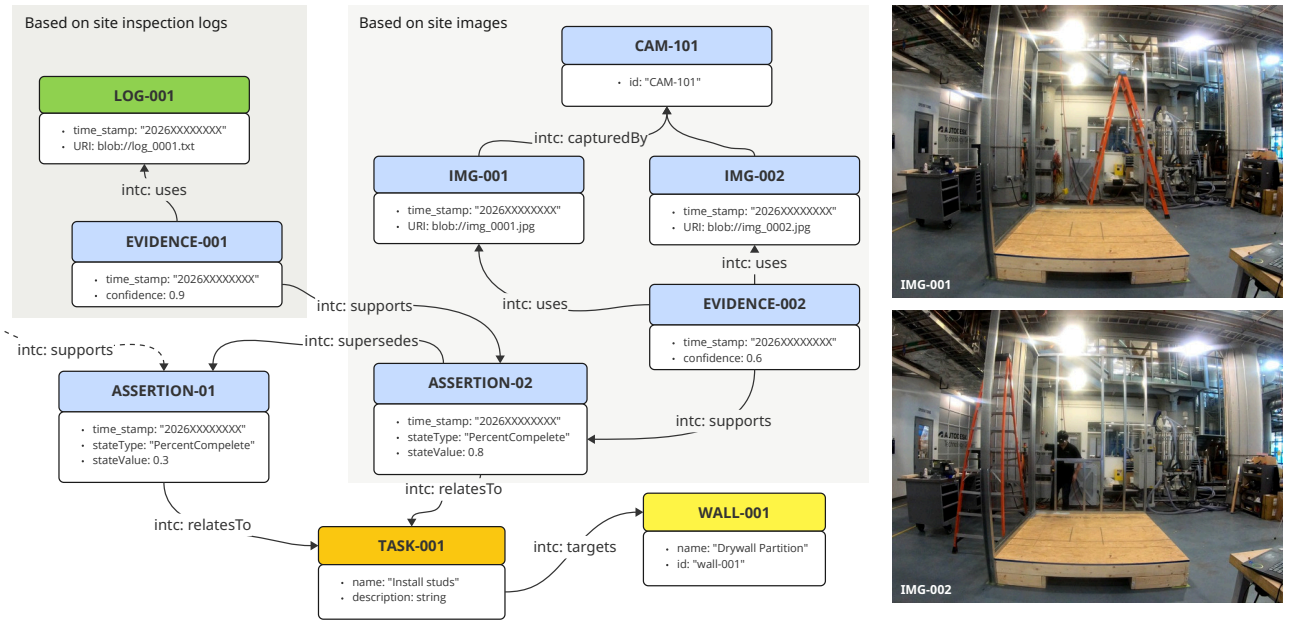


Figure 7: Example of ontology instantiation illustrating observation-to-insight pipeline in an experimental setting

Table 1: Summary of ontology validation competency questions

Category	Validation Purpose	Example Competency Question
Effective Progress State	Resolve multiple associated assertions into a single effective state (via validAt, confidence, etc.)	What is the current effective progress state of a given BuildingElement?
Temporal Reconstruction & Trajectories	Reconstruct progress as-of time t (via validAt); progress state evolution as a time series (trajectory, rates, stagnation)	What was the progress state and completion rate of a BuildingElement as of a time t?
Traceability & Evidence	Enable auditability: trace assertions to evidence, sensors, and raw data references	What evidence (observations) support a specific progress update?
Revisions & Updates	Track non-destructive revisions and state assertions history (when, what, who)	Was a progress state revised? what was the updated value? who initiated?
Spatial Context	Contextualize progress spatially (aggregate/filter by Zone/Space)	What is the progress of a given BuildingElement within a given Space?

```
MATCH (t:Task {taskId: "TASK-001"})<-[:RELATES_TO]-(sa:StateAssertion)
WHERE sa.stateType = "PercentComplete"
WITH sa ORDER BY sa.createdAt DESC, sa.confidence DESC LIMIT 1

MATCH (sa)-[r1:SUPPORTED_BY]->(ev:Evidence)
MATCH (ev)-[r2:USES]->(o:Observation)
OPTIONAL MATCH (o)-[r3:OBSERVED_BY]->(source)
OPTIONAL MATCH (sa)-[r4:RELATES_TO]->(t:Task)

WITH sa, t,
     collect(DISTINCT ev) AS evidence,
     collect(DISTINCT o) AS observations,
     collect(DISTINCT source) AS sources,
     collect(DISTINCT r1) AS supportedByRels,
     collect(DISTINCT r2) AS usesRels,
     collect(DISTINCT r3) AS observedByRels,
     collect(DISTINCT r4) AS relatesToRels

RETURN {
  nodes: {assertion: sa, task: t, evidence: evidence,
observations:observations, sources: sources},
  relationships: {supportedBy: supportedByRels,uses: usesRels,observedBy:
observedByRels,relatesTo: relatesToRels}
} AS subgraph
```

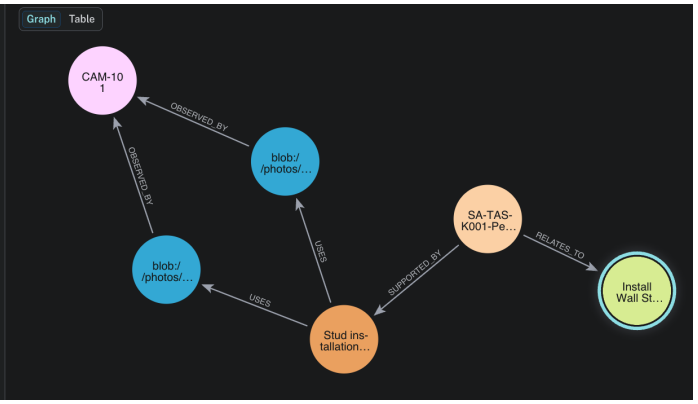


Figure 8: Example of CQ queries: retrieving latest state assertion and its chain of evidence for the data depicted in Figure 7

Validation

We adopt a Competency Question (CQ) methodology (Wiśniewski et al., 2019) to validate whether our proposed knowledge structuring can support the core representational requirements of the intended agentic workflow. The CQs were derived by first distilling the recurring information needs implied by the observation-to-insight pipeline (e.g., grounding, provenance) and then refining and prioritizing them through discussions with the domain researchers involved in the mock-up room deployment. A summary of the key patterns studied here is presented in Table 1. Each CQ is instantiated as Cypher query statements executed over the Neo4j instance, which contains data from the observation enrichment pipeline, as well as synthetic data needed to support the monitoring of the wall's construction progress. An example query associated with the scenario depicted in Figure 7 and it corresponding results are shown in Figure 8. To facilitate reproducibility of the validation results, QC evaluations are represented in a native LPG model (not RDF-based). The full list of questions studied in our experiments can be found in a designated public repository (GitHub, 2026).

Discussion

Our experiments indicate that our proposed ontology can support essential query patterns such as, "what changed," "why do we believe it changed," and "when did this belief become valid/invalid," using explicit representations. One of the key practical implications of our proposed ontology can be attributed to its support for traceable and auditable progress monitoring updates. First, it separates raw observations from interpreted claims about progress, which allows the system to retain historical assertions (including potentially conflicting ones) while updating the current estimate as further evidence becomes available. Second, the explicit linking of assertions to evidence and inference provenance makes each reported progress state inspectable: a user (or downstream analysis) can retrieve the originating sources (e.g., images, site logs), the intermediate evidence artifacts (enriched observation), and the method used to generate the claim (e.g., agent's model version). These aspects are crucial for construction practice, since field teams and project managers often have access to large volumes of signals (images, site reports, schedule logs), but the bottleneck lies in converting those signals into accountable estimates (state assertions) that can be traced back to evidence and acted upon. Moreover, progress updates and issue reports routinely require justification, especially when they trigger downstream actions such as schedule replanning, payment approvals, or rework decisions.

Our proposed representation is intended to be applicable across two complementary consumption modes. First, it can support deterministic workflows through formal queries embedded in conventional software. Second, it offers a structured representation for AI-enabled components and agentic workflows. Importantly, this dual use is enabled by a representation that remains relatively minimal: it captures the constructs required for traceability of evolving states without imposing an extensively granular design that can cause scalability issues for the broader system.

Finally, our evaluation primarily validates CQ answerability, however, does not directly quantify the advantage of the proposed model over existing data structures used in conventional schedule tracking systems. Although we demonstrate that the representation supports critical query patterns, we have not yet measured downstream impacts. In addition, our evaluation setup focuses on the construction of wall elements in a controlled environment. While this scenario can provide a clear sequence of visually observable states, other trade activities (e.g., mechanical/electrical rough-ins) may exhibit subtler visual indicators and greater ambiguity in mapping observations to relevant subjects or require additional sensing modalities. As such, the present contribution should be interpreted as a novel approach towards deviation assessment in construction progress.

Conclusions and Future Work

This paper introduced a semantic backbone for construction digital twins in the context of progress tracking. The proposed ontology provides a principled way to ground heterogeneous site observations into the planned project context by linking them to semantically-related entities (e.g., building elements or tasks). It enables the digital twin to maintain an evolving plan–vs–actual view as a set of time-scoped, evidence-backed assertions that remain queryable, interoperable, and inspectable as project knowledge is incrementally updated over time.

Future work should link the proposed representation to operational metrics and assess how effectively AI agents can exploit it, while clearly separating representational capability from agent performance. Larger-scale validation under more diverse sensing conditions remains necessary. The approach also opens avenues toward quality-aware progress monitoring and evidence-based causal analysis.

References

Avogaro, D., Berlato, M., and Zanchetta, C. (2024). Process scheduling application based on web ontology language representation of the ifc schema. In Construction Management Workshop, pages 125–135. Springer.

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., et al. (2016). Brick: Towards a unified metadata schema for buildings. In Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, pages 41–50.

Dridi, A., Patlakas, P., Lefrancois, M., Beach, T., Petr, H., and Vakaj, E. (2024). AEC3PO: A knowledge model for machine executable construction regulations. Available at SSRN 4784123.

Getuli, V. (2020). Ontologies for Knowledge modeling in construction planning: Theory and application. Firenze University Press.

GitHub (2026). EC3-2026-intc. `https://github.com/tomaslay/EC3-2026`. Accessed: 2026-01-27.

Gómez-Cabrera, A., Gutierrez-Bucheli, L., and Muñoz, S. (2023). Causes of time and cost overruns in construction projects: a scoping review. International Journal of Construction Management.

Hagedorn, P., Fauth, J., Zentgraf, S., Seiß, S., König, M., and Brilakis, I. (2025). OntoBPR: An ontology-based framework for performing building permit reviews using standardized information containers. Advanced Engineering Informatics, 66:103369.

Haller, A., Janowicz, K., Cox, S. J., Lefrançois, M., Taylor, K., Le Phuoc, D., Lieberman, J., García-Castro, R., Atkinson, R., and Stadler, C. (2018). The modular ssn ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. Semantic Web, 10(1):9–32.

Hasan, S. and Sacks, R. (2023). Integrating bim and multiple construction monitoring technologies for acquisition of project status information. 149(7). Publisher Copyright: © 2023 American Society of Civil Engineers.

Janowicz, K., Haller, A., Cox, S. J., Le Phuoc, D., and Lefrançois, M. (2019). SOSA: A lightweight ontology for sensors, observations, samples, and actuators. Journal of Web Semantics, 56:1–10.

Katsigarakis, K., Lilis, G., Rovas, D., González-Gerpe, S., Bernardos, S., Cimmino, A., Poveda-Villalón, M., and García-Castro, R. (2022). A digital twin platform generating knowledge graphs for construction projects. CEUR Workshop Proceedings.

Lilis, G. N., Giannakis, G. I., and Rovas, D. V. (2017). Automatic generation of second-level space boundary topology from ifc geometry inputs. Automation in Construction, 76:108–124.

Pal, A., Lin, J. J., Hsieh, S.-H., and Golparvar-Fard, M. (2023). Automated vision-based construction progress monitoring in built environment through digital twin. Developments in the Built Environment, 16:100247.

Pauwels, P. and Terkaj, W. (2016). EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcowl ontology. Automation in construction, 63:100–133.

Pfitzner, F., Braun, A., and Borrmann, A. (2024). From data to knowledge: Construction process analysis through continuous image capturing, object detection, and knowledge graph creation. Automation in Construction, 164:105451.

Rasmussen, M. H., Lefrançois, M., Schneider, G. F., and Pauwels, P. (2020). Bot: The building topology ontology of the w3c linked building data group. Semantic Web, 12(1):143–161.

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S., and Girolami, M. (2020). Construction with digital twin information systems. Data-centric engineering, 1:e14.

Schlenger, J., Pluta, K., Mathew, A., Yeung, T., Sacks, R., and Borrmann, A. (2025). Reference architecture and ontology framework for digital twin construction. Automation in Construction, 174:106111.

Wiśniewski, D., Potoniec, J., Ławrynowicz, A., and Keet, C. M. (2019). Analysis of ontology competency questions and their formalizations in SPARQL-OWL. Journal of Web Semantics, 59:100534.

Wu, Z. and Ma, G. (2023). Automatic generation of bim-based construction schedule: Combining an ontology constraint rule and a genetic algorithm. Engineering, Construction and Architectural Management, 30(10):5253–5279.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R., and Cao, Y. (2022). ReAct: Synergizing reasoning and acting in language models. In The eleventh international conference on learning representations.

Zheng, Y., Törmä, S., and Seppänen, O. (2021). A shared ontology suite for digital construction workflow. Automation in Construction, 132:103930.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

CONNECTING DOTS: HARNESSING DIVERSE PERSPECTIVES FOR THE INTEGRATION OF DIGITAL BUILDING PERMITS AND LOGBOOKS

Pedro Mêda[1], Judith Fauth[2]
[1]Project Manager/Invited Assistant Professor, CONSTRUCT, ICS – Institute for Sustainable Construction, University of Porto, Faculty of Engineering, Porto, Portugal
[2]Postdoctoral Researcher, Technical University of Munich, Munich, Germany

Abstract

The integration of Digital Building Permits (DBP) and Digital Building Logbooks (DBL) is increasingly recognised as essential for sustainable and data-driven construction and building lifecycle management. This study examines how European initiatives, including permitting platforms, logbook gateways, city Digital Twins and building stock observatories, contribute to DBP and DBL integration, often implicitly. Using a step-by-step thematic analysis of stakeholder and expert interviews, the research identifies shared data foundations, maturity gaps and governance challenges. The findings outline initial pathways for aligning fragmented initiatives and support the development of coherent and interoperable DBP and DBL ecosystems.

Introduction

The pace of digitalisation in the built environment sector has gained new momentum amid growing sustainability focus (Papadaki *et al.*, 2023). Requirements in terms of energy efficiency, the rational use of resources, and the implementation of more circular practices have led to new demands for data availability and accessibility, while also giving purpose and direction to the digitalisation of processes (European Parliament and European Union Council, 2022). Data-driven instruments, long assumed to be needed, are emerging, such as the Digital Product Passport for Construction (DPP_C) and the Digital Building Logbook (DBL) (Papadaki *et al.*, 2023). These instruments are intended to span many domains, aiming to meet the abovementioned requirements. Also, they are meant to work in combination with existing digital technologies such as Building Information Modelling (BIM) and Geographic Information Systems (GIS), to name a few (Calvetti, Mêda and Sousa, 2021). There is the intuition that several initiatives not specifically targeting DPP_C and DBL are already using or will need to use the same data. Not recognising the connections can introduce several constraints on the path towards efficiency, sustainability, and modernisation in the construction and built environment sectors. Therefore, this study seeks to identify whether there are connections to be made by considering diverse perspectives. Based on the observations, it aims to provide initial guidelines that support different paths to achieving a common goal, with a main focus on DBL.

The DPP_C focuses on product data. Within the context of the new Construction Products Regulation, it serves to disclose the declaration of performance and conformity, including the essential characteristics, and to enable harmonised marketing across the EU (European Parliament and the Council of the European Union, 2024b). Furthermore, the digital tools developed are designed to preserve the data and expand their use, as the products will be incorporated into construction projects and built objects (European Parliament and the Council of the European Union, 2024b). The DBL is building-centred (Mêda *et al.*, 2022). It is envisaged as a multi-level, multi-layer information container that collects, connects, and supports relevant construction-related data at land cadastre, built object, and building unit levels, to be made available for multiple purposes (Grow, 2023). Its initial motivation was the energy performance of buildings, establishing a common framework and data points for this specific output (Gómez-Gil, Espinosa-Fernández and López-Mesa, 2022). However, it was realised that, given the data requirements, a set of additional purposes could be achieved within the same effort (Gómez-Gil *et al.*, 2022). Notwithstanding, the energy-related purpose has been a major driver, and this is now part of the EU legal framework under the new Energy Performance of Buildings Directive (EPBD) (European Parliament and the Council of the European Union, 2024a). The observation of the relationships at the building level and construction process life cycle led to exploratory research on the alignment, benefits and synergies with another long-recognised research domain, critical for the efficiency of building construction; the building permitting process (Noardo *et al.*, 2022; Fauth and Seiß, 2023; Fauth *et al.*, 2025). Further, the EU's Renovation Wave strategy and the Affordable Housing plan heavily rely on data to increase knowledge of the European built stock, its uses, performance, and resource potential, among other key figures (European Commission, 2020, 2025). Digital instruments are needed to collect, manage and process that data. The DBL, the Digital Building Permit (DBP), other associated concepts and the Digital Twin (DTw) platforms are fundamental pieces to make it happen (CEN/TC 442, 2026)(Calvetti *et al.*, 2025).

The present study, grounded on the constructivist paradigm (Cobb, 2006), builds on recent works that have demonstrated the relationship between the DBL and the

DBP concepts and, subsequently, the added value of assessing and forecasting the potential synergies between them (Mêda, Fauth, Schranz, Sousa, *et al.*, 2024; Mêda, Fauth, Schranz, Urban, *et al.*, 2024). Moreover, this work aims to explore further how various initiatives currently underway can contribute to this integration (Fauth, J.; Mêda, P.; Gisbert Ferrer, P.; Flickenschild, M.; Babok, 2025). A step-by-step thematic analysis (Naeem *et al.*, 2023) is conducted based on interviews with qualified stakeholders with in-depth knowledge of the permitting and logbook domains. The selection is related to their involvement in the development and implementation of initiatives in affiliated domains. This enables qualitative assessments based on diagrams to define a conceptual model, supported by a maturity framework, for DBP and DBL integration. The aim is to discuss the need to set paths that can help industry stakeholders understand the landscape of concepts and how initiatives pointing to other goals can be adjusted or already partially cover the scope of what DBP and DBL want to achieve. In this sense, the research questions are as follows: Considering different existing initiatives, how do they perceive DBP and DBL integration? Considering their specific objectives, how can these initiatives contribute towards that integration? Which actions (maturity increments) can be done to foster that integration?

Following this Introduction, the present research starts with a background that systematises the problem and the contributions from previous research. The explanation of the methodology employed follows to present the results from the thematic analysis. A discussion on the main aspects to be considered as part of a maturity framework is provided, together with study limitations and future perspectives. The work finalises with a conclusions section in which the answers to the research questions are presented.

Background

The Transition pathway for Construction has paved the way for the industry to advance towards the green and digital transition goals (Papadaki *et al.*, 2023). Despite developments and evolutions in understanding, this document remains a cornerstone of the EU's strategy for a Sustainable Built Environment (SBE).

The accomplishment of an SBE relies on the ability to have available buildings/built stock data for a wide range of purposes, from theoretical models of energy efficiency to energy performance monitoring, space identification and their maintenance plans and requirements, bill of materials, quantities for reuse/recycling, change of ownership, dangerous substances in products, planning of investment needs, among others. The DBL, working as a DTw enabler at the building level (Staff, 2021), helps unlock the buildings' unawareness of these dimensions.

However, a DBL needs to be deployed early in the design phases and will need to interact with databases, processes, platforms, and tools to link, collect, organise, compile, and manage data points (Mêda *et al.*, 2022). While BIM is often identified as one of the main components of digitalisation strategies for construction, those often struggle to explore metadata or the Level of Information Need (LOIN) in the BIM model (ISO, 2018). Data-driven processes, such as building permitting, require a clear identification of the required data points (Bloch and Fauth, 2023). Significant research exists on the digitalisation of building permits, with some focusing on BIM (Ullah, Witt and Lill, 2022).

Interestingly, despite the research community's growing interest in DBL, this remains quite unknown (Fauth, J.; Mêda, P.; Gisbert Ferrer, P.; Flickenschild, M.; Babok, 2025), meaning that stakeholders recognise the need for a type of database or information container capable of capturing and retaining the necessary data, but don't assume it as a logbook. This realisation was only perceived during the study on the synergies between DBL and DBP (Fauth, J.; Mêda, P.; Gisbert Ferrer, P.; Flickenschild, M.; Babok, 2025). The study findings confirmed that integration between logbooks and permits is both feasible and emerging, creating opportunities to reduce duplication, automate compliance checks, and streamline processes. The benefits range from improved housing delivery to more accurate environmental performance and administrative efficiency. It is worth highlighting efficiency gains from faster procedures, as well as improved transparency and trust through centralised, validated data (Fauth, J.; Mêda, P.; Gisbert Ferrer, P.; Flickenschild, M.; Babok, 2025). However, not understanding how relationships and interactions work across different perspectives or purposes is more than just a challenge; it can lead to misunderstandings. These misunderstandings can result in shortcomings, inefficiencies, and, more critically, blockers and frustrations regarding the innovations being implemented.

Research on permits is more extensive, with multiple relevant contributions. Recently, several European projects have been finished. These have helped people understand the differences between EU countries and the plans to make building permits more digital (European Commission, 2023). In this context, and with relevance to the present study, four maturity levels have been established regarding DBPs. They are as follows: 1) paper-based, 2) digital paper, 3) digital BIM-based process, 4) digital BIM-based process with GIS (European Commission, 2023).

Methodology

As mentioned, the present research builds on previous work, including desk research (Moore, 2018) and literature reviews, to provide an overview of existing scientific studies and strategic documents at the EU level focusing on DBL, DBP, and their combination. As part of those studies, a series of semi-structured interviews was performed to capture impressions from a sample of relevant stakeholders directly involved in digitalisation initiatives, with additional insights from experts (Fauth, J.; Mêda, P.; Gisbert Ferrer, P.; Flickenschild, M.; Babok, 2025). These interviews lasted about 1 hour and were conducted in October/November 2025 (Table 1). A semi-structured interview guideline was intended to cover several topics related to permitting and logbook

integration, with a focus on housing projects. The interviewees gave their permission for transcription. To ensure confidentiality, sensitive data were anonymised.

Table 1: Example of a table where the caption should be on top of the table

	Stakeholder type	Exper-tise	Country Region	Experi-ence
1	Municipality Dep. Head	DBP	Estonia	10 years
2	Project manager	DBL	Flanders	10 years
3	Project manager	DBP/GIS	The Netherlands	10 years
4	Municipality Dep. Head	DBP	Austria	15 years
5	Project manager	EPBD	Europe	15 years

Given the specific contexts and objectives, different conventions were used to express similar topics. Therefore, a step-by-step thematic analysis process, as proposed by Naeem (Moore, 2018), was followed. This process comprises the following steps: 1) Transcription, familiarisation with the data and selection of quotations, 2) Selection of keywords, 3) Coding, 4) Theme development, 5) Conceptualisation through interpretation of keywords, codes and themes. Despite familiarity with the data, transcription and selection of quotations from the recordings were performed manually for each interview from start to finish. This was because the important information was scattered across different places rather than presented as answers to specific questions. The next step involved selecting relevant keywords to explain and define the objectives and key features of the initiatives. However, because the interviewees had different nationalities, it was necessary to take two distinct analytical steps to ensure everyone understood the topic the same way. 1) Relationship between different concepts in each initiative; 2) Classification of keywords in the following categories: “scope”, “data type”, “process/deliverable”, “concept”, “requirement/rule” and “technology”. Implementing these steps involved coding and theme development. An inductive coding approach was practised as the concepts emerged directly from the interview data. To ensure coding consistency, an audit trail was established. From the initial narrative descriptions of each initiative, keywords were refined to yield the final themes, along with diagrams summarising the descriptions. Specific colours and shapes are used to highlight the labelling in the diagrams. These elements are detailed in the following section.

Results

This section addresses the individual results based on the thematic analysis steps. Descriptive overviews of the initiatives and the simplified diagrams are presented. Due to space constraints, the description was preferred over the diagrams. However, a more complete graphical perspective is provided for the first case presented to exemplify the work performed. Additionally, authors use dashed lines on the diagrams to show ongoing and future developments. As mentioned before, from existing studies and the initiatives to be presented, the integration between the DBP and the DBL has not been the focus or the driving objective. The section ends with an analysis of all the cases, identifying the overlaps and setting a common framework.

Digital construction platform (Estonia)

Regarding the e-construction platform initiative in Estonia, here defined as digital construction platform (for generalisation purposes), the developments were focused on achieving a multi-module solution that could 1) Contribute to the establishment of a Digital Twin at the country scale; and 2) Set a governance framework for construction processes life cycle digitalisation. To put these aspects in place, BIM and GIS were considered, along with existing legacy documentation and data points, such as the national building registry, the cadastral database, and land spatial planning. Other relevant data types were considered, such as the ownership, tax data, utilities networks and consumption data (energy and water). Some of these legacy data types originate from building permitting processes; therefore, BIM-based permitting was implemented to streamline compliance checks and assessments for issuing permits and to make relevant data available for other phases of the construction life cycle. Although the DBL is not a clear and individualised concept in the system, many of the data types, data points and potential features are considered, making its materialisation as a “*de facto*” feasible concept (Figure 1).

In summary, this initiative demonstrates a high level of maturity in digitalisation and, more importantly, in the relationship between processes, requirements, data types, and legacy databases to materialise a Digital Twin at the country level. A strong concern was expressed regarding the necessity for the country to invest in achieving a common national information ground (for buildings and other construction types and layers), despite the potential interest in supporting owner-specific needs. The building dimensions are paramount in the strategy, and although the DBL was not recognised as a component of the system architecture, its role, positioning, and potential uses are recognised. Moreover, substantial DBL data and associated features are already embedded in the system.

Logbook-centered initiative (Flanders)

This initiative assumes as its main objective providing a single data gateway for building owners (residential and non-residential) to have access to relevant buildings' legacy data. Many features of the DBL are covered by this gateway, which centralises existing data from Governmental and municipal databases, including those for spatial planning and environment, heritage, building registry, energy and water supply and consumption, building quality, materials, and waste. It should be noted that this is official national data. However, building owners are at liberty to contribute additional information and data points to the gateway, including alterations and

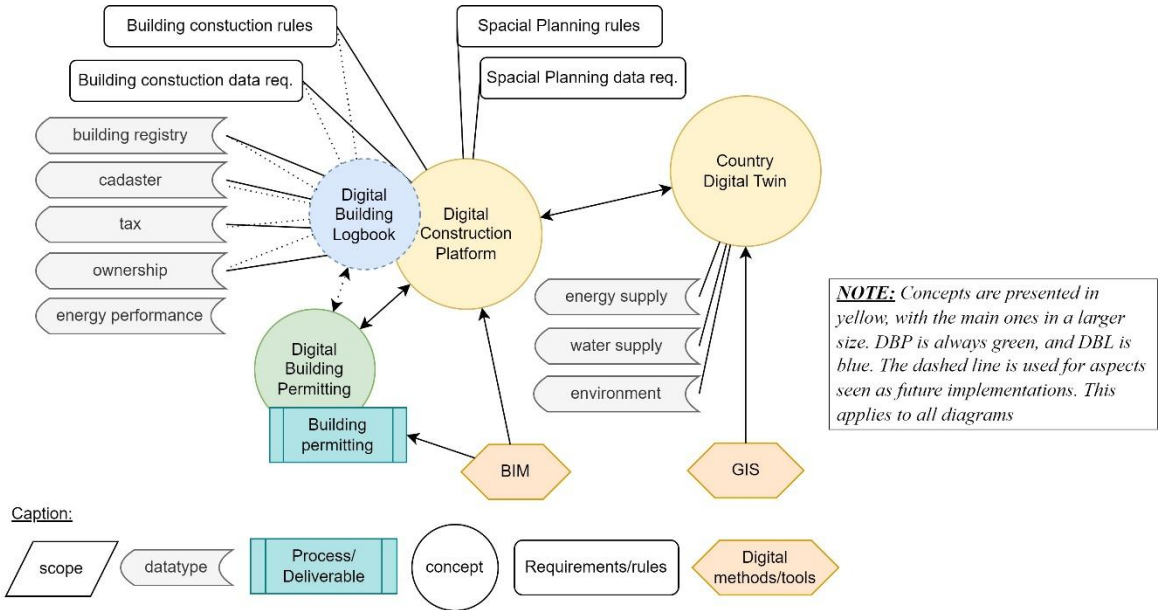


Figure 1: Schematic diagram of digital construction platform

maintenance actions. Following an initial development, setting a relationship with the building permitting processes was identified as a strategic step. In this respect, metadata and documentation for the initial permit and subsequent permits can be tracked. It is recognised that this DBL's primary concern is capturing legacy data and making it available. Improved features are required to capture additional data types and more dynamic data. In this respect, the forecasted steps align with the goal of developing a city platform where these features will be developed and integrated with the DBL (Figure 2).

In this case, the initiative began without searching for an integration between DBL and DBP. During development and implementation, it was recognised that this would not only make sense but could also add value to the initiative's governance.

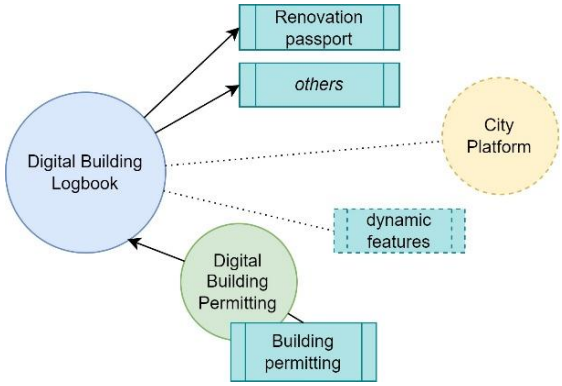


Figure 2: Simplified diagram for the initiative focused on the DBL.

City Digital Twin (The Netherlands)

The aim of this initiative is to build a platform that can materialise a digital twin at the city level. At the core of the goals is improving city operations in terms of efficiency, costs, and resource consumption. To accomplish them, several data sources were brought together from city maps, georeferencing systems, transport network information, and buildings' blueprints. The next step aimed to create a 3D city model. In this respect, and with regard to buildings, this requirement aligns with existing building permitting requirements, which already demand a BIM methodology and establish a set of information delivery specifications focused on openBIM. Linked data is used to aggregate data from the building level to the city level (Figure 3).

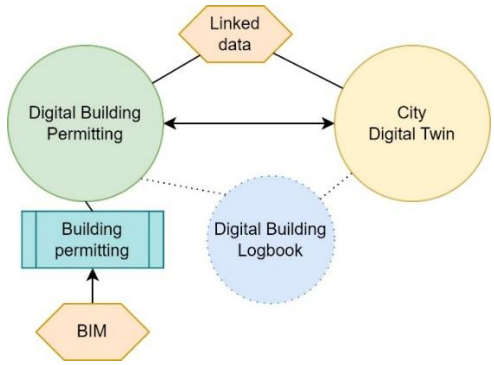


Figure 3: Simplified diagram for the initiative focused on the city's digital twin.

In this process, it made sense to integrate BIM-based permitting processes into the city's digital twin platform to enable observation of construction evolution related to DBP procedures and to streamline compliance checks for subjects such as spatial planning and available utilities. It is recognised that, for now, most available building data reflects the "as-design" condition, and only specific data points reflect the "as-built". However, this is part of the developments under consideration.

In summary, although the DBL was not recognised as a component for the city's digital twin systems architecture, its role, positioning, and potential uses were perceived. It was clear that logbook data and some features are already embedded in the system, although not structured with the vision of providing a structured Digital Building Logbook.

Single gateway for municipal permitting (Austria)

In this case, the initiative originated from the idea of centralising all permitting procedures on a single platform. Considering the wide range of permits that citizens can request from an authority, from kindergarten for children to car parking spots, urban waste containers, or construction operations, just to name a few, the objective was to provide a common access point. In this respect, building permits were identified as one of the relevant types to consider. The developments centred on dematerialising the process by establishing digital mechanisms and disclosing common rules for submissions by type. This new system links with the authority that receives all required documentation in PDF format to start the compliance checks. It is recognised that, at this moment, the system is working as intended, but there is the notion that improved digitalisation can be incorporated into building permitting procedures, namely through the use of BIM, augmented reality, and artificial intelligence (Figure 4). In this respect, considering the growing requirements for permitting in terms of data; as CO_2 consumption, materials used and recycling potential, among others; having all building-related data in one place — the logbook — would improve the initiative's goals.

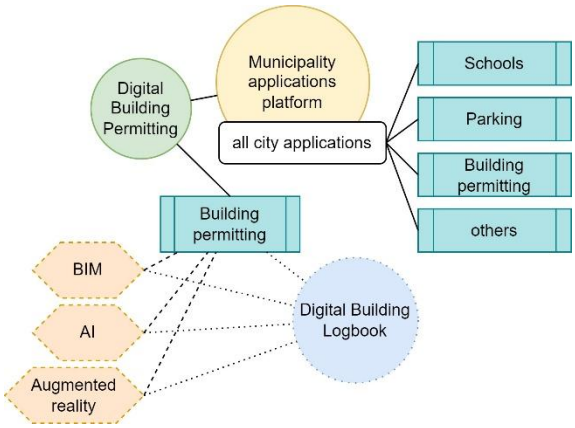


Figure 4: Simplified diagram for a single gateway for municipal permitting.

The building permitting processes require substantial documentation, and new requirements are being introduced that will increase the amount and complexity of that documentation and the compliance-checking procedures. It is acknowledged that digital technologies will be pertinent in addressing and rendering the new requirements tangible. The concept of a "data lake" storing all pertinent building information with the objective of streamlining the permitting process is both desirable and advantageous for a range of other requirements that extend beyond the scope of permitting. In this respect, it is important to be aware of how a DBL can quickly meet requirements for materials and recycling.

Building stock observatory (Europe)

Despite the different scope and dimensions, addressing this initiative is relevant given its alignment with the research's scope and the additional layer it adds to national-level initiatives. The EU building stock observatory is associated with the Energy Performance of Buildings Directive (EPBD) and aims to cover a broad range of energy-related topics, including energy consumption, building elements and installed technical building systems, energy performance certificates, nearly zero-energy buildings, and renovation rates. In this respect, renovation passports and the energy performance of buildings databases defined in Article 22 of the EPBD constitute key data sources (European Parliament and the Council of the European Union, 2024a). The EU building stock observatory combines data from member states on the energy performance of buildings, building types, building registries, and the number of permits by type. This information originates from multiple databases maintained by statistical bodies and authorities responsible for issuing energy performance certificates (EPCs). While these mechanisms are recognised as being important for the observatory's initiation, it is acknowledged that reducing data duplication, enhancing systematisation, and ensuring more reliable and accurate data are pivotal considerations. In this respect, the logbook is envisaged serving as the repository for the building's data, interacting with and collecting data from processes throughout the building's life cycle, such as permitting, renovations, maintenance, and other events (Figure 5).

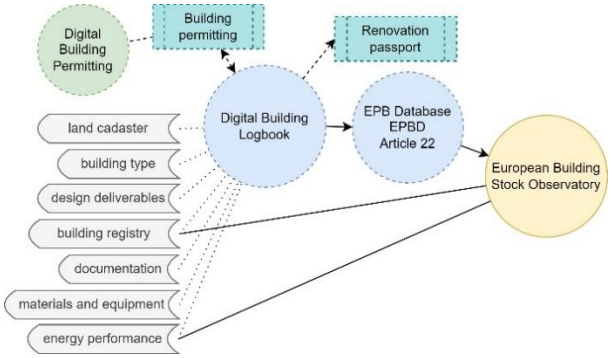


Figure 5: Simplified diagram focused on the Building Stock Observatory.

Despite the fact that this initiative is firmly rooted in the energy performance of buildings, it was acknowledged that its success will be contingent on the successful organisation of all relevant building data. Energy-related data constitutes only a fraction of this, with additional data sources including the technical solutions employed, equipment, spatial dimensions and orientation, among others. In this particular context, it is noteworthy that the DBL is mentioned in the EPBD in relation to the renovation passport. Considering that a substantial part of the observatory data comes from statistical offices, it is relevant to note that its origin lies in permitting processes. Therefore, although not explicitly mentioned, the outcomes of permitting play a critical role in enabling the observatory as it works today, suggesting that more digital processes in permitting would be beneficial.

Setting a common ground

Despite the varying scopes of the initiatives, a strong overlap and a sequencing of objectives can be established, as these work with building data and processes associated with building construction (building permitting) and operation (energy efficiency). Of greater significance is the observation that DBP and DBL are prevalent despite the variations.

The thematic analysis indicates that the integration of DBP and DBL is widely recognised as both desirable and feasible. However, it was not explicitly included in the initiative's initial objectives. Furthermore, in the post-interview period, integration is regarded as being of the utmost importance. This, despite the general absence of DBL realisation at the time when the development strategies for the initiatives were formulated. In most cases, this is justified because the DBL studies were still being developed or because the widespread understanding of its role was still very low.

It is also worth noting that existing data sources are already related, linked, or associated with DBP or DBL-like functionalities. Building registry data, ownership, tax, design deliverables, energy performance, or legal documentation for building use constitute good examples. However, in terms of support, these can still be paper-based, PDF-based (images or metadata), GIS- or BIM-based (metadata), or a combination of these. On another level, city or country existing data sources, such as land cadaster or spatial planning, as well as characteristics of facilities such as water and energy supply, environmental constraints, among many others, also provide similar support. This represents a critical aspect that enables not just logbooks and permits, but also all the objectives mentioned for each initiative. In this respect, the gaps regarding a common national information ground, the uneven maturity across municipalities/countries, the confusion about definitions (stakeholders may want a "data container" but not call it a logbook), the varying data quality ("as-design" vs "as-built"), the governance and ownership questions, the compatibility with legacy systems, and scalability of metadata/LOIN definitions should be assumed as main barriers and concerns. Figure 6 aims to combine, in a single diagram, all the concepts, data types, digital methods/tools, and processes/deliverables that were approached individually.

Discussion

Digital Building Permits and Digital Building Logbooks are common and are perceived as critical to the success of initiatives like the ones explored. Increasing the maturity of each concept and working on its integration will expand the capabilities of city and country platforms, boosting the data availability, relationship and accuracy for the European Building Stock Observatory and, more importantly, improving the knowledge on buildings and their units for a wide range of stakeholders and purposes, from ownership management, real estate, city governance, and networks capacity, among many others.

Previous studies (Fauth *et al.*, 2024, 2025) demonstrate that building permits vary substantially from a process perspective. However, the object is the same, meaning that there is a common ground for the data needed to run these processes. Digital Building Permits already have maturity levels associated. It is evident that the initiatives under consideration vary in their DBP maturity, as indicated by the available information. With regard to the DBL, the establishment of maturity levels remains to be accomplished, a prerequisite for the formulation of a maturity framework centred on their integration.

In this respect, the present research points out some relevant aspects that will need to be considered in that definition, such as:

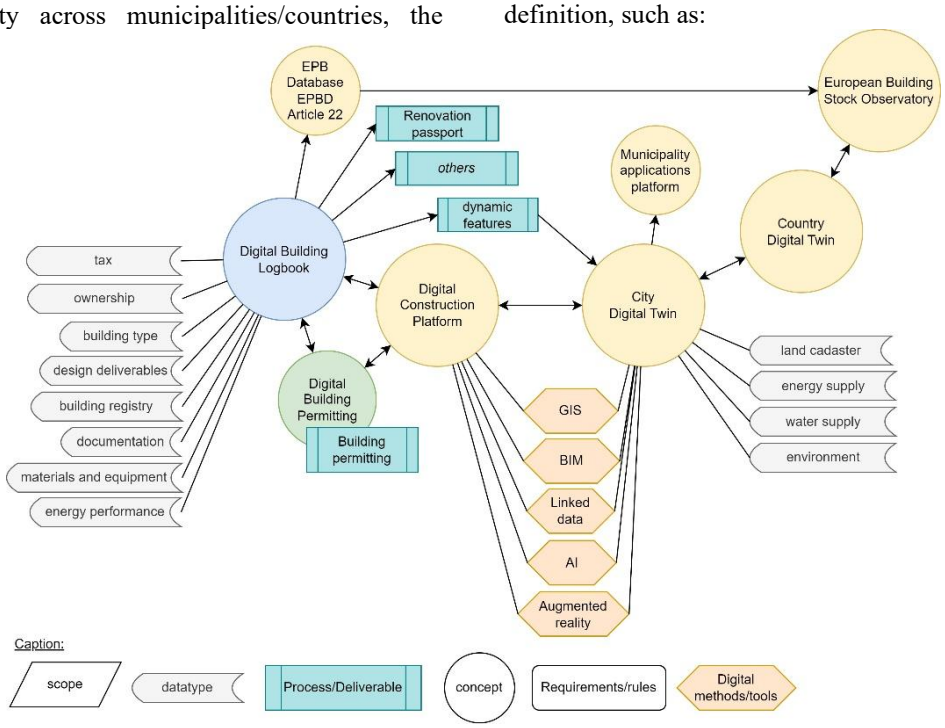


Figure 6: Complete diagram merging the initiatives.

- Identification of the life-cycle phases and associated data points covered, example “as-designed”, “as-built”, “as-use”, “all”, “only few”
- Data inputs, in relation to existing databases, such as land cadaster, permits, tax, ownership, others
- Data formats, in which form is the data associated to the DBL (similar to the DBP maturity levels)
- Use of GIS and BIM as supporting ground for the data (yes, no, to what extent)
- Level of integration, made visible to what extent the data overlaps are harmonised
- Data points status, dynamic vs static data
- Range of features covered or already implemented on the DBL (energy efficiency, ownership assessment, permitting, renovation passport, material passports, among others)

It is imperative to emphasise that the initial point of adoption, as evidenced by the diverse initiatives, has a substantial influence on subsequent integration, given their divergent approaches to the aforementioned aspects. In this regard, there is a perceived need to expand the studies and make a stronger push to define a comprehensive common framework for DBL.

Finally, it was observed that the motivations for the initiatives vary substantially. From country- and municipality-level strategic visions to meet legal requirements, there are several enablers for similar initiatives to take place. Notwithstanding, the clarification regarding DBP and DBL adoption and their integration is perceived as a critical enabler or barrier. Given the wide variety of legal diplomas involved, the EU, its member states, and municipalities' legal frameworks will need to provide stakeholders with clear guidance on how to approach these concepts.

The present research has limitations that are associated with the number of interviews and initiatives approached. In addition, the focus was more on the initiatives' organic aspects, seeking similarities rather than on the data points covered, which will be a topic for future research.

Working on all the abovementioned aspects, focusing on the existing data and how it can support the integration, and expanding knowledge of other initiatives are part of future research actions.

Conclusions

While previous studies were more theoretical, this study investigated existing real-life initiatives, bringing a new level of realisation on regarding the integration of Digital Building Logbooks and Digital Building Permits.

Considering the initial research questions, the following conclusions can be drawn:

- Integration between DBP and DBL is perceived as necessary and beneficial, but typically emergent rather than planned. Initiatives often embed DBL-like data without naming it, and they view DBP/DBL integration as a logical step to reduce duplication, enable compliance automation, and support policy goals.
- The perception varies depending on the scale/level: country platforms think in terms of governance and canonical data; city platforms focus on process simplification; observatories focus on data quality and standard datasets.
- It was observed that each initiative brings complementary assets:
  - Digital construction platforms bring governance, canonical identifiers, cross-domain linking, and scalability;
  - DBL gateways: owner-facing data aggregation, provenance/version control, owner workflows;
  - City Digital Twins: spatial integration, visualisation, operational data linking, as-built updates;
  - Permitting platforms: front-end submission workflows, compliance checking, documentation-based or BIM-based, metadata validation;
- Building stock observatories: regulatory requirements, standardised datasets, reporting needs that motivate upstream digitalisation.

As mentioned in the introduction, in addition to DBP and DBL, the Digital Product Passport for Construction integration within this framework is paramount and warrants further research.

Recognising the synergy between DBP and DBL is imperative, as only by overcoming the fragmentation of concepts can we avoid barriers to innovation and unlock the true potential of a modern, transparent built environment.

Acknowledgements

The authors would like to thank all interviewees, the High-Level Construction Forum and DG GROW. Pedro Mêda was supported by: UID/04708 of the CONSTRUCT - Instituto de I&D em Estruturas e Construções - funded by Fundação para a Ciência e a Tecnologia, I.P./ MCTES through the national funds and DPP4SBE project: “Digital Product Passports for Sustainable Building Environment” (COMPETE2030-FEDER-01456700) funded by PORTUGAL 2030 program; Judith Fauth: brain-SCC GmbH under Grant R&D project “Development of a process model and identification of innovative approaches to the building permit procedure in the building life cycle”.

References

Bloch, T.; Fauth, J. (2023) ‘The unbalanced research on digitalization and automation of the building permitting process’, *Advanced Engineering Informatics*, 58(June), p. 102188. https://doi.org/10.1016/j.aei.2023.102188.

Calvetti, D. *et al.* (2025) ‘Incremental digital twin framework: A design science research approach for practical deployment’, *Automation in Construction*, 170(2024).https://doi.org/10.1016/j.autcon.2024.105954.

Calvetti, D., Mêda, P. and Sousa, H. (2021) ‘Tech

enablers to the EU Renovation Wave: Framework-based on the Communication (2020) 662', in *38th CIB W78 conference on Information and Communication Technologies for AECO*. Luxembourg: LIST. http://itc.scix.net/paper/w78-2021-paper-061.

CEN/TC 442 (2026) *EN 18162 Building Information Modelling (BIM) - Digital twins applied to the built environment - Concept and definitions.* CEN.

Cobb, T. (2006) 'Constructivism', in *Encyclopedia of Language & Linguistics*. Elsevier, pp. 85–87. https://doi.org/10.1016/B0-08-044854-2/01593-5.

European Commission (2020) *A Renovation Wave for Europe*, *COM(2020) 662 final*. Brussels.

European Commission (2023) *Digital building permits toolkit for building public authorities*. Brussels. https://build-up.ec.europa.eu/en/resources-and-tools/publications/toolkit-digital-building-permits-municipalities.

European Commission (2025) *The European Affordable Housing Plan*. Brussels. https://housing.ec.europa.eu

European Parliament and European Union Council (2022) *2022 Strategic Foresight Report - Twinning the green and digital transitions in the new geopolitical context (COM(2022) 289 final)*. Brussels.

European Parliament and the Council of the European Union (2024a) *Directive (EU) 2024/1275 of the European Parliament and of the Council on the energy performance of buildings (EPBD)*.

European Parliament and the Council of the European Union (2024b) *Regulation (EU) 2024/3110 laying down harmonised rules for the marketing of construction products (CPR)*. Official Journal of the European Union.

Fauth, J.; Mêda, P.; Gisbert Ferrer, P.; Flickenschild, M.; Babok, C. (2025) *Synergies between Digital Building Logbooks and Building Permitting*. Brussels: High Level Construction Forum.

Fauth, J. *et al.* (2024) *Comparative study on building permit processes in Europe - European Network for Digital Building Permit EUnet4DBP Publication Series*. EUnet4DBP edn. https://doi.org/10.5281/zenodo.14178512.

Fauth, J. *et al.* (2025) 'Investigating building permit processes across Europe: characteristics and patterns', *Building Research and Information*, 53(4), pp. 417–434. https://doi.org/10.1080/09613218.2024.2400467.

Fauth, J.; Seiß, S. (2023) 'Ontology for building permit authorities (OBPA) for advanced building permit processes', *Advanced Engineering Informatics*, 58(October). https://doi.org/10.1016/j.aei.2023.102216.

Gómez-Gil, M. *et al.* (2022) 'The Digital Building Logbook as a gateway linked to existing national data sources: The cases of Spain and Italy', *Journal of Building Engineering*, 63(August 2022). https://doi.org/10.1016/j.jobe.2022.105461.

Gómez-Gil, M., Espinosa-Fernández, A. and López-Mesa, B. (2022) 'Review and Analysis of Models for a European Digital Building Logbook', *Energies*, 15, p. 24. https://doi.org/https://doi.org/10.3390/en15061994.

Grow, D. (2023) *Technical guidelines for Digital Building Logbooks - Guidelines to the Member States on setting up and operationalising digital building logbooks under a common EU framework*. Brussels.

Mêda, P. *et al.* (2022) 'A Process-Based Framework for Digital Building Logbooks', in EC3 (ed.) *Proceedings of the 2022 European Conference on Computing in Construction*. Ixia: EC3, https://doi.org/10.35490/EC3.2022.183.

Mêda, P., Fauth, J., Schranz, C., Urban, H., *et al.* (2024) 'Digital Building Permits and Digital Building Logbooks - Clustering the Challenges and Requirements for alignment', in *Proceedings of the 2024 European Conference on Computing in Construction*. Chania: EC3, https://doi.org/10.35490/EC3.2024.245.

Mêda, P.; Fauth, J.; Schranz, C.; Sousa, H., *et al.* (2024) 'Twinning the path of digital building permits and digital building logbooks – Diagnosis and challenges', *Developments in the Built Environment*, 20(December), p. 16. https://doi.org/10.1016/j.dibe.2024.100573.

Moore, N. (2018) 'Desk research', in C.U. Press (ed.) *How to Do Research: The Practical Guide to Designing and Managing Research Projects*. Cambridge: Facet publishing. https://doi.org/10.29085/9781856049825.

Naeem, M. *et al.* (2023) 'A Step-by-Step Process of Thematic Analysis to Develop a Conceptual Model in Qualitative Research', *International Journal of Qualitative Methods*, 22(October), pp. 1–18. https://doi.org/10.1177/16094069231205789.

Noardo, F. *et al.* (2022) 'Unveiling the actual progress of Digital Building Permit: Getting awareness through a critical state of the art review', *Building and Environment*, 213(January), p. 108854. https://doi.org/10.1016/j.buildenv.2022.108854.

Papadaki, I. *et al.* (2023) *Transition pathway for Construction*, *Directorate H – Ecosystems III: Construction, machinery & standardisation*. Brussels. https://ec.europa.eu/docsroom/documents/53854.

Staff, C. (2021) *Scenarios for a transition pathway for a resilient, greener and more digital construction ecosystem (Ares(2021)7679109) final)*. Brussels.

Ullah, K., Witt, E. and Lill, I. (2022) 'The BIM-Based Building Permit Process: Factors Affecting Adoption', *Buildings*, 12(1), p. 45. https://doi.org/10.3390/BUILDINGS12010045.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

PROJECT MANAGEMENT WITH IFC: TOWARD AN OPEN-BIM FRAMEWORK

Riccardo Padoan[1], Lidia Pinti[1], Jan Karlshøj[2]
[1]Department of Civil, Mechanical and Environmental Engineering, University of Trento, Italy
[2]Department of Civil and Mechanical Engineering, Technical University of Denmark, Denmark

Abstract

Successful construction management relies on interoperability, cost estimation, and scheduling. Current practices suffer from fragmented processes and data loss, due to proprietary silos and lack of automation. This research proposes a computational framework to optimize 4D and 5D BIM workflows by enriching the IFC schema. Through *IfcOpenShell*, the methodology provides semantic validation via IDS and collaboration via BCF, facilitating extraction of properties and quantities, scheduling, cost estimation, and resource allocation. Results indicate that an integrated OpenBIM framework using IFC as the primary source improves reliability. The contribution is a reproducible, open-source framework providing a methodology for multi-dimensional BIM management.

Introduction

Building Information Modelling (BIM) transformed the Architecture Engineering Construction (AEC) industry from a static 2D representation to a dynamic parameterized data environment. Despite the methodological maturity suggested by international standards like ISO19650:2018 series, a significant ”BIM gap” persists between theoretical frameworks and professional practice (Abanda et al., 2025). The industry relies on closed-proprietary software that creates rigid sequential dependencies that restrict interoperability (Lou et al., 2021) and facilitate the propagation of human errors (Sacks et al., 2018). Integrating BIM with process simulations can address these uncertainties, transforming the model into a decision-support environment capable of mitigating delays and overcoming the limitations of closed systems (Wang et al., 2014). Managing complex 4D and 5D datasets within these environments often requires intensive visual programming (Lou et al., 2021), such as *Dynamo*, to handle the overwhelming volume of repetitive tasks inherent in Quantitative Take Off (QTO) and Bill of Quantities (BoQ). Furthermore, empirical evidence suggests that BIM-based QTO is still rarely applied in practice due to the complexity of the software and incomplete data extraction methods (Alathamneh et al., 2024). The effectiveness of these workflows is inherently linked to the level of information needed which defines the usefulness of the model throughout the life-cycle (Pinti and Bonelli, 2022). In this context, the granularity of information is crucial to calibrate data density and prevent data gaps and over-modeling (Bolpagni and BIMe Initiative, 2016). Interoperability is facilitated by Industry Foundation Class (IFC), which provides machine interpretability of information and allows automation of workflows (Avogaro and Zanchetta, 2025). For project management purposes, entities such as *IfcWorkPlan*, *IfcWorkSchedule* and *IfcTask* regulate temporal aspects, while *IfcCostSchedule*, *IfcCostItem* and *IfcCostValue* provide the foundation for economic estimation. Recent advances, such as Information Delivery Specification (IDS) and BIM Collaboration Format (BCF), have further enhanced the ability to define computer-interpretable requirements and streamline model-based coordination (Fischer et al., 2025). To address these systemic inefficiencies, this research investigates the transition toward OpenBIM computational framework (McGinley et al., 2023). Although current IFC-based exchanges support object-level data, the integration of automated 4D and 5D logic remains insufficiently explored outside of proprietary silos (Avogaro and Zanchetta, 2025). This study proposes a methodology to reliably extract, manipulate, and validate project data by leveraging open-source logic. By bypassing the limitations of proprietary ecosystems, the research proposes a methodology toward the ”Total BIM” (Disney et al., 2022) vision, where the digital model serves as the primary source of truth for contractual and management purposes (Pinti and Bonelli, 2022). The paper is organized as follows: Section 2 provides a Literature Review of the IFC schema and construction management. Section 3 and 4 detail the Materials adopted and Key Concepts of the framework, while Section 5 describes the implemented Methodology. Finally, Section 6 and 7 present the Discussion of results and the Conclusion with future research directions.

Literature Review

IFC and Construction Management

The IFC, currently an ISO16739-1:2018 with *IFC4 ADD2 TC1*, provides an extensible, vendor-neutral schema designed to facilitate semantic interoperability within the AEC industry. Unlike proprietary formats, the IFC structure is organized into a four-level hierarchical architecture (Resource, Core, Interoperability, and Domain), allowing for the digital representation of complex construction domains, including actors, processes, and resources.

In the context of construction management, the schema lies in its ability to establish an explicit *IfcRelationship* between the physical component of the building, *IfcProduct*, and the management-related entities. This study exploits the ontological robustness of the IFC schema to synchronize three critical domains: temporal, economic, and resources-based.

IFC, IDS and BCF

The IDS represents a standard that defines Information Requirements (IR) as both human-readable and computer-interpretable (Fischer et al., 2025). The main purpose is to overcome the traditional limitations of exchanging requirements through non-interoperable formats, such as PDF documents, which often hinder automation and data accessibility. By specifying entity types, properties, and classifications, IDS enables automated compliance checks, ensuring that the IFC model contains the semantic richness necessary for 4D and 5D analysis (Fischer et al., 2025). For example, a gap addressed by design firms and public administration is the practical implementation of IR through IDS, particularly in linking Work Breakdown Structure (WBS) to IFC entity (deMarco et al., 2024). While the IFC schema defines the data model, the BCF standard serves as a model-based communication protocol (Lou et al., 2021). By linking IFC entities with BCF topics, stakeholders can communicate and collaborate without sharing the entire model, establishing traceability through *GlobalId*.

IFC and QTO

Despite the potential of BIM-based workflows, a recent systematic review suggests that QTO is still predominantly performed using 2D methods due to software complexity, modeling effort and data inconsistency (Alathamneh et al., 2024). The effectiveness of QTO depends on consistent data and standardization of information exchange (Pinti and Bonelli, 2022). IFC provides a concrete framework for defining measurable elements through *IfcQuantitySet* (Karlshøj and Proctor, 2025). The specification distinguishes between Base Quantities, which provide standardized geometric measures for early-stage estimation, and Custom Quantities, which allow the integration of local rules of measurement or specific professional requirements.

IFC and Project Scheduling

IFC provides a solution to data fragmentation by providing a standardized schema for time information (Xue et al. (2015); Avogaro and Zanchetta (2025)). *IfcWorkPlan* represents the central 4D entity, structuring all the information needed to complete complex activities. The scheduling hierarchy is managed through *IfcRelAggregates* which decomposes *IfcWorkPlan* into specific *IfcWorkSchedules*, and *IfcRelAssignsToControl* which links temporal elements directly to physical *IfcProduct* components. Individual activities are instantiated as *IfcTask* entities, organized through *IfcRelNests* and characterized by temporal data via *IfcTaskTime*. Logical dependencies are established through *IfcRelSequence*, while *IfcRelAssignsToProcess* maps *IfcProduct* instances to their corresponding *IfcTask* anchoring the physical geometry of the digital twin to the temporal workflow. *IfcWorkCalendar* defines standardized working (*WorkingTime*) and non-working (*ExceptionDate*) periods, providing precise control over working-time constraints and supporting accurate sequencing, thus strengthening the connection between temporal and physical dimensions in the 4D BIM environment (Noardo et al., 2022).

IFC and Cost Estimation

IFC schema supports financial information, through entities such as *IfcCostSchedule* for structuring, *IfcCostItem* for element-based assignment, and *IfcCostValue* for price definitions (Pishdad and Onungwa, 2024). Current research utilizes these capabilities to overcome the limitations of parameter-based 5D workflows. Cassandro et al. (2024a) proposed a methodology implemented through *IfcOpenShell* library that used *IfcCostItem* to establish a direct relationship between costs and IFC geometric entities, proving to be more effective than the 5D BIM approach based on simple assignment of parameters and attributes. The approach was validated in a second study on a structural model, demonstrating its effectiveness in optimizing cost estimates for public tenders and minimizing human error by anchoring *IfcCostItem* instances to official price lists (Cassandro et al., 2024b). *IfcCostSchedule* organizes *IfcCostItem* entities into a coherent hierarchy through *IfcRelAggregates*, ensuring the validity of financial data throughout the building's life-cycle. The entity supports hierarchical nesting through *IsNestedBy* and *Nests* relationships, ensuring that total costs remain the verifiable sum of subordinate items Cassandro et al. (2024a). *IfcCostItem* is linked to one or more *IfcCostValue* instances to define price breakdowns across materials, equipment, and labor, providing a native foundation for BoQ generation that current practice consistently replaces with external spreadsheet-based workflows.

IFC and Resource Allocation

Resource allocation is critical for successful project management, as it directly influences site safety, spatial coordination, and the feasibility of both schedules and budgets. Within the IFC schema, *IfcConstructionResource* provides the formal structure defining activity resources (Avogaro and Zanchetta, 2025). The schema supports complex resource containers through *IfcCrewResource* and *IfcSubContractResource*, which aggregate elementary resources - *IfcLaborResource*, *IfcConstructionEquipmentResource* and *IfcConstructionMaterialResource* - into functional units through *IfcRelNests*. By linking resources to specific tasks and costs, the schema provides the documentation necessary to justify expenses, verify work progress, and maintain a historical record of resource performance, which is important for transparent project reporting and future cost benchmarking (Pinti and Bonelli, 2022).

Research gap

Despite the ability of the IFC schema to support construction data management, current studies often treat model validation, project scheduling, and cost estimation as isolated processes (Avogaro and Zanchetta, 2025). Despite the expressive power of its scheduling, cost, and resource entities, these native capabilities remain largely underutilised in practice, consistently replaced by proprietary tools that fragment data across disconnected pipelines and never write information back into the schema. Recent contributions have successfully standardized IDS and cost estimation processes, demonstrating that machine-readable data is essential to reduce human errors and legal disputes in public tenders (Cassandro et al. 2024a; 2024c). However, a significant gap remains regarding the dynamic synchronization of time scheduling, cost estimation and resource allocation within a unified OpenBIM workflow. This fragmentation prevents the AEC industry from reaching the "Total BIM" vision (Disney et al., 2022), where the model serves as the only source of contractual and operational information (Pinti et al., 2022). The lack of a transparent methodology capable of orchestrating these different IFC domains remains the main barrier to a fully automated OpenBIM environment. This study addresses this deficit by proposing an integrated framework that leverages open-source computational logic to synchronize IR with multidimensional project control.

Materials

The proposed framework is developed within Free and Open Source Software (FOSS) ecosystem, with the aim to provide a transparent and accessible alternative to proprietary BIM tools. This choice is aligned with the current academic necessity of "Supporting the OpenBIM transition" (McGinley et al., 2023), ensuring open data and long-term interoperability. Adopting a "decoupled" approach, the framework performs data multidimensional (4D and 5D) management and enrichment, independently of the original authoring software avoiding the "black-box" logic of commercial tools. This ensures full compliance with OpenBIM standards and the ISO16739-1:2018 specification, through controllable and verifiable data manipulation. The core logic is implemented in *Python* (v3.10+) taking advantage of its extensive ecosystem in data science and AEC automation. To ensure maintainability reproducibility, the environment is managed via a *Virtual Environment* (.venv) architecture and a *requirements.txt* protocol, while version control is maintained via *Git*. The framework uses *IfcOpenShell* (v0.8.4) documentation as the core engine providing bidirectional access to IFC data structures and allowing users to navigate model hierarchies and resolve inverse relationships between physical components and their semantic properties (Noardo et al., 2022). The technological stack and the standardized protocols integrated into the framework are summarized in Table 1.

Table 1: Technological stack and data exchange protocols

Category	Component	Function
Rendering	*ThatOpen*, *Three.js*	Browser-based fragment rendering via *WASM* and GPU acceleration.
Validation	*IfcTester*, IDS	Automated compliance check of requirements via *XML* IDS files.
Interoperability	BCF, *JSON*	Issue tracking via *IfcOpenShell* and state persistence.
Analysis	*Pandas*, *NumPy*	IFC transformation into *DataFrames*.
Visualization	*Plotly*	Interactive 4D/5D charting.

Key Concepts

The proposed framework establishes a digital workflow designed to automate the validation and data enrichment of IFC models (Noardo et al., 2022). The process begins with model viewing and the validation, then the framework executes a set of algorithms to integrate 4D and 5D information into the system. The methodology is built upon seven fundamental concepts:

i. IDS and BCF: The framework treats (IR) as machine-readable document, verifying that the model contains the correct classifications and geometric properties before any calculation, through IDS (Fischer et al., 2025). If discrepancies are found, the BCF module enables a traceable feedback loop (Lou et al., 2021).

ii. QTO and Cost Estimation: To ensure accurate budget control, the framework performs an automated Quantity Take-Off. The total project cost is calculated by summing direct costs with indirect components (Liu et al., 2014):

$$C_{tot} = \sum_{i} \sum_{j=1}^{n} (Q_j \times R_j) + C_{indirect} \qquad (1)$$

where Q_j is the quantity extracted from the validated IFC geometry and R_j is the unit price mapped from standardized price lists. Each unit price comprises three cost components: $R_j = R_{mat} + R_{equ} + R_{lab}$ (materials, equipment, labour).

iii. Time Estimation: The framework calculates task durations using two methodologies:

- Synthetic Method: Derived from labour cost incidence (I_{man}) provided by regional price lists ($C_{man} = C_{tot} \times I_{man}$).
- Analytical Method: Based on specific production rates and unit labour costs (UP_{man}) directly associated with quantities ($C_{man} = Q \times UP_{man}$).

The total man-hours (H) are converted into working days (T_g) considering the size of the operative crew (N_{so}): $T_g = H/(8 \times N_{so})$.

iv. Program Evaluation and Review Technique (PERT) Scheduling: To account for construction uncertainty, durations are treated as random variables. The framework calculates the expected duration (T_A) using a beta probability distribution:

$$T_A = \frac{a + 4m + b}{6} \tag{2}$$

where m is the most likely duration (T_g), a is the optimistic estimate ($0.75m$), and b is the pessimistic estimate ($1.75m$). This approach ensures a robust Critical Path Method (CPM) evaluation and the identification of critical path activities.

v. Resource Allocation: Each task is linked to a multidimensional array of resources, including costs, materials, equipment and specialized labour. This exhaustive mapping is critical for site safety management, as it enables the identification of spatial-temporal overlaps between hazardous activities and ensures that safety protocols and required certifications are explicitly associated with each operational unit.

vi. Task Sequencing: Construction activities are sequenced through standard logical dependencies (F-S, S-S, F-F, S-F) and visualized through Gantt chart and *DataFrames*.

vii. Cash Flow: The framework aggregates the results of the 4D and 5D synchronization to perform automated financial forecasting. By setting incomes and expenses, the system generates cumulative cost distributions, allowing project managers to monitor the financial health of the project, identifying potential liquidity gaps and comparing budgets against real-time projections.

Methodologies

The workflow in Figure 1 guides the reader towards the logic of the framework, which illustrates the integration of IFC data validation and collaboration, followed by automated quantity extraction necessary for a dynamic project management phase.

IFC Export and Import

The framework approach begins with input standardization through two modules: export configuration and model import. To ensure data quality, the system generates *JSON* files that translate the *IFC4X3_ADD2* specifications into parameters ready for authoring software. Subsequently, the IFC import module loads the file and initializes the working environment. Using *IfcOpenShell*, the framework performs a structural audit that detects the schema version and verifies the integrity of the file. This step is crucial to confirm that the model is ready for semantic enrichment and 4D and 5D analysis.

3D Visualization Architecture

The 3D visualization module handles high-density IFC data by moving processing to a decoupled client-side pipeline. Data are passed as *NumPy uint8* memory blocks from the *Python* backend to a custom *React* component, bypassing the overhead of text-based encodings. The core engine, powered by *WebAssembly*, *Three.js*, and the *That Open Company* ecosystem, converts the model into *Fragments* to group identical geometries and to optimize *GPU* performance. This approach enables native browser-based rendering, eliminating the need for external BIM viewers. To ensure data persistence, camera matrices are serialized into *JSON*, while session stability is maintained through a memory management logic that purges geometry buffers from *VRAM* during module transitions.

IDS and BCF

The framework ensures data consistency through an automated workflow that bridges the gap between IR and collaborative management. This process relies on IDS to validate the model requirements and BCF to communicate discrepancies. The validation engine utilizes *IfcTester* to ensure that IR are machine-readable and to verify the compliance of the model. By parsing *XML*-based IDS files, the tool validates whether the model contains all the required information without manual intervention. Users can dynamically define rules based on IFC classes and *PropertySet*, which the system validates using a robust lookup strategy via *IfcOpenShell*. Using *ifcopenshell.util.element.get_psets*, the engine extracts granular properties and element attributes, transforming complex model hierarchies into a structured *Pandas DataFrame*. This approach identifies missing properties or invalid values, calculates real-time compliance metrics, and generates visual reports using *Plotly*. To close the coordination loop, the framework integrates a synchronization logic between the 3D viewer and the reporting module. When a non-compliant element is identified, the system captures the current state of the model viewer—including camera projection matrices—as standardized snapshots. These visual cues are associated with the specific validation failure. Leveraging the *IfcOpenShell* BCF module, the framework serializes these discrepancies into a *.bcfzip* archive. This package contains the necessary *markup.bcf* and *project.bcfp XML* structures, allowing the identified issues to be re-imported into any BIM authoring software. This ensures a feedback-based workflow where requirements are not only verified, but effectively communicated for resolution.

Properties and Quantities Extraction

The framework transforms IFC models into structured datasets through a schema-independent extraction pipeline based on the *ifcopenshell.util* module. The module dynamically extracts each *IfcPropertySet* instances attached to *IfcProduct* elements, resolving *IfcRelDefinesByProperties* relationships to capture property names, values, and data types with full *GlobalId* traceability. This approach

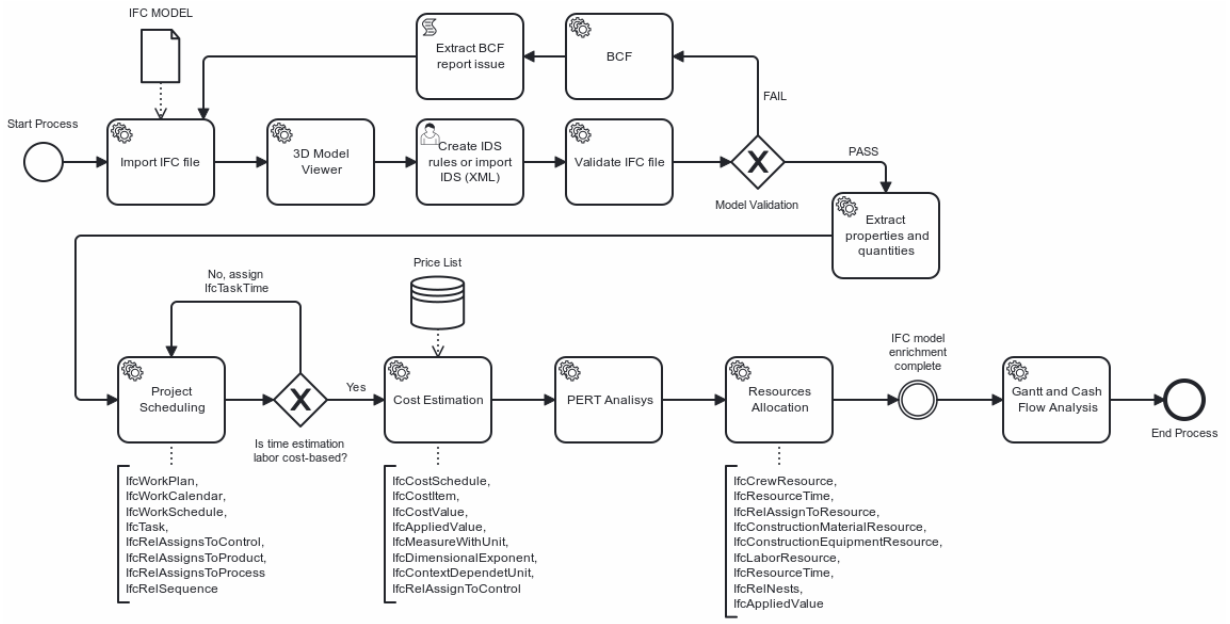


Figure 1: Framework logical process.

ensures compatibility with both standard property definitions (*Pset_**) and custom sets. Geometric quantities are retrieved by parsing *IfcElementQuantity* entities through the inverse *IsDefinedBy* attribute. The extraction logic processes physical quantity subtypes including *IfcQuantityArea*, *IfcQuantityVolume*, *IfcQuantityLength*, *IfcQuantityCount*, and *IfcQuantityWeight*, supporting both standard *BaseQuantities* and domain-specific (*Qto_**) sets. *Pandas* and *NumPy* transform IFC hierarchies into tabular *DataFrames*, enabling efficient grouping, filtering, and quantity calculations for cost estimation and scheduling. Each row represents a property or quantity instance with element identification, classification metadata, source discriminators, and value containers. Furthermore, the extracted quantity incorporates a diagnostic layer that flags potential duplicate entries across element classes. Records that share identical quantity values within the same IFC class are highlighted as potential duplicates. In the test model, for example, querying *NetArea* on *IfcSlab* elements on *F_00 Level* revealed a raw aggregate of $81,807.68\text{m}^2$ against a potential corrected value of $2,586.36\text{m}^2$, demonstrating the capacity to highlight suspicious duplicate quantities of the same element.

4D Scheduling

The temporal dimension is integrated into the model by implementing the IFC schema via the *ifcopenshell.api.sequence* module. The framework establishes a hierarchical structure in which an *IfcWorkPlan*, declared within the *IfcProject* via *IfcRelDeclares*, serves as the main container for multiple *IfcWorkSchedule* entities. Operational constraints are defined via *IfcWorkCalendar*. The framework utilizes *sequence.add_work_time* and *sequence.assign_recurrence_pattern* to specify working hours and recurring exceptions. Individual tasks are instantiated as *IfcTask* entities using *sequence.add_task*. The system supports recursive nesting via *IfcRelNests* via *nest.assign_object*, allowing complex work packages to be broken down into granular subtasks. Each task is associated with an *IfcTaskTime* object via *sequence.add_task*. Through *sequence.assign_process*, the framework identifies which elements are operated upon, while *sequence.assign_product* defines the physical output of each task, defining *IfcRelAssignsToProcess* and *IfcRelAssignsToProduct*. Task sequencing is managed through *IfcRelSequence* relationships created via *sequence.assign_sequence*. The framework supports standard dependency types, enabling dynamic CPM analysis directly within the OpenBIM environment.

5D Cost Estimation and BoQ

The cost estimation framework automates the generation of BoQ by synchronizing costs with the IFC geometric data using the *ifcopenshell.api.cost* and *ifcopenshell.api.control* modules. By importing an external price list (.csv) and instantiating an *IfcCostSchedule*, the framework creates an *IfcCostItem* for each price line using *cost.add_cost_item*, aggregating three distinct *IfcCostValue* entities representing materials, equipment, and labour costs via *cost.add_cost_value*. Each value is associated with an *IfcMeasureWithUnit* and to overcome naming discrepancies between model elements and price list descriptions, the system implements a fuzzy semantic mapping using *difflib.SequenceMatcher*. The pipeline first filters candidate price rows by IFC class using *IfcMatch* field, a classification column manually aligned to the price list structure, retaining only entries compatible with the element type. Within this, *difflib.SequenceMatcher* computes string similarity between the normalized element name and each filtered candidate, and the highest-scoring entry is selected. For example, element

"HEB Søjle_N:HE180B:371924" - following the convention *FamilyType:TypeName:GlobalId* - correctly matches *"Stålsøjle HE-B 180"* in the *Molio* price list despite the structural difference (similarity score = 0.47). Once validated, *control.assign_control* establishes the definitive link between the *IfcElement* and its *IfcCostItem*, which are logged and surfaced to the user, who can manually correct erroneous assignments by removing individual *IfcRelAssignsToControl* relationships and re-triggering the pipeline. Quantity extraction follows a hierarchical priority logic to ensure accuracy, where net measurements are preferred over gross ones. The framework automatically handles unit conversions by querying *IfcUnitAssignment*. The final economic output is structured as a relational *DataFrame* where costs are filtered and aggregated by *IfcBuildingStorey* using *IfcRelContainedInSpatialStructure* relationships. For each element, the unit price R_j is derived from the sum of the costs of materials, equipment, and labour retrieved from the corresponding *IfcCostItem*, ensuring hierarchical budget breakdowns by level and for the entire project.

Resource Allocation

The resource allocation module integrates construction resources with scheduling by instantiating *IfcConstructionResource* entities and linking them to scheduled tasks via the *ifcopenshell.api.resource* module, supporting *IfcCrewResource* and *IfcSubContractResource*. For each *IfcTask*, the framework determines its *ScheduleDuration* through a PERT analysis in which the total labour hours, extracted from the BoQ, are adjusted based on a defined *crew_size*. To ensure accuracy, the module queries *IfcWorkCalendar* to identify *WorkingTimes* and *ExceptionTimes*. The result is converted into an ISO 8601 duration string and assigned to an *IfcTaskTime* entity through a custom function. Using *resource.add_resource*, a parent resource container is instantiated with attributes, and an *IfcResourceTime* entity is created via *resource.add_resource_time* to store the total labour hours. The parent resource is then assigned to the task via *resource.assign_resource*, establishing an *IfcRelAssignsToResource* relationship. For crew-managed tasks, the parent resource is decomposed into child resources (*IfcLaborResource*, *IfcConstructionEquipmentResource*, and *IfcConstructionMaterialResource*) nested via *IfcRelNests*. Time and financial attributes are embedded in this hierarchy: *IfcLaborResource* instances receive an *IfcResourceTime* object where the *ScheduleWork* attribute reflects the adjusted man-hours, while cost data for all resources is stored via *IfcAppliedValue* entities within the *BaseCosts* attribute. This structure enables a recursive cascade where root tasks and *IfcWorkSchedule* are automatically updated by aggregating the *ScheduleStart* and *ScheduleFinish* dates of their nested sub-tasks, ensuring global time and economic consistency across the IFC model.

GANTT and Cash Flow analysis

The proposed framework processes the previous IFC entities to generate integrated Gantt charts and cash flow visualizations, by extracting task schedules and linking them to cost and resource data. The system reconstructs the project hierarchy by crossing *IfcRelNests* and *IfcRelAggregates* relationships, extracting data from each *IfcTask*. The costs are then mapped to the timeline by intersecting *IfcRelAssignsToProcess* (linking tasks to products) and *IfcRelAssignsToControl* (linking products to cost items), aggregating categorized materials, equipment, and labour expenses. Financial forecasts distribute costs proportionally across task durations rather than treating them as point-in-time events. For example, a multi-week construction activity spreads its total cost across the execution period based on the proportion of work hours in each time window. Processing these data streams through *Pandas*, construction costs are integrated with external incomes and expenses to generate cumulative cost distributions and net cash flow metrics, enabling accurate liquidity planning and early identification of funding gaps, providing a decision support for financial sustainability. This integrated approach establishes a bidirectional link between the IFC model and the project's temporal financial health, fulfilling the "Total BIM" (Disney et al., 2022) requirement of using the digital model as the primary contractual and management record.

Results and Discussion

The results shown in Table 2 demonstrate that the proposed framework bridges the gap between static IFC models and dynamic project management. By automating the "validation loop", only high-quality data populate the model, reducing the manual effort required for compliance checking. Unlike proprietary workflows, the proposed framework operates entirely within the native IFC schema; all scheduling, cost, and resource data are co-located within a single model and semantically linked through native IFC entities. This ensures that any design or on-site change propagates consistently, reducing the operational overhead of maintaining data consistency across multiple proprietary tools. Enriching 4D and 5D entities directly ensures that all project management data remain queryable, portable, and auditable, without requiring any software outside the open-source eco-system overcoming the fragmentation that OpenBIM tries to prevent (Sacks et al., 2018). The proposed framework provides immediate visibility into cost and schedule changes, and by preventing the "cascade effect" of unexpected expenses, the system enables reliable decision-making that protects the quality and overall performance of the project. However, significant challenges and limitations have been revealed in the implementation phase. In particular, the integration of 3D visualization components highlights the gap between BIM data and graphical engines. Furthermore, to overcome current gaps in the *IfcOpenShell* library with respect to specific modules, the framework implements custom algorithms to

Table 2: IFC 4D and 5D main enrichment contribution

Category	Enriched IFC Instance
Scheduling	#78807=IFCWORKPLAN('0NsExtC2f8D8WeUQZlTNzR', #78805, 'Construction WorkPlan', $, '', 'WP-01', '2026-01-14T15:40:00', (#15), $, $, $, '2026-01-14T08:00:00', '2027-01-14T17:00:00', .PLANNED.);
	#78811=IFCWORKCALENDAR('07lFbCthz3nwQka1zjY40E', #78805, 'Construction WorkCalendar', 'WC-01', $, $, (#78812), (#78815), .FIRSTSHIFT.);
	#78813=IFCRECURRENCEPATTERN(.WEEKLY., $, (1,2,3,4,5), $, $, $, $, (#78814));
	#78819=IFCWORKSCHEDULE('0menkKZ$51oAc_AcvL_qld', #78805, 'Construction WorkSchedule', $, $, 'WS-01', '2026-01-14T15:40:00', (#15), $, $, $, '2026-01-14T08:00:00', '2027-01-14T17:00:00', .PLANNED.);
	#78821=IFCRELNESTS('2zVMrPaEL37ex7xLms7nHC', #78820, $, $, #78811, (#78807, #78819));
	#78829=IFCTASK('1NN1H7z_LFCwmdHctg15XD', #78828, 'B_00 Slab', 'B_00 Slab Construction', $, '1.1', $, $, $, .F., 0, #79750, .NOTDEFINED.);
	#78835=IFCRELASSIGNSTOPROCESS('2JPu4ZsZn3OhG98crHGOnU', #78834, $, $, (#46908), $, #78829, $);
	#79005=IFCRELSEQUENCE('1BqjPS6fn74PiTOQRd4ylY', #79004, $, $, #78829, #78833, $, .FINISH_START., $);
Cost	#79007=IFCCOSTSCHEDULE('25qTTCIEzEcgVMOqXhCm3F', #78805, 'Price List Cost Schedule', $, $, $, .SCHEDULEOFRATES., $, $, '2026-01-14T15:48:31.239086');
	#79009=IFCCOSTITEM('10bAGvkHHEw8R1PlUeSNlN', #79008, 'Betonbj\X2\00E6\X0\lke 200x300mm', $, $, '04.10.82,01', $, (#79012, #79016, #79020), $);
	#79011=IFCRELASSIGNSTOCONTROL('0jRM67LPX9Ue3QN_$YIzt6', #79010, $, $, (#79009, …, #79707), $, #79007);
	#79012=IFCCOSTVALUE('Materials', $, IFCMONETARYMEASURE(4056.05), #79015, $, $, 'MATERIALS', $, .MULTIPLY., $);
	#79016=IFCCOSTVALUE('Equipment', $, IFCMONETARYMEASURE(2520.36), #79019, $, $, 'EQUIPMENT', $, .MULTIPLY., $);
	#79020=IFCCOSTVALUE('Labor', 'Labor Hourly Rate: 484.57 c/h', IFCMONETARYMEASURE(15113.7383), #79023, $, $, 'LABOR', $, .MULTIPLY., $);
	#79723=IFCRELASSIGNSTOCONTROL('0j$nKjAVz19OQC8faTeGYg', #79722, $, $, (#17791, …, #63008), $, #79009);
Resources	#79753=IFCCREWRESOURCE('1a_wSbOMz7EhoFL4luGDWg', #79752, 'Crew Resource for Task B_00 Slab', 'Management Type: Own crew...', $, 'RES-B_00 Slab-CRW', $, #79754, $, $, .NOTDEFINED.);
	#79754=IFCRESOURCETIME($, $, $, 'PT15H6M37S', $, $, $, $, $, $, $, $, $, $, $, $, $);
	#79756=IFCRELASSIGNSTORESOURCE('2iXBk4EBP3HAHE$_XGUVrV', #79755, $, $, (#78829), $, #79753);
	#79758=IFCCONSTRUCTIONMATERIALRESOURCE('26lXNDGc55C9z8aufxDE$L', #79757, 'Material - 78829-IfcSlab', 'Type: CONCRETE', $, 'MAT-78829-IfcSlab', $, $, (#79761), $, .CONCRETE.);
	#79760=IFCRELNESTS('3PGGyYVVXB4uYwydQvpI6Q', #79759, $, $, #79753, (#79758, #79763, #79767));
	#79763=IFCLABORRESOURCE('01D3bhwnj7POQu2zvO5Sdl', #79762, 'Labor - 78829-IfcSlab', 'Crew Size: 2 workers', $, 'LAB-78829-IfcSlab', $, #79765, (#79764), $, .CONCRETE.);
	#79767=IFCCONSTRUCTIONEQUIPMENTRESOURCE('2feoeXk_vAPBHTq8qrIptH', #79766, 'Equipment - 78829-IfcSlab', 'Type: PUMPING', $, 'EQP-78829-IfcSlab', $, $, (#79768), $, .PUMPING.);

manage specific hierarchical scheduling complexities. In addition, the intricate relationships and nesting of resource entities made the attribute manipulation complex. While *difflib.SequenceMatcher* provides a scalable baseline for semantic mapping, the test model revealed a structural limitation: across 1 006 elements assigned scores ranged from 0.12 to 0.51 (avg. 0.32), confirming that the primary safeguard is the IFC class pre-filter, while within-class ambiguity requires manual verification. This suggests that integrating standardized classification systems, such as WBS codes or Uniclass classification, would mitigate string similarity inaccuracies and improve reliability. Finally, while *GlobalId* instability across re-exports remains a known OpenBIM challenge, the framework ensures data persistence by establishing associations within a single pipeline, provided that the downstream workflow preserves native IFC entities.

Conclusions

This research proposed an integrated OpenBIM framework designed to transform digital models into active decision-support tools for integrated construction management. Responding to the operational workflows of design firms, public administrations, and construction site managers, the framework bridges regulatory requirements and construction practice. Its main contribution lies in its ability to mitigate repetitive tasks associated with manual data entry, often the main cause of human error and project delays. By providing an accessible alternative to expensive proprietary software, the framework makes high-level BIM management feasible even for smaller organizations and public administrations. Despite these advances, the framework has limitations: it assumes a single authoritative IFC file as its primary data source and complex architectural and Mechanical, Electrical, Plumbing (MEP) models could face challenges with highly complex geometries or intricate systems. The granularity of the data and non-linear relationship between components in these systems challenge data extraction and quantification, requiring more specialized analysis rules. Based on these findings, work is underway to evolve this framework into a user-friendly standalone tool. While qualitative validation is currently being conducted to assess its impact on professional workflows and its effectiveness in diverse project environments, quantitative validation will be conducted to measure the framework's performance in terms of time-saving and error reduction compared to traditional manual methods. Future implementations will address the above limitations and may extend to versioning control, IoT-based site monitoring, and Life Cycle Assessment (LCA) integration. The use of OpenBIM standards ensures high scalability and reproducibility, while the use of *Python*-based technologies ensures the framework's implementation in cloud environments, facilitating collaborative management. Furthermore, to enhance user accessibility, the implementation of a Natural Language Processing (NLP) interface is envisioned. This would allow non-expert users to query the IFC model in natural language (Nabavi et al.,

2023).

Data availability

Data are provided under a *Demonstration-Only Proprietary License* and will be made available on request.

References

Abanda, F. H., Balu, B., Adukpo, S. E., and Akintola, A. (2025). Decoding iso 19650 through process modelling for information management and stakeholder communication in bim. Buildings, 15(3).

Alathamneh, S., Collins, W., and Azhar, S. (2024). Bim-based quantity takeoff: Current state and future opportunities. Automation in Construction, 165, 105549.

Avogaro, D. and Zanchetta, C. (2025). Ifc and project control: A systematic literature review. Buildings, 16, 91.

Bolpagni, M. and BIMe Initiative (2016). The many faces of 'LOD'. First published 19 July 2016, updated 2025. Accessed: 21 January 2026.

Cassandro, J., Farina, A., Mirarchi, C., and Pavan, A. (2024a). Cost estimation practices: a comparative analysis of traditional 5d bim and ifc methods. Proceedings of the CIB W78 Conference 2024, October 1st-3rd 2024, Marrakesh, Morocco.

Cassandro, J., Mirarchi, C., and Pavan, A. (2024b). Ifc-based cost estimation: Application to a structural model. Proceedings of the 2024 European Conference on Computing in Construction, Chania, Crete, Greece July 15-17, 2024.

Cassandro, J., Mirarchi, C., Zanchetta, C., and Pavan, A. (2024c). Enhancing accuracy in cost estimation: structured cost data integration and model validation. Journal of Information Technology in Construction, 29, 1293–1325.

deMarco, G., Slongo, C., and Siegele, D. (2024). Enriching building information modeling models through information delivery specification. Buildings, 14, 2206.

Disney, O., Roupé, M., Johansson, M., and Domenico Leto, A. (2022). Embracing bim in its totality: a total bim case study. Smart and Sustainable Built Environment, 13(3), 512–531.

Fischer, S., Urban, H., Schranz, C., and Zucker, G. (2025). Bridging the gap between tabular information requirements and the information delivery specification (ids). Buildings, 15(7).

Karlshøj, J. and Proctor, J. (2025). Quantity take-off information delivery manual (IDM). buildingSMART International, (BR-2025-1024-CS).

Liu, H., Lu, M., and Al-Hussein, M. (2014). Bim-based integrated framework for detailed cost estimation and schedule planning of construction projects. 31st International Symposium on Automation and Robotics in Construction and Mining, ISARC 2014 - Proceedings.

Lou, J., Lu, W., and Xue, F. (2021). A Review of BIM Data Exchange Method in BIM Collaboration, pages 1329–1338. Proceedings of the 25th International Symposium on Advancement of Construction Management and Real Estate. CRIOCM 2020. Springer, Singapore.

McGinley, T., Negendahl, K., Smolira, P., Jakubowska, M., Vejlgaard, A., and Karlshøj, J. (2023). Supporting the openbim transition. Proceedings of the 2023 European Conference on Computing in Construction and the 40th International CIB W78 Conference. Heraklion, Crete, Greece July 10-12, 2023.

Nabavi, A., J. Ramaji, I., Sadeghi, N., and Anderson, A. (2023). Leveraging natural language processing for automated information inquiry from building information models. Journal of Information Technology in Construction, 28, 266–285.

Noardo, F., Wu, T., Arroyo Ohori, K., Krijnen, T., and Stoter, J. (2022). Ifc models for semi-automating common planning checks for building permits. Automation in Construction, 134, 104097.

Pinti, L. and Bonelli, S. (2022). A methodological framework to optimize data management costs and the handover phase in cultural heritage projects. Buildings, 12(9).

Pinti, L., Codinhoto, R., and Bonelli, S. (2022). A review of building information modelling (bim) for facility management (fm): Implementation in public organisations. Applied Sciences, 12(3).

Pishdad, P. and Onungwa, I. O. (2024). Analysis of 5d bim for cost estimation, cost control and payments. Journal of Information Technology in Construction (ITcon), 29, 525–548.

Sacks, R., Eastman, C., Lee, G., and Teicholz, P. (2018). Bim handbook: A guide to building information modeling for owners, designers, engineers, contractors, and facility managers. 3rd ed., john wiley sons, doi:10.1002/9781119287568.

Wang, W.-C., Weng, S.-W., Wang, S.-H., and Chen, C.-Y. (2014). Integrating building information models with construction process simulations for project scheduling support. Automation in Construction, 37, 68–80.

Xue, W., Wang, Y., and Man, Q. (2015). Research on information models for the construction schedule management based on the ifc standard. Journal of Industrial Engineering and Management, 8(3), 615–635.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

AN EXTENSIBLE GRAPHQL API FOR FINE-GRAINED ACCESS TO BUILDING INFORMATION MODELS

Nepomuk Wolf[1,2], Sebastian Esser[1,2], and André Borrmann[1,2]
[1] Chair of Computing in Civil and Building Engineering, Technical University of Munich, Germany
[2] TUM Georg Nemetschek Institute, Technical University of Munich, Munich, Germany

Abstract

Building information exchange in large projects remains largely file-based, although many applications require only small, task-specific subsets of data. This paper addresses the need for element-level, transactional access to building information in web-based BIM workflows, including visualization and model analysis. While direct IFC-to-web mappings are technically feasible, they often result in complex schemas and verbose queries. We present a simplified GraphQL API that enables fine-grained access to building information while abstracting IFC complexity. Through atomic geometry resources, flattened property access, and a modular schema structure, the approach improves usability, query readability, and application-oriented BIM data access.

Introduction

Efficient information retrieval is a fundamental requirement of modern Building Information Modeling (BIM) workflows. While the Industry Foundation Classes (IFC) standard provides a comprehensive framework for representing geometry, topology, and rich semantic information, current practice remains largely centered around file-based exchange. IFC models are typically transferred as monolithic STEP files, forcing client applications to load entire datasets, even when only a small subset of information is required. This paradigm creates information silos that hinder low-latency data access and complicate the integration of emerging workflows, such as web-based visualization, digital twins, and AI-assisted analysis.

As the Architecture, Engineering, Construction and Operation (AECO) industry moves toward higher levels of BIM maturity, the demand for element-level, transactional access to model data is increasing. Many practical use cases require only partial model views, such as web-based visualization, quantity takeoff, or coordination workflows. In these scenarios, transferring and processing complete IFC files introduces unnecessary overhead and latency.

Previous work by Wolf et al. (2025) demonstrated that a direct 1:1 mapping between the IFC schema and GraphQL is technically feasible. However, the resulting API closely mirrored the complexity of the IFC schema, so even simple queries required traversing deeply nested relationships such as property assignment or geometric representations. Moreover, the heterogeneity of IFC geometry forces clients to possess detailed prior knowledge of internal modeling constructs.

This observation parallels a broader development in the Linked Data ecosystem of the building industry, where direct schema mappings have increasingly been replaced by more abstract and modular representations of building information.

Inspired by this paradigm shift, the research presented in this paper moves away from strict schema mirroring and toward the design of a tailored GraphQL API for BIM data access. Instead of exposing the full internal structure of IFC, the proposed interface provides a lightweight, extensible skeleton that prioritizes simplicity and practical usability. Complex hierarchical constructs are flattened, geometry and properties are treated as atomic leaf nodes, and building elements are exposed as directly queryable resources. At the same time, the API remains agnostic to the underlying data source and can operate on traditional IFC STEP files as well as with alternative storage backends.

In contrast to Large Language Model (LLM)-based interfaces, which may support natural-language interaction but do not themselves guarantee deterministic or structurally validated access to BIM data, GraphQL provides a strongly typed and self-documenting interface with automatic query validation.

The research question of this paper is: How can a tailored GraphQL API support fine-grained, interoperable access to building information while reducing query complexity compared to direct schema mappings? The scope of this work is read-oriented access for AECO workflows, with a particular focus on spatial structure, properties, and geometry.

This paper presents the architecture and design decisions of this simplified GraphQL interface, whose public software repository is cited in (Wolf, 2026), and shows how modular schema extensions enable the integration of external data sources, such as data sheets, databases, or sensor platforms, without requiring disruptive API versioning.

Technical Background and Related Works

GraphQL is a web-based API paradigm that combines a schema definition language, a query language, and a runtime for resolving client-defined queries. Originally introduced to address limitations of REST-based data fetching, such as over- and under-fetching of resources, it has since become a widely adopted approach for structured

data access. GraphQL APIs are strongly typed and self-documenting, allowing client applications to precisely define the structure and scope of requested responses. These characteristics make GraphQL particularly relevant for BIM-related scenarios in which clients frequently require only a small, task-specific subset of a much larger building model.

In recent years, GraphQL has also attracted growing interest within the AECO research community, where its potential for structured and fine-grained access to BIM-related data has been explored (Alexiev et al., 2023; Werbrouck et al., 2019b,a). Beyond the BIM domain, studies have evaluated GraphQL as an alternative to REST-based APIs. While REST can achieve lower raw response times in static scenarios, GraphQL has been shown to use resources more efficiently when data requirements change frequently (Lawi et al., 2021). Its declarative query model also mitigates overfetching, which is particularly relevant for BIM applications with large models (Quiña-Mera et al., 2023).

From Monolithic IFC Mappings to Modular Semantic Representations

The move toward tailored, API-driven access to IFC data is part of a broader evolution in building data standards, which increasingly favor modular, element-level information retrieval over monolithic, file-based exchange (Zhu et al., 2025). Early attempts to integrate IFC with web technologies were led by the Linked Data community. The ifcOWL ontology provides a 1:1 bidirectional mapping of the EXPRESS-based IFC schema into RDF (Beetz et al., 2009). While semantically complete, this approach resulted in very large graphs and highly verbose query patterns, mainly due to ordered collections and objectified relationships (Pauwels et al., 2015). As a response, the Linked Building Data community shifted toward smaller modular ontologies such as the Building Topology Ontology (BOT) (Rasmussen et al., 2021) and the Building Element Ontology (Pauwels, 2018). This transition demonstrated the benefits of lightweight structural skeletons that preserve essential semantics while reducing query and modeling complexity.

GraphQL and Semantic Web Approaches for Building Data Access

Rather than treating GraphQL and Semantic Web technologies as competing paradigms, recent research has explored their integration. Several approaches propose GraphQL as a client-facing interface for RDF and SPARQL backends (Gleim et al., 2020; Karalis et al., 2023). Taelman et al. (2019) compare architectural patterns for bridging GraphQL and RDF, while Angele et al. (2022) demonstrate that GraphQL can reduce query complexity for RDF-based datasets and enable closer alignment with ontology-driven data models.

A similar challenge arises when exposing IFC data through GraphQL. Wolf et al. (2025) demonstrated that a direct 1:1 mapping from EXPRESS to GraphQL is technically feasible. However, the resulting schema inherited much of the structural complexity of IFC, leading to deeply nested queries and frequent reliance on inline fragments. In addition, GraphQL lacks native support for several EXPRESS features, which further complicates schema design and client-side querying. This limitation mirrors earlier experiences from the Linked Data community: preserving the entire source schema maximizes formal completeness, but often at the cost of practical usability.

Toward Modular and Transactional BIM Data Exchange

In parallel to these developments, the IFC 5 roadmap explicitly promotes modularization, normalization, and object-level data exchange to support transactional workflows and partial model access (van Berlo et al., 2021). This direction aligns closely with the design motivation of the present work.

Geometry representation poses a comparable challenge in both GraphQL- and RDF-based approaches. In ifcOWL, geometry is mapped directly according to the IFC schema, which leads to extremely large datasets and high structural complexity (Pauwels et al., 2017). Wagner et al. (2020) provide an overview of alternative geometry handling strategies in Linked Data. Bonduel et al. (2019) demonstrate the possibility of abstracting geometry into atomic entities and embedding external formats such as OBJ or glTF. These findings align with the design decisions presented in this work, which treat geometry as an atomic resource optimized for downstream visualization and analysis tasks.

Proposed GraphQL API Design

The proposed API design follows four primary objectives: (1) to enable fine-grained, element-level access to building information without requiring full-file transfers, (2) to reduce query complexity by abstracting verbose IFC relationship structures, (3) to remain agnostic to the underlying data representation in order to support heterogeneous backends, and (4) to provide a modular foundation that can be extended with domain-specific information without introducing breaking changes. These objectives guide the simplification of spatial structures, the flattening of property access, the abstraction of geometry representation, and the modular extension mechanism described in the following sections. Rather than adopting a strict 1:1 mapping of the IFC schema, the proposed design follows a handcrafted, lightweight core schema that captures only the essential structural concepts. This should be understood as a task-oriented abstraction rather than a complete reproduction of the original data model. Although the initial development was driven by IFC models and their STEP-based serialization, the resulting API is not tied to a specific source format. Instead, the modular architecture enables the resolution of data from heterogeneous backends, such as files in arbitrary formats (for example RDF, JSON,

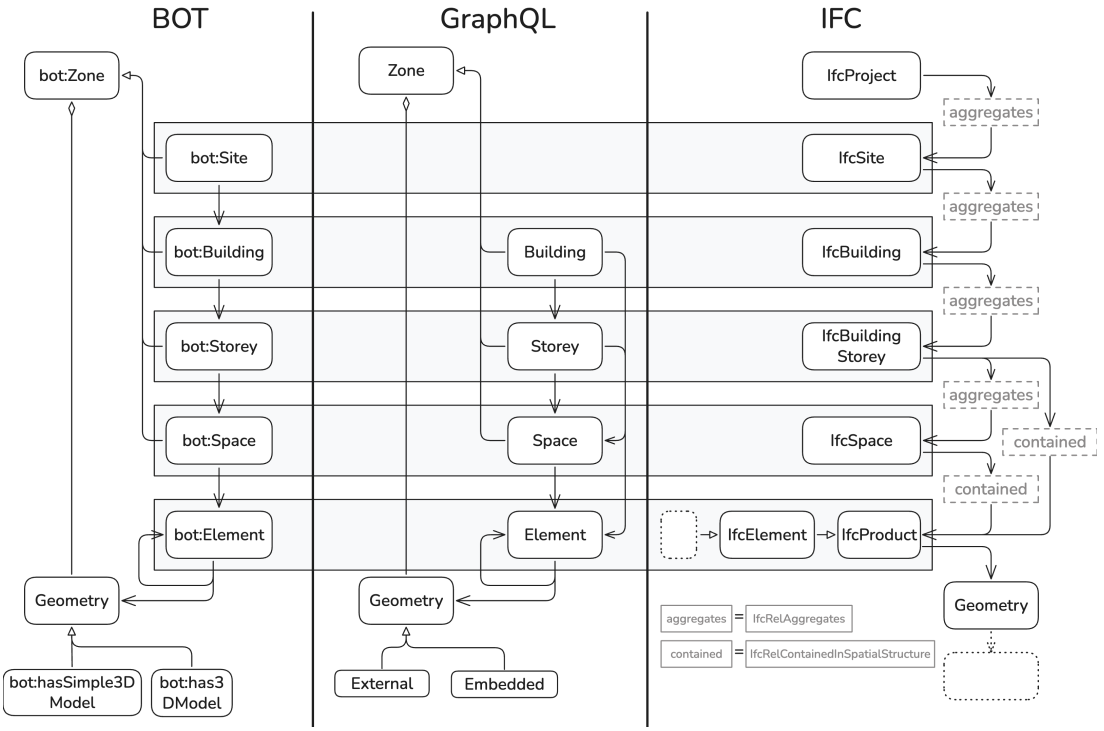


Figure 1: Similarities and differences between the spatial structure in BOT, IFC, and the GraphQL API

XML) or databases.

Spatial Structure and Topology

The core of the API design is a simplified spatial hierarchy that is conceptually aligned with the Building Topology Ontology (BOT) while remaining compatible with common IFC modeling practices. This structure provides a simplified navigation model for client applications and serves as the primary entry point for accessing spatially organized building data. To reduce traversal complexity, the API adopts a hierarchical zone model based on BOT and introduces a common `Zone` interface implemented by the `Building`, `Storey`, and `Space` types. In contrast to IFC, where spatial containment is expressed through objectified relationships (e.g., `IfcRelAggregates`), these relationships are represented directly as fields of the respective types. This significantly reduces query depth and improves readability while preserving the underlying semantic structure. Figure 1 illustrates the resulting schema structure and highlights the similarities and differences between BOT, IFC, and the proposed GraphQL API. Despite substantial simplification, the high-level organization remains closely aligned with IFC's spatial concepts. The most notable deviation is the elimination of intermediary relationship entities, which are replaced by direct associations exposed at the API level.

To support spatial reasoning and topological queries, the API additionally exposes explicit relationship fields between zones, including `partOf`, `intersects`, and `adjacent`. These relations follow the conventions established by BOT and enable expressive spatial queries without requiring clients to traverse low-level relational constructs. Building elements such as walls, windows, and slabs are represented using a generic `BuildingElement` type instead of directly mirroring the deep inheritance hierarchy of IFC. The concrete IFC subtype is preserved as a field on the `BuildingElement` type. This design trades explicit schema-level inheritance for a flatter and more compact API structure while trying to maintaining access to semantic type information. Element-level containment is supported through the `contains` and `partOf` fields, enabling the representation of nested component relationships, such as windows hosted within walls. In addition, building elements can be attached to any of the `Zone` subclasses (Building, Storey, Space), following the flexible attachment model defined in BOT. This differs from typical IFC usage patterns, where most building elements are associated with `IfcBuildingStorey` and furnishing elements with `IfcSpace`. Listing 1 shows an excerpt of the corresponding GraphQL schema definition.

Not all semantically rich IFC structures are directly exposed in the compact core schema. More complex concepts, such as material assignments or links to external classification systems, can instead be resolved dynamically through dedicated resolver functions. In this way, the simplified API surface remains stable, while the server can still retrieve and integrate richer information from IFC relationships or from additional backend systems when required.

Flattened Property Access

A direct 1:1 mapping of the IFC property hierarchy results in verbose queries that require traversing multiple intermediate entities such as `IfcPropertySet` and

```
interface Zone {
    id: ID!
    name: String
    geometry: Geometry
    contains: [Zone]
    elements: [BuildingElement]
    partOf:[Zone]
    intersects: [Zone]
    adjacent: [Zone]
}

type Building implements Zone{...}

type Storey implements Zone{...}

type Space implements Zone{...}
```

Listing 1: GraphQL schema excerpt for spatial structure and zone hierarchy

`IfcPropertySingleValue`. While this level of detail is valuable for authoring and editing workflows, it provides limited benefit for most read-oriented client applications and significantly reduces query readability.

To address this issue, the proposed API collapses objectified property relationships and exposes properties directly as a field of the `BuildingElement` type.

In IFC, properties are additionally organized into property sets (psets), which provide semantic grouping and contextual meaning. Preserving this information while maintaining a flattened access pattern can be achieved in two ways. The first approach is to explicitly model property sets as separate entities attached to each building element, each containing a collection of properties. The second approach is to associate a `pset` attribute directly with each property, thereby retaining grouping information without introducing an additional layer of nesting.

Both approaches preserve the original semantics, however, they differ in terms of query ergonomics rather than expressive power. Since typical client queries focus on retrieving individual properties rather than entire property sets, the second approach is adopted. This design minimizes structural overhead while still allowing semantic grouping information to be accessed when required.

Listing 2 shows that the association between properties and property sets is preserved in two complementary ways. First, the `pset` field can be explicitly requested as part of the `Property` object. Second, the `properties` field of the `BuildingElement` type supports filtering by property set name, enabling selective retrieval of properties belonging to a specific `pset`.

Atomic Geometry Representation

Geometry is one of the most relevant information categories in building models. It also represents one of the most complex aspects of the IFC data structure. A direct 1:1 mapping of IFC geometry entities into GraphQL needs to represent highly verbose and heterogeneous geometry representations, including boundary representations (BREP), swept solids, meshes, constructive solid geometry (CSG), and parametric extrusions. As a result, client applications either need prior knowledge of the underlying modeling paradigm or need to request multiple alternative representations simultaneously, in order to formulate valid queries.

```
type BuildingElement{
    geometry: Geometry
    guid: ID!
    name: String
    contains: [BuildingElement]
    partOf: [BuildingElement]
    properties(pset: String): [Property]
}

type Property {
    name: String
    value: JSON
    pset: String
}
```

Listing 2: GraphQL schema excerpt for building elements

This limitation conflicts with the declarative nature of GraphQL, where clients explicitly specify the expected response structure. Since the geometric representation type of an element is not known in advance, a fine-grained schema-level exposure of IFC geometry significantly increases query complexity.

To address this issue, geometry is treated as an atomic attribute within the API design. In this context, atomic means that the client accesses the geometry as a single leaf node rather than traversing the internal construction logic of the IFC model. Instead of exposing the internal structure, the complete geometry of an element is abstracted behind a single `geometry` field and can therefore be returned as a single unit in the GraphQL response. This design removes the need for the client to know in advance whether the underlying IFC geometry was modeled as BREP, extrusion, CSG, or another representation type.

This abstraction is intentionally lossy. Part of the expressive power of the original IFC representation is no longer directly available once geometry is exported into a downstream exchange format or otherwise reduced to an atomic resource. In particular, procedural and parametric geometry definitions are typically not preserved when exporting to commonly used geometry exchange formats, which limits the ability to reconstruct design intent or perform fine-grained semantic reasoning on geometric constructs. The same tradeoff applies more generally throughout the proposed schema design: direct access to the full internal IFC graph is reduced in favor of improved query readability and simpler client-side interaction.

```
"guid": "0Czv28JAHuep4",
"geometry": {
  "external": {
    "uri": "https://.../0Czv28JAHuep4.obj",
    "extension": "obj"
  },
  "embedded": {
    "payload": "
    v 1.22342604 2.82471817 0.00000000
    v 1.22342604 2.82471817 3.00000000
    v 0.54726129 0.30123695 3.00000000
    ...
    ",
    "encoding": "none",
    "extension": "obj"
  }
}
```

Listing 3: Response snippet with both external and embedded geometry

Geometry Transfer Strategies

The API supports two complementary strategies for transferring geometric data, depending on performance requirements and application context.

External Geometry: In this approach, geometry is stored in an external resource such as a file server or object storage system. The corresponding resolver returns a URI referencing the geometry file. This strategy enables efficient handling of large datasets, supports streaming and caching, and avoids inflating GraphQL response payloads.

Embedded Geometry: Alternatively, geometry can be embedded directly into the GraphQL response. Since GraphQL responses are (typically) encoded as JSON, the geometric representation must be serializable into a compatible format. This can be achieved in two ways:

- Text-based geometry formats, such as OBJ or WKT can be embedded directly as strings.
- Binary geometry formats can be encoded as a string (for example using Base64-encoding) and transmitted together with metadata describing the format and encoding method.

While Base64 encoding increases payload size and requires decoding on the client side, it enables the transfer of arbitrary geometry formats in a self-contained manner. For this reason, both embedded and externally referenced geometry representations are supported by the prototype API.

To demonstrate the flexibility of this abstraction, the prototype implementation supports multiple geometry formats, including OBJ, GLTF, GLB, WKT, BREP, and STL. The schema itself remains agnostic to specific formats and can be extended without modifying the core API definition. Each client application has full control over the format of the returned geometry and can choose between one of the supported formats. Listing 3 shows a response object, containing geometry in the OBJ format both as external and embedded representation.

Modular Schema Design

A central design principle adopted from the Linked Building Data community is to keep the core schema compact while enabling domain-specific extensions. This separation allows the API to remain lightweight and stable, while still supporting specialized application requirements such as energy analysis, cost estimation, or environmental assessment. GraphQL natively supports this paradigm, as types and fields can be extended without introducing breaking changes or requiring explicit API versioning.

Unlike REST-based interfaces, where modifications to response payloads often require version management to avoid disrupting existing clients, GraphQL enables clients to explicitly define the requested response structure. Newly added fields therefore remain invisible to existing client applications unless they are explicitly queried. This property enables incremental schema evolution while preserving backward compatibility.

Beyond the static core schema, the proposed implementation introduces a modular extension mechanism that dynamically loads schema fragments together with their associated resolver functions. Each extension contributes additional fields and data sources without modifying the base API definition. Resolver functions establish the connection between these abstract schema fields and concrete backend services and are executed only when the corresponding fields are requested by the client.

To demonstrate this mechanism, an example extension was implemented that augments the `BuildingElement` type with a `dataSheetURL` attribute. This field establishes a link between a building element and an external product data sheet, when such information is available. The referenced data is not part of the base API schema and is typically not contained in IFC models, yet it is highly relevant for many downstream workflows such as specification, procurement, and facility management.

Listing 4 illustrates the schema snippet required to extend the core API. In addition to the schema extension, a corresponding resolver function is provided that dynamically searches for and retrieves a matching product data sheet URL when the field is requested by a client. The same mechanism can also be used to expose more complex information that is not represented as a simple core field, such as material associations or external classification data. This example demonstrates how heterogeneous external information sources can be integrated transparently into the API without modifying the underlying core data model.

Evaluation and Discussion

This section presents a qualitative evaluation of the proposed API design with respect to query expressiveness, modular extensibility, and practical limitations. Rather

```
extend type BuildingElement {
    dataSheetURL: String
}
```

Listing 4: GraphQL schema extension for linking external product data sheets

than benchmarking response times or comparing implementation performance, the evaluation focuses on the usability of the schema through representative query scenarios and the prototypical API implementation. The selected examples were chosen to cover two central requirements of purpose-driven BIM data access: retrieving connected information across multiple relationship levels and integrating external information without modifying the core schema.

Query Example 1: Connected Data Retrieval

Many BIM workflows require access not only to individual elements but also to related entities connected through spatial, containment, or topological relationships. Typical examples include retrieving all components of a storey together with their subcomponents, or navigating from structural elements to embedded openings or installations. In conventional REST-based APIs, such workflows would require multiple sequential requests and additional client-side aggregation logic. The proposed GraphQL API enables retrieval of connected information across multiple relationship levels within a single query. Figure 2 illustrates this capability by querying all load-bearing walls of a given storey and the windows contained within each wall. The query traverses the spatial hierarchy and containment relations and explicitly specifies the required response structure, in this case *guid* and *name* for each wall and *geometry* for each window. This example was selected because it combines several of the central design goals of the proposed schema: element-level filtering, traversal across hierarchical and containment relationships, and access to geometry as an atomic resource. It therefore demonstrates how the API supports connected data retrieval in a compact and application-oriented way. This example also showcases the integration of geometry in the form of a reference to an external resource.

Query Example 2: Modular Data Enrichment via Schema Extensions

Many analysis and decision-support workflows require information that is not natively contained in IFC models, such as environmental indicators, cost metrics, or product-specific metadata. As described previously, the proposed API addresses this challenge through a modular extension mechanism that allows additional fields to be introduced.

Figure 3 shows the result containing information from the IFC model alongside additional information in the form of the reference to an external data sheet related to the element. This example was chosen to evaluate extensibility rather than traversal complexity. It demonstrates that the API can federate information from the core building model together with external resources through resolver-based schema extensions, without requiring changes to existing query structures or disruptive API versioning.

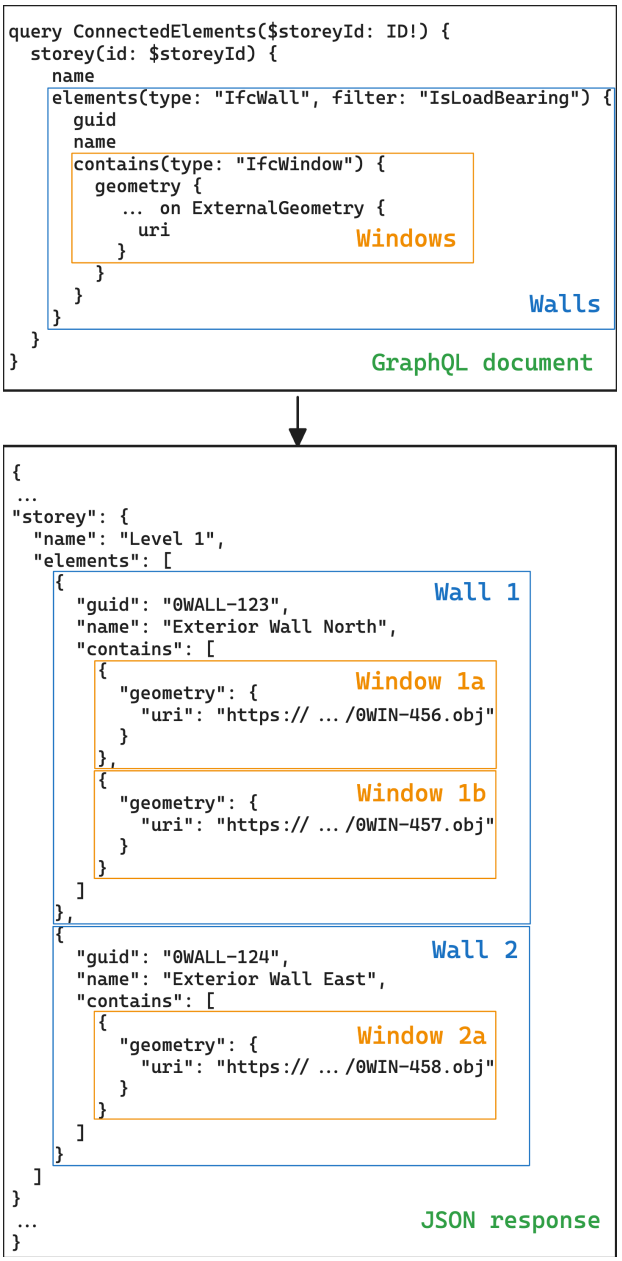


Figure 2: Requesting related elements with filtering

A key conceptual difference between the proposed GraphQL interface and SPARQL-based access to building data lies in the interaction model and query abstraction level. SPARQL provides direct access to RDF graphs and exposes the full underlying data model, which enables complex pattern matching but requires detailed knowledge of ontology structures. In contrast, the proposed API deliberately abstracts underlying data representations and exposes a task-oriented schema that is tailored to common BIM access patterns. The GraphQL interface encapsulates

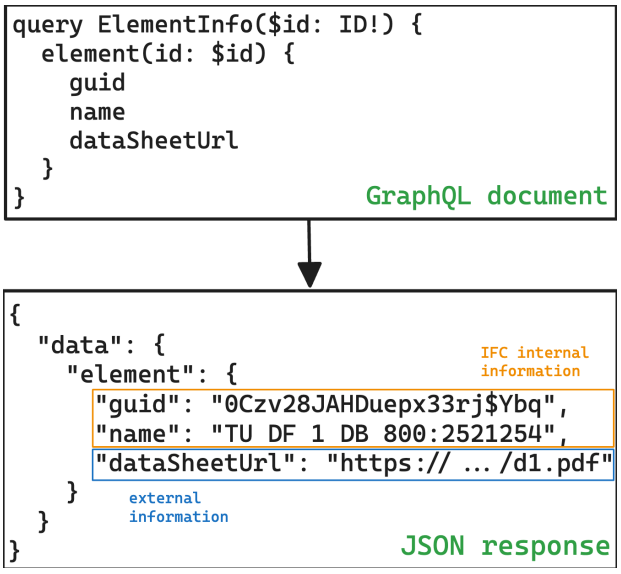


Figure 3: Query containing external information

complex graph traversal logic within resolver implementations, shifting complexity from the client to the server. At the same time, this abstraction comes at the cost of reduced direct access to the full underlying graph structure.

Scope and Limitations

The proposed API intentionally departs from a strict 1:1 mapping of the IFC schema in favor of a simplified data model. While this improves usability and query readability, it also implies that not all IFC semantics and modeling constructs are directly exposed at the API level.

The abstraction of geometry into atomic representations introduces additional limitations. Mesh-based exchange formats commonly used for visualization and data transfer, such as OBJ or glTF, do not preserve the full semantic and parametric structure of the original IFC geometry. Once converted, it is no longer possible to determine whether a shape was originally defined as BREP, CSG, or procedural description. It could be explored if extracting the geometry as a valid subset from the IFC model could be an alternative to typical geometry exchange formats.

Furthermore, certain geometric constructs cannot be represented exactly using mesh-based formats and require approximation. While these representations are sufficient for most visualization and downstream analysis tasks, they may lead to information loss in workflows that depend on higher-level modeling semantics. A detailed assessment of these trade-offs is provided in Wagner et al. (2020).

Several aspects did not fit the scope of this work and should be explored in future works.

First, the current implementation focuses on read-oriented access patterns. Future development should include support for GraphQL mutations to enable write operations, such as updating or creating (new) elements.

Second, the modular extension mechanism could be expanded to cover additional domain-specific use cases, for example quantity take-off, cost estimation, energy performance indicators, scheduling information, and sensor data integration.

Third, although the API design is resource-agnostic in principle, the current prototype only resolves data from IFC STEP files. Future work should extend the resolver implementations to support alternative IFC representations, such as graph databases, as well as native Linked Data representations.

Finally, this work does not address authentication, authorization, and access control mechanisms. Incorporating these aspects is essential for real-world deployment and represents an important direction for future research.

Conclusion

This paper presented a GraphQL-based API design for purpose-driven, fine-grained access to IFC building models. By moving away from a strict one-to-one mapping of the EXPRESS schema and instead adopting a flattened, modular interface inspired by Linked Building Data principles, the proposed approach reduces query complexity while supporting element-level, transactional access for read-oriented AECO workflows.

The presented schema design addresses two central challenges of IFC-based web access: verbose geometry retrieval and deeply nested property hierarchies. Treating geometry as an atomic, format-agnostic resource enables flexible integration of external and embedded representations, while flattened property access improves usability for common analysis and visualization workflows. In addition, the modular extension mechanism demonstrates how domain-specific information can be integrated dynamically without breaking existing clients or requiring API versioning.

The qualitative evaluation illustrates how GraphQL enables traversal of connected building information within a single request and supports the integration of external information through schema extensions. At the same time, the proposed design intentionally trades direct access to the full internal IFC structure for improved query readability and simpler client-side interaction. The resolver-based architecture further decouples the API schema from underlying data sources.

Overall, this work shows that GraphQL provides a practical and structurally explicit foundation for application-oriented access to building information models.

Acknowledgement

The research presented has been financially supported by the Federal Ministry of Education and Research (Bundesministeriums für Bildung und Forschung) in the frame of the project BauPuls360 under grant number 02K23A096, and by the German Research Foundation (DFG) in the frame of the Research Unit FOR 5672 (grant number 517965147). The responsibility for the content of this publication lies with the authors.

During the preparation of this paper, the authors used ChatGPT 5.3 to assist with writing and language refinement. All intellectual content, analysis, and conclusions remain the authors' responsibility.

References

Alexiev, V., Radkov, M., and Keberle, N. (2023). Semantic bSDD: improving the GraphQL, JSON and RDF representations of buildingSMART data dictionary. In Proc. of the LDAC Conference, pages 85–97.

Angele, K., Meitinger, M., Bußjäger, M., Föhl, S., and Fensel, A. (2022). GraphSPARQL: a graphql interface for linked data. In Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing, pages 778–785.

Beetz, J., van Leeuwen, J., and Vries, d. (2009). IfcOWL: A case of transforming EXPRESS schemas into ontologies. AI EDAM, 23:89–101.

Bonduel, M., Wagner, A., Pauwels, P., Vergauwen, M., and Klein, R. (2019). Including widespread geometry formats in semantic graphs using RDF literals. In EC3 Conference 2019, volume 1, pages 341–350. European Council on Computing in Construction.

Gleim, L. C., Holzheim, T., Koren, I., and Decker, S. J. (2020). Automatic bootstrapping of graphql endpoints for rdf triple stores. In ASLD 2020: Advances in Semantics and Linked Data — Joint Workshop Proceedings from ISWC 2020, volume 2722 of CEUR Workshop Proceedings, pages 119–134, Aachen, Germany. CEUR-WS.org.

Karalis, N., Bigerl, A., and Ngonga Ngomo, A.-C. (2023). Native Execution of GraphQL Queries over RDF Graphs Using Multi-Way Joins. In Knowledge Graphs: Semantics, Machine Learning, and Languages. IOS Press.

Lawi, A., Panggabean, B. L. E., and Yoshida, T. (2021). Evaluating GraphQL and REST API Services Performance in a Massive and Intensive Accessible Information System. Computers, 10(11):138.

Pauwels, P. (2018). Building element ontology. `https://pi.pauwel.be/voc/buildingelement/`, accessed 2026-04-03.

Pauwels, P., Krijnen, T., Terkaj, W., and Beetz, J. (2017). Enhancing the ifcOWL ontology with an alternative representation for geometric data. Automation in Construction, 80:77–94.

Pauwels, P., Terkaj, W., Krijnen, T., and Beetz, J. (2015). Coping with lists in the ifcOWL ontology. In Proceedings of the 22nd EG-ICE Workshop 2015, 13-15 June 2015, Eindhoven, The Netherlands, pages 111–120.

Quiña-Mera, A., Fernandez, P., García, J. M., and Ruiz-Cortés, A. (2023). GraphQL: A Systematic Mapping Study. ACM Comput. Surv., 55(10):202:1–202:35.

Rasmussen, M. H., Lefrançois, M., Schneider, G. F., and Pauwels, P. (2021). BOT: The building topology ontology of the W3C linked building data group. Semantic Web, 12(1):143–161.

Taelman, R., Vander Sande, M., and Verborgh, R. (2019). Bridges between GraphQL and RDF. In W3C Workshop on Web Standardization for Graph Data. W3C, pages 4–7.

van Berlo, L., Krijnen, T., Tauscher, H., Liebich, T., van Kranenburg, A., and Paasiala, P. (2021). Future of the Industry Foundation Classes: towards IFC 5. In Proc. of the 38th International Conference of CIB W, volume 78, pages 11–15.

Wagner, A., Bonduel, M., Pauwels, P., and Rüppel, U. (2020). Representing construction-related geometry in a semantic web context: A review of approaches. Automation in Construction, 115:103130.

Werbrouck, J., Senthilvel, M., Beetz, J., Bourreau, P., and Van Berlo, L. (2019a). Semantic query languages for knowledge-based web services in a construction context. In 26th International Workshop on Intelligent Computing in Engineering, EG-ICE 2019, volume 2394, page 3.

Werbrouck, J., Senthilvel, M., Beetz, J., and Pauwels, P. (2019b). Querying heterogeneous linked building data with context-expanded GraphQL queries. In 7th International Workshop on Linked Data in Architecture and Construction, volume 2389 of CEUR Workshop Proceedings, pages 21–34. CEUR-WS. org.

Wolf, N. (2026). graphql-building-api. Software repository. URL: `https://github.com/NepomukWolf/graphql-building-api`.

Wolf, N., Esser, S., and Borrmann, A. (2025). IFC2GraphQL: schema mapping and automated api construction from ifc to graphql. In Proceedings of the 42nd International CIB W78 Conference.

Zhu, J., Nisbet, N., Yin, M., Wei, R., and Brilakis, I. (2025). Releasing the power of graph for building information discovery. Automation in Construction, 172:106034.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AN INFORMATION MANAGEMENT FRAMEWORK FOR MULTI-USE CASE BUILDING AND DISTRICT DIGITAL TWINS

Joao Patacas, Edgar Segovia, Xiang Xie, Phil James, and Mohamad Kassem
Newcastle University, United Kingdom

Abstract

Digital twins are increasingly being adopted in the built environment to support data-driven operations, optimisation, and lifecycle decision-making. However, most current implementations remain limited to single use cases and centralised architectures, constraining scalability and interoperability at building and district scales. This paper proposes a use-case-driven information management framework for multi-use case building and district digital twins. The framework combines a layered architecture with a dual semantic modelling strategy that separates stable, standards-aligned core models from flexible, use-case-specific extensions. Semantic alignment is demonstrated through a real pilot. The applicability of the framework to support multi-use case integration will be demonstrated in future work.

Introduction

Digital twin (DT) technologies are increasingly recognised as a cornerstone of digital transformation in the built environment, enabling data-driven operations, optimisation, and strategic decision-making across the lifecycle of buildings and districts (Boje et al., 2020; Opoku et al., 2021). By creating a dynamic digital representation of physical assets and systems that is continuously updated through data streams, digital twins extend traditional Building Information Modelling (BIM) approaches by enabling real-time analytics, which can be used to improve the performance of the physical twin (Grieves & Vickers, 2017; Lu et al., 2020). As a result, DTs are being actively explored for applications such as energy monitoring and optimisation, predictive maintenance, renovation planning, energy grid stability, and resilience assessment at both building and district scales.

Despite this growing interest, most existing building and district digital twin implementations remain focused on individual use cases or specific stakeholder needs. Many solutions are developed as isolated or BIM-centric systems, tightly coupled to particular data sources or software platforms, which limits scalability and reuse across the many use cases during the building lifecycle (Volk et al., 2014; Pärn et al., 2017). This fragmentation limits the ability of digital twins to support integrated decision-making across multiple domains and time scales.

These challenges are further exacerbated at district scale, where digital twins must integrate data and models from multiple buildings owned and operated by different stakeholders. Centralised data architectures raise concerns related to data sovereignty, privacy, and trust (Batty, 2018; Otto et al., 2021). Conversely, fully decentralised approaches often struggle to achieve semantic interoperability and do not support coordinated analytics.

This paper addresses these challenges by proposing an information management framework designed to support multi-use case digital twins operating across building and district scales. The framework adopts a use-case-driven design methodology, ensuring that information requirements are derived from specific operational and strategic needs rather than exclusively from technical considerations. It combines this with a layered and federated digital twin architecture aligned with emerging standards and data space principles (IDSA 2021, ISO/IEC 2023).

The proposed framework is grounded in the experience and results of the WILSON project, a European research initiative focused on distributed data modelling and federated digital twinning for the sustainable operation and management of buildings and districts. This paper provides four main contributions to the field of building and district digital twins: (i) development and analysis of multi-use case information requirements; (ii) definition of a layered information management framework; (iii) specification of a dual semantic modelling strategy enabling local and federated digital twins; and (iv) illustration of the approach through a real pilot implementation.

Background and Related Work

Digital Twins in the Built Environment

The concept of the digital twin originated in manufacturing and aerospace domains and has since been adopted in the built environment to support various asset management and operational optimisation activities during building and infrastructure lifecycle (Tao et al., 2018; Grieves & Vickers, 2017). In this context, digital twins are commonly defined as digital representations of physical buildings or infrastructure assets that are dynamically linked to their physical counterparts through data streams.

In contrast to BIM, which primarily supports design and construction through static information models, digital twins emphasise dynamic behaviour, continuous data integration, and the ability to support simulation, prediction, and control during operations (Volk et al., 2014; Lu et al., 2020). Building-level digital twins are typically applied to energy monitoring, fault detection, predictive maintenance, and indoor environmental quality management, while district- and urban-scale digital twins extend these capabilities to support aggregated analyses and coordination across multiple assets (Batty, 2018).

Multi-Use Case Challenges

A persistent limitation of many digital twin implementations in the construction and facilities management domains is their focus on single use cases or isolated lifecycle phases. Different applications often require different data resolutions, update frequencies, and semantic interpretations, making it difficult to reuse data and models across use cases (Pärn et al., 2017).

For example, real-time operational optimisation requires high-frequency sensor data and control interfaces, whereas renovation planning and investment analysis depend on aggregated historical data and scenario-based simulations. Without a coherent information management and semantic alignment strategy, there is a risk that these requirements are addressed through parallel, redundant data pipelines.

Semantic Interoperability and Alignment

Semantic interoperability has long been recognised as a key challenge in construction informatics. While standards such as IFC have significantly improved data exchange, they do not fully address the integration of heterogeneous operational data sources or the needs of analytics-driven digital twins (Beetz et al., 2009; Pauwels & Terkaj, 2016).

Ontology-based approaches and semantic web technologies have been proposed to enhance interoperability by providing machine-interpretable structures that unify BIM, sensor data, geospatial information, and external datasets, enabling better knowledge representation and facilitating reasoning (Karabulut et al 2024). However, many existing approaches rely on monolithic or overly complex models that are difficult to adapt and reuse to different project and operational requirements.

Methodological Approach

Multi Use-Case-Driven Framework Design

The framework presented in this paper adopts a multi use-case-driven design methodology. Rather than defining information models and architectures in isolation, the approach starts from defining use cases that capture the goals, actors, processes, and decision contexts relevant to building and district management. The use case methodology considers the input from domain experts, providing general ideas and functional requirements, while system experts detail these use cases down to a level that can be used to specify interfaces, dedicated functionalities, data and service model exchanges. This methodology ensures that the resulting digital twin infrastructure directly supports practical needs and avoids unnecessary complexity.

Within the WILSON project, use cases were developed through an iterative process involving technical partners, pilot site operators, and other stakeholders. This process combined action design research principles (Sein et al. 2011) with established use-case modelling standards (BSI 2020), resulting in a consistent set of use-case descriptions covering functional and non-functional requirements. Each stage in the Action Design Research (ADR) framework adopted in use case development is summarised in Table 1.

Table 1: Action design research for use case development

Stage 1 - Problem formulation	Use case scoping: Identification of UC leaders and stakeholders
Stage 2 - Building, intervention and evaluation - Use case development	• Initial development of use cases • Use case workshops (round 1) • Further development of use cases • Use case workshops (round 2)
Stage 3 - Reflection and learning - Developing detailed user requirements	Develop final version of use cases, and identification of main system functionalities
Stage 4 – Formalisation & Reporting	Reporting in project deliverables, scientific publications, etc.

From Use Cases to Information Requirements

Each use case was analysed to identify its information requirements, including actors, scenarios, functional and data requirements. These requirements were then abstracted to identify common information needs across multiple use cases, as well as specific extensions required for particular applications. This analysis provides the basis for defining a shared information management framework capable of supporting multiple use cases without duplication.

Use Cases and Information Requirements

Overview of Building and District Use Cases

The information management framework proposed in this paper is grounded in a comprehensive set of building- and district-level use cases that reflect operational and strategic needs in the built environment. Following structured stakeholder engagement and iterative analysis, use cases and their scope was defined (Table 2). The identified use cases illustrate the diversity of objectives, time horizons, and data requirements that the proposed information management framework must support.

Stakeholders and Decision Layers

The identified use cases (Table 2) involve a wide range of stakeholders, each interacting with the digital twin at different decision layers. Facility managers and building operators typically engage with the digital twin on a short-term operational basis, using real-time or near-real-time information to adjust control strategies, diagnose faults,

or respond to alerts. Their primary concern is maintaining efficient and reliable building operation while ensuring occupant comfort.

Table 2: Overview of identified use cases

Use case	Scope
Decentralised multi-source data management	Develop a federated DT approach, to achieve interoperability among diverse data sources scattered across various platforms and formats.
Energy monitoring and optimisation	Optimize energy management at building and district levels, allowing owners and other key actors to make data-driven decisions to enhance energy performance, comfort, and environmental sustainability.
Data-driven network asset management	Provide energy network operators with tools for optimising asset management, ensuring energy availability and network resilience.
Predictive maintenance	Enable proactive maintenance by predicting equipment failures and optimising maintenance schedules, reducing downtime and costs.
Risk, resilience and biodiversity nexus	Enable building owners to mitigate climate change hazards by identifying environmental risks at building and district levels.
Renovation planning and design	Apply LCA for planning and design of energy-efficient renovations and retrofitting projects.
Technoeconomic investment planning for property owners	Evaluation of energy improvement interventions and economic feasibility at building and community levels.
Boost of green energy investments by market actors	Enhancing green energy investments by leveraging DTs and network planning algorithms

Asset owners and portfolio managers operate at a medium- to long-term decision layer. They use digital twin outputs to support renovation planning, lifecycle cost analysis, and investment decision-making. For these stakeholders, historical data, performance trends, and scenario-based simulations are more important than real-time control.

At district level, additional stakeholders come into play, including energy service companies, distribution system operators, and public authorities. These actors are interested in aggregated performance indicators, flexibility potential, and compliance with regulatory or policy objectives. The coexistence of these diverse stakeholders highlights the need for an information management framework that can present different views of the same underlying data while enforcing appropriate access controls.

Information Requirements Analysis

Analysis of the use cases reveals a set of common information requirements that cut across applications and scales. Static information forms the foundation of the digital twin and includes building geometry, construction properties, system topology, and asset metadata. This information is often derived from BIM models and asset registers and changes infrequently over time.

Dynamic information consists primarily of time-series data from sensors, meters, and building management systems. These data streams capture the operational state of buildings and systems, such as energy consumption, temperatures, equipment status, and control signals. Depending on the use case, data may be required at different temporal resolutions, ranging from seconds or minutes for control-oriented applications to hourly or daily aggregates for reporting and analysis.

Contextual information provides essential background for interpreting building performance. Examples include weather data, occupancy schedules, energy tariffs, and grid conditions. Many analytical models rely on such contextual data to generate accurate predictions or recommendations.

In addition to raw data, use cases require access to derived information, such as key performance indicators (KPIs), alarms, forecasts, and optimisation results. Managing these different categories of information in a consistent and reusable manner is a central challenge addressed by the proposed information management framework.

Cross-Use-Case Synergies and Conflicts

A key finding of the use-case analysis is the significant potential for synergy through shared information resources. For example, the same sensor data streams can support energy optimisation, fault detection, and indoor comfort analysis. Similarly, building geometry and system topology information is relevant for both operational optimisation and long-term renovation planning.

However, conflicts may arise when different use cases impose competing objectives or constraints. Energy optimisation strategies that minimise consumption may negatively affect occupant comfort, while aggressive demand response actions may increase equipment wear. These conflicts cannot be resolved solely at the data level; they require explicit governance, prioritisation, and negotiation mechanisms embedded within the digital twin framework.

Proposed Information Management Framework

Framework Overview

The proposed information management framework is designed to address the complexity and diversity of multi-use case digital twins in the built environment. It adopts a multi-layer model inspired by Semantic Web principles

and DT reference architectures (CEN 2026, IDSA 2021, ISO 2021, Karabulut et al. 2024) that separates concerns while enabling strong integration across layers.

Heterogeneous data streams originating from BIM models, IoT sensors, legacy BMS systems, and operational platforms are transformed into semantically enriched and machine-interpretable representations, ensuring that local and federated DTs operate with shared meaning, supporting reasoning, analytics, and standardised data exchanges. By allowing data, models, and services to be shared across multiple applications, the framework enhances scalability, enabling the addition of new buildings, data sources, or use cases without fundamental redesign. The framework comprises the logical layers described below (Figure 1).

Acquisition Layer

The acquisition layer forms the interface between the physical world and the digital twin. It is responsible for ingesting heterogeneous data from a wide range of sources, including BIM repositories, building management systems, IoT devices, and external data services, while preserving essential metadata such as timestamps, sensor calibration, and spatial context. Data ingestion follows the principles of ISO/IEC 30141:2024 - Internet of Things (IoT) — Reference architecture (ISO/IEC 2024) and ISO 12006 for BIM-based information management (ISO 2022). By aligning data acquisition with these standards, it ensures consistent data quality and traceability across both building and district DTs, forming the foundation for accurate semantic mapping and analytics.

Integration Layer

Data Space connectors are implemented following the International Data Spaces (IDS) Reference Architecture (IDSA 2021) and Gaia-X Federation Services (Gaia-X 2022). This layer ensures data sovereignty and standardised access through policy-based usage control and authentication. Personalised Data Hubs (PDHs) provide APIs for tool access while enforcing local governance (Ramachandruni et al. 2025).

Semantic Layer

The semantic layer manages semantic mediation and ontology alignment between local data and the Core DT Model. This process follows ISO/IEC 21838 – Top-Level Ontologies for Interoperability (ISO/IEC 2021), to translate data into structured RDF or JSON-LD formats. The layer comprises two components:

- Semantic Mediation Gateways, which perform schema mapping and transformation from raw BIM, sensor, and energy datasets to standard ontologies such as BOT, SSN/SOSA, SAREF, and BRICK;
- Vocabulary Hub, which serves as a central ontology and vocabulary registry that provides shared concepts, mappings, and equivalence rules across DTs, ensuring alignment and semantic consistency (Ramachandruni et al. 2025).

Knowledge Layer

This layer structures the semantic building and district DTs following the dual-layer ontology model (i.e. core plus use-case specific extensions) to support reasoning and federated queries across DTs. The knowledge layer defines how entities, properties, and behaviours are modelled, enabling both static and dynamic aspects (geometry, operational parameters, behavioural trends, and contextual conditions) to be represented in an interoperable format. This layer provides the structural and semantic basis for the building-level and federated district-level DTs.

Applications Layer

The applications layer enables reasoning, analytics, and integration across Digital Twins tools and services. It aims to support cross-building insights and district-level use cases aligned with ISO 23247's composability and interoperability of DTs, ensuring that DTs can seamlessly exchange data and coordinate analyses across building (local) and district-level domains (ISO 2021).

Governance Layer

The governance layer underpins the entire framework by defining rules and mechanisms for data access, sharing, and usage. It addresses critical issues such as privacy,

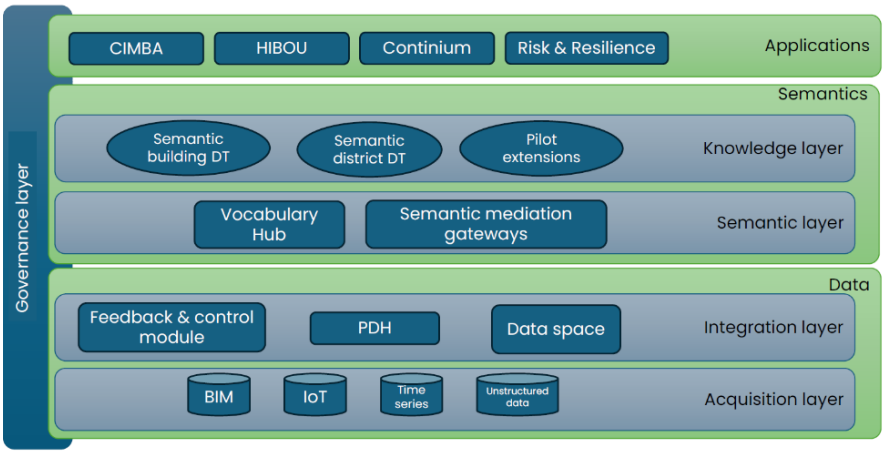


Figure 1 Proposed Information Management Framework

security, consent, and data sovereignty, to allow cross-building analytics and decision support while respecting organisational boundaries and regulatory constraints.

Semantic Local and Federated Digital Twins

Role of Semantic Alignment in Multi-Use Case Digital Twins

Semantic alignment is a central enabler of multi-use case digital twins, particularly in environments characterised by heterogeneous data sources, legacy systems, and multiple stakeholders. Without a shared semantic understanding, data originating from different buildings, systems, or domains cannot be meaningfully combined or reused across applications. This challenge becomes even greater when moving from building-level digital twins to federated district-level twins.

In the proposed framework, semantic alignment is treated as a continuous and systematic process embedded within the digital twin lifecycle. It provides the foundation for interoperability across data sources, analytical models, and services, enabling different use cases to operate concurrently on a shared information basis.

Dual-layer Modelling Strategy for Semantic Alignment

To balance interoperability, extensibility, and long-term maintainability, the framework adopts a dual-layer modelling strategy for semantic alignment. This strategy is particularly suited to construction and built-environment contexts, where stable reference models must coexist with project-specific, evolving information needs.

The core digital twin model captures fundamental concepts that are common across buildings, districts, and use cases. It includes the following entities: *District, Building, Space, System, Asset/Component, Sensor/Actuator, Observation*. The core model is intentionally kept compact, stable, and is aligned with internationally recognised standards and specifications used in construction and digital twin ecosystems. This alignment ensures long-term interoperability and supports integration with external platforms and systems (Table 3). Complementing the core model, context- and use-case-specific extension models are defined to represent additional concepts, attributes, or relationships required in specific use cases. These extensions may capture specialised system configurations, custom KPIs, domain-specific constraints, or locally relevant semantics that are not appropriate for inclusion in a generic core. Importantly, extensions are not isolated schemas: they are explicitly linked to the core model through well-defined semantic mappings. New use cases or data sources can be integrated by extending the model rather than modifying the core, reducing the risk of fragmentation and ensuring backward compatibility across the digital twin federation.

Table 3 – Alignment between DT core model entities and reference standards/ontologies

Core Entity	Reference Class / Standard Alignment	Description (ISO 2023, 2024, ETSI 2017, Brick 2023, W3C 2017, 2021)
District	CityObjectGroup	Represents an aggregation of multiple buildings and infrastructures (e.g. energy, mobility, environment) forming a district-level twin.
Building	bot:Building / IfcBuilding	Represents an individual physical building and its associated metadata such as ownership, use, and geolocation.
Space	bot:Space / IfcSpace	A physical or functional zone within a building (e.g. room, corridor, floor zone). Used to link spatial context with sensors and assets.
System	saref:DeviceGroup / brick:System/ IfcSystem	Represents technical systems such as HVAC, lighting, or electrical distribution that connect assets and sensors.
Asset / Component	saref:Device / brick:Equipment / IfcAsset	Represents components within a system (e.g. pump, fan, EV charger). Provide the link between physical equipment and their monitoring points.
Sensor / Actuator	sosa:Sensor/sosa: Actuator / IfcSensor/IfcActuator	Devices that capture observations or perform control actions on assets and systems.
Observation	sosa:Observation	A single measurement or data record generated by a sensor or actuator, including timestamp, unit, and value.

Semantic Alignment in Practice: Newcastle Helix

The Newcastle Helix innovation district is a mixed-use urban campus developed through a partnership between Newcastle University, Newcastle City Council and private stakeholders (Figure 2). Within the WILSON project, the pilot focuses on three representative buildings: Urban Sciences Building (USB), The Spark, and The Core. These three buildings provide a diversity of typologies and system maturities (higher education research facility, collaborative workspace, and commercial office respectively), which makes Helix an appropriate demonstrator for testing the semantic DT framework across heterogeneous operational contexts.



Figure 2 - Newcastle HELIX innovation district (UK)

The dual-layer approach to data modelling was applied to the Helix pilot considering:

- A core alignment layer, providing a common shared vocabulary and minimal schema.
- Pilot-specific extensions, representing Helix-specific concepts such as district heating nodes, site-wide EV infrastructure, and Urban Observatory (UO 2026) sensor feeds.

Figure 3 provides the UML diagram for the WILSON DT data model for Newcastle Helix pilot, highlighting the implementation of the proposed dual-layer modelling approach. The core model entities are represented in white, while pilot-specific extensions, representing Helix-specific concepts are highlighted in orange, represented as extensions (or inherited by) the core model.

The mapping of Helix datasets to the WILSON model follows these rules:

BIM data mapped to *Building / Space / Asset*. BIM models are mapped to the WILSON Building and Space classes. Asset registers (e.g. for AHUs, pumps, EV chargers) populate Asset entities with manufacturer and model metadata. The code snippet in TTL syntax in Listing 1 illustrates this example using existing ontologies:

Listing 1 - BIM data mapped to Building / Space / Asset

```
@prefix : <https://wilson-project.eu/helix/> .
@prefix bot: <https://w3id.org/bot#> .
@prefix schema: <https://schema.org/> .

:USB_Building  a bot:Building ;
    schema:name "Urban Sciences Building" ;
    bot:hasSpace :Some_USB_Lab ;
    bot:hasElement :USB_AHU_x .

:USB_Lab_2F  a bot:Space ;
    schema:name "Second Floor Lab" .

:USB_AHU_01  a schema:Product ;
    schema:manufacturer "Some Manufacturer" ;
    schema:model "Some model" .
```

BMS / IoT feeds mapped to *Sensor / Observation* (Timeseries). Real-time building management system (BMS) feeds, available at the USB and selected systems in other buildings, populate the Sensor and Observation entities. Where structured APIs are available (as in the Urban Observatory APIs (UO 2026)), the Entity–Feed–Timeseries chain is preserved to maintain semantic richness. Aggregated datasets (e.g. hourly averages) are represented as Observations, with metadata describing aggregation level and sampling frequency. The code snippet in TTL syntax in Listing 2 illustrates this example using existing ontologies:

Listing 2 - BMS / IoT feeds mapped to Sensor / Observation

```
@prefix : <https://wilson-project.eu/helix/> .
@prefix sosa: <http://www.w3.org/ns/sosa/> .
@prefix xsd:
<http://www.w3.org/2001/XMLSchema#> .

:USB_TempSensor_2F  a sosa:Sensor ;
    sosa:observes :AirTemperature ;
    sosa:madeObservation :Obs_2025_10_24_1200 .

:Obs_2025_10_24_1200  a sosa:Observation ;
    sosa:hasResult "21.3"^^xsd:float ;
    sosa:resultTime "2025-10-
24T12:00:00Z"^^xsd:dateTime .
```

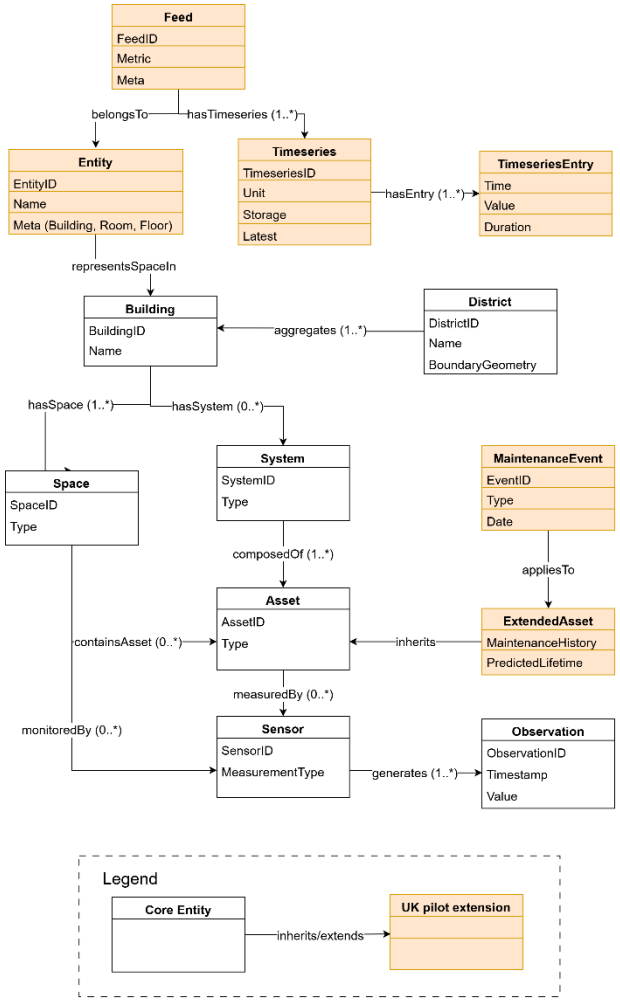


Figure 3 - Digital Twin data model for Helix UK pilot

Asset registers & maintenance logs mapped to *Asset / MaintenanceEvent.* Structured asset condition scores and lifecycle estimates are mapped to attributes of the Asset class (e.g. conditionScore,remainingLifetime). Maintenance actions are captured as MaintenanceEvent, linked to ExtendedAsset metadata. The code snippet using TTL syntax in Listing 3 illustrates this example using existing ontologies:

Listing 3 - Asset registers & maintenance logs mapped to Asset / MaintenanceEvent

```
@prefix : <https://wilson-project.eu/helix/> .
@prefix schema: <https://schema.org/> .

:Spark_Pump_03  a schema:Product ;
    schema:name "Heating Pump x" ;
    schema:manufacturer "Manufacturer Name" ;
    schema:model "Heating Pump model" ;
    schema:additionalProperty [
        schema:name "conditionScore" ;
        schema:value "0.80"
    ] .

:Spark_Pump_03_Maint_2025_05_14  a
schema:MaintenanceEvent ;
    schema:object :Spark_Pump_x ;
    schema:startDate "2025-01-01" ;
    schema:description "Bearing replacement" .
```

The Helix pilot illustrates several benefits of the proposed semantic alignment approach. Shared core semantics enable the same underlying data to support multiple use cases, while pilot-specific extensions preserve the semantic richness required for detailed local analyses.
From a scalability perspective, the dual modelling strategy reduces integration effort when adding new buildings or data sources, as only mappings to the core model are required. Additionally, it enables selective exposure of information at district level, supporting data sovereignty and compliance with the organisational constraints of the Helix district.

Discussion

The primary contribution of this work lies in the integration of a dual semantic modelling strategy with a federated digital twin architecture, grounded in real multi-use case requirements. While semantic models, BIM standards, and digital twins have been extensively studied individually, their systematic combination to support multiple building- and district-level use cases remains limited. Unlike monolithic or purely BIM-centric approaches, the proposed framework:

- Treats semantics as a key architectural concern, not merely a data integration layer;
- Separates stable, standards-aligned semantics from evolving, use case-specific requirements;
- Aligns with emerging federated data and data space paradigms, which are increasingly relevant for cross-organisational construction ecosystems.

Limitations and Practical Considerations

Despite its advantages, the proposed approach also presents challenges. Establishing and maintaining high-quality semantic mappings requires upfront effort and domain expertise. Legacy systems and incomplete BIM models may limit the availability of structured information, necessitating incremental or hybrid integration strategies.

Furthermore, while the framework supports federation conceptually and architecturally, its effectiveness will depend on stakeholder willingness to adopt shared governance principles and semantic conventions. Addressing these factors is an important focus for future work, both during and beyond the duration of the WILSON project.

Conclusions and Future Work

This paper has presented an information management framework for multi-use case building and district digital twins. By combining a use-case-driven design methodology with a layered and federated architecture, the framework addresses the complexity inherent in supporting diverse applications and stakeholders across multiple spatial scales.

Future work will focus on the implementation and evaluation of the framework in real-world pilots, to demonstrate how shared information resources, model management, and governance mechanisms can enable scalable and reusable digital twin implementations to support use cases at the local (building) and district level.

Acknowledgements

The authors gratefully acknowledge the support of the WILSON project (May 2024 – April 2028), funded by the European Union under Grant Agreement ID: 101147267, within the Horizon Europe framework (Call: HORIZON-CL5-2023-D4-02-01).

References

Batty, M. (2018) Digital twins. Environment and Planning B: Urban Analytics and City Science, 45(5), 817–820.

Beetz, J., van Leeuwen, J., & de Vries, B. (2009) IfcOWL: A case of transforming EXPRESS schemas into ontologies. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 23(1), 89–101.

Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020) Towards a semantic Construction Digital Twin: Directions for future research. Automation in Construction, 114, 103179.

Brick Consortium (2023) Brick Schema Specification, Version 1.3. Available at: https://brickschema.org (Accessed: 6 February 2026).

British Standards Institute. (2020). BS IEC SRD 62559-4:2020 Use case methodology. Best practices in use case development for IEC standardization processes and some examples for application outside standardization.

CEN (2026) EN 18162:2026 Building Information Modelling (BIM) - Digital twins applied to the built environment - Concept and definitions.

ETSI (European Telecommunications Standards Institute), 2017. ETSI TS 103 264 V2.1.1 – Smart Appliances; Reference Ontology and oneM2M Mapping (SAREF). Sophia Antipolis: ETSI. Available at: https://www.etsi.org/deliver/etsi_ts/103200_103299/103264/02.01.01_60/ts_103264v020101p.pdf (Accessed: 6 February 2026).

Gaia-X (2022). Policy Rules and Architecture. Gaia-X Association for Data and Cloud (AISBL). Available at: https://gaia-x.eu (Accessed: 6 February 2026).

Grieves, M., & Vickers, J. (2016) Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In Transdisciplinary Perspectives on Complex Systems.

IDSA (International Data Spaces Association) 2021. IDS Reference Architecture Model 4.0. IDSA, Dortmund. Available at: https://internationaldataspaces.org (Accessed: 6 February 2026).

ISO (International Organization for Standardization) 2021. ISO 23247:2021. Automation systems and integration — Digital Twin framework for manufacturing. ISO, Geneva.

ISO/IEC (International Organization for Standardization) 2021. ISO/IEC 21838:2021. Information technology — Top-level ontologies. ISO, Geneva.

ISO (International Organization for Standardization) 2022. ISO 12006-3:2022. Building construction — Organization of information about construction works — Part 3: Framework for object-oriented information. ISO, Geneva.

ISO (International Organization for Standardization) 2023 ISO/IEC 30173:2023 Digital Twin — Concepts and terminology.

ISO (International Organization for Standardization) 2023. ISO 19157-1:2023 Geographic information — Data quality. Part 1: General requirements. ISO, Geneva.

ISO (International Organization for Standardization) 2024. ISO 16739-1:2024. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries. Part 1: Data schema. ISO, Geneva.

ISO/IEC (2024). ISO/IEC 30141:2024 – Internet of things (IoT) — Reference architecture. Geneva: International Organization for Standardization and International Electrotechnical Commission.

Karabulut, E.; Pileggi, S. F.; Groth, P.; Degeler, V. (2024), Ontologies in Digital Twins: A systematic literature review, Future Generation Computer Systems, Volume 153, 2024, Pages 442-456, ISSN 0167-739X, https://doi.org/10.1016/j.future.2023.12.013.

Lu, Q., Xie, X., Heaton, J., Parlikad, A. K., & Schooling, J. (2020) From BIM towards digital twin: Strategy and future development for smart asset management. Studies in Computational Intelligence, 853, 392–404.

Opoku, D.-G. J., Perera, S., Osei-Kyei, R., & Rashidi, A. (2021) Digital twin application in the construction industry: A literature review. Journal of Building Engineering, 40, 102726.

Otto, B., et al. (2021) International data spaces—A distributed data ecosystem for sovereign data sharing. Procedia CIRP, 104, 112–117.

Pärn, E. A., Edwards, D. J., & Sing, M. C. P. (2017) The building information modelling trajectory in facilities management. Automation in Construction, 75, 45–55.

Pauwels, P., & Terkaj, W. (2016) EXPRESS to OWL for construction industry: Towards a recommendable ifcOWL ontology. Automation in Construction, 63, 100–133.

Ramachandruni, A., Sharma, R., Petrova, E., & Pauwels, P. (2025). Vocabulary Hub for Semantic Data Exchange and Interoperability in an AECO Domain-specific Dataspace. https://doi.org/10.35490/EC3.2025.340

Sein, Henfridsson, Purao, Rossi, & Lindgren. (2011). Action Design Research. MIS Quarterly, 35(1), 37. https://doi.org/10.2307/23043488

Tao, F., Qi, Q., Wang L., & Nee A.Y.C. (2019) Digital Twins and Cyber–Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison Engineering, 5(4), 653–661.

Urban Observatory (UO) 2026 – Explore data from sensor clusters monitoring urban environments in Newcastle. https://newcastle.urbanobservatory.ac.uk/

Volk, R., Stengel, J., & Schultmann, F. (2014) Building Information Modeling (BIM) for existing buildings — Literature review and future needs. Automation in Construction, 38, 109–127.

W3C (World Wide Web Consortium), 2017. Semantic Sensor Network Ontology (SSN) and Sensor, Observation, Sample, and Actuator Ontology (SOSA). W3C Recommendation, 19 October 2017. Available at: https://www.w3.org/TR/vocab-ssn/

W3C (World Wide Web Consortium), 2021 Building Topology Ontology (BOT) – a W3C Building Data Working Group draft ontology for the built environment. W3C, 28 June 2021. Available at: https://w3c-lbd-cg.github.io/bot

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AN IFC FRAMEWORK FOR INFORMATION AND DOCUMENT MANAGEMENT IN CONSTRUCTION PROCESSES

Beatrice Maria Toldo[1], Carlo Zanchetta[2]
[1]Department of Management and Engineering, University of Padua, 36100 Vicenza, Italy
[2]Department of Civil, Environmental and Architectural Engineering, University of Padua, 35131 Padua, Italy

Abstract

Information management of construction site documentation is crucial for cost control in construction companies. Currently, sites are often managed with traditional tools, leading to slow archiving processes and data inconsistencies. This highlights the need for a more efficient document-management system, where the IFC format can act as a standardized structure for organizing the data recorded in the documents. This paper proposes a conceptual framework and proof-of-concept implementation to optimize the use of site documentation by mapping and classifying key information. The structured extraction of data, as man-hours, enables integration with management control systems and BIM models, supporting more accurate evaluation of construction cost performance.

Introduction

The progressive digitalization of information flows represents a strategic cornerstone in the evolution of the construction sector toward Industry 4.0. The capability to collect data directly from construction sites and manage it efficiently is a fundamental requirement for the monitoring and control of contemporary projects, as well as a significant operational advantage. Effective information management enables the overcoming of traditional communication barriers and supports data-driven decision-making processes based on accurate and up-to-date information.

In recent years, the Building Information Modeling (BIM) methodology has significantly transformed the planning and monitoring of construction projects. Industry Foundation Classes (IFC) (EN ISO 16739-1:2024), an internationally recognized standard, constitute the principal open data schema for the representation and interoperable exchange of BIM information across heterogeneous software platforms and stakeholders, thereby enhancing collaboration and information consistency within the construction industry. A first stream of research focuses on the role of IFC as a standard for interoperability. The research highlights that, although the IFC provides a neutral framework for data sharing, its practical application is often hindered by inconsistent software implementations and potential data loss during data exchange. (Kone and Mahesh, 2025; Noardo et al., 2021). This becomes evident when maintaining consistency across multiple model versions or linking heterogeneous data, which often requires additional layers of data management or external databases to ensure reliable information management (Mastino et al., 2026; Pauwels et al., 2017). A second research area investigates the use of IFC for construction process modeling and project control. Recent studies have highlighted the importance of using IFC processes, such as *IfcTask*, *IfcProcedure*, and *IfcRelAssignsToProcess*, in bridging the gap between static product models and dynamic construction scheduling (Avogaro et al., 2024). Existing studies demonstrate that IFC-based models can facilitate the comparison between planned and actual progress. However, these approaches typically focus on predefined workflows, while the integration of site data, especially when derived from unstructured or semi-structured documents, remains a significant gap in the research (Avogaro and Zanchetta, 2025).

Despite the increasing adoption of BIM during the design phase, information management at the construction site level remains an area requiring further development, largely due to the high heterogeneity of data involved (Kerosuo et al., 2015).

At present, the management of temporal and economic progress on construction sites still relies predominantly on traditional, document-based project control approaches (e.g., PDF documents and spreadsheets). These are time-consuming and, if not properly managed within a centralized database, may lead to information loss (Dave et al., 2016). However, such approaches are inadequate for the dynamic nature of modern construction sites, which require a higher level of process digitalization. Labor costs account on average for between 30% and 50% of total project costs (Mohammed et al., 2024), while inefficient management of material resources can represent up to 70% of the overall project cost (Kasim et al., 2019). Current shortcomings in data collection and management systems arise from information fragmentation and manual reporting practices. Cost management, as well as the monitoring of construction progress over time, is often based on pre-existing documentation that frequently remains unused within company archives. This lack of integration between planning activities and on-site construction makes it difficult to assess project performance in real time, resulting in delayed reporting of schedule deviations and inaccurate evaluations of productivity.

To bridge this gap, the literature has proposed new field data collection systems that leverage emerging technologies (Deng et al., 2025), such as RFID for material tracking (Jain et al., 2024), or the use of tags for automated inventory identification via dedicated scanners. There are also examples of IoT sensors for the vertical tracking of materials on multi-story construction sites, which communicate via wireless protocols, ensuring the efficient transmission of field-collected data to a central platform (Gehring et al., 2025). Recent studies also explore Machine Learning and Natural Language Processing (NLP) to extract information from documents (Jafari et al., 2021; Singh et al., 2023). Although these approaches make it possible to transform vast amounts of unstructured digital data into actionable insights (Mahire et al., 2024), the research highlights that such approaches face persistent challenges related to the poor interpretability of models and the handling of linguistic ambiguity. In the specific context of the AEC sector, interoperability is further hindered by the fact that the IFC standard does not natively support dynamic data (time series) for systematically extracting and converting data from the site to Digital Twins (Du et al., 2024).

In this technological development context, BIM emerges as a data management tool capable of centralizing information from multiple stakeholders. In practice, it constitutes a single information database able to integrate not only physical quantities but also cost-related data for economic monitoring (Smith, 2014). To ensure interoperability and accurate information exchange among the various actors, IFC plays a central role. As a neutral and open data format, IFC enables interoperability across different platforms, allowing the storage and exchange of not only geometries but also complex metadata related to costs, schedules, and resources. The significance of IFC in project control lies in its hierarchical structure, which allows the association of specific properties (Property Sets) with each building entity, covering construction site documentation such as delivery notes, daily reports, and work progress status.

To address this gap, the paper investigates the following research questions:

RQ1: How can construction site documentation be semantically structured and integrated within the IFC schema?

RQ2: How can IFC entities be used to represent the relationships between documents, resources, activities, and time in construction processes?

This paper contributes to the existing body of knowledge by:

(i) proposing a semantic framework for integrating construction site documentation into the IFC schema;

(ii) formalizing the relationships between documents, activities, resources, and time within a standardized data model;

(iii) demonstrating the feasibility of using IFC as a process-oriented information system for construction management.

The novelty of the proposed approach lies in enabling the use of IFC as a process-oriented data infrastructure for integrating heterogeneous, document-based construction data into a unified and queryable information model.

Framework Overview

The proposed framework aims to support the document management of construction production and costs. The integration of documentation within the BIM environment, through the IFC standard, enables the transformation of documentary data into structured data. In this work, IFC4x3 ("IFC4.3.2.0 Documentation," 2024) has been used to implement the framework. Documentation is classified according to four categories of direct costs: materials, labor, equipment rentals, and subcontracting. From this perspective, site documentation is not regarded merely as a bureaucratic requirement, but as an informational input for the quantification of variable production costs. Each category is associated with specific types of site documents and corresponding IFC resource entities.

Although the framework is designed to be applicable to all cost categories, the present work focuses on labor-related documents for demonstration purposes. Labor is selected due to its significant impact on construction costs and the richness of information typically contained in daily site reports. The same methodological principles can be applied to the other categories within the proposed framework.

The framework developed in this study manages construction site documents by integrating the data they contain into the BIM model. The starting point is the analysis of labor data, historically the most complex cost item to monitor due to its dynamic nature. The framework identifies daily reports as the primary source of data.

Information relating to the document

In general, information relating to documents can be divided into information describing the document itself and information contained within the document. This division allows the relevant IFC classes to be identified, where the information can be entered. *IfcDocumentInformation* is the most suitable IFC class for collecting metadata relating to the document. This class is generic for any document. The actual content of the document is not defined in this specification, but its location is managed via the *Location* attribute. In this field, it is possible to transcribe the URL of the document, thus generating an external link to the database. The IFC file does not directly contain the document, thus avoiding becoming excessively large in terms of MB. Instead, the IFC file maintains a reference to the document. The standard provides several other attributes that allow the association and description of the information related to a document. Among the main ones are *Identification*, which enables to assign a unique identifier to the document, *CreationDate*, useful for indicating the date of creation, and *Name*, which allows to assign a name or a short description of the content.

The *IfcDocumentInformation* element can be linked to model elements via the *IfcRelAssociatesDocument* relationship, which allows a document to be associated with one or more subtypes of IfcObject. In this way, a single document can refer to several model objects at the same time, ensuring centralised and consistent information management.

This feature is particularly significant in complex operational scenarios, where a single document may contain data relevant to multiple components of the model. For example, in the context of labor management, documents such as daily report can record activities related to different tasks in the schedule, involving multiple construction elements and the resources associated with them. The ability to link a document to multiple objects ensures integrated and up-to-date traceability of information, improving data consistency and coordination between the various figures involved in the design and construction process.

Information contained in the document: resources

Regarding the information contained in the documents, the IFC standard provides a clear framework for mapping using specific entities related to tasks, resources and costs. Regarding labor resources, the main documents produced are daily reports and timesheets. The latter, which is mandatory for public works, is also widely used in private projects. These documents contain detailed information on the work carried out, the personnel involved, the hours worked and, where relevant, the equipment and materials used. In many cases, information regarding the labor of subcontractors can also be included.

Within the IFC model, this information is stored in resource-specific entities provided by the standard, including *IfcConstructionMaterialResource* for material resources, *IfcConstructionProductResource* for products, *IfcLaborResource* for labor, *IfcConstructionEquipmentResource* for equipment, and *IfcSubContractResource* for subcontracted work. All these entities are subtypes of *IfcConstructionResource*, from which they inherit common attributes useful for resource management.

The handling of activities performed is treated separately within the *IfcTask* class, which represents a unit of work within the project schedule. This structure allows for a conceptual link between the resources (labor, equipment, materials) and the actual task executed.

IfcLaborResource

Hours worked are associated with *IfcLaborResource* entities, enabling detailed and consistent tracking of work performed. The standard provides the Property Set *Pset_ConstructionResource*, which includes properties of relevance for reporting purposes. Among these, *ActualWorkTime* allows the recording of actual hours worked through an *IfcTimeSeries*, which defines time intervals associated with measurements. As a time series, *ActualWorkTime* can contain multiple values, effectively constituting a real-time record stored within the information model.

The *IfcIrregularTimeSeries* class, a subclass of *IfcTimeSeries*, is used to record data that varies over time. Although the recording of hours worked may generally be considered regular, variations in recording intervals can occur due to factors such as weather conditions, holidays, or unforeseen events on site. Using the *Values* attribute, a list of recorded values can be reported, each associated with its corresponding date of measurement. In practice, activities involving human resources often span multiple days or time intervals. As a result, several reports are typically produced to document the ongoing execution of the same task by the same personnel over successive days.

In this context, within the IFC standard, an *IfcIrregularTimeSeries* contains multiple *IfcIrregularTimeSeriesValue* entities, each corresponding to a measurement taken on a different day and including the total man-hours expended on the specific activity during the relevant period. Since this data originates from separate documentation, the *IfcLaborResource* entity is linked to multiple instances of *IfcDocumentInformation*, one for each report from which the information is derived. Figure 1 presents a representative example of this data structure. This approach enables hours worked to be associated with each resource in a structured and dynamic manner, ensuring transparency, traceability, and continuous updating of information within the Common Data Environment.

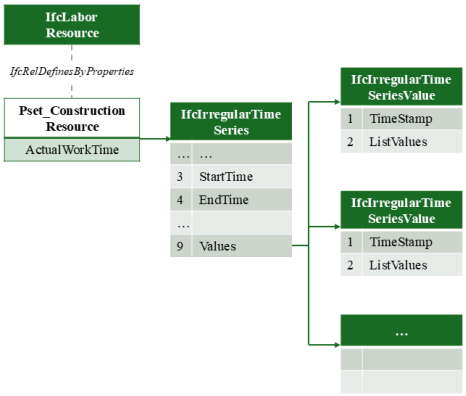


Figure 1: Mapping of temporal properties for an IfcLaborResource entity

IfcSubContractResource

Daily site reports also document the hours worked by subcontracted labor. In such cases, within the IFC standard, the reference class for the structured storage of this information is *IfcSubContractResource*, which represents the subcontractor responsible for the execution of the activities. The *IfcSubContractResource* entity is not used directly to record working hours, but serves as a reference organizational entity. For the purpose of capturing labor hours, one or more instances of the *IfcLaborResource* class are created, each linked to the corresponding *IfcSubContractResource* entity. In this way, human resources employed by the subcontractor are modeled similarly to those within the main contractor,

while maintaining a clear distinction from a contractual and organizational perspective.

Each subcontractor present on site is therefore represented by a dedicated *IfcSubContractResource* instance, to which the relative labor resources are linked through corresponding *IfcLaborResource* entities. This structure enables a clear distinction between the different companies involved, facilitating financial reporting as well as control, verification, and comparison of what is contractually agreed versus what is actually performed.
All resources employed in a construction site activity are, in addition to being linked to the relevant reference document or drawing, directly associated with the operational task to which they contribute. This approach ensures full traceability between planning, documentation, and the activities actually carried out on site.

IfcConstructionEquipmentResource

In most construction site daily reports, the equipment used to perform the work is also typically recorded. The IFC standard defines the *IfcConstructionEquipmentResource* entity to represent each machine or tool on site. Similar to labor, equipment usage can be tracked through the *Pset_ConstructionResource* property set, using *ActualWorkTime* and an *IfcIrregularTimeSeries* instance. This approach enables the documentation of actual usage time, links it to site reports, and associates it with the relevant operational activities.

Relationship between IfcResource and IfcTask

Within the IFC model, the *IfcTask* entity represents a project work unit and serves as the fundamental element for activity management. It can describe both high-level activities and more detailed operations, thanks to the possibility of instantiating additional *IfcTask* entities within it, thereby enabling a hierarchical structure of the work program (work breakdown structure). The association between activities and the resources employed is established through the *IfcRelAssignsToResource* relationship, which allows different resource types, such as labor, materials, and equipment, to be linked to the corresponding activity. This relationship makes it possible to specify which resources are used for a given task, facilitating the control of costs, quantities, and operational utilization. Moreover, *IfcTask* allows the recording of temporal parameters through the *IfcTaskTime* object, which collects all information related to the time dimension of the task. In particular, it is possible to define the scheduled start and finish dates and times of the activity (*ScheduleStart* and *ScheduleFinish*), as well as the corresponding actual values (*ActualStart* and *ActualFinish*), enabling a comparison between planned and actual work progress. These temporal values can also be determined and updated based on the time and availability information associated with the resources involved (Figure 2).

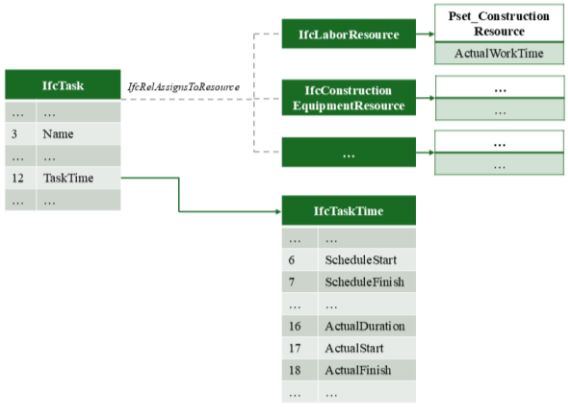


Figure 2: Relationship schema between IfcTask project activities and resource allocation

Relationship between IfcDocumentInformation and IfcTask

The above discussion addresses the organization and management of information contained within a single document. However, in the real context of a construction site, the situation is generally much more complex: numerous activities must be carried out over time, and the resources involved frequently interact, participating in multiple processes and operational phases. Consequently, it is essential to understand how to correctly manage the links between activities, resources, and documentation in such a complex and dynamic environment.

Specifically, regarding labor reporting, an operational document such as a site report may include information relating to work performed over several days. This information is therefore distributed across multiple separate documents, each corresponding to a specific date or period. In this scenario, the same *IfcTask* entity, representing a specific activity or process, can be linked to multiple documents, each documenting a portion of its progress or a particular time phase. Similarly, the same *IfcLaborResource* entity, representing a human resource, may be associated with multiple documents, as the same resource may contribute to different activities over multiple days (Figure 3).

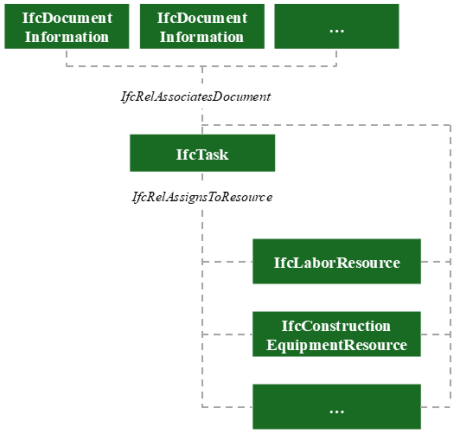


Figure 3: IfcLaborResource entity associated with multiple documents

The opposite situation can also occur: multiple activities are performed on the same working day and are consequently documented within a single reporting document. In this case, the single document is linked to multiple *IfcTask* entities, each representing a distinct task carried out during the same period (Figure 4). This configuration highlights a many-to-many relationship between documents and activities, which must be carefully managed to ensure the consistency and completeness of the information.

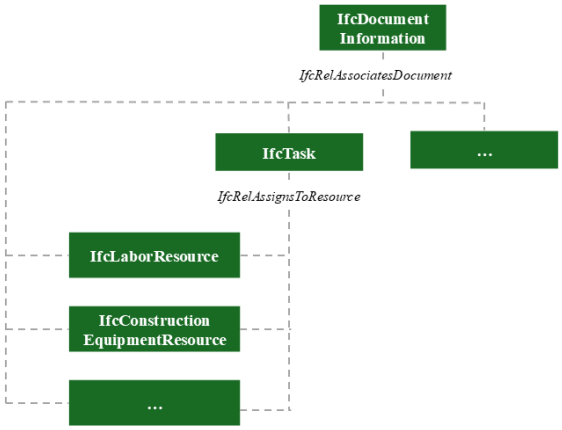


Figure 4: IfcDocumentInformation entity associated with multiple tasks

Results

This section presents the results of the study and provides a technical validation of the framework through a Proof-of-Concept (PoC), showing how the IFC standard enables the integration of different document-based data into a unified information structure. It was evaluated to verify the consistency and completeness of the proposed data mapping. In particular, the validation focused on the ability to correctly represent the relationships between documents, activities, and resources, as well as on the completeness of the mapping of information derived from the documents into IFC entities. It is assumed that two daily reports are available containing the following information. Table 1 presents the data for Daily Report No. 37, while Table 2 shows the data for Daily Report No. 38.

Table 1: Information contained in Daily Report No. 37

Task	Labor Resource hour	IfcConstruction EquipmentResource hour
Placement of formwork for first-floor columns	General construction worker (4 h)	Cold rental of electric saw (4 h)
Laying of bricks for construction of first-floor partition walls	Skilled construction worker (8 h)	Hoist rental with operator (8 h)

Table 2: Information contained in Daily Report No. 38

Task	Labor Resource hour	IfcConstruction EquipmentResource hour
Reinforcement for first-floor columns	General construction worker (6 h) Skilled construction worker (8 h)	wet rental of mobile crane (8 h)
Laying of bricks for construction of first-floor partition walls	Skilled construction worker (8 h)	Hoist rental with operator (8 h)

The figure available at the following link (Mapping) illustrates the resulting IFC schema after the mapping process. Each document is linked to the corresponding resources and tasks, which represent the activities carried out on site. A task is shared by both reports since it is carried out over both working days; therefore, the *IfcTask* entity is not duplicated, but linked twice. In this scenario, the information related to the resources connected to the task pertains to both document 37 and document 38. The resources involved are *IfcLaborResource* and *IfcConstructionEquipmentResource*. In both time series within the *Pset_ConstructionResource*, two values are recorded:

- In orange, the data from document 37 is shown, indicating 8 hours of work on 2025-09-10, with 8 hours of equipment usage.
- In purple, the data from document 38 is shown, indicating 8 hours of work on 2025-09-11, with 8 hours of equipment usage.

When the first *IfcIrregularTimeSeriesValue* for a specific activity is recorded, it corresponds to the *ActualStart* attribute of *IfcTaskTime* (number 17). If the activity is completed, the end date is recorded in the *ActualFinish* attribute (number 18). This allows a later comparison between planned and actual progress. In this case, the activity is assumed not to be completed yet, so the end parameter is not filled in. These temporal data can be compared with the project data contained in *IfcTaskTime* to monitor the progress of the activity. Additionally, *ActualStart* and *ActualFinish* can also be recorded for individual resources via *IfcResourceTime*, allowing the monitoring of each resource's specific duration.

In cases where different types of labor are used for the same activity, for example skilled workers and general laborers, two *IfcLaborResource* entities are created, one for each qualification, to which the worked hours are then assigned. This approach allows for more accurate resource management and ensures a consistent allocation of hourly costs at a later stage. In the upper part of the figure, this scenario can be observed. For the task "Reinforcement for first-floor columns", two *IfcLaborResource* entities are created, each containing the hours worked by the corresponding labor group.

Discussions

Based on the previously identified IFC classes and explained hierarchies, it is possible to develop an operational workflow for data integration. Although this research does not focus on tools for automatic data collection from documents, the framework serves as a technical proof of concept for automated BIM-based project control. The possible stages are: Data digitization (key information is extracted from documents); Mapping to IFC; Output for management control (once populated, the IFC model functions as a database for monitoring actual costs). This paper proposes a computational framework for integrating construction site documentation into BIM models based on the IFC standard, with particular emphasis on labor management. The current results should therefore be interpreted as a proof of concept rather than a comprehensive empirical validation, highlighting the need for further testing on real-world projects. This approach enables a structured mapping between unstructured documents and IFC entities, ensuring consistency in data representation. Although the framework does not currently rely on machine learning techniques, it provides a robust semantic structure that addresses the fragmented nature of traditional reporting.

Construction site documentation collected during operational activities, particularly daily reports, is often characterized by unstructured or manually compiled documents. The framework addresses the unstructured nature of construction site documentation, such as daily reports, by enabling their integration into a standardized BIM environment. Documents remain stored externally (e.g., within Common Data Environments), while relevant information is structured within the IFC model and linked through references. This approach supports interoperability across different information systems and maintains synchronization between BIM models and external data sources. Another significant outcome of this work concerns the management of labor and hours worked within the IFC model, closely aligned with the document and operational management structure described above. Human resources are modeled using the *IfcLaborResource* entity, while the recording and calculation of hours worked are managed through the Property Set *Pset_ConstructionResource*, in combination with the use of *IfcIrregularTimeSeries*. This configuration enables the representation of discontinuous and variable time trends, which are typical of construction site activities, while maintaining a consistent and formalized data structure for monitoring labor hours and costs. Once the actual hours are recorded in the *ActualWorkTime* parameter through the time series, a cost calculation system can be established based on the assignment of specific hourly rates for each resource, derived from company price lists or subcontracting contracts. This allows for the determination of variable production costs in real time by multiplying the unit cost by the hours worked directly within the IFC model. Such an approach overcomes the limitations of manual and fragmented reporting typical of traditional tools, such as spreadsheets.

At the same time, the association between *IfcTask* and *IfcTaskTime* enables a direct link to be established between planned activities and temporal data, facilitating integration with planning and scheduling systems. This connection allows the IFC model to be synchronized with external work planning tools, enabling a comparison between the theoretical schedule, the actual hours worked, and the site documentation, which remains archived in external systems via IFC references as described above.

The work presented shifts the focus from the automatic classification of documents to their informational integration into the BIM model. Unlike many existing approaches, which rely on text mining, machine learning, or natural language processing techniques applied to documents, the proposed methodology emphasizes the importance of mapping data onto standardized entities and relationships.

In this context, the theoretical contribution of the study lies in formalizing a semantic correspondence between construction site documentation and the IFC schema, explicitly linking documents, activities, resources, and time. The IFC model is thus not used solely as a geometric exchange format, but as a computational infrastructure for the management of process-related data.

Thanks to this integration, with all information contained within a single data model, it becomes possible to monitor construction site progress in real time, not only in terms of activity scheduling but also regarding the resources used and the associated costs. Although the research does not focus specifically on labor resource documents, the framework is applicable to any type of documentation, thereby providing a comprehensive and consistent view of the construction site, capable of supporting informed decision-making and integrated management strategies.

Furthermore, the centralized management of documents within the IFC model helps prevent the loss of critical information, which is typically associated with paper-based records or unstructured archives, ensuring that all information resources are stored, accessible, and correctly linked to the operational entities of the construction site. This approach significantly enhances data traceability, security, and continuity, promoting more effective and reliable management of the entire construction process.

Mapping and associating documentation with resources and activities modelled in the BIM model allows for significantly more detailed management of construction site progress. Thanks to this integration, the company can constantly monitor the actual status of the work and compare actual data with that planned in the project, thus obtaining an accurate view of the correspondence between planning and operational progress. This level of detail is particularly useful for more accurate management control, as it allows for the timely identification of any deviations from plans and targeted intervention to correct delays, optimise the use of resources or rebalance workloads. The methodology also provides a clear view of the costs incurred and resources used, not only at a global level, but also for each individual activity or planned work phase. Thanks to this granularity, it is

possible to carry out specific and targeted analyses, for example by evaluating the productivity of a team, comparing the hours worked with those scheduled, or analysing the impact of operational choices on overall costs. In this way, the integrated BIM model becomes a powerful decision-making tool, capable of supporting the strategic and operational management of the construction site, increasing transparency, reducing margins of error, and promoting a proactive approach to planning and controlling activities.

The proposed method prioritizes semantic structuring over automated extraction, distinguishing it from ML and NLP approaches. While these techniques efficiently process large amounts of unstructured data, they may struggle to ensure consistent and standardized outputs. In contrast, the IFC-based approach enhances semantic consistency and interoperability but requires more structured input and greater manual effort. For this reason, the two approaches should be seen as complementary rather than competing. Future research should focus on integrating their respective strengths.

Although the framework demonstrates the feasibility of integrating site documentation into the IFC schema, further research is needed to validate its robustness, scalability, and applicability in different contexts. Future studies should assess data completeness and consistency, as well as their effectiveness in supporting cost control and progress monitoring. In addition, integrating NLP techniques could enable automated extraction of information from unstructured documents. Finally, further developments should explore interoperability with external systems (e.g., CDEs, ERPs, planning tools) and extend the framework to include additional construction data, such as materials, equipment, and subcontracting.

Moreover, the integration of real-time data sources, such as IoT sensors and tracking systems, offers significant potential to develop dynamic, continuously updated BIM-based control systems aligned with actual site conditions.

Conclusions

This research defines a conceptual framework for document management and the integration of data derived from documentation within BIM models, in accordance with the IFC standard. The research demonstrates how IFC entities can act as structured data containers to support management control. The main objective is to move beyond the traditional approach based on mere document archiving, instead enhancing their value as a structured source of information useful for monitoring project performance. Analysis of the construction site documentation led to the identification of the relevant IFC classes, highlighting how the standard can operate as an environment for collecting, organizing and correlating information relating to resources, activities and time.

In particular, the combined use of the *IfcDocumentInformation*, *IfcTask*, and *IfcLaborResource* entities, together with *IfcIrregularTimeSeries*, enables a consistent representation of labor reporting while maintaining a direct link to the original documents stored in the Common Data Environment. These entities are complemented by additional resource classes dedicated to the management of materials, rentals, and subcontracts, such as *IfcConstructionMaterialResource*, *IfcConstructionProductResource*, *IfcConstructionEquipmentResource*, and *IfcSubContractResource*.

Within the context of this research, the IFC standard is not employed solely as a tool for geometric representation, but as a genuine information database for the management of construction processes. Although the proposed framework is demonstrated using labor as an example, it can be extended to other main categories of direct costs including materials, rentals, and subcontracts by applying the same principles of classification, mapping, and document association. This makes the approach suitable for integrated and comprehensive construction site management. Standardization of data also enhances information traceability and enables systematic comparisons between planned and actual outcomes, supporting more timely and accurate management control. While the proposed framework successfully maps construction site data into the IFC4x3 schema, it faces challenges regarding data stability and interoperability. Native GUIDs and custom properties are often subject to changes during model re-exports, potentially breaking semantic links. To mitigate this risk, it is advisable to use stable, user-defined identifiers, such as WBS codes or persistent project parameters. Furthermore, although this paper advocates for a rigid semantic structure to ensure the high data fidelity required for cost control, this inherently entails a high level of manual intervention and mapping effort. The first area for future research concerns the empirical validation of the framework through real case studies. Applying the methodology to projects of varying scale and complexity would allow for an assessment of the robustness of the proposed IFC mapping, as well as the completeness and reliability of the data supporting management control. At the same time, extending the approach to different types of documents could reveal critical issues related to format variability and the diverse operational practices adopted by companies. To reduce the burden of manual input, future developments may focus on the identification of automated tools for data extraction and their correct assignment to the corresponding IFC entities. Integrating ML and NLP algorithms could combine automation in data extraction with the semantic precision of IFC, merging operational efficiency and structured control. In this context, methods for verifying the accuracy of information mapping should be investigated. The adoption of models for text and document structure recognition could further reduce manual intervention and minimize the risk of errors. Finally, another area of development concerns integration with IoT technologies, allowing the inclusion of information directly obtained from the field and not contained in static documents.

References

Avogaro, D., Berlato, M., Zanchetta, C., 2024. Implementation of IFC-based process models for collaborative management of temporal scheduling through Linked Data.

Avogaro, D., Zanchetta, C., 2025. IFC and Project Control: A Systematic Literature Review. Buildings 16, 91. https://doi.org/10.3390/buildings16010091

Dave, B., Kubler, S., Främling, K., Koskela, L., 2016. Opportunities for enhanced lean construction management using Internet of Things standards. Autom. Constr. 61, 86–97. https://doi.org/10.1016/j.autcon.2015.10.009

Deng, R., Wang, K., Mao, Y., 2025. Real-Time Monitoring of Mobile Construction Resources Based on Multiple Object Tracking. J. Comput. Civ. Eng. 39, 04025059. https://doi.org/10.1061/JCCEE5.CPENG-6587

Du, S., Hou, L., Zhang, G., Tan, Y., Mao, P., 2024. BIM and IFC Data Readiness for AI Integration in the Construction Industry: A Review Approach. Buildings 14, 3305. https://doi.org/10.3390/buildings14103305

EN ISO 16739-1:2024. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries - Part 1: Data schema, 2024.

Gehring, M., Wala, J., Rüppel, U., 2025. IoT-altimeter in smart pallets for material tracking on multi-storey construction sites. Smart Constr. https://doi.org/10.55092/sc20250023

IFC4.3.2.0 Documentation [WWW Document], 2024. URL https://standards.buildingsmart.org/IFC/RELEASE/IFC4_3/ (accessed 1.28.26).

Jafari, P., Al Hattab, M., Mohamed, E., AbouRizk, S., 2021. Automated Extraction and Time-Cost Prediction of Contractual Reporting Requirements in Construction Using Natural Language Processing and Simulation. Appl. Sci. 11, 6188. https://doi.org/10.3390/app11136188

Jain, S., Sanchez, G., Sunil, T., Sharma, D.K., 2024. Radio frequency identification based materials tracking system for construction industry. Int. J. Reconfigurable Embed. Syst. IJRES 13, 85. https://doi.org/10.11591/ijres.v13.i1.pp85-95

Kasim, N., Kusumaningtias, R., Department of Accounting, Faculty of Economics,Universitas Negeri Surabaya, Jalan Ketintang, Kampus UNESA, Surabaya, 60231, MALAYSIA, Sarpin, N., Department of Construction Management, Faculty of Technology Management & Business, Universiti Tun Hussein Onn Malaysia,Parit Raja, Batu Pahat, 86400, MALAYSIA, 2019. Enhancing Material Tracking Practices of Material Management inConstruction Project. Int. J. Sustain. Constr. Eng. Technol. 10. https://doi.org/10.30880/ijscet.2019.10.02.006

Kerosuo, H., Miettinen, R., Paavola, S., Mäki, T., Korpela, J., 2015. Challenges of the expansive use of Building Information Modeling (BIM) in construction projects. Production 25, 289–297. https://doi.org/10.1590/0103-6513.106512

Kone, V., Mahesh, G., 2025. An ontology-driven bi-directional workflow for integrating project management data into the IFC standard. J. Inf. Technol. Constr. 30, 1768. https://doi.org/10.36680/j.itcon.2025.073

Mahire, S., Balaji Saravanan, U.K., M, M., N, P., 2024. A Comprehensive Study on Advancements in Text Mining and Natural Language Processing, in: 2024 International Conference on Emerging Research in Computational Science (ICERCS). Presented at the 2024 International Conference on Emerging Research in Computational Science (ICERCS), pp. 1–7. https://doi.org/10.1109/ICERCS63125.2024.10895330

Mastino, C.C., Vaičiūnas, J., Possidente, R., Frattolillo, A., Zavari, M., Da Pos, V., 2026. A Digital Twins Model Based on IFC Open BIM Models Managed on Web Platforms, in: Jurelionis, A., Fokaides, P.A., Mazzarella, L., Hartmann, T. (Eds.), Building Digital Twins, Lecture Notes in Civil Engineering. Springer Nature Switzerland, Cham, pp. 26–39. https://doi.org/10.1007/978-3-032-09040-9_3

Mohammed, M.N., Lalmi, R., Taher, M., Jweeg, M., Aljibori, H.S.S., 2024. Assessing and Identifying On-Site Labor Cost Challenges and Necessary Control Measures in Iraqi Construction Projects: A Case Study Approach, in: Hamdan, A., Harraf, A. (Eds.), Business Development via AI and Digitalization, Studies in Systems, Decision and Control. Springer Nature Switzerland, Cham, pp. 787–796. https://doi.org/10.1007/978-3-031-62102-4_66

Noardo, F., Arroyo Ohori, K., Krijnen, T., Stoter, J., 2021. An Inspection of IFC Models from Practice. Appl. Sci. 11, 2232. https://doi.org/10.3390/app11052232

Pauwels, P., Zhang, S., Lee, Y.-C., 2017. Semantic web technologies in AEC industry: A literature overview. Autom. Constr. 73, 145–165. https://doi.org/10.1016/j.autcon.2016.10.003

Singh, A.K., Pal, A., Kumar, P., Lin, J.J., Hsieh, S.-H., 2023. Prospects of Integrating BIM and NLP for Automatic Construction Schedule Management. Presented at the 40th International Symposium on Automation and Robotics in Construction, Chennai, India. https://doi.org/10.22260/ISARC2023/0034

Smith, P., 2014. BIM & the 5D Project Cost Manager. Procedia - Soc. Behav. Sci. 119, 475–484. https://doi.org/10.1016/j.sbspro.2014.03.053

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

BIM-BASED DESIGN CHANGE FRAMEWORK FOR ASSESSING MEP SYSTEM STABILISATION

Mawara Khan[1], Muhammad Salman Javaid[1], Eelon Lappalainen[1], Hisham Abou Ibrahim[2], Olli Seppänen[1]
[1]Aalto University, Espoo, Finland
[2]American University of Beirut, Lebanon

Abstract

Design changes during construction remain a challenge, as late-stage modifications to complex MEP systems impact cost and schedule. While BIM enables documentation of design evolution through versioned IFC files, existing tools focus on retrospective change detection rather than forward-looking stability assessment. This study introduces two metrics, Change Intensity (CI) and Stability Index (SI), to track MEP design evolution across IFC versions. CI aggregates the proportion of elements changed, while SI applies a logistic function to translate CI into an interpretable stability score. Eleven IFCs from a real project demonstrate the framework's ability to identify disruption events and track design convergence.

Introduction

Design instability during the construction phase remains a critical challenge in construction project management. Studies have shown that site managers spend more time due to design modifications and rework. This extra time can result in up to 5 to 40% cost overrun (Aslam et al., 2019). Notably, changes made to MEP systems near the completion of the project can be costly, as often these systems must be rerouted around previously installed structural or architectural components.

Last-minute modifications, particularly in MEP design, are affected by numerous circumstances. The prevailing belief among contractors is that MEP design can be modified nearly up to the commencement of installation without substantial detriment to the installation process (Valkonen et al., 2025). The selected installation method significantly impacts this; the typically low level of prefabrication allows contractors to alter plans nearly up to the commencement of installation (Valkonen et al., 2025). The late involvement of the MEP contractor, often used in many countries for business and legal reasons, creates pressure to make changes after the MEP contract has been secured (Valkonen et al., 2025). One of the main causes for delayed design changes is the postponement in approval of changes to the architectural design and its subsequent effect on the MEP design, which follows later in the sequence (Tillmann, 2020).

Leveraging BIM now allows detailed documentation of the evolution of the design through versioned modelling files. Additionally, several commercial software packages can assist users in identifying the changes that occurred between model versions through comparison of either element property or geometry (Bockstael and Issa, 2016). These capabilities provide transparency into how a design has evolved over time.

However, for design managers, simply knowing the number of changes does not provide a reliable and actionable insight. The conventional method of tracking changes and reacting to them does not inherently yield enough reliable and useful insights (Juszczyk et al., 2016). There is a need to understand whether the design is converging toward stability or if there are more impactful changes still expected. Scholars have suggested that if the design change management practices are improved, including predicting or forecasting change impacts, it can enhance overall project performance, reduce negative impact on cost and schedule, and help in coordination decisions (Wang et al., 2022, Motawa et al., 2007).

Despite recent advances in BIM-based change management, a gap remains in that the existing BIM comparison tools detect what changed between model versions, but are limited in assessing design maturity or distinguishing meaningful design changes from metadata updates (Lin et al., 2019). Furthermore, only a limited number of metrics exist to quantify the rate at which a design approaches stability or when convergence will occur.

The current research addresses this gap by proposing two complementary metrics, which are Change Intensity (CI) and Stability Index (SI), to track MEP design evolution across consecutive IFC versions. CI measures the percentage of elements added, removed, or changed from each updated version of a design, and then uses that information to apply a logistic function and compute an interpretable stability score (SI). The resulting CI/SI framework, combined with smoothing and trend analysis techniques, allows the management to analyse the historical patterns of design evolution as well as make predictions about the future stability of designs going forward.

The objective of this study is to develop a simple framework that tracks MEP design evolution across IFC versions and provides interpretable metrics for assessing design stabilisation, enabling project teams to make informed decisions about coordination timing and construction readiness.

Literature review

The detection of design changes within BIM has evolved significantly over the past decade. Lee et al. (2011) developed foundational metrics to quantify the similarities and differences between IFC files. Their research proposed five key metrics, which are similarity rate, matching rate, globally unique identifier (GUID) preservation rate, missing rate, and addition rate. They performed the analysis on 88 IFC files and demonstrated the potential for automated comparison methods in tracking design evolution across project versions. Building upon this foundation, various commercial tools provided clash detection and model coordination mechanisms (Bockstael and Issa, 2016). However, the focus of these tools is primarily on identifying what changed between model versions, and they do not assess design progression or predict stabilisation patterns (Abdulfattah et al., 2023).

Concurrently, the concept of digital twins has advanced from physical asset monitoring to evolve design process management. Boje et al. (2020) proposed directions for the semantic construction of digital twins that emphasises on the real-time data integration throughout the building lifecycle. To complement the design-based data integration, Dai et al. (2024) proposed digital twin modelling methods, which are based on IFC standards. In addition, the integration of BIM with lean principles has further demonstrated benefits, including waste reduction and enhanced visualisation (Sacks et al., 2010). However, these approaches typically address process management rather than quantitative design stability assessment.

Despite these advances, a critical gap persists in the literature. The existing methods focus on retrospective change identification rather than forward-looking stability assessment. For example, Abdulfattah et al. (2023) concluded that no established framework exists for quantifying design maturity progression across multiple model versions. Furthermore, the current comparison methods do not adequately distinguish between meaningful design changes, as indicated by Lee et al. (2011) who acknowledged that automated identification of modified versus newly generated information remains an unsolved challenge. Moreover, the methods for predicting design stabilisation timing are notably absent. While Saoud et al. (2017) proposed Design Structure Matrix approaches for change propagation prediction and subsequently, Chen and Whyte (2022) developed digital-twin-driven DSM methods; these approaches address change impact propagation rather than overall design convergence assessment.

More recently, researchers have explored approaches to managing time-evolving BIM data at the object level. Esser et al. (2022) proposed a graph-based version control method that represents IFC object networks as formal property graphs and describes model changes through graph transformations, enabling object-level synchronization rather than whole-file comparison. Their subsequent work extended this to event-driven collaboration environments for asynchronous BIM workflows (Esser et al., 2023). While these contributions advance the technical infrastructure for detecting granular changes between model versions, they focus on the mechanics of change detection rather than interpreting change patterns in terms of design stability. The present study complements this work by providing interpretable metrics that translate detected changes into actionable stability assessments.

Recent studies emphasize that artificial intelligence applications in construction should support practical decision-making (Pan and Zhang, 2021), yet current implementations prioritise technical sophistication over accessibility. This gap indicates an opportunity for frameworks that could convert the complex change data into actionable metrics for design managers and project stakeholders.

To address these gaps, this study introduces a framework comprising two complementary metrics, which are CI and SI. These metrics are designed to track MEP design evolution across consecutive IFC versions. The following section details the methodology for calculating these metrics and their application in assessing design maturity progression.

Methodology

The research was conducted using the Design Science Research Methodology as proposed by Peffers et al. (2007). The research process is structured into six sequential activities: problem identification, definition of objectives, artefact development, demonstration, evaluation, and communication, as shown in Figure 1.

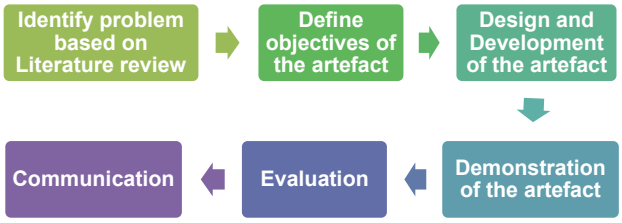


Figure 1. Research Method

The research addresses the problem that the existing BIM model comparison tools detect changes but do not assess whether the design is stabilising over time or not. To address this gap, a BIM-based analytical artefact was developed to quantify design change behaviour and assess stabilisation trends across model versions.

The artefact was iteratively developed and refined using a case study consisting of eleven consecutive IFC versions of a sprinkler system spanning approximately five months of design development. The sprinkler system was chosen because the changes made by architects and clients to the suspended ceiling and interior finishes cause design alterations in sprinkler systems, which make them one of the critical systems in MEP systems. This system is suitable for the demonstration of the problem. The Iterative refinement was guided by sensitivity analysis and empirical behaviour observed in the case study, consistent with DSR principles of relevance, rigor, and utility.

A positive side of this research approach is that it does not require extensive data sets like complex machine learning approaches, which need extensive training data, the CI and SI metrics. The current approach offers immediate interpretability with simple computation. The logistic transformation used in SI calculation provides an intuitive scale that facilitates communication among project stakeholders and supports decision-making in construction management.

Change Intensity (CI)

Design changes between consecutive IFC files are quantified using a CI metric, which captures the proportion of elements affected by different types of changes. CI is defined as (1):

$$CI = A + R + G + P \quad (1)$$

Where:

A = Percentage of added elements

R = Percentage of removed elements

G = Percentage of elements with geometric changes

P = Percentage of elements with property changes

The formulation provides a normalised measure of design activity between versions. CI reflects how much the model has changed without yet interpreting whether those changes indicate convergence or instability.

Stability Index (SI)

To interpret CI in terms of design stabilisation, SI was introduced. The SI transforms CI values into an interpretable stability signal using a two-step process:

Step 1: Raw Stability Index (Logistic Mapping)

A logistic function is used to map CI values to a bounded stability scale [0, 100], shown as (2):

$$SI_{raw} = \frac{100}{1+e^{\frac{CI-\theta}{s}}} \quad (2)$$

Where:

θ is the stability threshold,

S controls the steepness of the transition around the threshold

The logistic function was selected because design stabilisation is non-linear, i.e., small increases in CI near the threshold can significantly reduce confidence, while very low CI values rapidly increase perceived stability. Earlier studies have used logistic growth models as part of forecasting model comparisons to evaluate trend-based cost predictions (Narbaev and De Marco, 2014a).

Step 2: Smoothed Stability Index (Moving Average)

Single version stability measurements may be sensitive to isolated events such as bulk geometry updates or metadata changes. To reduce this sensitivity and reveal underlying stabilisation trends, a moving average is applied to the raw SI values, shown in (3):

$$SI_{smoothed} = \frac{1}{n}\sum_{j=i-n+1}^{i} SI_{raw,j} \quad (3)$$

where n is the window size.

In this study, n-3, meaning that stability is evaluated based on the three most recent versions. The moving average smooths short-term fluctuations and supports the interpretation of sustained convergence behaviour, which is more relevant for design management decisions than single version stability. Multiple practical descriptions show moving averages are widely applied in fields like finance, forecasting, business, and economics (smoothing trends, forecasting) (Hansun, 2013, Hyndman, 2025).

Trend-Based Prediction Logic

Rather than predicting absolute future design stability, the artefact supports trend-based anticipation of stability behaviour. A simple first-order extrapolation (2) is used to estimate near-term change intensity:

$$CI_{predicted} = CI_i + (CI_i - CI_{i-1}) \quad (4)$$

The predicted CI is then transformed using the same logistic mapping to obtain a predicted raw stability value. Final anticipated stability is expressed as a moving average combining recent and predicted values, shown in (5):

$$SI_{predicted} = moving\ average\ of\ (SI_{raw,recent} + SI_{raw,predicted}) \quad (5)$$

This mechanism does not claim deterministic forecasting but provides early warning signals indicating whether the design process is converging, persisting, or diverging.

Artefact Iteration

The artefact parameters were iteratively calibrated through sensitivity analysis:

- For Threshold and Steepness, twelve parameter combinations were tested (θ = 5-20%, s = 2-5). The final values θ = 10% and s = 3 were selected as they provided the most interpretable separation between stable and unstable design phases
- For the moving average window, sizes of n = 2, 3, 4, 5 were evaluated. A three-version window was selected as it balanced responsiveness with robustness against isolated spikes.

These iterations ensured that the artefact remains sensitive to meaningful design changes while avoiding overreaction to short-term anomalies.

Demonstration

The artefact was demonstrated on a real-world case study involving a sprinkler system from a Finnish under-construction sports arena project. The gross area of the building was 15000 sqm with eight storeys. The dataset consisted of eleven consecutive IFC versions spanning approximately five months of design development. The initial version of the dataset contained approximately

15,800 MEP elements, enabling analysis of stabilisation over time.

Evaluation

To assess the extent to which the artefact met its intended purpose, an evaluation process was carried out. Evaluation focused on the ability to identify convergence and divergence trends against updated design changes, and the interpretability of CI and SI metrics for design management decisions. This paper details the evaluation results along with the final design of the artefact and how those results were used to provide a practical BIM-based tool for evaluating the stabilisation of the MEP design phase/preliminary designs using open IFC data.

While the by Peffers et al. (2007) framework guided the sequential process of this research, from problem identification through communication, Hevner et al. (2008) provide a complementary set of guidelines for evaluating the rigour and relevance of design science outputs. Table 1 maps these guidelines to the current study to demonstrate that the developed artefact satisfies established quality criteria for design science research.

Table 1. Mapping of DSR guidelines to the study

DSR Guideline	Application in this study
Guideline 1: Design as an Artifact	The CI/SI framework is a purposeful artifact developed to quantify and track BIM design stability.
Guideline 2: Problem Relevance	Addresses the critical gap where existing BIM tools lack stability assessment for project management.
Guideline 3: Design Evaluation	The framework was demonstrated and evaluated on 11 real-world IFC versions to ensure empirical validity.
Guideline 4: Research Contribution	Provides novel interpretable metrics (CI and SI) that offer a new way to measure and visualize design stabilization.
Guideline 5: Research Rigor	Rigor is maintained through sensitivity analysis and the iterative calibration of threshold and steepness parameters.
Guideline 6: Design as a Search Process	The "search" involved testing 12 parameter combinations and evaluating various moving average window sizes to find the optimal configuration.
Guideline 7: Communication of Research	Findings are presented to both technical and management audiences through the detailed documentation in this paper.

Results

Change Intensity across IFC versions

Change Intensity was calculated for ten consecutive version transitions from 1 March 8 March to 15 July 26 July by aggregating added, removed, geometry-changed, and property-changed elements as shown in Figure 2.

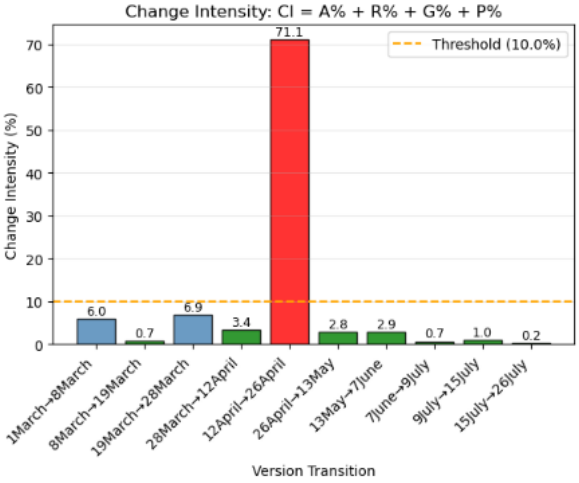


Figure 2. Change Intensity over revisions

A clear exception occurred during the 12 April → 26 April transition, where CI spiked sharply to approximately 71%. This spike was primarily driven by many elements undergoing property changes, indicating a bulk update event rather than gradual design development. Such behaviour reflects a disruptive intervention in the design process, for example, late requirement changes, system reconfiguration, or data restructuring. After this event of instability, CI values dropped rapidly and consistently, reaching near-zero levels by 15 July → 26 July, signalling a return to stable and controlled design changes.

Stability Index Behaviour

The raw SI, computed using the logistic function, responded sensitively to CI fluctuations as shown in Figure 3.

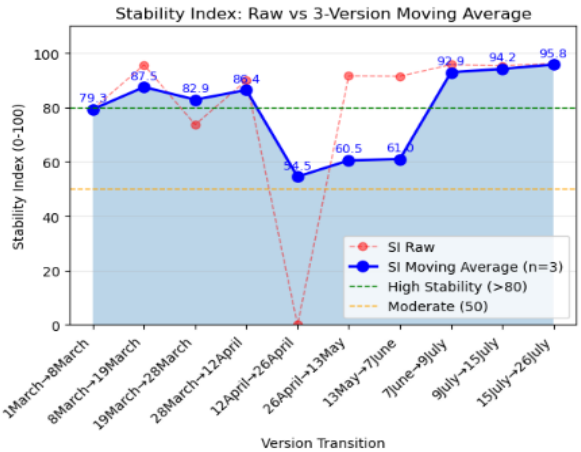


Figure 3. Stability Index

During early revisions, raw SI values remained mostly in the high-stability range (>80), reflecting low change intensity and increasing design stability. On 12 April 26 April, the extreme CI spike caused the raw SI to collapse to near zero, demonstrating the model's ability to

correctly penalise disruptive design behaviour. However, this sharp drop also illustrates the volatility of raw SI when isolated bulk changes occur.

To address this, a 3-versions moving average was applied. The smoothened SI successfully dampened abrupt fluctuations while preserving the overall trend. Following the disturbance on 12 April → 26 April, the smoothed SI gradually recovered, crossing back into the high-stability zone and reaching 95.8 on 15 July → 26 July. This progression reflects a transition from unstable redesign back to controlled convergence.

Stability Trend Prediction and Evaluation

As described earlier, there were eleven versions of IFCs that were used in the demonstration. Nine versions were used to calculate the stability trends, and the last two were used for the prediction of the stability index. Figure 4 indicates the trend analysis, which shows a positive SI trend (+1.6) combined with a negative CI trend (−0.7%) in the most recent versions. Based on linear extrapolation of CI and the smoothed SI trajectory, the predicted next version indicates continued stability.

Evaluating the trends based on the results shown in Table 2, it was noticed that the predicted stability points align with the raw/actual stability value of the last two versions, i.e., 9 Jul → 15 Jul difference is 1.1, and 15 Jul → 26 Jul difference is 0.5. The predicted SI for the last two versions remains within the high-stability regime, corresponding to an estimated ~80% stability confidence.

The results show different phases of design evolution that validate the framework's effectiveness. During initial revisions from 1 March to 12 April, CI values remained between 0.73% and 6.92% with raw SI exceeding 73%, which indicates controlled incremental development consistent. The versions from 12 April to 26 April represented a significant disruption event, with CI spiking to 71.06% and raw SI collapsing to zero, which was caused by bulk property updates or late requirement changes rather than gradual evolution. The 3-version moving average effectively reduced this instability, dropping gradually to 54.5 while maintaining trend visibility. After this abrupt disturbance, the design showed gradual stabilisation. CI decreased constantly from 2.84% to 0.23%, while smooth SI recovered from 60.5 to 95.8 by the final alteration. The positive SI trend (+1.6) combined with the negative CI trend (−0.7%) in recent versions demonstrates the framework's capacity to differentiate between normal iteration, disruptive interventions, and stabilisation patterns.

Table 2. Results of Raw and Predicted Stability Index

Version	CI (%)	SI_raw	SI_avg	Window
1 Mar → 8 Mar	5.96	79.3	79.3	[1..1]
8 Mar → 19 Mar	0.73	95.6	87.5	[1..2]
19 Mar → 28 Mar	6.92	73.6	82.9	[1..3]
28 Mar → 12 Apr	3.43	89.9	86.4	[2..4]
12 Apr → 26 Apr	71.06	0.0	54.5	[3..5]
26 Apr → 13 May	2.84	91.6	60.5	[4..6]
13 May → 7 Jun	2.88	91.5	61.0	[5..7]
7 Jun → 9 Jul	0.66	95.8	92.9	[6..8]
9 Jul → 15 Jul	0.96	95.3	94.2	[7..9]
15 Jul → 26 Jul	0.23	96.3	95.8	[8..10]

Discussion

The proposed research framework builds upon existing literature in the collaboration and coordination of BIM-based projects by integrating temporal analysis of multiple sequential versions of an IFC, rather than a single-instance comparison. While the commercial tools used to identify conflicts and clashes between elements of a built project are well known (e.g., Hasannejad et al. (2023)), the question of whether the design is stabilising or not has not been addressed within the academic literature. This distinction could influence decisions made by design managers regarding project coordination, timeliness of projects, and construction readiness. To

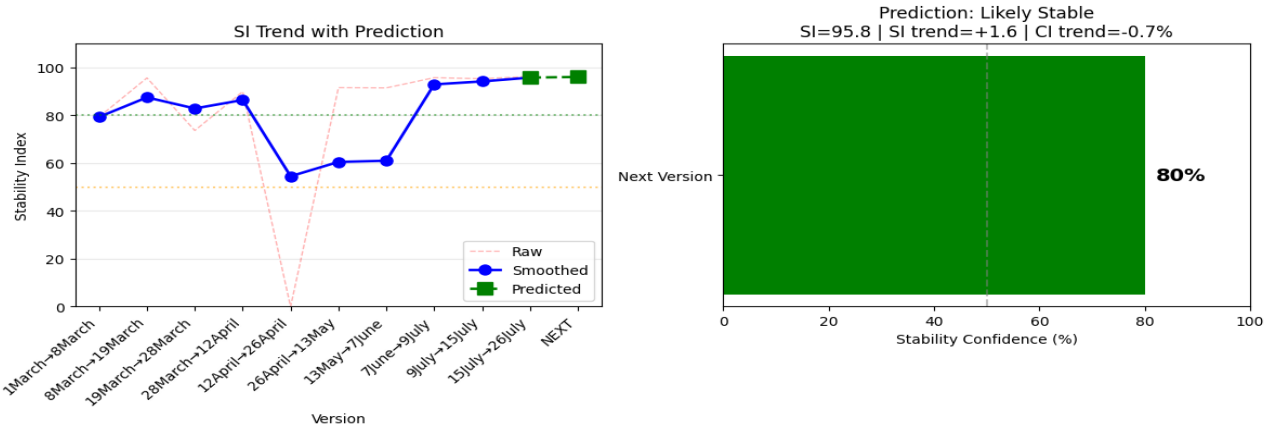


Figure 4. Stability Index trend with prediction

address this gap in the building and construction industry, 2 interpretable project metrics, CI and SI, are proposed with a demonstration of the framework.

The CI and SI metrics have been developed to be easily understood by project managers without requiring a statistical background to understand how the results are represented. For example, if a project manager sees a CI value of 15%, they know 15% of elements have changed; or if they see an SI value of 85%, they know that all project elements are in a stable condition. This transparency is not merely convenient, but it is essential for adoption in practice, where stakeholders need to trust and understand the basis for recommendations. The proposed metrics also do not rely on complex machine learning algorithms (Abdulfattah et al., 2023, Pan and Zhang, 2021) to determine how an element will behave in the future based on its previous history; therefore, no statistical expertise is required to understand the CI and SI metrics.

The study supports the existing research by extending element-based two-version IFCs comparison to an analysis of a series of consecutive IFC versions. Most of the change detection studies only catalogue differences between two versions (Moayeri et al., 2015), which provides valuable data and information, but provides limited information to reveal patterns of design evolution over time. This research demonstrates through tracking CI and SI over 11 versions of IFCs that the design process follows recognisable patterns of design. These patterns include controlled incremental refinement during initial phases, possible disruption events that produce temporary instability, and eventual convergence towards stability as a design develops.

This finding of this research aligns with the established understanding of iterative design processes in BIM-enabled projects, where coordination activities gradually resolve conflicts until the design reaches a sufficient level of maturity for execution (Ahmad et al., 2017). In complex project management, coordination noise is experienced during an iterative information flow. To address this noise, moving average filters are utilised. This approach is complemented by logistic growth models, which have been validated in previous literature for forecasting the progression of design stability (Largani et al., 2026, Narbaev and De Marco, 2014b). This prediction mechanism provides a potential solution for the challenges in construction readiness due to design changes, mentioned in recent researches (Radzi et al., 2025).

Limitations and Future Research

This study is based on one project only and focuses on fire sprinkler systems, which limits the generalizability of the results. The selected thresholds and weighting factors, while suitable and effective in this case, may require calibration for different project types and levels of model granularity. Additionally, the current CI formulation treats all geometric changes and all property changes as equally weighted. In practice, repositioning an element carries different design implications than modifying a pipe cross-section, yet both are captured identically as geometric changes. Similarly, updating a critical performance parameter such as a flow rate differs in significance from a metadata update such as a naming convention change.

To address these limitations, future studies can investigate weighted sub-categorisation of change types based on their functional relevance to the design intent. Furthermore, future research can apply the framework across multiple projects and building systems to validate its robustness and explore adaptive or learning-based threshold selection

Practical Implications

The framework provides a practical way for design managers to optimise coordination, timing, and resource allocation. A construction site with complete situation awareness of upcoming design changes can perform better than traditional sites in terms of rework reduction and cost overrun. The 10% CI threshold provides a simple decision rule, which is that if the transitions exceed this threshold warrant investigation into root causes and potential downstream impacts. This alertness helps the site manager to stay aware of the design change situation. The smoothed SI trend enables actionable scheduling of design reviews, and positive trends suggest approaching stability. While negative trends indicate continued volatility, requiring additional coordination meetings to stabilise the design. For this case project, the framework would have flagged the April disruption event in real-time and enabled the managers to inspect the cause and assess the impact before implementation. This approach aligns with research demonstrating that early conflict identification through BIM reduces costly on-site modifications (Khanzode et al., 2008). The prediction capability mentioned in the results of the case study estimates 80% stability confidence for subsequent versions. This stability confidence has the potential to support decisions regarding construction release timing and trade coordination scheduling.

Conclusions

This study presented a BIM-based framework for tracking MEP design development leveraging CI and SI metrics. CI quantifies element-level modifications across consecutive IFC versions, while SI applies a logistic transformation to produce a bounded stability score from 0 to 100%. The demonstration of eleven IFC versions from a real project demonstrated that the framework successfully identifies disruption events, tracks design convergence, and supports trend-based prediction of stabilisation. The approach addresses a critical gap in existing tools, which focus on retrospective clash

detection rather than forward-looking stability assessment, providing project managers with interpretable metrics for coordination timing and construction readiness decisions. Proposed future work can extend validation across multiple projects and building systems, integrate with scheduling tools, and explore adaptive thresholds for different project contexts.

Acknowledgement and Fundings

This research was conducted as part of the Research to Business (R2B) project BIMSpect. The authors gratefully acknowledge the financial support provided by Aalto University and Business Finland, which made this work possible. The authors also wish to thank all members of this R2B project, including industry piloting partners, practitioners, and team members, for their valuable contributions and collaboration. This work was partially supported by the Research Council of Finland, grant number 357307.

References

Abdulfattah, B. S., Abdelsalam, H. A., Abdelsalam, M., Bolpagni, M., Thurairajah, N., Perez, L. F. & Butt, T. E. 2023. Predicting implications of design changes in BIM-based construction projects through machine learning. Automation in Construction, 155, 105057.

Ahmad, T., Aibinu, A. & Thaheem, M. J. 2017. BIM-based iterative tool for sustainable building design: A conceptual framework. Procedia engineering, 180, 782–792.

Aslam, M., Baffoe-Twum, E. & Saleem, F. 2019. Design changes in construction projects causes and impact on the cost. Civil Engineering Journal, 5, 1647–1655.

Bockstael, D. & Issa, M. H. 2016. A methodology for contractor clash detection using building information modelling on commercial construction projects. Journal of information technology in construction (ITcon), 21, 233–249.

Boje, C., Guerriero, A., Kubicki, S. & Rezgui, Y. 2020. Towards a semantic Construction Digital Twin: Directions for future research. Automation in construction, 114, 103179.

Chen, L. & Whyte, J. 2022. Understanding design change propagation in complex engineering systems using a digital twin and design structure matrix. Engineering, construction and architectural management, 29, 2950–2975.

Dai, C., Cheng, K., Liang, B., Zhang, X., Liu, Q. & Kuang, Z. 2024. Digital twin modeling method based on IFC standards for building construction processes. Frontiers in Energy Research, 12, 1334192.

Esser, S., Vilgertshofer, S. & Borrmann, A. 2022. Graph-based version control for asynchronous BIM collaboration. Advanced Engineering Informatics, 53, 101664.

Esser, S., Vilgertshofer, S. & Borrmann, A. 2023. Version control for asynchronous BIM collaboration: Model merging through graph analysis and transformation. Automation in Construction, 155, 105063.

Hansun, S. A new approach of moving average method in time series analysis. 2013 conference on new media studies (CoNMedia), 2013. IEEE, 1–4.

Hasannejad, A., Sardrud, J. M. & Shirzadi Javid, A. A. 2023. BIM-based clash detection improvement automatically. International Journal of Construction Management, 23, 2431–2437.

Hevner, A. R., March, S. T., Park, J. & Ram, S. 2008. Design science in information systems research. Management Information Systems Quarterly, 28, 6.

Hyndman, R. J. 2025. Moving averages. International encyclopedia of statistical science. Springer.

Juszczyk, M., Tomana, A. & Bartoszek, M. 2016. Current issues of BIM-based design change management, analysis and visualization. Procedia engineering, 164, 518–525.

Khanzode, A., Fischer, M. & Reed, D. 2008. Benefits and lessons learned of implementing building virtual design and construction (VDC) technologies for coordination of mechanical, electrical, and plumbing (MEP) systems on a large healthcare project. Journal of Information Technology in Construction (ITcon), 13, 324–342.

Largani, B. S., Nasiri, M. M. & Jolai, F. 2026. Forecasting the completion time of construction projects using the moving average method. International Journal of Management and Decision Making, 25, 82–100.

Lee, G., Won, J., Ham, S. & Shin, Y. 2011. Metrics for quantifying the similarities and differences between IFC files. Journal of Computing in Civil Engineering, 25, 172–181.

Lin, J., Zhou, Y., Zhang, J. & Hu, Z. Classification and exemplary BIM models development of design changes. ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction, 2019. IAARC Publications, 122–127.

Moayeri, V., Moselhi, O. & Zhu, Z. 2015. Design change management using a BIM-based visualization model.

Motawa, I. A., Anumba, C. J., Lee, S. & Peña-Mora, F. 2007. An integrated system for change management in construction. Automation in Construction, 16, 368–377.

Narbaev, T. & De Marco, A. 2014a. Combination of Growth Model and Earned Schedule to Forecast Project

Cost at Completion. Journal of Construction Engineering and Management, 140.

Narbaev, T. & De Marco, A. 2014b. Combination of growth model and earned schedule to forecast project cost at completion. Journal of Construction engineering and management, 140, 04013038.

Pan, Y. & Zhang, L. 2021. Roles of artificial intelligence in construction engineering and management: A critical review and future trends. Automation in Construction, 122, 103517.

Peffers, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. 2007. A design science research methodology for information systems research. Journal of management information systems, 24, 45–77.

Radzi, A. R., Azmi, N. F., Kamaruzzaman, S. N., Algahtany, M. & Rahman, R. A. 2025. Challenges in construction readiness for BIM-based building projects. Journal of Asian Architecture and Building Engineering, 24, 1689–1704.

Sacks, R., Koskela, L., Dave, B. A. & Owen, R. 2010. Interaction of lean and building information modeling in construction. Journal of construction engineering and management, 136, 968–980.

Saoud, L. A., Omran, J., Hassan, B., Vilutienė, T. & Kiaulakis, A. 2017. A method to predict change propagation within building information model. Journal of Civil Engineering and Management, 23, 836–846.

Tillmann, P. A. 2020. Using the Last Planner System to tackle the social aspects of BIM-enabled MEP coordination. Canadian Journal of Civil Engineering, 47, 140–152.

Valkonen, T., Alhava, O., Viitanen, J. & Seppänen, O. 2025. Why do some projects prefabricate MEP while others do not?

Wang, R., Samarasinghe, D. A. S., Skelton, L. & Rotimi, J. O. B. 2022. A Study of Design Change Management for Infrastructure Development Projects in New Zealand. Buildings, 12, 1486.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DIGITAL TWIN VALUE CREATION AS A DYNAMIC SYSTEM: A CAUSAL LOOP DIAGRAM APPROACH

J Omar Abbas, Nicola Moretti, and Tim Broyd
University College London, London, United Kingdom

Abstract

Digital Twins (DTs) are promoted as a transformative paradigm for the built Environment, yet their value is articulated through benefit propositions that lack cost integration. This study proposes a new conceptualisation of DTs value creation for asset owners, based on System Dynamics (SD). DT costs and benefits are identified first and modelled as variables of four subsystems representing operational, human, financial and market mechanisms across the asset lifecycle. The results reveal different feedback loops of how value is realised. The contribution is a general system-level explanation of DT value that supports more informed investment decisions.

Introduction

Digital Twins (DTs) are increasingly positioned as a transformative paradigm for the Architecture, Engineering, Construction and Operation (AECO) industry, promising to enhance decision-making, performance, and value creation across the asset lifecycle. DTs integrate multiple digital technologies to enable continuous data exchange between physical assets and their digital representations, allowing organisations to monitor, analyse, and predict asset behaviour over time (Iqbal & Mirzabeigi, 2025). Despite growing interest and increasing experimentation, the value of DTs remains difficult to articulate and justify in practice (Opoku et al., 2023). Much of the existing discourse emphasises expected or potential benefits, often presented in static or descriptive terms, while providing limited insight into how benefits relate to incurred costs or how these elements interact over time and across organisational functions (Abbas et al., 2024).

The AECO industry has struggled to translate digital technological advances into productivity and performance improvements (Berlak et al., 2021). These challenges stem from inherent industry characteristics, including high uncertainty with work processes that are often unpredictable and difficult to standardise (Oesterreich & Teuteberg, 2016). Against this setting, Industry 4.0 principles have stimulated interest in digital technologies capable of addressing productivity and decision-making challenges in the AECO industry. Technologies such as Building Information Modelling (BIM), Artificial Intelligence (AI), and the Internet of Things (IoT), have been identified as key drivers of digital transformation in the AECO industry (Moshood et al., 2024). DTs have emerged as a prominent concept within this broader set of digital technologies, integrating many of them to support lifecycle-oriented asset management.

Problem statement and research gap

A central challenge lies in the lack of a clear and shared understanding of DT value. As a result, organisations often struggle to justify the additional investment burden associated with DT development and implementation (Opoku et al., 2023). Furthermore, there is no established framework or taxonomy that articulates the full spectrum of DT value delivered to different stakeholders across the asset lifecycle, leaving their true value largely undefined (Pregnolato et al., 2022).

Existing research has focused on describing value propositions or potential benefits of DTs, without empirical validation or explicit consideration of associated costs. For instance, in their formulation of ontological dimensions for DTs during the construction phase, Sacks et al. (2020) presented some benefits and described them as potential or expected. These benefits were not examined in terms of how they interact with costs within organisational settings. This highlights a gap in both research and practice which is the absence of a system-level explanation that links DT-enabled benefits and costs through explicit cause/effect mechanisms within organisational settings. This gap underscores the need for approaches that move beyond static representations of DT benefits and instead provide a systemic explanation of how DT value is created over time.

Research aim, questions, and objectives

This study adopts Causal Loop Diagrams (CLDs) within System Dynamics (SD) to move beyond the static definition of DT benefits and costs. The aim of this study is to develop a system-level conceptual model that represents how value from DTs emerge over time through the interaction of operational, human, financial and market mechanisms across the asset lifecycle, from an asset owner perspective. Guided by this perspective, the study addresses two Research Questions (RQs) and identifies four related Research Objectives (ROs) (see Table 1).

Table 1: Research Questions and Objectives of this study.

Research Questions	Related Research Objectives
RQ1: What DT benefits and costs are relevant for representing value creation from an asset owner perspective across the asset lifecycle?	**RO1**: to identify and synthesise the benefits and costs of DTs that are relevant to asset owners across the planning, delivery, operation and end-of-life phases.
	RO2: to structure these benefits and costs into an integrated system encompassing operational processes, labour dynamics, financial performance, and market outcomes, and to classify performance-related variables using Balanced Scorecard (BSC) perspectives (financial, customer, internal business, and innovation and learning).
RQ2: How do these benefits and costs interact through causal relationships and feedback mechanisms?	**RO3**: to develop a CLD that represents the causal relationships between DT capabilities and performance metrics, under the assumed case of active and consistent DT use across the asset lifecycle.
	RO4: to identify and describe the dominant feedback loops that shape how DT value evolves over time.

The remainder of this paper is structured as follows. The Related work section reviews DT literature in the built environment, focusing on how benefits and costs have been conceptualised. The Materials and methods outline the identification and synthesis of DT value elements, the CLD approach, and the asset owner system boundary and subsystems. The Results and discussion present and interpret the CLD and its dominant feedback mechanisms, with implications for asset owners' decision-making. The paper concludes by summarising the main contributions and outlining directions for future research.

Related work

The ISO/IEC 30173:2023 defines a DT as a "digital representation […] of a target entity […] with data connections that enable convergence between the physical and digital states at an appropriate rate of synchronization" (ISO, 2023: p. 6). There is also a broad agreement on three fundamental DT building blocks: the physical asset, its digital representation, and the data exchange linking both (Korenhof et al., 2021). While there is no technology stack prescribed for developing a DT, DTs could be characterised as a collection of technologies (Wei et al., 2022). Alongside BIM, which is widely regarded as a core modelling element of DTs (Revolti et al., 2024), key enabling technologies include IoT, AI, Cloud Computing and Extended Reality (XR) which encompasses Augmented Reality (AR), Virtual Reality (VR) and Mixed Reality (MR) (Attaran & Celik, 2023).

In their review of DT applications in the built environment, Kookalani et al. (2026) mapped use cases across lifecycle phases and sectors. During design and construction, applications include lifecycle planning, progress monitoring, and resource management, while in operation and maintenance they include fault detection, energy optimisation, and predictive maintenance. Across sectors, applications span infrastructure projects, buildings and urban development contexts, among others. The different applications are what ultimately lead to value creation, which has been identified as a fundamental property of DTs (Bolton et al., 2018). In the literature, the concept of value is frequently invoked but often used inconsistently, sometimes interchangeably with benefits only (Laursen & Svejvig, 2016). Value is instead understood as a relationship between benefits and costs, consistent with definitions that conceptualise value as benefits realised minus costs incurred (Morris, 2013).

Prior research has proposed several ways of classifying benefits. For example, they have been categorised according to strategic, tactical, and operational levels (Irani & Love, 2002), where strategic benefits are often intangible and difficult to quantify, and operational benefits are typically more tangible and measurable. Bradley (2010) further distinguishes between intermediate benefits and end benefits, arguing that end benefits are often realised only after a sequence of intermediate benefits has occurred. It was further delineated that enablers are inputs that make benefits possible, and in the context of this study, implementing a DT could be considered as such.

With respect to costs, studies have distinguished between capital expenditure (CapEx) and operational expenditure (OpEx) of DTs, identifying sensors, computing infrastructure, and software licensing as CapEx, and maintenance and server-related expenses as OpEx (McClenaghan et al., 2023). This study, however, was largely based on experimental settings and did not provide a comprehensive account of DT costs across the asset lifecycle. Oesterreich and Teuteberg (2018a) identified the costs of developing and utilising BIM from earlier Information Science domain studies, distinguishing between indirect costs, and direct costs. In this framing, direct costs are those directly attributable to operating the digital system (e.g., hardware and software costs), whereas indirect costs include human inputs (e.g., training, and human resource costs) and organisational costs associated with shifting to new workflows and practices (Love et al., 2013). More recently, Su et al. (2025) proposed a methodology for estimating DT costs in the manufacturing domain. While this study provides useful insights, it focuses exclusively on the development phase of a DT and is limited to a prototype-scale implementation in the manufacturing domain.

Overall, the literature reveals a strong emphasis on DT benefits, often presented as value propositions without empirical evidence, alongside a comparatively limited and fragmented treatment of costs (Abbas et al., 2024).

Moreover, existing studies did not provide system-level explanations that link benefits and costs or account for how these elements interact and evolve over time.

Materials and methods

This study adopts a qualitative multi-method approach. It builds on a previously published Systematic Literature Review (SLR) (Abbas et al., 2024), which identifies the benefits and costs of digital twins and informs the development of the CLD model presented in this paper. The model is used to explain how value emerges over time. The research schema is shown in Figure 1.

Identification and synthesis of benefits and costs

Additional studies were incorporated to expand and consolidate the list of benefits of DTs identified in Abbas et al. (2024), including cost-benefit analysis (CBA) models using SD for BIM (Oesterreich & Teuteberg, 2018a) and AR in the built environment (Oesterreich & Teuteberg, 2018b). Both BIM and AR are widely recognised as integral components of DT systems (Attaran & Celik, 2023). Incorporating these studies therefore allowed the synthesis to capture benefits that are either explicitly attributed to DTs or implicitly enabled through DT-related technologies. The synthesised benefits are illustrated in Table 2. Cost identification drew on DT cost studies (Oettl et al., 2023; Su et al., 2025), and BIM cost classifications (Oesterreich & Teuteberg, 2018a).

System Dynamics modelling

CLDs form a core component of the SD methodology and are commonly used as an initial step in modelling system behaviour. They provide a qualitative representation of a system by depicting variables and the causal relationships between them, including the polarity of each relationship and the presence of reinforcing or balancing feedback loops (Sterman, 2000). Variables can be either exogenous (external inputs that influence the system but are not influenced by it) or endogenous (internal system variables whose behaviour is determined by interactions with other variables) (Ford, 2019). Costs and benefits in this study are modelled as variables of the overall DT SD model. Vensim[1] was used to create the CLD.

The modelling approach in this study followed the first two stages of the SD process outlined by Sterman (2000). The first stage, problem articulation, is addressed through the definition of the research problem and the adoption of an asset owner perspective. The asset owner is defined as the long-term beneficiary of the asset's financial value, responsible for governing asset management policy and whole-life performance requirements (The Institute of Asset Management, 2024). This boundary is assumed to encompass facility management functions and the client/investment role. Delivery organisations such as designers, contractors, and operators may be outsourced, but their contributions are treated within the boundary because the asset owner remains ultimately accountable for lifecycle value outcomes. Occupiers/tenants are included as stakeholders within the boundary. For conceptual clarity, the model assumes a single asset owner and does not explicitly represent ownership transitions or split ownership. The boundary spans the asset lifecycle from initial DT investment through operation and end-of-life, with financial performance represented up to Earnings Before Interest and Taxes (EBIT) and excluding external macroeconomic factors outside organisational control such as inflation or interest rates.

The overall system is structured into four interacting subsystems adapted from earlier SD models of organisational performance (Clark & Augustine, 1992) and subsequent applications in the construction domain (Oesterreich & Teuteberg, 2018a; Oesterreich & Teuteberg, 2018b). The subsystems used in this study are:

1. Processes & Data (Information, data, systems, execution),
2. Personnel & Labour (People, skills, time),
3. Sales & Market (Output, quality, demand),
4. Finance & Accounting (Cost aggregation, profits).

Each variable was assigned to one of the four Balanced Scorecard (BSC) perspectives proposed by Kaplan and Norton (1992): financial (e,g., profitability and cost-related measures) customer (e.g., customer satisfaction, retention, and market share), internal business (e.g., operational efficiency), and innovation and learning (e.g., workforce development, and innovation capability). Using BSC with SD modelling has been used in prior work, for example to group performance metrics into performance drivers and outcome metrics before mapping

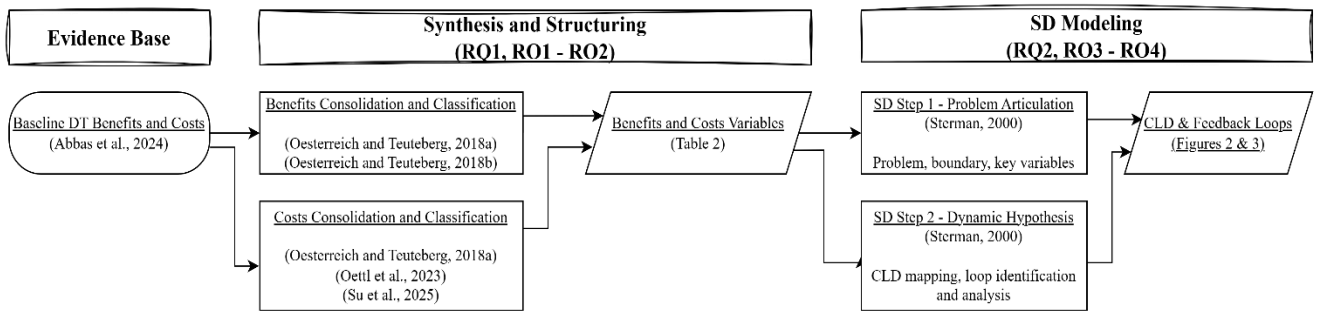


Figure 1: Flowchart illustrating the research schema.

[1] https://vensim.com/software/

their links in a CLD (Hristov et al., 2024). Utilising BSC classification at the CLD stage improves the model's relevance and interpretability for asset owners (and other stakeholders within this organisational boundary). Although these metrics are not quantified at this stage, the BSC grouping supports selection of quantifiable indicators for translating the CLD into a Stock and Flow Diagram (SFD), consistent with standard SD practice (Sterman, 2000).

The second stage, formulation of a dynamic hypothesis, involves identifying the key endogenous variables and feedback structures that are hypothesised to drive system behaviour over time and it is operationalised through the development of the CLD. To select the feedback loops for analysis, the Shortest Independent Loop Set (SILS) method is applied to identify a set of shortest, non-redundant loops that account for system behaviour (Oliva, 2004).

Results and discussion

Consolidation of DT value elements

To address RO1 & RO2, Table 2 summarises the variables included in the model, indicating their associated subsystem, endogenous or exogenous status, and Performance metric/perspective, along with the consolidated benefits from literature.

Consistent with the benefit realisation perspective (Bradley, 2010), some benefits could be interpreted as DT-based organisational enablers (depicted as DT capabilities thereafter), reflecting improved capabilities in monitoring, coordination, risk control, and asset management as examples. Other benefits are represented as internal business outcomes, including reduced execution time, reduced rework and error incidence, and improved output quality and output volume. Customer performance metrics reflect the organisation's relative position in the market, including reputation and market share, while innovation and growth performance metrics show employee and organisational growth.

Most financial variables, as illustrated in Table 2 and Figure 2, were derived from the literature. Additional elements were introduced to improve system-level representation, including cost aggregation (direct, overhead and indirect costs), and depreciation to derive EBIT. Cost variables represent both initial and ongoing investment expenditures, as well as costs that arise endogenously from system behaviour, such as staffing, training, and injury compensation costs. Costs are represented as both exogenous and endogenous variables, depending on their role in the system. Baseline investment and operational costs associated with DT deployment are treated as exogenous, while costs arising from workforce dynamics, business processes, and organisational change are modelled as endogenous.

System-level CLD and cross-subsystem integration

To address RO3, the CLD, illustrated in Figure 2, illustrates how DT capabilities influence different performance perspectives through interacting feedback loops across the asset lifecycle. These capabilities are represented as exogenous variables, and they are highlighted using dotted green boxes to distinguish them from other variables. The CLD indicates that DT value is not produced directly by technology deployment but emerges through the interaction of DT capabilities with operational, workforce, financial and market mechanisms. In the model, DT capabilities shape endogenous behaviour primarily through internal business metrics such as execution time, rework and error incidence, and output quality.

The *Process and Data subsystem* captures the primary pathways which DT capabilities affect internal business processes. For instance, capabilities related to information management, forecasting, coordination, and asset performance (exogenous capabilities) influence execution time, rework and error incidence, and output quality (internal business performance metric). For example, information and documentation quality, and site coordination effectiveness reduce rework and error incidence by limiting coordination errors, and this in turn improves output quality. Progress monitoring capability, and supply chain reliability reduce schedule slippage and execution time by enabling earlier detection of deviations. Asset performance and efficiency improve operational continuity by reducing downtime, thereby improving labour productivity. Several of the internal business performance variables are highly interconnected, reflecting the central role of operational performance as a mediator between DT and downstream customer and financial outcomes. Within this subsystem, execution time and rework and error incidence emerge as especially central, given their multiple connections to other performance metrics. In terms of cross-subsystem interactions, the Processes and Data variables receive incoming influences from only three variables in the Personnel and Labour subsystem: experience accumulation, labour productivity, and workforce skills.

The *Personnel and Labour subsystem* captures workforce-related mechanisms through which DT capabilities and internal business metrics affect productivity, safety, workforce stability and associated costs. Safety monitoring and prevention capability appears as a native variable of this subsystem, and it is the only DT capability in this subsystem reflecting its primary influence on workforce outcomes. This capability reduces accident frequency, and this improves labour productivity and enhances organisational reputation. Labour productivity is a highly central variable within this subsystem, linking workforce skills, safety outcomes, and staffing requirements. This position reflects its function linking workforce dynamics to financial outcomes.

The *Sales and Market subsystem* represents how internal business performance translates into customer outcomes over time. For example, improved output quality and reduced rework increase customer satisfaction, which supports order volume and strengthens organisational reputation. This in turn improves the firm's competitive position and market share. Environmental management

capability is the only DT capability in this subsystem, and it is shown to positively influence customer satisfaction.

Finally, *the Finance and Accounting subsystem* captures the cost and revenue consequences of DT development and implementation. Exogenous financial variables represent DT deployment and operation costs, such as data acquisition, software and platform, and security and data protection, and organisational and change management costs. Endogenous cost variables arise from internal system behaviour, including injury and compensation costs, hiring and recruitment costs, training costs, human resource costs. Direct costs, and overheads and indirect costs act as aggregation points capturing the cumulative effects of operational activities. Revenue is driven jointly by output volume and order volume. EBIT emerges from the interaction of revenue generation and cost accumulation.

Key feedback loops

To address RO4, applying the SILS method identified four reinforcing feedback loops, illustrated in Figure 3. Loop 1 (order volume - experience accumulation - output quality - customer satisfaction - order volume) is reinforcing: as demand increases, experience accumulates, which improves output quality and customer satisfaction, supporting further demand growth. This implies that DT value can compound over time when organisational learning is sustained.

Loop 2 (accident frequency - organisational reputation - employer attractiveness - staff turnover rate - experience accumulation - workforce skills - accident frequency) is also reinforcing and describes how safety retains and improves organisational reputation. Here, the exogenous safety monitoring and prevention capability acts as a lever that reduces accident frequency. This loop shows the

Table 2: Variables used in the CLD and corresponding Subsystem. Financial or cost items that are aggregated into direct costs, or overhead and indirect costs are excluded from the table and presented in Figure 2.

Subsystem	CLD Variable	Endogenous/ Exogenous	Performance metric/perspective	Syntehsised benefit from literature
Processes and Data	Information and documentation quality	Exogenous	DT capability	Improved information and documentation management
	Progress monitoring capability	Exogenous	DT capability	
	Forecasting and scenario capability	Exogenous	DT capability	Improved forecasting and scenario analysis
	Risk management capability	Exogenous	DT capability	Better risk management
	Site coordination effectiveness	Exogenous	DT capability	Improved site management
	Asset performance and efficiency	Exogenous	DT capability	Improved maintenance
	Supply chain reliability	Exogenous	DT capability	Improved supply chain management
	Resource allocation efficiency	Exogenous	DT capability	Better resource management
	Emergency response capability	Exogenous	DT capability	Better emergency and crisis management
	Environmental management capability	Exogenous	DT capability	Better environmental management
	Schedule slippage	Endogenous	Internal Business	Improved progress monitoring
	Risk exposure	Endogenous	Internal Business	
	Downtime	Endogenous	Internal Business	
	Execution time	Endogenous	Internal Business	Reduced execution time
	Rework and error incidence	Endogenous	Internal Business	
	Output quality	Endogenous	Internal Business	Improved output
	Output volume	Endogenous	Internal Business	
Personnel and Labour	Safety monitoring and prevention capability	Exogenous	DT capability	Better health and safety
	Labour productivity	Endogenous	Internal Business	Improved productivity
	Experience accumulation	Endogenous	Innovation & Learning	
	Workforce skills	Endogenous	Innovation & Learning	
	Staff turnover rate	Endogenous	Innovation & Learning	
	Employer attractiveness	Endogenous	Innovation & Learning	
	Employee satisfaction	Endogenous	Innovation & Learning	
	Required staff	Endogenous	Internal Business	
	Digital capability training	Exogenous	Innovation & Learning	
	Accident frequency	Endogenous	Internal Business	
	Injury severity	Endogenous	Internal Business	
Sales and Market	Customer satisfaction	Endogenous	Customer	Higher customer satisfaction
	Competitive position and market share	Endogenous	Customer	Competitive advantage gain
	Order volume	Endogenous	Customer	
	Organisational reputation	Endogenous	Customer	
Finance and Accounting	Depreciation	Exogenous	Financial	
	Direct costs	Endogenous	Financial	
	Overheads and indirect costs	Endogenous	Financial	
	Procurement effectiveness	Exogenous	DT capability	Improved procurement
	Revenue	Endogenous	Financial	
	EBIT	Endogenous	Financial	

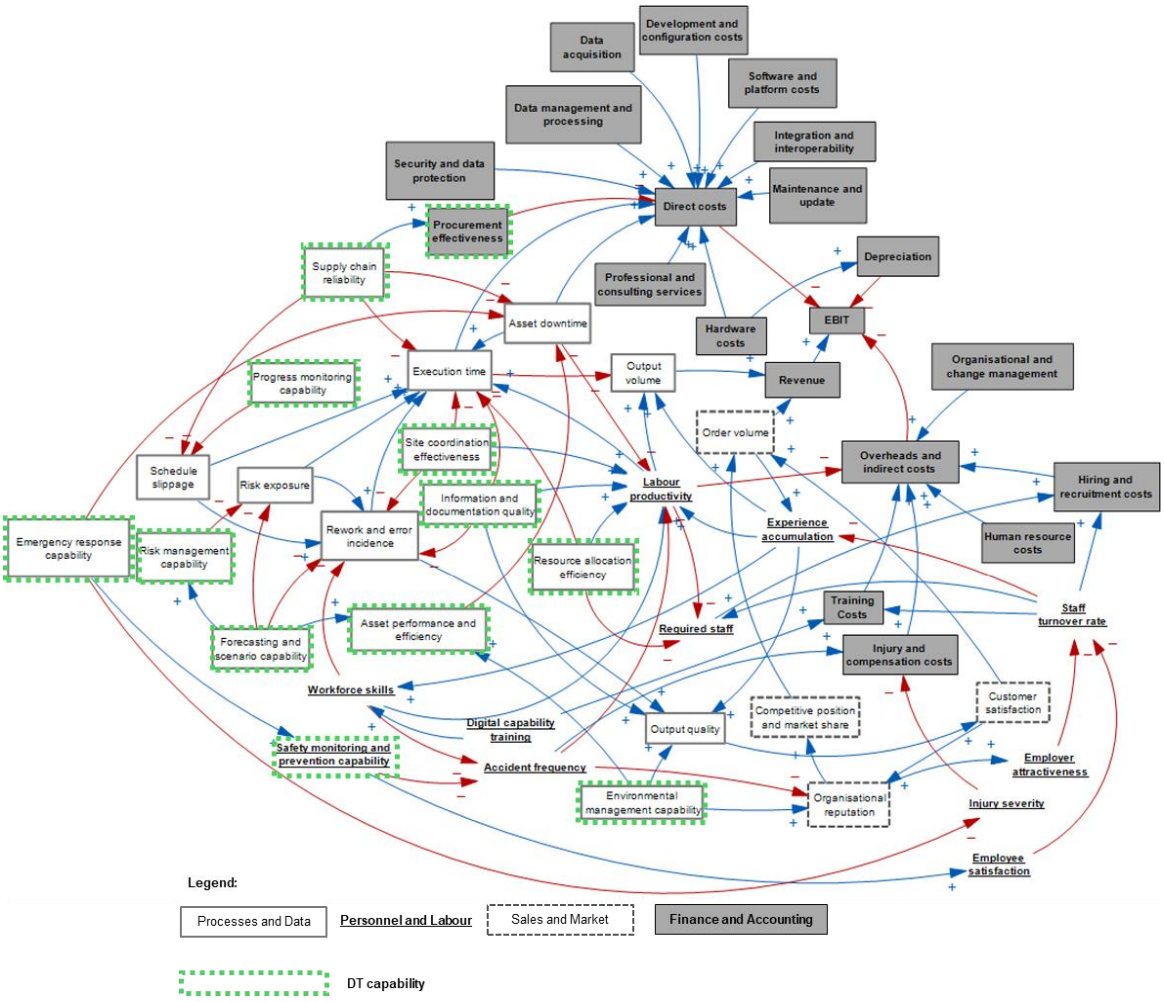


Figure 2: System-level CLD across the four subsystems.

significance of safety outcomes, and how they could shape a strategic organisational reputation.

Loop 3 (organisational reputation - competitive position and market share - order volume - experience accumulation - workforce skills - accident frequency - organisational reputation) caputres how market-facing outcomes intereact with safety and capability accumulation. This loop shows how organisational reputation can translate into competitive positioning and demand, while improving safety outcomes.

Loop 4 (organisational reputation - employer attractiveness - staff turnover rate - experience accumulation - workforce skills - accident frequency - organisational reputation) is also reinforcing and reflects a workforce stability mechanism along with improved safety outcomes. The loop also shows how organisational reputation can lead to capability accumulation by improving employer attractiveness.

Reflection and limitations

Given the limited evidence on realised DT value, this model was adapted from the SD-based CBA models developed by Oesterreich and Teuteberg (2018a) for BIM and Oesterreich and Teuteberg (2018b) for AR. Accordingly, this model should be interpreted as a conceptual hypothesis that organises the available evidence, rather than a definitive representation of DT value elements. It is presented as an initial gauging step to support progression to subsequent SD modelling stages.

Structurally, the three models differ in boundary choices and the explicit representation of technology capabilities. The BIM model is organised across three domains (Personnel and Labour; Customer and Market; Finance and Accounting), whereas the DT model adopts four subsystems, by introducing processes and data subsystem, to make DT capabilities clearer. The AR model also adopts four subsystems by introducing a Processes subsystem, which is comparable to the DT model's Processes and Data subsystem, although AR capabilities are not explicitly distinguished in the same way.

Across the market-facing domain, the DT model and the BIM model share comparable causal pathways. In the BIM model, environmental performance influences output quality and customer satisfaction, which then propagate to demand outcomes through customer base and orders. A similar mechanism appears in the DT model, where environmental management capability influences output quality and customer satisfaction, which then increases order volume. In the Personnel and Labour subsystem, the models also align in broad terms. Both BIM and AR models highlight safety-related mechanisms and their influence on productivity, and the BIM model

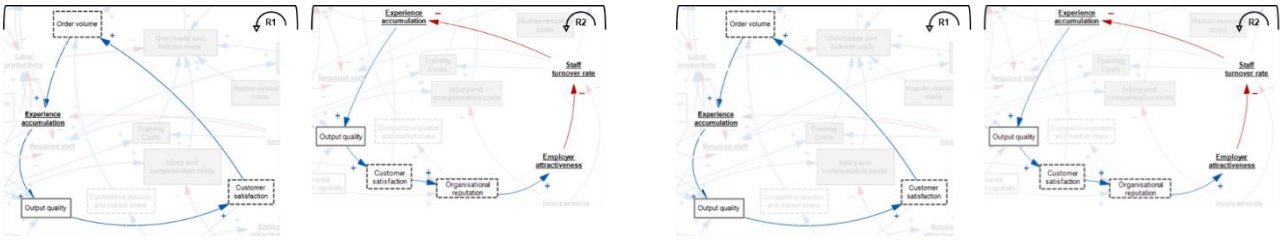


Figure 3: Four dominant feedback loops.

links injuries and accidents to corporate image and employer attractiveness. The DT model expresses a closely related pathway by linking accident frequency to organisational reputation and employer attractiveness.

The financial boundary and level of accounting detail differ across the models. The BIM model incorporates a more detailed financial structure extending beyond operating profit, whereas the present DT model adopts a simpler boundary with financial performance represented up to EBIT. The AR model applies a relatively simple financial structure as well and is therefore closer in scope to the DT model in its reliance on revenue, direct costs, and overheads and indirect costs. Despite differences in granularity, all three models share the same basic accounting logic in which value outcomes depend on revenue generation and cost accumulation across direct and indirect categories.

Unlike the other models, the DT model used BSC perspectives to structure the variables into recognisable performance domains, improving interpretability and clarifying how DT capabilities influence internal business processes and propagate to customer and financial outcomes. This could also provide a practical basis for subsequent quantification when translating the CLD into a SFD.

Regarding the limitations of this study, the benefits and costs were derived solely from secondary sources and may therefore not yet be comprehensive. The CLD built on earlier SD models developed for DT-enabling technologies, reflecting the limited availability of comparable SD studies. Third, this study only developed a CLD without progressing to later SD stages such as simulation, testing, or policy evaluation (Sterman, 2000), and within the CLD, the feedback loops analysed were not exhaustive. Furthermore, the financial boundary was limited to EBIT (rather than net profit), and the asset owner boundary excluded external macroeconomic factors such as inflation and interest rates, and this may affect realised value in practice.

Conclusion and future research

This study developed a CLD that explained how DT value evolves through interactions between operational, human, financial and market mechanisms across the asset lifecycle, from an asset owner perspective. In response to the literature gap identified, this study moves beyond itemised representations of DT benefits and costs by providing a system-level explanation (Sterman, 2000). In response to RQ1, the study identified and synthesised DT benefits and costs relevant to asset owners and structured them into an integrated system spanning four subsystems. For RQ2, the study developed the CLD structure linking DT capabilities to BSC perspective. The dominant feedback loops were also illustrated.

The study has implications for both research and practice. For academia, the model provides a conceptual basis for future empirical work on DT value mechanisms. The identified variables and CLD can be used in subsequent studies to test and validate these mechanisms for asset owners in the AECO industry. For practitioners, including asset owners and other stakeholders interacting with them, the CLD provides a tool for structuring investment decision-making while supporting more realistic expectations of value realisation over time. For policymakers, the results reinforce the importance of supporting not only technology deployment but also organisational capability building, to accelerate the realisation of value in DT systems.

Future work will initially ensure the completeness of benefits and costs of DTs by interviewing experts. Subsequently, the CLD will be tested and validated through stakeholder interviews and focus groups and then develop a SFD for quantitative simulation and scenario testing (Sterman, 2000). This will also include widening the financial boundary beyond EBIT and test sensitivity to external factors.

Acknowledgements

This research was funded partly by Bartlett School of Sustainable Construction, UCL.

References

Abbas, O.H., Moretti, N. and Broyd, T. (2024). 'Assessing the Benefits of Digital Twins Towards the Formulation of Whole Life Cycle Value Propositions: A Systematic Review'. In Proceedings of the 2024 European Conference on Computing in Construction.

Attaran, M. and Celik, B.G. (2023). 'Digital Twin: Benefits, use cases, challenges, and opportunities'. Decision Analytics Journal, 6, p.100165.

Berlak, J., Hafner, S. and Kuppelwieser, V.G. (2021). 'Digitalization's impacts on productivity: a model-based approach and evaluation in Germany's building construction industry'. Production Planning & Control, 32 (4), pp.335–345.

Bolton, A., Blackwell, B., Dabson, I., Enzer, M., Evans, M., Harradence, F., Keaney, E., Kemp, et al. (2018). 'The Gemini Principles: Guiding Values for the

National Digital Twin and Information Management Framework'

Bradley, G. (2010). Benefit Realisation Management: A Practical Guide to Achieving Benefits Through Change. Farnham: Gower.

Ford, D.N. (2019). 'A system dynamics glossary'. System Dynamics Review, 35 (4), pp.369–379.

Hristov, I., Cristofaro, M., Camilli, R. and Leoni, L. (2024). 'A system dynamics approach to the balanced scorecard: a review and dynamic strategy map for operations management'. Journal of Manufacturing Technology Management, 35 (4), pp.705–743.

Iqbal, F. and Mirzabeigi, S. (2025). 'Digital Twin-Enabled Building Information Modeling–Internet of Things (BIM-IoT) Framework for Optimizing Indoor Thermal Comfort Using Machine Learning'. Buildings, 15 (10), p.1584.

Irani, Z. and Love, P.E.D. (2002). 'Developing a frame of reference for ex-ante IT/IS investment evaluation'. European Journal of Information Systems, 11 (1), pp.74–82.

Kaplan, R. and Norton, D. (1992). 'The Balanced Scorecard: Measures that Drive Performance'. Harvard Business Review, 70 (1), pp.71–79.

Kookalani, S., Green, S., Luo, P., Alavi, H., Parn, E., Sun, Z. and Brilakis, I. (2026). 'Mapping digital twin applications in infrastructure and the built environment across research types, methods, sectors, phases, and scales'. Automation in Construction, 182, p.106778.

Korenhof, P., Blok, V. and Kloppenburg, S. (2021). 'Steering Representations—Towards a Critical Understanding of Digital Twins'. Philosophy & Technology, 34 (4), pp.1751–1773.

Laursen, M. and Svejvig, P. (2016). 'Taking stock of project value creation: A structured literature review with future directions for research and practice'. International Journal of Project Management, 34 (4), pp.736–747.

Love, P.E.D., Simpson, I., Hill, A. and Standing, C. (2013). 'From justification to evaluation: Building information modeling for asset owners'. Automation in Construction, 35, pp.208–216.

McClenaghan, A., Gopsill, J., Ballantyne, R. and Hicks, B. (2023). 'Cost Benefit Analysis for Digital Twin Model Selection at the Time of Investment.'. Procedia CIRP, 120, pp.1197–1202.

Morris, P. (2013). Reconstructing Project Management. Chichester: Wiley Blackwell.

Moshood, T.D., Rotimi, J.O.B., Shahzad, W. and Bamgbade, J.A. (2024). 'Infrastructure digital twin technology: A new paradigm for future construction industry'. Technology in Society, 77.

Oesterreich, T.D. and Teuteberg, F. (2018a). 'Looking at the big picture of IS investment appraisal through the lens of systems theory: A System Dynamics approach for understanding the economic impact of BIM'. Computers in Industry, 99, pp.262–281.

Oesterreich, T.D. and Teuteberg, F. (2018b). 'Why one big picture is worth a thousand numbers: measuring intangible benefits of investments in augmented reality based assistive technology using utility effect chains and system dynamics'. Information Systems and e-Business Management, 16 (2), pp.407–441.

Oettl, F., Eckart, L. and Schilp, J. (2023). 'Cost estimation approach of a digital twin implementation in industry'. Procedia CIRP, 118, pp.318–323.

Oliva, R. (2004). 'Model structure analysis through graph theory: partition heuristics and feedback structure decomposition'. System Dynamics Review, 20 (4), pp.313–336.

Opoku, D.-G.J., Perera, S., Osei-Kyei, R., Rashidi, M., Bamdad, K. and Famakinwa, T. (2023). 'Barriers to the Adoption of Digital Twin in the Construction Industry: A Literature Review'. Informatics, 10 (1).

Pregnolato, M., Gunner, S., Voyagaki, E., De Risi, R., Carhart, N., Gavriel, G., Tully, P., Tryfonas, T., Macdonald, J. and Taylor, C. (2022). 'Towards Civil Engineering 4.0: Concept, workflow and application of Digital Twins for existing infrastructure'. Automation in Construction, 141 (July), p.104421.

Revolti, A., Gualtieri, L., Pauwels, P. and Dallasega, P. (2024). 'From building information modeling to construction digital twin: a conceptual framework'. Production & Manufacturing Research, 12 (1), p.2387679.

Sacks, R., Brilakis, I., Pikas, E., Xie, H.S. and Girolami, M. (2020). 'Construction with digital twin information systems'. Data-Centric Engineering, 1, p.e14.

Sterman, J.D. (2000). Business dynamics: systems thinking and modeling for a complex world. Boston: McGraw-Hill.

Su, S., Nassehi, A., McClenaghan, A., Langridge, A. and Hicks, B. (2025). 'A methodology for estimating the cost of a digital twin'. Journal of Manufacturing Systems, 80, pp.841–858.

The Institute of Asset Management (2024). Asset Management - an anatomy. Bristol, UK: The Institute of Asset Management.

Wei, Y., Lei, Z. and Altaf, S. (2022). 'An Off-Site Construction Digital Twin Assessment Framework Using Wood Panelized Construction as a Case Study'. Buildings, 12 (5), p.566.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A MONITORING-ORIENTED DIGITAL TWIN PIPELINE FOR DATA-DRIVEN PREDICTION OF RESIDENTIAL PROSUMER ENERGY PROFILES

Amin Rahmani[1], Mahsa Moradian[1], Sonia Lupica Spagnolo[1], and J.J. McArthur[2]
[1] Politecnico di Milano, Milan, Italy
[2] Toronto Metropolitan University, Toronto, Canada

Abstract

Residential prosumer buildings with PV and high-load subsystems exhibit irregular demand, yet current residential digital-twin and baselining studies rarely address missing-data realism and causal deployment constraints. This paper presents a monitoring-oriented operation-phase digital twin analytics pipeline for an instrumented villa in Segrate, Italy, linking whole-building demand, PV output, pool-circuit power, and outdoor air temperature to a one-hour-ahead baseline task. A fixed Random Forest is evaluated across eight dataset configurations against persistence and Multiple Linear Regression. Under deployment-realistic causal conditions, the best configuration achieves RMSE 576.5 W, demonstrating the value of a traceable leakage-safe baselining workflow for residential prosumer settings.

Introduction

Digital twins, which were initially developed as models for the design phase, are now being used as effective tools to monitor and manage building operations (Oulefki et al., 2025; Mengual Torres et al., 2022). In this setting, data-driven energy baselining has emerged as a method to generate short-horizon reference predictions of building demand, supporting fault detection, performance tracking, and measurement and verification (Xie et al., 2023).

In the residential sector, smart home automation and IoT monitoring have been extensively studied, but residential digital twins are still less prevalent compared to their use in commercial facility management (Mengual Torres et al., 2022). Many existing residential studies concentrate on basic monitoring dashboards or the optimisation of specific components like batteries or heat pumps, lacking a robust and traceable connection between sub-metered devices and operational building-level representations (Xie et al., 2023). Residential prosumer buildings are also more difficult to model than conventional offices due to more stochastic occupancy patterns and the presence of onsite generation and high-load amenities, such as rooftop photovoltaics and swimming pools, which lead to non-linear and intermittent demand profiles necessitating comparison against both persistence and linear benchmark models to isolate the value of complex learners (Ghaemi et al., 2024). In the absence of a dependable short-horizon baseline that reflects these factors, building owners and facility managers do not have a clear reference point for detecting anomalies or assessing system upgrades (Ghaemi et al., 2024; Xie et al., 2023).

This gap is addressed in this paper using an instrumented residential villa in Segrate (Milan, Italy) as a documented operational-phase digital-twin testbed, linking heterogeneous IoT telemetry (whole-building electrical demand, PV production, swimming-pool circulation pump load, and outdoor air temperature) into a traceable data-processing pipeline that enforces domain constraints and prevents time leakage. This paper investigates a monitoring-oriented operation-phase digital twin analytics pipeline and documented residential testbed for causally consistent short-horizon baselining, addressing a gap narrower than a general digital-twin one: few residential prosumer studies jointly examine one-hour-ahead operational baselining, sub-metered telemetry, explicit missing-data policy, causal versus non-causal imputation, and comparison against persistence and linear regression within a leakage-safe design (Krishnan et al., 2024). Here, digital twin denotes a traceable operational representation linking telemetry, subsystem context, and analytics; it does not imply bidirectional control or formal BIM integration (Xie et al., 2023).

The one-hour-ahead horizon is chosen because the baseline serves as an operational reference for residual-based anomaly screening rather than long-range forecasting. At this horizon the baseline can support near-real-time fault detection in subsystems such as the pool circulation pump or the PV array, and it can provide building owners or facility managers with a timely reference for detecting deviations before they propagate into larger energy waste or equipment damage. The research question is whether adding sub-metered telemetry improves one-hour-ahead residential demand baselines over demand-only models and how that benefit varies across dataset configurations that differ in imputation strategy and missing-data policy.

Related work

Operation-phase monitoring pipelines and metadata traceability in residential digital twins

Operation-phase digital twin applications rely on robust data integration and metadata management to connect time-series data to building context (Xie et al., 2023). Recurring challenges in metadata consistency for smart-

building analytics motivate digital twin architectures that incorporate a traceable data integration layer linking building entities to operational data streams through curated metadata configurations (Balaji et al., 2018; Oulefki et al., 2025).

Data-driven energy prediction for occupied buildings

Building energy prediction has been extensively examined through three primary modelling paradigms: physics-based ("white-box"), data-driven ("black-box"), and hybrid ("grey-box") approaches. Reviews commonly observe that white-box models offer interpretability, but demand detailed and precise building parameters along with significant calibration efforts. In contrast, data-driven approaches are generally easier to implement when historical operational data are accessible and can deliver strong forecasting performance (Krishnan et al., 2024).

For operation-phase baselining and short-horizon forecasting, data-driven methods are frequently adopted as practical predictors, provided that issues such as data quality, missing values, and non-stationary operation (e.g., discontinuities from in-use residential operation) are appropriately addressed during preprocessing and evaluation (Krishnan et al., 2024). Within this framework, the methodological significance of baseline construction lies not only in final accuracy but also in how the inclusion of explanatory operational features, such as weather data, subsystem telemetry, and temporal context, contributes to the reliability and interpretability of predictive references relative to naïve benchmarks (Zhou, Yang & Wang, 2024). However, many studies still treat data handling and deployment constraints as secondary implementation aspects rather than as central methodological considerations (Krishnan et al., 2024; Xie et al., 2023).

The prosumer challenge and subsystem variability

Operation-phase analytics is more complex in residential prosumer contexts, where onsite generation and flexible or high-load subsystems lead to intermittent and non-linear demand patterns (Ghaemi et al., 2024). Recent peer-reviewed surveys on AI-enabled digital twins and data-driven energy management in buildings indicate that scalable applications such as forecasting and operational analytics require reliable data pipelines and consistent contextual information. Despite this, many residential studies continue to focus on component-level optimization, such as PV self-consumption, often without addressing the construction of a causally consistent whole-building operational baseline (compared against persistence and linear regression reference models) supported by sub-metered data (Xie et al., 2023).

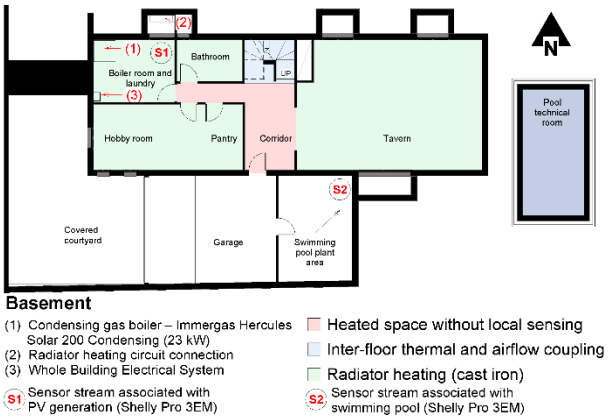
Recent residential short-term forecasting studies provide the closest benchmark context for this deployment-oriented case study. Kychkin & Chasparis (2025) conduct a broad comparative analysis of hourly and day-ahead load forecasting models for a small residential energy community, demonstrating that computationally efficient persistence-based and autoregressive models can match or outperform more complex deep learning approaches on aggregate community load. Popławski, Dudzik & Szeląg (2023) apply LASSO regression and Random Forest to forecast the net electricity balance of a single prosumer household with rooftop PV, using short training windows and both endogenous and exogenous scenarios. Neither study jointly addresses sub-metered high-load subsystem telemetry, explicit causal versus non-causal imputation policy, or a leakage-safe multi-configuration evaluation design under deployment-realistic missing-data conditions. The gap addressed here is therefore narrower than a general residential forecasting gap: this study examines one-hour-ahead operational baselining with sub-metered pool-circuit and PV telemetry under systematically varied imputation strategies and a chronologically enforced train/test split designed to reflect realistic deployment constraints.

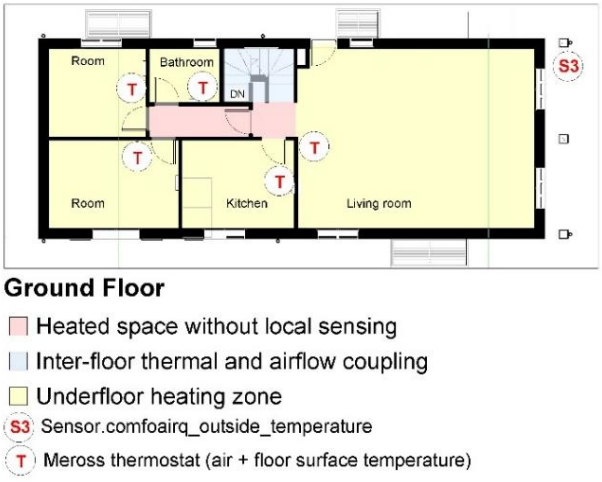
Methodology

This study implements a monitoring-oriented operation-phase digital twin analytics pipeline for operational energy baselining in a residential prosumer context following a three-stage empirical approach: characterising the case study, deploying the data integration and analytics architecture, and constructing and evaluating the operational baseline (Krishnan et al., 2024).

Case study apparatus: the monitored villa

The testbed is a detached residential villa situated in Segrate (Milan), Italy. It consists of three floors and has numerous sensors installed throughout (Fig. 1). Built in 1993, it falls under climatic zone E, with a heating season from 15 October to 15 April and 2404 degree days. The building envelope has been upgraded to comply with contemporary efficiency standards, with a calculated design heating load $Q_{H,des}$ of ~12.9 kW (approx. 44 W/m^2) according to its last Energy Performance Certificate (EPC). In the renovated sections, the transmittance of opaque envelope elements reaches values as low as 0.18 $W/(m^2K)$, reflecting a high level of thermal insulation. The net useful floor area is 291.5 m^2, spread across three levels (underground 105.8 m², ground floor 95.9 m^2, first floor 89.8 m^2). The heated gross volume is 1087 m^3, with a heat losing area/heated volume $S/V = 0.61\ m^{-1}$. In contrast to typical residential datasets that consider the building as a black box, this facility operates as a documented micro energy system with detailed information on both the building envelope and technical systems. A swimming pool and PV array complete the site (Fig. 2).





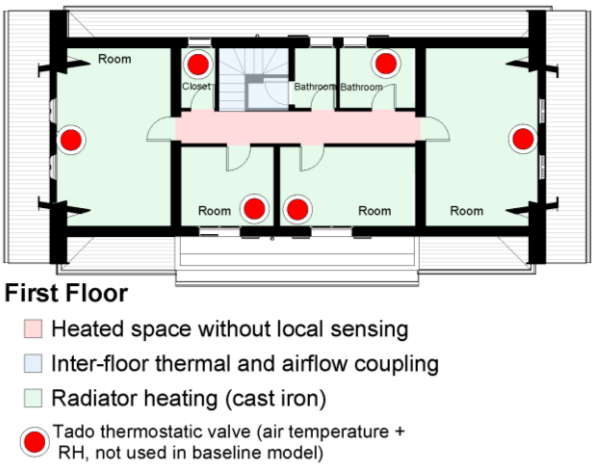


Figure 1. Villa floor plans

The outdoor swimming pool (approximately 50 m³) has a significant electrical load due to circulation and filtration. Ground-floor heating is provided by a 7-kW air-to-water heat pump with low-temperature hydronic underfloor heating, while the underground and first floors use a 23-kW condensing gas boiler with cast-iron radiators. The building also includes smart thermostatic control and a Zehnder ComfoAir Q heat-recovery ventilation unit.

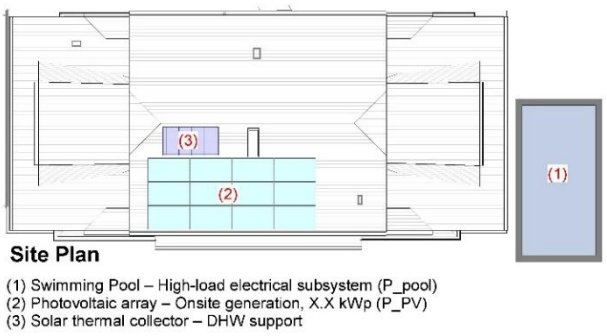


Figure 2. Villa site plan

Instrumentation consists of a heterogeneous IoT network including Shelly Pro 3EM energy meters, smart thermostats, and MVHR telemetry. The operational boundary is defined by four telemetry streams accessed through the *Home Assistant* platform: whole-villa electrical power *P_villa*, PV electrical power *P_PV*, pool circuit electrical power *P_pool*, and outdoor air temperature *T_outdoor*. Streams were retrieved as historical exports and processed in Python over a one-year period from 18 January 2025 to 18 January 2026, resampled to $\Delta t = 60$ minutes. Telemetry includes discontinuities and missing periods consistent with in-use residential operation, motivating explicit missing-data handling in the analytics pipeline.

Operation-phase digital twin architecture

Figure 3 shows the implemented architecture, in which Shelly and Zehnder sensors are integrated via the Home Assistant gateway and processed through a Python analytics pipeline. Sensor-to-subsystem association is maintained through a curated metadata configuration table that maps sensor identifiers to physical subsystems. The implemented architecture links the villa's IoT telemetry streams to a curated subsystem mapping within the analytics layer, enabling traceable association between sensor streams and operational building subsystems. This architecture supports traceable monitoring and baseline construction only; it does not implement bidirectional control, online actuation, or formal BIM-linked semantic integration.

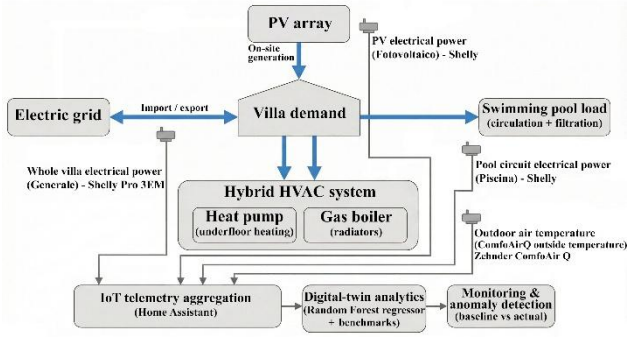


Figure 3. Energy system architecture and data pipeline integrating onsite generation, high-load subsystems, and hybrid HVAC operation with an IoT aggregation layer (Home Assistant) and a baseline-analytics module.

The data processing pipeline aligns asynchronous PV and whole-villa measurements to the nearest hourly timestamp, constructs datasets in multiple configurations based on feature availability (Table 1), imputes data using the noted missing-data policy, and generates engineered features to capture daily and seasonal patterns related to occupancy and solar exposure. For demand-only configurations, the feature set comprises the current demand reading, 1h, 24h, and 168h lags of demand, cyclic hour encodings (sine and cosine), day of week, month, weekend and public holiday indicators, a grid-export flag, and an imputation flag for the demand stream. For telemetry configurations, the feature set additionally includes the current readings and 1h, 24h, and 168h lags of P_PV, P_pool, and T_outdoor, together with per-stream imputation flags, yielding a feature vector of 28 variables for keep-imputed telemetry configurations and a reduced set for clean-only configurations where imputation flags are excluded.

The two missing data policies are as follows. For the first Ax series (causal imputation): missing values are imputed using only information available up to time t (bounded forward-fill plus signal-specific fallbacks):

$$\tilde{x}_t = x_t \quad (1)$$

if observed; otherwise

$$\tilde{x}_t = \tilde{x}_{t-1} \quad (2)$$

when the gap length $\leq H$, else

$$\tilde{x}_t = g(x,signal) \quad (3)$$

For example, PV approaches zero at night and the temperature was set based on an existing climate database. Signal-specific forward-fill bounds H were applied to each stream based on its physical rate of change. For P_villa, P_PV, and P_pool, H was set to 6 hours, reflecting the operational volatility of electrical signals where longer imputation would introduce implausible values. For T_outdoor, H was set to 24 hours, where temperature varies slowly enough that a longer fill window remains physically reasonable. Gaps exceeding these bounds fall back to the signal-specific default g(x, signal) defined in Eq. 3.

For the second Bx series (offline interpolation upper bound): for gaps up to 2 h, missing points are filled by linear interpolation between the nearest available samples x_a and x_b:

$$\tilde{x}_t = x_a + \frac{t-a}{b-a}(x_b - x_a) \quad (4)$$

with

$$b - a \leq 2 \text{ h} \quad (5)$$

Gap-filling and interpolation are applied separately within the training and test partitions to prevent test-period information from influencing the training data. In addition, both keep-imputed and clean-only data selection policies are considered to distinguish realistic deployment conditions from an upper performance bound.

Because interpolation uses both the past and the future endpoint within a gap, Bx represents an offline upper bound; Ax uses past-only information and therefore better approximates conditions where only information available up to time t is used (no future data).

Table 1: Data Treatments

Dataset	Imputation Pipeline	Missing Data Policy	Feature Set
A1	Pipeline A: causal-fill	Keep-imputed	Demand-only
A2	Pipeline A: causal-fill	Clean-only	Demand-only
A3	Pipeline A: causal-fill	Keep-imputed	Demand + telemetry
A4	Pipeline A: causal-fill	Clean-only	Demand + telemetry
B1	Pipeline B: time-based interpolation	Keep-imputed	Demand-only
B2	Pipeline B: time-based interpolation	Clean-only	Demand-only
B3	Pipeline B: time-based interpolation	Keep-imputed	Demand + telemetry
B4	Pipeline B: time-based interpolation	Clean-only	Demand + telemetry

The analytics layer accesses the processed feature vectors and constructs short-horizon operational energy baselines using a fixed Random Forest regressor (Krishnan et al., 2024).

Energy baseline prediction strategy

The analytics function seeks to develop a dynamic energy baseline model that estimates what the building's total consumption should be under current conditions based on historical behaviour. This is framed as a short-horizon baseline construction task (Krishnan et al., 2024). Let y_t denote the total electrical demand at time t. The baseline function f predicts the demand one hour ahead as:

$$\hat{y}_{t+1} = f(X_t) \quad (6)$$

Following the pipeline design, eight dataset configurations are constructed by combining feature availability (demand-only vs. demand augmented with PV, pool, and temperature telemetry), missing-data policy (keep-imputed vs. clean-only), and imputation strategy (Pipeline A: causal fill vs. Pipeline B: time-based interpolation). Imputation strategies are critical to ensure that missing data points or sensor errors do not introduce gaps that could adversely affect the predictive performance of the digital twin (Xie et al., 2023).

A data quality assessment of A1, the causal-fill, keep-imputed, demand-only dataset, indicates substantial operational missingness (e.g., demand missing/imputed 15.9%, PV missing 27.2%, and on-site temperature sensor missing 56.3%), leaving 38.4% fully clean rows; this motivates the explicit keep-imputed vs. clean-only policies and the imputation ablations.

These missingness rates should be understood as a property of an in-use residential monitoring environment rather than as a laboratory-grade sensing setup. In this case study, discontinuities are consistent with intermittent subsystem activity, telemetry or gateway interruptions, sensor outages, and periods in which parts of the monitoring pathway were unavailable. For the outdoor temperature stream specifically, the on-site sensor was

unreliable throughout the monitoring period, and the pipeline substituted values from a meteorological database where available; the reported 56.3% missingness rate reflects the combined proportion of rows where neither source provided a valid observation. For the PV stream, missingness appears as scattered individual hours consistent with intermittent gateway synchronisation rather than extended outage periods. This level of missingness is unusually high for a continuously monitored residential system, but it provides a challenging test of deployment-oriented baseline construction. From a deployment perspective, this means that a practically useful baseline cannot assume fully clean streams and must remain interpretable under missing and reconstructed observations. For this reason, missing-data handling is treated here as a first-order design variable rather than as a secondary preprocessing detail.

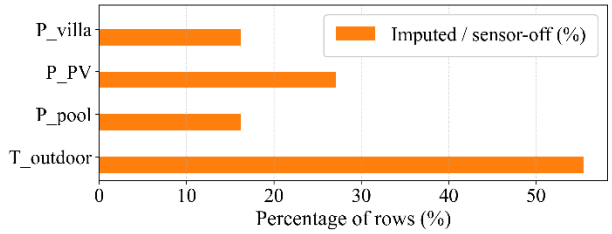


Figure 4. Missingness proxy by signal for dataset A3, expressed as the percentage of rows flagged as Imputed or sensor-off.

Pipeline A constitutes the primary deployable dataset family used for reporting results, while Pipeline B is included only as an offline upper-bound diagnostic that quantifies the performance cost of enforcing causal imputation under otherwise matched conditions; it is not presented as an independent contribution and its near-identical performance relative to Pipeline A is an expected and interpretable outcome rather than a redundancy. Data are split chronologically into 70% training and 30% testing to preserve temporal order and avoid look-ahead bias. To evaluate performance, we compare the Random Forest regressor against two benchmark models: (i) a naïve persistence model, which assumes that next-hour demand equals the current hour's demand (Eq. 7), and (ii) a Multiple Linear Regression (MLR) model trained on the same feature set to test whether simple linear relations are sufficient for short-horizon baselining.

$$\hat{y}_{t+1} = y_t \tag{7}$$

Together, these establish a minimum performance floor that must be exceeded to demonstrate predictive value (Zhou et al., 2024). A Random Forest regressor is used as a fixed learner to capture non-linear interactions between weather, prosumer subsystems, and temporal patterns (Krishnan et al., 2024). We keep the model class fixed (Random Forest) across all experiments to isolate dataset effects. Hyperparameters were tuned on training data only using *RandomizedSearchCV* with *30 iterations* and negative RMSE as the scoring metric, applied via *TimeSeriesSplit* with four folds to avoid look-ahead bias. The search space *covered n_estimators* in the range *300–1100*, *max_depth {None, 10, 20, 40, 60}*, *min_samples_split 2–30*, and *min_samples_leaf 1–20*. The resulting best hyperparameters were frozen per dataset configuration, and the test set was used only once for final evaluation.

Algorithm selection and validation

The Random Forest regressor is adopted as the primary learner because it can capture non-linear relationships in medium-sized tabular time-series data and remains robust with one year of operational observations. The learner is kept fixed across dataset variants to isolate the effect of dataset construction rather than algorithm choice. Feature importances are reported using impurity-based importance, defined as the normalized total reduction in split criterion across trees as implemented in scikit-learn (Pedregosa et al., 2011); these values should not be interpreted as equivalent to Pearson correlation. Performance is evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Tree-based ensembles are widely reported as strong baselines for tabular prediction under irregular target behaviour (Grinsztajn, Oyallon & Varoquaux, 2022).

Results

The Random Forest regressor was benchmarked against both persistence and MLR on the chronological test set, achieving an 8.4% RMSE reduction versus persistence (629.26 W → 576.50 W) and a 5.9% reduction versus MLR (612.0 W → 576.5 W) on A3. Feature importance analysis revealed that sub-metered PV and pool data were the most influential predictors. These findings demonstrate the practical utility of the monitored villa testbed and the proposed pipeline for short-horizon operational baselining in this residential prosumer case study.

Direct RMSE comparison across residential forecasting studies should be interpreted cautiously because buildings, climates, target ranges, horizons, and feature availability differ substantially; for this reason, the present study is positioned as a deployment-oriented case study rather than as a universal benchmark. The full one-year time series of all four input signals alongside whole-building demand is not shown here due to space constraints; signal quality, missingness patterns, and the operational range of each stream are characterised in the Methodology section and summarised in Fig. 4.

Fig. 5 presents the Pearson correlation heatmaps of the hourly processed datasets under causal fill, time-based interpolation, and clean-only selection, focusing on the relationships between subsystem telemetry and whole-villa electricity demand.

PV generation shows the strongest negative linear association with whole-building demand, which is expected because it reflects onsite generation rather than demand-side consumption. Outdoor temperature also shows a negative association consistent with seasonal heating behaviour. Pool-circuit power exhibits a weaker linear correlation, but this does not imply low predictive value because pool operation is intermittent and state-like rather than smoothly linear.

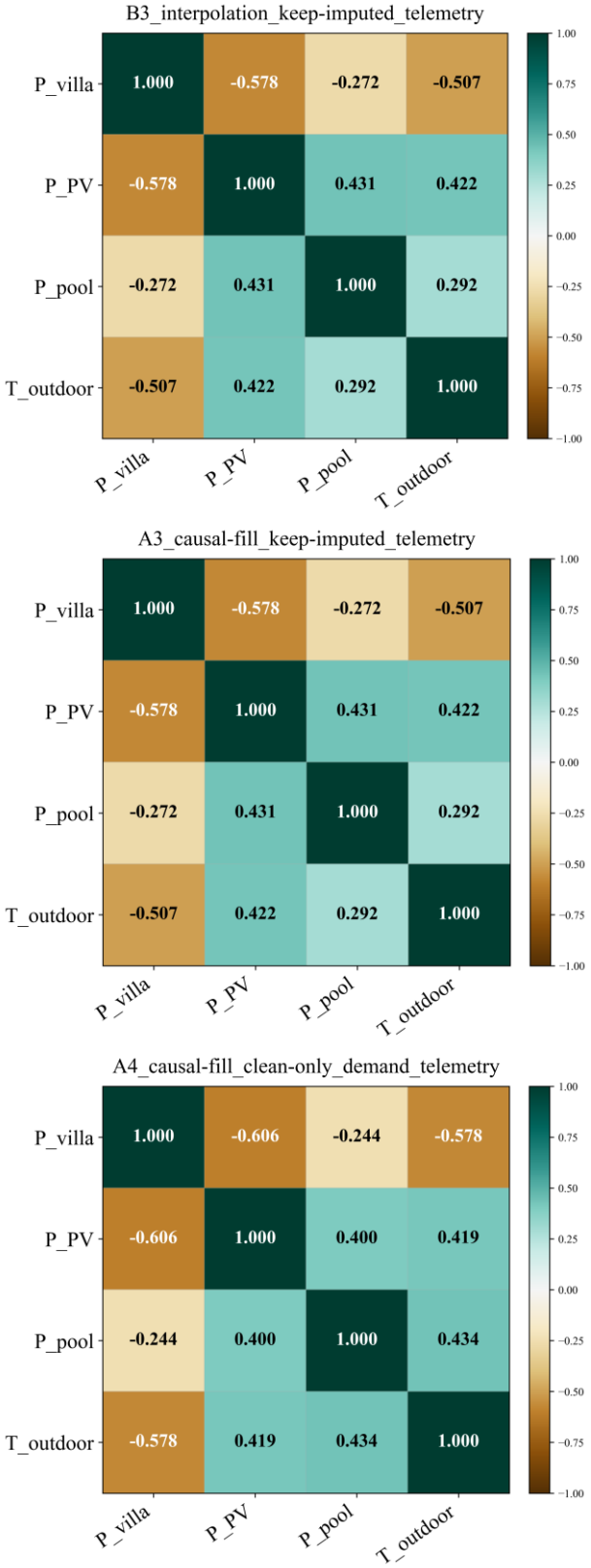


Figure 5. Pearson correlation heatmaps for selected processed datasets: B3 (top), A3 (middle), and A4 (bottom).

For this reason, feature-importance results and Pearson correlation should be interpreted as complementary rather than equivalent diagnostics. The near-identical correlation matrices for A3 and B3 are expected: in both configurations missing values are retained through imputation rather than removed, so the underlying observed data structure is identical and only the imputed segments differ; since these segments are small relative to the full dataset and consistent with the surrounding trend, the global linear association structure remains effectively unchanged. In contrast, the A4 and B4 configurations show noticeably different correlation values because clean-only filtering removes imputed rows and thereby alters the sample composition and temporal coverage. The maintained metadata mapping is used only to preserve subsystem traceability within the analytics layer. Correlation patterns are broadly similar across the compared datasets, while any differences introduced by clean-only filtering should be interpreted cautiously because dataset coverage and evaluated timestamps change across configurations.

Operational baseline model performance

Table 2 summarises model performance on the chronological test set (30% of the available data; 2529 samples for keep-imputed datasets) and reports Random Forest RMSE alongside those from the persistence and MLR benchmarks, showing that the datasets using the "Keep-imputed" data policy alongside "Demand + telemetry" features outperformed the others. Note that A4 and B4 have substantially lower usable hours than the other datasets, as reflected in the test-row counts reported in Tables 2 and 3, partially accounting for their decreased performance. This reduction in usable hours is not only a sample-size issue. Clean-only filtering also changes which operating regimes remain in the evaluated subset and therefore changes the task itself. In A4 and B4, the retained rows no longer represent the same full operational variability as the keep-imputed datasets. As a result, the predictive advantage of the richer feature set and the non-linear flexibility of Random Forest can be reduced, while a simpler model such as MLR may become comparatively more competitive on the filtered subset. This regime shift also accounts for the MAE results in Table 3, where RF underperforms MLR in A4 and B4; on a filtered subset with reduced operational variability, the non-linear flexibility of Random Forest provides no advantage over a linear model and the smaller sample size limits effective hyperparameter tuning. The eight configurations test three directional hypotheses embedded in the pipeline design. Keep-imputed configurations should outperform clean-only counterparts by preserving operational variability that filtering removes. Telemetry configurations should outperform demand-only configurations by providing subsystem-level predictive signal that demand history alone cannot capture. Pipeline A and B should converge under keep-imputed conditions because the imputation method has limited influence on the global data structure when imputed values are retained. Tables 2 and 3 confirm all three patterns.

Table 2: Summary of results – RMSE

Dataset	Test rows	RF RMSE (kW)	Persistence RMSE (kW)	MLR RMSE (kW)
A1	2529	0.597	0.629	0.634
A2	2118	0.652	0.681	0.655
A3	2529	**0.576**	0.629	0.612
A4	939	0.831	0.859	0.804
B1	2529	0.596	0.629	0.634
B2	2118	0.652	0.681	0.655
B3	2529	**0.573**	0.629	0.612
B4	939	0.831	0.859	0.804

We additionally report MAE (Table 3) to indicate typical behaviour error, while the RMSE values indicate the model sensitivity to extreme conditions.

Table 3: Summary of results – Mean Absolute Error

Dataset	Test rows	RF MAE (kW)	Persistence MAE (kW)	MLR MAE (kW)
A1	2529	0.366	0.378	0.426
A2	2118	0.431	0.440	0.473
A3	2529	**0.362**	0.378	0.409
A4	939	0.633	0.618	0.620
B1	2529	0.365	0.378	0.426
B2	2118	0.431	0.440	0.473
B3	2529	**0.361**	0.378	0.409
B4	939	0.633	0.618	0.620

The Random Forest regressor trained on A3 achieved MAE = 0.362 kW and RMSE = 0.576 kW, representing the best trade-off between accuracy and deployment realism among the tested configurations, since B3 achieves marginally lower error but relies on non-causal interpolation that is not available at deployment time. The building exhibits a skewed demand distribution with a median of 0.76 kW and a mean of 1.04 kW driven by occasional high-load and PV export events, making the RMSE of 0.576 kW a substantial fraction of typical instantaneous demand. This indicates that the baseline is suited to detecting large residual deviations consistent with equipment faults or sustained anomalous consumption but may not meet the accuracy requirements of fine-grained measurement and verification applications. In this paper, the baseline is intended as an operational reference for residual-based monitoring and measurement-oriented comparison, not as a full physical emulator of building behaviour. Its practical value lies in improving over persistence and MLR under leakage-safe and deployment-realistic conditions.

Adding telemetry (A3 vs A1) reduces RMSE by ~3.5%, while RF improves over persistence and MLR by 8.4% and 5.9%, respectively, quantifying the practical benefit of subsystem instrumentation under deployment-like constraints.

A3 and B3 show nearly identical results (RMSE ≤ 0.5%). The cost of enforcing deployable causality is therefore modest in this case study. Across all keep-imputed configurations, the Random Forest regressor outperforms the persistence benchmark, confirming predictive value beyond trivial extrapolation. Feature importance analysis ranks P_pool (32%), recent demand history (28%), P_PV (19%), and T_outdoor (12%) as the dominant explanatory variables, with temporal features making up the balance across all keep-imputed configurations. These rankings are consistent across all keep-imputed configurations; demand-only configurations assign full importance to demand history and temporal features by construction. The prominence of pool-circuit telemetry in impurity-based feature importance, despite its weaker Pearson correlation with demand, is plausible in this setting because the pool signal acts as a regime marker for intermittent switching behaviour and interacts non-linearly with other predictors. In other words, a feature can contribute strongly to tree-based prediction even when its standalone linear association with the target is modest. It should also be noted that impurity-based importance is known to inflate scores for intermittent or high-variance signals because such features create frequent and uneven splits across trees; the reported importance values should therefore be interpreted as indicative of predictive relevance within this specific Random Forest configuration rather than as absolute measures of physical influence.

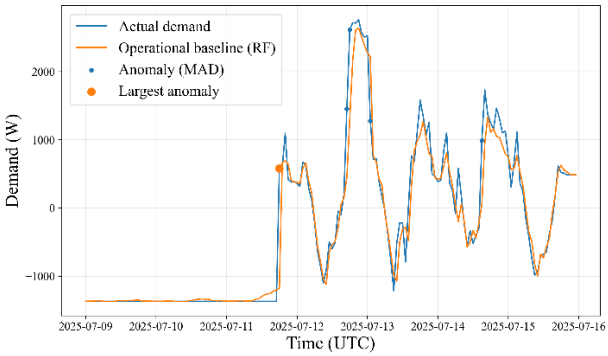


Figure 6. Sample week showing actual demand vs operational baseline (RF with A3 data configuration)

Fig. 6 illustrates a representative week of measured demand, and the one-hour-ahead operational baseline produced by the best-performing A3 configuration using Random Forest. The baseline reproduces daily structure and prosumer-specific effects such as PV self-consumption periods and pool-pump cycling. A prominent deviation is consistent with a documented circulation-pump fault due to voltage instability (resolved by installing a stabiliser in July 2025); such episodes are detected from the baseline residual using the MAD-based threshold defined in Eq. 9 and are flagged for separate post-hoc analysis. They are not removed from the test set used for RMSE and MAE

computation, and the current study does not implement an online retraining loop in which flagged anomalies are used to update the fitted baseline.

Anomalies are detected from the baseline residual rather than from raw demand. Let

$$r_{t+1} = y_{t+1} - \hat{y}_{t+1} \quad (8)$$

where y_{t+1} is the measured demand and $\hat{y}_{t+1}$ is the Random Forest baseline prediction. We compute the median residual $m = \text{median}(r)$ and $MAD = \text{median}(|\, r - m \,|)$, and flag time step $t + 1$ as anomalous if

$$|\, r_{t+1} - m \,| > K \cdot MAD \quad (9)$$

In this study we use $K = 9$, selected via a threshold sweep to reduce spurious detections during stable operation.

Conclusions

This study demonstrates that deployment-realistic missing-data handling and sub-metered subsystem telemetry are first-order determinants of operational baseline quality in residential prosumer buildings. The monitored villa testbed and the eight-configuration ablation show that imputation strategy and data policy shape predictive performance as much as feature selection, and that the cost of enforcing causal deployability is modest when imputed values are retained.

Key findings

Sub-metered subsystem telemetry, especially PV and pool-circuit power, improves predictive baselines when used within a controlled and leakage-safe dataset design. Baseline quality is strongly shaped by feature availability and missing-data policy rather than algorithm choice alone. Under keep-imputed causal conditions, A3 achieved the lowest deployable RMSE (0.576 kW). The A3 baseline residuals successfully flagged the documented circulation-pump fault in July 2025, confirming practical anomaly screening capability under in-use conditions.

Limitations and future work

Occupancy is not directly measured, so the current feature set relies on temporal proxies and may not distinguish between unoccupied and low-activity conditions. The present study is limited to monitoring and baseline construction on a single residential testbed. Accordingly, the current implementation remains an analytics-oriented operational representation rather than a closed-loop control system or a formally BIM-integrated workflow. Future work should validate the approach across multiple residential buildings, enrich behavioural and occupancy context, and extend the current offline pipeline toward continuous deployment, supervisory diagnostics, and optional richer semantic alignment were justified by the monitoring objective.

References

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Bergés, M., Culler, D., Gupta, R.K., Kjærgaard, M.B., Srivastava, M., & Whitehouse, K. (2018) Brick : Metadata schema for portable smart building applications, Applied Energy, 226, pp. 1273–1292.

Ghaemi, A., Rezgui, Y., Petri, I., Beach, T., & Ghoroghi, A. (2024) AI and Digital Twin Applications in Building Energy Management: A State-of-the-Art Review, IEEE Access, 12, pp.1-11.

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022) Why do tree-based models still outperform deep learning on typical tabular data? in Advances in Neural Information Processing Systems.

Krishnan, S., Divakaran, A., Rethnam, O.R., Markarian, E., Qiblawi, S., Thomas, A. & Azar, E. (2024) Evaluating the capabilities of surrogate modeling techniques in predicting hourly building energy consumption, in EC3 Conference 2024, 5, European Conference on Computing in Construction.

Kychkin, A., & Chasparis, G.C. (2025) Hourly Short Term Load Forecasting for Residential Buildings and Energy Communities. arXiv preprint arXiv:2501.19234.

Mengual Torres, S. G., Tzuc, O. M., Aguilar-Castro, K. M., Tellez, M. C., Sierra, J. O., Cruz, A. D. R. C. Y., & Barrera-Lao, F. J (2022) Analysis of Energy and Environmental Indicators for Sustainable Operation of Mexican Hotels in Tropical Climate Aided by Artificial Intelligence, Buildings, 12(8), p. 1155.

Oulefki, A., Kheddar, H., Amira, A., Kurugollu, F., Himeur, Y., & Bounceur, A., (2025) Innovative AI strategies for enhancing smart building operations through digital twins: A survey, Energy and Buildings, 335, pp. 115567–115567.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011) Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, pp. 2825–2830.

Popławski, T., Dudzik, S., & Szeląg, P. (2023) Forecasting of Energy Balance in Prosumer Micro-Installations Using Machine Learning Models. Energies, 16(18), 6726.

Xie, X., Merino, J., Moretti, N., Pauwels, P., Chang, J. Y., & Parlikad, A. (2023) Digital twin enabled fault detection and diagnosis process for building HVAC systems, Automation in Construction, 146, p. 104695.

Zhou, F., Yang, C. & Wang, Z. (2024) Prediction of building energy consumption for public structures utilizing BIM-DB and RF-LSTM, Energy Reports, 12, pp. 4631–4640.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A DECENTRALIZED DIGITAL TWINNING APPROACH FOR ASSESSING CIVIL ENGINEERING STRUCTURES

Kosmas Dragos, Thamer Al-Zuriqat, and Kay Smarsly
Institute of Digital and Autonomous Technology, Hamburg University of Technology,
Blohmstraße 15, 21079 Hamburg, Germany

Abstract

In wireless structural health monitoring (SHM), the focus is usually placed on reducing the power-consuming data transmission for preserving the limited resources of wireless sensor nodes. In this direction, embedded computing has been leveraged for distributing SHM data analysis tasks to the sensor nodes instead of transmitting raw data to centralized servers. This paper reports on decentralizing SHM tasks using reduced-order numerical models of structures, embedded into wireless sensor nodes as decentralized digital twins that use sensor data to assess the structural condition. The applicability of the proposed approach is demonstrated via laboratory validation tests.

Introduction

The traction that wireless structural health monitoring (SHM) has been gaining in structural maintenance since the end of the 20th century has mainly been attributed to the drawbacks of cable-based technologies for SHM, in terms of labor installation and cabling (Çelebi, 2002). Notwithstanding the significant benefits of wireless technologies, non-negligible drawbacks have been recognized already from early wireless SHM approaches, related to low transmission reliability and limited power autonomy (Lynch and Loh, 2006). The recent growing trends towards digitalization, fostering state-of-the-art technological concepts, such as Internet-of-Things systems (Al-Zuriqat et al., 2025), has brought wireless SHM once again to the forefront, frequently placed within the framework of emerging computing paradigms, such as embedded systems design and edge computing (Buckley et al., 2021).

Considerable research effort has been invested in tackling the weakest point of wireless SHM systems: The reliance on batteries, which limits the power autonomy. In this direction, the main schools of thought have focused on either integrating hardware components capable of generating power or reducing in-network wireless transmission both among wireless sensor nodes and between wireless sensor nodes and centralized servers. The former school of thought promotes the installation of energy harvesters for converting energy from ambient sources (solar, wind) to power, or – in some cases – exploiting the very vibration of structures for charging wireless sensor nodes (Park et al., 2008). Nevertheless, energy harvesting does not impose fundamental changes in the design and operation of wireless SHM systems. The latter school of thought builds upon embedded (or edge) computing, which aims to convert wireless sensor nodes from mere data collection devices to quasi-independent platforms capable of autonomously executing SHM tasks (Dragos and Smarsly, 2015). In this context, embedded/edge computing effects a decisive change on the design of wireless SHM systems, as well as on the overall SHM strategies.

Embedded computing in wireless SHM has been studied extensively over the last three decades. Seminal work proving the superiority of embedded computing in SHM over centralized approaches, in terms of energy efficiency, dates back to the beginning of the millennium with the works of Mitchel et al. (2000), Spencer Jr. et al. (2003), Lynch et al. (2004) and Wang et al. (2007), paving the way for introducing a plethora of embedded computing approaches, mostly from 2008 onwards. Common ground among most embedded computing approaches has been the extraction of features from sensor data (e.g. structural response data or environmental data) indicative of a “system state” (essentially, structural condition), which enables either identifying structural parameters or detecting structural damage. Examples include the extraction and tracking of modal parameters, reported by Zimmerman et al. (2008), the estimation of the damage location vector, introduced by Nagayama et al. (2009), and the computation of the transmissibility function between mobile sensor nodes, presented by Yi et al. (2010). In recent work, Quqa et al. (2021) have applied embedded damage detection based on the interpolation error when approximating deflection shapes using smooth functions. Other approaches for extracting features from sensor data have been based on the random decrement method (Friedmann et al., 2010), auto-regressive modeling (Kijewski-Correa and Su, 2009), and on cable-force estimation (Spencer Jr. and Cho, 2011). A special class of embedded computing approaches for wireless SHM has centered around machine learning, employed for system identification (Peckens et al., 2011), damage detection (Avci et al., 2018) and for fault diagnosis in the time domain (Smarsly and Law, 2014), in the frequency domain (Dragos et al., 2016) and in the coupled time-frequency domain (Fritz et al., 2022). Finally, embedded computing has been leveraged for eliminating synchronization errors, as part of on-board estimation of modal parameters (Dragos et al., 2018).

The literature review on embedded computing for wireless SHM reveals a trend in adopting data-driven analysis methods for on-board computations. However informative, data-driven analysis methods exhibit inherently limited information potential, due to only implicitly considering the physics dominating the structural behavior. In other words, the outcomes of data-driven analysis methods are essentially mathematical quantities, whose physical meaning cannot always be guaranteed. In this context, this paper reports on an embedded computing approach aiming to render on-board analysis "physics-informed". In particular, drawing from the state-of-the-art paradigm of digital twinning, the proposed approach focuses on decentralizing digital twins in an attempt to enable wireless sensor nodes to conduct complex SHM tasks on board with enhanced decision-making capabilities. Preliminary work on employing physics-based modeling for embedded computing has been conducted by the authors (Dragos and Smarsly, 2017; Dragos and Smarsly, 2024). Capitalizing on the preliminary work, the decentralized digital twinning approach builds upon embedding physics-based models that are computationally manageable and amenable to independent analysis on each wireless sensor node. The proposed approach is validated in laboratory tests on a shear-frame structure, demonstrating the capabilities of the decentralized digital twins in identifying structural damage.

The remainder of the paper starts with a description of the decentralized digital twinning approach, followed by the implementation and the laboratory validation test. The paper ends with the summary and conclusions, as well as with a brief discussion on potential future research.

Decentralized digital twinning

The decentralized digital twinning approach is presented in this section. First, the two steps towards producing physics-based models suitable for embedment into wireless sensor nodes (model-order reduction and model segmentation) are illuminated. Next, the overview of the approach is described.

Model-order reduction

The limited computational resources of wireless sensor nodes render the embedment of physics-based models, frequently used in structural engineering (mostly, finite element models), prohibitive. In an attempt to alleviate the computational burden of finite element (FE) models, model-order reduction (MOR) is employed. In civil engineering, the reduction usually involves representing physical structures through features, capable of capturing the structural behavior. The features are estimated either by transforming elaborate FE models, e.g., from the physical domain to the modal domain (representing a physical structure via its modal parameters) or by transforming structural response data, collected from a physical structure, from the time domain to the frequency domain (representing the structural behavior in terms of dominant frequency components, instead of raw structural response data). However, most MOR methods are associated with transforms that result in reduced-order models hardly representable in physical-coordinate systems. Consequently, coupling such reduced-order models with structural response data, corresponding to degrees of freedom with physical coordinates, is challenging.

To tackle the aforementioned challenge, the decentralized digital twinning approach, presented in this paper, employs a MOR method that partially retains representability in physical coordinates. Specifically, the dynamic condensation method is applied, which has been widely used in case studies of component mode synthesis (Bathe and Dong, 2014). Considering a physical structure, represented by a FE model of N degrees of freedom, the starting point for the dynamic condensation is the formulation of the equation of motion as follows:

$$\mathbf{M}^{N\times N}\ddot{\mathbf{x}}^{N\times 1}(t)+\mathbf{C}^{N\times N}\dot{\mathbf{x}}^{N\times 1}(t)+\mathbf{K}^{N\times N}\mathbf{x}^{N\times 1}(t)=\mathbf{p}^{N\times 1}(t). \quad (1)$$

The quantities $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$, represent the mass, damping, and stiffness matrices, respectively. The vectors $\ddot{\mathbf{x}}$, $\dot{\mathbf{x}}$ and $\mathbf{x}$, hold the acceleration, velocity, and displacement response data, the variable t denotes time, and $\mathbf{p}$ is the load vector. The N degrees of freedom are partitioned into two sets, the "retained" (r set) and the "condensed" (c set) degrees of freedom. Since the goal of MOR is to produce reduced-order models intuitive for SHM, the r set contains degrees of freedom measured by the SHM system, and the c set contains the rest of the degrees of freedom of the FE model. Thereupon, each of the matrices $\mathbf{A}$ in Equation 1 (with $\mathbf{A}$ = {$\mathbf{M}$, $\mathbf{C}$, $\mathbf{K}$}) and each of the vectors $\mathbf{a}$ (with $\mathbf{a}$ = {$\ddot{\mathbf{x}}$, $\dot{\mathbf{x}}$, $\mathbf{x}$}) are rearranged as shown below:

$$\mathbf{A}=\left[\begin{array}{c|c}\mathbf{A}_{cc} & \mathbf{A}_{cr}\\ \hline \mathbf{A}_{rc} & \mathbf{A}_{rr}\end{array}\right] \quad \mathbf{a}=\left\{\begin{array}{c}\mathbf{a}_c\\ \hline \mathbf{a}_r\end{array}\right\}. \quad (2)$$

Next, the submatrices of the c set are used for computing (i) the static-constraint modes $\mathbf{\Psi}_{cr}=(-\mathbf{K}_{cc})^{-1}\mathbf{K}_{cr}\mathbf{I}_{rr}$ and (ii) the eigenmodes $\mathbf{\Phi}_{cv}=[\boldsymbol{\varphi}_{cj}]$ ($j=1...v$) of the eigenproblem $\mathbf{K}_{cc}\boldsymbol{\varphi}_{cj}=\lambda_j\mathbf{M}_{cc}\boldsymbol{\varphi}_{cj}$, which together form the following transformation matrix $\mathbf{\Psi}_r$:

$$\mathbf{\Psi}_r=\begin{bmatrix}\mathbf{\Phi}_{cv} & \mathbf{\Psi}_{cr}\\ \mathbf{0}_{rv} & \mathbf{I}_{rr}\end{bmatrix}. \quad (3)$$

Finally, the transformation matrix is used for computing the reduced-order $r\times r$ model matrices, $\mathbf{K}_{red}$ and $\mathbf{M}_{red}$, as shown in Equation 4.

$$\mathbf{K}_{red}=\mathbf{\Psi}_r^T\mathbf{K}\mathbf{\Psi}_r \qquad \mathbf{M}_{red}=\mathbf{\Psi}_r^T\mathbf{M}\mathbf{\Psi}_r \quad (4)$$

As regards the damping matrix, due to the generally low vibration excitation of typical SHM applications, the Rayleigh-damping approximation is adopted, which is frequently applied in civil engineering. As a result, the damping matrix is considered proportional to the stiffness matrix ($\mathbf{C}_{red}=\beta\mathbf{K}_{red}$) with β representing the Rayleigh coefficient. Proportionality with the mass matrix, also applied in civil engineering, is not considered, in this case, for simplicity. Upon completing MOR, the reduced-order model is segmented to enable its use on board the wireless sensor nodes, as will be shown in the next subsection.

Model segmentation

The dynamic condensation produces matrices of the following layout:

$$\mathbf{M}_{red} = \left[\begin{array}{c|c} \mathbf{M}_{rr} & \mathbf{M}_{rv} \\ \hline \mathbf{M}_{vr} & \mathbf{I}_{vv} \end{array}\right], \quad \mathbf{K}_{red} = \left[\begin{array}{c|c} \mathbf{K}_{rr} & \mathbf{0}_{rv} \\ \hline \mathbf{0}_{vr} & \mathbf{\Lambda}_{vv} \end{array}\right]. \quad (5)$$

In Equation 5, $\mathbf{I}$ is the unitary matrix, $\mathbf{0}$ is a zero matrix, and $\mathbf{\Lambda}$ is a diagonal matrix holding the v eigenvalues from the eigenproblem, previously shown. Moreover, the vectors assume the form $\mathbf{x}_{red} = [\mathbf{x}_c, \mathbf{q}_v]^T$, where $\mathbf{q}$ is the vector of the generalized coordinates, corresponding to the v eigenvalues. By substituting the matrices and vectors of the reduced-order model into Equation 1 and by recasting the matrix-form equation as a set of two equations, the following formulation is obtained:

$$\begin{aligned} &\mathbf{M}_{rr}\ddot{\mathbf{x}}_r + \mathbf{M}_{rv}\ddot{\mathbf{q}}_v + \mathbf{K}_{rr}\left(\mathbf{x}_r + \beta\dot{\mathbf{x}}_r\right) = \mathbf{p}_r \\ &\mathbf{M}_{vr}\ddot{\mathbf{x}}_r + \ddot{\mathbf{q}}_v + \mathbf{\Lambda}_{vv}\left(\mathbf{q}_v + \beta\dot{\mathbf{q}}_v\right) = \mathbf{p}_v \end{aligned} \quad (6)$$

Since the top equation contains structural response data from the r set of degrees of freedom, this equation forms the basis for the rest of the computations. The first step towards model segmentation involves eliminating terms from the equation that are unknown. For example, in typical vibration-based SHM practices, acceleration response data is collected, while measuring velocities and displacements is relatively rare. Furthermore, estimating velocities and displacements via double integration is hardly accurate, due to the inevitable contamination of acceleration response data with noise and measurement errors. To remove the "unknown" velocities and displacements, the top part of Equation 6 is transformed into the frequency domain via the fast Fourier transform (FFT):

$$\mathbf{M}_{rr}\ddot{\mathbf{X}}_r + \mathbf{M}_{rv}\ddot{\mathbf{Q}}_v + \mathbf{K}_{rr}\left(\mathbf{X}_r + \beta\dot{\mathbf{X}}_r\right) = \mathbf{P}_r. \quad (6)$$

The FFT representations of acceleration, velocity and displacement response data are denoted as $\ddot{\mathbf{X}}$, $\dot{\mathbf{X}}$, and $\mathbf{X}$, respectively. The following equivalences between the FFT representations for each frequency component f, with respect to the natural frequency ω ($\omega = 2\pi f$), are considered:

$$\dot{\mathbf{X}}_r = \frac{\ddot{\mathbf{X}}_r}{\omega \mathrm{i}} + \varepsilon_1, \qquad \mathbf{X}_r = -\frac{\ddot{\mathbf{X}}_r}{\omega^2} + \varepsilon_2. \quad (7)$$

In Equation 7, ε_1 and ε_2 are error functions of the natural frequency, accounting for the inaccuracies stemming from the approximative character of the FFT and the effects of incomplete sampling, spectral leakage, and noise. Using Equation 7, Equation 6 is converted as follows:

$$\left[\mathbf{M}_{rr} + \left(\frac{\beta}{\omega \mathrm{i}} - \frac{1}{\omega^2}\right)\mathbf{K}_{rr}\right]\ddot{\mathbf{X}}_r = \mathbf{P}_r. \quad (8)$$

Representing an extension to the approach presented in Dragos and Smarsly (2024), MOR in this paper is applied in a pair-wise manner, i.e. in pairs of degrees of freedom, corresponding to locations measured by the SHM system, designating a "master" (w) and a "slave" (o) degree of freedom (DOF) in each pair (r = 2). The purpose of applying pair-wise MOR is to produce reduced-order models as lightweight as possible, while minimizing in-network communication. In fact, each reduced-order model is designed to be embedded into the wireless sensor node measuring the master DOF, and use the model with locally collected acceleration response data, as well as with data collected from the slave-DOF (typically, neighboring) sensor node. As a result, the matrices $\mathbf{M}_{rr}$ and $\mathbf{K}_{rr}$ are 2×2 with the layout shown below:

$$\mathbf{M}_{rr} = \left[\begin{array}{c|c} m_{ww} & m_{wo} \\ \hline m_{ow} & m_{oo} \end{array}\right], \quad \mathbf{K}_{rr} = \left[\begin{array}{c|c} k_{ww} & k_{wo} \\ \hline k_{ow} & k_{oo} \end{array}\right]. \quad (9)$$

By substituting the expressions of Equation 9 into Equation 6, and by performing some mathematical manipulations, the top part of the 2×2 matrix-form equation of motion assumes the following form:

$$\begin{aligned} &a_{ww} \cdot \ddot{X}_w + a_{wo} \cdot \ddot{X}_o = h_w \\ &a_{ij} = \left[m_{ij} + \left(\frac{\beta}{\omega \mathrm{i}} - \frac{1}{\omega^2}\right)k_{ij}\right]. \end{aligned} \quad (10)$$

In Equation 10, h_w denotes an error term, including the quantities: $h_w = P_w - (\mathbf{m}_{wv})^T\,\ddot{\mathbf{Q}}_v - (\beta\varepsilon_1 + \varepsilon_2)\cdot(k_{ww} + k_{wo})$. Each wireless sensor node is assigned the task to use the reduced-order model and locally collected acceleration response data $\ddot{X}_w$ to estimate the slave-DOF acceleration response data ($\ddot{X}_o$), by solving Equation 10 as:

$$\ddot{X}_o = \frac{h_w - \left[m_{ww} + \left(\frac{\beta}{\omega \mathrm{i}} - \frac{1}{\omega^2}\right)k_{ww}\right]\cdot \ddot{X}_w}{\left[m_{wo} + \left(\frac{\beta}{\omega \mathrm{i}} - \frac{1}{\omega^2}\right)k_{wo}\right]}. \quad (11)$$

Equation 11 represents a "segmented" reduced-order model, capable of being implemented into an independently operating wireless sensor node.

Following up on the previous discussion, the decentralized digital twinning approach is summarized in Figure 1. The starting point for applying the approach on each wireless sensor node involves the formulation, calibration and dynamic condensation (to master-slave DOF pairs) of a FE model representing the initial condition of the structure being monitored ("undamaged" structure). The calibration is achieved using preliminary acceleration response data from all degrees of freedom measured and by applying FE model updating. Then, a family of L FE model variants, corresponding to damage scenarios, are created by perturbing a set of structural parameters $\boldsymbol{\theta}$ and reduced, following the same condensation as in the undamaged structure. Finally, using newly collected acceleration response data, Equation 11 is applied for several FE model variants.

The error term h_w in Equation 11 is generally unknown, hindering the estimation of the slave-DOF responses. To overcome the computation difficulty, the term h_w is assumed to be random Gaussian, the acceleration response data is segmented into n time windows, and Equation 11 is applied to each time window separately to statistically limit the effect of the term h_w. As a result, a vector of slave-DOF response estimates is obtained

$\tilde{\mathbf{X}}_o = [\tilde{X}_{o1}, \tilde{X}_{o2}, ..., \tilde{X}_{on}]^T$ deviating from the respective acceleration response data $\ddot{\mathbf{X}}_o = [\ddot{X}_{o1}, \ddot{X}_{o2},..., \ddot{X}_{on}]^T$ by the error terms $\mathbf{h}_w = [h_{w1}, h_{w2},...,h_{wn}]$. The fitness of each model variant is determined upon the proximity between the vector $\tilde{\mathbf{X}}_o$ to the vector $\ddot{\mathbf{X}}_o$, computed by the coefficient of determination R^2, shown below, summed over a range Ω of frequency components, separately for the real and imaginary parts of the complex-valued $\tilde{\mathbf{X}}_o$ and $\ddot{\mathbf{X}}_o$:

$$R^2 = 1 - \frac{\sum_{j=1}^{n}\left(\ddot{X}_o - \tilde{X}_o\right)^2}{\sum_{j=1}^{n}\left(\ddot{X}_o - \bar{X}_o\right)^2}. \qquad (12)$$

In Equation 12, $\bar{X}_o$ is the mean slave-DOF response from the vector $\ddot{\mathbf{X}}_o$. The decentralized digital twinning approach is implemented in a prototype wireless SHM system and validated in a laboratory test, presented in the next section.

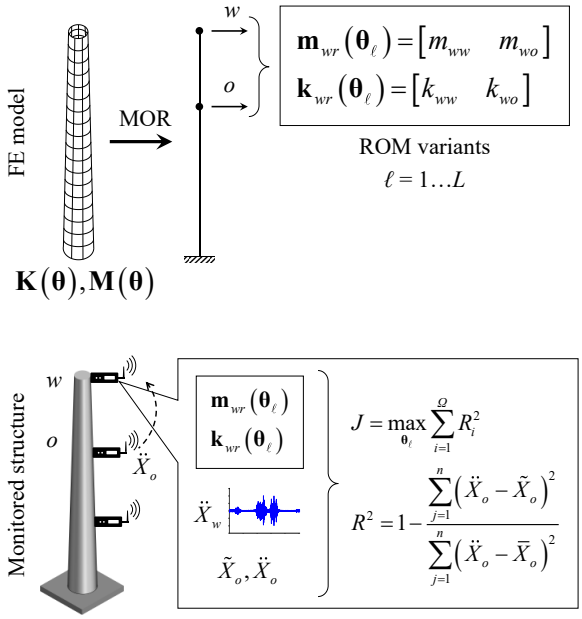


Figure 1: Overview of the decentralized digital twinning approach.

Implementation and validation

The prototypical implementation of the decentralized digital twinning approach involves the design of a relatively simple wireless SHM system, equipped with embedded physics-based models. In this section, first, the implementation details are illuminated, with emphasis on the software designed for the microcontrollers of the wireless sensor nodes. Thereupon, a small-scale laboratory test is presented, serving as proof of concept of the capability of the embedded physics-based models in identifying the structural condition.

Implementation in a prototype wireless SHM system

A key benefit of the decentralized digital twinning approach is the application of MOR in pairs of degrees of freedom, which allows a relatively minimal SHM system design, in terms of hardware. In fact, the proposed approach can be implemented even in SHM systems comprising only two wireless sensor nodes, one base station and one on-site computer (functioning as a server), as shown in Figure 2. The role of the base station and the on-site computer is secondary: The on-site computer is used for SHM system management (e.g. initialization, coordination) and data storage, and the base station acts as gateway node, enabling communication between the on-site computer and the wireless sensor nodes. Nevertheless, the wireless sensor nodes operate in a quasi-independent fashion, and the communication with the on-site computer is only performed when needed for system management and for transmitting data including results of the on-board computations for storage.

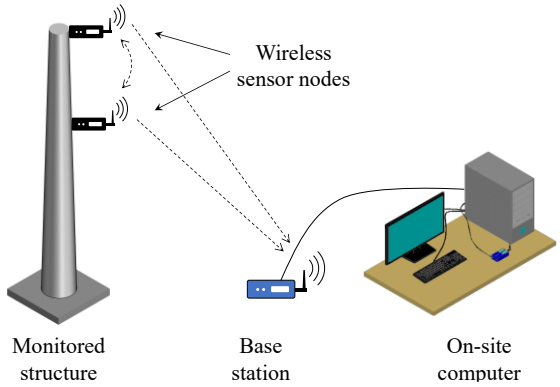


Figure 2: Layout of the prototype wireless structural health monitoring system.

Regarding the software implementation, embedded software for applying Equations 11 and 12 on board the wireless sensor nodes is exemplarily written in the Java programming language. The complexity of the software architecture is kept deliberately low to (i) preserve the limited computational resources of the sensor nodes and (ii) to facilitate the reproducibility of the code in other programming languages. An extract of the embedded software is shown in Figure 3.

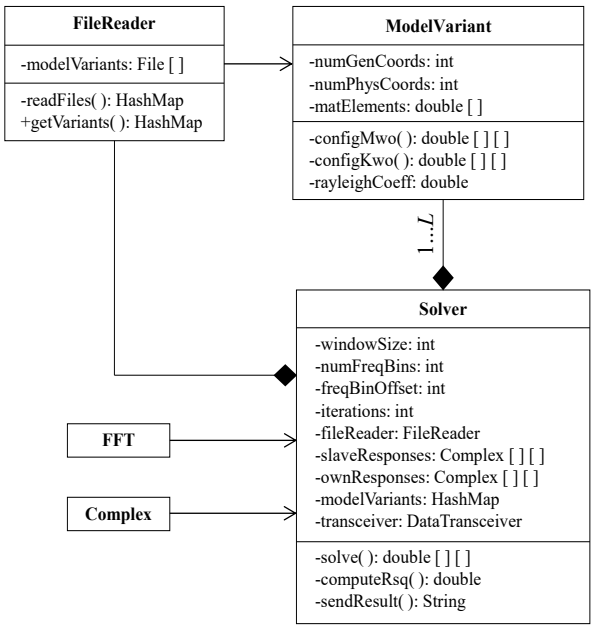


Figure 3: Extract of the embedded software.

The class `ModelVariant` realizes every reduced-order model variant, representing a damage scenario, the specifications of which are read as input streams by the `FileReader` class. Upon retrieving the responses $\ddot{\mathbf{X}}_w$ (`ownResponses` attribute) and the slave-DOF acceleration response data $\ddot{\mathbf{X}}_o$, the class `Solver` computes the estimates $\tilde{\mathbf{X}}_o$, using the method "`solve`" and the coefficient of determination for each model variant using the method "`computeRsq`". The outcome of the `Solver` class is communicated to the base station by invoking the method "`sendResult`". Finally, the FFT transform and the mathematical operations involving complex numbers are handled by the helper classes `FFT` and `Complex`, respectively. The prototype wireless SHM system is validated in a laboratory test, shown in the next subsection.

Validation test

The laboratory structure used for the validation test is shown in Figure 4. A 3-story shear-frame structure is employed, consisting of high-density-fiber (HDF) plates supported by aluminum columns. The density of the materials is assumed equal to 0.634 t/m^3 for the HDF and 2.70 t/m^3 for the aluminum, respectively. The elastic moduli are assumed equal to 3 MPa for the HDF and 70 MPa for the aluminum. Plate-to-column connections are fixed using two M3 screws per column. The dimensions of each plate are 300×200 (mm) (length × width) with a thickness of 18 mm and vertical inter-story spacing of 245 mm. The nominal cross-section dimensions of the columns are equal to 20×1.8 (mm) (width × thickness); however, the thickness of the columns is selected as "uncertain parameter" with high sensitivity and is used for calibrating the FE model. The column thicknesses, computed via model updating and used for MOR, are t_{z1} = 1.64 mm, t_{z2} = 1.79 mm and t_{z3} = 1.88 mm, for the first, second, and third story, respectively. At the base, the shear-frame structure is fixed on a 20-mm thick steel plate, which secures adequate clamping.

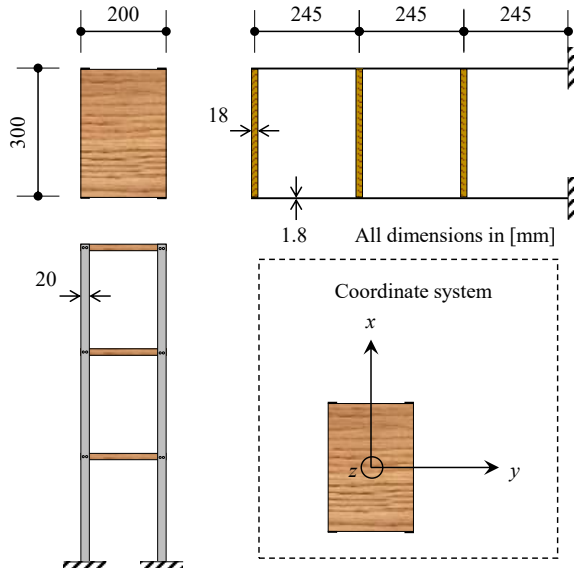


Figure 4: Laboratory shear-frame structure.

Two wireless sensor nodes are placed on the structure of type Oracle SunSPOT (Oracle Corp., 2010), featuring Java-based microcontrollers that enable the straightforward embedment of user-defined code. The software, presented in the previous subsection is embedded in the wireless sensor nodes, which are equipped with 3-axial accelerometers, capable of collecting acceleration response data at a sampling rate of 125 Hz, digitized with 8-bit resolution for a measurement range of ±2g. Despite the user-friendly interface of the SunSPOT wireless sensor nodes for embedding software, preliminary tests have revealed measurement inaccuracies, due to the low resolution of the accelerometer and the relatively high noise floor. To overcome the measurement inaccuracies, while leveraging the embedment capabilities of the SunSPOT sensor nodes, external measurement datasets are collected for each test case using high-accuracy wireless sensor nodes of type Lord Microstrain G-Link-200 (Parker Hannifin Microstrain Sensing, 2020). The external measurement datasets are subsequently deployed to the SunSPOT sensor nodes and used for applying the decentralized digital twinning approach.

The two wireless sensor nodes are placed on the first and second story, respectively, i.e. at heights of z_1 = 245 mm and z_2 = 450 mm. The sampling rate for the external measurement datasets is set equal to f_s = 128 Hz, and the measurement duration is 300 s for each test case. A window size of 2048 measurements is selected, resulting in n = 36 time windows with 50% overlap. Three test cases are considered: (i) undamaged ("intact") structure, (ii) minor damage, and (iii) major damage. For inflicting damage on the structure, the screws at the base of one column are loosened (one screw for "minor" damage and both screws for "major" damage). A family of L = 9 FE model variants are considered, including one FE model for the undamaged structure and 8 FE model variants, representing 4 different levels of damage per story for the bottom two stories. It should be noted that damage scenarios at the top story are not considered in this study, due to the negligible effect of top-story stiffness changes on the MOR of the FE model to the degrees of freedom of the bottom two stories. Following typical assumptions for shear-frame structure, the laboratory structure is modeled as a multi-DOF oscillator with one translational DOF per story in the x-axis direction. The translation in the y-axis direction is neglected due to the relatively high stiffness. Damage is simulated as stiffness reduction factors (0.9, 0.8, 0.7, 0.6) multiplied by the collective story stiffness ($4 \cdot k_c$, where k_c is the column stiffness). For selecting the range Ω of frequency components, the highest peaks in the FFT amplitude spectrum are located in each set of external measurements, and an offset of ±1 Hz is considered around each peak. The FFT amplitude spectra for the test cases (i) and (iii) are exemplarily illustrated in Figures 5 and 6. As can be seen in the FFT amplitude spectra, the effect of damage is relatively small in the second and third peak, leading to the conclusion that extending the range to include these peaks may be hardly sensitive to the damage scenario considered and are, therefore, excluded from Ω. All test cases are conducted by hammer-induced excitation. The results are summarized in Table 1.

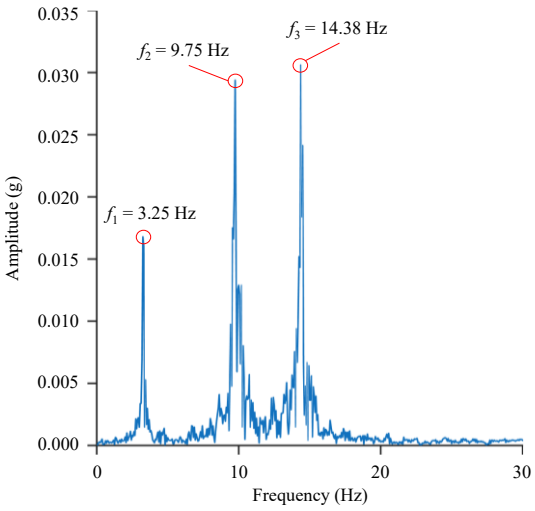


Figure 5: FFT amplitude spectrum for test case (i) (intact structure).

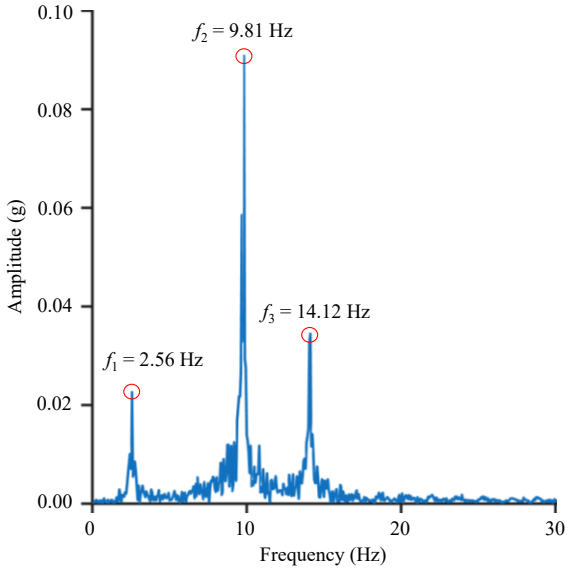


Figure 6: FFT amplitude spectrum for test case (iii) (major damage).

Table 1: $\sum R^2$ for all test cases and all model variants.

Model variant	Stiffness reduction factor	Test case		
		(i)	(ii)	(iii)
Intact	N/A	**58.86**	56.33	53.00
Damage on 1st story	0.9	58.67	56.79	54.71
	0.8	58.20	**56.91**	55.96
	0.7	57.46	56.69	56.77
	0.6	56.43	56.12	**57.13**
Damage on 2nd story	0.9	58.77	55.63	51.27
	0.8	58.37	54.41	48.64
	0.7	57.37	52.27	44.53
	0.6	55.23	48.40	37.75

The test results clearly show that the decentralized digital twinning approach yields the fittest model variant most compatible with the test case considered. In test case (i), the "intact" model variant receives the highest R^2 value, although the proximity with model variants representing minor damage is close, indicating the sensitivity of the approach to measurement inaccuracies. The fittest model variant for test case (ii) is the model with reduction factor 0.8 at the first story, which is expected, since the damage inflicted is minor. By contrast, the fitness of the significant-damage model variant with reduction factor 0.6 is prominent in test case (iii), where damage is indeed major. Nevertheless, the applicability of the decentralized digital twinning approach needs to be investigated in large-scale experiments, which will be covered in future work.

Summary and conclusions

Embedded computing has been employed in wireless structural health monitoring as a means of reducing the power-consuming transmission of large sets of structural response data. However, most embedded computing approaches rely on data-driven analysis, which yields limited information on the structural condition, due to the implicit consideration of the physics dominating the structural behavior. This paper has reported on embedding physics-based models for conducting SHM tasks directly on wireless sensor nodes, realized as decentralized digital twins. The proposed decentralized digital twinning approach relies on model-order reduction (to render the physics-based models – in practice, finite element models – computationally manageable by the sensor nodes) and model segmentation (to enable wireless sensor nodes to use the models quasi-independently). The decentralized digital twinning approach has been implemented in a prototype wireless SHM system and validated in a small-scale laboratory test, serving as proof of concept for the capability of the approach to correctly identify structural damage. Future work will focus on investigating the applicability limits of the decentralized digital twinning approach, via large-scale tests on real-world structures.

Acknowledgments

The authors gratefully acknowledge the support offered by the German Research Foundation (DFG) through grants SM 281/9-3, SM 281/33-1 and SM 281/44-1, as well as by the German Federal Ministry of Transport (BMV) under grant 01FV2059C. Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the DFG.

References

Al-Zuriqat, T., Chillón Geck, C., Dragos, K., Manolis, G. D., & Smarsly, K. (2025). Decentralized sensor fault diagnosis for wireless structural health monitoring systems using Artificial Intelligence of Things. In: 15th International Workshop on Structural Health Monitoring (IWSHM). Stanford, CA, USA, 09/11/2025.

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., and Inman, D.J. (2018). Wireless and real-time structural

damage detection: A novel decentralized method for wireless sensor networks. Journal of Sound and Vibration, 424, pp. 158-172.

Bathe, K.-J. and Dong, J. (2014). Component mode synthesis with subspace iterations for controlled accuracy of frequency and mode shape solutions. Computers and Structures, 139, pp. 28-32.

Buckley, T., Ghosh, B., and Pakrashi, V. (2021). Edge structural health monitoring (E-SHM) using low-power wireless sensing. Sensors, 21, 20, ID: 6760.

Çelebi, M. (2002). Seismic instrumentation of buildings (with emphasis on federal buildings). Technical report No. 0-7460-68170. Menlo Park, CA, USA: United States Geological Survey.

Dragos, K. and Smarsly, K. (2015). A comparative review of wireless sensor nodes for structural health monitoring. In: 7th International Conference on Structural Health Monitoring of Intelligent Infrastructure. Turin, Italy, 07/01/2015.

Dragos, K. and Smarsly, K. (2016). Distributed adaptive diagnosis of sensor faults using structural response data. Smart Materials and Structures, 25, 10, ID: 105019.

Dragos, K. and Smarsly, K. (2017). Decentralized infrastructure health monitoring using embedded computing in wireless sensor networks. In: Sextos, A. & Manolis, G. D. (eds.). Dynamic Response of Infrastructure to Environmentally Induced Loads. Pp. 183-201. Cham, Switzerland: Springer International Publishing AG.

Dragos, K. and Smarsly, K. (2024). A smart monitoring approach based on decentralized digital twins. In: 20th International Conference on Computing in Civil and Building Engineering (ICCCBE). Montréal, Canada, 08/25/2024.

Dragos, K., Theiler, M., Magalhães, F., Moutinho, C., and Smarsly, K. (2018). On-board data synchronization in wireless structural health monitoring systems based on phase locking. Structural Control and Health Monitoring, 25, 11, ID: e2248.

Friedmann, A., Mayer, D., and Kauba, M. (2010). An approach for decentralized mode estimation based on the random decrement method. Shock and Vibration, 17, 4-5, pp. 579-588.

Fritz, H., Peralta, J., Legatiuk, D., Steiner, M., Dragos, K., and Smarsly, K. (2022). Fault diagnosis in structural health monitoring systems using signal processing and machine learning techniques. In: Cury, A., Ribeiro, D., Ubertini, F., Todd, M. D. (eds.). Structural health monitoring based on data science techniques. Pp. 143-164.

Kijewski-Correa, T. and Su, S. (2009). BRAIN: A bivariate data-driven approach to damage detection in multi-scale wireless sensor networks. Smart Structures and Systems, 5, 4, pp. 415-426.

Mitchel, K., Sana, S., Liu, P., Cingirikonda, K., Rao, V.S., and Pottinger, H.J. (2000). Distributed computing and sensing for structural health monitoring systems. In: SPIE Smart Structures and Materials: Smart Electronics and MEMS. Newport Beach, CA, USA: 03/06/2000.

Nagayama, T., Spencer Jr, B.F., and Rice, J.A. (2009). Autonomous decentralized structural health monitoring using smart sensors. Structural Control Health Monitoring, 16, 7-8, pp. 842-859.

Lynch, J.P. and Loh, K.J. (2006). A summary review of wireless sensor networks for structural health monitoring. The Shock and Vibration Digest, 38, 2, pp. 91-128.

Lynch, J.P., Sundararajan, A., Law, K.H., Kiremidjian, A.S., and Carryer, E. (2004). Embedding damage detection algorithms in a wireless sensing unit for operational power efficiency. Smart Materials and Structures, 13, 4, pp. 800-810.

Oracle Corp. (2010). Sun SPOT eDEMO Technical Datasheet, 8th edition. Santa Clara, CA, USA: Sun Labs.

Park, G., Rosing, T., Todd, M.D., Farrar, C.R., and Hodgkiss, W. (2008). Energy harvesting for structural health monitoring sensor networks. ASCE Journal of Infrastructure Systems, 14, 1, pp. 64-79.

Parker Hannifin Microstrain Sensing (2020). Microstrain Sensing Product Datasheet: G-Link-200 Ruggedized Wireless Triaxial Accelerometer Node. Williston, VT, USA: Parker Hannifin Corp.

Peckens, C.A., Lynch, J.P., and Pei, J.-S. (2011). Distributed neural computations for embedded sensor networks. In: SPIE Smart Structures and Materials: Nondestructive Evaluation and Health Monitoring. San Diego, CA, USA, 03/06/2011.

Quqa, S., Giordano, P.F., Limongelli, M.P., Landi, L., and Diotallevi, P.P. (2021). Clump interpolation error for the identification of damage using decentralized sensor networks. Smart Structures and Systems, 27, 2, pp. 351-363.

Smarsly, K. and Law, K.H. (2014). Decentralized fault detection and isolation in wireless structural health monitoring systems using analytical redundancy. Advances in Engineering Software, 73, pp. 1-10.

Spencer Jr., B.F. and Cho, S. (2011). Wireless smart sensor technology for monitoring civil infrastructure: Technological developments and full-scale applications. In: 2011 World Congress on Advances in Structural Engineering and Mechanics (ASEM 11). Seoul, S. Korea, 09/18/2011.

Spencer Jr, B.F., Ruiz-Sandoval, M.E., and Kurata, N. (2004). Smart sensing technology: Opportunities and challenges. Structural Control Health Monitoring, 11, pp. 349-368.

Wang, Y., Lynch, J.P., and Law, K.H. (2007). A wireless structural health monitoring system with multithreaded

sensing devices: design and validation. Structure and Infrastructure Engineering, 3, 2, pp. 103-120.

Yi, X., Zhu, D., Wang, Y., Guo, J., and Lee, K.-M. (2010). Embedded transmissibility function analysis for damage detection in a mobile sensor network. In: SPIE Smart Structures and Materials: Nondestructive Evaluation and Health Monitoring. San Diego, CA, USA, 03/07/2010.

Zimmerman, A.T., Shiraishi, M., Swartz, R.A., and Lynch, J.P. (2008). Automated Modal Parameter Estimation by Parallel Processing within Wireless Monitoring Systems. ASCE Journal of Infrastructure Systems, 14, 1, pp. 102-113.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

DEEP LEARNING AND REALITY CAPTURE FOR AUTOMATED RAILWAY BRIDGE INSPECTION WITH BIM ENRICHMENT

Rafael Cabral[1,2], Ricardo Santos[1,2], Diogo Ribeiro[1,2]
[1] iBuilt, School of Engineering, Polytechnic of Porto, Portugal
[2] CONSTRUCT, Faculty of Engineering, University of Porto, Porto, Portugal

Abstract

Manual inspection remains subjective and poorly integrated with digital workflows. Although computer vision enables automated damage detection, most approaches fail to translate observations into parametric Building Information Modelling (BIM). For bridges, this limitation persists within Bridge Information Modelling (BrIM), where geometry, semantics, and lifecycle information must align. This paper addresses the gap through a four-stage framework: optimal reality modelling, hybrid damage perception, automated Scan-to-BIM, and semantic enrichment via IFC objects. Validated on operational railway bridges, the method achieves a damage-detection mAP50 of 0.621 and geometric deviations under 23 mm, enabling BrIM generation directly from visual data.

Introduction

Railway bridges are critical infrastructure assets whose safety and serviceability depend on regular condition assessment. This requirement is becoming increasingly pressing as the European railway network ages; currently, more than 35% of railway bridges in the European Union are over 100 years old, with many approaching the end of their design service life (Tomor, 2013). Current inspection practice relies predominantly on manual visual surveys, which are inherently subjective, difficult to reproduce, and constrained by access, safety, and time limitations. Large-scale studies have demonstrated substantial inter-inspector variability in condition ratings, directly affecting the reliability of bridge management systems (BMS) and long-term maintenance planning (Moore et al., 2001).

To address these limitations, recent progress in reality capture technologies, specifically Unmanned Aerial Vehicles (UAVs), Terrestrial Laser Scanning (TLS), and Mobile Laser Scanning (MLS), has enabled the rapid acquisition of high-resolution geometric and visual data (Rakoczy et al., 2025, Olaszek et al., 2024). In parallel, deep learning techniques have achieved remarkable performance in detecting and segmenting structural damage from imagery (Ribeiro et al., 2025). However, a significant discontinuity remains between detection and digitalization.

Recent reviews on Digital Twin implementation (Nichols et al., 2025) indicate that while visual data is abundant, it is rarely used to update the engineering models serving as basis for structural analysis. Most inspection pipelines end at the level of 2D images or unstructured point clouds, failing to translate perception results into structured, parametric, and computable Building Information Modelling (BIM) representations (Yang et al., 2025). Within the bridge domain, this paradigm is referred to as Bridge Information Modelling (BrIM), extending BIM principles to infrastructure assets. This gap limits BrIM adoption for automated Operation and Maintenance (O&M), as current Scan-to-BIM approaches struggle to generalize across complex typologies without substantial manual intervention (Patil and Kalantari, 2025).

Unlike prior works that focus on isolated tasks, this work presents an end-to-end methodology for generating "as-damaged" BIM models directly from multimodal reality capture data. The framework is validated by a reinforced concrete railway bridge study located in Porto, Portugal.

Related works

Bridge inspection has traditionally relied on manual visual surveys that are time-consuming, subjective, and difficult to standardize (Kruachottikul et al., 2021). Recent research focuses on using reality capture, deep learning, and BIM/BrIM to improve efficiency and data quality along different parts of the inspection workflow.

UAV photogrammetry and terrestrial laser scanning (TLS) are widely used to generate accurate 3D as-is models of bridges, often by combining multiple sensors to overcome occlusions and access limitations (Mirzazade et al., 2023). For railway bridges, multimodal TLS–UAV fusion -has been applied to create precise 3D as-is models and compare them to as-designed models for geometric deviation assessment (Cabral et al., 2023). Other studies use 3D point clouds for automatic bridge component recognition to support Scan-to-BIM of existing bridges (Xu et al., 2026).

Numerous works propose CNN-based systems to detect, segment, and classify multiple defect types in bridge images, including cracks, corrosion, rust, spalling, and other surface damage . Advanced networks such as BridgeNet for multiclass damage segmentation (Huang et al., 2024), BGInet for bi-task bolt and rust detection from UAV imagery (Xu et al., 2025), and attention-enhanced YOLO variants for RC bridges (Ruggieri et al., 2025) achieve high detection accuracy and real-time

performance. Several frameworks also address defect quantification, for example mapping image pixels to real-world dimensions or using orthophotos for crack measurements (Xu et al., 2025, Ayele et al., 2020).

A growing body of work links inspection data with BIM or digital twins to support asset management. Some studies integrate corrosion or surface defect measurements into BIM models to manage long-span or RC bridges more efficiently (Hattori et al., 2024, Cavieres-Lagos et al., 2024). Others develop BIM-based mixed-reality tools that overlay inspection information on 3D models to support on-site or remote bridge inspection and maintenance (Riedlinger et al., 2022, Zakaria et al., 2023).

A smaller set of studies tackles partial automation from vision-based damage detection to 3D localization and BIM linkage. Methods combining deep learning with Structure-from-Motion or reality capture integrate defect detections into 3D bridge models, reducing variability between images and enabling spatial recording of damage (Yamane et al., 2023). However, these workflows typically address either geometry reconstruction and component labelling (Kim et al., 2020) or damage detection and 3D localization (Yamane et al., 2024), rather than delivering a fully integrated, as-damaged BrIM that unifies multimodal capture, automated perception, Scan-to-BrIM, and semantic enrichment in a single, railway-bridge-specific framework.

Methodology

The proposed framework establishes a computational pipeline that connects raw sensor data to semantic BrIM models. As illustrated in *Figure 1*, the architecture is composed of four sequential modules: i) Optimal Reality Modelling; ii) Hybrid Damage Perception; iii) Automated Scan-to-BIM; and iv) Semantic BIM Enrichment via ray-casting and parametric object instantiation.

Optimal reality modelling

The first stage aims to generate high-fidelity 3D reality models suitable for geometric inference. Railway bridge environments are characterized by complex backgrounds, e.g. vegetation, catenary systems, and varying terrain, which introduce significant noise into both photogrammetric and LiDAR datasets.

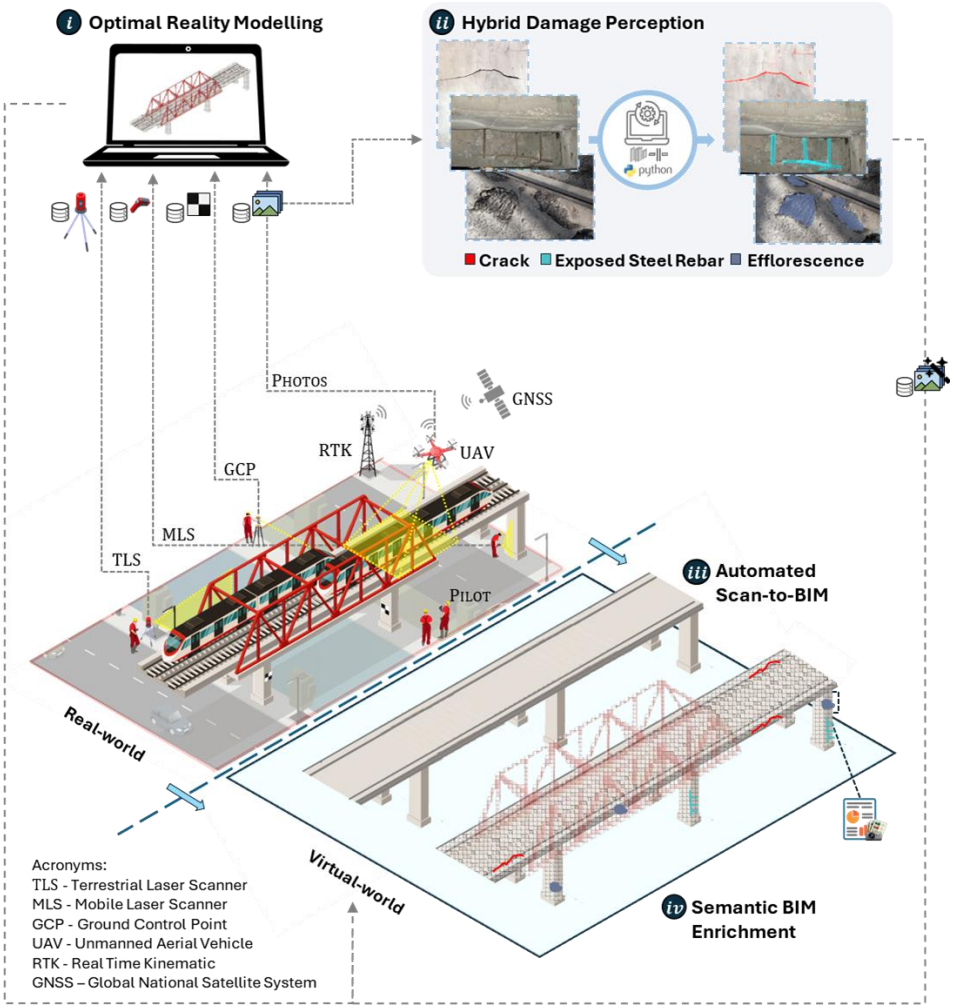


Figure 1: Overview of the proposed end-to-end framework, illustrating the four sequential stages: i) Optimal reality modelling, ii) hybrid damage perception, iii) Automated Scan-to-BrIM, and iv) semantic enrichment.

\

To mitigate these artefacts, a multi-modal fusion strategy was adopted that integrates background-aware filtering prior to reconstruction, illustrated in *Figure 2* (Cabral et al., 2025b). For photogrammetric data, a semantic segmentation network S processes each raw image $I \in R^{H\times W\times 3}$, generating a binary mask M where structural pixels are preserved ($M_{i,j} = 1$) and background pixels are suppressed ($M_{i,j} = 0$). This ensures that the subsequent Structure-from-Motion (SfM) pipeline reconstructs only relevant structural geometry. Simultaneously, LiDAR point clouds acquired via TLS or MLS undergo heuristic filtering based on multi-source geometric proximity (Cloud-to-Cloud distance).

Following data fusion, a Voxel Grid Filtering strategy addresses the non-uniform point densities inherent to multi-sensor capture. The 3D space is discretized into volumetric elements (voxels) V of dimension δ (e.g., 0.01 m). For a set of points $P = \{p_1, p_2, \dots, p_n\}$ within a voxel V_k, the retained point centroid $p_{centroid}$ is computed in Equation (1).

$$p_{centroid} = \frac{1}{N_k}\sum_{i=1}^{N_k} p_i \quad (1)$$

where N_k is the number of points in voxel V_k. This downsampling preserves the underlying geometric topology while normalizing density, ensuring computational tractability for the subsequent deep learning stages without compromising the geometric fidelity required for defect quantification.

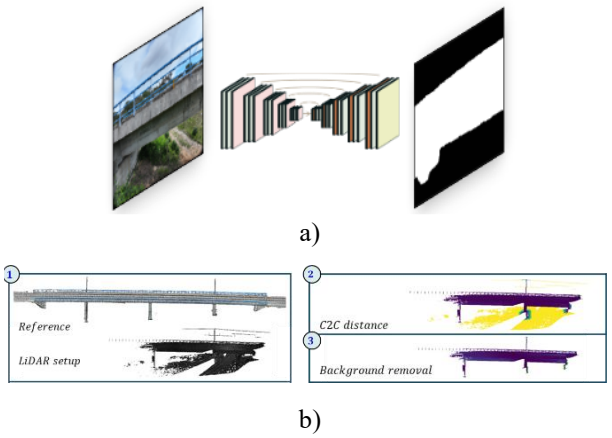


Figure 2: Optimal reality modelling strategy showing background removal for: a) imagery and b) LiDAR (adapted from Cabral et al., 2025b).

Hybrid damage perception

To achieve robust instance segmentation under operational conditions, the challenge of scale variation in high-resolution UAV imagery was addressed. Standard detectors often fail to identify fine cracks when images are downscaled. Thus, it is proposed a hybrid pipeline integrates YOLO11, Slicing Aided Hyper Inference (SAHI), and the Segment Anything Model (SAM) (Cabral et al., 2025a).

The detection phase utilises SAHI to systematically partition high-resolution images I_{HR} into overlapping patches P_k. Each patch is normalised to a 640×640 resolution, the native input size for the YOLO11 detector, ensuring that fine-grained features are preserved without the spatial degradation typically caused by global image downscaling. A YOLO11 detector, trained on railway-specific pathologies, processes each patch independently. The global detection set D_{global} is reconstructed by merging local predictions using Non-Maximum Suppression (NMS) to eliminate redundant bounding boxes.

To transition from coarse bounding boxes to pixel-perfect segmentation, the detected boxes B_{det} serve as geometric prompts for the Segment Anything Model (SAM), as depicted in *Figure 3*. Unlike traditional semantic segmentation networks trained on fixed classes, SAM leverages a Vision Transformer (ViT) backbone to generate zero-shot masks for each prompted instance. The segmentation objective is to minimize the boundary error between the predicted mask and the ground truth.

This hybrid approach ensures that fine-grained defects (e.g., hairline cracks) are detected at their native resolution while their boundaries are segmented with the precision of a foundation model.

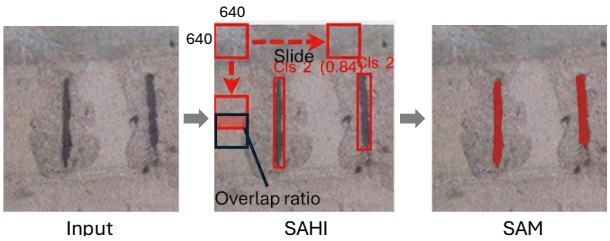


Figure 3: The hybrid perception pipeline: SAHI-assisted detection feeding into SAM-based segmentation.

Automated Scan-to-BrIM

The third stage automates the conversion of unstructured point clouds into parametric Bridge Information Modelling (BrIM) elements. It replaces manual modelling with a deep learning-based keypoint regression network trained on synthetic data.

Given the scarcity of labelled bridge point clouds, it employs a procedural generation engine to create a synthetic dataset. This engine, based on Revit® and Dynamo©, is used to randomise bridge geometrical and structural parameters. Nvidia Omniverse© is then leveraged to simulate laser scanning technology and generate semantically labelled point clouds. *Figure 4* illustrates the overall data generation and perception workflow.

The reconstruction network based on Yolo-Pose accepts point cloud slices S_i along the bridge alignment and regresses a set of structural keypoints $K = \{k_1, k_2, \dots, k_m\}$ defining the cross-sectional profile, as illustrated in *Figure 5*. The network is trained by minimising the Mean Squared Error (MSE) between predicted keypoints $\hat{K}$ and ground truth K_{GT} , as defined in Equation (2).

\

$$L_{keypoint} = \frac{1}{m}\sum_{j=1}^{m}\left|\left|\hat{k}_j - k_{GT,j}\right|\right|^2 \quad (2)$$

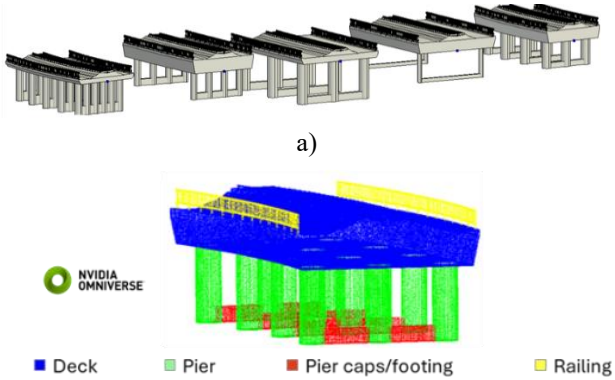


Figure 4. Synthetic bridge point cloud generation pipeline: a) bridge modelling and parameter randomisation using Revit® and Dynamo©; b) laser scanning simulation and semantic point cloud generation using NVIDIA Omniverse©.

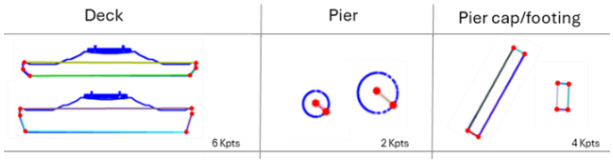


Figure 5: Labelled dataset for YOLO-Pose keypoints defining the cross-sectional profile.

Once inferred, these keypoints drive the instantiation of Industry Foundation Classes (IFC) entities. Structural elements are generated via lofting algorithms, ensuring the resulting model is fully parametric and interoperable, namely, IfcColumn, IfcBeam, IfCSlab and IfcWall (Amroune et al., 2022).

Semantic BIM Enrichment

The final stage transforms the geometric model into an "as-damaged" BrIM model by mapping 2D perception results into the 3D BIM environment.

Damage masks from Stage 2 are projected onto 3D space using a ray-casting algorithm. For a pixel u, v classified as damage, a ray $r(t)$ is cast from the camera center C through the pixel on the image plane. As defined in Equation (3).

$$r(t) = C + t \cdot K^{-1} \cdot [u, v, 1]^T \quad (3)$$

where K is the camera intrinsic matrix. The intersection of $r(t)$ with the 3D mesh geometry determines the spatial location of the defect. To consolidate detections from multiple overlapping images, a spatial clustering algorithm refined by Otsu thresholding is applied to filter low-confidence projections. *Figure 6* illustrate this process

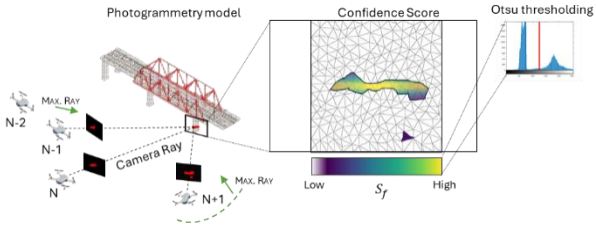


Figure 6: Multi-view ray-casting-based projection of 2D damage detections onto the 3D mesh for semantic enrichment.

The resulting 3D clusters are ortho-rectified to calculate engineering metrics (area, length, width). Finally, defects are instantiated as IfcBuildingElementProxy objects and hierarchically linked to their host structural element (e.g., IfcColumn) using the IfcRelContainedInSpatialStructure relationship.

This process establishes a direct link between visual inspection data and the BrIM model, enabling automated condition assessment and maintenance planning.

Validation and Results

Case Study Description

The proposed framework was experimentally validated on an operational reinforced concrete railway bridge located in Porto, Portugal, which is depicted in *Figure 7*. The structure consists of a single-track deck with a width of 11.33 m and a height of 0.80 m, supported by cylindrical piers (0.80 m in diameter and approximately 6.30 m in height), pier footing with a width of 0.40 m and a height of 0.80 m and integral abutments. The bridge exhibits heterogeneous degradation, including cracking, efflorescence, and exposed steel rebars.

This combination of structural complexity and varied damage mechanisms provides a case study for evaluating the robustness and reliability of the proposed automated inspection pipeline. All data acquisition activities were conducted in full compliance with national aviation regulations and railway safety requirements. The UAV flights were carried out with the necessary authorisations and under the supervision and support of the relevant infrastructure and railway authorities.



Figure 7: Aerial view of the bridge located in Porto (Portugal) used for experimental validation.

\

Reality Modelling Quality Assessment

The accuracy of the reality model serves as the foundational constraint for all downstream perception and reconstruction tasks. For the case study, the effectiveness of the proposed multi-modal fusion strategy, i.e. combining semantic background removal for photogrammetry with heuristic filtering for LiDAR, was evaluated based on geometric completeness.

Qualitative inspection of the fused point cloud confirms that the upstream filtering successfully suppressed high-frequency noise sources typical of railway corridors, as illustrated in *Figure 8*. This pre-processing yielded a structurally focused dataset in which the primary load-bearing elements (piers, deck slab, pier footings, and abutments) maintained surface continuity. Quantitatively, the voxel grid filtering strategy significantly reduced the data volume while homogenising point density.



a)



b)

Figure 8: geometric completeness: a) Raw reality modelling containing noise; b) Optimal reality model after filtering.

Damage Perception Performance

The damage identification performance was quantitatively evaluated using a hybrid deep learning pipeline designed to detect and segment three critical railway bridge pathologies, namely, cracks, efflorescence, and exposed rebar.

The model was trained using a multi-source transfer learning strategy to address the scarcity of labelled infrastructure data. The foundational weights were initialised from a model pre-trained on the MS COCO dataset, followed by fine-tuning on a curated subset of the public DACL10K bridge damage dataset (3,737 images). The final domain adaptation was performed on a custom high-resolution UAV dataset comprising 645 training images and 161 validation images (5280×3956 pixels), annotated with pixel-precise masks.

The proposed hybrid pipeline, which synergises a YOLO11 detector with Slicing Aided Hyper Inference (SAHI) and the Segment Anything Model (SAM), achieved a Mean Average Precision (mAP50) of 0.621 for detection across all classes on the validation set.

This step proved critical for the Scan-to-BIM stage by removing non-structural outliers, the effective geometric topology was preserved, allowing the keypoint regression network to operate on clean cross-sectional profiles.

In total, 170 distinct damage instances were identified and segmented, comprising 75 cracks, 13 efflorescence regions, and 82 areas of exposed reinforcement. Figure 9 qualitatively confirmed that the pipeline effectively identified the damage from the UAV image. A class-specific performance analysis highlighted the benefits of the hybrid architectural choices:

Cracks. The native YOLO11 output, enhanced by SAHI's high-resolution inference tiling, proved instrumental for linear defects. It achieved an Intersection over Union (IoU) of 0.495, successfully recalling narrow cracks that typically vanish during standard image downsampling.

Area Defects. For pathologies with complex morphologies but clearer boundaries, the two-stage refinement using SAM (prompted by YOLO bounding boxes) outperformed native segmentation. This approach yielded IoU scores of 0.331 for efflorescence and 0.205 for exposed steel rebars.

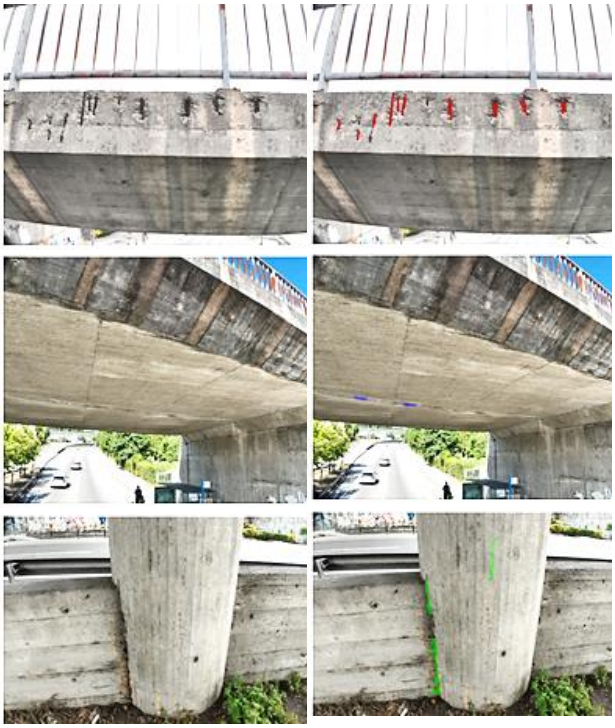
By producing well-defined masks, the framework ensures that metric quantification is derived from precise morphological geometry rather than coarse approximations.

Geometric Accuracy of Scan-to-BIM

The geometric validation of the generated Bridge Information Model (BrIM) was assessed by computing the signed Euclidean deviations between the generated IFC elements and the ground-truth reality capture point cloud. *Figure 11* depicts the BrIM model. The global mean geometric deviation for the bridge was approximately 14.46 mm. It is important to note that the upper surface of the deck was not modelled due to occlusion by the ballast and track superstructure; thus, this does not reflect in the deviation calculation. This accuracy aligns with the requirements for Level of Development (LOD) 300 (BIMForum, 2024), which requires elements be graphically represented as specific systems with precise quantities, sizes, shapes, locations, and orientations.

A component-level error analysis reveals distinct performance characteristics across structural elements. The pier caps and footings exhibited the lowest mean deviation (5.32 mm), followed closely by the abutments (9.29 mm). The cylindrical piers showed a mean deviation of 19.6 mm, while the deck slab presented the highest deviation at 22.89 mm. This confirms that synthetic domain randomisation is a viable strategy for overcoming the scarcity of labelled infrastructure datasets.

\



a) b)

Figure 9: Qualitative results of the hybrid perception engine: a) original image; b) prediction damage, i.e., cracks (green), efflorescence (blue), and exposed rebar (red).

Semantic Enrichment and Damage Quantification

The final phase of validation focused on the semantic enrichment capability, the transformation of the BrIM model from a simple structural component into an "as-damaged" BrIM model. The ray-casting and clustering algorithms successfully mapped the 2D perception results onto the 3D IFC entities, establishing a coherent spatial database of defects.



Figure 11: the automated generated parametric BrIM geometry (LOD 300).

The enriched BIM environment enabled the automatic extraction of aggregate damage metrics. *Figure 10* illustrates the damage mapped to the georeferenced BrIM model, with corresponding dimensional attributes and reference image as well as the corresponding orthophoto. For the case study bridge, the primary contribution is the instance-level granularity; each defect is instantiated as an IfcBuildingElementProxy and hierarchically linked to its host element (e.g., a specific pier). Beyond local statistics, the system quantified a total crack length of 13.82 m, an efflorescence surface area of 0.112 m², and an exposed reinforcement area of 0.241 m². This structured data format allows asset managers to query the model for element-specific condition ratings, bridging the gap between raw visual data and actionable maintenance decision-making.

Discussion

This research resolves the historical discontinuity between automated damage detection and structural digitalization by integrating multimodal reality capture with Bridge Information Modelling (BrIM). While conventional pipelines typically terminate at 2D imagery or unstructured point clouds, the proposed framework embeds quantified defects as *IfcBuildingElementProxy* objects. This approach transforms a static BrIM into an interoperable "as-damaged" model, supplying asset

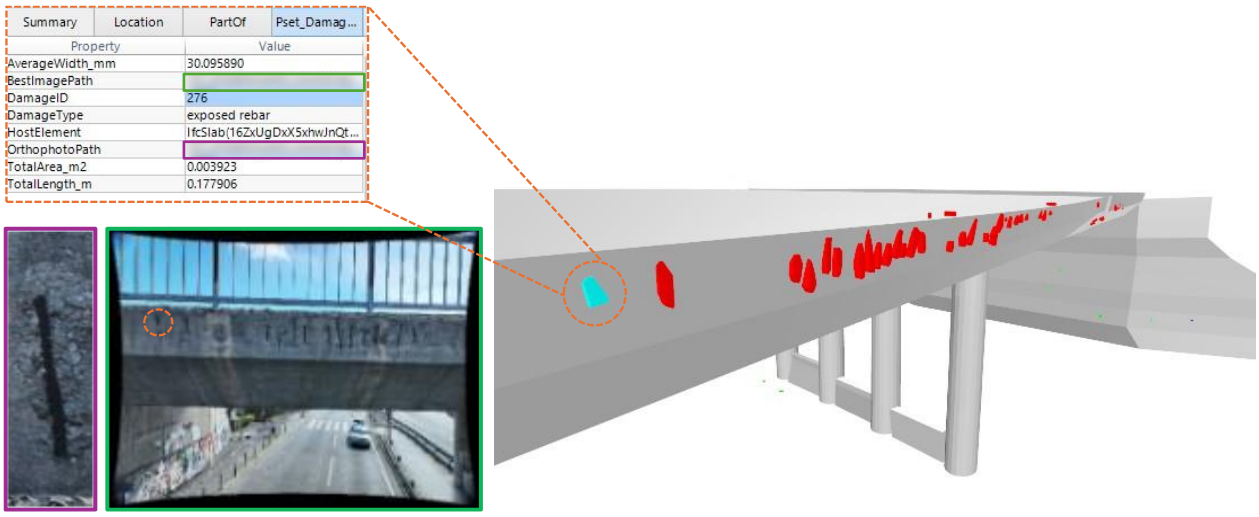


Figure 10: the "as-damaged" BrIM model visualizing mapped defects (light blue) alongside the associated property window displaying quantified dimensional attributes (e.g., area, length) and links to the source reference image and generated orthophoto.

\

managers with actionable, element-specific condition metrics for automated maintenance decision-making.

Experimental validation on an operational railway bridge demonstrated that high-fidelity reality modelling is a strict prerequisite for automated inspection in complex environments. By employing a multi-modal fusion strategy—combining semantic background removal for photogrammetry with heuristic LiDAR filtering—the framework successfully isolated structural geometry from background noise (e.g., vegetation, catenary systems). Consequently, the automated reconstruction achieved a global mean geometric deviation of 14.46 mm, satisfying Level of Development (LOD) 300 requirements.

The findings also highlight the inadequacy of standard object detectors for capturing fine-grained infrastructure pathologies, as standard image downscaling often eliminates hairline cracks. Integrating Slicing Aided Hyper Inference (SAHI) with YOLO11 preserved these linear defects at their native high resolution. Subsequent processing with the Segment Anything Model (SAM) generated zero-shot, pixel-perfect masks, ensuring damage quantification was driven by precise morphology rather than coarse bounding box approximations.

To address the scarcity of annotated 3D point clouds for Scan-to-BIM, the framework successfully implemented synthetic domain randomization. Generating simulated data via Revit®, Dynamo©, and NVIDIA Omniverse© proved highly effective for training accurate keypoint regression networks, yielding component-level deviations as low as 5.32 mm.

Despite robust performance, current limitations include the inability to model the bridge deck's upper surface due to track superstructure occlusions, alongside the methodology's exclusive focus on surface-level manifestations of damage.

Conclusions

This study presented and validated an end-to-end computational framework that effectively brings together multimodal reality capture and semantic Bridge Information Modelling (BrIM). By integrating AI-driven background removal, a hybrid YOLO11–SAHI–SAM perception engine, and synthetic-data-trained Scan-to-BIM reconstruction, the methodology overcomes the scalability limitations inherent to traditional manual inspection.

The experimental validation on the bridge located at Porto (Portugal) confirms that high-fidelity reality modelling is a prerequisite for automated inspection. The proposed multi-modal background filtering strategy proved essential for reducing geometric noise, enabling automated reconstruction algorithms to converge with millimetre-level accuracy (mean deviations < 23 mm).

Furthermore, the results demonstrate that standard object detectors are insufficient for the complexity of infrastructure pathologies. The integration of Slicing Aided Hyper Inference (SAHI) and the Segment Anything Model (SAM) was critical for decoupling detection from segmentation, enabling precise quantification of fine-grained defects, such as cracks, that are typically lost (due to resolution) in standard down sampling processes. Ultimately, the true engineering value of the framework lies in the semantic enrichment stage, which successfully transforms static a BrIM model into an "as-damaged" BrIM model. By embedding quantified defects as intelligent IfcBuildingElementProxy objects, the system delivers interoperable, structured data directly usable for predictive maintenance and asset management.

Future research will prioritise the comparative validation of this automated framework against traditional manual inspection benchmarks for the same asset. Additionally, the synthetic training strategy will be extended to encompass complex masonry and steel truss typologies, thereby enhancing model generalisation. Finally, efforts will be directed towards integrating subsurface non-destructive testing (NDT) data into the BrIM environment, providing a holistic view of structural health that extends beyond surface-level defects.

Acknowledgments

This work was financially supported by Funding-UID/04708 of the CONSTRUCT-Instituto de I&D em Estruturas e Construções-funded by Fundação para a Ciência e a Tecnologia, I.P./MCTES through the national funds, and co-financed by COMPETE 2030, by Portugal 2030, as well as, by the NORTE 2030 Programme through the European Regional Development Fund (ERDF), under project FireBIM, reference NORTE2030-FEDER-00358400. It was additionally supported by the grant UI/BD/150970/2021 Portuguese Science Foundation, FCT/MCTES.

References

Amroune, A., Cuillière, J. C. & François, V. (2022). Automated Lofting-Based Reconstruction of CAD Models from 3D Topology Optimization Results. Computer-Aided Design, 145, 103183.

Ayele, Y. Z., Aliyari, M., Griffiths, D. & Droguett, E. L. (2020). Automatic Crack Segmentation for UAV-Assisted Bridge Inspection. Energies, 13, 6250.

Bimforum. (2024). Level of development (LOD) specification [Online]. Available: https://bimforum.org/resource/lod-level-of-development-lod-specification/ [Accessed January 2 2025].

Cabral, R., Oliveira, R., Ribeiro, D., Rakoczy, A. M., Santos, R., Azenha, M. & Correia, J. (2023). Railway Bridge Geometry Assessment Supported by Cutting-Edge Reality Capture Technologies and 3D As-Designed Models. Infrastructures, 8, 114.

Cabral, R., Santos, R., Correia, J. & Ribeiro, D. (2025a). A Hybrid YOLO and Segment Anything Model Pipeline for Multi-Damage Segmentation in UAV Inspection Imagery. Sensors (Basel), 25.

Cabral, R., Santos, R., Correia, J. & Ribeiro, D. (2025b). Optimal reconstruction of railway bridges using a

\

machine learning framework based on UAV photogrammetry and LiDAR. Structure and Infrastructure Engineering, 173, 1-21.

Cavieres-Lagos, S., Muñoz La Rivera, F., Atencio, E. & Herrera, R. F. (2024). Integration of BIM tools for the facility management of railway bridges. Applied Sciences, 14, 6209.

Hattori, K., Oki, K., Sugita, A., Sugiyama, T. & Chun, P.-J. (2024). Deep learning-based corrosion inspection of long-span bridges with BIM integration. Heliyon, 10.

Huang, L., Fan, G., Li, J. & Hao, H. (2024). Deep learning for automated multiclass surface damage detection in bridge inspections. Automation in Construction, 166, 105601.

Kim, H., Yoon, J. & Sim, S. H. (2020). Automated bridge component recognition from point clouds using deep learning. Structural Control and Health Monitoring, 27, e2591.

Kruachottikul, P., Cooharojananone, N., Phanomchoeng, G., Chavarnakul, T., Kovitanggoon, K. & Trakulwaranont, D. (2021). Deep learning-based visual defect-inspection system for reinforced concrete bridge substructure: a case of Thailand's department of highways. Journal of Civil Structural Health Monitoring, 11, 949-965.

Mirzazade, A., Popescu, C., Gonzalez-Libreros, J., Blanksvärd, T., Täljsten, B. & Sas, G. (2023). Semi-autonomous inspection for concrete structures using digital models and a hybrid approach based on deep learning and photogrammetry. Journal of Civil Structural Health Monitoring, 13, 1633-1652.

Moore, M., Phares, B. M., Graybeal, B., Rolander, D., Washer, G. & Wiss, J. (2001). Reliability of visual inspection for highway bridges, volume I. Turner-Fairbank Highway Research Center.

Nichols, L., Ashmawi, A. & Nguyen, P. H. D. (2025). Digitalizing Bridge Inspection Processes Using Building Information Modeling (BIM) and Business Intelligence (BI). Applied Sciences-Basel, 15, 10927.

Olaszek, P., Maciejewski, E., Rakoczy, A., Cabral, R., Santos, R. & Ribeiro, D. (2024). Remote Inspection of Bridges with the Integration of Scanning Total Station and Unmanned Aerial Vehicle Data. Remote Sensing, 16, 4176.

Patil, J. & Kalantari, M. (2025). Automatic Scan-to-BIM-The Impact of Semantic Segmentation Accuracy. Buildings, 15, 1126.

Rakoczy, A. M., Ribeiro, D., Hoskere, V., Narazaki, Y., Olaszek, P., Karwowski, W., Cabral, R., Guo, Y., Futai, M. M. & Milillo, P. (2025). Technologies and platforms for remote and autonomous bridge inspection–review. Structural Engineering International, 35, 354-376.

Ribeiro, D., Rakoczy, A. M., Cabral, R., Hoskere, V., Narazaki, Y., Santos, R., Tondo, G., Gonzalez, L., Matos, J. C., Massao Futai, M., Guo, Y., Trias, A., Tinoco, J., Samec, V., Minh, T. Q., Moreu, F., Popescu, C., Mirzazade, A., Jorge, T., Magalhaes, J., Schmidt, F., Ventura, J. & Fonseca, J. (2025). Methodologies for Remote Bridge Inspection-Review. Sensors (Basel), 25, 5708.

Riedlinger, U., Klein, F., Hill, M., Lambracht, C., Nieborowski, S., Holst, R., Bahlau, S. & Oppermann, L. (2022). Evaluation of mixed reality support for bridge inspectors using BIM data: Digital prototype for a manual task with a long-lasting tradition. i-com, 21, 253-267.

Ruggieri, S., Cardellicchio, A., Nettis, A., Renò, V. & Uva, G. (2025). Using attention for improving defect detection in existing RC bridges. IEEE Access, 13, 18994-19015.

Tomor, A. (2013). Life-cycle assessment and deterioration models for masonry arch bridges. WIT Transactions on Ecology and the Environment, 179, 535-546.

Xu, J., Shi, M., Cabral, R., Ribeiro, D., Yu, W., Wu, H. & Narazaki, Y. (2026). Domain-generalizable point cloud instance segmentation of bridge components using class-balanced dynamic thresholding. Automation in Construction, 181, 106631.

Xu, T., Wu, Y., Qin, Y., Long, S., Yang, Z. & Guo, F. (2025). Automatic steel girder inspection system for high-speed railway bridge using hybrid learning framework. Computer-Aided Civil and Infrastructure Engineering, 40, 1508-1527.

Yamane, T., Chun, P.-J. & Honda, R. (2024). Detecting and localising damage based on image recognition and structure from motion, and reflecting it in a 3D bridge model. Structure and Infrastructure Engineering, 20, 594-606.

Yamane, T., Chun, P. J., Dang, J. & Honda, R. (2023). Recording of bridge damage areas by 3D integration of multiple images and reduction of the variability in detected results. Computer-Aided Civil and Infrastructure Engineering, 38, 2391-2407.

Yang, X. F., Castillo, E. D., Zou, Y., Wotherspoon, L., Yang, J. X. & Li, H. (2025). Automated Concrete Bridge Damage Detection Using an Efficient Vision Transformer-Enhanced Anchor-Free YOLO. Engineering, 51, 311-326.

Zakaria, M., Karaaslan, E. & Necati Catbas, F. Real-time ai-based bridge inspection using mixed reality platform. Structures Congress 2023, 2023. 120-131.

\

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

AUTOMATED GENERATION OF SEMANTIC IFC MODELS FROM LEGACY SEWER NETWORK DATA

Peter Gölzhäuser, Raquel Valles Gomez, Sabine Hartmann, Hendrik Morgenstern, and Katharina Klemt-Albert
Chair and Institute of Construction Management, Digital Engineering and Robotics in Construction, RWTH Aachen University, Aachen, Germany

Abstract

This paper presents an automated workflow for generating georeferenced IFC 4.3 models from legacy data structured according to the DWA-M 150 schema. The DWA-M 150 schema is an XML-based data exchange format for condition monitoring of drainage systems, which is used as a standard by German municipal utilities. The approach consists of three stages: geometry generation from master inventory data, semantic enrichment, and integration of inspection and condition assessment data. The resulting semantically enriched models provide a solid foundation for future digital twin applications in sewer infrastructure management.

Introduction

Germany's drainage and sewerage infrastructure faces the dual challenge of aging assets and necessary expansion (Berger et al., 2020). Furthermore, climate change exacerbates these pressures, with an increased frequency of heavy rain events heightening the risk of flooding (Blöschl et al., 2019). Against this context, the complexity of modern infrastructure assets, coupled with the high cost of labour, makes manual inspection and condition data interpretation increasingly impractical (Tse, 2016). Adopting a more data-driven and automated approach may offer a solution to this problem. However, the data landscape is characterised by fragmentation, often referred to as ''data silos'', which poses a significant obstacle to the adoption of advanced predictive maintenance strategies (Kretschmer et al., 2023). To address this, researchers propose standardized data models that enable both human comprehension and computational reasoning, facilitating seamless integration of heterogeneous data streams into a coherent system view (Du et al., 2023).

This paper presents an automated approach for generating georeferenced IFC 4.3 models from data based on the DWA-M 150 schema, including the visualization of inspection and condition assessment results. By transforming existing municipal sewer data into semantically enriched 3D models, this approach provides the necessary foundation for future digital twin applications.

Background

This section establishes the theoretical framework for the study. It contextualizes the current state of digital sewer data management, contrasting existing practices with the requirements of digital twins and the IFC 4.3 schema. It distinguishes between model generation methods and explains the principles of semantic enrichment to translate structured legacy data into 3D data models.

Digital transformation in sewer infrastructure management

The growing demand for integrated data management drives the construction sector towards the adoption of the *digital twin (DT)*, defined as a dynamic digital representation of a physical asset (buildingSMART International, 2020). This definition can be refined by distinguishing between a *Digital Shadow* (DS)—where data flows automatically from the physical to the digital object—and a full Digital Twin, which includes a bidirectional flow allowing for autonomous control (Ghorbani Bam et al., 2025). To operationalize these concepts, current research proposes a layered reference architecture consisting of Presentation, Platform, Transmission, and Data Acquisition layers (Smarsly et al., 2025). In line with this framework, our study falls within the platform layer, as it focuses on the elements that make up the digital replica, in particular the modeling and data storage elements required for a foundational digital representation.

Currently, sewage networks are typically structured topologically and described using a node-edge model. In this model, nodes represent structural elements such as manholes, inspection chambers and sewage treatment plants, while edges represent connected pipe sections (DWA, 2021). In Germany, the management and standardization of this data are governed by the DWA-M 150 guideline (*Datenaustausch für die Zustandserfassung von Entwässerungssystemen*) (DWA, 2010). In terms of spatial context, Geographic Information Systems (GIS) have become the standard at a macro level. However, they lack the geometric granularity required for component-level maintenance, a task for which Building Information Modeling (BIM) has been proven to be useful (Wang et al., 2019; Valles Gomez et al., 2025).

The role of Building Information Modeling and open standards

The release of IFC 4.3 as an ISO standard in 2024 (ISO 16739-1:2024) marked a milestone by extending the schema to include comprehensive support for infrastructure assets. While core entities like `IfcPipeSegment`,

`IfcWasteTerminal` and `IfcDistributionSystem` existed in earlier versions for vertical construction, IFC 4.3 introduces the specialized predefined types and linear referencing capabilities necessary to model civil works such as highway drainage or municipal utilities. This standardisation paves the way for a semantically structured 3D data model. Its open data structures allow for high interoperability, enabling the future integration of sensor data for managing municipal assets throughout their lifecycle (Edmondson et al., 2018).

To bridge the gap between existing infrastructure records and new digital standards, two methodological approaches can be differentiated:

1. **Scan-to-BIM (Reality Capture):** Processes involve capturing spatial data via Terrestrial Laser Scanning (TLS), photogrammetry, or SLAM-based mobile mapping to generate high-density point clouds, followed by a reconstruction process. Recent research has focused on automating the subsequent reconstruction process via Deep Learning architectures for semantic segmentation (Jäkel et al., 2025; Kellner et al., 2025). While effective for visible, aboveground structures such as bridges or facades, the lack of line-of-sight and the logistical constraints of physical access render reality capture and large-scale modeling for extensive underground utility networks highly challenging.

2. **Legacy-to-BIM (Information Reconstruction):** Often characterized as semantic enrichment in recent literature (Belsky et al., 2016; Jiang et al., 2023), reconstructs or augments models from existing datasets. The approach proposed in this paper uses the implicit semantics within legacy records (such as structured XML files like DWA-M 150) to algorithmically generate semantics and geometry. The research challenge lies in translating data attributes into the 3D shapes, properties, and topological relationships required by BIM standards.

While legacy data for sewer systems is often highly available and structured in domain-specific formats (e.g., DWA-M 150), there is currently no direct, standardized mapping to transform this data into the explicit geometry and semantic structure required by the IFC 4.3 standard.

Approaches to geometric and semantic enrichment

The concept of *semantic enrichment (SE)* has emerged as a response to the interoperability gaps inherent in digital construction workflow (Sacks et al., 2017). Addressing this challenge, Belsky et al. introduced ''SeeBIM'' (Semantic Enrichment Engine for BIM), a prototypical framework designed to interpret the implicit information contained in building models. Their rule-based inference engine utilizes domain expert knowledge—such as the connection constraints between precast concrete elements—encoded into ''IF-THEN'' logic to deduce new semantic objects and relationships from raw geometry. While highly effective, this approach fundamentally relies on inferring semantics from geometric and topological evidence within standard IFC files (Belsky et al., 2016). The workflow follows a logical path of Geometry → Topology → Semantics, deriving meaning from scanned shapes (Sacks et al., 2017). In the context of this study, however, the challenge is inverted. The source data (DWA-M 150) is semantically rich, but does not include a 3D-model. Consequently, the approach required for legacy sewer data must follow the path of Semantics → Topology → Geometry, utilising the structured attributes in the XML to generate the geometric representation.

This methodological inversion prompts a re-evaluation of the term ''semantic enrichment''. Bloch refines the prior definition by distinguishing between enrichment and enhancement. In her framework, *semantic enrichment* describes a process using inference engines to derive new information from the model itself (creating new knowledge), whereas *semantic enhancement* refers to the integration of external data sources into the model (linking existing knowledge) (Bloch, 2022). The workflow proposed in this paper occupies a hybrid position between these two concepts: it *enhances* the model by mapping external attributes to IFC property sets, yet it *enriches* the model by algorithmically generating 3D geometry and topological connections that did not previously exist in the 2D source data. Thus, while acknowledging the taxonomic distinction between enrichment and enhancement proposed by Bloch (2022), this paper adopts semantic enrichment as a comprehensive umbrella term to describe the hybrid process of both inferring model geometry and integrating external legacy records.

Research gap

Current research focuses on either digital twin dynamics (e.g., simulation, sensors) or geometric reconstruction (Scan-to-BIM), but lacks standardised methods for the automated semantic translation of legacy data sources into open BIM standards.

Since the DWA-M 150 schema provides rich, structured semantic data, converting these records to IFC 4.3 through a ''Legacy-to-BIM'' approach offers a pragmatic solution. This methodology is particularly relevant given the proliferation of Machine Learning techniques in both semantic segmentation and enrichment (Schönfelder et al., 2023). However, by their very nature, these approaches entail probabilistic uncertainties (Kellner et al., 2026). In contrast, the use of a well-structured database source enables deterministic, rule-based algorithms to create georeferenced semantic models with high accuracy, thereby circumventing the uncertainty associated with machine learning.

Methodology

This section presents the technical methodology for the automated generation of the IFC model and semantic en-

richment from DWA-M 150 data. The proposed approach maintains a separation of geometry and semantics, enabling both processes to run independently of each other. This allows for flexibility in scenarios where geometry originates from alternative sources, such as point clouds resulting from 3D scans, and facilitates regularly updating condition data resulting from new inspections. Figure 1 shows a schematic representation of the conversion pipeline.

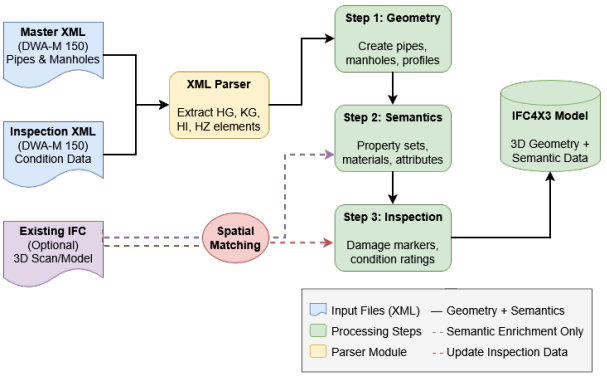


Figure 1: Overview of the conversion pipeline

Data basis and requirements

The process relies on input data structured according to the DWA-M 150 schema, which has been widely adopted by German municipalities and is integrated into most Sewer Information Systems. Although the algorithm presented in this paper was designed specifically for this standard, the underlying approach is transferable and adaptable to other structured data formats.

Within DWA-M 150, each infrastructure element is described by a set of standardized identifiers that are defined as mandatory or optional within the schema. Pipe sections (''Haltungen'' in German) use the identifier prefix `<HG>`, while nodes (''Knoten'' in German) use `<KG>`. The standard defines four element categories:

- **Basic Data** (HG/KG): Designation, material, dimensions, construction year
- **Geometry Data** (GO/GP): Coordinates and spatial information
- **Inspection Data** (HI/KI): Camera survey metadata and observations
- **Condition Data** (HZ/KZ): Damage classifications and assessments

The source file functions as a hierarchical data tree where the root element `<Data>` contains collections of pipes `<HG>` and nodes `<KG>`. Geometry is explicitly defined in `<GO>` (Geometry Objects) subsections, providing Easting, Northing, and Elevation values for each feature point `<GP>`.

Automated geometry generation workflow

The workflow is implemented as an automated converter in Python using IfcOpenShell for IFC 4.3.2.0 generation and NumPy for vector-based geometry calculations. Its geometry module handles coordinate transformations, profile creation, and extrusion operations. The converter also includes a parser module for the DWA-M 150 XML serialisation, which handles all element types (HG, KG, HI, HZ, and HM) and a mappings module for code translations and reference table handling.

Profile definition

The cross-sectional profile of each pipe is determined from the profile type code (`HG305`) and dimensional attributes (`HG306`, `HG307`). The converter supports three profile types:

- **Circular profiles** (code `DN`): Most common, represented as `IfcCircleHollowProfileDef`
- **Egg-shaped profiles** (code `EI`): Typical for combined sewers, created as custom parametric profiles using `IfcArbitraryProfileDefWithVoids` with standard DIN ratios.
- **Rectangular profiles** (code `RE`): Used for larger collectors, represented as `IfcRectangleHollowProfileDef`

Since the wall thickness is not specified in DWA-M 150, a diameter-dependent lookup table based on common precast concrete specifications is used (e.g., 60 mm for DN 150, 75 mm for DN 300–500, up to 220 mm for DN 2200), with linear interpolation for intermediate sizes.

Elevation handling

Missing Z-values are common in legacy datasets. The converter employs linear interpolation when multiple elevation points are known, assumes a flat profile when only one is available, and applies a configurable default value when no elevations exist. To ensure transparency, the converter uses a custom property set, `DWAM150_DataQuality`, to document the assumptions made.

Geometry extrusion

Pipes are modeled as `IfcPipeSegment` entities by extruding the assigned profile along the direction vector between consecutive geometry points. When a pipe section contains more than two geometry points, the converter creates multiple connected linear segments to approximate the polyline path. Manholes and inspection chambers are created as `IfcDistributionChamberElement` elements, with parametric geometry —consisting of a main shaft, an eccentric tapered cone, adjustment rings, and a cover plate— with a forced minimum height of 0.8 meters to prevent geometric degeneration when cover and invert elevations are identical.

Visual quality indicators

To ensure transparency regarding data completeness, color-coding distinguishes elements with complete data from those requiring assumptions:

- **Blue**: Pipes with complete data

- **Red**: Pipes with missing or interpolated values
- **Yellow**: Nodes with complete data
- **Orange**: Nodes with incomplete data

Semantic enrichment framework

The DWA-M 150 schema is optimised for data transport between systems, encoding attributes as 5-character field codes (e.g., `HG304` for pipe material) with values referencing internal code tables. For a functional digital twin, this technical encoding must be transformed into human-readable, queryable IFC property sets. The semantics module handles this transformation through code mapping, property set assignment, and spatial matching for data fusion.

Data mapping strategy

The DWA-M 150 standard encodes attributes using a two-level reference system. At the global level, field codes identify attribute types (e.g., `KG305` for the manhole material). At the local level, attribute values are numeric codes referencing lookup tables within the XML file. A mapping dictionary translates 5-character field codes into human-readable German property names. For example:

- `HG001` → ''Bezeichnung'' (Designation)
- `HG304` → ''Material''
- `HG305` → ''Profilart'' (Profile Type)
- `HZ001` → ''Station'' (Position along pipe)
- `HZ014` → ''Schadensklasse'' (Damage Classification)

Semantic values are resolved using a two-stage lookup: the parser first dynamically builds a dictionary from the XML's internal Reference Tables (`<RT>` elements), then falls back to built-in mappings for codes not defined in the source file. For instance, in the material field `HG304`, a value of ''10'' is resolved to ''Beton'' (Concrete), ''50'' to ''Steinzeug'' (Vitrified Clay), and ''70'' to ''Kunststoff'' (Plastic).

Spatial matching algorithm

The converter supports two strategies for assigning semantic data to geometry: name-based matching and spatial matching. By default, the converter uses the pipe designation (`HG001`) to find the corresponding IFC element by name. This works well when the same data source provides both geometry and attribute data. Spatial matching becomes necessary in several scenarios. First, when inspection data comes from a different source than the master data, such as when importing camera survey results from a third-party contractor. Second, when pipe names have changed between data exports or contain inconsistencies. Third, when working with geometry obtained from 3D laser scans, point clouds, or photogrammetry that does not include attribute information. The spatial matching algorithm builds an index tree of all geometric elements in the existing IFC model and calculates the geometric centroid, typically the midpoint between start and end points for pipe sections. First, the algorithm queries the spatial tree at the exact centroid location. If exactly one element intersects that point, a match is found. If no elements are found, the search radius expands exponentially—doubling from an initial radius of 10 cm (0.1, 0.2, 0.4 m, etc.) up to a configurable maximum (default: 5 m) to ensure that coordinate differences due to rounding or survey accuracy do not prevent valid matches. When multiple elements fall within the search radius, the match is considered ambiguous, and no automatic association is made. The converter reports these ambiguous cases so that operators can resolve them manually. The centroid-based approach was chosen for its simplicity. Alternative strategies such as bounding-box overlap, direction-vector similarity, or topological connectivity matching were not implemented in the current version, as the centroid approach proved sufficient for straight pipe segments in the test dataset. The limitations of this algorithm for curved segments and dense junction areas are discussed further in Section 4.5.

Property set assignment

Once a match is found, the attributes are assigned to the IFC object using custom property sets prefixed with `DWA_`. Custom property sets are needed because the IFC 4.3 standard does not define property sets that fully accommodate the DWA-M 150 domain model—for example, `Pset_Condition` provides only a single overall assessment field, whereas DWA requires a triple classification (tightness, structural stability, and operational reliability) with per-defect station coding. To ensure both usability and data lineage, two distinct property sets are created for each data category: one includes decoded values for end users, while the `*_raw` version preserves the original DWA codes for traceability and automated validation. For the static master data (''Stammdaten'' in German), which includes basic and geometry data elements, the following property sets are created:

- `DWA_Stammdaten`: Contains the decoded attribute values (e.g., Material: ''Concrete''). This is intended for end-users in BIM viewers.
- `DWA_Stammdaten_raw`: Preserves the original DWA code values (e.g., Material: ''10'').

For inspection and condition data, additional property sets are created following the same approach:

- `DWA_Inspektionsdaten` / `..._raw`: Inspection metadata.
- `DWA_Zustandsdaten` / `..._raw`: Condition observation data.
- `DWA_DamageReference`: Links damage markers to parent pipes with station, classification, and a human-readable damage description.

Inspection data integration

Beyond static master data, the converter supports inspection records `<HI>` and condition observations `<HZ>`. In

DWA-M 150, damage observations are recorded as linear stations — distances in meters from the inspection start point (typically the upstream manhole). The `HZ001` field stores this 1D station value, which is converted to a 3D world coordinate by interpolating along the pipe's direction vector. Individual damage observations are visualized as 3D ring-shaped color-coded markers, as shown in Figure 2, according to the German condition classification:

- **Class 0 (Green):** No damage / informational
- **Class 1 (Red):** Immediate action required
- **Class 2 (Orange):** Short-term action needed
- **Class 3 (Yellow):** Medium-term action
- **Class 4 (Grey):** Long-term monitoring

This enables operators to immediately identify critical sections requiring maintenance intervention.

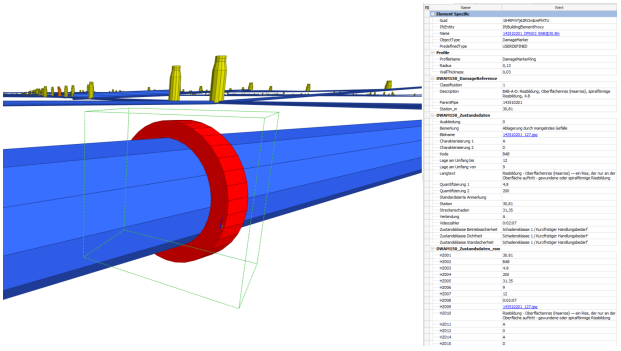


Figure 2: Damage Marker shown in BIMvision

Results and discussion

The workflow was validated using a combination of real-world and synthetic datasets. Real municipal master data from the German municipality of Alsdorf was used to test geometry generation and semantic enrichment with static master data. Although name-based matching is the primary method for data integration from the same source, the spatial matching algorithm was also tested using this dataset. Synthetic inspection data was generated to test the inspection data integration, covering the full range of damage classifications and several station positions across the pipes. The following sections evaluate geometric accuracy, semantic data verification, inspection data integration, and system performance, followed by a discussion of current limitations.

Geometric model accuracy

Georeferencing accuracy was validated by overlaying the generated IFC model on a Digital Terrain Model (DTM) created in Autodesk Civil 3D. The DTM was constructed using GeoTIFF elevation data obtained from the state geodata portal of North Rhine-Westphalia (Geoportal NRW), and orthophotos as visual overlay, referenced to the ETRS89 / UTM zone 32N coordinate system (EPSG:25832). The unmodified IFC file was then superimposed on the DTM using both Desite and Autodesk Infraworks, as illustrated in Figure 3. Coordinates from several pipe elements along the outer boundary of the network were extracted from the IFC file and from the combined model and verified against the original DWA-M 150 source data. This comparison confirmed that UTM coordinates were correctly preserved in the generated model.

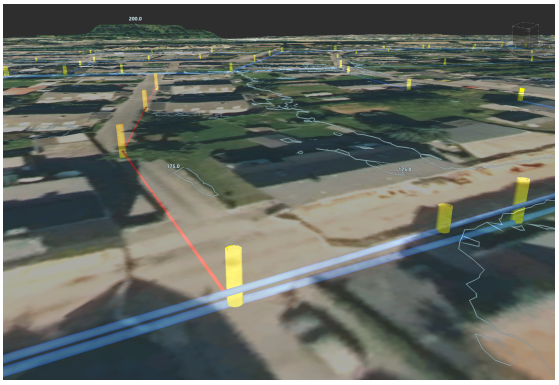


Figure 3: Coordination Model in Autodesk InfraWorks

Semantic data verification

The semantic enrichment was validated using real-world municipal datasets from the German municipality of Alsdorf, containing 701 pipe sections and 702 nodes. Both matching strategies were tested independently. Since both geometry and master data originated from the same DWA-M 150 source and were consistent, name-based matching achieved a 100% success rate, linking all 701 pipes and 702 nodes to their corresponding semantic attributes. Spatial matching achieved a success rate of 97.6% for pipes (684 of 701 matched), with 17 pipes (2.4%) remaining unmatched. 14 of these 17 pipes were curved segments with more than two geometry points, where the centroid-based matching fails due to geometric segmentation. The remaining three pipes were unmatched due to overlapping geometries at complex junction points, where multiple pipes shared near-identical centroids. For nodes, spatial matching achieved 99.6% accuracy (699 of 702). The resulting IFC models were inspected in BIMvision, where the dual property set structure enabled visual verification of key properties (such as material types), while retaining the raw data for quality checks.

Inspection data integration

To validate the inspection data integration, synthetic inspection data was generated following the official DWA-M 150 example file structure. The synthetic dataset was created by selecting 50 pipes from the Alsdorf master data and generating inspection records with DWA-M 150 element hierarchy: pipe sections (`<HG>`) with geometry objects (`<GO>/<GP>`), inspection headers (`<HI>`), and condition observations (`<HZ>`). The dataset contains 251 condition observations distributed across all five damage classifications. The converter successfully parsed all inspection records and generated correctly positioned damage markers with appropriate severity color-coding. Linear station values across the pipes (`HZ001`) were correctly trans-

formed to world coordinates via linear interpolation. The precision is limited only by the input coordinate resolution (typically millimetres in UTM) and the station value precision from the source data. However, the matching algorithms could not be tested using the synthetic data, as it was derived directly from the master data with identical pipe names and coordinates, resulting in a 100% matching rate.

An alternative validation method could include introducing controlled perturbations, like coordinate offsets (e.g., 0.5–2.0 m), renaming pipe identifiers to simulate naming inconsistencies between data sources, or merging inspection records from a genuinely independent data export. As no more heterogeneous datasets were available at the time of testing, this remains a limitation, and future work includes testing with real inspection data from third-party contractors, or alternatively, introducing synthetic perturbations to simulate real-world data inconsistencies.

System performance and quality assessment

The developed workflow successfully achieved its primary objective of automated generation and semantic enrichment of 3D sewer models from DWA-M 150 master data. The resulting model features a complete geometric representation of pipes and nodes, correct georeferencing, and comprehensive semantic information based on the DWA-M 150 standard embedded in property sets. This demonstrates the feasibility of automated, standard-based data transfer from DWA-M 150 to IFC using open-source tools such as `IfcOpenShell`. In terms of quality, the converter captured all objects from the source data, provided that the input data adhered to the standard. Semantic consistency was verified by checking that the attributes matched the DWA identifiers, with the raw values preserved in a separate property set to mitigate data loss from missing reference table entries in the source. The output adheres to IFC 4.3.2.0 and was validated using the buildingSMART Validation Service. The models were additionally loaded in BIMvision and Blender (Bonsai) to confirm practical interoperability. Performance was satisfactory, with geometry generation completing in approximately 15 seconds and semantic enrichment in approximately 18 seconds (name-based) or 14 seconds (spatial) on a workstation equipped with an AMD Ryzen 9 8945HS processor and 32 GB RAM.

Limitations and challenges

Regarding the geometric model accuracy, the system faces challenges primarily related to source data quality, such as missing height values, which necessitated defaults. The manholes are represented as simplified parametric geometries, consisting of cylindrical elements, a tapered cone, and a cover plate. This representation does not account for asymmetric bases, varying wall thicknesses, or non-circular shaft cross-sections. A realistic representation of the nodes would be more demanding. In DWA-M 150, curved pipes are represented by three or more geometry points. During IFC generation, these are approximated as multiple straight segments. While this linear representation facilitates the centroid-based matching algorithm, arc-based swept solids could represent curved pipe alignments more smoothly in future iterations. For the semantic enrichment, the centroid-based spatial matching algorithm shows limitations when handling curved pipe sections. The spatial matching algorithm calculates the centroid from all original geometry points, but this centroid may not fall within any single IFC segment's bounding box. It should be noted that the DWA-M 150 schema only provides discrete coordinate points, making linear approximation the only deterministic geometric representation; fitting arc or spline curves would require additional assumptions about curvature that are not supported by the source data. Matching precision could be improved by calculating separate centroids for each segment of curved pipes or by adopting alternative geometric matching strategies. As more diverse datasets become available in the future—including geometry from laser scans, where curved representations may exist—alternative matching strategies for this specific case could also be evaluated.

Regarding validation, the reliance on synthetic data means that the matching algorithms remain untested for inspection and condition data. The current implementation can therefore only be interpreted as a proof-of-concept for the pipeline architecture, with the matching algorithm requiring further validation as real-world datasets become available. Additionally, specific sewer properties like ''Flood Area'' lack native IFC entities, requiring mapping to IFC spatial entities (eg. `IfcSpatialZone` or `IfcGeographicElement`), potentially reducing interoperability with downstream analysis tools.

Comparison with existing approaches

Compared to proprietary Sewer Information Systems and transformation workflows, this open-source approach offers significant advantages in transparency and flexibility. The separation of geometry generation from semantic enrichment allows for the enrichment of any valid IFC geometry, such as that derived from 3D scans, provided the coordinates match. This makes the workflow particularly suitable for research and development contexts, where adaptability is required. Regarding transferability to other infrastructure domains, the workflow components differ in their degree of generalizability. The geometry generation engine (profile extrusion, coordinate transformation, elevation handling) and the spatial matching algorithm can be reused directly for any node-edge network topology. In contrast, the DWA-M 150 parser, the field code mapping dictionaries, and the reference table resolution logic are domain-specific and would require replacement when adapting the workflow to other standards (such as DVGW for gas and water networks). The IFC entity mapping (e.g., which DWA element types correspond to which IFC classes) is also domain-specific but follows a pattern that can be systematically replicated for other schemas.

Conclusion and outlook

This section consolidates the workflow's main contributions and outlines future development directions.

Key achievements

The study successfully developed a two-stage automated workflow that translates DWA-M 150 sewer infrastructure data into a fully structured, georeferenced, and semantically enriched IFC 4.3 model. This work validates the feasibility of bridging traditional Sewer Information Systems (KIS) and Building Information Modeling (BIM) using open standards (IFC, DWA-M 150).

Key technical achievements include decoupling geometric extrusion from semantic enrichment to enable data integration from heterogeneous sources. To facilitate this process, the study developed and evaluated name- and centroid-based spatial matching algorithms, achieving integration accuracies of over 97% for the spatial matching of pipe segments on real-world municipal datasets. Furthermore, the workflow effectively dismantles existing data silos by translating specific DWA-M 150 attributes into standardised, human-readable IFC property sets, while retaining the original codes for algorithmic validation. Finally, this approach has been extended to operational condition monitoring by translating 1D linear inspection stations into visual, colour-coded 3D damage markers. Having been validated against official buildingSMART configuration rules and coordinated with municipal digital terrain models, this Legacy-to-BIM methodology establishes a tangible, standard-compliant foundation for future predictive digital twin applications in underground infrastructure.

Future improvements

Technical improvements could address geometric and topological limitations. Currently, pipes are linearly simplified; integrating OpenCASCADE/STEP via `IfcOpenShell` would allow for complex profiles and curved segments. Connection geometry between pipes and nodes could also be automated for cleaner transitions. The spatial matching algorithm could be extended with directional similarity scoring and topological connectivity analysis to resolve ambiguous cases. To support hydraulic simulations, semantic flow definitions could be implemented using the IFC concept of ''Port Nesting''. Furthermore, hardcoded defaults for missing data could be replaced with dynamic calculations based on slope or rule-based plausibility checks. For inspection data, integration with the BIM Collaboration Format (BCF) would enable linking damage observations to external documentation such as inspection photographs and video frames. However, BCF currently lacks native fields for precise damage coordinates, requiring either format extensions or hybrid approaches.

Standards and digital twin potential

While the current approach aligns with DWA-M 150, the upcoming DWA-M 145-3 standard (incorporating GML for coordinates) will require adaptations in XML parsing logic, though the general methodology remains valid. The enriched IFC model serves as a foundation for a digital twin, enabling flood prevention simulations and integration with sensors for predictive maintenance. While developed for wastewater infrastructure, the methodology is transferable to other utility networks characterized by similar data structures: water distribution, gas networks, and district heating systems all employ node-edge topologies with standardized attribute schemas.

Acknowledgments

This research is part of the *KaSyTwin* project funded by the Federal Ministry of Transport (BMV) within the mFUND program under grant 01FV2059A.
The authors would like to thank *Regionetz GmbH* in Aachen, the *City of Alsdorf*, and *Stadtentwässerungsbetriebe Köln* for providing data and supporting the research efforts.

References

Belsky, M., Sacks, R., and Brilakis, I. (2016). Semantic enrichment for building information modeling. Computer-Aided Civil and Infrastructure Engineering, 31(4):261–274.

Berger, C., Falk, C., Hetzel, F., Pinnekamp, J., Ruppelt, J., Schleiffer, P., and Schmitt, J. (2020). Zustand der kanalisation in deutschland: Ergebnisse der dwa-umfrage 2020. KA Korrespondenz Abwasser, Abfall, 67(12/2020).

Bloch, T. (2022). Connecting research on semantic enrichment of bim - review of approaches, methods and possible applications. Journal of Information Technology in Construction, 27:416–440.

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., Lun, D., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Frolova, N., Ganora, D., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Volpi, E., Wilson, D., Zaimi, K., and Živković, N. (2019). Changing climate both increases and decreases european river floods. Nature, 573(7772):108–111.

buildingSMART International (2020). Enabling an ecosystem of digital twins.

Du, H., Wei, L., Dimitrova, V., Magee, D., Clarke, B.,

Collins, R., Entwisle, D., Eskandari Torbaghan, M., Curioni, G., Stirling, R., Reeves, H., and Cohn, A. G. (2023). City infrastructure ontologies. Computers, Environment and Urban Systems, 104:101991.

DWA, editor (2010). Datenaustauschformat für die Zustandserfassung von Entwässerungssystemen. Number 150 in DWA-Regelwerk Merkblatt DWA-M. DWA, Hennef, april 2010 edition.

DWA, editor (2021). Anforderungen an ein Datenmodell und Schnittstelle (Entwurf). Number DWA-M 145-3 in Kanalinformationssysteme. Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall, Hennef (Sieg), entwurf, september 2021 edition.

Edmondson, V., Cerny, M., Lim, M., Gledson, B., Lockley, S., and Woodward, J. (2018). A smart sewer asset information model to enable an 'internet of things' for operational wastewater management. Automation in Construction, 91:193–205.

Ghorbani Bam, P., Rezaei, N., Roubanis, A., Austin, D., Austin, E., Tarroja, B., Takacs, I., Villez, K., and Rosso, D. (2025). Digital twin applications in the water sector: A review. Water, 17(20):2957.

Jäkel, J.-I., Heinlein, E., Gölzhäuser, P., Kellner, M., Klemt-Albert, K., and Reiterer, A. (2025). How to digitise bridge structures—a systematic review of the status quo for creating digital bim models of existing bridge structures in the operational phase. Infrastructures, 10(3):47.

Jiang, S., Feng, X., Zhang, B., and Shi, J. (2023). Semantic enrichment for bim: Enabling technologies and applications. Advanced Engineering Informatics, 56:101961.

Kellner, M., König, T., Jäkel, J.-I., Klemt-Albert, K., and Reiterer, A. (2025). 3d bridge segmentation using semi-supervised domain adaptation. Automation in Construction, 172:106021.

Kellner, M., Schmitt, A., and Reiterer, A. (2026). Automatic generation of 3D bridge models from 3D point clouds. Results in Engineering, 29:109532.

Kretschmer, F., Franziskowski, S., Huber, F., and Ertl, T. (2023). Chances and barriers of building information modelling in wastewater management. Water science and technology : a journal of the International Association on Water Pollution Research, 87(7):1630–1642.

Sacks, R., Ma, L., Yosef, R., Borrmann, A., Daum, S., and Kattel, U. (2017). Semantic enrichment for building information modeling: Procedure for compiling inference rules and operators for complex geometry. Journal of Computing in Civil Engineering, 31(6).

Schönfelder, P., Aziz, A., Faltin, B., and König, M. (2023). Automating the retrospective generation of As-is BIM models using machine learning. Automation in Construction, 152:104937.

Smarsly, K., Ahmad, M. E., Peralta, P., Al-Zuriqat, T., Al-Nasser, H., Dragos, K., and Geck, C. C. (2025). Digital Twins, Architectures, and Elements in Civil Engineering – A Multivocal Literature Review. In Francis, A., Miresco, E., and Melhado, S., editors, Advances in Information Technology in Civil and Building Engineering, volume 628, pages 507–518. Springer Nature Switzerland.

Tse, J. H. (2016). Challenges of Adopting Knowledge-based Building Information Modeling for E&M Asset Management Supplemented with Mobile Solutions - A Case Study in Public Sewage Pumping Facilities. American Journal of Operations Management and Information Systems, 1:17–33.

Valles Gomez, R., Jamora, J. L., Gölzhäuser, P. J., Hartmann, S. V., Ravandi, A. B., Morgenstern, H. W. L., and Klemt-Albert, K. (2025). An ontology-based approach for sewer infrastructure management. In Tagungsband des 36. Forum Bauinformatik : 24.-26. September 2025. RWTH Aachen University.

Wang, H., Pan, Y., and Luo, X. (2019). Integrating BIM and GIS in green building certification: A building lifecycle perspective. Automation in Construction, 103:41–51.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

TOPOLOGY-DRIVEN SEMANTIC INTEGRATION OF INDEPENDENT BIM MODELS FOR RENOVATION ANALYSIS

Jacopo Cassandro and Claudio Mirarchi
Department of Architecture, Built Environment, and Construction Engineering (DABC),
Polytechnic of Milan, Milan, Italy

Abstract

This paper proposes a topology-driven semantic integration approach to align as-is and to-be BIM models in building renovation scenarios. Starting from independent IFC files, a surface-based topological analysis infers physically meaningful contacts between envelope objects using coplanarity, opposite orientation, vertical overlap, and an overlap-coverage metric. The detected adjacencies are then represented as cross-model semantic relations by instantiating `bot:Interface` links between the corresponding `bot:Element` individuals in a dedicated knowledge graph, preserving model independence. The resulting topology enables cross-model SPARQL queries to extract and aggregate energy-related data at the envelope-system level without IFC merging or manual data preparation.

Introduction

Building renovation workflows increasingly rely on Building Information Modeling (BIM) based digital representations to support design, analysis, and decision-making processes (Han et al., 2025; Ismail et al., 2017). In practice, however, renovation projects are typically characterized by multiple BIM models representing different project phases, such as the existing (as-is) condition and the proposed retrofit (to-be) intervention. These models are often developed at different time stages, so they are simultaneously independent of each other, stored in separate IFC files, and lack explicit machine-interpretable semantic relationships that describe how the introduced objects physically relate to the existing building (Pauwels and Terkaj, 2016; Rasmussen et al., 2017).

This fragmentation represents a significant challenge for automatic data analysis by machines, such as the extraction of energy assessment for the envelope system. While the physical reality of a renovated building envelope is composed of coherent construction systems (e.g., walls upgraded with insulation layers), this coherence is not explicitly represented in conventional BIM models. As a result, properties relevant for performance evaluation are distributed across independent models, preventing automated system-level reasoning and requiring manual interpretation, model merging, or ad hoc data preparation.

Existing approaches to BIM-based analysis often assume that relevant relationships are already encoded within a single model or can be derived through explicit semantic links. However, such assumptions rarely hold in renovation scenarios, where as-is and to-be models remain semantically detached. Geometry-based methods can detect spatial proximity or collisions, but they typically do not provide a structured mechanism to transform geometric evidence into persistent, machine-interpretable relations suitable for higher-level analysis (Gan et al., 2022). Conversely, ontology-based approaches offer rich semantic representations, yet they rely on the availability of explicit relationships that are not present in independent BIM models (Karmakar and Delhi, 2023; Moreira, 2024).

This paper addresses this gap by proposing a geometry-driven and ontology-based approach for aligning independent as-is and to-be BIM models through surface-level topological analysis and semantic linking. Rather than merging models or imposing predefined semantic structures, the proposed method infers physically meaningful adjacency relations directly from geometric surface evidence and explicitly encodes them as semantic relations, reusing the Building Topology Ontology (BOT) and Building Element Ontology (BEO). In this way, objects that are fragmented across different models but belong to the same physical construction system can be coherently related in a machine-interpretable manner.

The main contribution of this work lies in enabling cross-model reasoning over building envelope systems without compromising model independence. By transforming geometric contact evidence into explicit semantic constructs, the proposed approach allows distributed BIM data to be aggregated, queried, and exploited for performance-oriented analyses, such as energy evaluation(Utkucu et al., 2025; Massafra et al., 2024). The approach is demonstrated through a case study focusing on envelope retrofit, showing how independent BIM models can be semantically linked to support automated data extraction and downstream simulation workflows.

The paper is structured as follows. Section *Methodology* presents the proposed methodology, including geometry-driven topological analysis and semantic enrichment. Section *Topological analysis and overlap metrics* formalizes the algorithm for the analysis of objects surface-based matching and the associated overlap metric. Section *Case study and results* reports the case study and discusses the

results, while the final sections address semantic linking, energy-oriented data exploitation, and future work.

Background

Topology in BIM

Several studies have investigated the representation of topological relationships within BIM models to support advanced querying, navigation, and analysis tasks. Graph-based approaches encode building components and their relations as nodes and edges, allowing the computation of shortest paths, connectivity patterns, and higher-level structural properties (Han et al., 2025). Graph databases have been used to transform IFC models into object graphs, supporting advanced querying, algorithms, and topological analysis in IFC versions (Ismail et al., 2017). In parallel, IFC-based graph data models use weighted graph principles and automatic algorithms to extract and query topological relationships among 3D elements more efficiently than traditional 2D/3D representations (Khalili and Chua, 2015). More recent work has introduced the notion of Topological BIM (TBIM), a space-oriented BIM representation tailored for building performance management; TBIM relies on topological and conditional modeling principles to derive semantically enriched information models to build performance simulation (Massafra et al., 2024). Graph-based topological models have also been used to support semi-automated dataset generation for residential buildings, showing that explicit spatial graphs can serve as a robust backbone for downstream analytical tasks (Massafra, 2025).

Beyond purely topological representations, several approaches leverage geometry-based techniques—such as surface adjacency, coplanarity, and contact area analysis—to infer spatial relationships between building elements. These methods extend graph-based BIM representations by incorporating geometric evidence to identify physically compatible interfaces between components, for instance, in modular or surface-driven modeling strategies (Gan et al., 2022). However, geometry-based inferences are typically applied within a single BIM model and produce transient results that are not persisted as explicit semantic relations, limiting their applicability to cross-model renovation scenarios.

BIM-to-RDF

The translation of BIM data into semantic representations has been extensively explored as a means to improve interoperability, data integration, and querying capabilities. Ontologies such as IfcOWL enable a comprehensive mapping of the IFC schema into RDF triples, supporting formal reasoning and semantic querying. Pauwels and Terkaj (2016) formalized a standardized EXPRESS-to-OWL conversion procedure specifically targeting IFC, producing a recommendable ifcOWL ontology in OWL2 DL that balances fidelity to the original schema with practical usability. Due to the size and complexity of full IFC-to-RDF mappings, several lightweight ontologies have been proposed to represent core aspects of building topology and element classification in a more tractable way; notable examples include BOT and BEO ontologies, which focus respectively on spatial/elementary topology and on reusable element taxonomies (Rasmussen et al., 2017). More recently, domain-specific performance ontologies like the Holistic Building Performance Ontology (HBPO) provide unified schemas to connect BIM data with multi-domain building performance simulation tools (acoustics, lighting, energy) (Utkucu et al., 2025). In the Linked Building Data (LBD) context, these ontologies are combined with IFC and other domain schemas to support data-driven modeling of building topology, while frameworks such as Onto4Reuse address ontology reuse and alignment across BIM-related domains (Moreira, 2024). In parallel, semantic enrichment has been proposed as a strategy to reconstruct implicit or missing information in IFC-based models, particularly where exported data lack the relational structure needed for higher-level reasoning. Rule-based enrichment techniques infer missing topological relations, reorganise building hierarchies, and support automated code compliance checking by encoding domain knowledge as declarative rules (Karmakar and Delhi, 2023). Nevertheless, most of these semantic approaches assume that relevant relationships are explicitly present in the source models, or can be directly mapped from the IFC schema, and do not tackle the derivation of new topological relations from geometric evidence, nor the reconstruction of topology between independent as-is and to-be models. These approaches confirm that BIM models often lack explicit relational topology and that enrichment is necessary to make such information machine-interpretable. However, existing semantic enrichment methods are typically domain-specific, rely on rule sets tailored to particular use cases (e.g. regulatory checking or asset management), and operate on single BIM models. As a result, they offer limited support for renovation workflows in which as-is and to-be models are authored independently, stored in separate IFC files, and need to be semantically linked through derived topological relations.

Research gaps and research questions

The literature review highlights significant advances in graph-based BIM representations, geometry-driven spatial analysis, and semantic enrichment techniques. However, these approaches are typically developed in isolation and do not fully address renovation workflows characterized by the coexistence of independent as-is and to-be BIM models. In particular, there is a lack of methods capable of deriving persistent, machine- interpretable adjacency relations from geometric evidence and exploiting them for system-level performance analysis.

This research addresses the following questions:

- RQ1: How can adjacency relations between independent BIM models be inferred from surface-level geometric evidence?
- RQ2: How can such relations be semantically created

without model merging?
- RQ3: How can the resulting semantic model support performance-oriented data extraction?

Methodology

The proposed methodology enables the identification and semantic formalization of cross-model adjacency relations between building envelope objects belonging to independent as-is and to-be BIM models, which are typically disconnected in renovation workflows. The approach combines geometry-driven topological analysis with semantic knowledge graph representation, enabling aggregation of relevant properties (e.g., thermal transmittance) at the envelope-system level, despite their distribution across separate IFC files. This enables the use of domain-specific analysis tools (e.g., energy simulation software) to evaluate renovation scenarios without requiring the IFC model to be merged.

The proposed method transforms geometric surface-contact evidence into semantic envelope systems, thereby enabling the integration of distributed BIM data for energy performance evaluation. By decoupling geometric analysis from semantic representation, the approach preserves model independence while supporting cross-model reasoning and system-level assessment.

As illustrated in Figure 1, the workflow is structured into four main steps: independent IFC model, geometry-driven topological analysis and semantic translation, semantic enrichment through topological relation, and data extraction for domain-specific analysis.

Step 1 – Independent IFC model.

The IFC^{as-is} model is analysed and a corresponding IFC^{to-be} model is generated. Both models are treated as independent data sources, each containing specific objects and performance-related properties associated with different project phases. The objects of both models lack explicit relationships that a machine can interpret.

Step 2 – Geometry-driven topological analysis and semantic translation.

- *Step 2.a – Surface-based topological analysis.*
  A geometry-driven topological analysis is performed on the geometric representations of the as-is and to-be models to identify surface-level contact evidence between envelope objects. Topological adjacency is inferred from geometric conditions, including the detection of coplanar and oppositely oriented surfaces exhibiting a significant overlap area. This process produces a set of candidate adjacency relations expressed as pairs of element identifiers (GUID) and associated overlap metrics.
- *Step 2.b – Semantic representation of BIM data.*
  In parallel, the IFC^{as-is} and IFC^{to-be} models are independently translated into RDF reusing BOT and BEO ontology, generating two separate knowledge graphs (KG) that preserve the original model structure, geometry references, and performance-related properties.

Step 3 – Semantic enrichment based on geometry-driven topological analysis.

The geometric adjacency evidence obtained in Step 2.a is subsequently materialized as new semantic relations between the KG^{as-is} and KG^{to-be} generated in Step 2.b. Specifically, adjacency relations are instantiated as topological relations using the BOT ontology (e.g., `bot:Interface`), enabling the formal representation of cross-model envelope systems. This step represents the transition from geometry-driven evidence to explicit, machine-interpretable topological knowledge that can be queried, reasoned upon, and exploited in downstream analyses.

Step 4 – Data extraction for domain-specific analysis.

The enriched knowledge graph is queried using SPARQL to retrieve envelope-system-level data, including thermal properties originating from both the as-is and to-be models. These data are then provided as input to external energy analysis software to support detailed performance assessment of renovation strategies.

Topological analysis and overlap metric

This section builds upon the geometric processing and semantic preparation steps introduced in the previous methodology, focusing on the formal definition of topological analysis and its associated overlap metric. In the proposed approach, topological analysis is not assumed as an input condition but is derived from geometric surface evidence through a geometry-driven inference process. The objective is to identify physically meaningful adjacencies between building objects belonging to independent IFC_{as-is} and IFC_{to-be} models.

Geometry-driven topological analysis

Geometry-driven topological analysis is defined as the process of identifying adjacency relations between objects belonging to independent BIM models based on geometric surface evidence. Rather than relying on predefined semantic links or model merging, adjacency is inferred by analysing the spatial configuration and mutual compatibility of envelope surfaces.

This geometry-driven topological analysis enables the creation of cross-model relations that are not explicitly represented in the original IFC data, but are essential for system-level reasoning in renovation workflows.

Geometric preprocessing and surface abstraction

Each target building object (e.g., wall) is first converted from its native IFC geometric representation into a triangulated surface mesh. This tessellation step is required to homogenize heterogeneous IFC geometry definitions (e.g., swept solids, boundary representations) into a common computational format suitable for surface-based analysis.

Formally, each object Obj is represented as a set of trian-

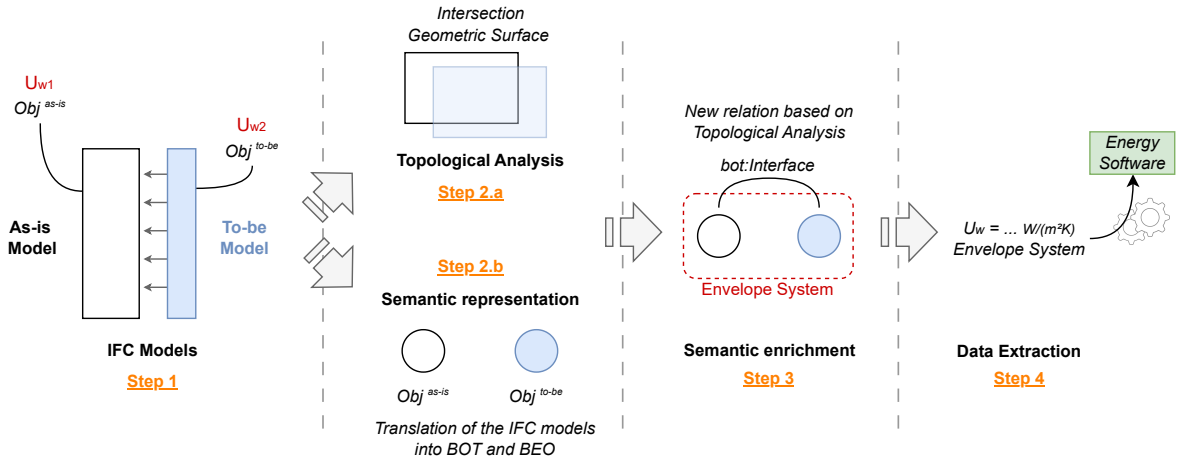


Figure 1: Methodology

gular faces:

$$\mathcal{T}(Obj) = \{t_k \mid t_k = (\mathbf{v}_{k1}, \mathbf{v}_{k2}, \mathbf{v}_{k3})\} \tag{1}$$

where $\mathbf{v}_{ki} \in \mathbb{R}^3$ are the triangle vertices expressed in a shared world coordinate system.
To reduce the computational complexity of subsequent operations, an axis-aligned bounding box (AABB) $B(Obj)$ is computed for each object and used for broad-phase spatial filtering. Only object pairs whose bounding boxes intersect are considered for further analysis.
For each triangle t_k, fundamental geometric descriptors are computed, including surface normal, area, and supporting plane. Individual triangles are then aggregated into *macro-faces* representing extended planar surface regions by grouping coplanar triangles with similar orientation. Each macro-face f is characterized by a unit normal $\mathbf{n}_f$, a supporting plane π_f, a total area $A(f)$, and a vertical extent $[z_{\min}, z_{\max}]$.
This surface abstraction step provides a compact and semantically meaningful representation of envelope geometry, enabling robust surface-level comparison across independent models.

Surface compatibility conditions

Let f_a and f_b be two macro-faces belonging respectively to an as-is object Obj_i^{as-is} and a to-be object Obj_j^{to-be}. The two surfaces are considered topologically compatible only if the following geometric conditions are simultaneously satisfied.

Opposite orientation

The two surfaces must face each other, ensuring the physical plausibility of a contact interface:

$$\angle(\mathbf{n}_{f_a}, -\mathbf{n}_{f_b}) < \varepsilon_\theta \tag{2}$$

where $\mathbf{n}_{f_a}$ and $\mathbf{n}_{f_b}$ denote the unit normal vectors of the two macro-faces.

Coplanarity

The supporting planes of the two surfaces must coincide within a geometric tolerance:

$$\text{dist}(\pi_{f_a}, f_b) < \varepsilon_d \tag{3}$$

where π_{f_a} denotes the plane associated with f_a and the distance is evaluated using a point of f_b.

Vertical overlap

To exclude geometrically coplanar but vertically misaligned surfaces, a vertical overlap constraint is introduced:

$$\min(z^a_{\max}, z^b_{\max}) - \max(z^a_{\min}, z^b_{\min}) > \tau_z \tag{4}$$

where $[z^a_{\min}, z^a_{\max}]$ and $[z^b_{\min}, z^b_{\max}]$ represent the vertical extents of the two macro-faces.
Together, these conditions ensure that only surface pairs capable of defining a physically meaningful interface are retained for further evaluation.

Overlap metric definition

For each pair of geometrically compatible macro-faces, a quantitative overlap metric is computed to assess the extent of the contact. Each surface is projected onto a local two-dimensional coordinate system defined on its supporting plane. The union of projected triangles forms a planar polygon representation P_f for each macro-face.
The effective contact surface area is computed as:

$$A_\cap(f_a, f_b) = \text{area}(P_{f_a} \cap P_{f_b}) \tag{5}$$

To provide a normalized and scale-independent measure, an overlap ratio is defined as:

$$\rho(f_a, f_b) = \frac{A_\cap(f_a, f_b)}{A(f_a)} \tag{6}$$

The overlap ratio expresses the proportion of the as-is surface that is covered by the to-be surface, enabling a quantitative comparison between different contact configurations.

Decision rule for adjacency inference

A topological adjacency relation between two objects Obj_i^{as-is} and Obj_j^{to-be} is inferred if at least one pair of their macro-faces satisfies the overlap condition:

$$\exists (f_a, f_b) \quad \text{s.t.} \quad \rho(f_a, f_b) \geq \tau \tag{7}$$

where τ is a user-defined threshold controlling the minimum required contact significance.
This rule allows multiple Obj_j^{to-be} to be associated with a single Obj_i^{as-is}, resulting in a one-to-many topological matching configuration. The inferred adjacency relations constitute the basis for subsequent semantic enrichment. Topological matching is thus formulated as a geometry-driven inference process, in which surface overlap metrics provide both qualitative adjacency relations and quantitative evidence supporting system-level aggregation.

Semantic linking and data exploitation

The set of inferred adjacency relations produced by the geometry-driven topological analysis is subsequently materialized in a semantic knowledge graph. To manage the complexity of IFC representations, the models are translated reusing the BEO ontology, which provides a lightweight taxonomy for objects such as `beo:Wall` and `beo:Insulation`, facilitating more efficient SPARQL querying compared to the direct translation of the EXPRESS schema into IfcOWL which could generate millions of RDF triples. Each objects pair $(Obj_i^{as-is}, Obj_j^{to-be})$ is mapped to its corresponding RDF individuals using the *IfcGlobalUniqueId* as a linking key. To maintain ontological correctness—given that `bot:adjacentElement` is formally restricted to Zone-to-Element relations—the physical contact between the Obj_i^{as-is} and Obj_j^{to-be} is represented through the `bot:Interface` class. This approach allows for the creation of a topological "bridge" between the two models, where a single interface instance is declared as `bot:interfaceOf` both the existing objects and the new retrofit objects. The inferred relations are asserted in a dedicated KG that operates as a federated layer over the original KG^{as-is} and KG^{to-be}. This architecture preserves the integrity of the original BIM data while enabling the explicit representation of cross-model adjacencies derived from geometric evidence. This semantic enrichment enables advanced cross-model querying of envelope systems through SPARQL. By relating objects coming from independent IFC models into a KG, specific data (e.g., *U*-values) can be structured, extracted, and used as input data for domain-specific simulation tools.

Case study and results

As a case study, a typical building envelope retrofit scenario was considered, featuring an existing wall with an opening (a window), as illustrated in Figure 2. Despite its simplicity, the case study already includes relevant complexities, such as multiple elements, the presence of an opening, and partial geometric misalignments (e.g., vertically offset objects). These aspects allow demonstrating the method's ability to correctly identify meaningful adjacencies while effectively excluding false positives. The IFC^{as-is} model represents the existing state, characterized by a multi-layered composition (internal plaster, masonry, and external plaster). The IFC^{to-be} model includes only the retrofit objects, namely an external insulation system applied to the existing wall and a final plaster finish.

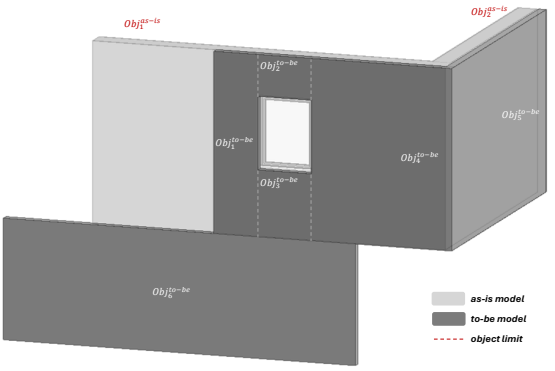


Figure 2: CaseStudy

To evaluate the proposed geometry-driven topological analysis, the retrofit objects were intentionally modeled as separate objects (Obj_1^{to-be}, Obj_2^{to-be}, Obj_3^{to-be}, Obj_4^{to-be}, and Obj_5^{to-be}) and some with a vertical offset relative to the existing objects (Obj_6^{to-be}). This modeling choice was adopted to test the algorithm's ability to distinguish between geometrically coplanar surfaces but with vertically misaligned surfaces and those with physically significant adjacencies, ensuring that only overlapping object surfaces are identified as valid contacts.
The case study was designed as a simplified yet representative scenario that reflects common building envelope renovation scenarios. Although limited in scale, it provides a controlled setting to evaluate the proposed method before its application to larger and more complex case studies within the ongoing research project.
Both models were exported using a shared coordinate system in IFC 4.3 ADD2 format and subsequently processed using a custom Python code (Python 3.10) to perform the geometry-driven topological analysis. In parallel, the IFC models were translated into lightweight semantic representations based on the BOT and BEO ontologies with the IFCtoLBD converter (Oraskari et al., 2024). The reuse of these ontologies provides a more compact and tractable structure compared to full ifcOWL representations, enabling efficient graph-based data management for the renovation process. The resulting RDF graphs were stored and queried using an Ontotext GraphDB.

Geometric matching results

The geometry-driven topological analysis was executed automatically on the two IFC models described in the pre-

vious section, following the methodology steps presented in the section *Topological Analysis and Overlap Metric*. The algorithm successfully identified the contact surfaces between existing objects (Obj_1^{as-is} and Obj_2^{as-is}) and new retrofit insulation objects (Obj_1^{to-be}, Obj_2^{to-be}, Obj_3^{to-be}, Obj_4^{to-be}, and Obj_5^{to-be}), while correctly excluding non-relevant surfaces (Obj_6^{to-be}).
No false-positive adjacencies were detected between geometrically coplanar but vertically misaligned surfaces (Figure 3).

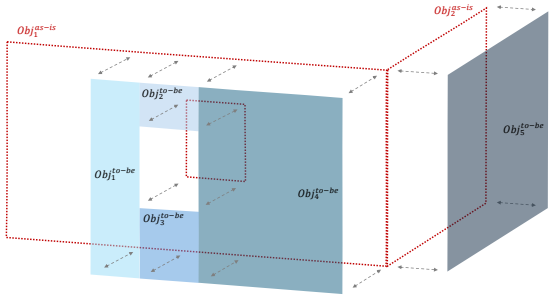


Figure 3: Geometry-driven topological analysis results.

The two existing walls of the IFC^{as-is} model were found to be adjacent to objects in the IFC^{to-be} model. In the scenario visible in Figure 2, a single existing object (Obj_1^{as-is}) was correctly associated with four different objects (Obj_1^{to-be}, Obj_2^{to-be}, Obj_3^{to-be}, and Obj_4^{to-be}), corresponding to the segmented area around the window opening. For all four objects, the coverage ratio computed with respect to the Obj_j^{to-be} surfaces exceeded 94%, indicating that each retrofit object is almost entirely supported by the existing wall, despite covering only a limited portion of its surface.
On the other side, a one-to-one correspondence between an existing object (Obj_2^{as-is}) and a retrofit object (Obj_5^{to-be}), with surface coverage ratios close to 100% on both sides, reflecting a full surface alignment.
The geometry-driven topological analysis results are summarized in Table 1, which respectively illustrate the contact surfaces identified between the existing objects (Obj_i^{as-is}) and the retrofit objects (Obj_j^{to-be}), including the corresponding surface coverage ratios for each detected adjacency. The results confirm that only geometrically and vertically aligned surfaces are selected, while areas associated with openings or without existing objects behind are correctly excluded. Furthermore, the results confirm the ability of the proposed algorithm to identify meaningful adjacency relations across independent IFC models and to quantify their geometric consistency through surface-based overlap metrics.

Semantic linking results

The adjacency relations inferred at the geometric level were automatically generated in a dedicated RDF case-study graph (KG^{link}). Starting from the set of matched object pairs produced by the geometry-driven topological analysis, the linking process was executed automatically.
Each pair of matched objects was identified across the independent KG_{as-is} and KG_{to-be} using the *IfcGlobalId* as a unique linking key. For each identified contact, a `bot:Interface` individual was automatically instantiated in KG^{link} and asserted as `bot:interfaceOf` both the existing `bot:Element` and the corresponding retrofit `bot:Element`. This approach enables the explicit and ontologically consistent representation of object-to-object adjacencies across independent BIM models.
All interface instances and their properties were generated directly from the output of the geometry-driven algorithm, ensuring full traceability between geometric evidence and semantic assertions (Figure 4). The resulting structure transforms a set of geometrically fragmented objects into a coherent, machine-interpretable system in which individual components are explicitly related through formal semantic constructs, thereby enabling consistent reasoning, querying, and exploitation via SPARQL.

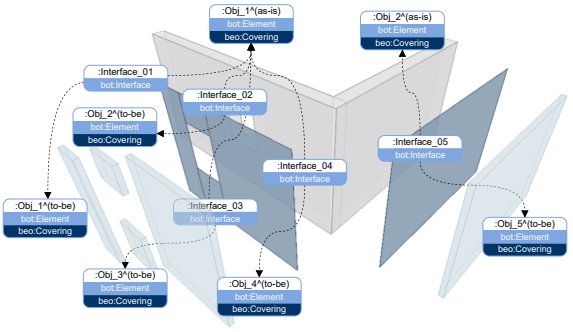


Figure 4: Ontology-based adjacencies as `bot:Interface` instances linking as-is and to-be objects.

Energy-oriented data extraction

The semantic linking introduced in the previous step enables the formulation of cross-model SPARQL queries that directly operate on physically meaningful object-to-object relations. By exploiting the `bot:Interface` individuals created in KG^{link}, existing and retrofit objects that are physically adjacent in the renovated envelope can be queried jointly, despite originating from independent KG^{as-is} and KG^{to-be} representations.
In this context, energy-related properties are retrieved by traversing the `bot:interfaceOf` relations and accessing the property states associated with the linked objects. Specifically, the thermal transmittance values of both the as-is and the to-be objects are extracted through their respective IFC property representations, using the *IfcGlobalId* as a stable linking key. This allows the selective retrieval of thermal properties only for object pairs that have been previously identified as physically adjacent through geometry-driven analysis.
As a result, thermal transmittance values that are distributed across separate BIM models can be aggregated at

Table 1: Geometry-driven topological analysis results with objects GUID and surface coverage ratios.

GUID Obj_i^{as-is}	GUID Obj_j^{to-be}	Surface Coverage Ratios
3CFf4AGlbEchvmFeRIHcCG	2FAjB5xJr21RdQcsKQE3TQ	Coverage Obj_1^{as-is}: 13.29% Coverage Obj_1^{to-be}: 100%
3CFf4AGlbEchvmFeRIHcCG	2FAjB5xJr21RdQcsKQE3UZ	Coverage Obj_1^{as-is}: 3.71% Coverage Obj_2^{to-be}: 100%
3CFf4AGlbEchvmFeRIHcCG	2FAjB5xJr21RdQcsKQE3Ut	Coverage Obj_1^{as-is}: 5.91% Coverage Obj_3^{to-be}: 100%
3CFf4AGlbEchvmFeRIHcCG	37NopSpLj37vnjyCBTBUoN	Coverage Obj_1^{as-is}: 40.14% Coverage Obj_4^{to-be}: 94.77%
3320QKjsPDs8e51pe4k$Fe	2ew7MBEPb01OZYaDjw2rmp	Coverage Obj_2^{as-is}: 100% Coverage Obj_5^{to-be}: 99.56%

the level of the renovated envelope assembly. The query returns, for each interface, the identifiers of the connected objects and their corresponding thermal transmittance values, providing a structured and machine-interpretable dataset that can be directly consumed for system-level energy performance evaluation. This demonstrates how geometry-derived semantic relations act as an enabling layer for performance-oriented data extraction without requiring IFC model merging or manual data preparation.
Importantly, the query operates at interface level rather than at individual object level, implicitly shifting the data extraction perspective from single BIM objetcs to physically reconstructed envelope systems. The query is independent of specific IFC schemas or modeling conventions and can be reused across different renovation scenarios, provided that geometry-derived interfaces have been materialized in the linking graph. In this context, a SPARQL query was defined to retrieve, for each detected interface, the identifiers of the connected as-is and to-be objects with their associated thermal transmittance values (Listing 1). The query selectively extracts thermal properties only for object pairs that have been previously identified as physically adjacent through geometry-driven analysis, ensuring that energy-related data are aggregated at the envelope-system level rather than at individual object level.

```
PREFIX rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bot:    <https://w3id.org/bot#>
PREFIX props: <http://lbd.arch.rwth-aachen.de/props#>
PREFIX opm:    <https://w3id.org/opm#>
PREFIX schema: <http://schema.org/>
SELECT DISTINCT ?interface ?guidA ?U_A ?guidB ?U_B
WHERE {
  ?interface a bot:Interface .
  ?interface bot:interfaceOf ?objA, ?objB .
  FILTER(STR(?objA) < STR(?objB))
  ?objA props:globalIdIfcRoot ?gA .
  ?gA opm:hasPropertyState ?sA .
  ?sA schema:value ?guidA .
  ?objB props:globalIdIfcRoot ?gB .
  ?gB opm:hasPropertyState ?sB .
  ?sB schema:value ?guidB .
  ?objA ?pA ?propA .
  FILTER(CONTAINS(LCASE(STR(?pA)), "thermaltransmittance")) .
  ?propA opm:hasPropertyState ?psA .
  ?psA schema:value ?U_A .
  ?objB ?pB ?propB .
  FILTER(CONTAINS(LCASE(STR(?pB)), "thermaltransmittance")) .
  ?propB opm:hasPropertyState ?psB .
  ?psB schema:value ?U_B .
}
ORDER BY ?guidA
```

Listing 1: SPARQL Code for extracting thermal transmittance values

Discussion

The case study demonstrates the effectiveness of the proposed approach in identifying physically meaningful adjacency relations between objects belonging to independent BIM models and in enabling their subsequent semantic integration for performance-oriented analyses. Despite the limited scale of the test scenario, the results confirm the robustness of the geometry-driven topological analysis in accurately detecting real physical contacts while systematically excluding irrelevant adjacencies, such as geometrically coplanar but non-touching surfaces and regions associated with openings.
A key contribution of the proposed method lies in its ability to operate across independent as-is and to-be models without requiring prior alignment, model merging, or manual re-modelling. Through geometry-driven topological analysis and semantic linking, the method explicitly encodes which objects from different models are physically related in the built environment. This capability is particularly relevant in renovation scenarios, where the physical reality of the building envelope is distributed across multiple BIM representations that remain semantically disconnected in conventional workflows.
By creating object-to-object adjacencies as explicit semantic relations, the method enables the reconstruction of coherent envelope systems that exist in reality but are not directly represented in the virtual models. In practice, multiple retrofit components, such as insulation panels, can be correctly interpreted as belonging to the same functional

wall assembly, despite being modeled as separate objects in the to-be model. This transformation allows the machine to reason about which objects jointly contribute to the same physical construction package, overcoming a well-known limitation of geometry-only or model-centric BIM analyses.

From a data exploitation perspective, this semantic reconstruction directly supports the automated extraction and aggregation of performance-related properties. System-level data, such as thermal transmittance (U-values), can be computed by combining properties originating from distinct models but semantically linked through the identified interfaces. As a result, the output of the proposed workflow constitutes a ready-to-use, structured dataset that can be consumed by external energy simulation engines and emerging Digital Twin platforms, without additional manual interpretation or data preparation.

More broadly, the approach contributes to bridging the gap between raw geometric evidence and higher-level semantic representations that are required for advanced analysis and decision-making. This enables comparison of renovation strategies, supporting building-level decision-making. By making physical relationships between independently modeled objects explicit and machine-interpretable, the proposed method enhances interoperability, traceability, and automation in renovation-oriented BIM workflows.

Conclusion

This paper presented a geometry-driven and ontology-based approach to identify and semantically represent adjacency relations between independent BIM models in renovation scenarios, enabling cross-model reasoning and reconstruction of envelope systems without model merging. By combining surface-based topological matching with semantic linking through the BOT ontology, the proposed method enables the explicit representation of physical relations not encoded in BIM data, allowing fragmented objects to be integrated into coherent building systems.

The case study demonstrates the robustness of the approach in detecting adjacencies while excluding irrelevant configurations. It further shows that fragmented retrofit objects can be correctly interpreted as part of coherent envelope systems, reflecting real construction logic rather than the modeling constraints of individual BIM models.

Through the creation of object-to-object adjacencies as machine-interpretable semantic relations, the proposed workflow bridges the gap between raw geometric evidence and structured semantic data. This enables the automated aggregation and extraction of performance properties, supporting system- and building-level performance analysis and enabling comparison of alternative renovation scenarios, while providing a ready-to-use input for energy simulation, and emerging Digital Twin platforms, without requiring manual model alignment or re-authoring. In conclusion, the proposed approach enables a shift from object- to system-level reasoning, which is essential for supporting renovation decision-making processes.

Despite the promising results, some limitations can be identified. The accuracy of the geometry-driven topological analysis depends on the quality of the IFC geometric representations, which may vary across authoring tools and levels of detail. Furthermore, the present validation is limited to a small-scale case study and does not provide a comprehensive evaluation of computational performance on large and highly detailed models. Future research will focus on validating the approach on larger and more complex case studies, such as multi-wall assemblies, multiple openings, and non-orthogonal geometries, as well as on systematically evaluating its computational performance and scalability in real-world renovation scenarios.

References

Gan, V. J. L., Wong, J. K. W., and Tse, K. T. (2022). A graph-based data model for modular building design and analysis. Automation in Construction, 140:104337.

Han, Q., Xie, X., and Lu, W. (2025). Unified network-based representation of bim models for efficient data analysis. Automation in Construction, 165:105321.

Ismail, A., Nahar, A., and Scherer, R. (2017). Application of graph databases and graph theory concepts for advanced analysing of bim models based on ifc standard. In Proceedings of the 25th EG-ICE International Workshop.

Karmakar, A. and Delhi, V. S. K. (2023). Restructuring elements in ifc topology through semantic enrichment for automated compliance checking. In Proceedings of the 40th International Symposium on Automation and Robotics in Construction (ISARC), pages 1–8.

Khalili, A. and Chua, D. K. H. (2015). Ifc-based graph data model for topological queries on building elements. Journal of Computing in Civil Engineering, 29(3):04014046.

Massafra, A., Jabi, W., and Gulli, R. (2024). Topological bim for building performance management. Automation in Construction, 166:105628.

Massafra, A., A.-H. D. H. S. L. . J. W. (2025). Semi-automated dataset generation for residential buildings using graph-based topological modelling. Buildings, 15(8):1283.

Moreira, J., D.-A. P. P. B. E. v. E. T. (2024). Onto4reuse: Towards an ontology reuse framework for knowledge-intensive software engineering. In Proceedings of the Joint Ontology Workshops (JOWO).

Oraskari, J., Bonduel, M., McGlinn, K., Pauwels, P., Priyatna, F., Wagner, A., Kukkonen, V., Steyskaland, S., Lehtonen, J., and Lefrançois, M. (2024). Ifctolbd: Ifctolbd v 2.44.0.

Pauwels, P. and Terkaj, W. (2016). Express to owl for construction industry: Towards a recommendable and usable ifcOWL ontology. Automation in Construction, 63:100–133.

Rasmussen, M. H., Lefrançois, M., Bonduel, M., and Pauwels, P. (2017). The building topology ontology. Semantic Web, 9(2):1–11.

Utkucu, D., Sacks, R., et al. (2025). Ontology for holistic building performance modeling and analysis. Automation in Construction, 175:106197.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

LATENCY-AWARE CONSTRUCTION ACTIVITY RECOGNITION USING KINEMATIC AND PHYSIOLOGICAL SENSORS

Umutcan Kılıç and Semra Çomu
Boğaziçi University, Türkiye

Abstract

Wearable sensors are increasingly being used to monitor worker activity and safety. While Human Activity Recognition (HAR) studies predominantly rely on kinematic data, physiological signals remain mostly underexplored due to the delayed nature of physiological responses. This study investigates the fusion of physiological and kinematic modalities by systematically evaluating temporal offsets to compensate for physiological latency. Results show that classification performance is sensitive to delay magnitude, with optimal performance achieved using an 8-second window and a 4-second time lag. Based on these findings, a latency-aware fusion protocol is proposed to enable robust multimodal HAR in wearable-based worker monitoring systems.

Introduction

The construction industry is characterized by dynamic and hazardous working environments, making occupational safety and health (OSH) a significant concern. To mitigate OSH-related risks, various methods have been proposed to monitor the physical and physiological status of the workforce (Antonaci et al., 2024). Enabling continuous and non-intrusive worker monitoring, Human Activity Recognition (HAR), which refers to automatic identification of human activities and movements using machine learning and deep learning algorithms, has emerged as a prominent approach for data-driven safety assessments and productivity tracking (Sedaghati et al., 2025).

HAR systems can be categorized into two types based on data modalities: visual and non-visual (Sun et al., 2022). Compared to vision-based systems, non-visual HAR solutions are cost-effective, capable of fine-grained recognition, less vulnerable to environmental conditions, and better at protecting worker privacy, making them attractive options for construction settings (Sun et al., 2022).

Non-visual HAR systems are predominantly based on kinematic data derived from Inertial Measurement Units (IMUs), such as accelerometers and gyroscopes (Shin et al., 2025). While effective for classifying physical movements such as walking, bending, or standing, kinematic sensors are inherently limited to describing motion-related patterns. They provide minimal insight into the physiological effects of construction tasks on workers' bodies, such as physical exertion, fatigue, or physiological stress, which are critical metrics for comprehensive safety monitoring (Kazar & Comu, 2021). On the other hand, physiological data, including Heart Rate (HR), Blood Volume Pulse (BVP), Electrodermal Activity (EDA), and Skin Temperature (ST), offer a promising complementary data stream to fill this gap. However, despite its potential to provide a holistic view of worker well-being, physiological data remains mostly unexplored in the context of activity recognition models (Demrozi et al., 2020).

One of the primary barriers to effectively integrating physiological data into HAR models is temporal misalignment. Unlike kinematic signals, which change instantaneously with body movement, physiological responses exhibit biological latency. For instance, while individual variability exists, a cardiovascular response to physical exertion typically occurs a few seconds after the onset of the stressor (Javorka et al., 2003). Similarly, EDA responses typically exhibit a latency of 2 to 6 seconds following a sudden physiological stressor (Yu et al., 2024). Consequently, standard sensor fusion approaches that align data streams strictly based on timestamps result in a mismatch between the physical action and the biological reaction, introducing noise into the dataset rather than enhancing it.

This study aims to address the synchronization challenge by investigating different temporal alignment strategies for multimodal HAR. We utilize machine learning (ML) to test the performance of proposed data fusion strategies and compare different data modalities. In particular, we systematically analyze the impact of different window sizes and the introduction of artificial time-lags to the physiological data stream, aiming to bridge the gap between physical action and physiological reaction. Based on this analysis, the study proposes a latency-aware fusion protocol designed to enable robust multimodal worker monitoring.

Methodology

The experimental study was conducted in a controlled environment at the Structure Laboratory of Boğaziçi University, spanning approximately 660 m² with a ceiling height of 12 meters. The experimental dataset was collected from five healthy male subjects, with no record of an injury or condition that might affect their postures or physiological responses, during independent sessions.

Before each session, the participants were briefed regarding the experiment and provided their informed consents. Table 1 summarizes the demographic details of participants, including age, height, and weight.

Table 1: Participant details

ID	Age (years)	Weight (kg)	Height (cm)
1	26	91	193
2	25	105	187
3	57	98	180
4	58	59	170
5	26	63.5	174

To capture a comprehensive set of kinematic and physiological data, the participants were instrumented with two wearable devices. An Apple iPhone 16 Pro Max smartphone (Apple Inc., Cupertino, CA, USA) was positioned at the lumbar region using an adjustable cloth belt to record gross body movements via its Inertial Measurement Unit (IMU). Because of its proximity to the human body's center of mass, this placement enables effective analysis of overall movement patterns (Wlodzimierz, 2018). The "Physics Toolbox Sensor Suite" application was utilized to log tri-axial gyroscope and accelerometer data (both at 100 Hz), storing the files locally in Comma Separated Values (CSV) format. Simultaneously, an Empatica EmbracePlus smartwatch (Empatica Inc., Boston, MA, USA) was worn on the non-dominant wrist to acquire kinematic data associated with hand movements and physiological signals. Although the smartwatch hardware includes a gyroscope, only tri-axial accelerometer data was recorded (at 64 Hz) due to access restrictions imposed by the "Academic & Basic Research" plan of Empatica. Physiological data included Blood Volume Pulse (BVP) recorded at 64 Hz, Systolic Peaks, EDA recorded at 4 Hz, and ST recorded at 1 Hz. The data collected by the smartwatch was retrieved via the manufacturer's dedicated data access platform as Avro files and subsequently converted to CSV format for further processing. Based on raw Systolic Peaks data, instantaneous HR values were derived by computing inter-beat intervals (IBI) and converting them to beats per minute. HR values outside the physiologically plausible range of 40–200 bpm were discarded as artifacts. The resulting irregularly sampled HR measurements were transformed into a continuous signal using linear interpolation and resampled at 10 Hz to obtain a uniformly spaced time series suitable for temporal analysis and multimodal synchronization. In addition, ST signals, originally recorded at 1 Hz, were upsampled via linear interpolation to a uniform sampling frequency of 4 Hz to prevent data loss at sliding window boundaries.

To capture physiological responses under varying levels of physical exertion, participants followed a physically demanding experimental protocol lasting approximately 30 minutes, designed to simulate continuous construction work. Each session began with a 5-minute seated rest baseline measurement phase (no movement or talking) to capture resting levels of EDA, ST, and cardiovascular signals derived from BVP, including systolic peaks and IBI (Posada-Quintero et al., 2018; Damoun et al., 2024). For analysis purposes, the first and last 30 seconds of the baseline measurement phase were omitted to account for participant settling time. Following the initial omitted period, the subsequent 240 seconds were labeled as the sitting baseline. Any remaining data from the seated baseline measurement was excluded from the analysis to prevent data leakage.

The experiment sessions were structured into four identical activity blocks and were performed across four custom-built stations designed to replicate common construction-related tasks. The first station involved six repetitions of walking on a narrow wooden plank (measuring 2.95 m in length and 0.15 m in width, elevated 25 cm above ground level) to challenge balance and stability. For safety purposes, buffer planks with a width of 20 cm were installed on both sides; however, participants were instructed to walk strictly on the central narrow plank. The second station consisted of two repetitions of accessing a mezzanine platform (measuring 6.3 m by 2.0 m at a height of 2.8 m) via a spiral staircase. Upon reaching the platform, participants walked to its end before descending back to ground level. The third station required participants to climb a standard scaffold (measuring 2.7 m by 0.8 m), featuring two working levels at heights of 0.60 m and 2.55 m. Participants ascended the scaffold using a ladder, reached the upper working level, and descended. The fourth station focused on material handling and involved transferring four separate loads (one 10 kg metal plate, one 5 kg metal plate, and two 1 kg wooden plates) from a platform at 30 cm height onto a pallet truck bed positioned at 10 cm height. Participants then pulled the loaded pallet truck over a distance of approximately 25 meters before returning to the starting position to unload the weights.

Each activity block began with a 60-meter regular walking segment, followed by the execution of all four stations in sequence, and concluded with an additional visit to the first station. After completing the first two blocks, participants took a one-minute standing rest period to emulate a brief work break. After this short recovery phase, the remaining two blocks were performed. The experiments were concluded with a one-minute seated rest period to capture post-activity physiological recovery dynamics. Figure 1 illustrates the content of an experimental session.

To establish precise ground truth for activity labeling, the experimental sessions were recorded on video at 60 fps using an Apple iPhone 16e (Apple Inc., Cupertino, CA, USA) and a GoPro Hero 4 (GoPro Inc., San Mateo, CA, USA). The smartphone captured the third-person view, while the GoPro recorded the first-person view of a given participant. Third-person videos were embedded with real-time timestamps with millisecond precision using the "Timestamp Camera Basic" application. Subsequently,

first- and third-person recordings were synchronized using visual cues via CapCut software.

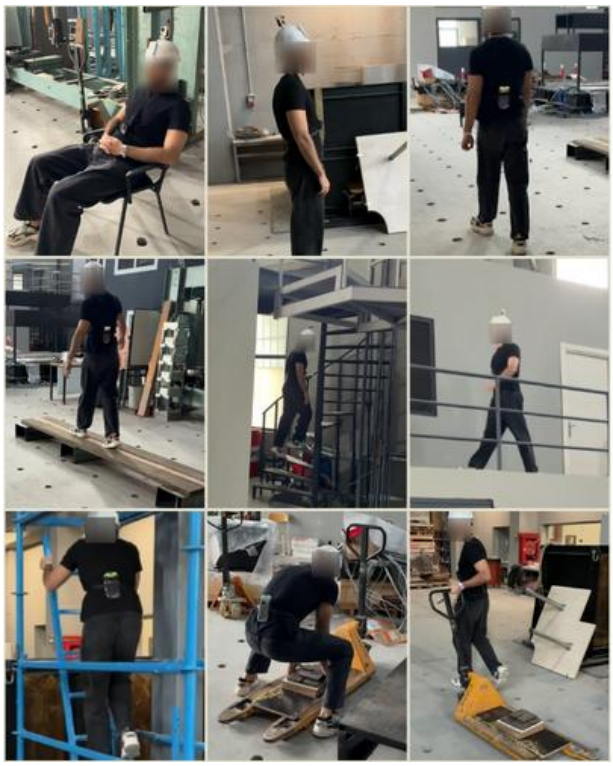


Figure 1: Typical content of an experimental session

Ground truth labeling was performed manually using ELAN software, where activity start and end points were annotated relative to the video timeline (ELAN, 2025). Based on the video footage, various activity types were labeled as shown in Table 2. All annotations were performed by trained annotators following a predefined labeling protocol, which was designed to minimize inter-rater variance by establishing rules that define exact start and end moments of activities. Interruptions during the experiment or other unclassified movements were labeled as junk and excluded from the analysis.

Table 2: Activity Types and Annotation Labels

Activity	Types	Label
Sitting	-	Sitting
Standing	-	Standing
Walking	Normal, Plank, Mezzanine Platform, Scaffolding	Walking
Climbing Stairs	-	N_Stairs
Climbing Ladder	-	V_Stairs
Bending	Hip Hinge, Squat	Bending
Cart	Pulling, Pushing	Cart
Junk	-	Junk

To mitigate temporal discrepancies caused by the Variable Frame Rate (VFR) inherent in smartphone video encoding while converting video-relative timestamps to Unix timestamps for feature extraction purposes, we implemented a piecewise linear interpolation-based conversion algorithm. To serve as anchor points, we extracted the first and the last frames of a given video, as well as a reference frame every 3600 frames (approximately 60 seconds) using FFmpeg (Version 6.0). For each extracted frame, we recorded the exact start time of the frame relative to the video start and the Unix timestamp embedded within the frame, both with millisecond precision. The algorithm was then applied to convert the manually annotated video-relative activity start and end timestamps into sensor-compatible Unix timestamps using (1),

$$T_{unix}(t) = U_i + \frac{(U_{i+1} - U_i)}{(V_{i+1} - V_i)} \cdot (t - V_i) \qquad (1)$$

where t represents the annotated video-relative timestamp, and (V_i, U_i) and (V_{i+1}, U_{i+1}) denote the video-relative and Unix timestamps of the two consecutive anchor frames enclosing t, respectively. This approach effectively compensated for the non-linear time drift associated with VFR, ensuring millisecond-level alignment between the visual ground truth and the multimodal sensor data.

Following the conversion of all ground truth annotations to Unix timestamps, the synchronized multimodal sensor data were segmented using a sliding window approach. Sliding window segmentation is commonly employed in time-series machine learning to compute window-level feature representations, where each window contains a fixed-length segment of data and adjacent windows overlap by a predefined percentage to preserve temporal continuity and increase the number of training samples. Based on the selected window lengths and overlap ratios, sliding windows were generated over the entire duration of each experimental session independently for each participant.

Since experiment sessions followed a continuous flow, a consensus-based labeling strategy was employed to assign activity labels at the window level. Rather than isolating different activities, we generated windows with selected lengths and overlap percentages from the start of a session, marked with the start Unix timestamp of the sitting baseline measurement. For each window, ground truth labels corresponding to the annotated activities were examined, and a window was assigned a class label only if at least 75% of its duration was occupied by a single activity. Windows that did not meet this criterion were considered mixed and excluded from the analysis. This threshold was empirically selected as a trade-off between label purity and data retention, aiming to reduce noise introduced by transitional movements while preserving a sufficient number of valid samples.

For each window, statistical and temporal features were extracted independently from each sensor channel. For BVP, HR, EDA, and ST signals, extracted features included mean, standard deviation, maximum, range,

energy (2), slope (defined as the difference between the last and first sample within each window, normalized by the window duration), and delta (defined as the difference between consecutive windows).

$$E = \frac{1}{N}\sum_{i=1}^{N} x_i^2 \quad (2)$$

Baseline-based z-score normalization was applied to level-dependent physiological features (mean, standard deviation, maximum, and energy) using (3) to reduce inter-subject variability and ensure comparability of the data across participants, whereas change-based features (range, slope, and delta) were computed in the original signal scale to preserve absolute variation within each window.

$$z_i = \frac{(x_i - \mu)}{\sigma} \quad (3)$$

where x_i denotes the physiological feature for the ith window, and μ and σ are the mean and standard deviation of the feature calculated over the baseline measurement period, respectively.

For kinematic data (accelerometer and gyroscope), axis-wise (x, y, z) features (mean, standard deviation, minimum, maximum, and range) and magnitude-based features including signal magnitude area (SMA), vector magnitude–based features (mean and standard deviation), and linear or angular jerk (approximated using first-order differences of acceleration and angular velocity signals, respectively) were extracted.

Table 3 and Table 4 summarize the feature sets generated from the multimodal sensor data.

Table 3: Physiological feature set and applied normalization methods

Sensor	Features	Normalization
BVP	Energy, Mean, Standard Deviation	z-Score
BVP	Range	None
HR	Maximum, Mean	z-Score
HR	Range, Slope	None
EDA	Energy, Maximum, Mean, Standard Deviation	z-Score
EDA	Range, Slope	None
ST	Mean	z-Score
ST	Delta	None

To explicitly investigate the effect of biological response latency on activity classification, we utilized a systematic parameter search strategy. While the sensor data were segmented using multiple window lengths (1, 2, 4, and 8 seconds) combined with overlap ratios of 50% and 75%, we simultaneously introduced temporal offsets ranging from 0 seconds (no delay) to 8 seconds in 1-second increments. This resulted in a total of 72 distinct window configurations. In this protocol, kinematic features were extracted synchronously with the ground truth activity labels to capture immediate physical motion. In contrast, physiological features were extracted from a time window shifted forward by the delay amount, aiming to align the delayed biological response with the causing physical activity. This approach allows for a controlled assessment of how explicit temporal alignment influences classification performance, independent of window size.

To evaluate the contribution of physiological signals to HAR models, we compared the performance of models trained on six different sensor combinations. This approach resulted in 240 distinct model configurations: for kinematic-only models, 3 sensor combinations with 8 window configurations each; and for models involving physiological data, 3 sensor combinations with 72 window configurations each. Based on the extracted feature vectors for each model configuration, activity classification models were trained using a Random Forest (RF) classifier with 100 estimators (trees), which is widely adopted as a strong baseline in machine learning (ML)–based HAR studies (Demrozi et al., 2020). To avoid data leakage and ensure robust, subject-independent evaluation, a Leave-One-Subject-Out (LOSO) cross-validation scheme was employed. In each fold, data from one participant were reserved for testing, while data from all remaining participants were used for training. Model performance was evaluated using standard classification metrics, including accuracy, precision, recall, macro and weighted F1-scores, as well as confusion matrices, enabling both class-wise and overall performance assessment. This procedure was iterated for all 240 model configurations, enabling a comparative assessment to identify the optimal temporal alignment strategy and to see the performance differences between different sensor configurations. Since all five experimental sessions had similar durations, F1-scores were reported as the average performance across all LOSO folds, unless otherwise specified.

Table 4: Kinematic feature set

Sensor Type	Feature Type	Features
Accelerometer	Axis-wise (x, y, z)	Maximum, Mean, Minimum, Range, Standard Deviation
Accelerometer	Magnitude-based	Linear Jerk, Mean Vector Magnitude, Signal Magnitude Area, Standard Deviation of Vector Magnitude
Gyroscope	Axis-wise (x, y, z)	Maximum, Mean, Minimum, Range, Standard Deviation
Gyroscope	Magnitude-based	Angular Jerk, Mean Vector Magnitude, Signal Magnitude Area, Standard Deviation of Vector Magnitude

Results and discussion

First, we analyzed the distribution of ground truth activity labels to characterize the dataset. Based on the video annotations, we calculated the combined total duration of the five experiment sessions to be 9756.833 seconds. As summarized in Table 5, "Walk" constituted the largest portion of the data with 4277.145 seconds (43.84%), followed by "Junk" segments and "N_Stairs".

Following the exclusion of invalid segments (i.e., Junk) and the removal of transitional periods where 75% label consensus was not met, valid windows were retained for model input. Table 6 summarizes the resulting window counts across different activity types for various window lengths at 50% overlap. On the other hand, setting the overlap to 75% resulted in approximately twice the number of analysis windows compared to the values given in Table 6.

The resulting class distribution reveals a natural imbalance often encountered in HAR literature (Lim et al., 2021). For instance, the dataset used by Sikder & Nahid (2021) exhibits an imbalance ratio of 8.41:1 (max:min), which is comparable to the 11.76:1 ratio observed in our dataset. “Walking” emerged as the dominant class, accounting for approximately 51% of the total samples. In contrast, short-duration or less frequent activities such as “Bending” and “V_Stairs” represented approximately 5% and 6% of the samples, respectively. This observed imbalance underscores the necessity of utilizing macro-averaged metrics alongside weighted metrics to ensure a fair assessment of the model's capability to recognize less frequent activities.

To evaluate the contribution of physiological signals to HAR models, we first compared the performance of models trained on different data modalities, fused using a simple fusion strategy (no temporal offsets), with window lengths of 1, 2, 4, and 8 seconds, and 50% and 75% overlap. The average classification performances of the resulting models are summarized in Figure 2.

As anticipated, scenarios relying solely on physiological data yielded the lowest results (Average Macro F1: 16.29%). Notably, a theoretically uniform random classifier would achieve an F1 score of 14.29%, which is only 12.28% worse than the physiological-only model. This marginal difference confirms that physiological signals alone are insufficient for distinguishing complex locomotor activities. In contrast, the model utilizing only the smartphone IMU data from the lumbar region, a standard configuration in HAR literature, achieved an average Macro F1 score of 69.61%, establishing a strong baseline. For instance, the RF-based LOSO model trained by Chen and Shen (2017) using a similar sensor configuration and placement achieved an average Macro F1 score of 67.21% while classifying 5 activities, indicating that our model performed comparably to previously reported results. The highest performance was observed in the configuration that combined kinematic data from both the smartphone and the smartwatch (Average Macro F1: 79.17%). On the other hand, the simple fusion of physiological signals into this optimal kinematic model resulted in a slight performance decline (Average Macro F1: 77.54%). These findings support the claim that the simple fusion of inherently latent physiological data may introduce noise or redundant information, thereby degrading the classifier's predictive capability rather than enhancing it.

Subsequently, we investigated the interaction between model performance and artificial temporal delays. Figure 3 illustrates the fusion model's sensitivity to different temporal delays across various window sizes and overlap ratios. Based on the results, we identified two key trends.

First, larger time windows consistently yielded higher performance. Specifically, the 8-second window achieved the highest Macro F1 scores across nearly all delay settings, with a few exceptions marginally favoring the 4-second window. This superiority is likely attributable to the ability of longer windows to simultaneously capture slow-varying physiological signals and sudden kinematic changes associated with an activity within the same window.

Secondly, the introduction of artificial delays revealed a surprising but expected concave relationship with performance. While simple fusion (0s delay) provided satisfactory results, particularly for longer windows, introducing a temporal shift to the physiological data significantly enhanced Macro F1 scores. For instance, the highest performance observed in this study was obtained at a 4-second delay with an 8-second window and 75% overlap (Macro F1: 81.22% ± 7.64%), highlighting the potential performance gains of integrating physiological sensors under a latency-aware fusion framework. In comparison, the simple fusion counterpart without temporal delay (0-second delay with an 8-second window and 75% overlap) achieved a lower Macro F1 of 80.12% with a higher standard deviation (± 8.14%), indicating that compensating for physiological latency not only improves overall accuracy but also enhances the consistency of the model across different subjects. However, further increasing the delay almost always led to a degradation in classification performance. This finding aligns with established physiological literature, which suggests that physiological responses typically lag behind the onset of a physical stressor by approximately 2–6 seconds (Javorka et al., 2003; Yu et al., 2024).

Lastly, we compared the confusion matrices and class-wise performance metrics for the overall best-performing model (latency-aware sensor fusion model utilizing an 8-second time window, 75% overlap, and a 4-second temporal shift) and its kinematic-only counterpart (Smartphone + Smartwatch). As illustrated in Figure 4 and Figure 5, both models demonstrated strong diagonal dominance. However, the inclusion of temporally aligned physiological data yielded noticeable improvements in specific activities. A closer inspection of the class-wise performance reveals that the most significant gains were observed in the “Cart” and “Walking” activities. These two activities present a critical challenge for kinematic-only models: the body posture similarity between the activities. As a result, kinematic sensors alone fail to

Table 5: Descriptive statistics of annotated activty classes

Activity	Count	Total Duration (s)	Avg. Duration (s)	Min. Duration (s)	Max. Duration (s)	Distribution (%)
Walking	488	4277.145	8.765	0.435	103.649	43.84%
Junk	486	1683.380	3.464	0.068	94.000	17.25%
N_Stairs	33	1078.334	32.677	3.333	117.434	11.05%
Standing	84	751.653	8.948	5.900	11.800	7.70%
Cart	32	666.601	20.831	3.500	31.800	6.83%
Bending	105	585.852	5.580	0.834	23.470	6.00%
V_Stairs	40	390.132	9.753	5.559	14.951	4.00%
Sitting	5	323.736	64.747	54.423	71.011	3.32%

capture the physical effort required to maneuver an approximately 110 kg pallet truck. In contrast, the latency-aware fusion model integrates the physiological markers of this high-exertion task, allowing the classifier to distinguish it from walking, despite the ambiguous kinematic cues. As a result, the number of "Walking" activities falsely classified as "Cart" decreased from 50 to 8, increasing the F1-score for the "Cart" activity from 60.83% to 67.14%. This suggests that while kinematic sensors effectively capture the distinct motion trajectories of the activities, they may struggle to differentiate the subtle intensity variations associated with them. The physiological signals, acting as an indicator of effort, provide the necessary complementary information to resolve these ambiguities.

Table 6: Total number of activity samples for different window lengths at 50% overlap

Window (s)	1	2	4	8
Walking	8065	3790	1629	650
Standing	2121	1046	505	239
N_Stairs	1416	664	294	102
Cart	1301	637	302	135
Bending	1073	480	180	67
V_Stairs	742	351	156	54
Sitting	593	293	142	67
Total Samples	15311	7261	3208	1314

These class-specific improvements highlight the mechanism behind the performance boost. The 4-second temporal offset effectively synchronizes the delayed physiological response with the immediate kinematic changes. Without this alignment (as observed in the simple fusion case discussed earlier), the physiological marker of a demanding activity registers after the activity label has ceased, creating noise. By shifting the windows, we ensure that the classifier receives both the motion pattern of an activity and the corresponding metabolic effect simultaneously. Consequently, the fusion model does not merely rely on how the subject moves, but also how much effort the movement entails, leading to a more robust and context-aware recognition system.

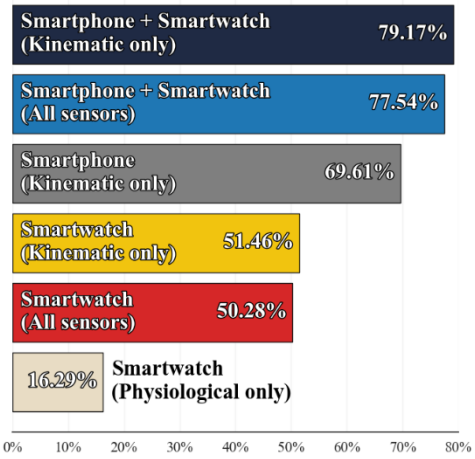


Figure 2: Average Macro F1-scores across various sensor combinations and their simple fusion

As a final step, we measured the average inference time of the best-performing model to evaluate the feasibility of the proposed approach for real-world applications. Using a standard desktop CPU (2.50 GHz, 16 threads), the model classified a single data window in approximately 0.06 milliseconds on average. This low computational cost indicates that the proposed latency-aware fusion model is lightweight enough to be deployed on edge devices or microprocessors available in modern wearable devices with negligible delay, satisfying the critical requirement for real-time safety monitoring in construction environments.

Ultimately, this study demonstrates that bridging the temporal gap between instantaneous motion and delayed physiological response is essential for developing context-aware HAR systems capable of distinguishing

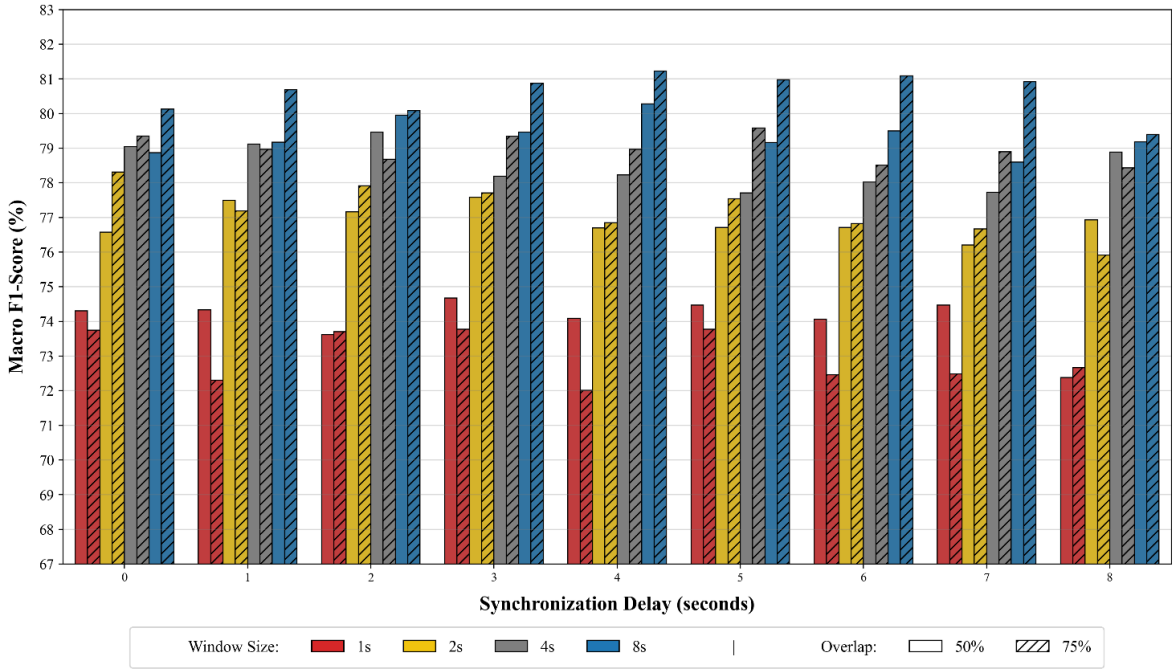
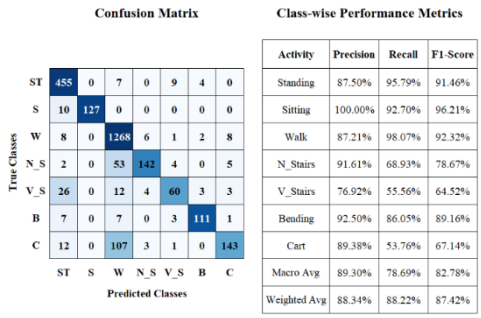


Figure 3: Performance variations of the fusion model as a function of artificial temporal synchronization dealys for different windows sizes and overlap ratios

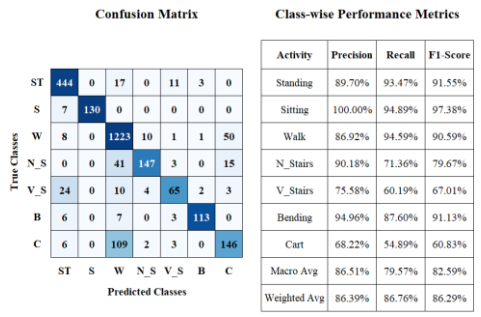
activities defined not just by how we move, but by how much effort is exerted. The proposed latency-aware fusion model achieved a Macro F1 score of 81.22% and a Weighted F1 score of 86.79% (Pooled Macro F1 score of 82.78% and Pooled Weighted F1 score of 87.42%), underscoring its potential to enhance construction site safety through robust worker monitoring.



Class-wise Performance Metrics

Activity	Precision	Recall	F1-Score
Standing	87.50%	95.79%	91.46%
Sitting	100.00%	92.70%	96.21%
Walk	87.21%	98.07%	92.32%
N_Stairs	91.61%	68.93%	78.67%
V_Stairs	76.92%	55.56%	64.52%
Bending	92.50%	86.05%	89.16%
Cart	89.38%	53.76%	67.14%
Macro Avg	89.30%	78.69%	82.78%
Weighted Avg	88.34%	88.22%	87.42%

Figure 4: (a) Confusion matrix and (b) class-wise performance metrics for the latency-aware sensor fusion model using an 8-second time window, 75% overlap, and a 4-second temporal delay.



Class-wise Performance Metrics

Activity	Precision	Recall	F1-Score
Standing	89.70%	93.47%	91.55%
Sitting	100.00%	94.89%	97.38%
Walk	86.92%	94.59%	90.59%
N_Stairs	90.18%	71.36%	79.67%
V_Stairs	75.58%	60.19%	67.01%
Bending	94.96%	87.60%	91.13%
Cart	68.22%	54.89%	60.83%
Macro Avg	86.51%	79.57%	82.59%
Weighted Avg	86.39%	86.76%	86.29%

Figure 5: (a) Confusion matrix and (b) class-wise performance metrics for the kinematic-only model using an 8-second time window and 75% overlap

Conclusion

This study investigated the impact of temporal alignment strategies on the performance of multimodal Human Activity Recognition (HAR) systems for construction worker monitoring, focusing on the fusion of kinematic and physiological data. By systematically evaluating different window sizes and artificial temporal delays, the study aimed to address the synchronization challenges caused by the latency of physiological responses.

In total, we trained and evaluated 240 distinct model configurations to systematically investigate the impact of temporal alignment. The results demonstrated that the temporal alignment of data streams significantly affects the classification performance of multimodal models. The latency-aware fusion model, utilizing an 8-second time window with 75% overlap and a 4-second temporal delay, achieved the highest performance with a Pooled Macro F1-score of 82.78% and a Pooled Weighted F1-score of 87.42%, significantly outperforming both the kinematic-only counterpart and the simple fusion approach. These improvements are largely attributed to the effective compensation of physiological latency, which ensures that the delayed metabolic markers of physical exertion are correctly aligned with the corresponding motion data.

Furthermore, the analysis revealed that the inclusion of temporally aligned physiological data is particularly critical for distinguishing activities that share similar kinematic patterns but differ in physical intensity. For instance, the latency-aware model successfully differentiated between "Walking" and high-exertion "Cart" pulling tasks, reducing misclassification rates where kinematic-only models failed due to postural similarities. In contrast, the simple fusion of physiological signals without temporal correction resulted in a slight

performance degradation compared to kinematic-only models, suggesting that misaligned physiological data acts as noise rather than an informative feature.

While these findings highlight the strong potential of latency-aware multimodal sensor fusion for robust worker monitoring, certain limitations should be acknowledged. First, the experiments were conducted in a controlled laboratory environment with a limited sample size of five healthy male participants (N=5). Although sufficient for a proof-of-concept temporal alignment analysis using a Leave-One-Subject-Out validation scheme, this small sample size inherently restricts the statistical power required for formal significance testing (e.g., paired t-tests) between alternative model configurations. For the generalizability of the results and robust validation of the optimal physiological temporal delay, future research should utilize a more diverse population varying significantly in age, baseline fitness levels, body mass index (BMI) and construction experience. Secondly, although the activities were designed to simulate realistic construction tasks, the controlled nature of the study does not fully replicate the complex, unstructured, and dynamic conditions of an active construction site, where factors such as environmental conditions and worker interactions may influence kinematic and physiological sensor readings.

By addressing these challenges, multimodal HAR systems can be further refined to provide a more reliable and comprehensive solution for occupational safety and health management in the construction industry.

References

Antonaci, F. G., Olivetti, E. C., Marcolin, F., Castiblanco Jimenez, I. A., Eynard, B., Vezzetti, E., & Moos, S. (2024). Workplace well-being in industry 5.0: A worker-centered systematic review. Sensors, 24(17), 5473. https://doi.org/10.3390/s24175473

Chen, Y., & Shen, C. (2017). Performance analysis of smartphone-sensor behavior for human activity recognition. IEEE Access, 5, 3095–3110. https://doi.org/10.1109/access.2017.2676168

Damoun, N., Amekran, Y., Taiek, N., & El Hangouche, A. J. (2024). Heart rate variability measurement and influencing factors: Towards the standardization of methodology. Global Cardiology Science and Practice, 2024(4). https://doi.org/10.21542/gcsp.2024.35

Demrozi, F., Pravadelli, G., Bihorac, A., & Rashidi, P. (2020). Human activity recognition using inertial, physiological and environmental sensors: A comprehensive survey. IEEE Access, 8, 210816–210836. https://doi.org/10.1109/access.2020.3037715

Javorka, M., Žila, I., Balhárek, T., & Javorka, K. (2003). On- and off-responses of heart rate to exercise – relations to heart rate variability. Clinical Physiology and Functional Imaging, 23(1), 1–8. https://doi.org/10.1046/j.1475-097x.2003.00460.x

Kazar, G., & Comu, S. (2021). Exploring the relations between the physiological factors and the likelihood of accidents on construction sites. Engineering, Construction and Architectural Management, 29(1), 456–475. https://doi.org/10.1108/ecam-11-2020-0958

Lim, X. Y., Gan, K. B., & Abd Aziz, N. A. (2021). Deep CONVLSTM network with dataset resampling for Upper Body Activity Recognition using minimal number of IMU sensors. Applied Sciences, 11(8), 3543. https://doi.org/10.3390/app11083543

ELAN (Version 7.0) [Computer software]. (2025). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Posada-Quintero, H. F., Reljin, N., Mills, C., Mills, I., Florian, J. P., VanHeest, J. L., & Chon, K. H. (2018). Time-varying analysis of electrodermal activity during exercise. PLOS ONE, 13(6). https://doi.org/10.1371/journal.pone.0198328

Sedaghati, N., Ardebili, S., & Ghaffari, A. (2025). Application of human activity/action recognition: A Review. Multimedia Tools and Applications, 84(28), 33475–33504. https://doi.org/10.1007/s11042-024-20576-2

Shin, J., Hassan, N., Miah, A. S., & Nishimura, S. (2025). A comprehensive methodological survey of human activity recognition across diverse data modalities. Sensors, 25(13), 4028. https://doi.org/10.3390/s25134028

Sikder, N., & Nahid, A.-A. (2021). Ku-Har: An open dataset for heterogeneous human activity recognition. Pattern Recognition Letters, 146, 46–54. https://doi.org/10.1016/j.patrec.2021.02.024

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., & Liu, J. (2022). Human action recognition from various data modalities: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–20. https://doi.org/10.1109/tpami.2022.3183112

Wlodzimierz, E. S. (2018). Center of mass of the human body helps in analysis of balance and movement. MOJ Applied Bionics and Biomechanics, 2(2). https://doi.org/10.15406/mojabb.2018.02.00057

Yu, X., Lu, J., Liu, W., Cheng, Z., & Xiao, G. (2024). Exploring physiological stress response evoked by passive translational acceleration in healthy adults: A pilot study utilizing electrodermal activity and heart rate variability measurements. Scientific Reports, 14(1). https://doi.org/10.1038/s41598-024-61656-5

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

INTEGRATING TIME, COST, AND RESOURCES IN BIM: A PROCESS-ORIENTED FRAMEWORK

Davide Avogaro[1] and Carlo Zanchetta[2]
[1]Department of Management and Engineering, University of Padova, 36100 Vicenza, Italy
[2]Department of Civil, Environmental and Architectural Engineering, University of Padova, 35131 Padova, Italy

Abstract

Current practices in scheduling, cost estimation, and resource accounting rely on fragmented workflows, unclear use of Industry Foundation Classes (IFC), and proprietary tools that hinder interoperability. These limitations lead to error-prone, non-standardized processes. This study proposes a framework based on IFC4.3 that defines construction processes and process templates to integrate time, cost, and resource data within IFC. A web-based prototype demonstrates the approach, enabling automatic generation of WBS, CBS, Bills of Quantities, and 4D simulations. Results confirm the feasibility and completeness of IFC for these domains, providing a foundation for interoperable, standardized project-control tools and future real-case implementations.

Introduction

Effective project control is essential for achieving objectives related to time, cost, quality, and safety. This discipline relies on systematic processes for planning, measuring, monitoring, and managing project performance. The present study focuses specifically on time scheduling, cost estimation, and resource accounting. Traditional methods for managing these domains are often labor-intensive, error-prone, and inefficient (Sheik et al., 2023). In particular, time scheduling frequently depends on the subjective experience of planners rather than on objective and verifiable data (Ciribini et al., 2015). Current practices generally do not support an integrated approach to cost, time, and resource management. Consequently, calculations are duplicated, multiple software tools—often incompatible unless part of the same proprietary ecosystem—must be used, and the same information is stored across separate databases, increasing the risk of data inconsistency (Al-Sinan et al., 2024). When such procedures are not aligned with a common standard such as IFC (International Organization for Standardization, 2024), collaboration becomes fragmented, data exchange becomes unreliable, and stakeholders cannot rely on a shared, consistent representation of project information (Lu et al., 2022).

This paper proposes a process-oriented, integrated framework for managing costs, tasks, and resources based on the IFC4.3 standard (International Organization for Standardization, 2024) through the conceptualization of process templates. The primary contribution of this work is a fully IFC-compliant framework that demonstrates how IFC classes can work together to support the use case of integrated management of time scheduling, cost estimation, and resource accounting. The framework provides a standardized, repeatable, and open-source, data-driven approach. In the ongoing discussion and formalization of the IFC5 standard, the development and dissemination of use cases and frameworks are particularly important. The emerging IFC5 version aims to preserve the semantic scope of IFC4.3 (van Berlo et al., 2021); however, certain branches, such as *IfcProcess* and *IfcControl*, may require reconsideration. Contributions such as the present work can help identify strategical or critical aspects, potential gaps, and redundancies within the standard, thereby supporting informed feedback for future revisions.

Additionally, a prototype web application is presented to demonstrate the practical feasibility and potential of this framework in real-world applications.

Background

Traditional practices rely on manually measuring elements from CAD drawings and calculating task costs, resource usage, or task progress using spreadsheets. Even when Building Information Modeling (BIM) workflows are adopted, practitioners frequently depend on proprietary software or closed ecosystems that limit interoperability among stakeholders. In many cases, BIM is used primarily for quantity extraction, while subsequent data processing continues to be performed in spreadsheets. In other cases, organizations introduce user-defined parameters to encode 4D (planning-related BIM data) and 5D (cost-related BIM data) information, enabling integration with external software tools or spreadsheets through export tables or direct file import. However, these approaches lack standardization and consequently hinder interoperability. Nevertheless,

proprietary software continues to dominate the market, likely due to its current robustness and maturity.

Among approaches based on the IFC standard, a systematic literature review (Avogaro & Zanchetta, 2025) demonstrates that existing research provides limited guidance on managing these domains individually, and even less on their integrated handling. The same study also highlights the lack of tools capable of supporting comprehensive, IFC-based management of time, cost, and resources. While adopting IFC is a crucial step toward achieving interoperability in data management, it is not sufficient to enable effective collaboration on its own. Professionals also require practical tools for handling IFC data and supporting collaborative workflows. In this context, web-based platforms represent an effective solution (Park et al., 2017), as they enable remote collaboration, facilitate the reuse of shared information, and provide intuitive interfaces that do not require specialized programming skills.

Relevant contributions in the literature include (Cassandro et al., 2025), who propose a resource-centric approach that integrates building geometry with scheduling and cost data using ontologies as a semantic foundation. In addition, (Sheik et al., 2023) developed an automated progress monitoring framework based on IFC, leveraging classes such as *IfcTask*, *IfcCostItem*, and *IfcConstructionResource*, and implemented a supporting web-based application. Moreover, Blender, when used with the Bonsai extension (IfcOpenShell, 2025), provides dedicated functionalities for time scheduling, cost estimation, and resource accounting based on IFC classes.

Methodology

A previous study (Avogaro et al., 2024) demonstrated that the IFC standard provides a high level of product definition, offering a substantial number of properties that cover most construction domains. This enables the definition of *IfcElementType* entities containing the shared information required to describe their corresponding *IfcElement* instances. In practical terms, this allows the creation of reusable product templates that can be applied across multiple projects, thereby reducing professional effort and enhancing interoperability and collaboration. The same logic can be extended to the construction processes associated with specific element types. In most cases, all instances of a given product type follow a similar sequence of construction processes. This observation led to the conceptualization of *process templates*, defined as the digital representation of all processes required to construct a given product. Two categories of processes are identified: (1) *construction processes*, which consume resources to produce a building element, and (2) *other processes*, such as transport or curing, which do not directly contribute to the physical creation of the product.

Construction processes typically consume four types of resources: human resources, materials, products, and equipment. Each resource is associated with a cost, and the aggregation of these costs determines the total cost of the process. However, only human resources and equipment contribute to time consumption, thereby defining the duration of each process. Within a process template, multiple processes may occur sequentially or in parallel, and the total duration of an element is determined by the time elapsed between the start of the first process and the completion of the last.

Processes are generally associated with different phases of a project. For example, within a process template for a concrete wall, concrete casting belongs to the structural phase, whereas painting belongs to the finishing phase. To ensure effective project scheduling, each task must be correctly assigned to its corresponding phase. This hierarchical organization defines a Work Breakdown Structure (WBS) for the project. Similarly, a Cost Breakdown Structure (CBS) aligned with the WBS is required to enable consistent cost management across different phases and processes.

Following the conceptualization of the framework based on project management principles, a detailed analysis of the IFC4.3 standard (buildingSMART Technical, 2024) was conducted to assess its capability to represent these data structures and relationships. The results of this analysis are presented in the following section.

Subsequently, a prototype web application was developed to test the proposed framework and to provide an open-source environment for managing and inputting the associated data. The implementation details are described in the Implementation section.

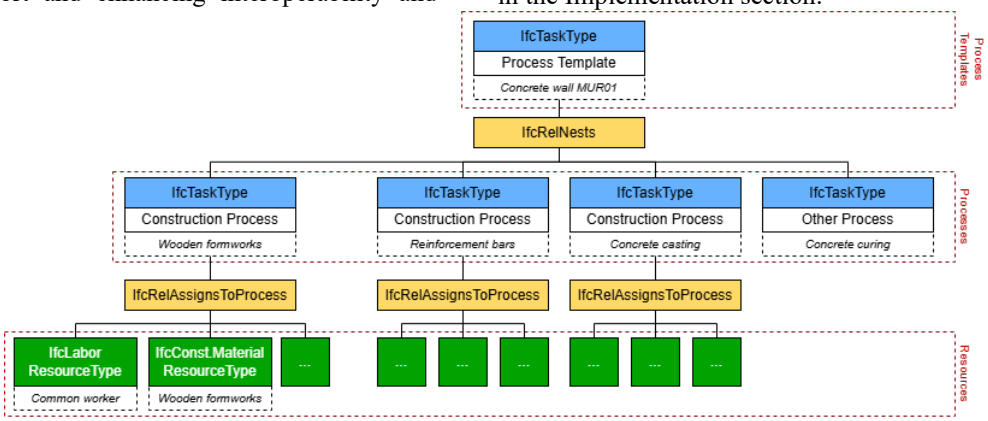


Figure 1: Process Template structure

Analysis

Within the IFC data model, a process template can be represented by an *IfcTaskType* entity that aggregates other *IfcTaskType* entities corresponding to individual processes, as illustrated in Figure 1. Each construction process consumes different types of resources, which in IFC are modeled as subclasses of *IfcConstructionResourceType*. These include *IfcLaborResourceType*, *IfcConstructionMaterialResourceType*, *IfcConstructionProductResourceType*, and *IfcConstructionEquipmentResourceType*, representing human resources, materials, products, and equipment, respectively. In this structure, a construction process represented by an *IfcTaskType* is associated with one or more *IfcConstructionResourceType* entities, as shown in Figure 1.

To support the standardized definition of construction processes, price lists can be used as databases that provide detailed descriptions of construction activities. These lists typically include the unit cost of each activity, the required quantities of resources, and the associated costs for each resource, as illustrated in Figure 2. Additionally, information on the time consumption of labor and equipment resources provides a robust basis for calculating man-hours or machine-hours, enabling accurate estimation of the time required to complete a unit of work.

Construction Process					
Code	**Description**			**Meas. unit**	**Price**
VEN25-01.08.02b	PRE-MIXED DURABLE CONCRETE XC1 workability class S3 (semi-fluid), exposure class XC1, C 25/30			m^3	129,21
Resources					
Code	**Description**	**Qta**	**Meas. unit**	**Unit Price**	**Price**
VEN25-RU.01.04.a	GENERAL CONSTRUCTION WORKER	0,150	h	27,33 €	4,10 €
VEN25-PR-B.02.03.b	PRE-MIXED DURABLE CONCRETE XC1 workability class S3 (semi-fluid), exposure class XC1, C 25/30	1,050	m^3	114,02 €	119,72 €
VEN25-AT.14.05.00	HOT RENTAL OF ELECTRIC CONCRETE VIBRATOR	0,150	h	35,96 €	5,39 €

Figure 2: Example of price voice data

By leveraging *IfcConstructionResourceType* entities, all relevant price list data can be encoded within the IFC model. In subsequent stages, these data are combined with the quantities associated with each *IfcElement* to derive overall costs and durations. Specifically, the *BaseCosts* attribute of an *IfcConstructionResourceType* is used to represent the unit cost of a resource, while the *BaseQuantity* attribute defines an *IfcPhysicalQuantity* that expresses the resource consumption per unit of product. The *Name* attribute of this quantity specifies the relevant measured parameter (e.g., *GrossVolume*) or a calculation formula, as illustrated in Figure 3.

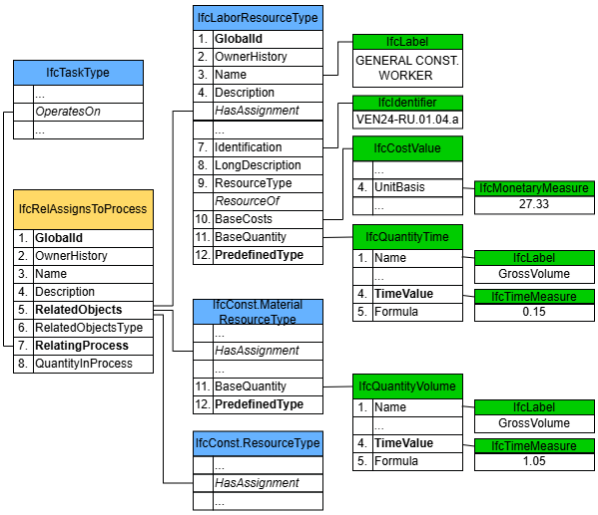


Figure 3: Use of the IfcConstructionResourceType subclasses

The total unit cost of a product corresponds to the sum of all associated resource costs. In this regard, a process template is considered as a structured arrangement of *IfcTaskType* and *IfcConstructionResourceType* entities that enables the standardized representation of price-list information within the IFC schema.

In the preceding paragraph, the procedure for encoding price list data into the IFC data model using process templates was described. The following section explains how to relate and instantiate the classes of a process template with specific *IfcElement* instances. The objective is to associate one or more *IfcBuildingElement* instances with an instantiated process template. As illustrated in Figure 4, the hierarchical structure of the process template represented through *IfcTaskType* entities is replicated using *IfcTask* entities, which now contain the actual duration of each process. The sum of these durations yields the total duration of the process template. A parallel structure is created for *IfcCostItem* entities, where the *IfcCostItem* corresponding to the process template aggregates the values of the individual process-level *IfcCostItem* instances.

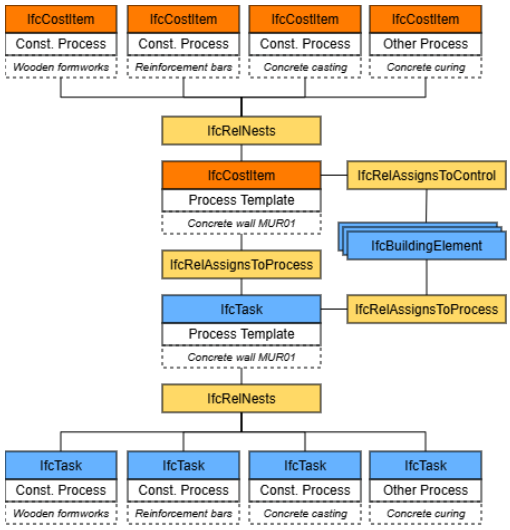


Figure 4: Symmetry in the hierarchical structure of IfcTask and IfcCostItem entities

Once the quantities of the various *IfcBuildingElement* instances associated with a process template are known, they are aggregated—through a scripted procedure—to enable the calculation of the corresponding resource quantities. For each *IfcConstructionResourceType* defined within the process template, the *BaseQuantity* attribute is retrieved, allowing identification of both the target parameter specified in the *Name* attribute of the *IfcPhysicalQuantity* (e.g., *GrossVolume*) and the quantity of resource consumed per unit of product. The target parameter is then extracted from the quantities of the building elements and used in the calculation by multiplying it by the consumption rate per unit of product. The resulting value is assigned as the *BaseQuantity* of the instantiated *IfcConstructionResource*. Based on this quantity, an *IfcCostItem* is created for the resource, and its *IfcCostValue* is computed by multiplying the resource quantity by the unit cost defined in the corresponding *IfcConstructionResourceType* (*BaseCosts* attribute). This procedure is repeated for each *IfcConstructionResourceType* associated with the process template. All resource-level *IfcCostItem* entities are then aggregated under a parent *IfcCostItem*, which sums their values and is associated with the corresponding *IfcTask* representing the construction process. This logic is consistent with the ISO 9000 definition of processes and with the ICOM process diagram. The resulting structure is presented in Fig. 5.

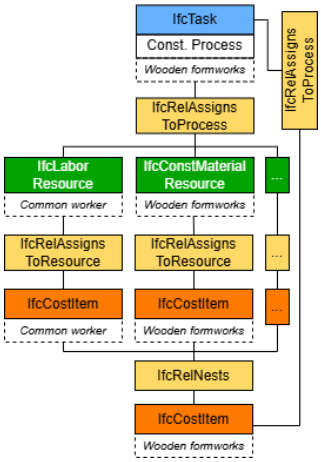


Figure 5: Relationships between IfcConstructionResource, IfcTask, and IfcCostItem entities

The duration of each *IfcTask* corresponding to a construction process is determined by the duration of the resources that consume time. As shown in Figure 2, resources such as labor and equipment require measurable time inputs. Using this information, it is possible to calculate the number of man-hours or machine-hours required to perform each construction process. However, these values represent workloads, not durations. For example, eight man-hours performed by two workers correspond to a four-hour task duration. The *Usage* attribute of *IfcConstructionResource* enables the use of the *IfcResourceTime* class, which includes the *ScheduleUsage* attribute that specifies the number of resources working concurrently. The ratio between *ScheduleUsage* and the required work quantity yields the resource-based duration. For each construction process, a script identifies the longest resource duration, which becomes the duration of the corresponding *IfcTask*.

Other processes—such as maturation or transportation—that do not rely on resource consumption may also have predefined durations and costs, which contribute to the overall duration and total cost of the process template.

Each process within a process template may include a defined starting date, from which its finishing date can be derived by adding the corresponding duration. Tasks may also incorporate sequencing relationships, constraints, and other scheduling logic. All temporal information is stored in the *IfcTask* entities and their associated *IfcTaskTime* objects. These data can be used to generate Gantt charts and construct a WBS. Furthermore, the hierarchical organization of *IfcCostItem* entities enables the creation of a CBS that mirrors the WBS, supporting procedures for monitoring actual costs over time and improving the overall quality of project control. Finally, for each task in the WBS, it is possible to extract a detailed list of resources and thereby generate task-specific Bills of Quantities.

Implementation

To evaluate the proposed framework, the authors developed a prototype application. The system is implemented as a Single-Page Application (SPA) using the React framework. The web application relies on libraries from That Open Company (*That Open Company*,

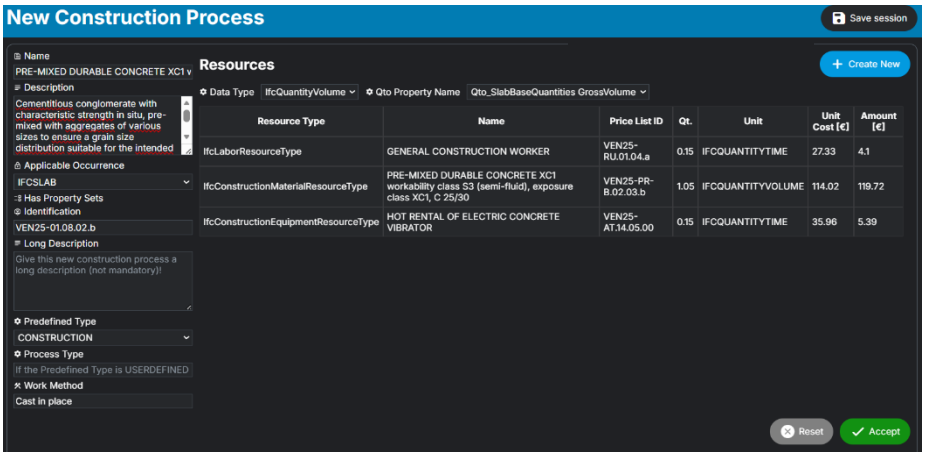


Figure 6: New Construction Process page

2025) to visualize IFC models and to read, write, and manage IFC objects.

The first interface page enables users to create new processes, as shown in Figure 6. The user interface allows the input of all attributes of an *IfcTaskType* entity. Among the most relevant attributes, *ApplicableOccurrence* specifies the IFC class to which the process applies (e.g., *IfcWall*), while *PredefinedType* allows the application to distinguish between construction processes (e.g., CONSTRUCTION, INSTALLATION, etc…) and other types of processes (e.g., MOVE, ATTENDANCE, etc…). For construction processes, the user can define the physical quantity to be considered (e.g., *IfcQuantityVolume*) and the corresponding property (e.g., *GrossVolume*), including formulas where necessary (e.g., "Qto_WallBaseQuantities GrossVolume − Qto_WallBaseQuantities GrossSideArea × 0.1"). The application provides selectable options for physical quantities and properties based on the IFC class specified in *ApplicableOccurrence*. These data are stored in the *BaseQuantity* attribute of the associated *IfcConstructionResourceType* entities. Within the same interface, the user can also define the list of *IfcConstructionResourceType* entities and their attributes. Upon saving, all information is written to an IFC-SPF file stored on a server, which functions as a shared database to support data reuse across users and to minimize redundant input.

The second interface page mirrors the structure of the first but is dedicated to creating new Process Templates, as shown in Figure 7. As before, the user completes a set of attributes for an *IfcTaskType* entity; however, *PredefinedType* is pre-filled as USERDEFINED, and the *ProcessType* attribute is set to PROCESSTEMPLATE. The user can assemble the set of processes that compose the template by selecting from those previously stored in the IFC-SPF database. Once completed, the template is saved to the shared database. An additional interface page presents multiple synchronized panels. The first panel, shown in Figure 8, provides a 3D model viewer alongside a list of available process templates. After selecting one or more model elements, the user can associate a process template with them.

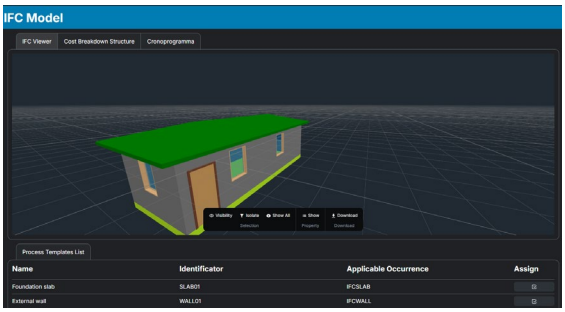


Figure 8: IFC viewer integrated with the process template assignment interface

For each construction process, the application opens a dialog prompting the user to specify the number of resources that contribute to time consumption, enabling accurate calculation of process duration, as shown in Figure 9. The dialog also requires the assignment of the process to a Work Breakdown Structure (WBS) work package, ensuring alignment with the Cost Breakdown Structure (CBS). For non-construction processes, the application instead requires manual input of duration and cost.

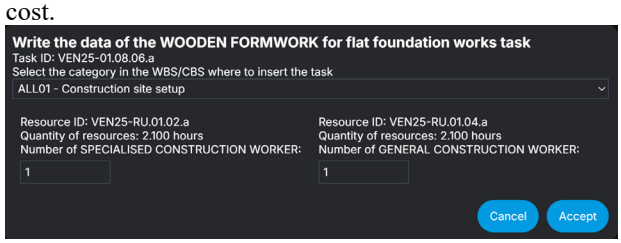


Figure 9: Resources consumption dialog

Results

The prototype enables automated cost estimation, time scheduling, and resource accounting based entirely on data. A second panel presents the Cost Breakdown Structure (CBS), mirroring the hierarchy of the Work Breakdown Structure (WBS) and reporting the cost associated with each process, as shown in Figure 10. A further panel provides an overview of the resources assigned to each process.

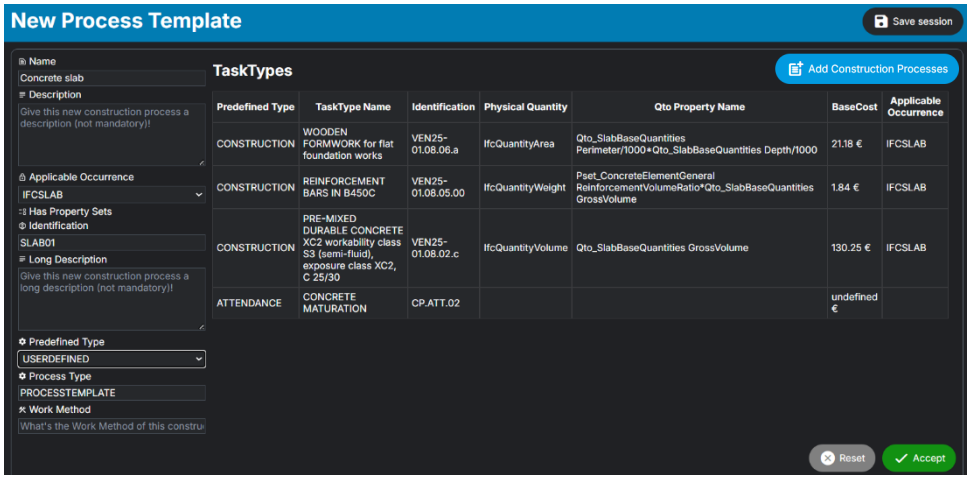


Figure 7: New Process Template page

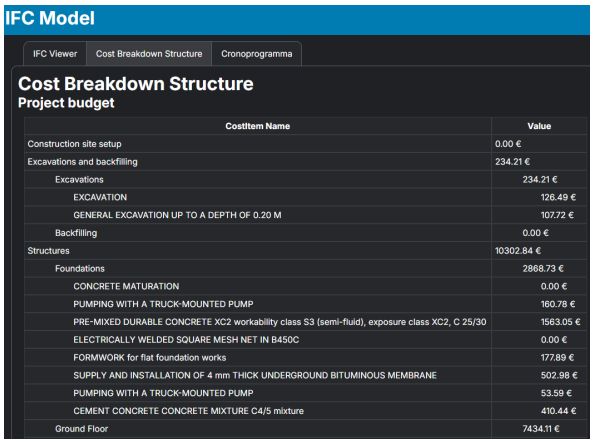


Figure 10: Cost Breakdown Structure page

The third panel displays the WBS, aligned with the CBS, and visualized through a Gantt chart representation, as illustrated in Figure 11. Within this panel, users can refine task durations, assign constraints, reorder tasks, and adjust dependencies through intuitive drag-and-drop interactions. Although the automatically calculated durations provide a reliable baseline, professional judgment is often required to adapt them to project-specific conditions; therefore, manual editing is supported. All modifications are not only reflected in the user interface but are also written back into the IFC model.

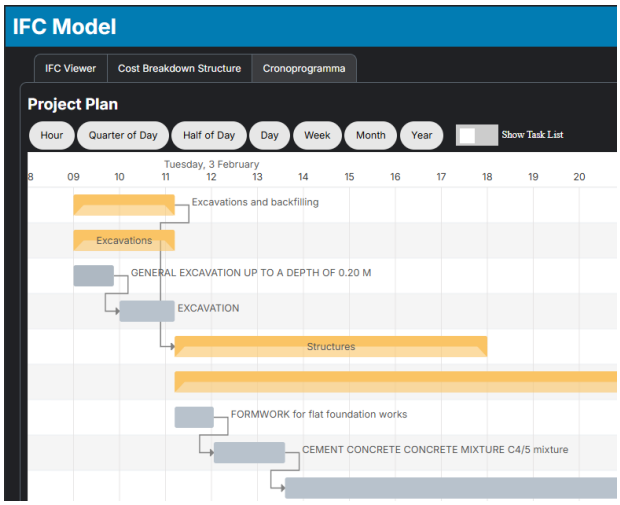


Figure 11: Work Breakdown Structure page trough a Gantt representation of the IfcTask

An additional interface page is dedicated to 4D simulation, allowing users to visualize the temporal progression of the project through animated model playback, as shown in Figure 12.

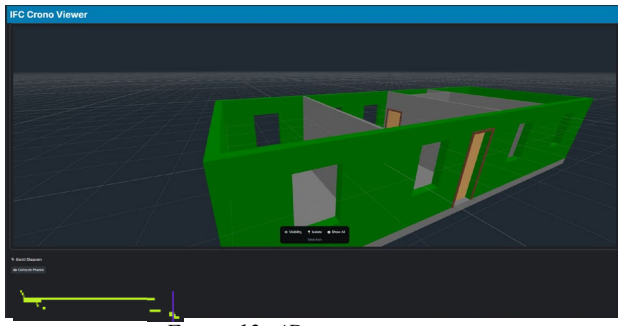


Figure 12: 4D animation page

All IFC files generated by the prototype—both those used as a shared database and those produced as project models—were successfully interpreted using external tools such as KITModelViewer (Institute for Automation and Applied Informatics (IAI), 2025).

Discussion

The testing phase conducted so far has focused on relatively simple building elements, including cast-in-place walls, slabs, windows, and doors. For cost data, the 2025 Veneto Region Price List (Veneto Region, 2025) was adopted, as it was considered appropriate for the scope of the initial experiments. Given that price lists vary significantly across regions and countries, future research will investigate the application of the methodology to real-world case studies, additional IFC classes, and alternative price lists. Although the workflow has been designed to be flexible and adaptable, further validation is required.

The developed prototype also highlights the relevance of web-based applications, which eliminate the need for specialized software installation and provide readily accessible tools for practitioners. The use of open-source libraries, such as those provided by That Open Company, enables professionals with programming expertise to develop customized solutions, helping to address the current lack of IFC-based software tools in these domains. Moreover, as these tools rely on open standards rather than proprietary ecosystems, they support the adoption of IFC and enhance interoperability and standardization across the sector.

The study emphasizes the importance of data reuse and sharing to reduce redundant work and improve interoperability, both through the conceptualization of process templates and the architecture of the prototype. However, the current implementation presents limitations in this regard. The IFC-SPF file used as a database does not provide the robustness of a dedicated database system: concurrent access by multiple users may result in overwriting conflicts, and essential functionalities—such as data consistency checks, input validation, and error handling—are not inherently supported. Future work will focus on developing a dedicated price list database based on Semantic Web technologies using the IfcOWL ontology (buildingSMART Technical, 2025). This approach would ensure consistency with the IFC data model while leveraging Semantic Web capabilities to

enable real-time collaboration among stakeholders (Yang et al., 2021).

To further reduce manual effort, future developments will also explore automated procedures for converting price lists into IFC-based databases, thereby minimizing the workload required from professionals. At the same time, the prototype requires further refinement to improve the user experience and to better address the complexity and variability of real-world projects. For example, it is necessary to support exceptions and variations by allowing users to define additional processes for specific elements that deviate from the standard process template. Once the prototype reaches a more stable stage, it will be possible to evaluate its effectiveness through quantitative metrics such as time savings, productivity gains, and error rates.

Furthermore, the study demonstrates the central role of an effective user interface. The prototype shows that even a complex structure of IFC classes can be navigated and managed by users without prior knowledge of the standard. This is particularly relevant for professionals without programming skills and for public administrations that require BIM data to be delivered in open formats.

Despite the identified technical and functional limitations, the prototype successfully validates the proposed methodology and demonstrates the feasibility of developing applications entirely based on IFC. More importantly, it highlights the potential of the IFC standard to support integrated cost, scheduling, and resource management, and underscores the need for consistent and shared research on standardized methods for its application.

Conclusions

The literature indicates that BIM methodologies can overcome traditional practices for time scheduling, cost estimation, and resource accounting, which are often time-consuming, error-prone, and based on subjective experience rather than data. However, current BIM practices in these domains still rely heavily on proprietary ecosystems, manual procedures, or user-defined approaches that hinder interoperability, lack standardization, and fragment data across multiple environments. The adoption of a common standard such as IFC, which encompasses these domains, offers a promising solution. Nevertheless, previous research (Avogaro & Zanchetta, 2025) highlights key barriers to the widespread use of IFC, including uncertainty regarding the correct use of IFC classes and the appropriate structuring of project-control data within the schema, as well as the lack of software tools capable of effectively managing such data through the IFC standard.

This study presents a process-oriented framework for the integrated management of time scheduling, cost estimation, and resource accounting based on the IFC4.3 standard. By introducing the concept of process templates and systematically mapping them to IFC entities, the research demonstrates that IFC can support a coherent, data-driven representation of construction processes, including their associated resources, costs, and temporal attributes. The proposed methodology enables the standardization and reuse of construction knowledge through template-based structures, reducing redundancy and improving consistency across projects. The implementation of a web-based prototype further validates the practical feasibility of the approach, showing that key project-control outputs—such as WBS, CBS, Bills of Quantities, and 4D simulations—can be generated automatically from IFC-compliant data. These results confirm that IFC can function not only as a data exchange format but also as a central data model for integrated project management workflows.

At the same time, the study identifies several challenges that must be addressed to enable real-world adoption. These include the need for more robust data management infrastructures, improved handling of concurrent data access, and enhanced support for variability and exceptions in construction processes. Furthermore, additional validation through more complex case studies and diverse price lists is required to assess scalability and generalizability.

Overall, this work lays the foundation for interoperable, standardized, and fully data-driven project-control tools, supporting a transition from fragmented and proprietary workflows toward more collaborative and transparent practices in the construction industry. It provides both a methodological foundation and a proof of concept demonstrating that integrated, standardized, and interoperable project-control workflows based entirely on IFC are achievable and beneficial, and it outlines a pathway for future research and tool development aimed at reducing inefficiencies, enhancing collaboration, and promoting the adoption of open standards in the sector.

References

Al-Sinan, M. A., Bubshait, A. A., & Aljaroudi, Z. (2024). Generation of construction scheduling through machine learning and BIM: A blueprint. Buildings, 14(4), 934.

Avogaro, D., Berlato, M., & Zanchetta, C. (2024). Process Scheduling Application Based on Web Ontology Language Representation of the IFC Schema. 125–135.

Avogaro, D., & Zanchetta, C. (2025). IFC and Project Control: A Systematic Literature Review. Buildings, 16(1), 91. https://doi.org/10.3390/buildings16010091

buildingSMART Technical. (2024, aprile). IFC 4.3 ADD2. IFC 4.3 ADD2 - 4.3.2.0 [Official]. https://standards.buildingsmart.org/IFC/RELEASE/IFC4_3/

buildingSMART Technical. (2025, luglio 1). ifcOWL. ifcOWL. https://technical.buildingsmart.org/standards/ifc/ifc-formats/ifcowl/

Cassandro, J., Hagedorn, P., Sigalov, K., Gatto, C., Mirarchi, C., Pavan, A., & König, M. (2025). Semantic

integration of cost, time and geometry for a resource-centred perspective of managing construction projects. 6, 0–0.

Ciribini, A. L. C., Ventura, S. M., & Paneroni, M. (2015). Implementation of an open and interoperable process to optimise design and construction phases of a residential building project: A case study using BIM in a public procurement. 32, 1.

IfcOpenShell. (2025). Bonsai. https://bonsaibim.org/index.html

Institute for Automation and Applied Informatics (IAI). (2025). KITModelViewer (Versione 7.4) [Software]. https://www.iai.kit.edu/english/1302.php

International Organization for Standardization. (2024). ISO 16739-1. https://www.iso.org/standard/84123.html

Lu, C., Wang, X., Nie, L., Bao, W., Huang, X., Yang, F., Wang, B., & Huang, C. (2022). Frame design of the BIM based budget software of nuclear power projects. Journal of Computational Methods in Science and Engineering, 22(4), 1195–1207.

Park, J., Cai, H., Dunston, P. S., & Ghasemkhani, H. (2017). Database-supported and web-based visualization for daily 4D BIM. Journal of Construction Engineering and Management, 143(10), 04017078.

Sheik, N. A., Veelaert, P., & Deruyter, G. (2023). Exchanging progress information using IFC-based BIM for automated progress monitoring. Buildings, 13(9), 2390.

That Open Company (Versione 2.4.0). (2025). [Software]. https://github.com/ThatOpen/

van Berlo, L., Krijnen, T., Tauscher, H., Liebich, T., Van Kranenburg, A., & Paasiala, P. (2021). Future of the industry foundation classes: Towards IFC 5. 78, 11–15.

Veneto Region. (2025). Veneto Region Price List (Versione 3.3.1) [Dataset]. https://prezzario.regione.veneto.it/

Yang, B., Dong, M., Wang, C., Liu, B., Wang, Z., & Zhang, B. (2021). IFC-based 4D construction management information model of prefabricated buildings and its application in graph database. Applied Sciences, 11(16), 7270.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A SEMANTIC DATA MODEL FOR INTEROPERABLE EXCHANGE OF TEST TRENCH VERIFICATION DATA

Faith Tangara[1], Léon olde Scholtenhuis[1], and Ramon ter Huurne[1]
[1]University of Twente, Enschede, The Netherlands

Abstract

In the Netherlands, nearly 50,000 utility strikes occurred in 2024, costing over €57 million in direct repairs. Test trenches provide centimeter-accurate verification of underground utility positions, yet this valuable data remains trapped in organizational silos, leading to repeated excavations. This paper introduces UTT (Universal Test Trench), a semantic data model extending the Dutch IMKL standard for interoperable exchange of 3D verification data. Developed using OWL/RDF with SHACL validation constraints, UTT provides a unified vocabulary for test trench observations, positions, deviations, and measurement quality. Future work includes prototype implementation and field evaluation.

Introduction

The underground infrastructure of cities contains a complex network of cables and pipes that deliver essential services, including electricity, gas, water, telecommunications, and sewerage. In the Netherlands alone, this network extends over 1.7 million kilometers. As cities pursue energy transitions and climate resilience programs, subsurface construction activity intensifies significantly. Current projections suggest that within the coming decade, one-third to half of urban roads will require reconstruction to accommodate new infrastructure systems including expanded electrical capacity, hydrogen pipelines, and enhanced drainage (COB, 2024).

This intensifying activity correlates with rising damage incidents. In 2024, the Netherlands recorded nearly 50,000 utility strikes at a rate of 137 daily occurrences, with direct repair costs exceeding €57 million (Digitale Infrastructuur, 2024). Including indirect costs such as service disruptions, project delays, and third-party damages, the true economic impact is estimated at 29 times higher (Makana et al., 2018). The root cause lies fundamentally in insufficient localization insights: the Cables and Pipelines Information Centre (KLIC) provides 2D maps based on administrative records that often deviate by over one meter horizontally and typically lack depth information entirely (Yan et al., 2019).

To mitigate this geometric uncertainty, contractors routinely excavate test trenches that expose underground utilities for direct measurement. A typical test trench is one to two meters deep and two to three meters long, allowing surveyors to record positions with centimeter-level accuracy using total stations, GNSS/RTK receivers, or increasingly, LiDAR scanning and photogrammetry. Yet excavating these trenches is itself costly, time-consuming, and carries risk of damaging the very utilities being located, making repeated excavation at the same location inherently undesirable. Despite this, the verified data remains trapped in silos within individual organizations. Once the trench is backfilled, measurements are rarely preserved in structured formats or shared with other stakeholders who may later excavate the same location.

The Dutch Centre for Underground Construction (COB) conducted feasibility studies in 2023 and 2024 that confirmed the value proposition of a ‘test trench database’(COB, 2023, 2024). This database should enable unified exchange of trial trench data among stakeholders in the Dutch utility sector. The feasibility studies found broad stakeholder support and identified standardization of test trench data as the critical prerequisite for implementation. It was recognized that without a common data model, interoperability of data would be compromised, hampering the exchange of test trench data.

This paper addresses this standardization gap by introducing UTT (Universal Test Trench), a semantic data model that extends the Dutch IMKL standard to capture 3D test trench verification data. While 3D utility data is not yet common practice, it is rapidly becoming essential as technologies such as LiDAR and advanced surveying tools gain widespread adoption. The ambition of UTT is therefore not only to standardize current field observations but also to provide future-proof infrastructure for the next generation of high-resolution utility mapping.

The remainder of this paper is organized as follows: the literature review examines related work in infrastructure ontologies and utility data standards; the methodology section describes the ontology development approach; subsequent sections present the UTT design, validation approach, and evaluation results; finally, discussion and conclusions address practical implications and future directions.

Literature Review

Standardization and Ontologies

Standardization is critical to facilitate communication between stakeholders in fragmented industries like construction (Howard & Björk, 2008). Without shared standards, organizations develop their own digital models using different taxonomies and vocabularies for the same assets (A Ter Huurne & Olde Scholtenhuis, 2018) . An ontology addresses this by providing an explicit specification of a shared conceptualization (Studer et al., 1998), encoding agreed concepts based on shared worldviews (El-Diraby & Osman, 2011). Semantic web technologies, particularly OWL and RDF, offer significant advantages over conventional relational or XML-based approaches for this purpose. Linked data principles enable distributed data integration without requiring a single centralized database, making them well-suited to multi-stakeholder environments where organizations maintain their own data stores (Berners-Lee, 2006).

Ontology in Infrastructure Engineering

Semantic web technologies have found application in infrastructure data management. In building information modeling, ifcOWL demonstrates ontology-based data integration (Pauwels & Terkaj, 2016).

For bridge infrastructure, linked data approaches support condition assessment (Hamdan & Scherer, 2020). The CityGML UtilityNetworkADE provides 3D utility representation within city models, with stronger geospatial integration than building-centric approaches (Biljecki et al., 2018 ; Kutzner et al., 2020).

Subsurface utilities remain underrepresented in semantic approaches. While above-ground infrastructure benefits from visual inspection and established BIM workflows, underground assets present unique challenges: they are invisible after installation, subject to undocumented modifications, and often predate digital documentation practices. These characteristics make semantic interoperability valuable yet technically challenging.

Utility Data Standards

Several utility data standards have emerged, each serving a particular purpose yet failing to accommodate the concepts required to comprehensively describe 3D test trench verification information.

At the European level, INSPIRE Utility Networks provides cross-border data exchange concepts for utility network topology (European Commission, 2013). CityGML's UtilityNetworkADE extends this to 3D city model integration, enabling rich geometric representation of underground networks (Kutzner et al., 2020). In the Netherlands, IMKL 3.0 provides the national standard for utility data exchange and maps to INSPIRE for European interoperability (Geonovum, 2022). However, IMKL primarily supports 2D registration and lacks concepts for verification observations or measurement uncertainty. Ter Huurne (2019) developed an OWL-based extension addressing semantic integration in utility asset management, demonstrating the feasibility of linked data approaches in this sector.

PAS 128 defines quality levels for utility detection but functions as an operational directive for survey procedures rather than a data exchange standard (BSI, 2014). International efforts like MUDDI address registry-level data but not verification observations. The gap between registered positions and actual field conditions remains outside existing standards.

While ontologies exist for building and bridge infrastructure, and data standards exist for utility network registration, no semantic model addresses the specific requirements of test trench verification data, including observed positions, deviations from registry data, measurement uncertainty, and verification provenance. UTT addresses this gap.

Methodology

The central goal of ontology development is achieving a shared conceptualization that enables meaningful data exchange between stakeholders who currently use incompatible representations. The methodology follows the Linked Open Terms (LOT) approach (Poveda-Villalón et al., 2022), a systematic framework comprising four iterative phases: requirements specification, implementation, publication, and maintenance. This paper covers the first three phases. LOT was selected

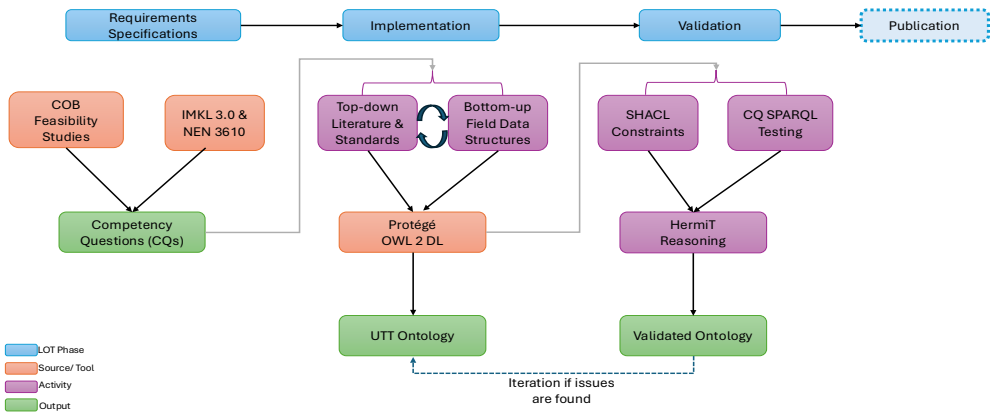


Figure 1:UTT ontology development process following the LOT methodology

because it integrates competency question formulation with standard ontology engineering practices, providing a reproducible process where requirements are formalized as testable queries against the resulting model.

Ontology Requirements Specification

Requirements were gathered through three complementary approaches. First, analysis of two COB feasibility study reports from 2023 and 2024, which collectively involved workshops with 15 organizations including utility network owners, contractors, surveying firms, and Kadaster, identified functional requirements from industry stakeholders. Second, systematic review of IMKL 3.0 documentation and NEN 3610 (the Dutch base model for geo-information) established alignment requirements with existing national standards. Third, iterative discussions with two industry partners from a utility contractor organization provided input on practical data capture workflows and current data formats in use.

The requirements specification yielded the following scope: the ontology must represent (1) the physical test trench excavation with its geometry and dimensions; (2) individual utility observations within the trench; (3) verified 3D positions with quantified accuracy; (4) deviations between observed and registered positions; (5) measurement methods and equipment; and (6) verification provenance including performer, date, and commissioning organization.

Requirements were formalized as competency questions (CQs) that the ontology must answer. CQs function as testable specifications: each translates to a SPARQL query that verifies whether the ontology captures required concepts. Table 1 lists the ten competency questions that guided development.

Table 1:Competency questions defining ontology requirements

ID	Competency Question
CQ1	What classes represent the test trench and its observations?
CQ2	What properties capture depth information for utilities?
CQ3	What measurement methods are defined in the ontology?
CQ4	What accuracy and quality properties are available?
CQ5	What properties link observations to IMKL concepts?
CQ6	What properties capture deviation between registered and observed?
CQ7	What provenance properties track verification performers?
CQ8	What supporting data file formats are defined?
CQ9	What IMKL utility types (thema) can be referenced?
CQ10	Is the trench class properly linked to spatial features?

Ontology Implementation

Implementation combined top-down and bottom-up development strategies (Noy & Mcguinness, 2001). The top-down approach defined concepts derived from IMKL documentation, NEN 3610 spatial object patterns, and PROV-O provenance structures. The bottom-up approach incorporated terminology extracted from existing test trench reports, surveyor workflows, and data formats currently used in Dutch practice.

The ontology was developed using Protégé 5.5 and encoded in OWL 2 DL following ISO 19150-2 transformation rules for geographic information. This standard provides patterns for converting UML-based geographic information models to OWL, ensuring compatibility with Dutch geo-information infrastructure. SHACL (Shapes Constraint Language) shapes define validation constraints for data quality enforcement (W3C, 2012).

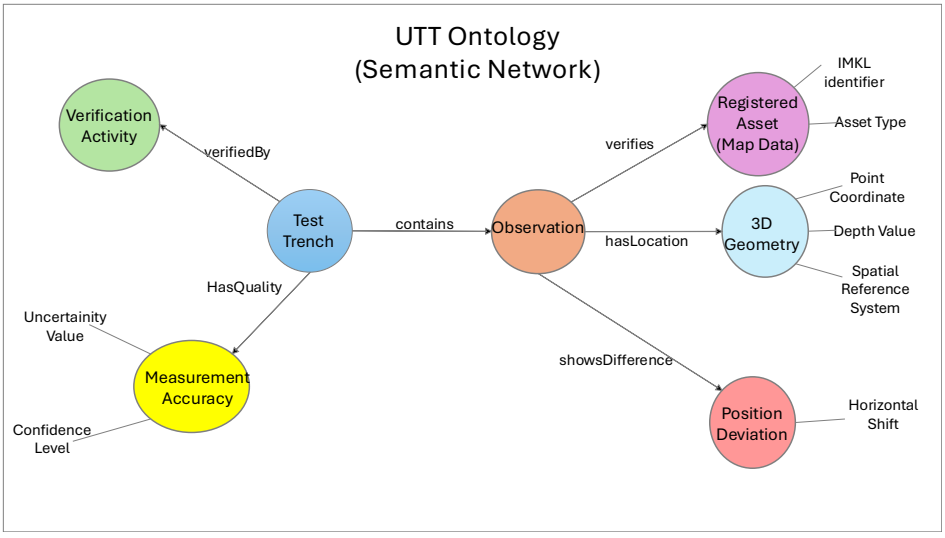


Figure 2:UTT ontology conceptual overview with core classes and relationships

The implementation reuses established ontologies following linked data best practices. GeoSPARQL provides spatial representation capabilities including geometry types and spatial relations. PROV-O (Provenance Ontology) supplies patterns for tracking data

lineage and attribution. SKOS (Simple Knowledge Organization System) structures controlled vocabularies for measurement methods and file formats.

Table 2: Namespace prefixes and URIs

Prefix	Namespace URI
utt:	https://w3id.org/utt/def/
imkl:	https://data.geostandaarden.nl/imkl/def/
geo:	http://www.opengis.net/ont/geosparql#
prov:	http://www.w3.org/ns/prov#
skos:	http://www.w3.org/2004/02/skos/core#

UTT Ontology Design

The UTT ontology comprises 13 classes, 21 object properties, and 17 datatype properties organized into seven functional modules. The complete ontology totals 378 RDF triples with an additional 275 triples in the IMKL subset and 175 triples defining SHACL validation shapes. Figure 2 presents the semantic network showing class relationships and IMKL integration points.

Test Trench and Geometry

The `utt:Proefsleuf` (Test Trench) class serves as the central container for the verification process.

Unlike standard IMKL definitions which focus on the utility network topology, UTT treats the trench as a distinct geospatial feature. A key design decision was to define this class as a subclass of `geo:Feature` (GeoSPARQL). This inheritance allows the trench to possess its own 3D geometry `utt:sleufGeometrie` (Trench Geometry), independent of the utilities contained within it.

UTT represents geometries using GeoSPARQL's `geo:hasGeometry` pattern. Trench geometries are stored as 3D Well-Known Text (WKT) polygon literals representing the excavation footprint, while observed utility positions use 3D WKT point geometries for discrete measurements or 3D linestring geometries for utility segments visible across the trench. Geometries are typed as `geo:wktLiteral` and annotated with their coordinate reference system. For Dutch geo-infrastructure integration, geometries use the Rijksdriehoek coordinate reference system (EPSG:28992) with NAP-referenced heights. This approach enables GeoSPARQL spatial queries such as `geo:sfWithin` and `geo:sfIntersects` while maintaining direct compatibility with GIS tools that consume WKT.

This structural separation is critical for spatial analysis: it allows the ontology to query which assets fall within the excavation volume, regardless of whether they were originally documented in the registry. The class creates a context for areas where utilities were expected but not found by defining the exact boundaries of the verified space.

The use of Dutch-language class identifiers and the Rijksdriehoek coordinate reference system reflects the decision to align UTT with IMKL, the mandatory national standard for utility data exchange in the Netherlands. All classes and properties carry bilingual annotations in both Dutch and English through rdfs:label declarations. The coordinate reference system is specified on individual geometry literals following GeoSPARQL conventions rather than encoded as a structural constraint, permitting adaptation to other national coordinate systems at the instance level.

Observation and Registry Alignment

The `utt:Waarneming` (Observation) class bridges the gap between the physical reality and the administrative registry. While existing standards like MUDDI focus on the asset itself, UTT distinguishes the act of observing from the asset observed. This separation enables the system to handle unknown or "orphan" cables that do not exist in the KLIC registry.

To ensure interoperability, the observation class utilizes the `utt:waargenomenType` (observation) property, which is mapped directly to the `imkl:Thema` (Utility Type) code lists. This design choice ensures that a field surveyor's classification (e.g., "Low Voltage Cable") is semantically identical to the registry's classification, facilitating automatic matching. The `utt:komtOvereenMet` (matches) property explicitly links the observation to a specific `imkl:KabelOfLeiding` (Cable or Pipeline) instance using the NEN 3610 Unique Identifier (GUID). This mechanism effectively connects the physical observation to the official digital record to enable precise validation.

Position Deviation

In many existing utility models, accuracy is treated as a simple attribute. However, to support engineering decisions, UTT models deviation as a distinct class rather than a data property. This reification strategy was chosen to capture the complex vector relationship between the planned and actual location. By treating deviation as a class, the ontology can store multiple dimensions of the discrepancy specifically `utt:horizontaleAfwijking` (horizontal shift) and `utt:verticaleAfwijking` (vertical shift) along with the `utt:afwijkingRichting` (directionality). This allows for querying, such as detecting consistent depth errors in specific asset types.

Verification Provenance

The `utt:VerificatieActiviteit` (Verification Activity) class identifies the specific events and actors responsible for the data collection, ensuring traceability for the verification process (see Figure 3). Modeled as a subclass of `prov:Activity` from the W3C PROV-O standard, this class serves as the connecting node between the physical trench and the administrative data.

This module explicitly links the verification to the `utt:Landmeter` (Surveyor), defined as a subclass of `prov:Agent`, and the `utt:meetmethode` (Measurement Method). By structuring this information as semantic relationships rather than simple text strings, the ontology enables advanced data filtering based on the

acquisition source. For instance, data consumers can distinguish between measurements captured via RTK-GNSS versus those verified by a Total Station, allowing for automated reliability assessments based on the equipment's inherent precision capability.

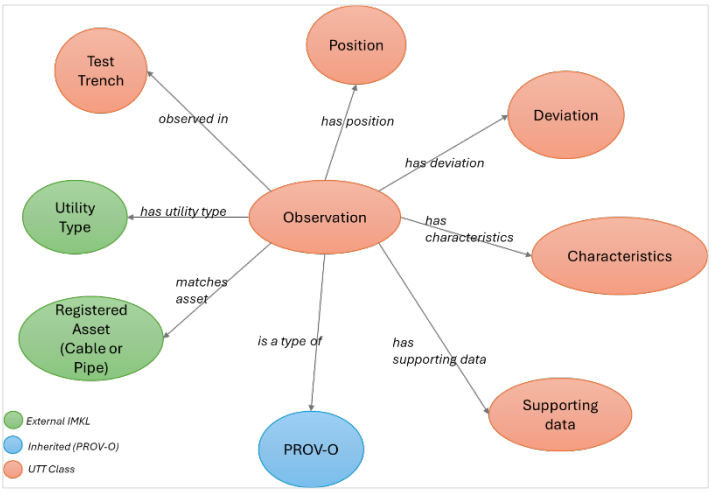


Figure 3:Observation module linking field data to IMKL

Data Validation with SHACL

The ontology includes SHACL (Shapes Constraint Language) shapes that enforce data quality rules. These shapes validate structural completeness, cardinality constraints, and value ranges.

The `utt:ProefSleufShape` (Test Trench Shape) requires each test trench to have exactly one geometry, one verification activity link, and at least one observation. Dimensional properties are constrained to reasonable ranges: trench length up to 50 meters, width up to 10 meters, and depth up to 10 meters.

The `utt:NauwkeurigheidShape` (Accuracy Shape) mandates horizontal uncertainty between 0 and 1.0 meter and vertical uncertainty between 0 and 0.5 meter. These ranges reflect typical achievable accuracies with professional surveying equipment while flagging implausible values.

Severity levels distinguish between violations (structural requirements) and warnings (recommended best practices). For example, linking an observation to a registered IMKL asset is recommended (severity: Info) but not required, acknowledging that unregistered utilities are sometimes discovered.

IMKL Integration Strategy

Integration with IMKL follows a complementary approach rather than modification of the base standard. UTT imports an OWL subset of IMKL classes transformed following ISO 19150-2 patterns. This subset includes core network classes *(Utiliteitsnet (Utility Network), KabelOfLeiding (Cable or Pipeline), Leidingelement (Pipeline Element))*, depth classes, and relevant code lists.

UTT reuses IMKL code lists for utility types *(Thema),* accuracy classifications *(NauwkeurigheidXYvalue (Accuracy XY Value), NauwkeurigheidDiepteValue (Accuracy Depth Value))*, depth reference points *(DiepteAangrijpingspuntValue (Depth Reference Point Value))*, and pipe materials *(PipeMaterialTypeIMKLvalue)*. This reuse ensures observations can be categorized consistently with registry data and supports queries that span both data sources.

The `utt:komtOvereenMet` (Matches) property creates explicit links between UTT observations and IMKL registered assets. This enables deviation calculations and supports use cases where users need to compare verified positions against registered positions for specific utilities.

Query Capabilities & Validation

The ontology enables SPARQL queries answering the competency questions in Table 1. The following query demonstrates retrieval of observations deviating more than 50cm from registered positions:

```
PREFIX utt: <https://w3id.org/utt/def/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?trench ?observation ?depth ?utilityType
WHERE {
  ?trench a utt:Proefsleuf ;
          utt:heeftWaarneming ?observation .
  ?observation utt:heeftPositie ?position ;
               utt:waargenomenType ?utilityType .
  ?position utt:diepteTovMaaiveld ?depth .
  FILTER (?depth > 1.5)
}
```

Figure 4:SPARQL query for CQ3 - finding utilities deeper than threshold

This query returns all observations where utilities were found deeper than 1.5 meters below ground level, along with their utility type classification. The results enable analysis of depth distribution patterns across utility types and geographic areas. To calculate position deviations;

```
PREFIX utt: <https://w3id.org/utt/def/>

SELECT ?observation ?hDeviation ?vDeviation
WHERE {
  ?observation a utt:Waarneming ;
               utt:heeftAfwijking ?deviation .
  ?deviation utt:horizontaleAfwijking ?hDeviation ;
             utt:verticaleAfwijking ?vDeviation .
  FILTER (?hDeviation > 0.5 || ?vDeviation > 0.3)
}
```

Figure 5:SPARQL query for CQ7 - calculating position deviations

This query identifies observations where registered positions deviated significantly from field-verified positions and supports systematic accuracy analysis of registry data.

The logical consistency was verified using HermiT reasoner in Protégé 5.5 to confirm no contradictory axioms exist and there are no unsatisfiable classes. The competency question was verified by translating each CQ into a SPARQL query. All ten CQs from Table 1 were

successfully answerable using the implemented ontology to confirm functional completeness for the defined scope.

To demonstrate the representational capacity of the ontology, Figure 6 presents an RDF/Turtle serialization of a single test trench containing two utility observations.

```
@prefix utt:  <https://w3id.org/utt/def/> .
@prefix imkl: <https://data.geostandaarden.nl/imkl/def/> .
@prefix geo:  <http://www.opengis.net/ont/geosparql#> .
@prefix xsd:  <http://www.w3.org/2001/XMLSchema#> .

utt:trench_A1 a utt:Proefsleuf ;
  utt:sleufLengte "2.5"^^xsd:decimal ;
  utt:sleufBreedte "2.0"^^xsd:decimal ;
  utt:sleufDiepte "1.8"^^xsd:decimal ;
  utt:sleufGeometrie [
    a geo:Geometry ;
    geo:asWKT "POLYGON Z((155000.0 463000.0 0.0,
      155002.5 463000.0 0.0, 155002.5 463002.0 0.0,
      155000.0 463002.0 0.0, 155000.0 463000.0
      0.0))"^^geo:wktLiteral ] ;
  utt:heeftWaarneming utt:obs_01 , utt:obs_02 ;
  utt:geverifieerdDoor utt:activiteit_01 .

utt:obs_01 a utt:Waarneming ;
  utt:waargenomenType imkl:laagspanning ;
  utt:heeftPositie [
    a utt:WaargenomenPositie ;
    utt:diepteTovMaaiveld "0.85"^^xsd:decimal ;
    utt:positieGeometrie [
      a geo:Geometry ;
      geo:asWKT "POINT Z(155001.2 463001.0
        -0.85)"^^geo:wktLiteral ] ] ;
  utt:heeftAfwijking [
    a utt:PositieAfwijking ;
    utt:horizontaleAfwijking "0.35"^^xsd:decimal ] .
```

Figure 6:RDF/Turtle serialization of a test trench instance with two utility observations

The trench instance is defined with its dimensions and a polygon geometry in three-dimensional coordinate space representing the excavation footprint, with depth captured separately. The first observation records a low-voltage electricity cable at 0.85 meters depth. The associated `PositieAfwijking` (Position Deviation) quantifies that this cable was 0.35 metres horizontally displaced from its registered position. The second observation records a telecommunications cable at 1.10 meters depth without a deviation instance, indicating an unregistered utility with no corresponding registry entry.

Discussion

This paper introduced UTT as a semantic data model addressing the absence of a standardized vocabulary for test trench verification data. Through its development, this study contributes to both the utility data modeling literature and to practice.

To the literature, UTT demonstrates that 3D field verification observations can be represented as linked data by extending existing geo-information standards. Where current standards such as IMKL, INSPIRE, and CityGML UtilityNetworkADE address network registration and topology (Kutzner et al., 2020 ; Geonovum, 2022) UTT introduces complementary concepts for the verification layer: what was actually found in the field, how accurately it was measured, and how it deviates from official records. The reification of position deviation as a distinct class, rather than a simple attribute, enables systematic analysis of registry accuracy patterns, a capability absent in existing utility data models. The integration of PROV-O provenance patterns with domain-specific measurement quality further demonstrates how general-purpose semantic web vocabularies can be combined with domain standards for infrastructure applications.

For Dutch practice, UTT directly supports the national test trench database initiative identified by COB (2024) as critical for reducing repeated excavations. The alignment with IMKL code lists ensures that field observations use identical classifications to the KLIC registry, enabling automatic matching. The explicit uncertainty values in the `Nauwkeurigheid` (Accuracy) class go beyond IMKL's categorical accuracy classes, supporting risk-based excavation planning where engineers can assess whether prior observations meet their precision requirements.

More broadly, test trench excavation is common practice in utility infrastructure management internationally, and the discrepancy between registered and field-verified positions is not confined to the Netherlands. The Dutch-specific elements of the current implementation are limited to the IMKL code lists for utility type classification and the Rijksdriehoek coordinate reference system, both of which are instance-level parameters rather than structural features of the schema. Adaptation to other national contexts would involve substituting these classifications and coordinate declarations while the class hierarchy and property structure remain intact. Because IMKL maps to the European INSPIRE Utility Networks specification, UTT inherits an existing interoperability pathway across jurisdictions. The ontology thus provides a reusable semantic layer for subsurface digital twins, distinguishing as-found observations from as-registered data.

This study has several limitations. The current design focuses on point and line observations; continuous corridor surveys along extended utility routes may require additional geometric patterns. The ontology has been validated for logical consistency and competency question coverage but not yet with large-scale production datasets from multiple organizations. Future work will focus on prototype implementation of the ontology within a triple store, demonstrating data exchange between organizations, and evaluating its merit in operational utility localization practice.

Conclusions

This paper presented UTT, a semantic data model for interoperable exchange of test trench verification data in the utility infrastructure sector. The ontology extends IMKL with concepts for test trenches, field observations, position deviations, quantified accuracy, and verification provenance. SHACL shapes enforce data quality constraints appropriate to the domain.

By providing a standardized vocabulary and validation framework, UTT addresses a critical gap in the utility data

infrastructure that contributes to repeated excavations and damage incidents. The model provides a foundation for developing subsurface digital twins where verified, provenance-tracked observations complement existing registry data.

Future work will deploy the ontology within a GeoSPARQL-enabled triple store to evaluate federated querying across organizational boundaries. The data model will be validated using field datasets from utility contractor organizations to assess representational adequacy across varying trench configurations and utility types. Structured evaluation workshops with industry stakeholders will determine whether the ontology supports decision-making in excavation planning

The ontology is available under Creative Commons Attribution 4.0 license and can be accessed here.

Acknowledgments

This research is conducted within the ZoARG research program. The authors thank participants from industry partners for their input during stakeholder workshops.

References

Berners-Lee, T. (2006). Linked Data - Design Issues. W3C Design Issues. https://www.w3.org/DesignIssues/LinkedData.html

Biljecki, F., Kumar, K., & Nagel, C. (2018). CityGML Application Domain Extension (ADE): overview of developments. Open Geospatial Data, Software and Standards, 3(1), 1–17. https://doi.org/10.1186/S40965-018-0055-6

BSI. (2014). PAS 128:2014 Specification for underground utility detection, verification and location.

COB. (2023). Haalbaarheid proefsleuvendatabase. https://www.cob.nl/document/haalbaarheid-proefsleuvendatabase/

COB. (2024). Haalbaarheid proefsleuvendatabase, fase 2. https://www.cob.nl/document/haalbaarheid-proefsleuvendatabase-fase-2/

Digitale Infrastructuur. (2024). Infographic graafschades 2024.

El-Diraby, T. E., & Osman, H. (2011). A domain ontology for construction concepts in urban infrastructure products. Automation in Construction, 20(8), 1120–1132. https://doi.org/10.1016/j.autcon.2011.04.014

European Commission. (2013). INSPIRE Data Specification on Utility and Government Services – Technical Guidelines. https://knowledge-base.inspire.ec.europa.eu/publications/inspire-data-specification-utility-and-government-services-technical-guidelines_en

Geonovum. (2022). Informatiemodel Kabels en Leidingen (IMKL) - versie 3.0. https://docs.geostandaarden.nl/kl/imkl/

Hamdan, A.-H., & Scherer, R. J. (2020). Integration of BIM-related bridge information in an ontological knowledgebase. https://github.com/Alhakam/bridgeOntology

Howard, R., & Björk, B. C. (2008). Building information modelling – Experts' views on standardisation and industry deployment. Advanced Engineering Informatics, 22(2), 271–280. https://doi.org/10.1016/j.aei.2007.03.001

Kutzner, T., Chaturvedi, K., & Kolbe, T. H. (2020). CityGML 3.0: New Functions Open Up New Applications. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science, 88(1), 43–61. https://doi.org/10.1007/s41064-020-00095-z

Makana, L., Metje, N., Jefferson, I., Sackey, M., & Rogers, C. (2018). Cost estimation of utility strikes. Infrastructure Asset Management, 7(2), 64–76. https://doi.org/10.1680/jinam.17.00033

Noy, N. F., & McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory.

Pauwels, P., & Terkaj, W. (2016). EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. Automation in Construction, 63, 100–133. https://doi.org/10.1016/J.AUTCON.2015.12.003

Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., & García-Castro, R. (2022). LOT: An industrial oriented ontology engineering framework. Engineering Applications of Artificial Intelligence, 111, 104755. https://doi.org/10.1016/J.ENGAPPAI.2022.104755

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. Data & Knowledge Engineering, 25(1–2), 161–197. https://doi.org/10.1016/S0169-023X(97)00056-6

Ter Huurne, R. B. A. (2019). Modelling utilities by developing a domain ontology. University of Twente. https://research.utwente.nl/en/publications/modelling-utilities-by-developing-a-domain-ontology/

Ter Huurne, R. B. A., & Olde Scholtenhuis, L. L. (2018). Digitization for integration: Fragmented realities in the utility sector. In Proceedings of the 34th Annual ARCOM Conference, Belfast, UK.

W3C. (2012). OWL 2 Web Ontology Language Document Overview (Second Edition). https://www.w3.org/TR/owl2-overview/

Yan, J., Jaw, S. W., Soon, K. H., Wieser, A., & Schrotter, G. (2019). Towards an Underground Utilities 3D Data Model for Land Administration. Remote Sensing, 11(17), 1957. https://doi.org/10.3390/rs11171957

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

COMPARATIVE ANALYSIS OF UWB AND IMU FOR CONSTRUCTION SAFETY MONITORING

Serdar Bayraktar and Semra Çomu
Boğaziçi University, Istanbul, Türkiye

Abstract

This study compares Ultra-Wideband (UWB) location-based tracking against Inertial Measurement Unit (IMU) gait-based activity recognition for construction safety monitoring. Experiments were conducted in a controlled environment simulating construction hazards (walking on planks, scaffolding, and mezzanines). Results indicate a trade-off: IMU classifiers effectively detect kinematic changes (81.6% accuracy) but fail in spatial awareness, whereas UWB successfully identifies elevation zones (92.3% accuracy) but suffers from vertical signal noise. Rather than viewing these limitations as failures, this study establishes the operational boundaries of standalone UWB and IMU systems in vertical environments, providing empirical baseline parameters for future sensor fusion design.

Introduction

The construction industry is known for its hazardous work environment (Soltanmohammadlou et al., 2019). In 2023, the European Union reported the highest incidence of non-fatal accidents at work within the construction sector, recording 2,899 accidents per 100,000 employed persons (Eurostat, 2025). These statistics highlight the need for targeted interventions to improve the health and safety standards of the construction sites. One important aspect is tracking the site personnel which can help to inhibit occurrence of the accidents by taking the right measures, also enabling the rapid emergency response if an accident occurs. Traditional safety monitoring relies on direct human observations and could not provide real-time coverage of large and complex construction sites (Akhavian & Behzadan, 2016; Zhang et al., 2018). However modern technologies allow real-time tracking of workers, material and equipment. Among these, Ultra-Wideband (UWB) for localization and Inertial Measurement Units (IMU) for activity recognition have emerged as prominent solutions, offering different approaches to safety monitoring.

The UWB positioning system is widely considered as standard for real-time tracking in construction, ensuring centimeter-level error with a clear line of sight (LOS) (Naheem & Kim, 2022). Unlike other wireless technologies such as BLE and Wi-Fi, which rely on Received Signal Strength Indicator (RSSI), UWB operates based on time-of-flight (ToF) measurements. Its short wavelength and high bandwidth contribute to robustness against multipath fading and interference when compared to alternative wireless technologies (Rao et al., 2022). In construction environments, UWB enables reliable real-time positioning of workers, materials, and equipment, particularly in open and unobstructed settings.

Despite its advantages over other wireless technologies, UWB faces limitations under Non-Line-of-Sight (NLOS) conditions, as its positioning accuracy degrades when signals are obstructed by reinforced concrete elements and steel reinforcements commonly found on construction sites (Li & Li, 2025). NLOS leads to disappearances of worker's positions. Another drawback of UWB is the requirement of system setup. The UWB system comes with an upfront cost and needs time and effort to set the anchor sensor throughout the construction site with precise georeferencing before tracking can begin.

An IMU is a micro-electromechanical system (MEMS) sensor package that consists of a 3-axis Accelerometer, 3-axis Gyroscope and generally a magnetometer to measure linear acceleration, angular velocity and magnetic field strength, respectively. By integrating these measurements, position relative to a starting position can be determined (Rao et al., 2022; Leitch et al., 2023). IMU data offers a complementary perspective by capturing both the type of activity and its spatial context (Trigo et al., 2025). Unlike location-based systems, IMUs are not affected by NLOS conditions, enabling uninterrupted tracking (Li & Li, 2025), and their integration into smartphones further improves accessibility. Nevertheless, IMU-based approaches are limited to relative tracking rather than absolute positioning.

In safety-critical construction applications where absolute positioning is required, UWB enables proactive safety measures such as virtual fencing and collision avoidance (Teizer et al., 2008).However, environmental clutter significantly degrades accuracy; metallic obstructions can increase positioning errors from 0.4 m to 1.2 m due to multipath effects (Shahi et al., 2012). Geometric constraints also play a role, as linear anchor layouts cause Dilution of Precision (Maalek & Sadeghpour, 2013). Kılıç et al. (2025) demonstrated that for the MDEK1001 sensors, body placement is decisive: head-mounted tags achieve >99% signal reception by maintaining Line-of-Sight.

In contrast, IMU-based approaches are primarily employed to identify and classify worker activities rather than to provide absolute positioning. Akhavian & Behzadan (2016) and Zhang et al. (2018) demonstrated that sensors embedded in smartphones can reliably recognize construction tasks. However, IMUs do not offer a fixed spatial reference; Li & Li (2025) explain this limitation through accumulated errors, where sensor noise and environmental vibrations progressively distort the estimated trajectory, causing rapid drift and limiting the suitability of standalone IMUs for independent georeferencing.

Consequently, this study presents a comparative analysis of two commonly used approaches: i) infrastructure-dependent UWB systems, which require prior installation, and ii) IMU-based tracking available on standard smartphones. While both sensing modalities have been explored in general tracking applications, this work focuses on experimentally evaluating their performance across different walking modes in complex, multi-level construction environments (e.g., scaffolds, planks, and mezzanine levels). Rather than aiming to identify a single comprehensive solution for continuous safety monitoring, the goal is to examine the practical trade-offs between location-based positioning (UWB) and motion-based recognition (IMU), and to highlight the conditions under which each approach performs more reliably.

Methodology

To evaluate the different methods of classification of walking activity, a multi-sensor data collection was held in a controlled structural engineering laboratory environment. This experiment involved 5 male participants, composed of graduate students and structures lab technicians with ages ranging 25 to 60 years.

IMU of Apple iPhone devices was utilized. 3-axis accelerometer and 3-axis gyroscope of the IMU captured linear acceleration and angular velocity data with App Store achievable smartphone application. The IMU data was captured at a sampling rate of 100 Hz. Two smartphones were placed on the participants' bodies. One of the smartphones was placed on the lumbar region of the participants to capture data close to the body center of mass. The other smartphone was placed in the right front pocket of the participant, representing the practical sensor placement (Figure 1).



Figure 1: Participant instrumented with a UWB tag (hardhat) and smartphones (pocket and lumbar) for data collection.

Location tracking within the controlled experiment setup was performed using 11 Qorvo MDEK1001 Ultra-Wideband (UWB) Transceiver Development Kit modules manufactured by Qorvo. Nine of the UWB sensors were assigned as anchor nodes and were positioned within the experiment area to ensure full coverage possible. The distances between the anchor nodes were measured with a total station and the relative coordinates of the nodes were entered into the Qorvo application available in the App Store. One of the UWB sensor was assigned as a tag and was mounted on the front of the hardhat worn by the participants (Figure 1). Another UWB sensor was assigned as a listener and connected to a Raspberry Pi 5 microcomputer. The listener sensor communicated with the tag sensor and received the relative position among the experiment setup. The location data was collected during each participant's experiment and saved in the CSV format at the end of each trial.

Each participant followed the same predefined circuit four times for their turn. The circuit included four walking scenarios: normal walking, elevated walking, plank walking and scaffolding walking. Normal walking refers to participant’s walking on the ground level. Elevated walking refers to participant’s walking on the mezzanine level. Plank walking refers to participant’s walking on a plank. The plank has 3 m length, 15 cm width and 20 cm height above the ground. Scaffolding walking refers to walking on a scaffold structure.

The ground truth was established by capturing video recordings of all the trials. The video recordings were captured in 720p at 60 fps. Video recordings were later used for labeling the IMU and UWB data.

The raw data collected from the multi-sensor systems were processed and synchronized to enable a direct comparison between the IMU-based system and the UWB-based system. All processing steps were implemented using Python.

The video recordings were labeled using ELAN - Linguistic Annotator software to establish the ground truth (Figure 2). Precise start and end timestamps were generated for five different walking activities on different surfaces: normal walking on ground level(n_walk), walking on a plank(p_walk), walking on mezzanine level(e_walk), walking on upper level of the scaffolding

(s_walk_upper), and walking on lower level of the scaffolding (s_walk_lower). The raw IMU and UWB data were synchronized with these labels using Unix timestamps, and non-relevant segments were removed from the dataset.

A 4th-order Butterworth low-pass filter with a cutoff frequency of 30 Hz was applied to both accelerometer and gyroscope signals to minimize the impact of environmental noises. The IMU data were segmented into sliding windows of 2.0 seconds with 50% overlap. For each window, statistical and frequency-domain features were extracted from the three orthogonal axes of the accelerometer and gyroscope, as well as their calculated magnitudes. A total of 13 features were computed per channel, including mean, standard deviation,

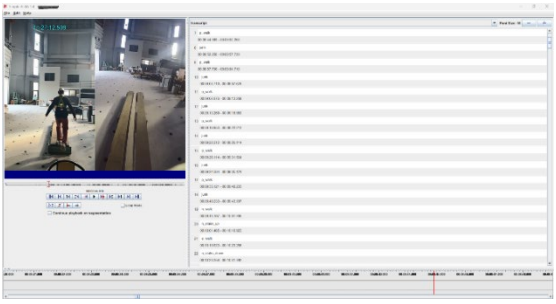


Figure 2: Ground truth labeling interface using ELAN software to segment walking activities

root mean square (RMS), skewness, kurtosis, interquartile range, Shannon entropy, dominant frequency, and spectral entropy.

In terms of determining the optimal configuration for activity recognition, a systematic grid search analysis was conducted. We evaluated sliding window durations ranging from 0.5 to 3.0 seconds, overlap ratios between 25% and 75%, and various feature subsets. Additionally, multiple machine learning algorithms such as linear models (Logistic Regression, LDA), probabilistic models (Naive Bayes), distance-based methods (k-NN), and tree-based ensembles (Random Forest, XGBoost) were compared. The Random Forest classifier (N = 100 trees), paired with a 2.0-second window and 50% overlap, consistently yielded the highest F1-score and generalization performance. Consequently, this configuration was selected for the final analysis. Performance was evaluated using a Leave-One-Subject-Out (LOSO) cross-validation scheme.

The UWB data was processed using a rule-based spatial algorithm that relies on the participant's position relative to the zones defined in the experimental setup. To mitigate signal jitter and non-line-of-sight (NLOS) outliers caused by environmental factors, a context-aware smoothing pipeline was implemented. Rather than applying a blanket filter, the raw coordinate data was first segmented based on hardware disconnections (time gaps > 1.5 seconds) and trial state changes. A rolling median filter (window size = 5) was then applied independently to each segment. This approach prevented the blurring of spatial boundaries between distinct activities while effectively removing anomalous positional spikes.

The UWB tags were mounted on the participants' hardhats. To estimate the vertical position relative to the walking surface Z_{foot}, the participant's standing height ($H_{participant}$) was subtracted from the sensor's Z-coordinate (Z_{tag}), as formulated in Equation (1):

$$Z_{foot} \approx Z_{tag} - H_{participant} \quad (1)$$

The core classification logic involved checking whether the coordinates (X, Y, Z_foot) fell within predefined 3D bounding boxes representing the walking zones. A significant challenge was the inherent noise in the UWB vertical measurements. Statistical analysis indicated a standard deviation of approximately sigma ≈ 0.6 m in the Z-axis for ground-level activities. To address this issue in a systematic manner and avoid arbitrary parameter selection, an empirical sensitivity analysis was conducted to determine appropriate spatial tolerances (buffers). As illustrated in Figure 3, multiple scenarios were evaluated by testing three horizontal (X-Y) to vertical (Z) expansion ratios (1:2, 1:3, and 1:4), systematically increasing the horizontal tolerance in 0.05 m increments. For each configuration, Macro F1-score and overall accuracy were computed to identify the most suitable noise mitigation threshold. The 1:3 expansion ratio demonstrated the most robust performance, establishing a clear "elbow point" at the 0.4 m horizontal mark where the Macro F1-Score plateaued. This suggests a reasonable balance between noise reduction and spatial precision, leading to the selection of final tolerances of 0.4 m horizontally and 1.2 m vertically.

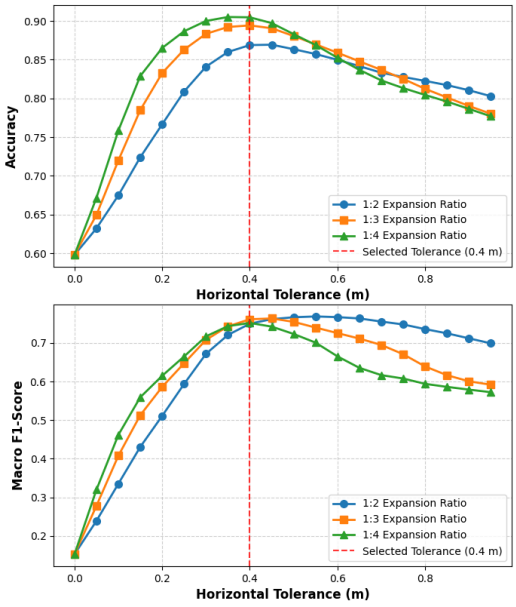


Figure 3: Empirical sensitivity analysis illustrating the effect of horizontal tolerance increments on Accuracy (top) and Macro F1-Score (bottom) across three spatial expansion ratios.

It is noted that these tolerance values are specific to the sensor noise characteristics and the spatial layout of this specific construction site. The point-based spatial predictions were aggregated to match the IMU processing window. The UWB data were segmented into 2.0-second

windows with a 50% overlap, and the final label for each window was determined via majority voting of the rule-based predictions within that window.

Results

In this section, the classification performance of the proposed systems is presented. First, we analyze the impact of sensor placement (Lumbar vs. Pocket) on activity recognition. Second, we compare the best-performing inertial configuration against the UWB rule-based system.

Table 1: Performance comparison of IMU sensor placement (Lumbar vs. Pocket) (N = 4)

Placement	Accuracy	Precision	Recall	F1-Score
Lumbar	**79.57%**	**0.327**	**0.358**	**0.340**
Pocket	75.20%	0.301	0.331	0.314

Table 2: Comparative performance of the Lumbar IMU and UWB systems (N = 5)

System	Accuracy	Precision	Recall	F1-Score
IMU	81.56%	0.443	0.376	0.372
UWB	**92.28%**	**0.713**	**0.808**	**0.732**

In terms of determining the optimal IMU placement, we compared the lumbar region and right front pocket positions. Due to data loss for one participant in the pocket dataset, this specific comparison was restricted to a subset of participants (N =4) to ensure a valid direct comparison. Table 1 summarizes the performance metrics. The Lumbar placement consistently outperformed the Pocket placement across all metrics.

We next compared the better-performing IMU configuration (Lumbar) against the UWB rule-based spatial classifier. Since complete lumbar IMU data were available for all participants, this comparison was conducted using the full dataset (N = 5). The results, presented in Table 2, demonstrate a significant performance divergence between the two approaches.

Although the IMU system has a relatively high overall accuracy (81.56%), its F1-score has a low value (0.372). The confusion matrices (Figure 4) provide insight into this discrepancy. The IMU-based system failed to capture spatial context; samples belonging to Elevated Walk (e_walk) and Scaffolding Walk (s_walk) were predominantly misclassified as Normal Walk (n_walk).

On the other hand, the UWB-based system achieved a dramatically higher F1-score of 0.732. It correctly identified 99% of the (e_walk) samples. The UWB-based system successfully distinguishes the mezzanine level from the ground floor.

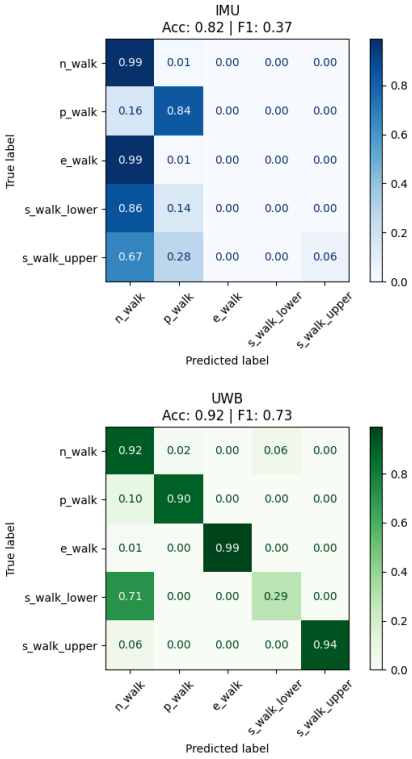


Figure 4: Confusion matrices for IMU(top) and UWB(bottom).

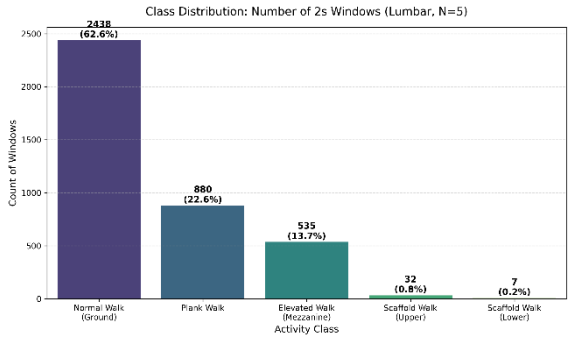


Figure 5: Class Distribution

Discussion

The primary objective of this study was to evaluate and compare UWB and IMU technologies for enhancing construction safety, specifically in monitoring worker activities across varying elevation levels. By analyzing sensor performance in controlled experiment conditions, this research sought to identify the operational limitations and complementary strengths of each system. Within this context, the results highlight significant trade-offs between the two systems.

One of the most essential outcomes of this study is the sensor placement's impact on the result. The lumbar placement outperformed the pocket placement which is consistent with the literature. Teng et al. (2024) highlighted the pelvis region as the single optimal sensor location due to its proximity to the body's center of mass. In their analysis of 12 body locations, the pelvis region consistently outperformed the upper thigh.

The observed discrepancy exists in the IMU results between high Accuracy (81.6%) and low F1-Score (0.37) can be attributed to class imbalance. As Figure 5 shows, Normal Walking (n_walk) constitutes the vast majority of the data windows. Consequently, the IMU classifier frequently predicts Normal Walking, inflating accuracy while failing to detect minority classes, as reflected by the low F1-score. This output aligns with the findings by Zimbelman & Keefe (2021) who observed that forestry workers spent 93% of their time in a 'Clear' (idle) state, which similarly degraded the model's ability to detect rare activities.

The UWB system successfully resolved the height awareness opposed to the IMU, achieving a Recall of 0.808. However, it reveals a specific weakness in differentiating between scaffold levels. While the system perfectly separated the Mezzanine (e_walk) from the Ground (n_walk), it struggled to separate the Lower Scaffold (s_walk_lower) from the Upper Scaffold (s_walk_upper). This limitation is directly attached to the spatial rule used to mitigate sensor noise. In this experimental setup, the optimal vertical tolerance exceeded the height difference between scaffold levels, resulting in class overlap.

The most critical finding is the IMU's inability to distinguish between ground level normal walking and elevated walking. Despite achieving 81.56% accuracy, the IMU system exhibited nearly zero recall for elevated walking classes due to the absence of absolute spatial awareness. This is expected, as the kinematic gait patterns of walking on a concrete ground floor and a concrete mezzanine are largely indistinguishable from an inertial sensing perspective. This limitation aligns with findings by Naheem and Kim (2022), who emphasized that while IMUs are effective for relative step detection, they require an external reference like UWB to resolve global positioning and correct for drift.

The analysis demonstrates the distinct operational boundaries of each system. IMU excels at detecting activity dynamics such as distinguishing the balance corrections in (p_walk) vs. (n_walk) but is blind to location. UWB excels at detecting general zones (Ground vs. Mezzanine) but is limited by the trade-off between noise mitigation and spatial resolution.

Practical Implementations and Real-World Limitations

While the controlled laboratory environment provided a necessary baseline for isolating sensor capabilities, deploying these technologies on active construction sites introduces practical complexities.

The current study utilized a localized network of nine UWB anchor nodes to ensure full coverage of a single laboratory space. Scaling this infrastructure to a continuously evolving, multi-story construction site presents significant logistical challenges. Maintaining precise georeferencing for dozens of UWB anchors across an active site requires constant recalibration and high upfront costs. In contrast, the IMU approach is highly scalable, as it leverages standard smartphones already carried by site personnel, requiring zero additional infrastructure.

The UWB system achieved a high spatial accuracy (92.3%) in a controlled setting. However, real-world construction sites are characterized by dynamic clutter such as moving heavy machinery, dense worker interactions, and temporary material stockpiles. The dynamic clutter induces Non-Line-of-Sight (NLOS) conditions. Furthermore, as previously established in literature (Shahi et al., 2012), a critical operational challenge for UWB is multipath interference caused by metallic obstructions. Reinforced concrete and steel scaffolding are common in vertical construction hence these dynamic obstructions and reflections are expected to degrade the baseline accuracy established in this study.

To address these real-world interferences, future deployments should monitor UWB data quality in real-time. When the positioning signal is stable and physically consistent, UWB spatial tracking remains the primary source. However, when the signal exhibits physically implausible movements, the system should dynamically shift reliance to IMU channel.

Conclusion

This study presented a comparative analysis of UWB-based localization and smartphone IMU-based gait analysis in terms of construction safety monitoring. The experimental outputs highlight a distinct trade-off between activity recognition and spatial context. While the IMU-based system effectively detected kinematic changes such as balance differences on narrow planks, its sensitivity to location was worse. The IMU-based analysis failed to distinguish between normal ground-level walking and elevated walking. Conversely, the UWB system provided critical spatial awareness, successfully identifying elevated zones like mezzanines. However, it performed worse in identifying walking on scaffolding due to noise buffering limits.

Rather than framing these limitations as shortcomings of the technologies, this study aims to better understand their practical operating conditions in complex, multi-level construction environments. The findings highlight key limitations, such as the spatial ambiguities associated with IMU-based tracking and the vertical noise observed in UWB measurements. By documenting these behaviors, the study provides a useful empirical reference for future work. Building on these insights, future research can explore more practical sensor fusion approaches that combine the complementary strengths of IMU-based activity recognition and UWB-based positioning, with the goal of improving reliability in real-world construction settings.

References

Akhavian, R., & Behzadan, A. H. (2016). Smartphone-based construction workers' activity recognition and classification. Automation in Construction, 71, 198-209.

Eurostat. (2025). Accidents at work - statistics by economic activity. Eurostat Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_-_statistics_by_economic_activity

Kılıç, U., Çomu, S., & Daniş, F. S. (2025, July). Evaluation of Different UWB Tag Placements on Construction Workers for Enhancing Safety. In EC3 Conference 2025 (Vol. 6, pp. 0-0). European Council on Computing in Construction.

Leitch, S. G., Ahmed, Q. Z., Abbas, W. B., Hafeez, M., Lazaridis, P. I., Sureephong, P., & Alade, T. (2023). On indoor localization using wifi, ble, uwb, and imu technologies. Sensors, 23(20), 8598.

Li, Z., & Li, D. (2025). UWB and IMU Fusion Construction Worker Localization Method Based on Transformer Correction. Journal of Computing in Civil Engineering, 39(4), 04025034.

Maalek, R., & Sadeghpour, F. (2013). Accuracy assessment of Ultra-Wide Band technology in tracking static resources in indoor construction scenarios. Automation in Construction, 30, 170-183.

Naheem, K., & Kim, M. S. (2022). A low-cost foot-placed UWB and IMU fusion-based indoor pedestrian tracking system for IoT applications. Sensors, 22(21), 8160.

Rao, A. S., Radanovic, M., Liu, Y., Hu, S., Fang, Y., Khoshelham, K., ... & Ngo, T. (2022). Real-time monitoring of construction sites: Sensors, methods, and applications. Automation in construction, 136, 104099.

Shahi, A., Aryan, A., West, J. S., Haas, C. T., & Haas, R. C. (2012). Deterioration of UWB positioning during construction. Automation in Construction, 24, 72-80.

Soltanmohammadlou, N., Sadeghi, S., Hon, C. K., & Mokhtarpour-Khanghah, F. (2019). Real-time locating systems and safety in construction sites: A literature review. Safety science, 117, 229-242.

Teizer, J., Venugopal, M., & Walia, A. (2008). Ultrawideband for automated real-time three-dimensional location sensing for workforce, equipment, and material positioning and tracking. Transportation Research Record, 2081(1), 56-64.

Teng, S., Kim, J. Y., Jeon, S., Gil, H. W., Lyu, J., Chung, E. H., ... & Nam, Y. (2024). Analyzing Optimal Wearable Motion Sensor Placement for Accurate Classification of Fall Directions. Sensors, 24(19), 6432.

Trigo, C., Della Vecchia, P., Crenna, F., Torres-Pardo, A., Gómez, J. A., Ratto, G., ... & Torricelli, D. (2025). Benchmarking of IMU-Based Gait Event Detection Algorithms Across Diverse Terrain Conditions. IEEE Transactions on Neural Systems and Rehabilitation Engineering.

Zhang, M., Chen, S., Zhao, X., & Yang, Z. (2018). Research on construction workers' activity recognition based on smartphone. Sensors, 18(8), 2667.

Zimbelman, E. G., & Keefe, R. F. (2021). Development and validation of smartwatch-based activity recognition models for rigging crew workers on cable logging operations. Plos one, 16(5), e0250624.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

OPENBIM-BASED GREEN PUBLIC PROCUREMENT CODE CHECKING: RULE MAPPING, INFORMATION DELIVERY SPECIFICATIONS AND GHERKIN TESTS

Carlo Zanchetta[1], Valeria Galleran[2], Thomas Krijnen[3] and Carlo Pavan[4]
[1]Università degli Studi di Padova, Italy
[2]Building Engineer, Rovigo, Italy
[3]AECGeeks, Eindhoven, The Netherlands
[4]120 grammi, Treviso, Italy

Abstract

This paper investigates openBIM-based Green Public Procurement (GPP) and Environmental Product Declaration (EPD) code checking and the suitability of Industry Foundation Classes (IFC) for supporting such checks. The coverage of Italian national requirements is assessed by concepts in the IFC schema and validation procedures such as Information Delivery Specification (IDS). The work was carried out under a research project funded by Northern Italy Venetian Regional Authority (CAM.bio project). For requirements exceeding IDS capabilities: graph-based checks, calculations, and facets outside IDS scope, a Gherkin-style validation combines human-readable scenarios with imperative code. Validation on IFC models is presented, alongside reflection and conclusions.

Introduction

The problem of implementing a digital strategy to monitor EPD and GPP processes is highly discussed both in relation to openBIM approaches (Aragon et al, 2025) standardization of rule checks (Tomczak, 2022) and alignment with Digital Product Passport (DPP) requirements. (Hamidkhaniha et al., 2025)

Several approaches have been developed to trace EPD and GPP requirements in a digital environment highlighting the necessity to adopt rule checking procedures. The majority of the approaches depends on naming standards and/or the definition of proprietary attributes. This causes the impossibility of standardizing those procedures and overpass national or local approaches. Due to the DPP directive this standardization and national codes alignment is necessary to fully support an EU policy for EPD monitoring, neutral and independent digitalization, computational code checking. (Honic et al., 2024). Based on this background, the study focused two research questions:

- RQ1. Is IFC a standard robust enough to support regulatory GPP requirements validation procedures?
- RQ2. Can IFC-based regulatory verification procedures be fully translated into standardized and vendor-neutral procedures such as IDS?

State of the art

The Industry Foundation Classes (IFC) are the predominant open standard to exchange Building Information by means of digital models (BIM). IFC is often regarded as a comprehensive and rich ontology, but in terms of producing data, the standard has limitations when highly specific information is required and proprietary authoring tools adopt specific dialects. At the same time, individual modeler preferences or the intent behind a data exchange has pronounced consequences on the way data is structured and exported. For example, Noardo et. al (2019) documents such tool limitations in the context of geospatial engineering. Gerbino et al. (2021) specifically looks at some of the difficulties in retaining material data across exports.

Limitations on information level available in IFC standard affect the code checking procedures that can be activated. There is a variety of standards and approaches to define information requirements on BIM models. Tomczak et al. (2022) identifies Spreadsheet, Product Data Templates, Data Dictionaries, IDS, mvdXML, idmXML, Level of information need, IFC Property templates and Linked Data + SHACL. These approaches follow a *declarative* approach, the requirements are stated as the intent that needs to be checked. This as opposed to an imperative approach that states the way how to perform the check. A somewhat hybrid form is Gherkin, which combines a natural language textual scenario description coupled with granular implementation steps in an imperative programming language. Gherkin for IFC has been described first in Moult and Krijnen (2020). It has since then been actively employed in the buildingSMART International (bSI) validation service[1].

When removing approaches without software support, approaches in which the semantics are not computer interpretable or that are more focused on the shape of the data, the approaches include IDS, mvdXML, SHACL and Gherkin. Jaud et al. (2025) specifically highlights the

[1] https://validate.buildingsmart.org

limitations of IDS compared to mvdXML, in particular with respect their precision: where mvdXML enables to stipulate the exact graph pattern to require, IDS is more focused on the interpretation of that pattern in end-user facing semantics and does not differentiate for example whether a property is defined on the occurrence or on the type-level. On the other hand, SHACL, described for its relevance to the AEC domain in Oraskari et al. (2021), requires moving the data to the language of the semantic web (RDF; Resource Description Format) which comes with implications in terms of increased dataset size and lacks a rich ecosystem to perform geometrical analysis (Pauwels et al. 2017). mvdXML has considerably less implementation support than IDS, higher complexity and no version-independence due to its focus on and tight coupling with IFC graph patterns.

The necessity for robust data validation is foundational in particular in the sustainability domain because of its transition toward an automated, BIM-based decision-support ecosystem, driven by increasingly stringent Green Public Procurement (GPP) requirements. Current research emphasizes the integration of Artificial Intelligence (AI) and Life Cycle Assessment (LCA) within the BIM environment to streamline the complex trade-offs required for regulatory compliance (Li et al., 2024). This paradigm shift is most critical during the Early Design Stage (EDS), where model-driven material selection dictates the building's ultimate environmental load and its alignment with mandatory sustainability criteria (Kamari et al., 2022). Such aspects state the necessity to rely on standardized datasets.

Methodology

To support the necessity to adopt standardized and openBIM code checking procedures related to EPD and GPP requirements, a study on the alignment with EPD information requirements was developed highlighting the percentage of attributes that are missing in the current IFC data model. To solve the problem of standardizing GPP/EPD information lacks in the IFC schema buildingSMART Data Dictionary (bSDD) technology was used to pivot on neutral attributes on an independent but standardized extension of the data model. The second part of the study was then dedicated to evaluate IDS robustness in supporting IFC based code checking procedures in the sustainability domain. Using standardized information requirements, an IDS-based code checking procedure was tested. Similarly to what was experimented in the first phase also in this phase the methodology was developed to assess the maturity of the standard by determining the actual percentage of code checking procedures that are supported by IDS .

An operational answer to these limits was developed and tested to demonstrate the applicability of the approach in a Common Data Environment (CDE) developed in alignment to those standards.

The analysis of the percentage of requirements and of procedures covered by bSI standards is then presented and commented on as a result of the experimentation and of the level of robustness of the schema.

The rationale behind the chosen methods is to rely on a procedure that includes at least 67% of EPD domain definitions that can be translated into IFC protocols. Furthermore, it must be demonstrated that the missing information can be integrated using only standardized protocols. Similarly, an information technology that allows for the integration of model compliance verification procedures using only open and neutral protocols is considered as promising. To demonstrate this assumption, in the third phase of the study integration scenarios, to properly support the openBIM code checking procedure, were tested.

Experimentation

First phase: specification

This phase defined the interoperability framework for linking CAM environmental requirements to IFC in a structured and verifiable way. The study focused only on mandatory requirements relevant to the design and construction phases, excluding non-binding provisions related to candidate qualification and procurement. Each criterion was systematically indexed within a spreadsheet, and its requirements were summarized highlighting qualitative and quantitative contents.

Subsequently, each item was evaluated and categorized into one of four levels of 'IFC Compatibility: fully IFC compliant, partially IFC compliant, potentially IFC compliant (partially or fully but characterized by high implementation complexity or requirements related to subjectivity of modeling), and non-IFC compliant (qualitative requirements or elements unrelated to information modeling). This assessment was based on three key factors: the formalization of CAM specifications into information requirements covered by IFC; the effective optimization of data entry and validation processes within current BIM applications; the operational feasibility for the stakeholders involved.

To ensure the practical feasibility and reliability of the proposed framework, the study focused exclusively on '*Fully*' and '*Partially*' compliant criteria. Consequently, both 'Non-IFC Compliant' items and those labeled as '*Potentially Compliant*'—due to their excessive technical barriers and operational complexity—were excluded from the research scope.

Development of a specific IFC Mapping and bSDD

The next step was to determine how the selected criteria could be mapped within the IFC schema. IFC4 ADD2 TC1 (version 4.0.2.1) was adopted as the reference schema, balancing standard maturity with compatibility with current industry tools. The mapping analysis process also aimed to:

- identify specific criteria requiring the development of new Classes or Property Sets (Psets) to address information which were not currently included in the IFC schema;
- identify possible difficulties in data writing procedures for operators with BIM Authoring applications (data assignments to abstract classes such

as groups or systems, document associations to elements, data assignments to material layers).

The analysis revealed that, of the total IFC-compliant criteria developed, 8% of the criteria can be fully mapped according to the schema, while the remaining 92% can be mapped according to other openBIM standards, by defining custom Psets and by developing a bSDD classification dictionary.
It also emerged that more than 50% of the criteria were affected by potential operational issues such as assignment to IfcProject, -Site or -Building which may not exist as such in the native environment, usage of document associations and classifications on objects and materials for which tools have limited support, or decompositions such as material layers.

A significant challenge arises from the diverse nature of CAM criteria, which often stipulate specific requirements for construction materials, such as defined recycled content. These requirements encompass a wide array of numerical, textual, and boolean parameters. Furthermore, certain criteria demand properties not natively supported by the IFC schema, specialized quantity take-offs from filtered object selections, or the inclusion of technical documentation linked directly to model entities.

A technical mapping sheet was developed to define, for each criterion, the relevant IFC entities, materials, properties, and document associations. To support consistent data population and semantic interoperability, CAM information was also structured within a dedicated "CAM Classification" bSDD dictionary covering elements, materials, and properties for the purpose of consistency and adoption in tools that support bSDD.

Analysis of validation procedures of each criterion

Finally, the specification phase defined validation rules and tolls needed for each criterion. Validation was initially formalized using the Information Delivery Specification (IDS) format, but the analysis showed that several criteria exceeded current IDS capabilities:

Validation of document association: most of the criteria required a document association validation related to objects or material layers. However, IDS does not support the verification of external files linked via *IfcDocumentReference*;

- Disjunctive logic (OR conjunction);
- Conditional logic (IF-THEN) validation: where the validation of an object's property depends on conditions defined in a (separated or related) entity;
- Element-to-Material recursion: while IDS can verify that an element associated with "Material X" possesses specific properties, it is impossible to shift focus to the material and apply constraints to the material entity itself;
- Data aggregation and computational criteria: performing project-wide calculations, such as verifying that the cumulative weight of all recyclable materials exceeds 70% of the total project mass.

To overcome these limitations, the research implemented an IDS multi-applicability approach and custom scripts to parse the IFC model, extracting data from both entity and material levels to perform the required validation. The analysis concluded by assigning each criterion to one of three automated validation scenarios tailored to the specific complexity of the data to be verified: IDS-exclusive, script-exclusive, blended validation.

Second phase: validation

To develop the GPP-CAM validation phase, multiple IFC model prototypes were created within an authoring tool according to defined rules and mapping in technical sheet to serve as a baseline for subsequent operations as: test data writing and checking limitations, identify formal control strategies according to the effectiveness of IDS rules, define an overall validation workflow.

An operational guide was then developed to support practitioners and streamline the workflow by sharing a general know-how of IDS rules structure and convey multi-step procedure of validation, including best practices for correct model setup and data population to ensure the accurate generation of the CAM information database according to the CAM Classification bsDD.

Operational and software limitations

A preliminary audit of the informational content of the test models revealed significant deficiencies in data quality and completeness.

Recognizing that many errors stem from the inherent limitations of commercial BIM authoring tools rather than user oversight, the methodology incorporated a 'post-procedural editing' phase (authoring&integration). This ensures that models failing validation due to software constraints can be automatically refined and corrected before the final assessment. Consequently, it became necessary to develop dedicated implementation tools within both desktop applications and CDE.

Recurring limitations in commercial BIM authoring softwares that hinder the direct implementation of CAM requirements included:

- *material classification*: difficulty in assigning standardized classification codes directly to material layers;
- *parameter duplication*: issues with redundant project parameters during the IFC merging phase;
- *entity and group classification*: constraints in mapping classifications to complex element groups;
- *document association*: lack of streamlined workflows for linking external documentation to model objects;
- *generic connectivity*: limitations in creating generic connection relationships between elements;
- *schema conversion*: difficulties in converting IfcGroup entities into specific IFC schema subclasses.

Formal control strategies: Editing for validation

To address these software limitations, the project initially tested advanced IFC model manipulation within the open-source Blender environment. By leveraging the integration of the IfcOpenShell library and the graphical

interface provided by Bonsai add-on, the following operations were successfully implemented:

- *document association*: deployment of IfcDocumentReference entities and their corresponding IfcRelAssociatesDocument relationships to ensure formal linkages between external sources (e.g., certifications, technical data sheets) and individual model entities.
- *database structuring*: refinement of the model's internal data structure to ensure full compliance with the IFC schema, thereby correcting inconsistencies generated during export from proprietary software;
- *classification mapping*: assignment of one or more classification codes to model elements using IfcClassificationReference entities and IfcRelAssociatesClassification relationships, facilitating interoperability with external classification systems such as the bSDD.

However, the operational complexity of manual editing highlighted a need for more efficient, less error-prone methods to support large-scale implementation. Consequently, the project shifted toward automated IFC-editing solutions through the development of dedicated 'IfcPatch' recipes in Python. These scripts allow for the automated post-processing and refinement, ensuring a standardized data structure regardless of the constraints imposed by the original authoring tool.

IDS Rule Architecture and Hierarchical Validation

Given the limitations of the IDS format, IDS rules were organized according to a predefined schema where each CAM criterion is decomposed into one or more conditions applicable to the IFC data. For every criterion verifiable via IDS, a multi-specification rule applicability approach was developed to determine compliance. These specifications are structured into three hierarchical levels to enable a step-by-step verification process:

- Level 0 - Applicability: These specifications define the relevance of a CAM rule to the analyzed model. If these conditions are not met, the criterion is either inapplicable to the project or essential preliminary information is absent.
- Level 1 - Entity Classification: These rules ensure that the targeted IFC entities are correctly classified according to the bSDD dictionary. Failure at this level indicates that entities are misidentified, thereby precluding further requirement validation.
- Level 2 - CAM Criterion Validation: These specifications verify whether the IFC model fulfills the actual environmental requirements of the CAM criterion. Non-compliance at this stage necessitates data integration or design adjustments.

This tiered structure facilitates a granular diagnostic process, identifying whether a failure stems from a lack of relevance (Level 0), a classification error (Level 1), or a direct violation of criteria requirements (Level 2).

Orchestrating IDS for conditional logic

The Multiple Applicability Approach was also implemented to overcome the structural constraints of the IDS standard, specifically the lack of support for dependencies or conditional IF-THEN logic. Instead of a single monolithic check, CAM requirements are decomposed into sequential, interdependent IDS files as:

- Triggering Specification: The first IDS file verifies the presence and attributes of the "source" element (e.g., IF an IfcWall is classified as 'LoadBearing').
- Targeted Specification: If the first check is satisfied, a second specification validates the required conditions on the related element (e.g., THEN the associated IfcMaterial must have a specific recycled content property).

Under this approach, CAM Compliance of IF-THEN logic is formally achieved through two valid states:

- Full Compliance: Both specifications are successfully validated, confirming the condition was triggered and the requirement met.
- Non-Applicability: The first specification is not satisfied, indicating the triggering condition is absent (e.g., no LoadBearing walls exist), and thus the requirement is not invoked.

Table 1: Synthesis of validation outcomes

Result level	Status	Meaning
Level 0 Fail	Not applicable	The rule does not apply to the model
Level 1 Fail	Classification Error	Entities are found but not correctly mapped to bsDD
Level 2 Fail	Non-Compliant	Data is present but does not meet CAM thresholds
All Level Pass	Compliant	The model satisfies the CAM criterion

Definition of an Automated Validation Framework

Finally, each criterion was assigned to one of three operational validation scenarios based on its technical complexity:

- Scenario 1: Hybrid Validation (IfcPatch + IDS + Script): The most robust approach. It combines model normalization, standard attribute checking, and complex computational logic for maximum accuracy and traceability.
- Scenario 2: Standard-Centric (IfcPatch + IDS): Suitable for criteria fully expressible through the IDS schema. It ensures high compatibility with international tools but is limited to non-computational requirements.
- Scenario 3: Script-Driven (IfcPatch + Script): Utilized for performance-based criteria that fall outside the IDS scope. It relies on direct analysis of the IFC

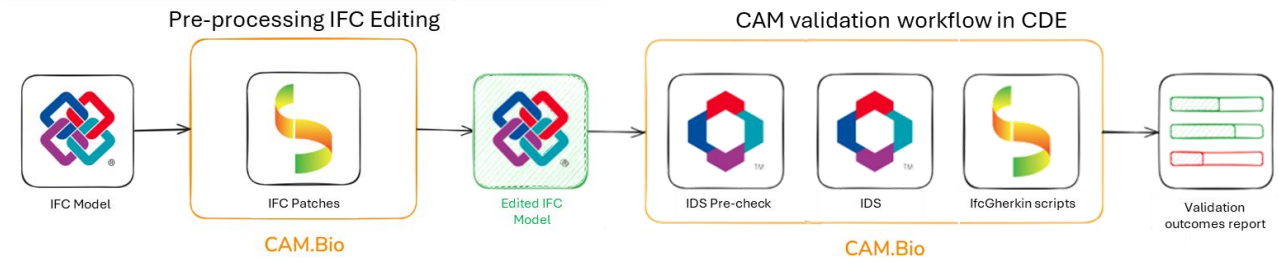


Figure 1: Schematic overview of the contributions of the CAM.bio project wrt. the validation workflow

structure to validate complex environmental constraints and relationships.

This analysis enabled the definition of a comprehensive validation framework that encompasses all potential scenarios, allowing it to adapt seamlessly to each specific criteria by integrating standard protocols with custom computational logic.

- **Phase 1: Pre-processing (IfcPatch):** Uses the IfcOpenShell library to normalize data, correct structural errors, and inject missing relationships or attributes into the IFC model.
- **Phase 2: IDS Pre-check:** A custom module verifies the presence of mandatory information required for formal validation, preventing execution errors.
- **Phase 3: Standard IDS Validation:** Executes .ids files to verify criteria that can be formally expressed via the standard buildingSMART XML schema.
- **Phase 4: Advanced Scripting (IfcGherkin/IfcOpenShell):** Python scripts perform complex verifications unreachable by IDS, such as Boolean logic on materials, cross-entity spatial checks, and geometric quantity aggregations**.**

The defined four-phase validation process can be made scalable and seamlessly integrated into the CDE platform by encapsulating the entire validation workflow (IfcPatch, IDS pre-check, IDS, and custom scripts) within a Docker microservice exposed via RESTful APIs.

openCDE Platform: Implementation and Visualization

To ensure CAM requirements are manageable by non-technical stakeholders and to demonstrate the feasibility of the proposed framework, the research evolved a CDE prototype. The Xeokit viewer was customized to expose CAM entities often hidden in traditional BIM viewers (e.g. TypeEnums). This included a reorganized tree hierarchy for IfcSpatialZone and IfcBuildingSystem, and direct visibility of material-level properties and external documentation (IfcDocumentReference) reflecting CAM Classification bSDD. The platform includes an IDS Authoring Environment for rule creation and a Validator UI for real-time compliance feedback. Third phase: integration and authoring

Given the limitations and dialects of IFC authoring tools, a process is in place to unify and harmonize the data ingested into the system. In the platform (named CAM.bio – the italian word for “change”) this is built around the idea of IfcPatch[2], which is an existing module in the IfcOpenShell family of tools that consists of ‘recipes’ to fix or alter IFC content and write out a revised version. These recipes are implemented in python using imperative code. In our particular case we built a minimal graph rewriting DSL to declaratively encode the intended transformations. In particular it was found that certain tools were unable to provide classification references to material definitions. As a work-around the modeller was instructed to write these as a material property using particular conventions that would be picked up by the IfcPatch recipe. See Figure 2 for part of the implementation, Figure 3 and Figure 4 for the before and after as a minimal graph rendering.

Model validation

As introduced earlier in the text there has been extensive research on validating information requirements. Many of these can also be used to validate design requirements such as CAM, because they use a similar vocabulary. For this project, several have been considered.

IDS is used in the project, it is limited in scope and expressiveness, but well-tested and widely implemented. An additional approach needs to be selected that is richer

```
select()\
  .where(type='IfcMaterialProperties')\
  .where(Name='ClassificazioneCAM')\
  .select('Material', 'Material').up()\
  .select('Properties')\
  .where(length=1)\
  .select(0)\
  .where(Name='ClassificationCode')\
  .select('NominalValue')\
  .select(0, 'Value')\
  .remove()\
  .construct({'IfcExternalReferenceRelationship': {
    'RelatedResourceObjects': ['{Material}'],
    'RelatingReference': {
      'IfcClassificationReference': {
        'Identification': '{Value}',
        'Description': lambda di:
            bsdd.code_to_description[di['Value']],
        'ReferencedSource': {
          'IfcClassification': {
            'Name': 'CAMClassification'
          }
        }
      }
    }
}})
```

Figure 2: Code produced to rewrite material classifications

[2] https://docs.ifcopenshell.org/ifcpatch.html

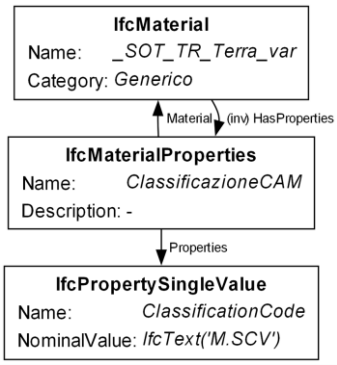


Figure 3: 'Before' situation with classification as property

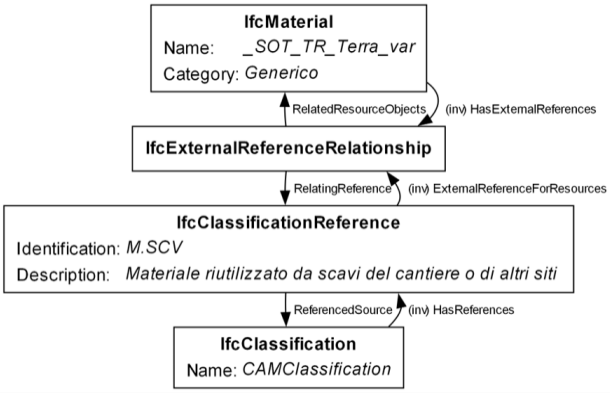


Figure 4: 'After' situation when rewritten

in expressiveness to support implementing the required constraints.

That is also not surprising given that IDS is intended to control the availability of information on top of the modelling mechanisms that IFC and standard model view definitions already propose. mvdXML has the possibility to precisely stipulate graph patterns, but is still lacking many of the constructs needed for the ambitions set out in this paper such as aggregate functions to sum areas together and the ability to assess and compare multiple focus nodes in one scenario.

Naturally, the option exists to implement rules in proprietary software, but this lacks the portability for moving such checks to a web-based platform and is not vendor neutral. Another option is to not use a pre-existing approach at all, and to simply use an imperative programming language to implement the required functionality, which is flexible but comes at a prohibitive cost in terms of shared understanding between domain stakeholders.

The approach selected in the end is Gherkin, which consists of a high level human readable set of scenarios (an example for this project is given in Figure 5) that describe test logic for shared understanding between stakeholders, which are paired with an implementation in (in this case) the Python programming language for the actual execution. As such it can be seen as a middle ground between expressiveness, portability and ease of understanding.

In Gherkin scenarios are broken up into steps, these steps are matched to parameterized implementation functions, as such there obvious advantages: people do not need to understand program code, but can follow the high-level specification; as such there is a shared understanding of the requirements between software engineers as well as other stakeholders; the textual breakdown into statements offers modular reuse of functionality. In order to best understand the technology it's best to look at the code repository that contains the body of rules developed for the validation service[3]. In Figure 6 one sees a table of the functional steps employed in the validation service. This table highlights the relative strength of the Gherkin approach for IFC model verification: Gherkin allows a similar structure of graph pattern building using entity and attribute selection (not too dissimilar from EntityRule and

```
00  @E00160
01  Feature: Rule 2.3.3.a - Riduzione effetto isola calore estiva e inquinamento atmosferico
02
03    Scenario: New construction projects and Urban renovation projects must have: green areas
04                >= 60% permeable land surface
05
06      Given an .IfcGeographicElement/TERRAIN. or .IfcPavement. or .IfcSlab/PAVING. Or …
07
08      # Filter based on classification
09      Given a relationship .IfcRelAssociatesClassification. to .the above. from …
10      Given .Identification. ^is^ 'E.TER' or 'E.SSN' or 'E.SPA' or 'E.SSI' or 'E.SCA' …
11      # Move back up to the instances
12      Given the instances '3' steps up
13
14      Given the property value for 'NetArea'
15      Given the sum of all values at depth 0 [stored as 'GreenAreas']
16
17      Given The result for 'SuperficiePermeabile' from 'Rule_232' [stored as
                     'SuperficiePermeabile']
18
```

Figure 5: Gherkin scenario text implemented for the CAM.bio project for model validation. Given statements correspond to the selection of data from the model (applicability); Then statements perform validation (requirement)

[3] https://github.com/buildingsmart/ifc-gherkin-rules

Predicate / operation	Usage
An .{entity_opt_stmt}.	233
.{attribute}. ^{comparison_operator:equal_or_not_equal}^ {value}	192
Its attribute .{attribute}.	166
A {file_or_model} with {field} '{values}'	80
There must be {constraint} {num:d} instance(s) of .{entity}. …	58
A relationship .{relationship}. {dir1:from_to} .{entity}. {dir2:from_to} …	27
…	…
It must have no arc segments that use colinear points after taking …	1
All edges must form a single connected component	
The boundaries of the face must conform to the implicit plane fitted through …	

Figure 6: Usage count of Gherkin implementation function highlighting very common mvdXML-like traversal operations and specialistic single-use steps for specific imperative validations

AttributeRule in MvdXML) and at the same time using highly specialistic predicates that also involve the wide availability of python modules to build rich behavior for the breadth of disciplines represented in the IFC schema and complexity of downstream verification use-cases.

An example of such a scenario that has been built for the CAM.bio project is depicted in Figure 5. The first lines (line 0-5) are descriptive. In line 8 an initial selection of entity instances is made from the model. Their classification is looked up and filtered. For the elements with matching classifications we sum their areas. This is then compared to the 'SuperficiePermeabile' which is another rule implemented in similar fashion.

Discussion and results

Regarding BIM-compliance analysis, the structural limitations identified within the schema are as follows:

- 44% of the criteria cannot be fully implemented due to the absence of specific entities for environmental and infrastructural contexts (e.g., river ecosystems, groundwater). These entities were introduced in the IFC 4.3 version, rendering the current version insufficient for direct ontological mapping. Future research is to be directed at this.
- The current breakdown of IfcMaterial does not allow for an analytical definition of the internal composition of mixtures or composite materials, complicating the verification of chemical-physical performance criteria.

Regarding requirements that are potentially or partially IFC-compliant but were excluded due to implementation difficulties, the following are highlighted:

Regulatory requirements involving subjective or qualitative evaluations (e.g., prevention, mitigation, analysis of existing greenery) that cannot be translated into standardized parametric properties.

- Compliance with external regulations outside the scope of this study; simplifications were adopted, transitioning to manual verification.
- Comparisons between different temporal phases, which are complex to develop and risk imposing workflows that are disproportionately burdensome for the operator relative to the benefits of the verification. This type of analysis could be further explored in future research.
- Criteria requiring the identification of "technical strategies" that conceptually lack a unique digital identity. In these cases, simplification was preferred to avoid the definition of overly extensive custom IfcPropertySet that would be difficult to standardize.

The analysis revealed that, of the total criteria defined by the CAM regulation, more than 76% are IFC compliant. Considering the research screening for determining the project's scope, based on the exclusion of non-mandatory and difficult-to-implement criteria, it emerges that IFC-compliant CAM criteria represent a total of 56% of the total CAM regulation criteria (46% fully IFC Compliant, 10% Partially IFC Compliant, 21% Potentially IFC Compliant, and 23% Non-IFC Compliant.

This study excluded several criteria from its scope, specifically those pertaining to urban and territorial planning. Such criteria necessitate the calculation of spatial proximities between the project site and local public services, a requirement that can be effectively addressed through enhanced interoperability between IFC and GIS models.

Concerning the IFC-mapping analysis, the study identified that out of the 816 entities comprising the IFC4x2 schema, a subset of 39 entities and their respective TypeEnum is sufficient to fulfill the CAM regulatory requirements. These requirements primarily map to the following categories: IfcMaterial, IfcDocumentReference, IfcSpatialElement, IfcSystem, and IfcBuildingElement sub-classes, including their relative predefined types.

The primary challenge encountered during this mapping phase was the necessity for granular discretization of entities based on their ontological and functional characteristics (e.g., IfcMaterial naming conventions, IfcSystem, and IfcBuildingElement) or according to specific document contents (e.g., IfcDocumentReference). This granularity is essential for executing automated filtering operations and rigorous data validation. Additionally, the definition of custom Property Sets proved critical, especially for capturing data related to recycled, recovered, and by-product content within products and materials.

In this context, the development of a dedicated bSDD was fundamental. To define the requirements dictated by the CAM criteria, the dictionary incorporates 144 class codes, categorized as follows (25 codes for the discretization of document types within the IfcDocumentReference entity; 43 for the characterization of building element features; 8 mapping to various IfcSpatialElement sub-classes; 54 for the classification of diverse IfcMaterial types; 14 defining IfcSystem sub-classes)

The project ultimately identified 104 distinct properties. While 13 of these attributes are natively available within the IFC schema, the remaining 91 were specifically introduced via the cam-bSDD.

Concerning Validation, the analysis of the 38 CAM criteria revealed a high dependency on model normalization and multi-layered validation framework. Specifically, 36 criteria (95%) required corrective support via IfcPatch to ensure data consistency. More than 87% of CAM criteria needed Gherkin as a fallback due to technical requirements and entity comparisons not supported by IDS format only.

Table 2: CAM validation procedure breakdown

Validation Component	Criteria Count (out of 38)	% of Total
IfcPatch Support (Pre-processing)	36	95%
IDS Pre-check	21	55%
IDS-only Validation	5	13%
Script-only Validation	14	37%
Hybrid Validation (IDS + Script)	19	50%

Conclusions

CAM.bio contents were fully aligned with the Italian regulatory framework in force at the time of completion (D.M. 23/06/2022 n° 256). As is often the case in building construction field, new legislation updating CAM criteria has taken effect from February 2026.

In this context, the project serves as a baseline reference to support Italian current implementation and the transition to the updated regulatory framework, and provides an example of structured bSDD dictionary development and as a reference for international researchers.

The study establishes limits of the current version of IFC and IDS in supporting computational and openBIM based code checking procedures based on the information requirements needed by the national implementation of EPD directive.

Acknowledgments

The study was developed in relation to a research project funded by Veneto regional authority (northern Italy area). Further details can be found at the following links:
https://cam-bio.it
https://identifier.buildingsmart.org/uri/cadlinesw/cambio/1.0

References

Aragón, A., Spudys, P., Pupeikis, D., Nieto, Ó., & Garcia Alberti, M. (2025). Bridging interoperability gaps between LCA and BIM: Analysis of limitations for the integration of EPD data in IFC. Buildings, 15(15), 2760.

Gerbino, S., Cieri, L., Rainieri, C., & Fabbrocino, G. (2021). On BIM Interoperability via the IFC Standard: An Assessment from the Structural Engineering and Design Viewpoint. Applied Sciences, 11(23), 11430. https://doi.org/10.3390/app112311430

Hamidkhaniha, S., Martinez, A. G., & Sedaghat, M. (2025, December). Integrating Digital Product Passports and BIM for Circular Construction: A Review of Challenges and Opportunities for Data-Driven Material Reuse and Sustainability Assessment. In IOP Conference Series: Earth and Environmental Science (Vol. 1568, No. 1, p. 012026). IOP Publishing.

Honic, M., Magalhães, P. M., & Van den Bosch, P. (2024). From data templates to material passports and digital product passports. A circular built environment in the digital age, 79.

Jaud, S., Marie, S., & Pinzenoehler, A. (2025). The Quest for the OpenBIM Exchange Requirements Checking Language: Comparing mvdXML, IDS and Gherkin. In EC3 Conference 2025 (Vol. 6, pp. 0-0). European Council on Computing in Construction.

Kamari, Aliakbar & Kotula, Bartlomiej & Schultz, Carl. (2022). A BIM-based LCA tool for sustainable building design during the early design stage. Smart and Sustainable Built Environment. 10.1108/SASBE-09-2021-0157.

Moult, Dion, and Thomas Krijnen. "Compliance checking on building models with the Gherkin language and Continuous Integration." EG-ICE 2020 Workshop on Intelligent Computing in Engineering, Proceedings. 2020.

Noardo, F., Arroyo Ohori, K., Biljecki, F., Krijnen, T., Ellul, C., Harrie, L., & Stoter, J. (2019, June). GeoBIM benchmark 2019: design and initial results. In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives (No. 2/W13, pp. 1339-1346). ISPRS.

Li, J., Liu, Z., Han, G., Demian, P., & Osmani, M. (2024). The Relationship Between Artificial Intelligence (AI) and Building Information Modeling (BIM) Technologies for Sustainable Building in the Context of Smart Cities. Sustainability, 16(24), 10848. https://doi.org/10.3390/su162410848

Oraskari, J., Senthilvel, M., & Beetz, J. (2021, October). SHACL is for LBD what mvdXML is for IFC. In Proc. of the Conference CIB W (Vol. 78, pp. 11-15).

Pauwels, P., Krijnen, T., Terkaj, W., & Beetz, J. (2017). Enhancing the ifcOWL ontology with an alternative representation for geometric data. Automation in Construction, 80, 77-94.

Tomczak, A., v Berlo, L., Krijnen, T., Borrmann, A., & Bolpagni, M. (2022, November). A review of methods to specify information requirements in digital construction projects. In IOP Conference Series: Earth and Environmental Science (Vol. 1101, No. 9, p. 092024). IOP Publishing.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

NOT ALL LINKS ARE EQUAL: LINK EXPRESSIVENESS AND DISCOURSE SCOPE IN BIM-GIS INTEGRATION

Judith Krischler[1], Sebastian Schilling[2], Maximilian Sternal[3], Jakob Taraben[1] and Christian Clemen[2]
[1]Bauhaus-Universität Weimar, Weimar, Germany
[2]Hochschule für Technik und Wirtschaft Dresden - University of Applied Sciences, Dresden, Germany
[3]Technische Universität Berlin, Berlin, Germany

Abstract

Linking is increasingly adopted in BIM-GIS integration to connect heterogeneous datasets without forcing full schema conversion. Application scenarios are the linked transfer e.g. of BIM and environmental data for permitting or the linking segmented 3D point clouds. However, links are often treated as purely technical connectors, while their interpretability remains implicit. We distinguish closed, shared, and open discourse space and analyze how common link paradigms support the ability to reconstruct sender intention. We also discuss practical trade-offs regarding implementation effort, tooling, and adoption. This paper is intended to help readers select appropriate linking mechanisms depending on BIM-GIS integration goals.

Introduction

Linking has become a central strategy in BIM-GIS integration, particularly where heterogeneous data structures, scales, and modeling assumptions make full schema unification impractical. A linked approach relies on establishing relationships between heterogeneous datasets instead of fully unifying their underlying data models or use naturally imperfect format conversions (Eriksson et al., 2020; Korkmaz and Basaraner, 2024). Instead of converting data between formats such as IFC and CityGML, many integration approaches rely on explicit or implicit links to establish relationships between models, elements, or representations (Liu et al., 2017; Guyo et al., 2021; Beck et al., 2020). If we assume a simple communication model involving a sender and a receiver, the sender creates both the application models and BIM/GIS links. After the data transfer, the receiver must identify the relevant parts of the application models and the intended meaning of the links. The receiving software might e.g. be a Common Data Environment, BIM-coordination software or a web-GIS.

Where the semantic intent behind a link cannot be recovered, e.g., which coordinate transformation was applied or what part-whole relationship was assumed, the receiving system either propagates a silent misinterpretation or must recompute the original analysis (if known) from scratch. The receiver cannot determine whether their interpretation corresponds to the one under which the link was created.

Semantic mismatches between established standards reflect not only technical modeling differences but fundamentally divergent domain perspectives: GIS adopts a global, landscape-oriented view prioritizing spatial relationships and regional context, while BIM employs a detailed, object-centric perspective focused on individual building components (Kolbe and Donaubauer, 2021; Tan et al., 2023; Sani et al., 2022).

An integration by linking both domains must therefore consider different semantic depths and requires flexible integration approaches. In this regard, Mohammed et al. (2026) propose multiple levels of BIM-GIS integration quality and suggest that higher integration levels rely on more expressive linking mechanisms, while also highlighting the need for clearer methodological guidance and standardization.

Despite their importance, links are often treated as technical implementation details or a matter of computational efficiency (Napps et al., 2026). As a result, assumptions about relationship semantics, scope of validity, and interpretability remain implicit, limiting transparency, reproducibility, and reuse of integration results. This is particularly problematic in BIM-GIS scenarios, where relationships may encode non-trivial analytical decisions - such as georeferencing strategies, spatial abstractions, or part-whole interpretations - that cannot be reconstructed from geometry or attributes alone.

Background and Research Questions

Previous work by Krischler et al. (2025) introduced and operationalized four link types - geospatial, geometrical, topological, and mereological - to support context-sensitive BIM-GIS integration within container-based exchange models, compliant to DIN 18290 (Multimodel-Container). The link types do not prescribe a specific technology or schema, but describe what kind of relationship is being expressed and which assumptions are made explicit. The Multimodel-schema was chosen here, as it is successfully implemented and used in (German) construction software for tendering, awarding contracts, and billing.

Geospatial links explicitly relate datasets through their geographic positioning by defining a common spatial reference or transformation parameters, enabling consistent co-registration across BIM and GIS models.
Geometrical links associate datasets via auxiliary (proxy) geometric representations, such as footprints or bounding volumes, to reference spatial subspaces where no shared

identifiers or semantic alignment exist. **Topological links** describe spatial relationships such as containment, intersection, or adjacency independently of exact geometry, allowing robust connections between entities of different dimensionality. **Mereological links** capture part-whole relationships between heterogeneous representations, enabling the association of datasets that describe overlapping or corresponding parts of the same real-world entity at different levels of detail. While this work demonstrated how expressive links can be formalized, it also revealed that identical link structures can have very different meanings depending on where and how they are interpreted.

This paper takes a step back and asks: why do link modeling approaches differ so fundamentally in their expressive power and interpretability? We argue that the key, yet underexplored factor, is the discourse space in which links operate. By making discourse scope explicit, we aim to clarify the relationship between link modeling paradigms, semantic expressiveness, and practical applicability.

The paper addresses the following research questions:

RQ1: How do commonly used link modeling approaches in BIM-GIS integration differ in their level of link expressiveness for geospatial, geometrical, topological, and mereological links?

RQ2: How does link expressiveness influence the ability to reconstruct sender intention and semantic context?

RQ3: What practical requirements and trade-offs are associated with realizing different levels of link expressiveness in BIM-GIS workflows?

Analytical Framework

This paper follows an analytical and comparative research approach. It first introduces discourse space as an analytical framework to distinguish between closed (local), shared, and open (global) semantic contexts in which links between BIM and GIS data are created and interpreted.

Building on this framework, commonly used link modeling approaches from research, standardization, and practice are analyzed comparatively with respect to their level of link expressiveness. Existing approaches are examined and positioned within the discourse space framework, allowing their underlying assumptions, expressive capabilities, and limitations to be discussed systematically.

The analysis further examines how different levels of link expressiveness affect the reconstruction of sender intention and semantic context, and concludes with a discussion on practical requirements and trade-offs for realizing expressive links in BIM-GIS workflows.

Discourse Spaces in BIM-GIS Integration

We define discourse space as the semantic scope within which the meaning of a link is interpretable without additional external knowledge. Prior work implicitly reflects different semantic depth of link interpretation (Bormann et al., 2021), ranging from schema-internal relationships (e.g. through Application Domain Extensions) to ontology-grounded semantic links (Beck, 2023; Hagedorn et al., 2023; Davar Panah, 2025; Schilling and Clemen, 2023). However, these scopes are rarely made explicit as design decision. Three discourse spaces are distinguished within this paper (Figure 1). The primary formal criterion for our "discourse spaces" is where (!) the link schema is located: in the application model (*closed*), in a standalone, use-case-specific link schema (*shared*), or in a formal, publicly accessible ontology in the sense of the Semantic Web (*open*), and whether the link is self-descriptive rather than just serving as a pointer. The W3 community describes "self-descriptive" data as a defined label, definition and comment within the representation itself without relying external documents for description (World Wide Web Consortium, 2014).

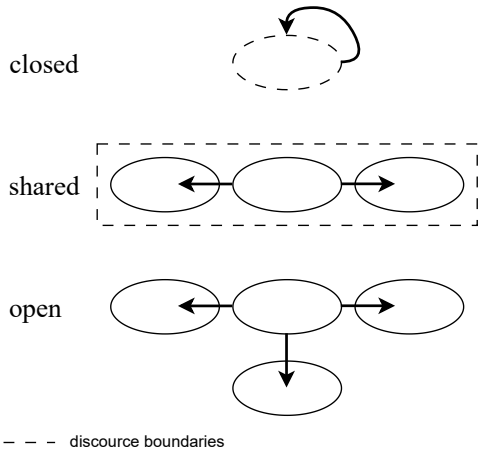


Figure 1: Link paradigms covering local, shared, and open contextual discourse.

Closed (Local) Discourse refers to links whose meaning is interpretable only within a single file, dataset, or schema. Semantics are implicit and depend on internal conventions, such as schema structure, attribute naming, or software-specific behavior. Interpretation requires knowledge of the originating data model, as no externalized semantic contract or self-descriptive semantic content exists within the link itself (e.g. a relational database schema).

Shared Discourse denotes a bounded semantic context established across multiple datasets within a project or integration setting (e.g. within information containers, where there is a semantic agreement on the linkage between datasets). Here, link meaning is self-descriptive and explicitly agreed upon through project conventions, container schemas, or link models, but remains valid only within that context.

Open (Global) Discourse describes links whose semantics are interpretable independently of a specific project and whose links are fully self-descriptive. Meaning is grounded in shared, publicly defined vocabularies or ontologies, typically using globally resolvable identifiers and interpretation does not rely on bilateral agreements (e.g. public ontologies following the FAIR-principles).

These discourse spaces are not mutually exclusive but rep-

Table 1: Categories of Link Expressiveness and Their Impact on Intention Reconstruction

Category	Low Expressiveness	High Expressiveness	Impact on RQ2
Direction	Non-directional	Explicitly directional	Clarifies information flow
Cardinality	Binary (1:1, 1:N)	N-ary, attributed	Preserves complex patterns
Architecture	Embedded	Reified with metadata	Enables rich context
Semantic Strength	Implicit/local	Standardized/global	Ensures consistent interpretation
Link Depth	Document-level	Attribute-level	Supports precise integration
ID Scope	Local	Global/persistent	Enables long-term validity

resent increasing degrees of semantic explicitness and external interpretability.

Link Expressiveness

In this paper, link expressiveness denotes the ability of a link representation to make relationship semantics explicit and machine-interpretable. This includes, but is not limited to:

- explicit relationship type and link directionality
- support for qualified and n-ary relations
- attachment of parameters and constraints
- explicit scope and validity context
- provenance and derivation metadata

Expressiveness is not inherently desirable in all cases as it usually introduces trade-offs in complexity, tooling and governance. The following sections analyze how expressiveness manifests across discourse spaces and link modeling paradigms.

Comparative Analysis of Link Expressiveness (RQ1)

This section presents a comparative analysis of commonly used link modeling approaches in BIM-GIS integration, examining how they differ in their ability to explicitly express geospatial, geometrical, topological, and mereological relationships. The analysis is structured along the previously introduced discourse spaces and positions existing approaches accordingly, thereby making their expressive capabilities and limitations directly comparable.

Closed Discourse: Schema-Inherent and Implicit Links

In closed (local) discourse, links are embedded within data models as references, foreign keys, or structural relationships. Common examples include relational database joins or GML XLinks. These mechanisms typically allow only minimal expressiveness: a link indicates that a relationship exists, but not what kind of relationship it is or why it was established.

Directionality, cardinality, and constraints are either implicit or fixed by the schema. For BIM-GIS integration, this limits expressiveness, particularly for geometrical and mereological (part-whole relationship) links, which often require additional parameters or contextual assumptions.

Local discourse links are easy to implement and widely supported, but they externalize semantic interpretation to documentation or human expertise. Figure 2 shows an exemplary relational database schema for buildings and parcels. The keys do not carry a self-explanatory semantic meaning. The link merely establishes a technical association whose interpretation depends entirely on understanding the specific schema design and naming conventions used by the database creators. Local discourse approaches exist primarily in everyday engineering practice. They are rarely represented in research, as published work focuses on more sophisticated approaches.

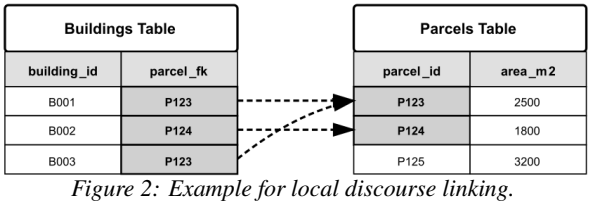


Figure 2: Example for local discourse linking.

Shared Discourse: Objectified Links in Integration Contexts

Shared discourse denotes a bounded semantic context established across multiple models within a project or integration setting. Within this scope, link meaning is fixed through explicit agreements, schemas, or link models that enable consistent interpretation among participating datasets, but remains valid only within that specific integration context.

Examples from research and standardization include information containers that intend to support the integration of heterogeneous information artifacts by organizing them in a common container while preserving their original representations. The ISO 21597 standard (Information Container for Linked Document Delivery, ICDD) introduces an explicit mechanism for representing relationships between documents by modeling links as independent, objectified entities. A similar approach, implemented in the DIN-standard 18290 (Multi-Model Container, MMC), enables links to carry types, metadata, and provenance information, supporting traceability and reproducibility in multi-model environments (Fuchs, 2015). Figure 3 shows an example link between a building model and a parcel in an MMC.

By separating information objects from their relationships, both standards establish a clear conceptual distinction between domain data and integration semantics. Links are

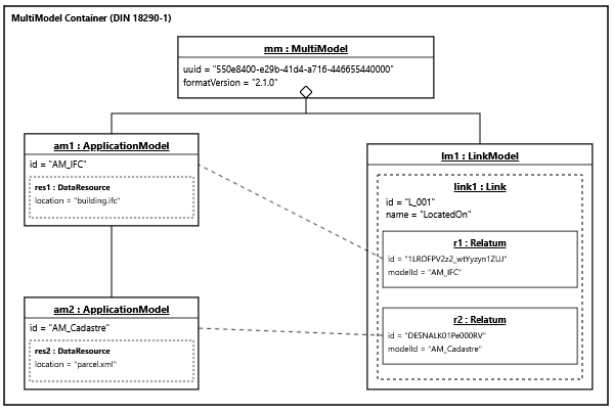


Figure 3: Example for shared discourse linking.

modeled as explicit objects with identifiers, types, and metadata, allowing richer expressiveness than schema-inherent references. In this discourse space, geospatial links can explicitly encode transformation parameters, geometrical links can reference spatial subsets, and topological or mereological relationships can be typed and directed. However, the semantics of these link types are typically defined by project agreements or domain-specific conventions rather than globally standardized vocabularies. Former research has focused on exploiting and extending both schemas to allow more semantic expressiveness, as by Hagedorn et al. (2023) or Krischler et al. (2025) and to include queries as external links.

Shared discourse enables consistent interpretation across participating tools and stakeholders, but links remain context-bound and require governance within the integration setting.

Open Discourse: Ontology- and Graph-Based Linking

Open discourse is realized through semantic web and graph-based approaches, where links are expressed using globally defined predicates and identifiers. Relationships are first-class entities that can be independently queried, extended, and reasoned over. However, the choice for a graph-based approach does not necessarily imply semantic expressiveness, but rather the quality of the used ontology (Akbarieh et al., 2025).

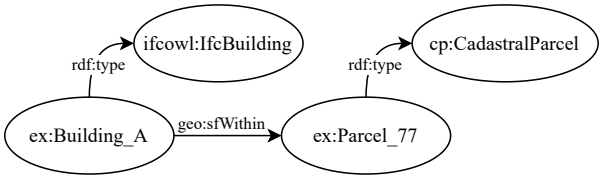


Figure 4: Example for open discourse linking.

For example, semantic web approaches that use URI-identified concepts from established ontologies, such as linking a building element to a concept from the Building Topology Ontology (BOT) (Rasmussen et al., 2020) using *bot:hasSpace*, or links expressed using standardized relationship types from domain-independent vocabularies like Dublin Core (*dcterms:isPartOf*) or schema.org (*schema:containedInPlace*) carry semantics that are globally documented and machine-interpretable without requiring knowledge of the specific BIM-GIS integration context. Works from Hbeich and Roxin (2020); Ji et al. (2024); Wang and Issa (2020) use ontological approaches to overcome integration challenges. Senthilvel (2024) presents a comprehensive approach for managing and validating semantically rich, objectified links using information containers and the BuildLinks Ontology, with a strong focus on provenance, link semantics, and standard compliance. This work allows for an open discourse as it relies on shared metadata understanding as well as semantic web technologies, even if using information containers as carrier.

This paradigm enables the highest degree of expressiveness for all four BIM-GIS link types, including qualified relations, temporal validity, and formal constraints. Mereological and topological semantics can be grounded in axiomatized vocabularies, enabling inference beyond explicitly stated links. The cost of this expressiveness lies in higher conceptual complexity, infrastructure requirements, and limited industry adoption.

Impact on Intention Reconstruction (RQ2)

Intention reconstruction refers to the ability of a receiving system to infer the semantic intent behind a link without reliance on external documentation or institutional memory. It operates on three levels: machine-readability of the link predicate, semantic reconstructability based on a shared vocabulary, and recovery of the originating analytical decision (why the link was created and under which assumptions it remains valid). The categories in Table 1 primarily address the first two levels, though all three matter in practice. Machine-readability suffices for visualization, semantic reconstructability is required for cross-organizational coordination, and intention recovery becomes critical where a link recomputed by a downstream actor may encode different assumptions than the sender intended. The link expressiveness as characterized in Table 1 influences directly whether reconstruction is possible.

Each category contributes differently to reconstruction, e.g. direction clarifies whether information flow can be traced unambiguously and cardinality determines whether complex integration patterns can be faithfully preserved or must be simplified. Architecture governs whether contextual metadata, such as provenance, derivation method, or quality indicators, can be attached to links. Semantic strength implies whether relationship meaning is locally implicit or grounded in standardized vocabularies interpretable without prior schema knowledge. Link depth affects the granularity at which relationships are expressed, from entire documents to individual attributes or geometric sub-elements. Finally, ID scope determines whether identifiers remain valid only within a single file or persist across projects and over time.

These categories do not operate independently. As Table 2 illustrates, concrete linking mechanisms combine them in

characteristic patterns corresponding to the three discourse spaces. Mechanisms in schema-based closed discourse, such as `IFCExternalReference`, or `XLink`, typically exhibit unidirectional linking, limited cardinality, and isolated semantic strength, which severely constrains intention reconstruction to the mere existence of a relationship. In shared discourse, container-based mechanisms such as ICDD linksets and MMC links achieve higher semantic expressiveness through objectified link architecture, explicit directionality, m:n cardinality and standardized semantics within their container schemas, enabling receiving systems to recover relationship type, parameters, and provenance. In open discourse, OWL-based mechanisms combine all high-expressiveness characteristics from Table 1 with globally persistent identifiers, enabling full externalization of sender intention through publicly defined vocabularies that support independent interpretation, validation, and inference.

The following examples demonstrate linking in the three discourse spaces and discuss the respective implications for intention reconstruction.

Closed Discourse: Linking Buildings with Cadastral Parcels

The following Listing 1 shows an example within the IFC data model. A direct reference from an *IfcSite* entity to a cadastral parcel identifier retrieves land ownership information. The cardinality is typically 1:1, occasionally 1:n for connecting parcels. The IFC entity references external object *DESNALK01Pe000RV*, interpretable only within IFC's internal semantics. The link purpose is not explicit; intention must be inferred from metadata, file names, or workflow conventions. The link is unidirectional (cadastral file has no knowledge of this reference), yet well-established in practice.

In a closed discourse, only the information contained inside the schema can be reconstructed. Information outside of this, such as whether the target reference is an object or an attribute, and its meaning, cannot be reconstructed because, unlike the sender, the receiver does not know the schema of the link target. Similarly, the meaning of the link cannot be restored. The sender knows that a building is going to be constructed on the parcel of land to which the link is attached. However, the link does not convey this knowledge, so the recipient cannot understand it.

Listing 1: External link to another file in IFC

```
#33=IFCSITE('2PKyxyUDfD6wqc9p_1NzER',...);
#70=IFCDOCUMENTREFERENCE($,'
    DESNALK01Pe000RV',$,$,#72);
#71=IFCRELASSOCIATESDOCUMENT('1R7R97$
    uLAAv4wci$KGwn8',#18,$,$,(#33),#70);
#72=IFCDOCUMENTINFORMATION('d3L4G','1206
    _ACW.xml',$,'https://www.example.org/'
    ,$,$,$,$,$,$,$,$,$,'2025-11-25','
    2026-11-24',$,$);
```

Shared Discourse: Geospatial Linking for Unified Positioning

Listing 2 shows an example link within a container for the exchange of models between project partners. Information about the georeferencing is required to derive knowledge or understand the link. The link is part of the link model in a multi-model container with extended link types by Krischler et al. (2025). Links are stored in a separate file, but are only valid within the container context. In a shared discourse, the recipient can understand the meaning of a link because the link type and its attributes explicitly specify it. Receiving software can interpret the georeferencing intention based on the link type and additional information, such as inner and outer poses. However, the recipient can only interpret link information that is defined within the context. If the sender uses link information from outside the container context, the recipient cannot interpret it. In this example, the information can be used to transform the data into another coordinate system to visualize all geometrical models in the MMC together.

Listing 2: Extended geospatial link in MMC with additional information for georeferencing

```
<GLM:GeospatialLink type="pose">
  <GLM:GeospatialRelatum m="ContainerCRS"
      type="wktCrs"/>
  <GLM:GeospatialRelatum m="BuildingModel"
      type="innerPose" innerPoseId="#22"/>
  <GLM:outerPose>
    <GLM:Pose origin="
        411251.200,5654378.125,119.300"
       x-axis="-0.305,-0.952,0"
       z-axis="0,0,1" scale="1,1,1"/>
  </GLM:outerPose>
</GLM:GeospatialLink>
```

Open Discourse: Linking Building Parts with Point Clouds

The Listing 3 shows an example using ontologies for linking within a global context. A wall was scanned using laser scanning. The point cloud is linked as a 3D representation of the wall. The wall has a bounding box to get only the relevant part of the point cloud. The point cloud can be retrieved from a URL as a stream.

Listing 3: Link between wall and point cloud using ontologies

```
ex:Wall_23 a bot:Element;
  bot:has3DModel ex:PC_2 .
ex:PC_2 a geo:Feature;
  geo:hasGeometry ex:PC_BBox;
  dcterms:created "2025-09-07"^^xsd:date;
  dcterms:format "application/octet-stream";
  ex:url <https://pc.example.org/pc_2.laz> .
ex:PC_BBox a geo:Geometry;
  geo:asWKT "POLYGON Z((16.0 51.0 100.0,
  ...))"^^geo:wktLiteral .
```

Table 2: Categorization of Linking Mechanisms in BIM-GIS Integration: The different Mechanism are assigned to Discourse Spaces (Disc.) for open, shared and closed

Mechanism	Disc.	Direction	Card.	Archit.	Sem. Strength	Depth	ID Scope
Schema / Encoding-based							
IfcExternalReference	○	unidirectional	1:n	direct	isolated	Object	Global (URI)
GML XLink	○	unidirectional	1:n	direct	isolated (XML)	Object / Property	File / Global (URI)
CityJSON Extensions (lynks schema)	○	unidirectional	1:n	direct	generalized	Object / Property	File
Information Container-based							
MMC Links	◒	directional	m:n	objectified	standardized	File / Object	Container
ICDD Link-set	◒/●	bidirectional	m:n	objectified	standardized	File / Object	Container
Semantic Web-based							
OWL ObjectProperty	●	bidirectional	m:n	objectified	standardized	Object	Global (URI)

Discourse Spaces: ○ closed ◒ shared ● open

Metadata, like format and recording date, can be added for additional information, for example, with the Dublin Core terms ontology. In the open discourse, all information of the sender's intention can be reconstructed, because the meaning of each class and predicate of a link can be retrieved from the web. The sender is not limited to a certain amount of possible link information, like in schemas with a closed discourse.

Discussion: Practical Requirements and Trade-Offs (RQ3)

Different discourse spaces imply distinct practical trade-offs:

Closed discourse offers low implementation effort and broad tool support, but systematically shifts interpretation costs downstream. While well-suited to integration scenarios that are temporally bounded (e.g. a single project phase), organizationally contained (few stakeholders involved), and functionally routine (stable workflows and well-known schemas with little risk of unknown data sets that might necessitate extensions), closed discourse provides only limited potential for reuse of both the linked data and the links themselves. Because link semantics remain implicit and context-bound, the sender's intention, such as why a specific spatial, topological, or geometric relationship was established, cannot be reconstructed without gaps once the link is detached from its original environment. As a result, analytical decisions embedded in the linking process (e.g. spatial analyses, topological evaluations, or transformation derivations) are not reproducible by receiving systems and must be recomputed by downstream actors. This introduces redundancy and constrains the derivation of new knowledge, as inference requires substantial additional effort or manual interpretation, and it raises uncertainty about the dataset as a whole. The interpretation burden therefore remains with human expertise rather than being encoded in a machine-interpretable form, and the contextual knowledge that motivated linking decisions persists only implicitly, relying on external documentation, tacit assumptions, or institutional memory.

Shared discourse balances semantic expressiveness and practical feasibility, making it particularly approachable for industry-oriented BIM-GIS workflows. It enables integration without requiring full data homogenization or migration into a unified semantic model, allowing datasets to remain in their native formats while supporting explicit, structured linking. Integration can therefore be carried out within a bounded, project-specific environment in which a shared "project truth" is established, without generalizing relationships beyond the immediate context unless required. At the same time, reliance on open and standardized container and link models ensures that integration artifacts remain extensible and interpretable for future use. By externalizing sender intention through objectified link structures, such as geometric transformations, topological relations, derivation provenance, or quality indicators, shared discourse reduces redundancy in downstream processing. Analytical results produced during link creation, for example spatial or topological analyses, can be directly consumed by receiving software systems instead of being recomputed, which might lead to amortizing integration effort across multiple stakeholders and project phases. This moderate increase in architectural complexity is justified where recurring integration tasks, cross-organizational coordination, and quality assurance demand a controlled yet flexible semantic environment.

Open discourse offers the highest potential for long-term interoperability, reuse, and machine-supported reasoning, but it also introduces substantial practical barriers for current industry adoption. In contrast to closed or shared discourse, it typically presupposes that data sources are available in RDF, that suitable ontologies (T-boxes) are selected, aligned, or extended for project-specific purposes, and that instance data (A-boxes) are populated accordingly. This form of semantic engineering requires expertise, tooling, and organizational maturity that are not yet widely established in the AEC sector, resulting in significantly higher upfront effort and complexity. Where applied successfully, open discourse enables integration beyond project-specific contexts, allowing links to remain interpretable across organizational and institutional bound-

aries without bilateral semantic agreements. Typical application domains include legal traceability, long-term information retention for asset and infrastructure management, and the consolidation of heterogeneous models across administrative levels such as cities or regions. Moreover, open discourse enables automated reasoning and inference, supporting use cases such as consistency checking, completeness verification, and the derivation of implicit relationships. The investment in semantic infrastructure is therefore justified primarily when links are intended to function as persistent, reusable knowledge assets, enabling machine-accessible representations of sender intention, analytical provenance, and contextual validity constraints at scale.

Conclusion

This study examined the significance of links in BIM-GIS integration from the perspective of the discourse space. It demonstrated that the semantic value and practical utility of links are not determined solely by the technology, but by the context in which they arise and are interpreted. By introducing closed, shared, and open discourse as an analytical framework, the paper provided a systematic basis for comparing heterogeneous link modeling approaches and positioning existing research accordingly (RQ1). This perspective allows integration mechanisms to be evaluated not only by their formal capabilities, but by their suitability for different organizational, temporal, and functional contexts.

The analysis showed that link expressiveness directly affects the ability to reconstruct sender intention and semantic context (RQ2). In closed discourse, intent remains implicit and bound to local schemas or workflows, limiting reproducibility and reuse. Shared discourse partially externalizes intention through objectified links and contextual metadata, enabling consistent interpretation within bounded integration settings. Open discourse fully externalizes semantic intent through globally defined vocabularies and formalized relationships, enabling independent interpretation, validation, and inference beyond project-specific boundaries.

From a practical perspective (RQ3), the three discourse spaces represent distinct trade-offs rather than a linear maturity path. Closed discourse offers low entry barriers but high downstream interpretation costs; shared discourse provides an industry-viable balance between expressiveness and feasibility; and open discourse enables long-term interoperability and reasoning at the cost of substantial semantic infrastructure and expertise. Instead of advocating a single integration paradigm, this paper contributes decision support for selecting appropriate linking strategies based on required semantic depth, reuse potential, and organizational readiness.

A key limitation of this work concerns the validation of link quality and sender intention reconstruction. While the paper shows conceptually how discourse space influences the externalization of semantic intent, it does not define formal metrics to assess how accurately sender intention can be reconstructed by receiving systems. Sender intention is often tacit and context-dependent, making objective evaluation difficult without a clear ground truth. In practice, reconstruction can only be assessed indirectly, for example by evaluating whether links can be interpreted without auxiliary documentation or analytical results reproduced.

Outlook

Future work will focus on the development of an optimized link meta-model capable of operating coherently across closed, shared, and open discourse spaces while remaining compatible with existing standards and domain data models. A meta-model would define a technology-agnostic conceptual core for links, capturing relationship types, scope of validity, sender intention, and contextual metadata independently of their concrete encoding. This enables the same conceptual link to be instantiated in different technical serializations - such as schema-inherent references, container-based link models, or ontology-based representations - depending on project needs, supporting progressive semantic enrichment without forcing premature data homogenization. The proof of whether the choice of closed, shared, or open can be provided technically by ensuring that a receiving software system can interpret the links and use them for analysis and visualization. At the same time, it must be ensured that the chosen link approach is adopted by the software industry and regularly used in real BIM-GIS projects. In parallel, emerging AI-based methods offer promising opportunities to support the improvement of weakly structured or semantically inconsistent datasets and to generate links by automated methods. Potential applications include the normalization of naming conventions, the identification of implicit semantic patterns, and the assisted mapping of project-specific attributes to standardized vocabularies, as well as supporting validation and refinement of link models, e.g. by suggesting missing relations or highlighting inconsistencies.

AI Declaration

During the preparation of this work, the authors used AI tools for grammar and spelling check, paraphrase and reword. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

Akbarieh, A., Osadcha, I., Farghaly, K., Bosché, F., and Puust, R. (2025). Development of a Built Environment Ontology Lookup Service (BE-OLS) with a New Ontology Evaluation. In Proceedings of the 2025 European Conference on Computing in Construction.

Beck, F. (2023). Context-sensitive linking of heteroge-

neous information models from the building and the urban domain. PhD thesis, Technische Universität München.

Beck, F., Borrmann, A., and Kolbe, T. H. (2020). The need for a differentiation between heterogeneous information integration approaches in the field of BIM-GIS Integration: A literature review. In ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, volume VI-4-W1-2020.

Borrmann, A., Abualdenien, J., and Krijnen, T. (2021). Information containers providing deep linkage of drawings and bim models. In Proceedings of the CIB W78 - LDAC 2021, pages 823–835.

Davar Panah, S. A. (2025). Untersuchung einer Erweiterung des CityJSONStandards für die Verknüpfung heterogener Daten im Bauwesen. Master's thesis, Bauhaus-Universität Weimar, Weimar.

Eriksson, H., Johansson, T., Olsson, P.-O., Andersson, M., Engvall, J., Hast, I., and Harrie, L. (2020). Requirements, Development, and Evaluation of A National Building Standard—A Swedish Case Study. ISPRS International Journal of Geo-Information, 9(2).

Fuchs, S. (2015). Erschließung domänenübergreifender Informationsräume mit Multimodellen: = Access of cross-domain information spaces using multi-models. PhD thesis, TU Dresden, Dresden.

Guyo, E., Hartmann, T., and Ungureanu, L. (2021). Interoperability between BIM and GIS through open data standards: An overview of current literature. In Proceedings of the 9th Linked Data in Architecture and Construction Workshop.

Hagedorn, P., Senthilvel, M., Schevers, H., and Verhelst, L. (2023). Towards usable ICDD containers for ontology-driven data linking and link validation. In LDAC 2023 - Linked Data in Architecture and Construction.

Hbeich, E. and Roxin, A. (2020). Linking BIM and GIS standard ontologies with Linked Data. In Proceedings of the 8th Linked Data in Architecture and Construction Workshop, pages 126–135.

Ji, Y., Wang, Y., Wei, Y., Wang, J., and Yan, W. (2024). An Ontology-based Rule Mapping Approach for Integrating IFC and CityGML. Transactions in GIS.

Kolbe, T. H. and Donaubauer, A. (2021). Semantic 3D City Modeling and BIM. In Urban Informatics, pages 609–636. Singapore.

Korkmaz, Ö. and Basaraner, M. (2024). Ontology Based Integration of BIM and GIS for the Representation of Architectural, Structural and Functional Elements of Buildings. In Proceedings of the 31st International Workshop on Intelligent Computing in Engineering, pages 49–55.

Krischler, J., Schilling, S., Taraben, J., Sternal, M., Clemen, C., and Koch, C. (2025). A Standards-Based Approach to BIM-GIS Integration: Extending the Multi-Model Container Schema. In LDAC 2025 - Linked Data in Architecture and Construction.

Liu, X., Wang, X., Wright, G., Cheng, J. C., Li, X., and Liu, R. (2017). A State-of-the-Art Review on the Integration of Building Information Modeling (BIM) and Geographic Information System (GIS). ISPRS International Journal of Geo-Information, 6(2):53.

Mohammed, P., Behan, A., O'Sullivan, D., and McAuley, B. (2026). Level of Integration (LoInt): A Conceptual Framework for BIM-GIS Data Integration. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVIII-4/W17-2025.

Napps, D., Winck, G., and König, M. (2026). Comparative evaluation of resource description framework and labeled property graphs for BIM-based early design decisions. Automation in Construction, 186:106854.

Rasmussen, M. H., Lefrançois, M., Schneider, G. F., and Pauwels, P. (2020). BOT: The Building Topology Ontology of the W3C Linked Building Data Group. Semantic Web, 12(1):143–161.

Sani, M. J., Musliman, I. A., and Abdul Rahman, A. (2022). IFC to CityGML Conversion Algorithm Based on Geometry and Semantic Mapping. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLVI-4/W3-2021:287–293.

Schilling, S. and Clemen, C. (2023). TerrainTwin - Microservice Architecture for the Integration of Geodata (GIS) and Building Models (BIM) Using Link Models. Proceedings of the FIG Working Week.

Senthilvel, M. (2024). Linking and managing heterogeneous data using information containers : leveraging linked data for BIM Stage 3 CDEs. PhD thesis, RWTH Aachen University.

Tan, Y., Liang, Y., and Zhu, J. (2023). CityGML in the integration of BIM and the GIS: Challenges and opportunities. In Buildings, volume 13, page 1758.

Wang, N. and Issa, R. R. A. (2020). Ontology-based integration of BIM and GIS for indoor routing. In Construction Research Congress 2020: Computer Applications, pages 1010–1019.

World Wide Web Consortium (2014). Best Practices for Publishing Linked Data: W3C Working Group Note 09 January 2014.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

TOWARDS A MULTI-DOMAIN SEMANTIC ARCHITECTURE FOR GRAPH-BASED ORCHESTRATION OF DECENTRALISED ADDITIVE MANUFACTURING

Lukas Kirner, Alex Donkers, and Ekaterina Petrova
Eindhoven University of Technology, Eindhoven, The Netherlands

Abstract

Construction-scale additive manufacturing (AM) is often executed in decentralised, multi-actor project environments, where production information is distributed across heterogeneous systems. This fragmentation hinders automation, coordination, and life-cycle traceability. This paper proposes a multi-domain semantic system architecture to support graph-based integration and orchestration of decentralised AM workflows. The architecture separates system topology, functional service interfaces, information requirements, and runtime provenance. Using a concrete 3D printing use case, the approach demonstrates how modular software components can be integrated through a shared digital-twin interface while maintaining end-to-end traceability of generated and consumed artefacts.

Introduction

Construction sectors in many developed countries share persistent structural challenges, including rising costs, long project durations, shortages of skilled labour, and increasing regulatory requirements related to sustainability and life-cycle assessment (Abdelalim et al., 2025; Cabeza et al., 2014). At the same time, the demand for affordable housing and resource-efficient construction continues to grow. These pressures challenge traditional construction methods and motivate the exploration of alternative production paradigms (Kedir and Hall, 2021).

Construction-scale additive manufacturing (AM) such as concrete 3D printing, has emerged as a promising approach to address some of these challenges (Lim et al., 2012; Buswell et al., 2018). By enabling digitally driven fabrication processes for building components, AM offers potential benefits including reduced labour dependency, shorter construction times, and material-efficient geometries (Hager et al., 2016). At the same time, the strong reliance on digital planning, parameterisation, and process control introduces complex information dependencies between design, material definition, and execution.

The implementation of AM on construction scale is usually integrated within temporary, multi-actor project environments. Design, material definition, machine control, and quality assurance are often handled by different stakeholders and supported by heterogeneous software systems. As a result, production-related information is generated across organisational and temporal boundaries, while increasing requirements for documentation and life-cycle assessment demand consistent traceability of design decisions and production-related information (Wu et al., 2016).

While practical challenges that relate to printing hardware and material behaviour are prominent, efficient data management and process coordination represent equally significant hurdles (Jaskuła et al., 2024). Information is frequently exchanged through isolated files, proprietary software tools, or manual handovers, resulting in limited interoperability and weak integration between design, execution, and as-built documentation. These limitations hinder automation, scalability, and the consistent availability of information required for life-cycle assessment (Cabeza et al., 2014).

Semantic models and linked data approaches have been proposed to address these challenges by enabling formal, machine-interpretable representations of entities, relationships, and dependencies across organisational and system boundaries. Graph-based representations are particularly well suited for decentralised settings in which no single system can act as a central source of truth (Pauwels et al., 2017).

Despite this potential, the practical use of semantic approaches in decentralised AM remains limited. Existing ontologies and data models typically address isolated aspects of the process chain, such as geometry, manufacturing activities, or sensor observations (Kirner et al., 2024a), while lacking a coherent architectural separation between system integration, functional interfaces, information dependencies, and execution traces. As a result, semantic models often become either difficult to operationalise or insufficient to support cross-system coordination and traceability.

This paper addresses this gap by proposing a multi-domain semantic system architecture for decentralised AM workflows. The architecture explicitly separates system topology, functional service descriptions, information requirements between modules, and runtime provenance. Using construction-scale concrete 3D printing as a representative use case, the approach demonstrates how existing semantic standards can be combined into an actionable architecture that supports integration, coordination, and end-to-end traceability across the AM process chain.

Related Work

Additive manufacturing

Construction-scale AM transfers layer-wise fabrication principles from industrial AM to the production of building components and, increasingly, full-scale construction, with early work on large-scale automated material deposition such as Contour Crafting and concrete printing systems (Khoshnevis, 2004; Lim et al., 2012). Over the past decade, concrete 3D printing (3DCP) has become the dominant form of construction-scale AM due to its poten-

tial to reduce formwork, enable material-efficient geometries, and increase automation (Buswell et al., 2018).

Typical 3DCP workflows span digital design, toolpath generation, material preparation, and layer-wise deposition with subsequent quality assurance. While geometry-to-machine translation resembles industrial CAM pipelines, construction-scale AM is strongly influenced by time-dependent material behaviour, such as buildability and inter-layer adhesion, which couples digital models, process parameters, and execution (Mechtcherine et al., 2020). Challenges in data management and process coordination include fragmented information flows, limited interoperability, and insufficient traceability of material definitions, parameters, and as-built outcomes (Bos et al., 2016; Jaskuła et al., 2024).

Semantic and ontology-based approaches in AM

To address gaps in data availability, data-centric frameworks in AM focus on the structured capture of process information. Deetman et al. introduce a database framework for 3D concrete printing to store cross-disciplinary process data for analysis and reproducibility focussing on storage and schema (Deetman et al., 2026).

For semantic integration across tools and domains, ontology-driven approaches have been suggested to structure and integrate knowledge in AM. Li and Petzold present a mixture-of-domain documentation framework that captures cross-domain knowledge in a shared knowledge graph to support semantic retrieval and question answering across robotics, materials, and BIM domains (Li and Petzold, 2025). Domain-specific semantic models have also been introduced, for example for wire arc additive manufacturing and composite extrusion, to formalise process knowledge and enable semantic interoperability across tools (Ahmad et al., 2025). Gouttebroze et al. propose a semantic standard and process ontology for additive manufacturing that aligns standard terminology with formal process descriptions (Gouttebroze et al., 2023).

System integration and system-of-systems approaches

The increasing distribution of digital tools in construction has motivated system-of-systems approaches for integrating heterogeneous services. Hagedorn et al. propose a system-of-systems architecture for deploying containerised construction IT services, focusing on modular integration and explicit service interfaces across independently managed components (Hagedorn et al., 2025). In a more general setting, Feng et al. present an ontology-based approach for modelling system-of-systems architectures and operational views to provide a shared semantic foundation for describing interacting systems (Feng et al., 2023).

These approaches address deployment and architectural modelling of distributed systems, but do not explicitly model information dependencies between tools or capture execution-time provenance needed for traceability across workflows.

Methodology

This work develops a multo-domain semantic system architecture to support graph-based coordination and traceability in decentralised construction-scale AM. The methodology to achieve this follows four steps: (I) use-case driven examination of an exemplary AM process chain and involved software systems, (II) identification of coordination and traceability concerns across system boundaries, (III) structuring of these concerns into distinct semantic domains, and (IV) alignment of each domain with existing standards and ontologies to form an extensible architectural framework.

I Use-case driven system examination

To ground the methodology in practice, the approach starts from a representative decentralised AM workflow derived from the EU-funded project "Additive To Predictive Manufacturing" (AM2PM)[1]. The focus is placed on a possible design–material–optimisation sub-workflow that supports predictive and sustainability-driven decision-making, while abstracting from the physical printing process itself. This scoping reflects the intention to examine information dependencies between software modules rather than machine control or sensing.

The considered sub-workflow couples material design, structural optimisation, and life-cycle assessment. A concrete mix design module produces a mix specification that includes, among others, cement content and material properties such as compressive strength. The cement component of the mix is linked to environmental product declaration (EPD) data, enabling the association of embodied CO_2 emissions with the mix composition. These material properties are required inputs for a shape optimisation module, which aims to minimise material usage for a slab structure by exploiting the mechanical performance of both the printed lost formwork and the subsequently poured concrete. Equipment type, staff, material quantities and other practical parameters are assessed in a simulation module. The involved modules and their interactions are illustrated in a simplified manner in Figure 1, where modules interact with a decentralized graph-based integration domain. Brown arrows represent two-way interaction between a module and the integration domain, and contain data conversion. Solid green arrows represent semantic relationships between concepts in a module, whereas dotted green arrows represent mappings between such semantic concepts. An optimization module that feeds a control module is using the integration domain. The latter is not the primary focus of this work.

Based on the material parameters and structural requirements, the optimisation module determines an efficient geometry, which directly influences the amount of material to be printed and cast. Whether material quantities are later measured during printing or estimated through predictive simulation is outside the scope of this work. In-

[1] https://www.am2pm-project.eu/

stead, attention is placed on how the required information is made available and propagated across modules. Given access to both the optimised geometry and the concrete mix specification, a life-cycle assessment (LCA) module can estimate the environmental impact of the slab, dominated by cement-related emissions.

In such a setting, changes in concrete composition affect structural behaviour, geometric optimisation results, material quantities, and downstream environmental assessments. These interdependencies are realised across multiple software modules that operate independently but rely on shared information artefacts.

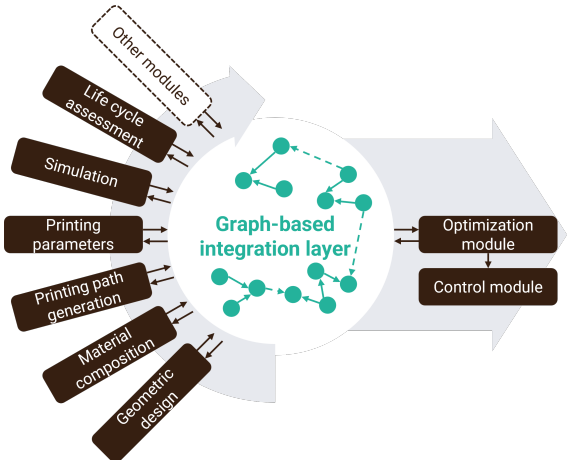


Figure 1: Simplified AM2PM module interaction.

II Coordination and traceability concerns

From the use-case driven system examination, a set of recurring and characteristic coordination and traceability concerns emerge. These concerns arise from the fact that multiple independent software systems contribute to a shared production outcome, while relying on different representations, responsibilities, and temporal perspectives on the same information artefacts.

A first class of concerns relates to information dependencies across system boundaries. Design artefacts, material definitions, optimisation results, and environmental data are produced and consumed by different modules, often at different points in time. Dependencies between these artefacts are typically implicit, making it difficult to determine whether all required inputs are available, which upstream changes necessitate recomputation, or which downstream results are affected by modifications in design or material parameters. Such dependencies are relevant at design time and influence validation, scheduling, and coordination decisions before any execution takes place.

A second class of concerns concerns traceability across execution and decision contexts. Results generated during execution, simulation, or optimisation need to be traceable back to the design-time inputs, assumptions, and configurations that led to them. This traceability extends beyond runtime logging and includes the ability to associate outputs with the specific versions of input artefacts and system configurations that were in effect when they were produced. Without explicit representation of these relations, it becomes difficult to support reproducibility, accountability, and life-cycle assessment in workflows where responsibilities and data are distributed across multiple systems.

Addressing these issues requires a structuring of system-related knowledge that separates stable system descriptions, information dependencies, and execution-time records.

III Derivation of the semantic architecture

The identified coordination and traceability concerns cannot be handled well within a single, undifferentiated semantic model. Design-time dependencies, stable system descriptions, and execution-time records differ in how long they remain valid, who is responsible for them, and how often they change. Treating these aspects in the same way risks either mixing volatile execution data into long-lived system descriptions or failing to represent dependencies that are needed for coordination before execution.

To address this, the proposed architecture separates semantic responsibilities into four domains with distinct scope, temporality, and mutability (see Table 1). The domains distinguish between system structure (*D1*), functional interfaces between systems (*D2*), information requirements and dependencies (*D3*), and provenance records of concrete executions (*D4*). This separation follows patterns observed in the use-case driven examination rather than assumptions imposed by a specific ontology.

By separating these concerns, each domain can be modelled at an appropriate level of abstraction. System topology and interfaces change slowly and support integration, information requirements capture dependencies that must be known before execution, and provenance records document what actually happened. This allows the domains to evolve independently while remaining linked, which is essential in decentralised workflows with multiple interacting systems.

IV Ontology selection and integration strategy

The proposed architecture is instantiated using a deliberately small set of established ontologies, selected for their conceptual fit with individual domains rather than for exhaustive domain coverage. This choice defines a minimal semantic starting point. Fig. 2 shows the integration of these ontologies.

System topology (D1)

At the system topology domain, structural relationships between software components are represented using *SAREF for System* (SAREF4SYST, prefix `s4syst`) (Lefrançois, 2019). It provides explicit concepts for systems, connection points, and connections, making it suitable for modelling decentralised software modules and their integration while focussing on a relatively stable system structure that aligns with the slow-changing nature of topology descriptions.

Table 1: Domain separation and according to scope, temporality and mutability (non-exhaustive).

Domain	Scope	Temporality	Mutability
D1: System topology	system structure	mostly static	slow
D2: Functional description	interface	mostly static	slow
D3: Information requirements	workflow dependency	logical	medium
D4: Provenance and traceability	execution record	historical	immutable

Functional service description (D2)
Functional interfaces are described using the W3C *Web of Things (WoT) Thing Description* (prefix `td`) (Charpenay et al., 2023). Thing Descriptions provide a machine-interpretable, declarative description of interaction affordances exposed by a system, including *properties*, *actions*, and *events*. In this architecture, the Digital Twin interface is represented as a `td:Thing` whose affordances describe access to shared project artefacts (e.g., concrete mix specifications, optimisation results) without prescribing orchestration logic or encoding execution-time provenance. This makes WoT TD suitable for complex, decentralised module ecosystems while keeping the functional contract separated from design-time information dependencies (*D3*) and runtime traces (*D4*).

Information requirements and dependencies (D3)
Design-time information dependencies are captured in the information requirement domain using the IoC Process Ontology (prefix `ioc`) (Kirner et al., 2024b) in alignment with the *WoT* (Charpenay et al., 2023). An `ioc:Process` is aligned with a `td:Thing`, while specific operations are modelled as WoT affordances. Information requirements are specified at the level of these affordances using `ioc:hasInputInformation` and `ioc:hasOutputInformation` to indicate which `ioc:Information` artefacts are required or produced by a given action, independent of their runtime access or generation (e.g. a topology optimisation action requiring a concrete mix specification). Information artefacts are treated as logical coordination objects supporting dependency reasoning, completeness checks, and change impact analysis prior to execution. Artefacts may be described using domain ontologies, but their internal semantics are not modelled at this domain. The separation of D3 from execution-time provenance enables validation and orchestration without reliance on historical execution data. In practice, explicitly modelled information requirements can be validated using SHACL shapes (Knublauch and Kontokostas, 2017). The domain itself may require future extensions by more formal requirement or constraint ontologies.

Provenance and traceability (D4)
Execution-time behaviour and traceability are represented using the W3C Provenance Ontology *PROV-O* (prefix `prov:`) (Lebo et al., 2013). PROV-O provides a standard model for representing activities, entities, and agents, and for expressing how information artefacts are generated, used, and derived over time. In the proposed architecture, provenance is used exclusively to record immutable execution histories, linking concrete runs of distributed modules to the specific input artefacts and system configurations that produced them. This makes PROV-O well suited for auditability, reproducibility, and life-cycle relevant traceability, while remaining strictly separated from coordination and dependency modelling.

Table 2: Overview of used ontologies for the proposed semantic architecture.

Ontology	Main classes and purpose
D1: `s4syst:`	*s4syst:System*, *s4syst:Connection/Point* System topology: software modules and connections (Lefrançois, 2019)
D2: `td:`	*td:Thing*, *td:PropertyAffordance*, *td:ActionAffordance*, *td:EventAffordance* Functional interfaces: interaction affordances and specific endpoints (Charpenay et al., 2023)
D3: `ioc:`	*ioc:Process*, *ioc:Information* Information requirement: Expresses design-time input/output dependencies (Kirner et al., 2024b) validated using SHACL (Knublauch and Kontokostas, 2017)
D4: `prov:`	*prov:Entity*, *prov:Activity*, *prov:Agent* Immutable execution history: Branches with used/generated artefacts and responsible agents for traceability (Lebo et al., 2013)

Cross-domain integration principles
The architecture deliberately avoids merging ontologies or introducing inheritance relations across domains. Each domain defines its own relations and responsibilities, while instances may be referenced across domains when appropriate (e.g. a software system described at the topology domain acting as a provenance agent at runtime). Relations that describe execution-time behaviour are expressed exclusively at the provenance domain. This separation prevents semantic overlap between design-time coordination and runtime history, while allowing referenced domain ontologies to evolve independently.
The selected vocabularies are not assumed to be sufficient in the long term. Instead, the architecture is designed to allow extensions and the integration of additional domain-specific ontologies as requirements evolve. Table 2 summarises the ontologies and standards used at each domain of the proposed architecture and their respective roles.

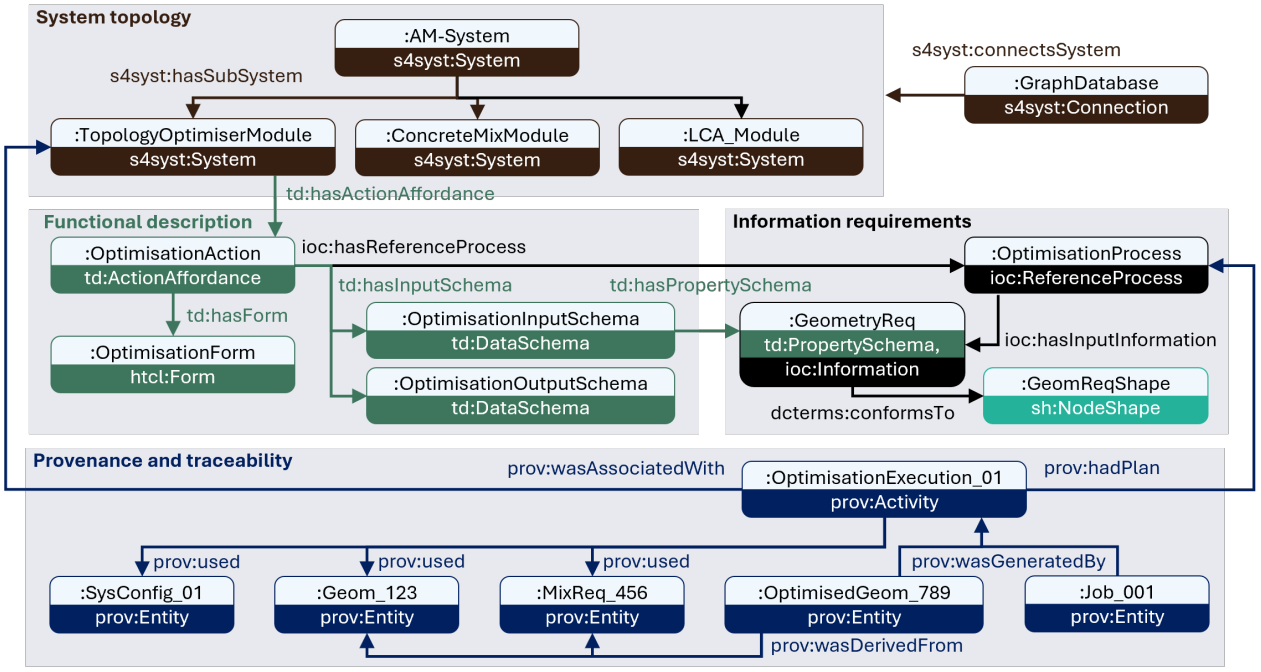


Figure 2: System integration using proposed ontologies, including a demonstrator instantiation

Evaluation and Demonstrator

Having defined the semantic architecture and its ontology alignment, the evaluation demonstrates its instantiation in the AM2PM use case through RDF instances, exemplary SPARQL queries, and a visualised provenance-aware information flow, and presents the semantic model together with its implementation in a platform prototype.

Instance-level semantic model for the AM2PM use case

The system and topology domain (D1) instantiates the Digital Twin Orchestrator as a `s4syst:System` and `td:Thing`. Exemplarily, the material design, topology optimisation, and LCA modules are attached via `s4syst:hasSubSystem`. Interactions between these components are mediated through an explicit `s4syst:Connection`, enabling the model to describe both structural composition and communication paths without mixing functional roles. Each module is additionally a `prov:Agent`, allowing execution responsibilities to be linked directly to provenance records.

The interaction domain (D2) is realised by a `td:ActionAffordance` bound to an HTTP endpoint via `hctl:hasTarget` and `htv:methodName`. The action is linked through `wp4-ext:hasReferenceProcess`, connecting the API operation to the abstract process description. Its payload is constrained by `td:hasInputSchema` and `td:hasOutputSchema`, where inputs such as *:geometry* and *:mixRequirements* are defined as IRIs, enforcing a by-reference data exchange pattern. This separates interface invocation from data ownership and allows the same service interface to operate over distributed or externally managed artefacts.

The information domain (D3) specifies required and produced artefacts independently of execution. The reference process declares its contract using `ioc:hasInputInformation` and `ioc:hasOutputInformation`, linking, for example, to *inst:GeometryInfo*, *inst:ConcreteMixReqInfo*, and *inst:OptimisedGeometryInfo*. Each information object is annotated with `dcterms:conformsTo` to reference a SHACL shape, enabling validation of the referenced resources without embedding the data into the invocation payload. As a result, information completeness and structural validity can be checked before execution and reused consistently.

The provenance domain (D4) models specific process executions as `prov:Activity` and links them to the reference process via `prov:hadPlan`. Used inputs are instantiated with `prov:used`, generated results with `prov:wasGeneratedBy`, and input-result dependencies with `prov:wasDerivedFrom`, making the full derivation chain between input versions and outputs explicitly queryable. Input and output artefacts are represented as `prov:Entity`. As example, revising a concrete mix parameter such as its compressive strength results in a new mix entity that is `prov:used` in a subsequent optimisation run. This leads to a modified optimised geometry, which in turn affects the life-cycle assessment results derived from that geometry. The corresponding result entities can be linked via `prov:wasRevisionOf`, while preserving `prov:wasDerivedFrom` relations to the exact input versions, enabling causal tracing of how changes in material parameters propagate through optimisation and LCA.

Query-based semantic analysis

To demonstrate the usability of the proposed approach, a set of simple questions covering all four domains is answered using SPARQL queries. The queries reflect typical information needs of users and software agents interacting with the Digital Twin system. Listing 1 shows the additional prefixes used.

Listing 1: Additional prefixes for subsequent SPARQL queries.

```
dcterms:<http://purl.org/dc/terms/>
hctl: <https://www.w3.org/2019/wot/hypermedia#>
htv: <http://www.w3.org/2011/http#>
omg: <https://w3id.org/omg#>
sh: <http://www.w3.org/ns/shacl#>
```

The first question addresses the system topology (D1): *Which modules exist and how are they connected?* The corresponding SPARQL query (Listing 2) returns all connected modules with their labels and descriptions and serves as an entry point for discovering the current system structure.

Listing 2: Query for retrieving all connected modules (D1).

```
SELECT ?subsystem ?label ?description
WHERE {
  ?syst s4syst:hasSubSystem ?subsystem;
        s4syst:connectedThrough ?connection.
  ?subsystem rdfs:label ?label;
  dcterms:description ?description;
  s4syst:connectedThrough ?connection .}
```

The second question focuses on the available interaction methods, such as callable API endpoints (D2). Using one of the modules identified in Listing 2, the question is: *Which actions can be invoked on the TopologyOptimiser module and what input do they expect?* The query returns the available actions, for example *inst:RunOptimisationForm*, an HTTP POST form that triggers the topology optimisation. Labels and `dcterms:description` support not only invocation but also discovery of these endpoints, for instance by software agents or agent-based toolchains.

Listing 3: Query to the form for a specific affordance (D2).

```
SELECT ?action ?actionl ?actionDesc ?method ?target
     ?contentT
WHERE {
  inst:TopologyOptimiserModule td:hasActionAffordance
       ?action.
  OPTIONAL {?action rdfs:label ?actionl.}
  OPTIONAL {?action dcterms:description ?actionDesc.}

  ?action td:hasForm ?form.
  OPTIONAL {?form htv:methodName ?method.}
  OPTIONAL {?form hctl:hasTarget ?target.}
  OPTIONAL {?form hctl:forContentType ?contentT.}}
```

The third question addresses information requirements and constraints (D3): *What is the needed content of infromation the input needs to perform the requested action? How must it be sturctured?* Starting from the action identified in Listing 3, the query follows its link to the associated `ioc:ReferenceProcess` (aligned to `prov:Plan`). Using `ioc:hasInputInformation`, it retrieves the strcture of the required information artefacts together with their SHACL-based constraints. The use of D3 and the required domain ontologies to concisely describe the expected data structures is subject to future work.

Listing 4: Query to retrieve information requirements.

```
SELECT ?info ?fieldName ?dataT ?shape
WHERE {
  inst:OptimisationAction ioc:hasReferenceProcess ?
      proc.
  ?proc ioc:hasInputInformation ?info.
  OPTIONAL {?info rdfs:label ?fieldName.}
  OPTIONAL {?info td:datatype ?datatT.}
  OPTIONAL {?info dcterms:conformsTo ?shape.}}
```

The fourth question addresses provenance after a process execution (D4): *Which executions produced which results, and which data influenced these outcomes?* In the example query, the resulting `omg:Geometry` is used as an entry point to retrieve the activity that generated it and the associated execution context. This includes the used input data, the reference process or plan that was followed, and the responsible module (`prov:Agent`), in this case `inst:TopologyOptimiser`. For simulation workflows, this supports comparing outputs produced under the same reference process but with different inputs, enabling systematic analysis of how parameter changes affect the results e.g. for additive manufacturing.

Listing 5: Query to retrieve provenance information after a sucessfull process execution.

```
SELECT ?activityl ?inputs ?result ?agentl
WHERE {
  ?result a omg:Geometry.
  ?result prov:wasGeneratedBy ?activity.
  ?activity prov:used ?inputs;
    prov:hadPlan ?reference;
    prov:wasAssociatedWith ?agent.
  OPTIONAL { ?activity rdfs:label ?activityl.}
  OPTIONAL { ?agent rdfs:label ?agentl.}}
```

The presented questions can be asked independently, for example by different systems or software agents with distinct responsibilities. They can also be applied in the given sequence, progressing from a general understanding of the system, to concrete instructions for its use, to its information requirements and constraints, and finally to the interpretation of its outputs. This progression is essential for supporting simulation workflows and, ultimately, can enable future predictive manufacturing applications.

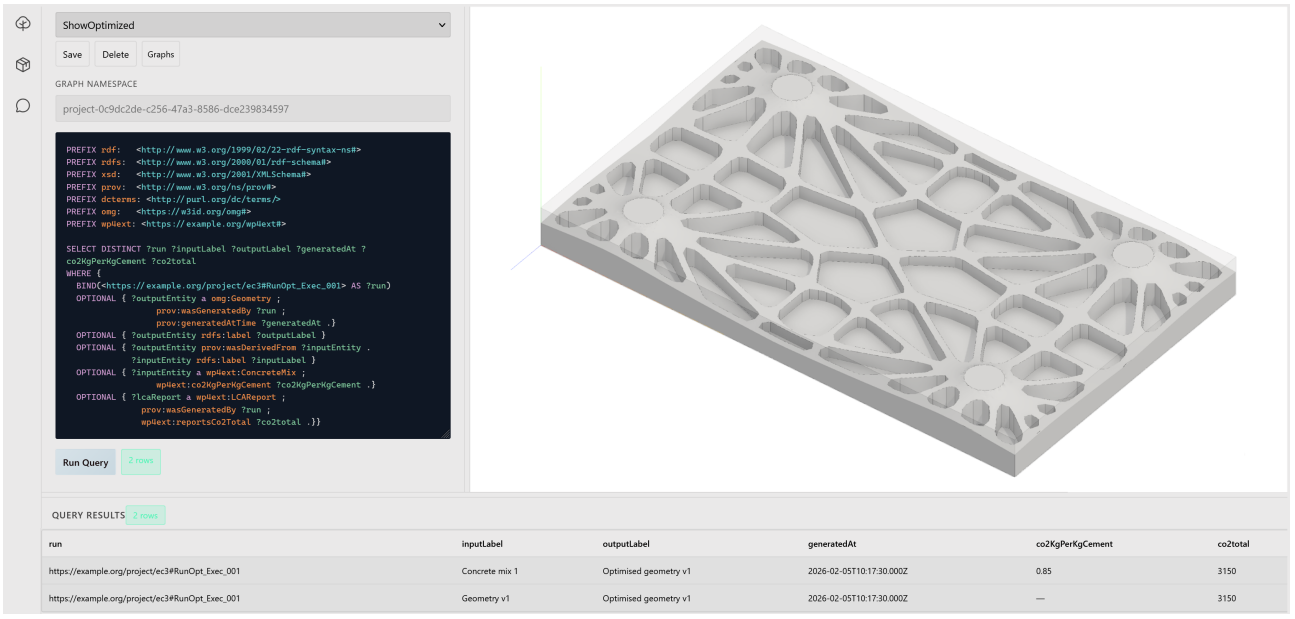


Figure 3: Web platform linking original and optimised geometry to multi-domain semantic data.

Platform-based visualisation and comparison

To demonstrate what the multi-domain semantic architecture can achieve for user-facing applications, the presented exemplary A-Box is stored in a triple store of a linked data–based web platform. The application exposes a query endpoint that can be controlled by its viewer component (see Fig. 3). On the left, a SPARQL query is used to retrieve the output of an optimisation run involving the concrete mixing, topology optimisation, and LCA modules. Here, we focus on the query logic of D3 and D4. The 3D viewer shows in transparent (as hint) the original, non-optimised geometry and highlights the results of the optimised run. Below, the query results are shown in a table, so users can access relevant information about the completed optimisation run, such as its IRI, the used cement mix, the time the geometry was generated and resulting CO_2-equivalent.

Results & Discussion

The evaluation shows that the semantic integration of distributed modules, information requirements, and provenance records enables the comparison of two variants that satisfy the same functional and structural requirements while differing in material usage and CO_2-equivalent emissions, and that a simple web-based 3D viewer can make these differences accessible to users.

Beyond this specific example, queries across domains illustrate how the approach supports typical coordination and analysis tasks, such as discovering available modules, checking required information, and tracing how results were produced. This indicates that the architecture is not limited to a single workflow, but can serve as a general integration pattern to enable system-of-systems approaches. At the same time, the study also reveals clear limitations. The current evaluation is restricted in scale, while the modelling effort is significant, especially for defining information requirements and maintaining alignments between vocabularies. Constraints are currently limited to basic validation and need to be refined to remain both precise and usable in practice. In addition, issues such as governance, trust, and security were not addressed, but may become critical in cross-organisational settings. In short, the approach can be a viable basis for coordination and traceability in decentralised AM workflows, but its practical adoption depends on reducing modelling effort, strengthening constraint definitions, and validating the approach in larger and more heterogeneous settings.

Conclusion

This paper presented a multi-domain semantic architecture for decentralised additive manufacturing workflows. The approach separates system structure, interfaces, information requirements, and provenance. The evaluation shows that this separation supports integration, comparison of results, and traceability across tools. The current work is limited to a prototype and a small use case. Future work will address larger and more distributed system landscapes in AM, where coordination may also involve agentic AI components. Building on this, simulation-based workflows that rely on explicit system and information modelling will be developed, with the goal of supporting predictive manufacturing (PM).

Acknowledgments

This work has received funding from the European Union's Horizon Europe research and innovation programme under the Pathfinder project "AM2PM" (Grant Agreement No. 101162318).

References

Abdelalim, A. M. et al. (2025). An analysis of factors contributing to cost overruns in the global construction industry. Buildings, 15(1):18.

Ahmad, M. E., Peralta Abadía, J. J., Chmelnizkij, A., Reimann, J., Hildebrand, J., Bergmann, J. P., and Smarsly, K. (2025). Ontology-based knowledge representation for wire arc additive manufacturing and composite extrusion modeling. In EG-ICE 2025, Glasgow, United Kingdom. University of Strathclyde Publishing.

Bos, F. P., Wolfs, R. J. M., Ahmed, Z. Y., and Salet, T. A. M. (2016). Additive manufacturing of concrete in construction: potentials and challenges of 3d concrete printing. Virtual and Physical Prototyping, 11(3):209–225.

Buswell, R. A., Leal da Silva, W. R., Jones, S. Z., and Dirrenberger, J. (2018). 3d printing using concrete extrusion: A roadmap for research. Cement and Concrete Research, 112:37–49.

Cabeza, L. F., Rincón, L., Vilariño, V., Pérez, G., and Castell, A. (2014). Life cycle assessment (lca) and life cycle energy analysis (lcea) of buildings and the building sector: A review. Renewable and Sustainable Energy Reviews, 29:394–416.

Charpenay, V., Lefrançois, M., Villalón, M. P., and Käbisch, S. (2023). Web of Things (WoT) Thing Description (TD) Ontology. W3C Editor's Draft. Editor's Draft, RDF vocabulary.

Deetman, A., Bos, D., Lucas, S. S., Salet, T., and Wolfs, R. (2026). A database framework for 3d concrete printing. Results in Engineering, 29:108669.

Feng, Y., Zou, Q., Zhou, C., Liu, Y., and Peng, Q. (2023). Ontology-based architecture process of system-of-systems: From capability development to operational modeling. Applied Sciences, 13(9):5419.

Gouttebroze, S., Friis, J., Hovig, E. W., and Boivie, K. (2023). Toward semantic standard and process ontology for additive manufacturing. IOP Conference Series: Materials Science and Engineering, 1281(1):012014.

Hagedorn, P., Petrova, E., and König, M. (2025). A system-of-systems approach for deploying containerized construction digital twins using linked data. In Advances in Information Technology in Civil and Building Engineering - Proceedings of ICCCBE 2024, Lecture Notes in Civil Engineering, pages 478–491.

Hager, I., Golonka, A., and Putanowicz, R. (2016). 3d printing of buildings and building components as the future of sustainable construction? Procedia Engineering, 151:292–299.

Jaskuła, K., Kifokeris, D., Papadonikolaki, E., and Rovas, D. (2024). Common data environments in construction: state-of-the-art and challenges for practical implementation. Construction Innovation.

Kedir, F. and Hall, D. M. (2021). Resource efficiency in industrialized housing construction – a systematic review of current performance and future opportunities. Journal of Cleaner Production, 286:125443.

Khoshnevis, B. (2004). Automated construction by contour crafting—related robotics and information technologies. Automation in Construction, 13(1):5–19.

Kirner, L., Jung, V., Oraskari, J., and Brell-Cokcan, S. (2024a). Enhancing robotic steel prefabrication with semantic digital twins driven by established industry standards. Automation in Construction, 167:105699.

Kirner, L., Oraskari, J., Wildemann, P., and Brell-Cokcan, S. (2024b). Internet of Construction Process Ontology (ioc) v 0.5.

Knublauch, H. and Kontokostas, D. (2017). Shapes Constraint Language (SHACL). W3C Recommendation.

Lebo, T., Sahoo, S., and McGuinness, D. (2013). PROV-O: The PROV Ontology. W3C Recommendation. W3C Recommendation, 30 April 2013.

Lefrançois, M. (2019). SAREF4SYST: An Extension of SAREF for Typology of Systems and Their Interconnections. ETSI Technical Specification TS 103 548. Version 1.1.2, ETSI Smart Applications REFerence ontology extension.

Li, C. and Petzold, F. (2025). Ontology-driven mixture-of-domain documentation: A backbone approach enabling question answering for additive construction. Buildings, 15(1):133.

Lim, S., Buswell, R. A., Le, T. T., Austin, S. A., Gibb, A. G. F., and Thorpe, T. (2012). Developments in construction-scale additive manufacturing processes. Automation in Construction, 21:262–268.

Mechtcherine, V., Bos, F. P., Perrot, A., Leal da Silva, W. R., Nerella, V. N., Fataei, S., Wolfs, R. J. M., Sonebi, M., and Roussel, N. (2020). Extrusion-based additive manufacturing with cement-based materials—production steps, processes, and their underlying physics: A review. Cement and Concrete Research, 132:106037.

Pauwels, P., Zhang, S., and Lee, Y. (2017). Semantic web technologies in aec industry: a literature review. Automation in Construction, 73:145–165.

Wu, P., Wang, J., and Wang, X. (2016). A critical review of the use of 3-d printing in the construction industry. Automation in Construction, 68:21–31.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

A MULTIMODAL HYBRID RETRIEVAL FRAMEWORK FOR DIGITAL TWINS IN CIVIL AND STRUCTURAL ENGINEERING

Marcel Aubart[1,2] Jyrki Oraskari[1,3], Jakob Beetz[1,3] and Jörg Blankenbach[1,2]
[1]Cluster of Excellence CARE, TU Dresden and RWTH Aachen University, Germany
[2]Geodetic Institute and Chair for Computing in Civil Engineering & GIS, RWTH Aachen University, Germany
[3]Design Computation, RWTH Aachen University, Germany

Abstract

Digital twin adoption is hindered by fragmented symbolic and subsymbolic construction data from heterogeneous sources. We propose a hybrid knowledge-graph and vector-database architecture that integrates a GeoSPARQL-enabled RDF store with semantic embeddings to unify technical documentation and support both precise reasoning and fuzzy retrieval. A hybrid RAG workflow on PDF-based data demonstrates improved ranking quality and contextual relevance compared to single-paradigm retrieval. This lays the groundwork for future work on infrastructure-focused digital twins. Comprehensive support for multimodal data and case studies, such as minimally invasive bridge strengthening and scaffold-free robotic assembly, is left as a planned continuation of the research.

Introduction

Digital twins have emerged as one of the most transformative paradigms in engineering, construction, and infrastructure management. They promise to provide a continually updated and data-driven representation of physical assets, enabling predictive maintenance and sustainable rehabilitation. Despite their potential, the integration of digital twins into practical workflows remains limited. A major barrier is the heterogeneity of data sources: engineering projects generate symbolic data (design models, ontologies, metadata) and subsymbolic data (images, sensor readings, point clouds). These data differ in format, structure, and semantics, making unified access and analysis challenging.

Within the Climate-Neutral and Resource-Efficient Construction (CARE) initiative, we aim to address these challenges by developing comprehensive data integration platforms. The goal of this project is to develop a unified data framework, as shown in Figure 1, that ensures both current and legacy construction data comply with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. By establishing standardized data models, metadata structures, and interoperability mechanisms, the project seeks to enable new automation- and sustainability-focused workflows across rehabilitation and construction processes.

As the first step, this paper presents a prototype implementation of the CARE hybrid symbolic–subsymbolic architecture and assesses its retrieval layer using multimodal documents containing text, tables, and embedded images. The current study aims to validate the architectural foundations and retrieval performance that underpin the broader digital twin vision of the CARE initiative. Support for multimodal data such as IFC models, point clouds, and sensor streams is planned for future work.

Within this scope, the paper addresses the following research questions:

- RQ1: To what extent does ontology-driven alignment improve retrieval accuracy and robustness under incomplete or missing metadata in infrastructure-like document corpora?
- RQ2: How effectively can a hybrid knowledge-graph + vector-database architecture support retrieval for infrastructure digital twins compared to single-paradigm retrieval-augmented generation (RAG) approaches?

Based on these research questions, the following hypothesis is evaluated:

- H1: A hybrid knowledge-graph + vector-database architecture enables more accurate and context-rich retrieval for infrastructure-like document collections than single-paradigm RAG approaches.

These research questions highlight the need for retrieval architectures that bridge heterogeneous construction data. This study therefore evaluates the proposed hybrid approach against non-hybrid RAG methods.

Background and Literature Review

Digital Twins in Infrastructure

Digital twins are virtual replicas of physical assets that integrate real-time data streams with simulation models. In infrastructure, they are used for monitoring bridges, tunnels, and buildings (Gao et al., 2021). However, most implementations remain siloed, focusing on specific data types (e.g., structural health monitoring) rather than holistic integration. This leads to the research question of developing a methodology to integrate various information

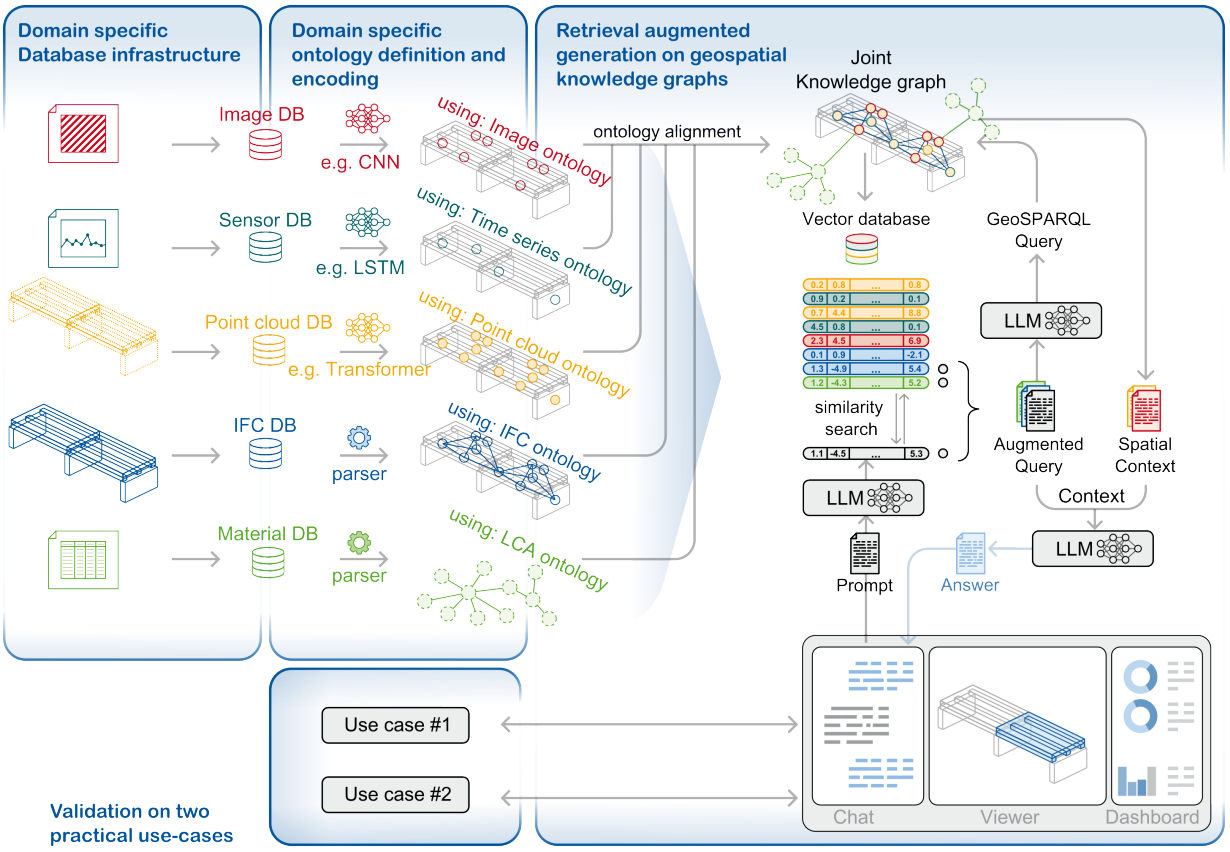


Figure 1: Schematic overview of the CARE Scheme, a hybrid data integration pipeline for construction intelligence.

and simulation models, as well as heterogeneous datasets, into a digital twin, thereby improving the confidence in the integrated simulation model and its predictions (Ye et al., 2019).

Challenges of Data Integration

Infrastructure projects increasingly rely on large, heterogeneous datasets originating from design processes, construction activities, and operational monitoring systems. These datasets differ not only in format but also in semantics, temporal resolution, and uncertainty, creating substantial challenges for integrated data management. A key difficulty arises from the coexistence of symbolic data, including CAD and paper drawings, Building Information Modeling (BIM), ontologies, environmental and geospatial (GIS) data and metadata, and subsymbolic data, such as sensor measurements, site photographs, point clouds, inspection and maintenance records, and time-series data streams (Kaden et al., 2022). Symbolic data are typically structured, semantically explicit, and governed by domain-specific rules, while subsymbolic data are continuous, high-dimensional, and often noisy or incomplete (Pauwels et al., 2017). Integrating these fundamentally different representations requires reconciling discrepancies in granularity, data quality, and interpretability, which traditional database systems are not designed to handle (Platzer, 2025). Conventional relational databases assume stable schemas and consistent semantics, making them poorly suited to the dynamic, multimodal nature of infrastructure data. As a result, workflows often become fragmented: design teams operate within GIS, BIM, or CAD environments, while monitoring systems generate sensor data that remains disconnected from the symbolic representation of the project (Sacks et al., 2018). This fragmentation limits the ability to perform cross-domain analysis, such as automated progress monitoring, structural health assessment, or predictive maintenance. For example, linking point clouds or site imagery to BIM elements requires complex spatial and semantic alignment procedures, which are prone to uncertainty and information loss (Biljecki et al., 2015).

Interoperability challenges further exacerbate the problem. Standards such as Industry Foundation Classes (IFC)[1], City Geography Markup Language (CityGML)[2], and LAS[3] offer only partial solutions, and linked-data mappings like ifcOWL (Beetz et al., 2009) fail to resolve cross-domain interoperability. Inconsistencies in metadata, versioning, and implementation across software platforms hinder seamless data exchange (Gröger and Plümer, 2012). In addition, subsymbolic data sources, particularly those generated by Internet of Things (IoT) sensors or mobile devices, often lack standardized schemata alto-

[1] https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes/
[2] https://www.ogc.org/standards/citygml/
[3] https://www.ogc.org/standards/las/

gether, complicating their integration into existing information models.

Recent research emphasizes the need for hybrid data platforms capable of handling both symbolic and subsymbolic modalities (Platzer, 2025). Approaches include graph-based data models, Semantic Web technologies, and machine-learning-based feature extraction, matching, and correlation across domains. These methods show promise in enabling digital twins, which require continuous synchronization between physical assets and their virtual representations (Fuller et al., 2020). However, significant challenges remain, including scalability constraints, inconsistent metadata standards, and the difficulty of aligning machine-interpretable semantics with real-world observations. Addressing these issues is essential for achieving robust, data-driven decision-making throughout the life cycle of infrastructure systems.

Existing Retrieval Augmented Generation (RAG) Approaches

Recent advances in Artificial Intelligence (AI) and data management have enabled the integration of symbolic and subsymbolic data in infrastructure projects. RAG is a method where an AI model retrieves relevant external information and uses it to produce more accurate, up-to-date, and grounded responses. RAG approaches include vector-, keyword-, hybrid, graph-, multi-hop, agentic, structured-data, multimodal, and summarization-centric retrieval methods. Among these, VectorRAG, GraphRAG, and HybridRAG represent the core semantic, structured, and combined retrieval strategies.

While each offers unique strengths, they also exhibit limitations that motivate ongoing research into more unified multimodal architectures.

Knowledge graphs remain a foundational technology for representing structured, semantically rich information. They encode entities and relationships in a transparent, machine-interpretable form, enabling reasoning and complex semantic queries, and add an additional layer of compatibility through ontology-based constraint checking. Emerging work centers on improving knowledge graphs with RAG pipelines. However, they rely heavily on complete and precise metadata; when annotations are missing or inconsistent, their performance suffers.

In contrast, vector databases excel at handling unstructured or subsymbolic data such as images, documents, and sensor streams. They embed data into high-dimensional vector spaces, enabling similarity search and pattern recognition without requiring explicit semantic labels. Still, recent studies indicate that vector-embedding–based RAG approaches underperform compared to GraphRAG and Hybrid GraphRAG (a.k.a. HybridRAG): GraphRAG achieves higher context and answer relevance, while Hybrid GraphRAG attains higher factual correctness (Ahmad et al., 2025). HybridRAG architectures combine vector-based retrieval with knowledge-graph reasoning and large language models (LLMs), enabling systems that leverage both latent similarity and explicit semantics, demonstrating how integrating knowledge graphs into RAG pipelines improves precision, context awareness, and explainability in real-world applications (Yuan et al., 2024). These systems represent an early yet significant step toward multimodal data spaces capable of unifying symbolic and subsymbolic data. However, applications in infrastructure are constantly evolving, with open challenges including ontology alignment, scalable graph construction, and the integration of high-volume sensor data.

In general, recent research indicates that GraphRAG-style hybrid architectures are the most promising path forward (see Table 1). They combine the strengths of knowledge graphs (semantic structure, reasoning) with those of vector databases (pattern recognition, scalability), offering a more holistic approach to data integration. Yet, substantial work remains to adapt these methods to the unique demands of infrastructure engineering, where data heterogeneity, uncertainty, and scale are particularly pronounced.

CARE Data Acquisition

The Climate-Neutral and Resource-Efficient Construction (CARE) Cluster of Excellence (EXC)[4] provides a unifying research environment for developing scalable, interoperable digital twin technologies across infrastructure domains. A central requirement for validating the hybrid symbolic–subsymbolic integration proposed in this work is the availability of real-world datasets that combine structured engineering knowledge with heterogeneous sensor observations. CARE's life-size bridge demonstrator fulfills this role by offering a controlled yet realistic environment in which multimodal data can be collected, aligned, and evaluated (Herbers et al., 2024).

The openLAB[5] research bridge is a 45m, three-span prestressed concrete structure designed explicitly as a full-scale experimental platform. Its construction integrates multiple structural typologies and intentionally embedded damage mechanisms, enabling the systematic study of deterioration processes and monitoring strategies. Each span represents a distinct research scenario: legacy-inspired defects and reduced shear capacity in the first span, state-of-the-art construction with embedded fiber-optic "Smart Tendons" in the second (Richter et al., 2024), and an innovative rapid-assembly system with Ultra-High-Performance Fiber-Reinforced Concrete (UHPFRC) joints in the third. This diversity ensures that the resulting dataset captures a wide range of structural behaviors and material responses relevant to both existing and future infrastructure.

A defining characteristic of the openLAB demonstrator is its comprehensive monitoring strategy, which begins during precast fabrication and continues throughout construc-

[4] https://www.exzellenzstrategie.de/en/clusters-of-excellence/care/

[5] https://tu-dresden.de/bu/bauingenieurwesen/imb/forschung/openLAB

Table 1: Comparison of Knowledge Graphs, Vector Databases, and Hybrid RAG Approaches

Criterion	Knowledge Graphs (KG)	Vector Databases	HybridRAG Systems
Data Type Strength	Structured, symbolic, semantic data	Unstructured, subsymbolic data (text, images, sensor streams)	Both symbolic and subsymbolic data
Semantics	Explicit, machine-interpretable	Implicit, encoded in embeddings	Combines explicit semantics with latent similarity
Reasoning	Strong logical and spatial reasoning	Weak; no native reasoning	Enhanced reasoning via KG + LLM integration
Robustness to Missing Metadata	Low; requires complete annotations	High; embeddings tolerate noise	Medium–High; vector retrieval compensates for missing metadata
Explainability	High transparency	Low interpretability	Medium–High (KG-backed explanations)
Scalability	Moderate; graph operations can be expensive	High; optimized for large-scale retrieval	Moderate; depends on pipeline design
Typical Use Cases	BIM integration, geospatial reasoning	Similarity search, anomaly detection, multimodal retrieval	Digital twins, semantic search, multimodal analytics

tion, reference operation, and controlled damage loading. The instrumentation combines global measurements, such as accelerations, inclinations, and environmental conditions, with dense local sensing, including strain gauges, displacement transducers, and nearly one kilometer of distributed fiber-optic sensors (DFOS). The DFOS network provides continuous, high-resolution strain profiles along reinforcement and concrete surfaces, enabling precise detection of cracks, stiffness changes, and localized damage. Beyond sensor streams, the dataset also includes a wide range of engineering data representations essential for digital-twin workflows, such as IFC models, Computer-Aided Design (CAD) drawings, point clouds, image collections, Portable Document Format (PDF)[6] construction plans, and tabular sensor exports (e.g., .csv files). These complementary modalities enrich the symbolic layer with geometric, semantic, and visual context, enabling cross-domain alignment and multimodal reasoning. This multimodal sensing environment produces a uniquely rich dataset that spans symbolic construction metadata, geospatial information, and high-dimensional subsymbolic sensor streams.

For CARE-DS and CARE-DIVE, the openLAB bridge serves as a critical validation case. Its heterogeneous data landscape mirrors the challenges encountered in real infrastructure systems, including inconsistent metadata, varying spatial granularity, and overlapping effects of environmental and structural influences. The demonstrator therefore enables systematic testing of ontology-driven data harmonization, multimodal embedding generation, hybrid retrieval, and uncertainty-aware reasoning. Moreover, integrating all monitoring data into an as-is digital twin provides a reference model against which hybrid queries and RAG-based workflows can be evaluated.

In summary, the openLAB research bridge constitutes a foundational dataset for the CARE ecosystem. Its combination of symbolic engineering knowledge, subsymbolic sensor observations, and lifecycle-oriented digital twin integration provides the ideal environment for developing and validating the hybrid data integration methodologies proposed in this work

[6] https://www.iso.org/standard/75839.html

Objectives

The overarching objective of this research is to develop a hybrid symbolic–subsymbolic query layer that enables holistic, multimodal data integration for digital twins. This query layer aims to unify structured, semantically rich information (e.g., ontologies, BIM models, geospatial metadata) with unstructured or subsymbolic data sources (e.g., sensor streams, images, point clouds). By bridging these two representational paradigms, the project seeks to overcome current fragmentation in infrastructure data workflows and support more robust, transparent, and scalable digital twin applications (Pregnolato et al., 2022).

This objective reflects the growing need for digital twins that can reason over both explicit engineering knowledge and real-world observations, enabling more accurate monitoring, simulation, and decision-making across the life cycle of infrastructure assets.

Heterogeneous data infrastructures include image, sensor, point cloud, IFC, and material databases. The data is processed via machine learning models and parsers, then semantically aligned using domain-specific ontologies. The resulting representations are fused into a Joint Knowledge Graph, enabling structured data generation and downstream interpretation by an LLM. Outputs are visualized through a dashboard interface that combines chat, a 3D viewer, and analytical components. Lastly, the architecture will be validated across two practical use cases, demonstrating scalable, ontology-driven decision support in infrastructure contexts.

Methodology and Work Plan

The methodological structure of this project is organized around two tightly coupled pillars: The data space, which provides the semantic and data backbone, and the hybrid query interface, which implements the hybrid symbolic–subsymbolic query layer. Each pillar is operationalized through a set of concrete work steps. The data space encompasses the development of the domain-specific data infrastructure and ontology. The hybrid query interface (DIVE) builds on this foundation to implement hybrid retrieval mechanisms and validates them on two real-world

use-cases.
Those two big pillars are further divided into four steps, each addressing a critical component of the hybrid symbolic–subsymbolic data integration framework. They can be sorted like:
Data space: Semantic and Data Infrastructure

- Domain-specific database infrastructure
- Definition and encoding of the domain-specific ontology

Hybrid query interface: Hybrid Retrieval and Application Layer

- Retrieval-augmented generation (RAG) on geospatial knowledge graphs and visualization
- Validation on two practical use-cases

Together, those four steps form a pipeline that spans semantic modeling, multimodal encoding, hybrid retrieval, and real-world validation. The workflow is intentionally modular to ensure scalability, extensibility, and compatibility with future digital twin applications.

Domain-Specific Data Infrastructure

The first work step establishes the project's technical foundation by developing a domain-specific, file-type–aware data infrastructure. Infrastructure data is inherently heterogeneous, spanning CAD/BIM models, point clouds, sensor streams, inspection images, geospatial metadata, and textual documentation. Each modality requires its own storage strategy, indexing mechanism, and metadata model. To ensure long-term accessibility and interoperability, all components of the data space rely on open, license-free formats such as IFC, GeoJSON, LAS/LAZ, and open sensor metadata standards.
The focus of this step is the construction of specialized, modality-specific databases, each optimized for its respective data type—for example:

- a spatially indexed point-cloud store
- an IFC/BIM repository,
- an image archive with embedded metadata
- a time-series database for sensor data.

These stores remain independent at the technical level, preserving their native structure and performance characteristics. Interconnectability is introduced later through semantic identifiers, not by forcing premature unification. This design ensures that downstream components—such as visualization tools (e.g., point-cloud rendering in the viewer) or hybrid retrieval pipelines—can operate efficiently on each modality.
A particular methodological focus lies on representing uncertain geospatial and temporal metadata, which is common in legacy datasets and field-collected information. We investigate unified approaches, including probabilistic spatial envelopes, temporal validity intervals, and fuzzy metadata descriptors. These representations allow downstream processes to reason about incomplete or ambiguous information without discarding valuable data.
The outcome of this step is a fully operational, open, and extensible data space that provides file-type–specific storage and indexing, supports FAIR data principles, and enables seamless interconnectability across modalities. This backbone forms the structural basis for the symbolic and subsymbolic layers developed in subsequent work steps.

Domain-Specific Ontology Encoding and Subsymbolic Modeling

The second work step builds on the data backbone by introducing the hybrid architecture's semantic and subsymbolic layers. Although symbolic representations provide structure and interpretability, they cannot fully capture the richness of unstructured or subsymbolic data such as images, point clouds, or time-series sensor streams. We therefore develop:
Domain-specific ontologies that formalize structural components, materials, interventions, sustainability indicators, and lifecycle processes. Ontology engineering follows established best practices, including a clear purpose and scope, competency questions, ontology reuse when possible, appropriate levels of abstraction, consistent naming, modular design, examples and usage patterns; SHACL-based constraint definitions, and publishing in accessible formats. Alignment with standards such as IFC, GeoSPARQL and SOSA/SSN[7] ensures interoperability with BIM, geospatial, and sensor vocabularies.
Lightweight, domain-tailored encoder models that generate compact embeddings for each modality. Examples include CLIP-style models (Radford et al., 2021) for image–text alignment and DistilBERT-based encoders for textual descriptions (Sanh et al., 2020). These embeddings enable fuzzy semantic search, similarity retrieval, and cross-modal reasoning.
The methodological approach begins with the creation of modality-specific datasets, including annotated images, inspection records, vibration or strain time-series, and textual descriptions. Depending on data availability, different training paradigms are employed:

- fully supervised learning for a well-annotated dataset
- pretraining for image-to-language tasks
- self-supervised learning for large unlabeled texts
- contrastive learning for cross-modal alignment

Each encoder is trained in isolation to ensure modularity and avoid negative transfer. The resulting models are optimized for efficiency, enabling real-time vector generation suitable for integration into vector databases and RAG pipelines down the line.

[7] https://www.w3.org/TR/vocab-ssn/

We also aim to develop an API layer that exposes embedding generation and ontology-based reasoning as services. This ensures that downstream components—such as the hybrid query interface in CARE-DIVE—can seamlessly access both symbolic and subsymbolic representations.
The outcome is a coherent semantic and subsymbolic layer that complements the data infrastructure from the previous step and forms the conceptual foundation for hybrid symbolic–subsymbolic querying.

Retrieval augmented generation (RAG) on geospatial knowledge graphs

The hybrid retrieval layer constitutes the methodological core of the architecture. Its purpose is to combine symbolic reasoning over a geospatial knowledge graph with subsymbolic similarity search in a vector database, enabling retrieval-augmented generation (RAG) that is both semantically grounded and robust to incomplete metadata.
At the symbolic level, domain concepts, structural elements, materials, and spatial relations are encoded in a GeoSPARQL-enabled knowledge graph. This representation supports semantic queries, topological reasoning, and ontology-driven inference, allowing the system to retrieve contextually relevant information even when explicit metadata is sparse or inconsistent. The graph serves as the structured backbone for aligning heterogeneous data sources and for expanding retrieval results with semantically related entities.
At the subsymbolic level, documents and other unstructured data are embedded using lightweight encoder models and indexed in a vector database. Dense similarity search provides fuzzy retrieval capabilities that complement the strict semantics of the knowledge graph. This enables the system to surface relevant content even when terminology varies, metadata is missing, or the user query is ambiguous.
The hybrid retrieval mechanism integrates both layers into a unified query pipeline. Vector search first identifies high-similarity candidates, which are then enriched through graph-based contextual expansion. The combined result set is passed to a large language model for downstream reasoning and answer generation. This design balances precision and contextual richness: vector search ensures broad coverage of potentially relevant content, while the knowledge graph injects semantic structure and domain specificity into the retrieved context.
The key innovation of the proposed approach lies in the tight coupling of semantic graph reasoning with subsymbolic similarity search. Unlike conventional RAG pipelines, that rely solely on vector embeddings, the hybrid architecture incorporates explicit ontology-driven constraints, spatial relations, enabling more interpretable and context-aware retrieval.
Although the present prototype focuses on text-based retrieval, the architecture is explicitly designed to accommodate multimodal extensions. Ongoing work explores the integration of embeddings for images, point clouds, and IFC-based geometric entities, enabling cross-modal retrieval and geospatially grounded RAG workflows. In this context, multimodal representations are aligned within a shared latent space using cross-modal contrastive learning. Modality-specific encoders—such as vision transformers for images and point-based neural networks for 3D geometry—are trained to map corresponding observations of the same structural element to nearby locations in the embedding space while separating non-matching pairs (Sun et al., 2024). This CLIP-style formulation provides a principled mechanism for bridging modality gaps and learning representations that are semantically comparable across data types. While these components are currently under development and not yet evaluated quantitatively, they establish a clear methodological pathway toward future multimodal digital-twin representations, including 2D–3D correspondences, spatially anchored observations, and multimodal retrieval workflows.

Validation and Demonstration

To ensure practical relevance, the system will be validated through two representative use cases: sustainability life-cycle assessment and structural circularity. These scenarios pose distinct but complementary challenges, allowing the framework to be evaluated across different data modalities, analytical workflows, and decision-making contexts. The validation focuses on three dimensions. First, feasibility is assessed by determining whether the system can reliably provide the precise subset of data required for each task, including checks on metadata completeness, cross-modal alignment quality, and the efficiency of retrieving relevant information. Second, robustness is evaluated by testing performance under varying data-quality conditions, such as incomplete metadata, noisy sensor inputs, and ambiguous visual observations. The hybrid architecture is expected to maintain resilience by combining symbolic reasoning with subsymbolic similarity cues. Third, transferability examines whether workflows developed for one asset type can be adapted to other asset types with minimal reconfiguration, a prerequisite for scaling the approach to large assets in the future.

Preliminary Results

The preliminary results focus on validating the retrieval layer of the proposed hybrid symbolic–subsymbolic architecture, addressing RQ1 and RQ2 through an empirical comparison of retrieval strategies on infrastructure-like document corpora enriched with embedded images and tables. The evaluation targets the core architectural assumption that combining ontology-driven graph reasoning with vector-based similarity search improves retrieval quality over single-paradigm approaches.
A first prototype implements a HybridRAG architecture by integrating dense vector search using Qdrant[8] with a structured semantic context stored in a Neo4j[9] graph

[8] https://qdrant.tech/
[9] https://neo4j.com/

Table 2: Comparison between the vector-based RAG, the graph-based RAG, and the HybridRAG

Model/Metric	Query hits	avg. MRR	avg. retrieved
VectorRAG	14.71	0.755	93.1
GraphRAG	17.34	0.699	185.7
HybridRAG	18.83	0.812	93.1

database. PDF-based documents are automatically ingested, segmented, and embedded, with visual content processed through a self-hosted Qwen3-Vl[10] model, before being indexed. Retrieval combines similarity-based candidate selection with graph-driven contextual expansion, after which a self-hosted Ollama Mistral 7B[11] large language model performs downstream reasoning over the retrieved content. To evaluate retrieval performance, HybridRAG was tested on the Open RAG Bench dataset[12], which consists of 1000 PDF documents and 3000 benchmark queries, and was compared against the VectorRAG and GraphRAG baselines. The results are summarized in Table 2. HybridRAG achieves the highest mean reciprocal rank (MRR) and query hit rate, indicating superior ranking quality and retrieval accuracy. These findings directly address RQ1, demonstrating that ontology-driven alignment improves robustness and accuracy when retrieving relevant information from heterogeneous document collections with uneven or incomplete metadata. In relation to RQ2, the comparison shows that while GraphRAG retrieves a broader contextual neighborhood, HybridRAG provides a more balanced trade-off between contextual richness and retrieval precision. By combining explicit semantic structure with subsymbolic similarity measures, the hybrid approach enables more context-aware retrieval than either paradigm in isolation. Overall, the observed improvements support H1, confirming that a hybrid knowledge-graph + vector-database architecture enables more accurate and context-rich retrieval for infrastructure-like document collections. Beyond this validated pipeline, the system is currently being extended to support multimodal capabilities. A Three.js[13]-based 3D viewer has been implemented and connected to the retrieval backend (Figure 2), establishing the technical interface for interacting with geospatial and structural data. In parallel, the document ingestion workflow is being generalized to spatial data sources such as point clouds and IFC models, with early prototyping indicating that geometric entities can be decomposed into semantically meaningful units, embedded using modality-specific encoders, and linked to symbolic graph structures. These developments are ongoing and have not yet been evaluated quantitatively; they demonstrate the architectural readiness of the proposed approach for future multimodal digital-twin applications.

[10] https://github.com/QwenLM/Qwen3-VL
[11] https://ollama.com/library/mistral:7b
[12] https://github.com/vectara/open-rag-bench
[13] https://threejs.org/

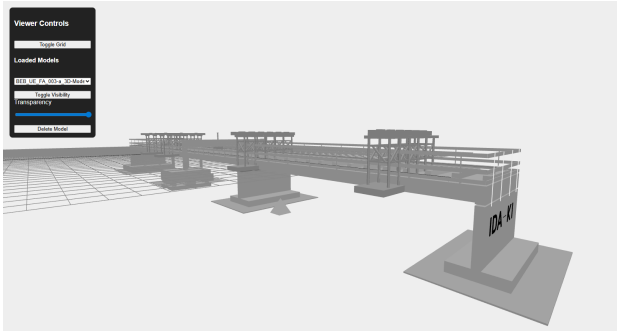


Figure 2: Image of the OpenLAB Bridge in the Three.js based web-viewer

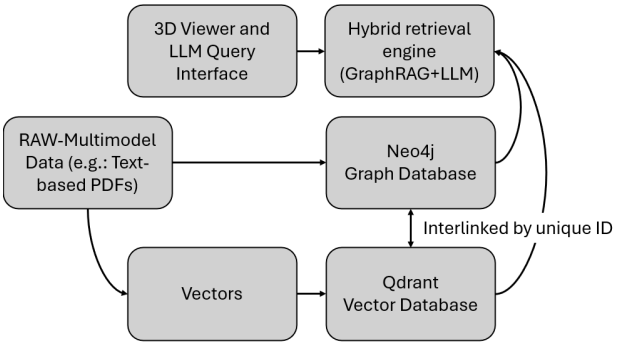


Figure 3: Overview of the current Retrieval and Generation Pipeline

Conclusion and Outlook

This paper presented a prototype hybrid symbolic–subsymbolic retrieval architecture for infrastructure digital twins and evaluated its effectiveness on infrastructure-like document corpora. The results demonstrate that ontology-driven alignment significantly improves retrieval accuracy and robustness under conditions of incomplete or heterogeneous metadata (RQ1). Furthermore, the empirical comparison confirms that a hybrid knowledge-graph + vector-database architecture outperforms single-paradigm RAG approaches in terms of ranking quality and contextual relevance (RQ2). The findings show that integrating knowledge-graph structures with vector-based semantic retrieval yields significantly more accurate and contextually richer access to technical documentation than single-paradigm approaches. These results validate the hybrid retrieval layer as a reliable and scalable foundation for more advanced digital-twin workflows. While the evaluation is conducted within the CARE ecosystem and informed by infrastructure demonstrators such as the openLAB research bridge, the core methodological contributions—ontology-driven data alignment, hybrid retrieval architectures, and RAG-based reasoning—are not CARE-specific. These components are transferable to other infrastructure domains, provided that domain-specific vocabularies and metadata schemas are adapted accordingly. Future work will extend the validated re-

trieval backbone toward fully multimodal digital twins, incorporating IFC models, point clouds, construction plans, and sensor data through cross-modal embedding alignment and geospatial reasoning. Additional validation will focus on operational sustainability use cases, including life-cycle assessment and circularity analysis, as well as on assessing the transferability of the proposed workflows across different asset types. Together, these developments aim to establish scalable, interoperable, and robust digital twin ecosystems capable of supporting data-driven infrastructure management.

Acknowledgments

This work was funded by the German Research Foundation (DFG) under the Cluster of Excellence CARE: Climate-Neutral And Resource-Efficient Construction (EXC 3115), project number 533767731.

References

Ahmad, S., Nezami, Z., Hafeez, M., and Zaidi, S. A. R. (2025). Benchmarking vector, graph and hybrid retrieval augmented generation (rag) pipelines for open radio access networks (oran). DOI: 10.48550/arXiv.2507.03608.

Beetz, J., van Leeuwen, J., and Vries, d. (2009). Ifcowl: A case of transforming express schemas into ontologies. AI EDAM, 23:89–101.

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., and Çöltekin, A. (2015). Applications of 3d city models: State of the art review. ISPRS International Journal of Geo-Information, 4(4):2842–2889.

Fuller, A., Fan, Z., Day, C., and Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. IEEE Access, 8:108952–108971.

Gao, Y., Qian, S., Li, Z., Wang, P., Wang, F., and He, Q. (2021). Digital twin and its application in transportation infrastructure. In 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI), pages 298–301. DOI: 10.1109/DTPI52967.2021.9540108.

Gröger, G. and Plümer, L. (2012). Citygml–interoperable semantic 3d city models. ISPRS Journal of Photogrammetry and Remote Sensing, 71:12–33.

Herbers, M., Bartels, J.-H., Richter, B., Collin, F., Ulbrich, L., Al-Zuriqat, T., Chillón Geck, C., Naraniecki, H., Hahn, O., Jesse, F., Smarsly, K., and Marx, S. (2024). openlab – eine forschungsbrücke zur entwicklung eines digitalen brückenzwillings. Beton- und Stahlbetonbau, 119(3):169–180. DOI: 10.1002/best.202300094.

Kaden, R., Clemen, C., Seuß, R., Blankenbach, J., Becker, R., Eichhorn, A., Donaubauer, A., Kolbe, T. H., and Gruber, U., editors (2022). Leitfaden Geodäsie und BIM : Version 3.1 (2022). Wißner-Verlag, Augsburg.

Pauwels, P., Zhang, S., and Lee, Y.-C. (2017). Semantic web technologies in aec industry: A literature overview. Automation in Construction, 73:145–165. DOI: 10.1016/j.autcon.2016.10.003.

Platzer, A. (2025). Intersymbolic ai. In Margaria, T. and Steffen, B., editors, Leveraging Applications of Formal Methods, Verification and Validation. Software Engineering Methodologies, pages 162–180, Cham. Springer Nature Switzerland.

Pregnolato, M., Gunner, S., Voyagaki, E., De Risi, R., Carhart, N., Gavriel, G., Tully, P., Tryfonas, T., Macdonald, J., and Taylor, C. (2022). Towards civil engineering 4.0: Concept, workflow and application of digital twins for existing infrastructure. Automation in Construction, 141:104421. DOI: 10.1016/j.autcon.2022.104421.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., and Sutskever, I. (2021). Learning transferable visual models from natural language supervision. DOI: 10.48550/arXiv.2103.00020.

Richter, B., Messerer, D., Herbers, M., Speck, K., Laukner, J., Gläser, C., Jesse, F., and Marx, S. (2024). Monitoring of a prestressed bridge girder with integrated distributed fiber optic sensors. Procedia Structural Integrity, 64:1208–1215. SMAR 2024 – 7th International Conference on Smart Monitoring, Assessment and Rehabilitation of Civil Structures; DOI: 10.1016/j.prostr.2024.09.168.

Sacks, R., Eastman, C., Lee, G., and Teicholz, P. (2018). BIM handbook: A guide to building information modeling for owners, designers, engineers, contractors, and facility managers. John Wiley & Sons.

Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2020). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. DOI: 10.48550/arXiv.1910.01108.

Sun, Q., Wang, J., Yu, Q., Cui, Y., Zhang, F., Zhang, X., and Wang, X. (2024). Eva-clip-18b: Scaling clip to 18 billion parameters. DOI: 10.48550/arXiv.2402.04252.

Ye, C., Butler, L., Calka, B., Iangurazov, M., Lu, Q., Gregory, A., Girolami, M., and Middleton, C. (2019). A digital twin of bridges for structural health monitoring. In Structural Health Monitoring 2019. DOI: 10.12783/shm2019/32287.

Yuan, Y., Liu, C., Yuan, J., Sun, G., Li, S., and Zhang, M. (2024). A hybrid rag system with comprehensive enhancement on complex reasoning. DOI: 10.48550/arXiv.2408.05141.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

TOWARD A BIM-CENTRIC DIGITAL TWIN FOR SUSTAINABLE MUSEUM OPERATION: THE MACA LIVING LAB CASE STUDY

Silvia Meschini[1], Viviana Vaccaro[1], Lavinia Chiara Tagliabue[1], Silvana Dalmazzone[1], Ferruccio Damiani[1], Gianluca Torta[1], Stefano Rinaldi[2]
[1]University of Turin, Turin, Italy
[2]University of Brescia, Brescia, Italy

Abstract

Digital Twins (DTs) couple BIM and IoT for sustainable building operation, yet many implementations remain siloed, hard to scale, and difficult to interpret. This paper presents a stepwise DT prototype for an environmental-education facility, integrating indoor/outdoor sensing (CO_2, PM, temperature, humidity, occupancy) with a Revit BIM model. A Dynamo–Python pipeline maps cloud time-series to shared BIM parameters, enabling in-model visualization, threshold alerts, and KPI reporting. Initial deployment delivers the sensing-and-display setup and a reusable BIM data schema ready for automated synchronization. The lasting contribution is a replicable BIM-centric implementation method for room- and zone-level monitoring in sustainable cultural buildings.

Introduction

DTs are increasingly recognised as dynamic digital representations that are continuously updated by data streams from the physical asset and used to support monitoring, analysis, and decision-making throughout the lifecycle (Grieves and Vickers, 2017; Fuller et al., 2020). In the built environment, DT research has accelerated as BIM (Building Information Modelling), while rich in geometry and semantics, remains essentially static and insufficient to represent operational conditions required in operations and maintenance (O&M) (Sacks et al., 2020; Jiang et al., 2021). A common response is to integrate BIM with Internet of Things (IoT) sensing, cloud data platforms, and analytics to produce operational DTs capable of performance tracking, anomaly detection, and optimisation (Eneyew et al., 2022).

Despite the rapid growth of BIM–IoT DT prototypes, the literature consistently highlights limitations that hinder repeatability and scaling: (i) interoperability and semantic misalignment between BIM schemas and heterogeneous IoT platforms; (ii) fragile, project-specific mapping between sensor identifiers and BIM elements; (iii) ad-hoc synchronisation pipelines that are difficult to maintain; and (iv) limited support for how stakeholders interpret and act on DT outputs (Pauwels et al., 2017; Boje et al., 2020). Recent literature confirms that these issues remain central, especially where heterogeneous data structures, non-standardised integration procedures, and weak information governance limit the transition from data collection to operational decision support (Tang et al., 2019; Mannino et al., 2021; Liu et al., 2025; Zhang et al., 2025). Several contributions argue that semantic enrichment and standards-aware information management are necessary for DT ecosystems to become reliable beyond prototypes (Boje et al., 2020; buildingSMART International, 2023). Recent work on BIM-IoT-enabled smart and sustainable buildings shows that real-time monitoring, in-model visualisation, and sensor-linked dashboards are increasingly common, but that maintainable integration logic and robust spatial contextualisation remain insufficiently clarified, particularly in existing buildings and operational settings (Valinejadshoubi et al., 2021; Piras et al., 2025). Standards and positioning efforts (e.g., Industry Foundation Classes (IFC) and ISO information management frameworks) stress that DT value depends on disciplined information modelling, governance, and transparent data exchange mechanisms rather than tool-specific integrations (ISO, 2018; ISO, 2018a; buildingSMART International, 2020). This direction is also coherent with ongoing CEN/TC 442 standardisation efforts on digital twins in the built environment, which emphasise the relationship between BIM, structured information, and interoperable exchange environments (CEN/TR 18077, 2024; EN 18162, 2026).

Cultural and educational buildings introduce additional challenges and opportunities. Retrofitting sensing infrastructure in existing museums must minimise invasiveness and accommodate visitor flow, while the DT can simultaneously support operational decision-making and public sustainability awareness through “human-in-the-loop” feedback (Turco et al., 2022). In this context, indoor environmental quality (IEQ), including thermal-hygrometric comfort and air quality, becomes a dual objective: protecting occupants (often children) and turning the building into a learning environment. For particulate matter (PM2.5, PM10) and other pollutants, internationally recognised guideline thresholds provide a robust reference for interpretability and alerting logic (World Health Organization, 2021).

This paper reports a DT setup designed as a progressive, stepwise implementation for a museum Living Lab, the Italian (MAcA) Museo A come Ambiente based in Turin, where sensing infrastructure, data management, BIM preparation, and automated synchronisation are introduced in a controlled sequence. The current phase focuses on the area where the sensing infrastructure and

visitor-facing displays are deployed, so that the end-to-end sensing-data-BIM pipeline can be validated, establishing the technical foundations required for scaling it to the full asset. While the five-layer architecture provides the overall DT framework for structuring sensing, data management, integration, digital representation, and stakeholder interaction in the museum case study, the main contribution of the paper lies in its BIM-centric implementation for maintainable room and zone level monitoring, where spatial contextualisation, progressive deployment, and stakeholder interpretability must coexist. More specifically, the contribution lies in the combination of persistent spatial identifiers, room-zone-level data binding, a writable shared-parameter schema, an incremental snapshot-to-live synchronisation logic, and a dual representation strategy in which BIM supports spatial-semantic contextualisation while external platforms support longitudinal analytics. The paper does not present a generic BIM-IoT monitoring platform, but an early-stage operational DT prototype whose advancement lies in making room-/zone-level climate monitoring more maintainable and interpretable through persistent spatial identifiers, BIM-native contextualisation, and interoperable state transfer. Accordingly, the paper illustrates (i) the BIM model and shared-parameter data schema, (ii) the configured cloud-to-BIM synchronisation pipeline, and (iii) intended use cases together with a KPI-driven evaluation strategy.

Case study and project context

The case study is the Living Lab established at MAcA, which is the first museum in Europe entirely dedicated to environmental issues, offering interactive exhibits, labs, and educational trails aimed at spreading environmental culture and supporting experiential learning. The museum (Figure 1) is in the north-west of Turin and in a post-industrial regeneration area facing Parco Dora with an overall surface of 3,500 square meters, including indoor and outdoor areas. It comprises two architectural bodies, a historic building and a newer volume, connected by an internal garden.



Figure 1: MAcA museum overview

Lying within the Po Valley, a densely populated and industrialised basin, particulate matter (PM) and other pollutants frequently exceed regulatory thresholds due to adverse meteorological and morphological conditions that reduce atmospheric mixing, especially in winter (ARPA Piemonte, 2024). Thus, MAcA is a meaningful testbed for setting up a DT focusing on indoor/outdoor microclimate and air quality awareness. Visitor-facing interpretation can connect measured values to broader environmental challenges, while operational monitoring can support comfort and risk mitigation in crowded educational spaces (World Health Organization, 2021). MAcA also provides an excellent "human–data interaction" environment as its mission explicitly targets schools and families through hands-on education. Furthermore, the museum runs structured programmes designed for school audiences. In 2024 it hosted over 34,400 visitors with 26,400 people coming from schools and groups (Associazione A come Ambiente, 2024). This is particularly relevant for a DT aimed at linking occupancy patterns and IEQ conditions (e.g., CO_2, particulate matter, thermal-hygrometric comfort) as comfort and air quality have both health and educational implications.

The development of the MaCA Living Lab is under the umbrella of the European Project NODES Spoke 4 "Digital innovation toward sustainable mountain", promoting digital technologies and services to improve sustainability and resilience in mountain territories and connected communities (ECS-NODES, 2025a). In particular, the INTERFACE flagship which aims to support the regeneration and repopulation of mountain territories through models sensitive to ecological context and participatory approaches capable of adapting to changing environmental conditions (ECS-NODES, 2025b). The MAcA Living Lab is used as a demonstrator where BIM-IoT integration, data-driven communication and monitoring workflows can be validated in a controlled public building and then transferred to other contexts (e.g., schools, cultural facilities, and tourism infrastructures), consistent with the mission of Spoke 4 (ECS-NODES, 2025a; ECS-NODES, 2025b).

Methodology and DT architecture

This work adopts a replicable methodology to set up a BIM-centric DT for sustainable cultural and educational buildings, combining (i) a stable semantic and spatial backbone in BIM, (ii) a heterogeneous IoT sensing infrastructure, and (iii) an automated synchronisation workflow that collects time-varying measurements into BIM parameters for contextualized interpretation, alerting and reporting. The approach is consistent with DT literature that frames built-asset DTs as layered cyber–physical ecosystems, where separation of concerns supports scalability and long-term maintainability (Fuller et al., 2020; Eneyew et al., 2022), and where semantic enrichment and standards-aware information management are essential to move beyond fragile prototypes (Boje et al., 2020; buildingSMART International, 2023). The long-term target is a whole-building DT, while the deployed implementation focuses on the currently sensorised ground floor as an initial deployment domain to validate mapping rules, data schema, and interpretability before scaling additional spaces. The implementation framework is designed along five steps, better described in the following sub-sections.

1. Sensing and stakeholder interfaces.

2. DT system architecture.
3. BIM model and data schema.
4. Data integration and synchronization pipeline.
5. Intended use cases and evaluation strategy.

Step 1 - Sensing and visualization interface

The first step establishes DT's physical observability and its human-facing interaction points. At this stage, the monitored variables and sensing strategy are selected to capture IEQ and occupancy-related dynamics that are relevant to comfort, air-quality awareness, and operational monitoring. In parallel, the approach defines how the DT will communicate these measurements to different stakeholder groups through appropriate interfaces, ensuring that collected data can be interpreted and acted upon rather than remaining isolated technical signals. Detailed descriptions of the deployed sensing assets and stakeholder interface components are reported in Step 2 within the DT layered architecture. This dual interface setup addresses a common barrier in DT adoption, where monitoring data exists but are not effectively translated into decision support and stakeholder understanding (O'Donnell et al., 2013; buildingSMART International, 2020).

Step 2 - DT system architecture

Step 2 defines the DT architecture as a separation-of-concerns system and specifies the data contract that links physical sensing, cloud time-series identifiers and BIM entities. Currently, the system is framed as an early-stage operational DT prototype, combining a monitored physical environment, a spatial-semantic digital counterpart, updateable data links, and user-oriented interpretative outputs, while live cloud synchronisation is still being progressively activated. Although the current implementation uses Revit and Dynamo as the orchestration environment for BIM-native write-back and inspection, the architecture is not intended as a fully centralised DT repository: longitudinal measurements remain managed in the external time-series layer, while BIM stores selected operational states for spatial contextualisation, traceability, and communication. The DT architecture is structured in five layers (Figure 2).

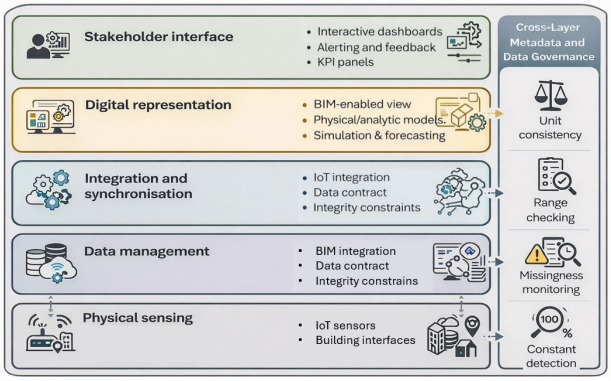


Figure 2: Layered DT architecture and data flow

This separation is widely adopted to support scalability and avoid tight coupling to vendor platforms (Fuller et al., 2020; Eneyew et al., 2022). Several contributions argue that semantic enrichment and standards-aware information management are necessary for DT ecosystems to become reliable beyond prototypes (Boje et al., 2020; buildingSMART International, 2023). In addition, the data contract includes explicit integrity constraints (e.g., unit consistency, range checks, missingness and constant-value detection) to prevent misleading interpretations and to support quality-controlled DT visualisations.

Physical sensing layer

This layer acquires IEQ and occupancy-related signals. The deployed devices measure indoor temperature, relative humidity (RH), carbon dioxide (CO_2), particulate matter (PM2.5, PM10), and occupancy indicators, complemented by outdoor boundary conditions via a weather station. The sensing stack combines KNX building automation devices and LoRaWAN (Long Range Wide Area Network) components to cover both stable IEQ monitoring and dynamic people-flow estimation. Sensor and device types, quantities, and monitored variables are summarised in Table 1 and are deployed across the sensorised domain to capture both environmental state and contextual drivers (occupancy/activity), enabling interpretation of microclimate variation as a function of use patterns. This heterogeneous configuration is designed to support both operational monitoring and educational interpretation by linking microclimate variations to occupant behaviour.

Table 1: Sensing devices deployed in the Living Lab

Sensor type	Qty	Variables measured	Technology
Weather station	1	Outdoor temperature, illuminance, wind, rain	KNX
Outdoor humidity	1	Outdoor RH	Analog (0–10V / 4–20mA) / integration to BMS
Air quality probes	5	PM2.5, PM10	KNX
Indoor multisensors	4	Temperature, RH, CO_2	KNX
Presence/activity sensors	4	Presence/activity, illuminance, noise proxy; (some nodes also expose pressure/dew point signals	KNX
IoT controller/gateway	4	Fieldbus integration (RS485), device bridging metadata	LoRaWAN + RS485
3D ToF people counters	4	People counting / flow direction	LoRaWAN

Data management layer

Sensor measurements are stored as time series in a provider-managed cloud environment. This layer decouples high-frequency ingestion from downstream consumers and provides a single authoritative data source with timestamps and metadata (e.g., sensor identifiers, variable types). The cloud layer decouples sensor ingestion from downstream consumers and supports historical queries and auditing. For the current validation phase, an exported time-series snapshot is used to test the downstream DT linkage. Currently, programmatic access to these cloud time series is under activation. The paper focuses on the integration contract (identifiers, schema and workflow) so that ingestion can be activated without re-authoring the BIM model once API (Application Programming Interface) credentials and endpoints are available.

Integration and synchronisation layer

This layer operationalises the cloud-to-BIM bridge. The proposed workflow is based on Autodesk Dynamo (i.e., visual programming) combined with embedded Python scripts to (i) query the BIM model for persistent space-level identifiers (MAcA_SpaceId), (ii) construct authenticated API requests to retrieve time-series values, (iii) validate and preprocess records (timestamp checks, missing values, unit normalisation, optional aggregation), and (iv) write validated values into BIM shared parameters. The workflow is designed to be inspectable and modular, supporting gradual extension to additional sensors and variables (Eneyew et al., 2022). In the current validation phase, the integration layer operates in a snapshot-based mode using exported time-series datasets. The same modules are designed to transition to authenticated live API queries once endpoints and credentials are available.

In practice, the integration workflow is implemented as a Dynamo graph operating within the Autodesk Revit environment, where Python scripts are used to parse time-series snapshots, apply temporal filters and threshold logic, and write aggregated results directly into BIM shared parameters associated with spaces and zones. The visual programming layer supports transparency and inspection of the data flow, while embedded scripts encapsulate reusable logic for data validation and aggregation.

Digital representation layer

This layer hosts the DT's digital counterpart as a BIM-based operational interface. The Autodesk Revit model functions as the stable spatial and semantic reference where measurements are contextualised, aggregated by zones, and visualised through native constructs such as schedules, view filters and colour-coded overrides. This aligns with research showing that spatially contextualised performance interpretation improves the usability of measured data and supports diagnostics (O'Donnell et al., 2013).

The detailed implementation of the Autodesk Revit model, including modelling scope, shared-parameter schema, sensor-to-model mapping rules, and standards-aware design choices, is presented in Step 3.

Stakeholder interface layer

This layer provides human-facing interpretation for both staff and visitors. The Living Lab deploys five 27-inch displays driven by local mini-PCs, supported by supervision/webserver components, to expose simplified indicators and narratives derived from the monitored variables. This interface layer is treated as integral to the DT ecosystem as it acts as a Human–Data Interaction (HDI) layer that exposes IEQ and occupancy dynamics in near real time, improving interpretability and bridging the gap between measurements and actionability of DT outputs (O'Donnell et al., 2013; buildingSMART International, 2020).

Table 2: summarises the deployed visualization and supervision components

Component	Qty.	Role
LCD monitor (27'')	5	Visitor facing real-time visualisation; staff situational awareness
Mini-PC	5	Local compute for visualisation and interaction
Supervision/energy management software	1	Integration layer; dashboards; device management
Webserver for remote or local interface	1	Remote or local access; integration services

Step 3 - BIM model and data schema

The third step focuses on the design of the BIM model as a semantic backbone for operational data binding and defines a parameter-based data schema to support persistent integration of environmental monitoring information. The BIM model is intentionally developed to prioritise operational interpretability and long-term maintainability rather than geometric detail, in line with operations-oriented BIM practices in facility management contexts (Sacks et al., 2020; Jiang et al., 2021). At the current stage, the model encompasses the sensorised ground-floor spaces, with explicit room and zone definitions to ensure consistent spatial contextualisation, aggregation, and reporting of monitored conditions.

Rather than establishing a one-to-one linkage between individual sensing devices and BIM elements, the adopted strategy associates monitored variables directly with spatial entities. Each space acts as a semantic container for a predefined set of environmental parameters, enabling the aggregation of sensor data at the room or zone level. This approach reduces coupling between physical devices and the BIM model, increasing robustness to sensor replacement, relocation, or changes in the sensing infrastructure, while preserving spatial meaning and continuity of historical records (Pauwels et al., 2017; Boje et al., 2020).

The shared-parameter schema reflects this aggregation-oriented logic and includes space-level parameters

representing monitored environmental quantities (e.g., temperature, relative humidity, CO_2 concentration, PM2.5, PM10, illuminance), along with a timestamp parameter recording the last valid data update. Additional boolean or state parameters are used to encode threshold-based anomaly conditions derived from recent time windows, supporting alerting and reporting without requiring persistent storage of full time series within the BIM environment. This schema enables BIM-native views, filters, and colour-coded representations once synchronised values are written to the model (O'Donnell et al., 2013). The parameters are designed to be writable by external automation scripts, enabling repeatable BIM updates without manual interaction. In the current validation phase, time-series snapshots imported from external data sources are used both to verify data-to-space mapping integrity prior to live ingestion and to test transferability toward open representations, including IFC property sets (ISO, 2018; ISO, 2018a; buildingSMART International, 2020).

The same schema was therefore mapped to a custom IFC property set (PSet_MAcA_IoT) through a DefaultUserDefinedParameterSets configuration, exporting room-level entities as IfcSpace objects with MAcA_SpaceId, timestamp, and selected IEQ variables (Figure 3). In the BAR zone, for example, Z_Bar - CO2, Z_Bar - Temperatura, and Z_Bar - Umidità were exported as Bar_CO2, Bar_Temperature, and Bar_Humidity, while the same logic was replicated for CUPOLA and VERDE. Outdoor variables remain in the time-series layer in the current phase; accordingly, the interoperable exchange model prioritises room-level indoor states through IfcSpace-linked custom properties, while external boundary conditions remain outside the primary IFC export. Currently the prototype focuses on writable implementation logic, stable spatial identifiers, and standards-oriented exchange of selected operational states rather than on bSDD for formal semantic harmonisation.

Step 4 - Data integration and synchronization pipeline

Building on the shared-parameter schema defined in Step 3, Step 4 configures the automated data synchronisation pipeline responsible for injecting time-varying environmental measurements into the BIM model. The workflow is implemented using Autodesk Dynamo in combination with embedded Python scripts and is designed to operate at the space and zone level rather than at the individual sensor level (Figure 4).

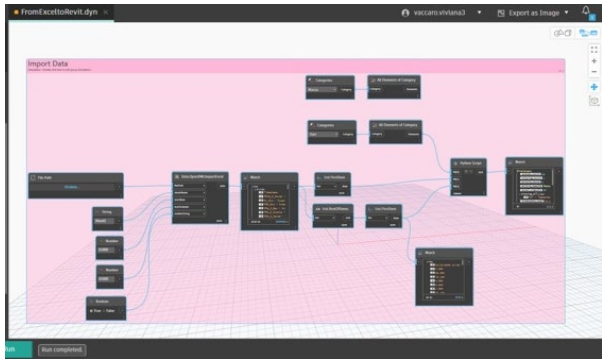


Figure 4: Dynamo based synchronization workflow

This design choice ensures robustness to changes in the sensing infrastructure while preserving the spatial interpretability of monitored conditions.

The synchronisation pipeline performs four main functions: (i) retrieval of structured monitoring data from external sources (currently via spreadsheet-based snapshots, with cloud-based time-series endpoints planned); (ii) temporal filtering and preprocessing of measurements, including timestamp parsing, unit normalisation, and aggregation over configurable time windows (e.g., rolling 24-hour periods); (iii) evaluation of threshold-based conditions to identify anomalous

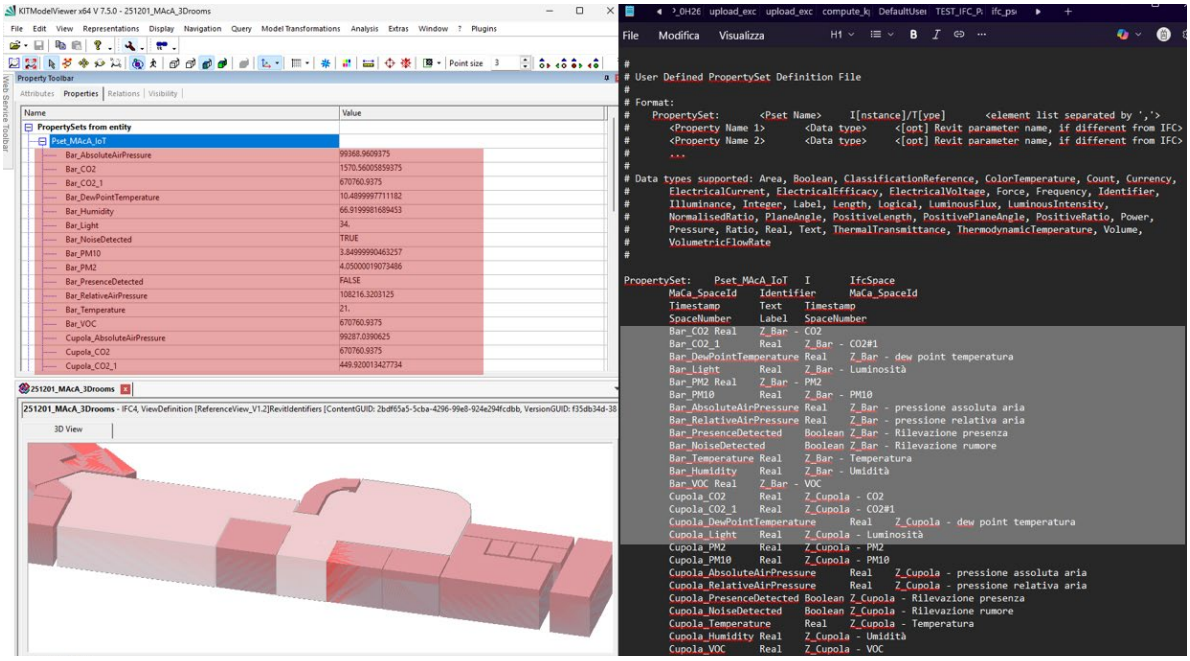


Figure 3: Custom PSet_MAcA_IoT definition and IFC-based export of room-level monitored states

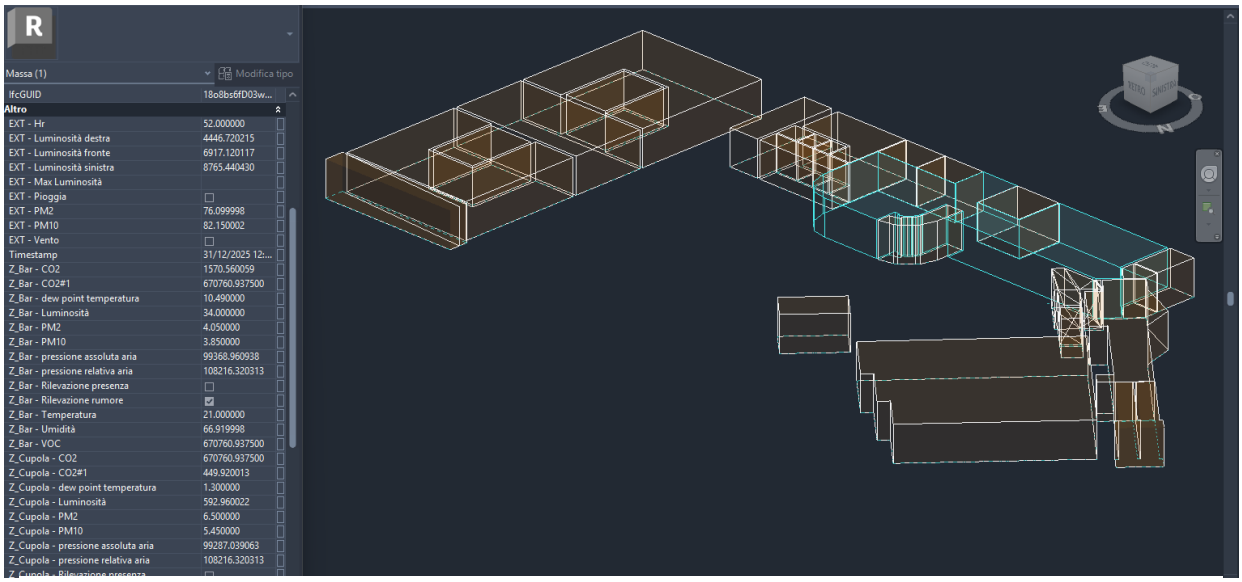


Figure 5: BIM contextualisation of monitoring data via shared parameters

environmental states; and (iv) writing of validated and aggregated values directly into predefined BIM parameters associated with rooms or zones, including measured quantities, last-update timestamps, and alert indicators. Rather than persisting full time series within the BIM environment, the pipeline selectively materialises operationally relevant summaries (e.g., maximum values over the last 24 hours and corresponding alert states). This approach enables BIM-native constructs, such as drafting views populated with automatically generated text-based annotations, view filters, and colour-coded overrides, to function as lightweight operational dashboards, supporting spatially contextualised interpretation of building performance without overloading the model with high-frequency data (Figure 5) (O'Donnell et al., 2013).

Within this workflow, Dynamo scripts directly write alert summaries and exceedance information into dedicated drafting views, allowing the BIM model to function as a self-contained, read-only operational reporting interface without relying on schedules or external dashboards.

At the current stage, live cloud access has not yet been activated; however, the synchronisation workflow is fully implemented and executable. Once data endpoints and authentication credentials are provided, the pipeline can be enabled without requiring modifications to the BIM model or its parameter schema, as the data-to-space association logic is already embedded in the operational configuration defined in Steps 2 and 3. This separation between model structure and data access supports incremental deployment and facilitates the progressive evolution of the system toward a fully operational DT. In this configuration, the time-series layer remains the authoritative source for longitudinal measurements and temporal analytics, while BIM hosts selected synchronised states for spatial-semantic contextualisation and operational interpretation.

Step 5 - Intended use cases and evaluation strategy

The last step of the methodology defines the intended DT capabilities and the evaluation protocol to be applied as integration progresses from snapshot-based validation to live synchronisation. The proposed DT supports three complementary functions: (i) operational observability of IEQ at zone level through continuous sensing; (ii) BIM-centric contextualisation, where monitored values are linked to spatial entities via persistent space-level identifiers (MAcA_SpaceId) and shared parameters; and (iii) stakeholder communication, where technical measurements are translated into readable indicators for educators, visitors, and facility staff. Evaluation is defined as an intended, KPI-driven strategy to be executed once live synchronisation is enabled, structured along three dimensions:

1. Pipeline performance and data reliability: quality criteria for time-series ingestion (latency, completeness, integrity);
2. Spatial correctness and interpretability: SpaceId mapping correctness; clarity of schedules and colour-coded views for staff;
3. Stakeholder usefulness: qualitative feedback on readability, perceived utility, and support to operational or educational decisions.

KPI logic and indicator interpretation will be grounded in recognised guidance for air quality and comfort, while preserving transparency on assumptions and data provenance.

Discussion

The first results support the feasibility of the proposed BIM-centric DT approach. A one-week monitoring excerpt sampled at 15-minute intervals was used as a testbed to verify end-to-end linkage between time-series measurements and the BIM environment, and descriptive statistics indicate meaningful room/zone-level IEQ dynamics. The test confirmed the successful population of shared parameters, timestamps, and alert-related fields from external records, showing that monitored values can be spatially contextualised in BIM without relying on full time-series storage within the model. This is relevant

because the proposed logic reduces dependence on device-specific BIM bindings and supports more robust room- and zone-level interpretation under sensor replacement or reconfiguration. The validation also confirms that data-quality assurance and metadata governance (e.g, covering variable naming, units, channel definitions, and consistent cross-zone mapping) are first-class DT requirements before thresholding logic or public-facing indicators are operationalised (Boje et al., 2020; buildingSMART International, 2023). Accordingly, the workflow incorporates integrity checks on units, ranges, missingness, and constant values to reduce misleading interpretations and false alerts. On the integration side, a proof-of-concept BIM linkage was achieved: DT fields were instantiated as shared parameters in Revit, and a Dynamo-Python workflow was configured and tested to connect the exported snapshot to BIM parameters, enabling automated population and BIM-native representations. This confirms BIM's role as a spatial and semantic backbone for operational data binding, supporting zone aggregation and "where/which space" reasoning that complements conventional dashboards and improves interpretability of measured performance (O'Donnell et al., 2013). In this respect, the contribution is not the generic coexistence of BIM and IoT, but a more maintainable operational logic in which monitored values are bound to persistent spatial entities rather than to device-specific model objects, reducing fragility under sensor replacement or reconfiguration and enabling interoperable room-level state externalisation through IFC. A key implication is that the DT should be implemented as a dual representation system: BIM-based views for spatial traceability and contextual diagnosis, and time-series dashboards for temporal analytics, comparisons, and anomaly cues. Thresholds for alerting and exceedance logic will be anchored to authoritative guidance and widely adopted standards, including WHO guidance for particulate matter (World Health Organization, 2021), (ASHRAE, 2023; ISO, 2025) for thermal comfort, and CO_2 used as a ventilation/occupancy proxy with known limitations (Lu et al., 2022). In this direction, the installed museum screens, currently limited to vendor-default static content, will be progressively repurposed to display curated near real-time dashboards aligned with acquisition frequency, using read-only visualisations, role-based access control, and content governance. Overall, the main scaling challenges are robust metadata and point naming, systematic data-quality assurance, secure and maintainable cloud access, and stakeholder-oriented visual design for public-facing displays.

Conclusions

This paper presented a BIM-centred, stepwise framework for integrating, contextualising, and operationally interpreting environmental monitoring data in a living-lab setting. Through the MAcA case study, the work showed how a DT can be implemented by using BIM not as a passive geometric repository, but as an active semantic interface for spatially grounded monitoring, alerting, and reporting. Rather than pursuing a device-centric or fully control-oriented architecture, the proposed approach intentionally operates at room/zone level, aligning data representation with practical facility-management reasoning and stakeholder interpretation.

The contribution lies not only in the use of a five-layer DT framework, but in its BIM-centric operationalisation for room- and zone-level monitoring, combining persistent spatial identifiers, a writable shared-parameter schema, an incremental path from snapshot-based validation to live synchronisation, and preliminary IFC-based transferability. In the current implementation phase, this approach has been validated through snapshot-to-BIM integration using a Dynamo-Python workflow, confirming that BIM-native views can support early operational interpretation without requiring invasive model re-authoring. The study also indicates that effective DT adoption in cultural or public buildings depends on balancing technical integration with interpretability, governance, and maintainability. In this configuration, BIM is not treated as a fully centralised repository of DT information, but as a spatial-semantic interface coupled with external time-series infrastructures and interoperable exchange mechanisms. Future work will extend live cloud ingestion, consolidate IFC-aligned interoperable exchange, and refine role-based interfaces for operational and public-facing use.

Acknowledgments

Publication part of the project NODES funded from the MUR - M4C2 1.5 of PNRR funded by the EU - NextGenerationEU (Grant agreement no. ECS00000036).

References

ASHRAE (2023) ANSI/ASHRAE Standard 55: Thermal Environmental Conditions for Human Occupancy (2023 edition).

Associazione A come Ambiente (2024) Relazione e Bilancio al 31 dicembre 2024 (Progetto di Bilancio).

Boje, C., Guerriero, A., Kubicki, S. and Rezgui, Y. (2020) Towards a semantic Construction Digital Twin: Directions for future research. Automation in Construction, 114, 103179.

buildingSMART International (2020) Enabling an Ecosystem of Digital Twins: A Positioning Paper. buildingSMART International. https://www.buildingsmart.org/wp-content/uploads/2020/05/Enabling-Digital-Twins-Positioning-Paper-Final.pdf (Accessed: 4 February 2026).

buildingSMART International (2023) Enabling an Ecosystem of Digital Twins – An Update (White Paper). buildingSMART International. https://www.buildingsmart.org/wp-content/uploads/2023/09/Enabling-an-Ecosystem-of-Digital-Twins_An-Update_Whitepaper_bSI_2023_v2.pdf (Accessed: 4 February 2026).

ECS-NODES (2025-a) 4. Digital innovation toward sustainable mountain (Spoke 4). Nord Ovest Digitale e Sostenibile (NODES). https://ecs-nodes.eu/en/4-digital-and-sustainable-mountain

ECS-NODES (2025-b) INTERFACE. Digital and participatory ecological transition for mountain communities (Flagship Project). https://ecs-nodes.eu/en/4-digital-and-sustainable-mountain/spoke-4-flagship-projects

Eneyew, D.D., Capretz, M.A.M. and Bitsuamlak, G.T. (2022) Toward smart-building digital twins: BIM and IoT data integration. IEEE Access, 10, pp.130487–130506.

European Commission, JRC (2021) Level(s) indicator 4.1: Indoor air quality. European Commission JRC. https://susproc.jrc.ec.europa.eu/product-bureau/sites/default/files/2021-02/UM3_Indicator_4.1_v1.1_37pp.pdf (Accessed: 4 February 2026).

Fuller, A., Fan, Z., Day, C. and Barlow, C. (2020) Digital Twin: Enabling technologies, challenges and open research. IEEE Access, 8, pp.108952–108971.

Grieves, M. and Vickers, J. (2017) Digital Twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. In: Kahlen, F.-J., Flumerfelt, S. and Alves, A. (eds.) Transdisciplinary Perspectives on Complex Systems. Cham: Springer.

Hosamo, H.H., Imran, M., Tchir, Cardenas-Cartagena J., Svennevig P.R., Svidt K. and Nielsen H.K. (2022) A review of the digital twin technology in the AEC-FM industry. Advances in Civil Engineering, 2022, 2185170.

Kang, T.W. and Mo, Y. (2024) A comprehensive digital twin framework for building environment monitoring with emphasis on real-time data connectivity and predictability. Developments in the Built Environment, 17, 100309.

ISO (2018) ISO 19650-1: Organization and digitization of information about buildings and civil engineering works, including BIM - Part 1.

ISO (2018a) ISO 16739-1: Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries - Part 1.

ISO (2025) ISO 7730:2025 Ergonomics of the thermal environment. Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. https://www.iso.org/standard/85803.html (Accessed: 4 February 2026).

Jiang, F., Ma, L., Broyd, T. and Chen, K. (2021) Digital twin and its implementations in the civil engineering sector. Automation in Construction, 130, 103838.

Liu, X., Antwi-Afari M.F., Li J., Zhang Y., and Manu P. (2025) BIM, IoT, and GIS integration in construction resource monitoring. Automation in Construction, 174, 106149.

Lu, X., Pang, Z., Fu, Y. and O'Neill, Z. (2022) The nexus of the indoor CO_2 concentration and ventilation demands underlying CO_2-based demand-controlled ventilation in commercial buildings: A critical review. Building and Environment, 218, 109116.

Mannino, A., Dejaco, M.C. and Re Cecconi, F. (2021) Building Information Modelling and Internet of Things Integration for Facility Management. Literature Review and Future Needs. Applied Sciences, 11(7), 3062.

O'Donnell, J., Maile, T., Settlemyre, K. and Haves, P. (2013) A visualization environment for analysis of measured and simulated building performance data. In: IBPSA Building Simulation 2013. Chambéry, France: IBPSA.

Piras, G., Agostinelli, S. and Muzi, F. (2025) Smart Buildings and Digital Twin to Monitoring the Efficiency and Wellness of Working Environments: A Case Study on IoT Integration and Data-Driven Management. *Applied Sciences*, 15(9), 4939.

Pauwels, P., Zhang, S. and Lee, Y.-C. (2017) Semantic web technologies in AEC industry: A literature overview. Automation in Construction, 73, pp.145-165.

Sacks, R., Brilakis, I., Pikas, E., Xie, H.S. and Girolami, M. (2020) Construction with digital twin information systems. Data-Centric Engineering, 1, e14.

Tang, S., Shelden, D.R., Eastman, C.M., Pishdad-Bozorgi, P. and Gao, X. (2019) A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. *Automation in Construction*, 101, pp. 127-139.

Turco, M. L., Tomalini, A., and Pristeri, E. (2021). IoT and BIM Interoperability: Digital Twins in Museum Collections. Heritage Within European Research Project, 285-292.

Valinejadshoubi, M., Moselhi, O., Bagchi, A. and Salem, A. (2021) Development of an IoT and BIM-based automated alert system for thermal comfort monitoring in buildings. Sustainable Cities and Society, 66, 102602

World Health Organization (2021) WHO global air quality guidelines: particulate matter, ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. https://www.who.int/publications/i/item/9789240034228 (Accessed: 4 February 2026)

Zhang, X., Zheng, J., Li, P., Yang, Y., Ye, Y., Causone, F. and Shi, X. (2025) A review of building digital twins: Framework and enabling technologies. *Journal of Building Engineering*, 111, 113117

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

TOWARDS SEMANTIC VOXEL-BASED REASONING FOR ROBOTIC FLEET MANAGEMENT IN CONSTRUCTION

Chu Han Wu[1, 2], Pranav Shevkar[1, 2], and Sigrid Brell-Cockan[1, 2]
[1]Cluster of Excellence CARE, TU Dresden and RWTH Aachen, Germany
[2]Chair for Individualized Production, RWTH Aachen University, 52074 Aachen, Germany

Abstract

Current robotic fleet management in construction primarily relies on 2D maps and geometry-based planners, with limited integration of process context and machine capabilities. This restricts scalability and adaptability in complex, dynamic sites. This paper proposes a semantic voxelisation framework that combines 3D voxel models with knowledge of construction processes, resources, and capabilities using IOC, BOT, and CaSkMan ontologies. The knowledge for capability-aware planning and task scheduling is queried using SPARQL and translated into domains and constraints for the planning framework. The contribution is a unified semantic-to-robotic pipeline enabling context-aware planning and scheduling for holistic fleet management.

Introduction

The increasing adoption of mobile robots and autonomous fleets in construction sites introduces new challenges for planning, scheduling, and coordination in highly dynamic, spatially complex environments. Unlike structured industrial settings, construction sites evolve continuously, involve heterogeneous resources, and impose strong dependencies between space, process, and machine capabilities. Current robotic fleet management systems predominantly rely on 2D geometric maps and local navigation constraints, which limit their ability to reason about volumetric space, task context, and resource suitability at a global planning level.

Parallel to advances in construction robotics, the construction informatics community has developed rich semantic representations of the built environment and construction processes, notably through ontologies such as the Building Topology Ontology (BOT), the Internet of Construction (IOC), and Capabilities and Skill of Machines, CaSkMan (CSS). However, these semantic models are typically used for information management, coordination, or analysis and are rarely integrated directly into task-planning and robotic fleet-management pipelines.

This paper argues that semantic voxelisation can serve as an effective intermediate representation between construction knowledge models and robotic systems. By discretising the construction site into a 3D voxel grid enriched with semantic information about space, processes, and resources, it becomes possible to perform capability-aware reasoning at multiple spatial and temporal scales. Tasks with explicit spatial and resource requirements can be matched against available voxel regions using semantic queries, while the resulting spatial constraints can be projected into representations consumable by robotic middleware such as ROS 2 and also industrial robotics.

Automated planning and scheduling are a branch of artificial intelligence that uses reasoning for generating a sequence of tasks that will lead from an initial state to a desired goal state.

We present a semantic voxelisation framework that integrates IOC, BOT, and CSS ontologies as a reasoning layer for task, process, and resource determination. The framework supports querying feasible workspaces and resource allocations via SPARQL, translating 3D semantic voxel information into 2D navigation costmaps, and providing higher-level inputs for fleet scheduling and task allocation. A prototype implementation demonstrates how semantic reasoning can inform navigation and fleet management decisions without replacing existing robotic planners.

The contribution of this work is a unified semantic-to-robotic pipeline that enables context-aware planning and holistic fleet management in construction environments, bridging the gap between construction knowledge representation and operational robotic systems.

State of the art

Fleet Management and Robotic Navigation in Construction

Current robotic fleet management systems in construction predominantly rely on 2D occupancy grids or topological maps derived from SLAM or BIM floor plans (Vega Torres et al. 2023). These representations are primarily geometry-centric and are decoupled from construction processes, task semantics, and machine capabilities. As a result, navigation and task allocation are typically treated as separate problems, handled by rule-based dispatchers or centralized schedulers (Fragapane et al. 2021).

Semantic Models and Digital Twins

Research on semantic digital twins and Linked Building Data (LBD) has introduced ontologies such as BOT, IFC-derived schemas, and construction process models to describe building topology, assets, and workflows (Boje

et al. 2020). These models are largely used for documentation, monitoring, or offline analysis. While not unheard-of, there are attempts of using semantic data and IFC information for robotic planning such as works from (Crespo et al. 2020) and (Byers and RazaviAlavi 2022).

Limitations of Existing Approaches

Despite progress in robotics and semantic modelling, current approaches remain fragmented across geometric modelling, task planning, and robotic execution, limiting their effectiveness in dynamic construction environments. In particular, the lack of integration between spatial context, process knowledge, and execution frameworks restricts context-aware and adaptive decision-making:

- Spatial representations are largely 2D and static, ignoring volumetric constraints.
- Task planning is largely a manual process and does not explicitly account for site conditions, machine capabilities, or spatial affordances.
- Semantic reasoning is typically disconnected from robotic navigation and scheduling systems, such as ROS2.

This gap motivates a unified representation that bridges 3D space, semantics, and robotic execution.

Methodology

Overall System Architecture

The proposed approach introduces a semantic voxel layer, integrating ontologies from different domains with the Construction Site Voxel ontology as an intermediate representation between construction knowledge models and robotic fleet management systems. The framework integrates a voxelised spatial model of the construction site, semantic annotations using construction ontologies, and a reasoning pipeline to derive task-, resource-, and space-aware planning inputs.

The Construction Site Voxel Ontology, CVX

The **CVX Ontology** provides a semantic representation of construction sites based on voxelised volumetric space. Its primary purpose is to bridge geometric discretisation with construction semantics, enabling capability-aware planning, scheduling, and robotic fleet management. CVX acts as an intermediate semantic layer between building topology (BOT) (Rasmussen et al. 2020), construction processes (IOC) (Kirner et al. 2024), and capabilities and skills for machine capabilities (CSS) (Christian Diedrich et al. 2022), allowing volumetric regions of a construction site to be queried, reasoned over, and translated into robotic navigation and scheduling inputs.

Unlike traditional 2D grid maps or purely geometric voxel models, CVX treats each voxel as a first-class semantic entity, capable of representing spatial location, occupancy state, temporal validity, and process relevance. This enables integrated reasoning over where tasks can be performed, when space is available, and which resources are suitable. An overall concept of the ontology and its links to other ontologies are shown in Figure 1.

cvx:Voxel

The class *cvx:Voxel* represents an atomic volumetric unit of space within a construction site. Each voxel corresponds to a bounded 3D region and serves as the fundamental spatial abstraction in CVX. Voxels are uniquely identified and persist across time, allowing them to be referenced consistently in semantic queries, scheduling decisions, and robotic planning pipelines. The *cvx:voxelAttribute* represents non-tangible, contextual, and informational properties associated with a voxel that are relevant for planning, navigation, and fleet management. Unlike geometric or topological properties, voxel attributes capture dynamic, sensed, or inferred data

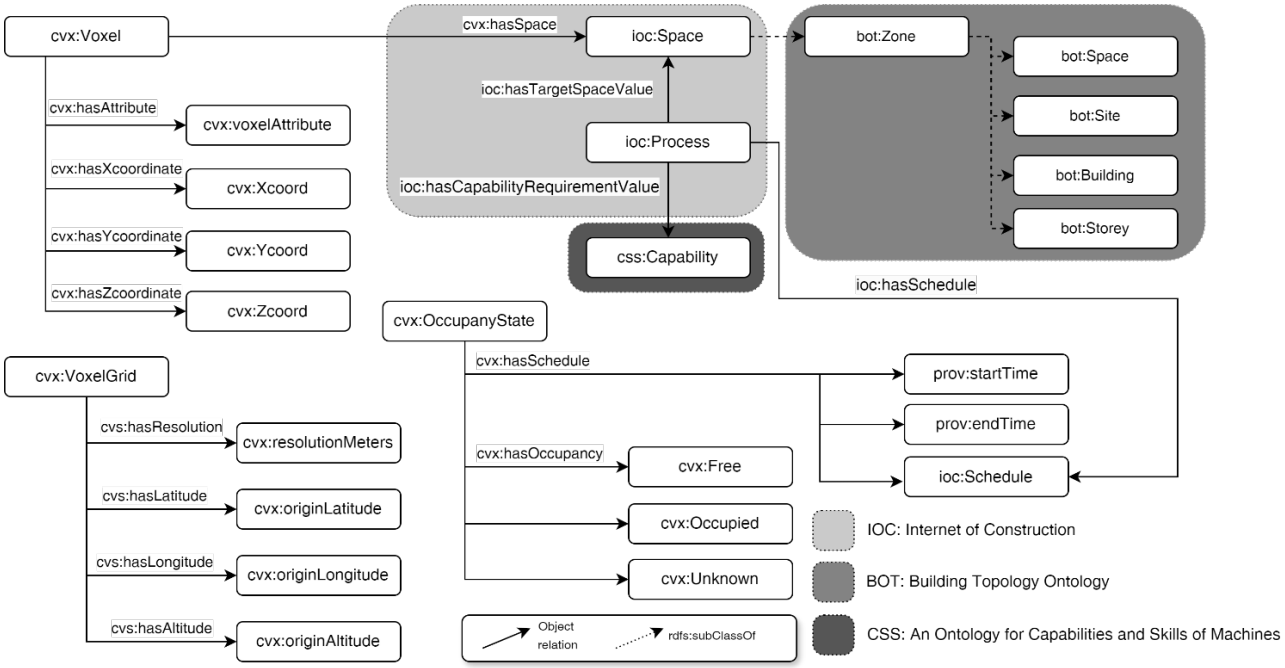


Figure 1: A first CVX ontology concept integrating IOC, CSS and BOT

that may influence robotic decision-making without constituting a physical obstruction.

cvx:VoxelGrid

The class *cvx:VoxelGrid* represents the global discretisation context within which voxels are defined. It captures parameters necessary to interpret voxel coordinates in real-world space. The key properties include *cvx:hasResolution*, specifying voxel size in meters, *cvx:originLatitude*, *cvx:originLongitude*, and *cvx:originAltitude*, defining the spatial reference origin of the grid. This design allows CVX to align voxelised construction sites with georeferenced building and site models, supporting integration with BIM, GIS, and digital twin systems.

cvx:OccupancyState

The class *cvx:OccupancyState* models the usage and availability of a voxel over time. It enables CVX to distinguish between static geometry and temporally evolving construction site conditions. A voxel's occupancy state is linked via *cvx:hasOccupancy* and can take one of three values:

- *cvx:Free* – the voxel is available for traversal or task execution
- *cvx:Occupied* – the voxel is currently in use or obstructed,
- *cvx:Unknown* – the voxel's state cannot be reliably determined.

This tri-state representation reflects real construction environments, where information may be incomplete or uncertain.

Data Storage

Voxel data and associated semantic annotations are persisted in an RDF triple store (e.g., GraphDB, Apache Jena, Fuseki, or Blazegraph) following Linked Building Data (LBD) modelling conventions. Voxels are represented as uniquely identified spatial entities and linked to building topology (BOT), construction processes (IOC), and machine capabilities (CSS). OWL DL axioms are used to capture class hierarchies, spatial affordances, and capability constraints, enabling lightweight semantic reasoning beyond purely syntactic SPARQL queries. SPARQL is subsequently employed for explicit selection and filtering of voxels that satisfy combined spatial, process, and resource conditions. Temporal aspects of voxel occupancy and usage are represented using time-indexed relations, allowing the semantic model to distinguish between current, planned, and historical states of the construction site.

AI-Based Reasoning

Planning Domain Definition Language (PDDL) (Haslum et al. 2019) is employed to model the current construction site state, the intended construction workflow, and the capabilities and status of the robots using the data from the CVX voxel ontology. This symbolic representation enables matching task requirements with available resources and reasoning over spatial feasibility to compute a valid task allocation. The planning output is then translated into an optimized execution schedule, which is consumed by the robot fleet manager. The fleet manager is responsible for low-level execution, including collision avoidance, navigation in dynamic environments, and handling unforeseen on-site changes. In the current implementation, this reasoning is limited to constraint extraction and does not yet include full symbolic plan generation.

Planning Methods

AI-based planning is formulated through two complementary specifications: a domain description and a problem definition. The domain captures the rules of the system, including available actions that are defined by preconditions and effects, objects, and logical predicates that represent world states. The problem specification provides the initial state and the desired goal state (Amer et al. 2021). A planning algorithm operates on both the domain and problem representations to synthesize a sequence of actions that transitions the system from the initial configuration to the goal (Sherkat et al. 2022).

Hierarchical Task Networks (HTN) is a technique that solves complex problems by decomposing high-level, abstract tasks into smaller and executable primitive actions (operators). It uses domain-specific methods to break down tasks, significantly reducing the search space for efficient planning in robotics (Höller et al. 2014). This enables the planner to exploit construction-specific structure and generate solutions that are aligned with established construction practices.

Temporal Planning methods generate action sequences while explicitly accounting for time-related constraints such as action durations, deadlines, and concurrency. These methods enable robots or other agents to (1) adhere to precedence relations such as "Action A must finish before Action B starts" or temporal requirements like "Complete Task before 3 PM", and (2) to coordinate concurrent activities if they do not conflict for optimizing resources. In this context, actions are modelled as temporally extended processes with defined start and end times, finite durations, and effects that may occur at specific time points or continuously over their execution. A combination of numeric and temporal planning can be used to model numeric changes over time, which is useful for resource management (e.g., the battery level) using planners such as OPTIC (Karpas and Magazzeni 2020). A key limitation of temporal planning is that action durations are typically modelled as estimates rather than guaranteed execution times. In real-world construction environments, the actual duration of robotic actions can deviate significantly due to factors such as environmental uncertainty, human intervention, dynamic obstacles, and resource conflicts. These discrepancies necessitate continuous execution monitoring and, when deviations exceed acceptable bounds, the ability to trigger online plan adaptation or re-planning to ensure timely and safe task completion (Staniaszek et al. 2025).

While this section introduces symbolic planning methods such as HTN decomposition and temporal planners (e.g., OPTIC), these are not fully integrated into the current

prototype. The present work focuses on the semantic voxel reasoning layer and its translation into navigation-relevant representations (e.g., costmaps) for execution within ROS2 Nav2. The described planning methods represent a higher-level extension, where the derived spatial and resource constraints can be consumed by external planning frameworks for task allocation and scheduling.

Voxelisation of the construction site to fleet management

Semantic Voxel to Navigation Representation

In the proposed framework, voxels are not treated merely as discretised geometry, but as semantic mediation units between the Linked Building Data layer and fleet-level navigation and task allocation. Each voxel is modelled as a first-class semantic individual with explicit spatial extent and temporal validity, linked to its surrounding context through LBD-aligned relations such as *bot:containsElement* and *bot:intersectsZone*. Rather than encoding navigation logic directly, voxels act as semantic carriers of affordances and constraints. Through reasoning over CVX, BOT, IOC, and CSS, voxels inherit process- and resource-dependent semantics, such as work zones, safety buffers, or machine-restricted areas, with *ioc:process* instances associated with CSS machines and robots so that voxel states are interpreted relative to the capabilities and roles of specific fleet members, rather than as globally fixed occupancy values.

This semantic enrichment allows a voxel to represent not only whether space is free or occupied, but how it may be used by whom and when. For example, the same ground voxel may be traversable for a transport robot, temporarily restricted during a lifting process, or costly under degraded 5G conditions encoded via *cvx:voxelAttribute*. The ontology is queried using SPARQL to extract voxels satisfying fleet-relevant constraints, producing a temporally indexed set of 3D semantic voxels projected into a 2D costmap. Costs are derived from aggregated semantic attributes such as process activity, machine compatibility, or non-tangible signals, making the costmap directly usable by fleet planners while remaining grounded in the ontology. A Hybrid-A* algorithm then finds paths between all feasible start and goal positions for each capable robot, with metrics such as distance, travel time, cost, and battery consumption used to optimise the plan. Task allocation and scheduling are calculated from these estimates, with each robot assigned a priority reflecting task criticality, temporal dependencies, and resource constraints to resolve conflicts in shared or congested areas. Continuous monitoring detects deviations from estimated costs or travel times, enabling dynamic re-planning while ensuring collision-free paths that respect process constraints and overall mission objectives.

Use Case

To evaluate the proposed semantic voxel-based reasoning and planning framework, we consider a representative robotic construction scenario involving a heterogeneous multi-robot team executing a logistic workflow on a construction site for transportation. The use case is designed to simulate real construction environments. Sequential process dependencies, heterogeneous robot capabilities and coordinated multi-robot actions. The scenario considers a dynamic construction site where material handling and safety monitoring are performed by a heterogeneous robotic fleet, consisting of:

- 2 Unmanned Guided Vehicles (UGVs) dedicated to material transport within the site.
- One UGV dedicated to human presence monitoring, ensuring that restricted areas remain free of personnel during automated operations.

The site is represented as a voxelised three-dimensional spatial grid, where each voxel captures semantic, temporal, and environmental properties. The voxel grid forms the basis for navigation planning, task allocation, and safety enforcement. The use case serves as a concrete demonstration of how semantic voxel layer, *cvx* enables context aware task allocation and fleet coordination beyond conventional map-based approaches. In the present prototype, fleet coordination is demonstrated at the level of shared spatial constraints, task-dependent zone reservations, and robot-specific access conditions derived from the semantic voxel model. The validation therefore emphasizes coordination support through task feasibility and time-indexed navigation representations, rather than a full experimental benchmark of decentralized or reactive multi-robot coordination behaviours.

Voxel Grid Generation

The construction site is modelled and discretised into a 3D voxel grid using a custom C# script, in which each voxel represents a bounded volumetric region of space. In addition to geometric information, the script extracts and associates semantic metadata, including building elements derived from IFC models and machine-related information. The resulting A-Box is then translated to Turtle (.ttl) format and uploaded to the triplestore. Each voxel or voxel-based zone is subsequently assigned a contextual state, including:

- Occupancy state (Free, Occupied, Unknown)
- Process allocation (reserved by planned or executing tasks)
- Environmental metadata, i.e., 5G signal strength
- Temporal validity

Voxels are generated deterministically based on grid coordinates and world-space positions, ensuring stable identifiers across different systems. Figure 2 depicts the construction model, and Figure 3 shows the construction site after voxelisation. From this voxelised spatial–temporal representation, a binary navigation map is generated by querying voxels that are semantically classified as free at a specific time instant, as illustrated in Figure 4.

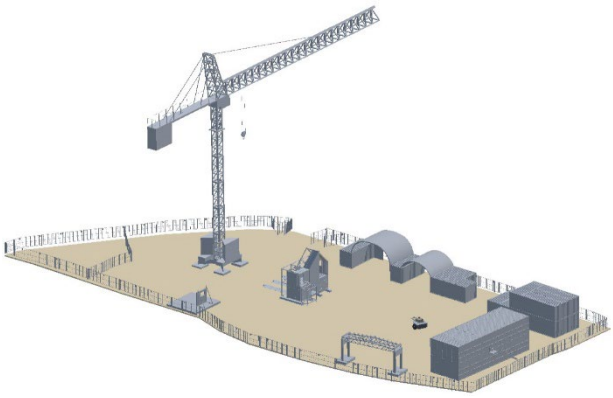


Figure 2: A model of the construction site



Figure 3: A voxelised construction site.

5G Signal Strength as a Spatial Cost Factor

A site-wide 5G signal strength map is integrated into the voxel grid as an environmental attribute. Signal strength varies spatially due to obstructions, distance to base stations, and site topology. This information is used to derive a communication-aware navigation cost map. The on-site signal strength is recorded using a Qualcomm



Figure 4: A 2D costmap queried from the knowledge graph.

SM8450 MTP (Mobile Test Platform), tested on Ericson NR Sub6 RU 3700MHz radio unit. Voxels with weak 5G coverage incur higher traversal cost. Transport UGVs prefer routes that maintain stable connectivity to ensure:

- reliable task updates
- safety monitoring
- remote intervention if required

By representing 5G signal strength as a spatially and temporally grounded attribute, the ontology enables context-aware path planning without hard-coding communication assumptions into the navigation stack. Figure 5, the top left of the figure shows a real test on the testbed using the MTP. On the right of the figure shows the distribution of the 5G signal strength. As tested in this paper (Behjati et al. 2022), it is shown that any Reference Signal Received Power (RSRP) above the value of -79dBM is able to provide a reliable connection for navigation and human observation, it is set that any signal strength less than the value of -80dBm has a high cost in the navigation map as shown in. The data is integrated into the knowledge graph and queried as voxels, the output is shown at the bottom left of Figure 5.

Robots, Capabilities, and Skills

At the semantic level, robots are modelled as resources that provide capabilities, which are realized through skills

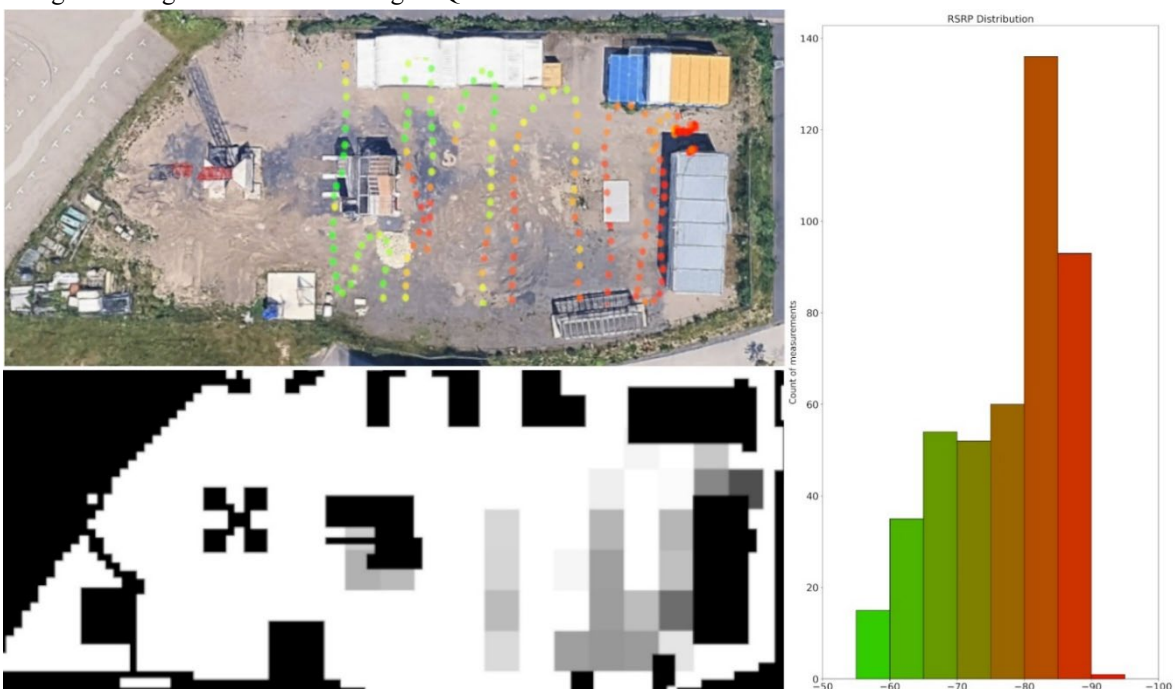


Figure 5: Top, Result of real test of the 5G signal strength on the test site. Bottom, a 2D costmap queried from the knowledge graph with 5G signal strength attribute. Right, a histogram of the data distribution.

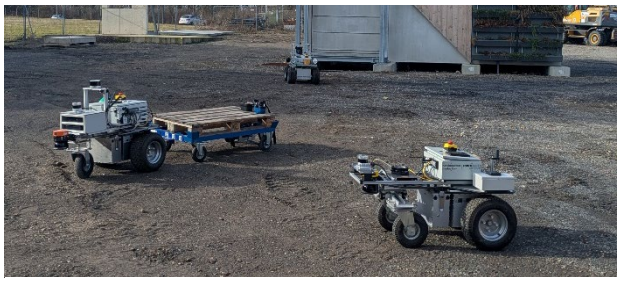


Figure 6: UGVs responsible for transport (left) and human observation

and accessed via skill interfaces. The two transport UGVs provide a Transport Capability, realised through:

- Locomotion skills (for navigation)
- Towing skills (for pallet dragging)
- Payload handling skills (for coupling and release)

The monitoring UGV provides a Safety Monitoring Capability, realised through:

- Locomotion skill (for navigation)
- Perception skill (for human detection)

enabling it to detect human presence within restricted zones. This separation allows tasks to be assigned based on capabilities, while execution selects appropriate skills depending on spatial and environmental conditions. Figure 6 shows the robots that are used for transportation and human observation. The robot on the left has a pallet that the robot can couple to and used to transport items while the robot on the right has an extra 3D LiDAR scanner and on-board computer with GPU for human detection algorithm.

Real-Time Updates and Consistency

The framework supports dynamic construction environments through incremental, event-driven updates of the semantic voxel model. Changes detected from sensors such as camera, LiDAR, or robot feedback are mapped to affected voxels, updating occupancy states, temporal validity (*cvx:hasOccupancy*), and contextual attributes without recomputing the full grid. These updates are propagated to the RDF store. The planning layer maintains consistency through triggered replanning and rolling-horizon execution, where only affected regions are reconsidered. At the execution level, the fleet manager continuously updates navigation costmaps derived from the semantic voxels, ensuring that robot behaviour remains aligned with the latest site conditions while preserving responsiveness.

Spatial-Temporal Map via contextualised working space

The proposed system maintains a spatial-temporal representation of the working space by contextualising spatial occupancy and 5G signal strength within a unified knowledge graph. Occupancy is updated from robotic perception and human-in-the-loop annotations, while communication data is integrated as a dynamic environmental layer. Unlike conventional layered costmaps, these attributes are modelled as first-class semantic entities with explicit temporal validity, enabling consistent querying of historical, current, and predicted site states.

Based on this contextualised representation, robotic task planning and execution are performed using standard frameworks such as Nav2 in ROS2. The knowledge graph acts as an intermediary semantic layer, where relevant spatial subsets are selected via SPARQL queries and lightweight semantic inference to identify regions satisfying task- and capability-specific requirements. These are then translated into navigation-compatible representations (e.g., occupancy grids or costmaps) for a given time horizon and task context, enabling capability-aware task allocation, temporal conflict avoidance, and context-aware navigation where traversal costs reflect both geometric and non-tangible factors.

Voxel-based reasoning refers to the process of evaluating semantically enriched voxels to derive planning-relevant constraints for task allocation and navigation. Each voxel encodes spatial, temporal, process, and contextual attributes, such as assigned processes, zones, machine capability requirements, occupancy state, and 5G signal strength. For example, a transport task requiring a UGV with a specific payload capacity will only be allocated to a voxel route where the associated capability constraints are satisfied, the occupancy state is unobstructed for the relevant time window, and signal strength exceeds the threshold required for teleoperation. This directly shapes both task allocation and the costmap weights used for navigation, meaning semantic voxel attributes are not merely annotations but active inputs for planning.

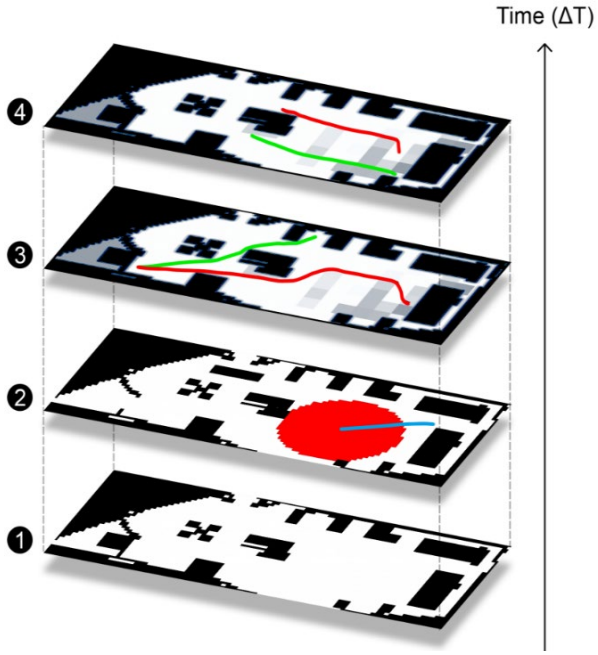


Figure 7: A voxel-based costmap for navigation queried from the knowledge graph at discrete time instance for initial navigation (1), human observation (2) and multiple and parallel processes (3,4).

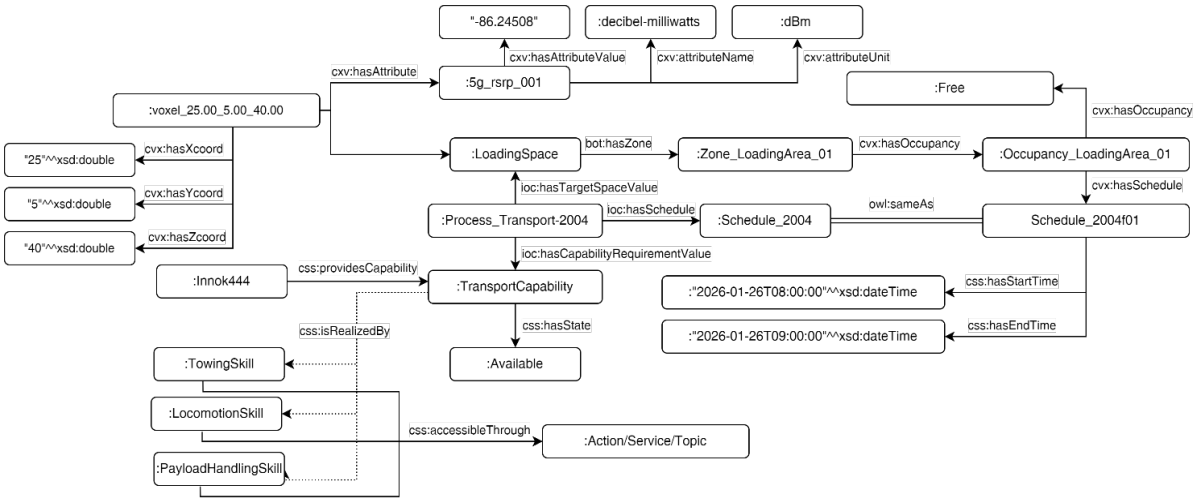


Figure 8: A excerpt of the instantiated voxel, linked to process, zones and machine capabilities.

As tasks are executed, routes, execution states, and temporal dependencies are written back into the knowledge graph, closing the perception–planning–execution loop. This supports continuous refinement of the spatial-temporal model and coordination among multiple agents operating in the same environment.

Figure 7 illustrates snapshots of navigation maps at different time periods. Map 1 shows a conventional binary occupancy map, while Map 2 incorporates human observation zones and a UGV navigation path. Maps 3 and 4 include 5G signal strength with two UGVs assigned transport tasks at different time intervals, where the planner avoids low-signal regions. While the resulting maps resemble layered costmaps (Lu et al. 2014), the key distinction is that this framework generates time-indexed spatial representations, rather than aggregating layers into a single static grid. Figure 8 shows an instance of a voxel linked to an IOC process and BOT zone, with the process associated with CSS-defined machine capabilities.

Conclusion

This paper presents an initial concept for a semantic, spatial-temporal framework for contextualising robotic working spaces in dynamic construction environments. By integrating occupancy states, task semantics, and attributes into a knowledge graph, the proposed approach enables robots to reason over space not only as a geometric entity but as a temporally evolving and context-dependent resource. This allows navigation and transport planning to move beyond static map assumptions and towards situation-aware decision-making.

The contextualised working space acts as a semantic intermediary between perception, planning, and execution. Standard robotic navigation stacks, such as Nav2 in ROS 2, can be leveraged without modification by exporting task- and time-specific spatial representations derived from the knowledge graph. In parallel, execution feedback and transport process updates are written back into the graph, enabling continuous synchronisation between the physical site and its digital representation. This bidirectional coupling supports coordination among heterogeneous agents, including UGVs and human observers, within a shared operational model. However, the current implementation focuses on generating semantically grounded constraints and navigation representations, while full integration with symbolic planning methods such as HTN or temporal planners (e.g., OPTIC) remains part of future work.

By treating spatial regions as first-class semantic entities rather than passive costmap layers, the framework supports traceability, explainability, and extensibility across planning horizons and task types, providing a foundation for planning under real-world constraints common to construction sites, such as dynamic obstacles, intermittent connectivity, and human presence.

Future Works

As a first approach, the proposed framework establishes the groundwork for a structured research roadmap. The following directions outline the planned extensions to mature the system toward real-world deployment. Several extensions can be pursued to further enhance the proposed spatial-temporal framework and broaden its applicability in real-world construction environments.

First, the framework can be extended through formal integration with symbolic planning models, such as Hierarchical Task Networks (HTN) and classical PDDL. Spatial-temporal regions represented in the knowledge graph can be abstracted into planning predicates, enabling high-level task decomposition and constraint-aware reasoning. HTN models are particularly promising for representing construction workflows, where transport, assembly, and inspection tasks follow structured yet adaptable procedures. Such an integration would allow semantic spatial constraints and communication conditions to directly influence task planning and execution at the symbolic level.

Second, the 3D nature of the voxel representation can be further exploited beyond conventional 2D costmap applications. Extending the framework to full 3D spatial–temporal planning would enable coordination of heterogeneous platforms operating across multiple elevation layers, including quadrupedal mobile robots capable of stair traversal, automated cranes transporting materials across building stories, and aerial robots operating in the airspace above the site.

Third, with sufficient data, learning-based occupancy prediction becomes feasible. Graph Neural Networks (GNNs) can be leveraged to model spatial-temporal dependencies between adjacent regions, robotic activities, and construction processes. By conditioning predictions on process states and task progress, the system could anticipate future occupancy changes and proactively adjust navigation and scheduling strategies. This would shift the framework from reactive planning towards predictive and anticipatory behaviour.

Finally, the proposed approach can be generalised into a knowledge-driven AI fleet management framework. By unifying semantic spatial representations, communication context, task models, and execution feedback, the system can serve as a coordination layer for heterogeneous robotic fleets. Such a framework would support dynamic task and resource allocation, and human-robot coexistence across large-scale sites, positioning the knowledge graph as the central decision backbone for adaptive, multi-agent construction robotics. Future evaluation would include explicit multi-robot coordination experiments, such as shared corridor negotiation, synchronized task execution, and conflict resolution in overlapping work zones, to quantify the impact of semantic voxel framework on coordination performance.

Acknowledgement

This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 3115 – 533767731.

References

Amer, Fouad; Koh, Hui; Golparvar-Fard, Mani (2021): Automated Methods and Systems for Construction Planning and Scheduling: Critical Review of Three Decades of Research. In Journal of Construction Engineering and Management 147. DOI: 10.1061/(ASCE)CO.1943-7862.0002093.

Behjati, Mehran; Zulkifley, Muhammad Aidiel; Alobaidy, Haider A. H.; Nordin, Rosdiadee; Abdullah, Nor Fadzilah (2022): Reliable Aerial Mobile Communications with RSRP & RSRQ Prediction Models for the Internet of Drones: A Machine Learning Approach. In Sensors (Basel, Switzerland) 22 (15). DOI: 10.3390/s22155522.

Boje, Calin; Guerriero, Annie; Kubicki, Sylvain; Rezgui, Yacine (2020): Towards a semantic Construction Digital Twin: Directions for future research. In Automation in Construction 114, p. 103179. DOI: 10.1016/j.autcon.2020.103179.

Byers, Gareth; RazaviAlavi, SeyedReza (2022): Layout modelling of the built environment for autonomous mobile robots using Building Information Modelling (BIM) and simulation. In mocs, pp. 201–208. DOI: 10.29173/mocs283.

Christian Diedrich; Alexander Belyaev; Rafi Blumenfeld; Juergen Bock; Stephan Grimm; Jesko Hermann et al. (2022): Information Model for Capabilities, Skills & Services. DOI: 10.13140/RG.2.2.30098.53440.

Crespo, Jonathan; Castillo, Jose Carlos; Mozos, Oscar Martinez; Barber, Ramon (2020): Semantic Information for Robot Navigation: A Survey. In Applied Sciences 10 (2), p. 497. DOI: 10.3390/app10020497.

Fragapane, Giuseppe; Koster, René de; Sgarbossa, Fabio; Strandhagen, Jan Ola (2021): Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda. In European Journal of Operational Research 294 (2), pp. 405–426. DOI: 10.1016/j.ejor.2021.01.019.

Haslum, Patrik; Lipovetzky, Nir; Magazzeni, Daniele; Muise, Christian (2019): An Introduction to the Planning Domain Definition Language. Cham: Springer International Publishing.

Höller; Daniel, ller; Gregor, Behnke; Pascal, Bercher; Susanne, Biundo (2014): Language Classification of Hierarchical Planning Problems. In : ECAI 2014: IOS Press (Frontiers in Artificial Intelligence and Applications).

Karpas, Erez; Magazzeni, Daniele (2020): Automated Planning for Robotics. In Annu. Rev. Control Robot. Auton. Syst. 3 (1), pp. 417–439. DOI: 10.1146/annurev-control-082619-100135.

Kirner, Lukas; Oraskari, Jyrki; Wildemann, Peter; Brell-Cokcan, Sigrid (2024): Internet of Construction Process Ontology (ioc) v 0.5: Zenodo.

Lu, David V.; Hershberger, Dave; Smart, William D. (2014): Layered costmaps for context-sensitive navigation, pp. 709–715. DOI: 10.1109/IROS.2014.6942636.

Rasmussen, Mads Holten; Lefrançois, Maxime; Schneider, Georg Ferdinand; Pauwels, Pieter (2020): BOT: The building topology ontology of the W3C linked building data group. In Semantic Web 12 (1), pp. 143–161. DOI: 10.3233/SW-200385.

Sherkat, Shermin; Wortmann, Andreas; Wortmann, Thomas (2022): Potentials of Symbolic AI Planning for Construction.

Staniaszek, Michal; Brudermüller, Lara; You, Yang; Bhattacharyya, Raunak; Lacerda, Bruno; Hawes, Nick (2025): Time-bounded planning with uncertain task duration distributions. In Robotics and Autonomous Systems 186, p. 104926. DOI: 10.1016/j.robot.2025.104926.

Vega Torres, M. A.; Braun, A.; Borrmann, A. (2023): OGM2PGBM: Robust BIM-based 2D-LiDAR localization for lifelong indoor navigation. Available online at https://arxiv.org/pdf/2308.05443.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

STRUCTURAL HEALTH MONITORING OF BRIDGES: INTEGRATING OPENBIM, GIS, AND IOT IN THE BREGANZE BRIDGE CASE STUDY

Idriss Tchaheu Tchaheu ,Francesca da Porto ,Andrea Gennaro, Giorgia Marcellino, Carlo Zanchetta
University of Padua, Italy

Abstract

Current bridge Structural Health Monitoring (SHM) workflows suffer from semantic disconnection, missing metadata validation, and inefficient model-updating. Addressing these deficiencies, this paper introduces a cloud-based platform. The developed workflow uses Information Delivery Specifications (IDS) for metadata integrity, nests sensors and hosts via IfcElementAssembly and IFC model updates using IfcOpenShell. Validated through the Breganze Bridge case study, the platform enables 3D model visualization, sensors measurements plots and threshold-based alerting. The lasting contribution is an OpenBIM SHM framework that integrates metadata validation, sensor-host linking, and threshold evaluation, bypassing proprietary lock-in and the dual-database integration step from model updating workflows.

Introduction

Bridges increasingly require structural health monitoring (SHM) to ensure their safety, operational performance, and long term reliability, particularly as they age and are exposed to growing load demands(Sun et al., 2025). The Architecture, Engineering, and Construction (AEC) industry has seen the widespread adoption of Building Information Modeling(Panah & Kioumarsi, 2021), with recent studies successfully demonstrating the integration of BIM and SHM by using 3D models for sensor placement and data visualization(Biagini et al., 2023; Huang et al., 2024; Li et al., 2022).

However, these BIM-SHM studies prioritize 3D geometric representation without verifying that the model and sensors contain the necessary metadata (properties, materials). Furthermore, although sensors are installed in close physical proximity to structural elements, they remain semantically disconnected entities within the BIM model, lacking formal relationships. Existing workflows typically plot database values without updating the sensor elements in the BIM model itself. Few studies have addressed model updating. Huang et al. (2024) demonstrated updates within Revit using Dynamo, while Marcellino et al. (2025) proposed IFC-based workflow with dual-database integration and file regeneration that may constrain performance for high-frequency updates (e.g., intervals shorter than one hour), requiring appropriate optimization strategies.

This work investigates the development of a cloud-based platform for Structural Health Monitoring of bridges, validated through a real-world case study of the Breganze bridge. The core contribution is a methodology with four integrated components. First, the integration of Information Delivery Specification (IDS) to validate BIM model metadata integrity before integration. Second, semantically nests sensor instances within their host structural elements using IfcElementAssembly, establishing explicit machine-readable relationships that transform sensors from isolated objects into integral components of the monitored structure thereby enabling automated queries and contextual interpretation of monitored elements. Third, updating the IFC model by directly querying the sensor database and modifying the IFC file using IfcOpenShell, thereby overcoming the proprietary lock-in of Revit-dependent workflows and bypassing the dual-database integration and IFC model regeneration step required by prior work. Finaly, visualizes the enriched model and sensor data in a cloud-based platform with BIM-GIS-IoT integration, where users can access the spatial context of the bridge model within the GIS MapLibre open-source interface and switch to other interfaces to interact with the 3D BIM model and visualize sensor measurements enabling to understand and evaluate localized structural element behavior such as displacement and tilt angles as well as environmental conditions including water level, relative humidity, and temperature.

The structure of this paper is as follows: a literature review on SHM, BIM-SHM integration, and BIM-GIS integration for SHM; followed by a description of the adopted methodology; then the presentation of the applied project case study; and finally, the conclusion.

Literature Review

Structural Health Monitoring (SHM)

Structural health monitoring (SHM) is defined as the process of implementing an autonomous damage detection strategy for aerospace, civil, and mechanical engineering infrastructure. It involves the installation of sensors, at key locations along the structure. These sensors monitor a range of variables, including loads, strains, displacements, accelerations, temperatures, and chemical composition of concrete and steel (Figueiredo et al., 2013). SHM has been extensively applied in several

engineering fields to track, identify and measure the responses of structures (Gharehbaghi et al., 2022).

Evolution from Traditional BMS to BIM-Based Monitoring

Despite the availability of monitoring technologies, traditional Bridge Management Systems (BMS) often lack 3D geometric representation and rely on manual data entry (Buuveibaatar et al., 2025). Recent studies suggest that the integration of BIM and SHM provide more sophisticated capabilities for bridge health monitoring addressing the limitations of conventional SHM approaches(Angelosanti et al., 2023; Huang et al., 2024).

Panah and Kioumarsi (2021) highlighted how BIM facilitates SHM by making sensor data accessible, practical, and comprehensible, offering improved means of visualization, interpretation, and documentation. Furthermore, having a 3D model is recognized as essential for BIM-SHM integration. For existing infrastructure, many studies rely on the Scan-to-BIM approach to recreate accurate digital representation for integration with SHM(Biagini et al., 2023; Fawad et al., 2025). As Huang et al. (2024) demonstrated, a 3D virtual model aligned with physical sensors enables precise sensor integration and allows for the colored visualization of strain distribution and residual fatigue life directly within the BIM environment, facilitating a deeper understanding of local bridge component behavior.

However, early implementations of BIM-SHM integration often relied on proprietary formats to build either desktop or web-based platforms. For instance, Li et al.(2022) proposed a BIM-SHM workflow centered on the 3D Revit model as a visual reference, where sensor data from the bridge is managed through a SQL Server database. A Revit plugin developed in C# allow users to view real-time measurement plots in conjunction with the 3D model, rather than updating the model geometry itself. Similarly, Fawad et al. (2023) developed a cloud-based IoT platform for bridge SHM that used a 3D Revit model, where clicking on the represented sensor objects enables access to real-time visualization dashboards for strain, temperature, and acceleration data. Expanding upon previous approaches toward model updating, Huang et al. (2024) employed a methodology utilizing a 3D Revit model, a MySQL database, and MATLAB for processing sensor data. In this framework, Dynamo was used not only to display measurement plots but also to update the instance parameters of sensor elements in Revit with real-time strain and residual fatigue values, ensuring the 3D model remained synchronized with the physical structure. While these studies demonstrate BIM-SHM integration with some plotting measurements and one going beyond to update the model they all share the limitation of relying on proprietary data formats. This dependency creates software lock-in and presents challenges for interoperability.

Among the open standards that support vendor-neutral interoperability across the Architecture, Engineering and Construction (AEC) industry, the sector relies primarily on the Industry Foundation Classes (IFC) standard used to represent both geometry and semantic data. Originally developed with a focus on buildings, the schema has evolved to integrate infrastructure domains as well. This evolution is reflected in the IFC 4.3.2 release published in 2023 by buildingSMART International, which introduces several bridge types and their associated structural elements (Teixeira et al., 2024). In 2024, IFC 4.3 became officially ISO 16739-1:2024. Despite the international recognition and widespread adoption, recent research highlights limitations of IFC bridge applications. Teixeira et al. (2024) point out that many commercial structural analysis software tools still lack full support for importing IFC bridge models; to prevent data loss, they propose a dedicated tool for interpreting bridge element semantics, geometry, and material properties. Additionally, Nakamura et al. (2025)observe that IFC does not currently provide schemas for maintenance and inspection information essential for bridge management. Lee et al.(2025) further notes that the IFC schema lacks definitions for the entities and property sets needed to integrate national design standards, as in the case of elements such as joints and approach slabs particularly common in Korean bridges, leading to the definition of custom extensions to meet local infrastructure requirements. Alongside these limitations in IFC bridge applications, the IFC schema also fails to represent the overall sensor network of an SHM system. Smarsly & Tauscher (2015) noted that while basic sensor types can be mapped, monitoring-related information such as the configuration, interaction protocols, and network topologies of sensors cannot be formally described using standard IFC entities. To address this gap, they proposed a conceptual approach for extending the IFC object model for semantic modeling of monitoring-related information. Building on this work, Theiler & Smarsly(2018) proposed the IFC Monitor schema extension, which defines dedicated entities for sensor nodes, sensor networks, and SHM systems, along with an enumeration for network topology. Their extension demonstrates that IFC can be expanded to represent not only individual sensors but also the topological structure of sensor networks. However, despite this advancement, the validation of the IFC Monitor remains limited on a laboratory-scale prototype of a shear frame structure, rather than a real building or infrastructure asset.

BIM-GIS Integration for SHM

While BIM-based SHM approaches have demonstrated advantages in visualizing and managing sensor data, their limitation lies in the lack of geographical context. As Bjelotomić Oršulić et al. (2024) argue, the integration of GIS and BIM is becoming increasingly important for the digitalization of the built environment. In the context of infrastructure monitoring, Gao et al. (2024) highlight that GIS integration into a digital platform allows operators to examine a structural element in the BIM model while simultaneously assessing its spatial context through the GIS interface. This spatial awareness is important for effective SHM.

Malekloo et al. (2020) conducted a comparative analysis of standalone GIS, standalone SHM, and integrated SHM-GIS approaches across multiple criteria, including structural damage assessment, spatial analysis, database management, decision-making, risk assessment, scalability, and cost. They demonstrated that while GIS alone lacks structural damage assessment capabilities and SHM alone lacks spatial analysis, their integration unifies all evaluated functionalities into a single, comprehensive framework for bridge management.

Methodology

The proposed methodology integrates IFC model, GIS, and sensors into a unified web-based platform for bridge SHM. While lightweight formats such as JSON offer advantages for web-based data integration over IFC files and have been used successfully used for BIM-IoT integration(Chatsuwan et al., 2025), few research work rely on IFC.JSON. Moreover, converting the original IFC-STP to IFC.JSON adds an additional step to the workflow and potential semantic loss. Therefore, IFC is adopted as the data backbone throughout the entire workflow.

The workflow comprises four sequential main stages as illustrated through the IDEF0 process map in Figure 1. The First stage consists of the definition of information requirements and model production, where both human-readable and machine-readable information requirements are defined using the Information Delivery Specification (IDS) standard. These specifications formally declare which properties and attributes must be present in the

BIM model to support SHM integration. With the human readable specification, a digital representation of the bridge, including its sensor instances, is developed in Revit and exported to IFC4.3 the most recent open standard for infrastructure data exchange.

Following the model production, the second stage involves the model validation and semantic enrichment. The exported IFC model is validated against the machine-readable specification using the IFCTester tool within BonsaiBIM. If validation fails, the tool displays the non-compliant elements and properties. Corrections are made either by directly editing the properties in BonsaiBIM and saving the IFC model, or by returning to Revit to adjust the model, re-exporting it to IFC, and re-running the validation. This iterative process continues until the model successfully passes the compliance check. After model validation, semantic enrichment is performed to establish explicit, relationships between sensors and their host structural elements. Addressing the lack of explicit sensor-host relationships in the current IFC 4.3 schema, this methodology adopts IfcElementAssembly as a pragmatic modeling pattern to nest sensors within their structural hosts, enabling meaningful machine-readable relationships and automated queries. To achieve the semantic nesting, a custom metadata property is attached to each host element, declaring which sensor instances belong to it. A script then parses this metadata, matches sensors to their respective hosts. The resulting assembly preserves the spatial context of the host element while linking sensors as integral components of the monitored element.

Moving to third stage of the methodology, the database system adopted to store the sensors is a relational database implemented on MySQL. Having the static IFC model and sensor values stored in the database, the model is updated with significant sensor readings using a MySQL python connector in conjunction with IfcOpenShell. Embedding important sensor readings into

the IFC model within each IfcSensor instance generates a temporally consistent view of the element status over a specific period of time. This allows users to select a sensor in the 3D model, inspect its properties, and assess localized element behavior or environmental conditions without reviewing the measurement plots.

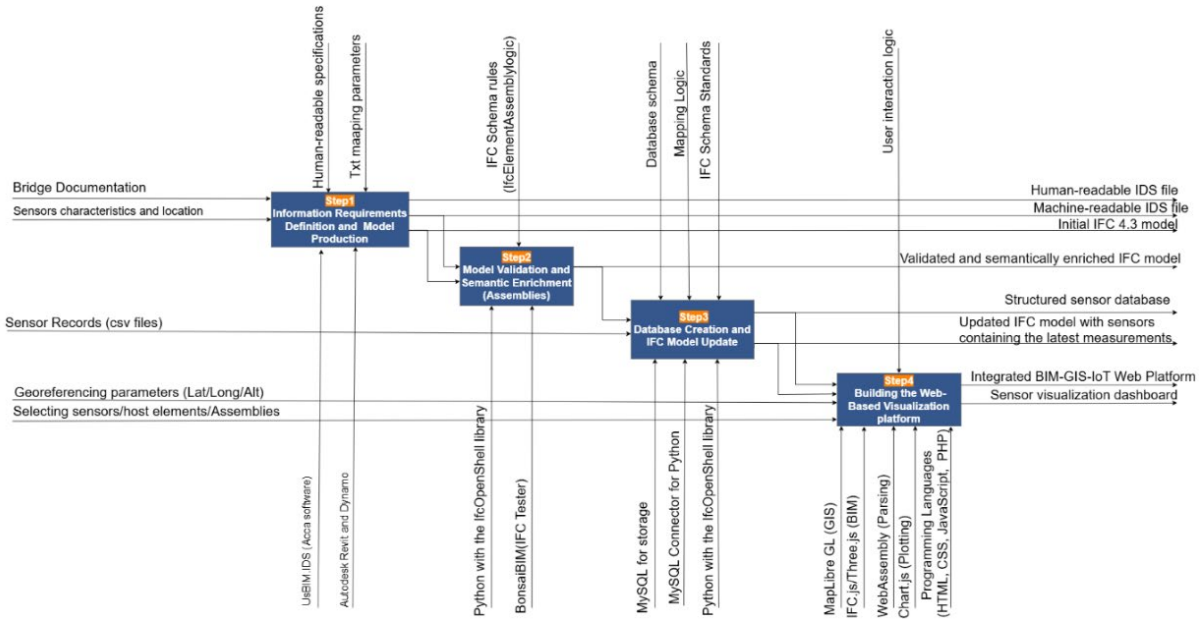


Figure 1:Process map of the methodology adopted

Finally, a web platform is developed to enable the integration of GIS, IFC model, and sensors following a three-tier client-server architecture. The architecture is structured into three distinct layers, the frontend layer serves as the client-side interface for user interaction and visualization; the backend layer manages server-side processing and the logic required for data retrieval; and the database layer provides the environment for data storage and management.

Result and Discussion

Case Study Context and Model Validation

The project case study is a reinforced concrete arched bridge. Various sensor types are installed at different locations on the bridge, including accelerometers, potentiometers, inclinometers, temperature sensors, water level sensor, and thermohydrometer. Prior to bridge model production, the human and machine-readable specifications was defined using the Information Delivery Specification (IDS) standard through the usBIM.IDS tool. Figure 2 illustrates the human-readable requirements defined for the accelerometers sensor metadata.

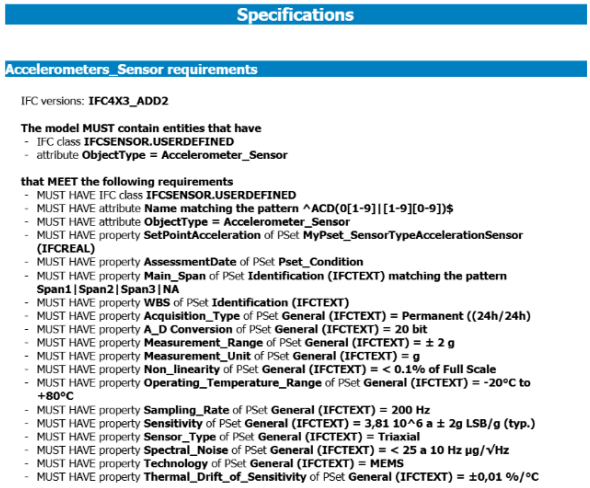


Figure 2: Human-readable requirements defined for the accelerometer sensor metadata

Following the information requirements definition, the 3D bridge model including sensors and metadata from the human-readable specification was created in Revit then exported to IFC4.3, and audited against the machine-readable requirements using IFCTester in BonsaiBIM. Figure 3 presents the validation outcome for the accelerometer sensors. The results demonstrate that the model successfully passed all compliance checks, thereby verifying that the sensors correctly adhere to the defined information requirements.

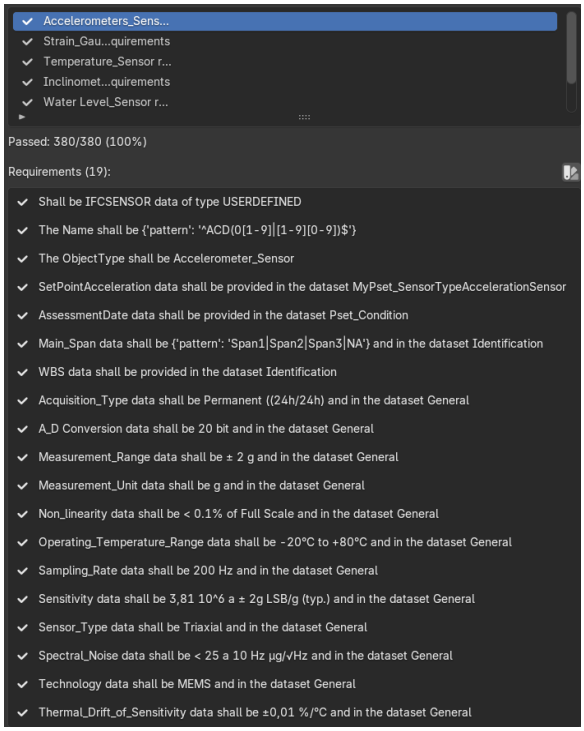


Figure 3: Human-readable requirements defined for the accelerometer sensor metadata

Semantic Linking

Following the successful validation of the model, the next step is to aggregate sensor elements with their hosts under IfcElementAssembly. An initial inspection of the IFC model within widely used BIM viewers, such as BIM Vision and BIMcollab Zoom, confirms that the physical and hierarchical relationships between the host elements and their associated sensors are not natively reflected.

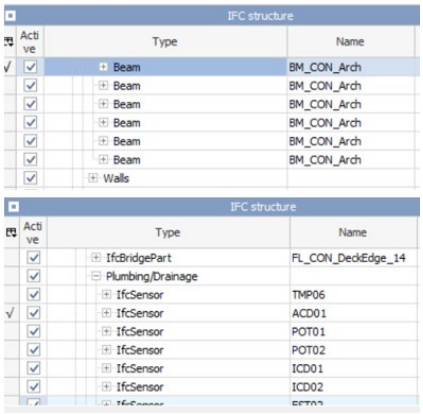


Figure 4:Beams and sensors elements hierarchical representation

Figure 4 displays the default spatial-tree representation, where beams and sensors appear as separate groups without any explicit semantic linkage. To address this, the adopted approach relies on metadata added to each host element to identify which sensors must be grouped with that specific element. This information is provided through the property “Sensors_Instances_Hosted” within the “Data” propertyset of the host element, as illustrated in Figure 5.

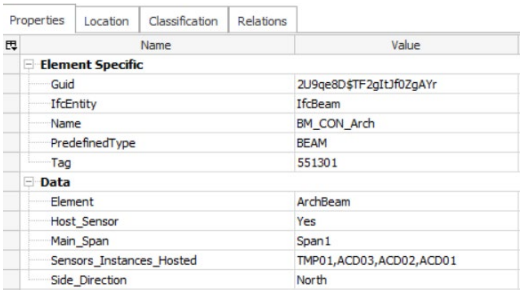


Figure 5:Identification of "Sensors_Instance_Hosted" property to establish linkage between sensors and their host

The overall workflow adopted to aggregate sensors to their host using IfcOpenShell library is illustrated in Figure 6.



Figure 6:Workflow to semantically link sensors and host element

The entities of the IFC model are iterated, and sensors and potential hosts are categorized into two groups. All IfcSensor entities are indexed by their Name attribute and Globally Unique Identifier (GUID) and stored in a dictionary. Then, during the matching phase, a list of candidate host entities is iterated, and retrieve the value of the custom property "Sensors_Instances_Hosted". This value is parsed into a list of unique identifiers, each cross-referenced against the sensor index to map each sensor directly to its specific host element. Finally, the matched elements (host element and its corresponding sensors) are aggregated into an IfcElementAssembly.The developed script successfully establishes a semantic link between sensors and their host. The result demonstrates interoperability compatibility across different BIM viewers, as shown in Figure 7 (BIM Vision) and Figure 8 (BIMcollab Zoom), where sensors (ACD01, ACD02, ACD03, TMP01) and their host element (BM_CON_Arch) are now grouped together under the assembly (BM_CON_Arch_ASB_12).

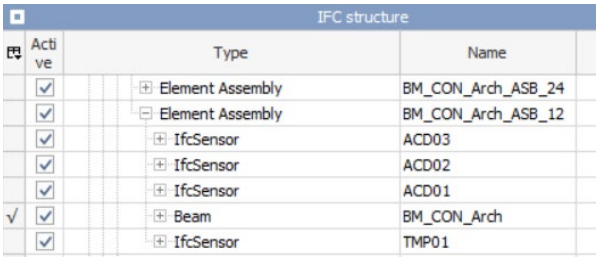


Figure 7:Resulting IfcElementAssembly of a beam element with sensors within BIM vision

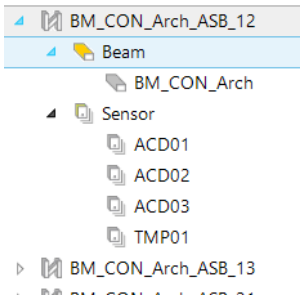


Figure 8:Resulting IfcElementAssembly of a beam element with sensors within BIMcollab Zoom

Database and Model Updating

The relational database designed on MySQL for storing sensor data is illustrated by the entity-relationship model represented in Figure 9.

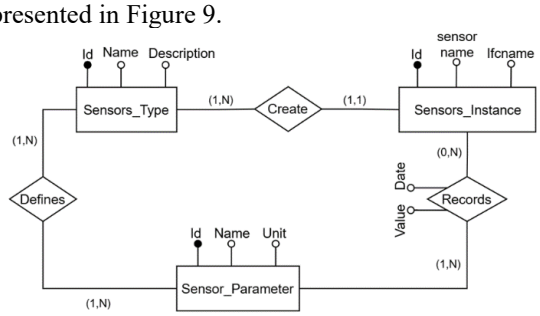


Figure 9:Entity relation model for the sensors database

The data model is organized around three core entities: Sensor_Type, representing the device category; Sensor_Instance, representing the specific physical devices installed on the structure; and Sensor_Parameter, which defines the measured physical quantities. Additionally, the "Records" relationship functions as a transactional entity, storing time-series measurements for each sensor instance

Regarding the IFC model update of IfcSensor instances with significant readings, they are updated with the latest, minimum, and maximum values over the synchronization interval, along with their respective timestamps. The workflow adopted to perform the operation is illustrated in Figure 10.

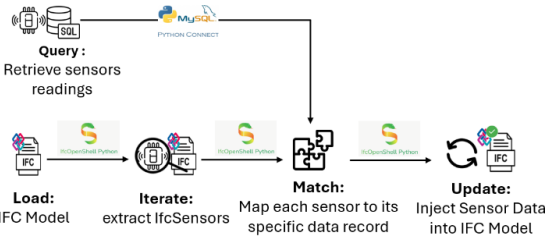


Figure 10:Workflow to update the IFC model with sensor reading

The script developed using IfcOpenShell performs the update as follows. First, the script queries the database to retrieve the values from the appropriate tables. Next, loads the IFC model and iterates through its entities to extract all IfcSensor elements.

During the matching and update phase, the script maps each sensor name from the database to the corresponding IfcName of the sensor within the model. Finally, the model update is performed by injecting the new measurement values and timestamps into the specific IfcProperty fields. The synchronization interval is one hour. Values are overwritten with each update and historical data is not retained within the IFC file. Instead, the complete time-series history is stored in the relational database. Identical IFC sensors that do not receive new sensor data are bypassed and remain untouched. The implementation of this approach is highlighted at Figure 11 and Figure 12, where Figure 11 presents the initial state of a potentiometer sensor (POT01) with default displacement value.

MyPset_Date	
AssessmentDate_Latest	21/05/2022
AssessmentDate_Max	21/05/2022
AssessmentDate_Min	21/05/2022
MyPset_SensorTypeDisplacementSensor	
SetPointDisplacement_Latest	0.01
SetPointDisplacement_Max	0.01
SetPointDisplacement_Min	0.01
Pset_ConstructionOccurence	
AssetIdentifier	POT01

Figure 11:Initial state of the POT01 sensor property before update

Figure 12 presents the successfully updated properties, with the new value and timestamp queried from the database.

MyPset_Date	
AssessmentDate_Latest	2024-12-09 01:59:59
AssessmentDate_Max	2024-12-09 01:59:58
AssessmentDate_Min	2024-12-09 01:12:29
MyPset_SensorTypeDisplacementSensor	
SetPointDisplacement_Latest	0.385801
SetPointDisplacement_Max	0.385806
SetPointDisplacement_Min	0.385636
Pset_ConstructionOccurence	
AssetIdentifier	POT01

Figure 12:Updated property of the POT01 sensor

Web Platform

To make the enriched model accessible, a three-tier web platform was developed as illustrated in Figure 13.

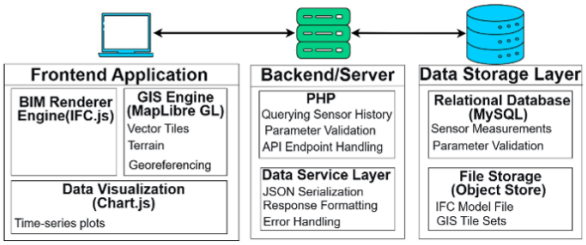


Figure 13:Updated property of the POT01 sensor

The architecture utilizes a MySQL database for data storage, a PHP backend for data processing, and a JavaScript-based frontend. To achieve BIM-GIS-IoT integration, the user interface leverages MapLibre for geospatial rendering, IFC.js for native 3D BIM interaction, and Chart.js for plotting sensor measurements Figure 14 illustrate the bridge model within the GIS context.

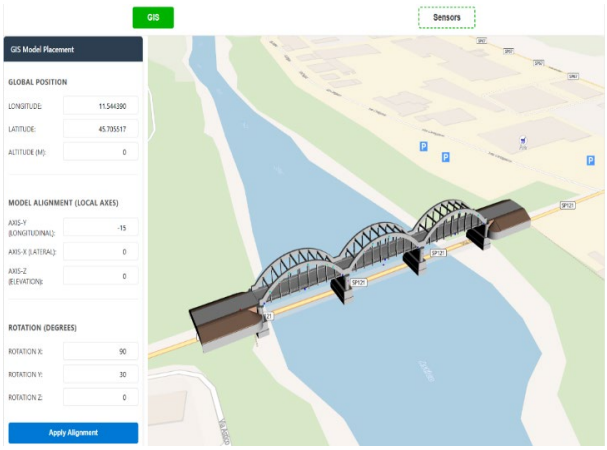


Figure 14:Bridge model into the GIS context

The sensor management interface provides 3D visualization, visibility filtering and element property inspection. Additionally, by interacting with the bridge geometry, users can retrieve measurement data; double-clicking a specific sensor or host element automatically generates the corresponding measurement plots for detailed analysis, as illustrated in Figure 15 for the case of an inclinometer being double-clicked.

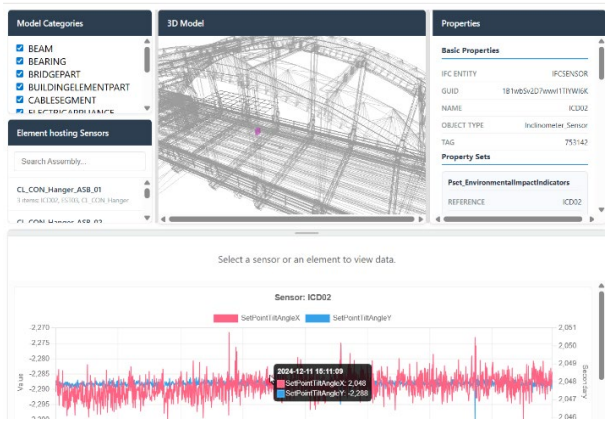


Figure 15:Data plot of an inclinometer measurements by double clicking on the sensor in the 3D model

Similarly, the list of host elements and their semantically nested sensors under IfcElementAssembly is provided in a dedicated window. Selecting an assembly automatically highlights the host element and displays plots for all associated sensors.

Furthermore, a threshold evaluation system was developed for static measurements to visually identify structural host elements exceeding predefined limits within a specified timeframe. Figure 16 shows the 3D model where host elements below the limit appear in green, while those above the threshold are highlighted in red



Figure 16:Threshold-based highlighting of bridge elements in the 3D model

The threshold manager dashboard illustrated in Figure 17 allows users to input threshold values and timeframes, displaying which inclinometer sensors that exceed or remain below the five-degrees limit.

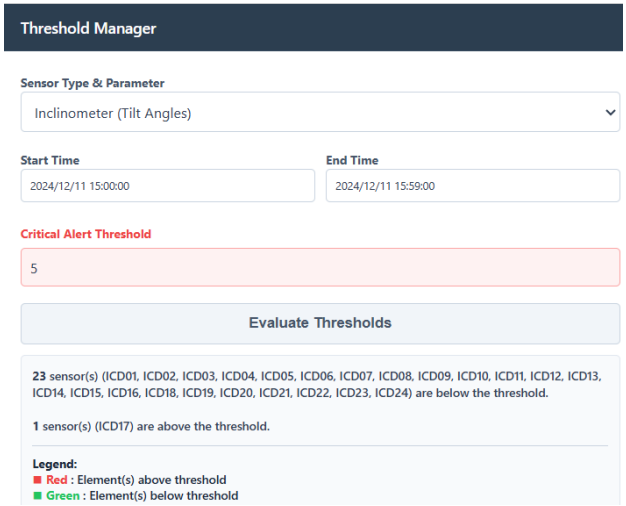


Figure 17:Threshold Manager showing inclinometer sensors above and below the critical limit

Conclusion

This study addressed critical needs identified in both literature and practice by implementing a web-based platform that integrates BIM, GIS, and IoT sensors for bridge monitoring. The platform integrates model checking using Information Delivery Specifications (IDS) to ensure data quality prior to integration. Furthermore, the use of the existing IfcElementAssembly entity links sensors with their host elements, enabling meaningful, machine-readable relationships that support automated querying. The proposed workflow also eliminates the reliance on proprietary data formats and dual-database synchronization for the model updating of IfcSensor instances. By injecting significant measurement values directly into the sensors, generates a temporally consistent view of the host element or environment status over a specific period. This allows users to select a sensor in the 3D model, inspect its properties directly without needing to review a separate measurement plots window. In addition, visualizing the bridge model within the GIS environment provides essential spatial context, while the sensor management interface allows users to interact with the 3D model and plot sensor measurements for detailed analysis. The platform enables the evaluation of both localized element behavior such as displacement tracked via potentiometers and tilt angle measured by inclinometers on specific structural elements and contextual environmental conditions, including relative humidity, temperature, and water levels. Additionally, a threshold-based alert system automatically highlights host elements in the 3D model when sensor readings exceed predefined limits, providing intuitive visual warnings. Nevertheless, some limitations must be noted. Because the semantic linking relies on exact metadata matching, the process can fail if sensor names are missing or incorrect. Future studies could therefore incorporate spatial collision detection algorithms to link sensors to elements based purely on their physical 3D location. Furthermore, while direct static sensor measurements provide a good understanding of localized element behavior, the global structural assessment remains incomplete, future research could integrate experimental data from Operational Modal Analysis (e.g., natural frequencies and mode shapes) with numerical FEM data to evaluate global structural behavior. Despite these current limitations, the proposed workflow provides a highly reliable, open-standard foundation for bridge SHM leveraging openBIM.

Acknowledgments

This work was conducted as part of the Indice Project, Task 2.3, Deliverable 2.3.1, focused on defining Digital Twin solution based on openBIM and IoT protocols. Gratitude is expressed to the BIM team of ITS Engineering Company for their support.

AI Declaration

The authors used generative AI tools to improve the readability and linguistic quality of the manuscript, as well as to assist with code development.

References

Angelosanti, M., Currà, E., & Sabato, A. (2023). BIM oriented applications of structural health monitoring based on magnified digital image correlation point-clouds. Automation in Construction, 148, 104754. https://doi.org/10.1016/j.autcon.2023.104754

Biagini, C., Aglietti, A., & Bongini, A. (2023). A BIM-BASED APPROACH TO THE MANAGEMENT OF HISTORIC BRIDGES. https://doi.org/10.36253/979-12-215-0289-3.110

Buuveibaatar, M., Shin, S., & Lee, W. (2025). Digital Twin Framework for Road Infrastructure Management.

Applied Sciences, 15(10), 5765. https://doi.org/10.3390/app15105765

Chatsuwan, M., Ichinose, M., & Alkhalaf, H. (2025). Enhancing Facility Management with a BIM and IoT Integration Tool and Framework in an Open Standard Environment. Buildings, 15(11), 1928. https://doi.org/10.3390/buildings15111928

Fawad, M., Salamak, M., Chen, Q., Uscilowski, M., Koris, K., Jasinski, M., Lazinski, P., & Piotrowski, D. (2025). Development of immersive bridge digital twin platform to facilitate bridge damage assessment and asset model updates. Computers in Industry, 164, 104189. https://doi.org/10.1016/j.compind.2024.104189

Fawad, M., Salamak, M., Poprawa, G., Koris, K., Jasinski, M., Lazinski, P., Piotrowski, D., Hasnain, M., & Gerges, M. (2023). Automation of structural health monitoring (SHM) system of a bridge using BIMification approach and BIM-based finite element model development. Scientific Reports, 13(1), 13215. https://doi.org/10.1038/s41598-023-40355-7

Figueiredo, E., Moldovan, I., & Marques, M. B. (2013). Condition Assessment of Bridges: Past, Present and Future. Universidade Católica Editora, Unipessoal, Lda.

Gharehbaghi, V. R., Noroozinejad Farsangi, E., Noori, M., Yang, T. Y., Li, S., Nguyen, A., Málaga-Chuquitaype, C., Gardoni, P., & Mirjalili, S. (2022). A Critical Review on Structural Health Monitoring: Definitions, Methods, and Perspectives. 29(4), 2209–2235. https://doi.org/10.1007/s11831-021-09665-9

Huang, D., Lu, Y., Hua, L., Lin, S. T., Deng, J., & Chen, S. (2024). Building information modeling supported bridge structural health diagnosis and prognosis. Structural Health Monitoring, 25(2), 684–701. https://doi.org/10.1177/14759217241293460

Lee, W., Yu, Y., Koo, B., Choi, B., & Lee, S. (2025). IFC Schema Extension for BIM-based Automated Code Checking of Korean Design Standards for Bridges. https://doi.org/10.22260/ISARC2025/0121

Li, X., Xiao, Y., Guo, H., & Zhang, J. (2022). A BIM Based Approach for Structural Health Monitoring of Bridges. KSCE Journal of Civil Engineering, 26(1), 155–165. https://doi.org/10.1007/s12205-021-2040-3

Malekloo, A., Ozer, E., & Al-Turjman, F. (2020). Combination of GIS and SHM in Prognosis and Diagnosis of Bridges in Earthquake Prone Locations, A Short Survey. In Smart Grid in IoT-Enabled Spaces: The Road to Intelligence in Power (Malekloo, A., Ozer, E., Al-Turjman, F, pp. 139–154). CRC Press. https://doi.org/10.1201/9781003055235-7

Marcellino, G., Toldo, B. M., & Zanchetta, C. (2025). Relational Database in Digital Platforms for Monitoring: Verona Arena Case-study. Proceedings of the 2025 European Conference on Computing in Construction, Computing in Construction, 6. https://doi.org/10.35490/EC3.2025.198

Nakamura, K., Tsukada, Y., Teraguchi, T., Kurokawa, C., & Imai, R. (2025). Integrated data model for bridges with 3D geometry and maintenance information. Computer-Aided Civil and Infrastructure Engineering, 40(27), 4994–5009. https://doi.org/10.1111/mice.70084

Panah, R. S., & Kioumarsi, M. (2021). Application of Building Information Modelling (BIM) in the Health Monitoring and Maintenance Process: A Systematic Review. Sensors, 21(3), 837. https://doi.org/10.3390/s21030837

Smarsly, K., & Tauscher, E. (2015). IFC-BASED MONITORING INFORMATION MODELING FOR DATA MANAGEMENT IN STRUCTURAL HEALTH MONITORING. https://doi.org/10.25643/bauhaus-universitaet.2823

Sun, Z., Mahmoodian, M., Sidiq, A., Jayasinghe, S., Shahrivar, F., & Setunge, S. (2025). Optimal Sensor Placement for Structural Health Monitoring: A Comprehensive Review. Journal of Sensor and Actuator Networks, 14(2), 22. https://doi.org/10.3390/jsan14020022

Teixeira, G. P., Ribeiro, J. C. L., César Júnior, K. M. L., Nunes, L. A., Carvalho, J. M. F., Oliveira, D. S., & Nalon, G. H. (2024). Interoperability level for bridge structural analysis from the IFC data interpretation. Gestão & Tecnologia de Projetos, 19(3), 49–67. https://doi.org/10.11606/gtp.v19i3.227401

Theiler, M., & Smarsly, K. (2018). IFC Monitor – An IFC schema extension for modeling structural health monitoring systems. https://doi.org/10.1016/j.aei.2018.04.011

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

A UTILITY-AWARE KNOWLEDGE GRAPH TO SUPPORT MULTI-SCALE RESIDENTIAL ENERGY MODELLING

Cathal Hoare[1], Sharon Coffee[2], Divyanshu Sood[2], Tiziana Margaria[1] and James O'Donnell[3]
[1] Department of Computer Science and Information Systems, University of Limerick, Ireland
[2] School of Material and Mechanical Engineering, UCD Energy Institute, University College Dublin, Ireland
[3] School of Engineering, University of Limerick, Ireland

Abstract

Building operations account for major global energy use, yet decarbonisation is hindered by fragmented data silos. This paper presents a utility-aware knowledge graph developed for multi-scale residential energy modelling. By leveraging Linked Data principles and the DDIM server, the framework integrates building-level performance data with national electricity distribution networks. Using Ireland as a case study, the methodology semantically associates 2.3 million residences with over 46,000 low-voltage transformers through a novel enrichment process. This scalable approach enables planners to perform complex queries for grid capacity analysis, bridging the gap between individual building physics and urban-scale energy simulation.

Introduction

Building operations accounted for 30% of global energy consumption in 2021 (Skillington et al., 2022; Bacheva and Raposo Grau, 2025). Ireland, like many countries, is on-track to miss decarbonisation goals, a potential significant cost to the exchequer (Environmental Protection Agency, 2025). Ireland is also suffering a housing crisis, with development of residential stock failing to meet demand (Social Justice Ireland, 2025). Infrastructure problems, such as insufficient electricity supply capacity, are a contributing factor to both of these issues (Irish Academy of Engineering, 2025). Data-driven decisions are now vital to address the issues of global warming, building comfort and stock availability (Skillington et al., 2022). By fusing various information sources, such as building data, energy infrastructure and population data, and applying Machine Learning (ML), planners can perform sophisticated analyses, to identify energy performance trends and optimize large-scale carbon reduction strategies (Ali et al., 2021a).

Access to accurate information is of fundamental importance to all stakeholders in this process, including planners and policy makers (Madnick et al., 2009). The information that must be managed is, however, complex and heterogeneous, both in content and organization (Bonilla et al., 2019). Information sources are dynamic, evolving, and multifaceted; information can be distributed across multiple sources providing related, yet heterogeneous collections of data that inform utility companies, planners and other stakeholders (Boddy et al., 2007). National and district level views of the same entities are managed by separate formats and systems, requiring expensive data reorganization in order to extract relevant information for planners and policy makers (Swan and Ugursal, 2009; Nacht et al., 2023). Together these issues describe rich, timely data sources whose utility is diminished by their disjointed and heterogeneous nature. This work presents an approach that seeks to overcome these difficulties and increases the value of data by integrating diverse sources into a homogeneous data source that can be extended to include new sources as they become available (Li et al., 2019).

Several advances have enabled solutions to these problems. Linked data provides a mechanism to interlink disparate datasources and interrogate them using semantic queries - that is queries that are of associative and contextual nature (Li et al., 2019). Linked data allows all relationships in a datastore to be explicitly defined. This facilitates the development of flexible queries, and admits reasoning about both data and their relationships (Ngomo et al., 2014). Linked data also permits the integration and exchange of data from a variety of sources and domains (Consoli et al., 2017). Digital twins for multi-scale energy modelling integrate Linked Data and ML to bridge the gap between building-level physics and urban-scale dynamics. Linked Data, utilizing RDF and ontologies, provides a semantic framework to unify heterogeneous data sources, such as IoT sensors, GIS, and weather APIs, ensuring interoperability across scales. Meanwhile, ML models accelerate simulations by acting as surrogate models for complex thermodynamic equations. By training on historical and real-time semantic data, these twins enable predictive analytics for grid stability and demand response. This synergy allows for a "glass-box" approach where data remains structured and searchable, while ML provides the computational power to optimize energy distribution in complex, multi-layered environments (Santos et al., 2025).

This research paper details the creation of a utility-aware knowledge graph designed to improve multi-scale energy modelling within the built environment. By utilizing Linked Data principles, the authors successfully integrate fragmented datasets involving building characteristics, population statistics, and electrical grid infrastructure. The study specifically demonstrates this approach

in Ireland, connecting millions of residential properties to their nearest low-voltage transformers to create a unified view of energy demand. This framework addresses the problem of functional silos by allowing planners to perform sophisticated, data-driven analyses of decarbonisation strategies and grid capacity. Ultimately, the methodology provides a scalable solution for integrating heterogeneous information sources to support sustainable urban planning across different jurisdictions.

The paper is organized as follows. First, the background explores functional silos in the energy modelling domain and the application of Linked Data and standardized ontologies. The methodology then details the development of the utility-aware knowledge graph, focusing on the core context and the DDIM server used for managing data fusion. A central component describes the enrichment process that semantically links building assets to low-voltage electricity distribution networks. The results section examines the national-scale graph generated using Irish datasets, including GeoDirectory and ESB Networks. The study's capabilities are then validated through competency questions that demonstrate sophisticated reasoning. Finally, the paper discusses the methodology's global scalability before concluding with future research directions.

Background

Multi-scale residential energy-modelling is significantly hindered by functional silos, which are compartmentalised operating units isolated from their environment (Ahmad et al., 2018; Ali et al., 2021b). These silos result in horizontal fragmentation between project disciplines and vertical fragmentation between individual building and district-scale representations (Ali et al., 2024; Balaji et al., 2016). Data integration remains a vexing problem because building information is complex, dynamic, and distributed across multifaceted sources in incompatible formats (He et al., 2023). Planners frequently encounter data inconsistencies and a lack of standard building identifiers, leading to property discrepancies across datasets (Hoare et al., 2022; Pauwels et al., 2017). Access to high-quality energy data is further restricted by privacy and protection policies, which may exclude critical information or compromise its quality through anonymisation (Pritoni et al., 2021; Rasmussen et al., 2021). These hurdles necessitate tedious manual data reorganisation and collection efforts that are both time-consuming and costly (Sood et al., 2023; Tardioli et al., 2017). Consequently, these integration issues create a persistent gap between detailed building energy modelling and traditional urban planning practices (Wang et al., 2024; Wu et al., 2024).

Linked Data principles address the persistent challenge in multi-scale residential energy-modelling of functional silos by providing a mechanism to interlink disparate, heterogeneous data sources into a unified "Web of Data" (Pauwels et al., 2017; Hoare et al., 2022). By utilizing the Resource Description Framework (RDF) and unique Uniform Identifiers (URIs), this approach admits semantic queries of an associative and contextual nature, allowing all relationships in a datastore to be explicitly defined (Hoare et al., 2022; Bizer et al., 2009).

The sector has operationalized this through standardized ontologies:

- BOT provides a lightweight core for building topology (Rasmussen et al., 2021).
- Brick standardizes building automation metadata (Balaji et al., 2016).
- SAREF enables interoperability among smart appliances (Daniele et al., 2015).

A significant practical application is the use of federated SPARQL queries to integrate Ireland's national transmission grid data with residential housing stock archetypes (Hoare et al., 2022). This allows for utility-aware, multi-scale reasoning where data owners maintain local control while planners perform national-level scenario exploration for decarbonisation (Hoare et al., 2022).

Utility information in Linked Data is represented through specialized ontologies that bridge the gap between building-level assets and national infrastructure (Hoare et al., 2022). While standard building schemas like Brick often lack grid-specific concepts (He et al., 2023), extensions such as `SARGON` incorporate electrical grid automation by mapping CIM and IEC 61850 data models into RDF formats (Haghgoo et al., 2020).

Crucially, this information is structured using a hierarchical *common context* spine; individual buildings are semantically linked to geographical small areas and then to transmission nodes (Hoare et al., 2022; He et al., 2023).

Methodology

We now turn to describing the development of a utility aware knowledge graph (KG) to support multi-scale energy modelling. The KG is structured by defining a number of common contexts to create vertical hierarchies of relationships between represented entities and horizontal relationships to define equivalence between disparate representation of the same physical entities. While exemplified using the Irish context (and relevant datasets), the approach can be applied across jurisdictions where suitable datasets are available. Having described the common context, the process used to integrate building and electricity distribution network information is described. The section will conclude with a brief description of DDIM server, used to manage the data fusion and access process.

Core context

A graphical representation of the context model along with the organization of this project's data through the use of the ontology are shown in Figure 1 (A). The ontology is deliberately sparse to maintain clarity. An entity represents some physical entity or space in time. Each instance

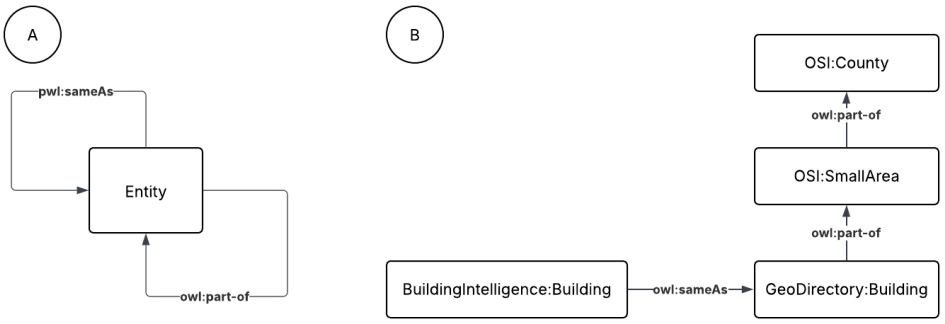


Figure 1: A. Context model and B. Relationships to data sets used in this project

of an entity is described by a class, for example, a building, region or nation. Together, these classes describe a set of layered contexts, each representing a different scale. Information sources are associated with some entity in each layer. This association with context allows sources to be intra-connected at a particular scale, and inter-connected between scales. Figure 1 (B) illustrates the use of the context model to place buildings in geographical hierarchies and relate building representations to each other across datasets. Equivalence is expressed using the *owl:sameAs* relationship; we use the relationship to denote that buildings in different datasets represent the same physical entity (Halpin et al., 2010). The use of *owl:sameAs* can give rise to semantic ambiguity where datasets' properties lists overlap. This issue is overcome by assigning a source authority ranking for datasets as they are added to the graph. This feature is provided through the DDIM's data upload forms. Hierarchies are created using the *owl:part-of* relationship. In Figure 1 building representations are in the geographical context of a small area which in turn is part of a county. In this way, various transitive relationships can be created to support aggregation and other analysis operations over geographical contexts.

Managing data fusion and access

The Dynamic District Information Model (DDIM) (Project, 2023a) is used to mine and manage relationships between the siloed data sources. For this project, the DDIM uses the core context to defined relationships between geographical areas and locations to provide a common context between data sources. This context is captured through a knowledge graph defiend in Neo4J (with NeoSemantic plugin). The server supports forms based no-code implementation of the graph, where forms are used to upload datasets and create owl:sameAs and owl:part-of relationships between columns in the uploaded data. the The server implementation consists of three key functional areas:

- Creation of 'Core Spines' (describe by the context model) using a form based process integrates uploaded information. These spines allows information to be associated with entities in various contexts, facilitating querying across these spaces;
- A series of interfaces to allow client software to query (through RESTful or SPARQL based interfaces);
- A DCAT compliant data catalogue;

The server is implemented using the Django Framework (Project, 2023b) with RESTful (DjangoRESTful, 2023) extensions. This framework also provides user management and security. Neo4J and NeoSemantic (Project, 2023a) is installed within the same domain. Queries are submitted through the Django server to this database.

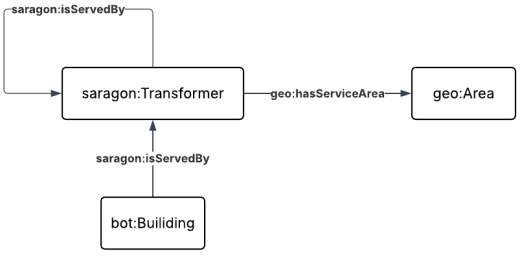


Figure 2: Additional graph structure added post enrichment.

Enriching KG with building/LV relationships

We now examine the steps taken to integrate building and electricity distribution network information. In this instance, the information used was provided by ESB Networks as part of their *Capacity Heatmap for Network Planning* tool (ESB Networks, 2026). This provides a wealth of information about the location and capacity of transformers on the Irish electricity distribution network. One purpose of the Capacity Heatmap is to guide customers who wish to make a new connection to the network. In this use, customers are advised to locate the transformer closest to the required connection; we employ the same heuristic in this work. Low Voltage (LV) transformers were associated with building data (from Geodirectory (GeoDirectory, 2022a)) and areas representing each transformer's serviced area were created using the following steps:

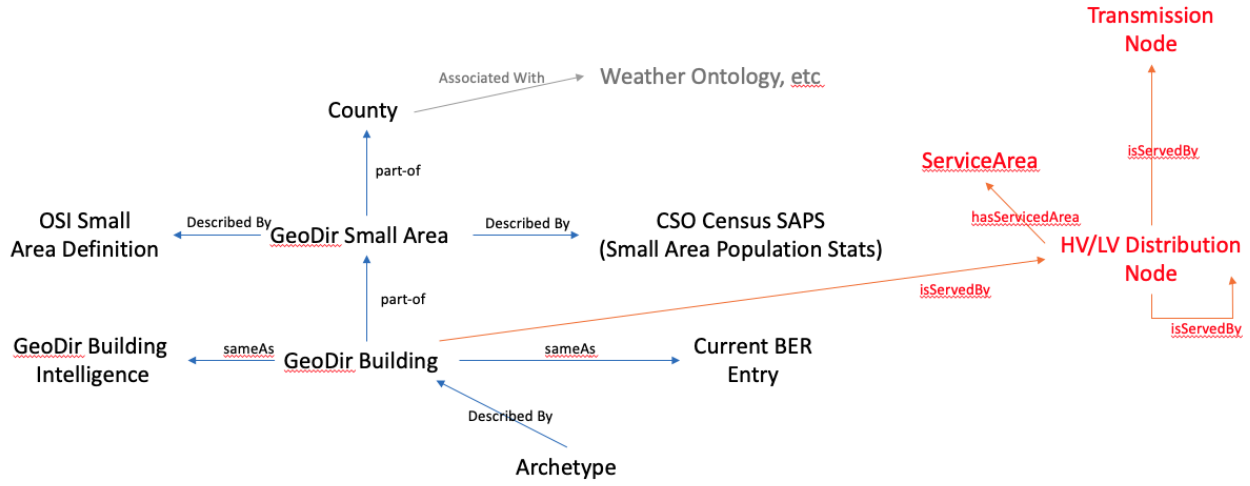


Figure 3: Representation of relationships between datasets in National Scale Residential Energy Graph. Elements added as part of the enrichment process are shown in red.

1. Entity definitions

Let B be the set of buildings:

$$B = \{b_i \mid i = 1, \ldots, n\}$$

where $b_i = (\mathbf{p}_i, \mathscr{C}_i)$, with $\mathbf{p}_i \in \mathbb{R}^2$ as coordinates and $\mathscr{C}_i$ as characteristics.
Let T be the set of transformers:

$$T = \{t_j \mid j = 1, \ldots, m\}$$

where $t_j = (n_j, \kappa_j, \mathbf{l}_j)$, with n_j as name, κ_j as capacity, and $\mathbf{l}_j \in \mathbb{R}^2$ as location.

2. Relationship assignment

For each building b_i, the closest transformer t^* is found via:

$$f(b_i) = \arg\min_{t_j \in T} d(\mathbf{p}_i, \mathbf{l}_j)$$

This establishes the relation $(b_i, \text{isServedBy}, f(b_i))$.

3. Geometric representation

The service area for transformer t_j is defined by the convex hull of the points S_j it supplies:

$$S_j = \{\mathbf{p}_i \mid f(b_i) = t_j\}$$

$$\text{CH}(S_j) = \left\{ \sum_{k=1}^{|S_j|} \alpha_k \mathbf{p}_k : \sum \alpha_k = 1, \alpha_k \geq 0 \right\}$$

4. Knowledge graph insertion

The graph $\mathscr{G}$ is updated such that:

$$\mathscr{G} \in \{(b_i, \text{isServedBy}, t_j), (t_j, \text{hasServiceArea}, \text{CH}(S_j))\}$$

These steps were implemented as a Python script which output two files, a text file containing Cypher queries to create a relationship between each building and the transformer that it is served by. The second output file is a GeoJSON (Internet Engineering Task Force (IETF), 2016) formatted definition of the convex hull defined areas that describes the area containing buildings serviced by a transformer. In the Irish context, 46,527 transformers were associated with over 2.3 million residences.

Figure 2 shows the structure of the developed graph extension when the LV/Building relationship process is processed. Transformers and their associated connections to buildings are described using the SmArt eneRGy dOmain oNtology (SARAGON) (Haghgoo et al., 2020) ontology. Transformers can be served by higher voltage transformers and in turn serve buildings. Each transfomer is also associated with a geographical area representing the convex hull described by the buildings it serves. This area can be used for visualisation and other purposes.

Results

This paper continues by examining the generated utility aware graph. The data sources and resulting graph structure will be described in the Irish context before examining the outputs Building/LV enrichment process. The convex hull outputs will be described before a number of competency questions will demonstrate the complexity of queries that the knowledge graph can inform. Finally, the potential to use this methodology in other jurisdictions will be discussed.

Structure of knowledge graph

The DDIM and core context were used to create a national scale residential energy graph for Ireland. The graph structure is used to inform energy modelling. Subsequently, the graph was enriched to include information about the Irish electricity distribution network using the steps described a-priori. Development of the graph followed the following steps:

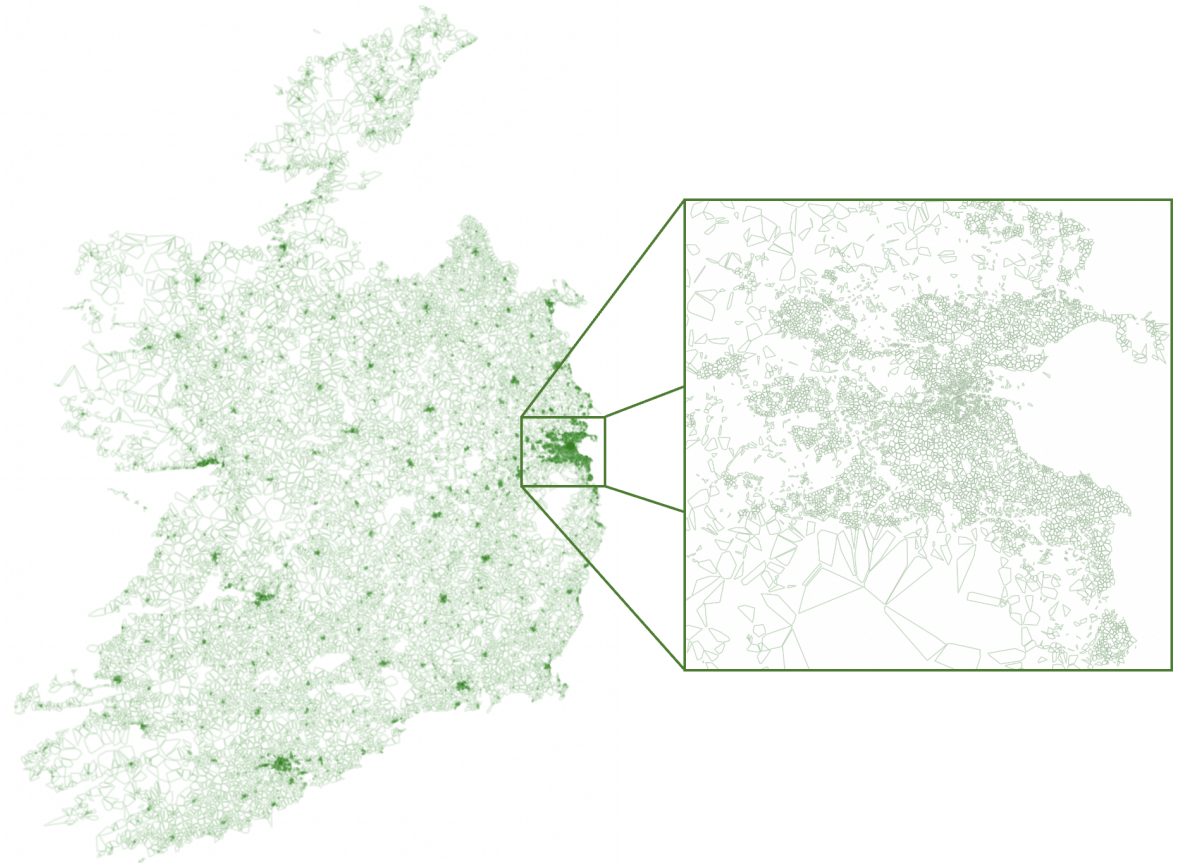


Figure 4: Map generated from presentation of derived convex hulls representing areas served by LV transformers. Clusters of smaller areas correspond to urban areas with higher densities of residential properties.

- Collection of relevant data sources.
- Identification of vertical and hierarchical relationships that could be exploited to support dataset fusion using the core context.
- Creation of the graph using the DDIM server.

The following datasets were identified:

- Geodirectory's *GeoAddress (GeoDirectory, 2022a)* and associated *BuildingIntelligence (GeoDirectory, 2022b)* – comprehensive list of approximately 2.3 million buildings, including location information and residential building properties. BuildingIntelligence includes energy modelling relevant information, for example, heating fuel.
- SEAI's *National BER Research Tool (SEAI, 2023)* - contains anonymised data used to calculate buildings' EPC certificate. The database contains data collected during building surveys. There are entries for approximately 60% of the Irish residential building stock.
- Irish Central Statistics Office (CSO) *Small Area Population Statistics (SAPS) (Office, 2022)* – 15 themes describe various statistics garnered from the CSO's census. These include socio-economic information about housing, commuting, motor cars and other energy modelling resources.

The Knowledge Graph contains other relevant information, including geographic area definitions from Ordinance Survey Ireland and developed building archtypes.

The DDIM was used to create the graph structure shown in Figure 3. Horizontal *same-as* relationships were identified between, building representations in the BER, GeoAddress and Building Intelligence datasets. Each building (represented by the GeoAddress entry) was associated with a building archtype. A vertical *core context* was created where each building was placed in the context of a small area using part-of relationships. Small areas were similarly placed in counties. These geographic contexts were chosen as they are widely used by various data providers. However, finer grained contexts including streets and townlands could also have been created.

The graph was subsequently enriched by creating associations between buildings and transformers (shown in red in Figure 3). Having followed the process described a priori. This added nodes for each transformer and a representation of the area they serve. Relationships between building representations and transformers were also created.

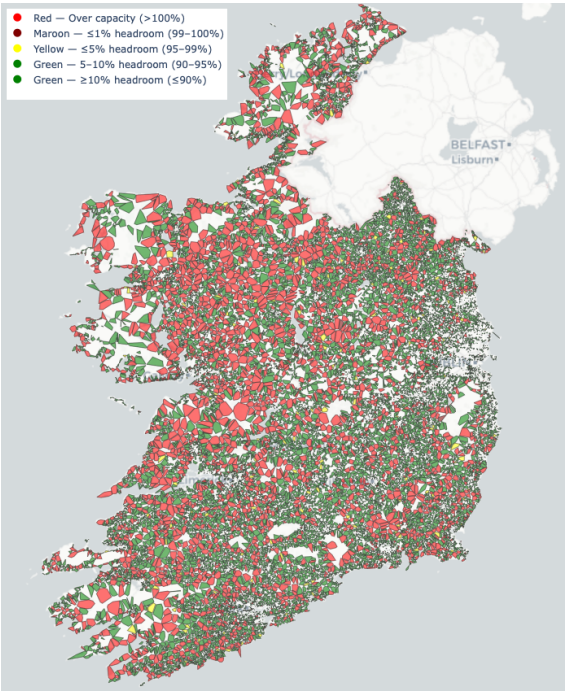


Figure 5: esults of multiscale energy modelling for scenarios including heatpump and EV adoption by residential households.

Visualising the relationships between LV transformers and buildings

Figure 4 presents a visualisation of the areas served by individual LV transformers at various scales. A distinct contast exists between urban areas characteristed by denser clusters of smaller areas and remote low-population areas (west coast and mountainous areas such as South of Dublin). This becomes more apparent when the map is zoomed to an urban areas such as Dublin (inset, Figure 4. As these areas are defined as GeoJSON polygons, colours can be assigned to each area depending on the form of analysis applied; for example, the visualisation can be used to show which LV transformers will experience capacity issues under various scenarios. The use of GeoJSON is also advantageous as web-based interfaces can be created dynamically zoom in and out across multiple-scales.

An example of this application is shown in Figure 5. In this isntance, building characteristics (fabric, occupancy and heating system) are aggregated into representative archetypes and converted to hourly electricity demand. Electrification scenarios such as heat-pump adoption and EV charging are applied as behavioural state changes within the housing stock. The resulting demand is spatially allocated to low-voltage substations, allowing feeder peak load and transformer overload to be calculated.

Competency questions for knowledge graph

Scoping and design of the KG was guided by developing numerous competency questions (Keet and Khan, 2024). These questions were also used to validate the capabilities of the graph. We now present two of these along with their results to demonstrate the ability of the KG to reason in sophisticated ways about knowledge relevant to multi-scale energy modelling in Ireland.

CQ1: Average residential buildings per LV node

The following Cypher query was used to report on statistics related to this query:

```
MATCH (b:Building)-[:NEAREST_LV_STATION]->(
lv:LVStation)

WITH lv, count(b) AS buildingCount

RETURN avg(buildingCount) AS
avgBuildingsPerLVNode,
min(buildingCount) AS minBuildings,
max(buildingCount) AS maxBuildings,
count(lv) AS totalLVNodes
```

The average number of connections was found to be 47.0 buildings per LV node with a standard deviation of 44.5. In total 42,235 LV nodes were found to supply at least one residential building. This can be explained by some LVs supplying commercial buildings and the use of the nearest LV to a building heuristic.

CQ2: Aggregate connectivity summary

Provide a summary table that describes overall connectivity between LV nodes and residential buildings.

```
MATCH (lv:LVStation)
OPTIONAL MATCH (lv)<-[:NEAREST_LV_STATION]-
                        (b:Building)
WITH lv, count(DISTINCT b) AS buildings

OPTIONAL MATCH (lv)-[:SERVES_SMALL_AREA]->
                        (sa:SmallArea)
WITH lv, buildings, count(DISTINCT sa) AS
                            smallAreas

OPTIONAL MATCH (lv)-[:HAS_ARCHETYPE_COUNTS]
->(ac:ArchetypeCounts) WITH lv, buildings,
smallAreas, count(DISTINCT ac) AS
                          archetypeNodes
WITH buildings + smallAreas +
     archetypeNodes AS totalConnections

RETURN avg(totalConnections) AS
avgConnectivity, min(totalConnections)
AS minConnectivity, max(totalConnections)
AS maxConnectivity, percentileCont(
totalConnections, 0.5) AS
medianConnectivity
```

In summary, overall connectivity is described thus: average 53.9, median 41.0, range 0–619. The sophistication of the query and its ability to bridge disparate siloed information sources is demonstrated.

Applying this methodology in other jurisdictions

The methodology is designed for global scalability. Access to appropriate datasets represents the greatest challenge. For example, in the United Kingdom, access to GeoPlace's National Address Gazetteer would yield a dataset of addresses equivalent to GeoDirectory, while EPC data can be accessed through the Department for Levelling Up, Housing & Communities' Energy Performance of Buildings Data datasets. By mapping these and other datasets to common contexts, the no-code form based approach used by the DDIM allows domain experts to create utility-aware knowledge graphs in any jurisdiction. Larger datascales at national level are supported by the ability of the DDIM to operate as a federated node in distributed deployment. Using building archetypes further enables national-scale energy simulations for diverse decarbonisation planning.

Conclusions

This paper addresses the fragmentation of AECOO data silos by developing a utility-aware knowledge graph (KG) for multi-scale residential energy modelling. By leveraging Linked Data principles and the Dynamic District Information Model (DDIM), the research successfully integrated building-level characteristics with national electricity infrastructure. The methodology was validated in the Irish context, semantically linking 2.3 million buildings to 46,527 Low Voltage (LV) transformers through a novel automated enrichment process. Results from competency questions demonstrated the KG's ability to perform sophisticated reasoning, revealing an average of 47.0 residential buildings served per LV node. This framework effectively bridges building physics with urban-scale dynamics, offering a globally scalable approach for planners to optimize grid capacity and decarbonisation strategies. Future work will focus on integrating real-time sensor data and machine learning to enhance predictive analytics for grid stability

Acknowledgments

This publication has emanated from research supported partly by a grant from the Sustainable Energy Authority of Ireland (SEAI) under the Research, Development and Demonstration Funding Programme Grant No. 22/RDD/1300. The opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the SEAI. Anthropic Claude used to improve the grammar and clarity of some text here-in; all content are the authors' responsibility.

References

Ahmad, T., Chen, H., Guo, Y., and Wang, J. (2018). A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. Energy and Buildings, 165:301–320.

Ali, U., Bano, S., Shamsi, M. H., Sood, D., Hoare, C., Zuo, W., Hewitt, N., and O'Donnell, J. (2024). Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. Energy and Buildings, 303:113768.

Ali, U., Shamsi, M. H., Bohacek, M., Purcell, K., Hoare, C., and O'Donnell, J. (2021a). Gis-based multi-scale residential building energy modeling using a data-driven approach. In Proc. IBPSA 2021, volume 17 of Building Simulation, pages 1115–1122, Bruges, Belgium. IBPSA.

Ali, U., Shamsi, M. H., Hoare, C., Mangina, E., and O'Donnell, J. (2021b). Review of urban building energy modeling (ubem) approaches, methods and tools using qualitative and quantitative analysis. Energy and Buildings, 246:111073.

Bacheva, T. S. and Raposo Grau, J. F. (2025). Embodied impacts in buildings: A systematic review of life cycle gaps and sectoral integration strategies. Buildings, 15(10).

Balaji, B., Bhattacharya, A., Fierro, G., et al. (2016). Brick: Towards a unified metadata schema for buildings. In Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys '16), pages 41–50.

Bizer, C., Heath, T., and Berners-Lee, T. (2009). Linked data - the story so far. International Journal on Semantic Web and Information Systems (IJSWIS), 5(3):1–22.

Boddy, S., Rezgui, Y., Cooper, G., and Wetherill, M. (2007). Computer integrated construction: A review and proposals for future direction. Adv. Eng. Softw., 38(10):677–687.

Bonilla, J., Perez, J., Pinto-Seppa, I., and Zarli, A. (2019). Articial intelligence - research and innovation needs of manufacturing, energy intensive industries, bio-based industries and construction. Technical report, The European Construction Technology Platform - and the Energy Efficient Building.

Consoli, S., Presutti, V., Recupero], D. R., Nuzzolese, A. G., Peroni, S., Mongiovi', M., and Gangemi, A. (2017). Producing linked data for smart cities: The case of catania. Big Data Research, 7:1 – 15.

Daniele, L., den Hartog, F., and Roes, J. (2015). Created in close interaction with the industry: The smart appliances reference (saref) ontology. In Proceedings of the 7th International Workshop on Formal Ontologies Meet Industry (FOMI 2015), pages 100–112. Springer.

DjangoRESTful (2023). Django rest framework. `https://www.django-rest-framework.org`. [Online; accessed 27-Feb-2023].

Environmental Protection Agency (2025). EPA projections show Ireland off track for 2030 climate targets. `https://www.epa.ie/news-releases/news-releases-2025/epa-projections-show-ireland-off-track-for-2030-clima` `php`. Accessed: 2026-02-26.

ESB Networks (2026). Network Capacity Heatmap. Ac-

cessed: 2026-02-23.

GeoDirectory (2022a). Geoaddress smart data. https://www.geodirectory.ie/products-services/geoaddress-smartdata. [Online; accessed 27-Feb-2023].

GeoDirectory (2022b). Geobuilding intel. https://www.geodirectory.ie/products-services/geobuilding-intel. [Online; accessed 27-Feb-2023].

Haghgoo, M., Sychev, I., Monti, A., and Fitzek, F. H. (2020). Sargon – smart energy domain ontology. IET Smart Cities, 2(4):191–198.

Halpin, H., Hayes, P. J., McCusker, J. P., McGuinness, D. L., and Thompson, H. S. (2010). When owl:sameas isn't the same: An analysis of identity in linked data. In Patel-Schneider, P. F., Pan, Y., Hitzler, P., Mika, P., Zhang, L., Pan, J. Z., Horrocks, I., and Glimm, B., editors, The Semantic Web – ISWC 2010, pages 305–320, Berlin, Heidelberg. Springer Berlin Heidelberg.

He, F., Wang, D., and Sun, Y. (2023). Ontology integration for building systems and energy storage systems. In Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '23), pages 445–448.

Hoare, C., Aghamolaei, R., Lynch, M., Gaur, A., and O'Donnell, J. (2022). A linked data approach to multi-scale energy modelling. Advanced Engineering Informatics, 54:101719.

Internet Engineering Task Force (IETF) (2016). RFC 7946: GeoJSON Specification. Accessed: 2026-02-23.

Irish Academy of Engineering (2025). Rebalancing ireland's energy policy. Technical report, Irish Academy of Engineering, Dublin, Ireland.

Keet, C. and Khan, Z. (2024). Discerning and characterising types of competency questions for ontologies.

Li, Y., García-Castro, R., Mihindukulasooriya, N., O'Donnell, J., and Vega-Sánchez, S. (2019). Enhancing energy management at district and building levels via an em-kpi ontology. Automation in Construction, 99:152 – 167.

Madnick, S. E., Wang, R. Y., Lee, Y. W., and Zhu, H. (2009). Overview and framework for data and information quality research. J. Data and Information Quality, 1(1).

Nacht, T., Pratter, R., Ganglbauer, J., Schibline, A., Aguayo, A., Fragkos, P., and Zisarou, E. (2023). Modeling approaches for residential energy consumption: A literature review. Climate, 11(9).

Ngomo, A.-C. N., Auer, S., Lehmann, J., and Zaveri, A. (2014). Introduction to Linked Data and Its Lifecycle on the Web, pages 1–99. Springer International Publishing, Cham.

Office, C. S. (2022). Small Area Population Statistics. Accessed: 2026-02-23.

Pauwels, P., Zhang, S., and Lee, Y.-C. (2017). Semantic web technologies in aec industry: A literature overview. Automation in Construction, 73:145–165.

Pritoni, M., Paine, D., Fierro, G., et al. (2021). Metadata schemas and ontologies for building energy applications: A critical review and use case analysis. Energies, 14(7):2024.

Project, D. (2023a). Django project. https://www.djangoproject.com. [Online; accessed 27-Feb-2023].

Project, D. (2023b). Django project. https://www.djangoproject.com. [Online; accessed 27-Feb-2023].

Rasmussen, M. H., Lefrançois, M., Schneider, G. F., and Pauwels, P. (2021). Bot: the building topology ontology of the w3c linked building data group. Semantic Web, 12(3):437–461.

Santos, R., Costa, Á., Silva, J., et al. (2025). Integrating machine learning and digital twins for enhanced smart building operation and energy management: A systematic review. Urban Science, 9(6):202.

SEAI (2023). Ber research too. https://ndber.seai.ie/BERResearchTool/ber/search.aspx. [Online; accessed 27-Feb-2023].

Skillington, K., Crawford, R. H., Warren-Myers, G., and Davidson, K. (2022). A review of existing policy for reducing embodied energy and greenhouse gas emissions of buildings. Energy Policy, 168:112920.

Social Justice Ireland (2025). National housing plan 2025-2030: Submission to department of housing, local government and heritage. Technical report, Social Justice Ireland, Dublin, Ireland.

Sood, D., Alhindawi, I., Ali, U., McGrath, J. A., Byrne, M. A., Finn, D., and O'Donnell, J. (2023). Simulation-based evaluation of occupancy on energy consumption of multi-scale residential building archetypes. Journal of Building Engineering, 75:106872.

Swan, L. G. and Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renewable and sustainable energy reviews, 13(8):1819–1835.

Tardioli, G., Kerrigan, R., Oates, M., O'Donnell, J., and Finn, D. (2017). A data-driven modelling approach for large scale demand profiling of residential buildings. In Proceedings of BS2017: 15th Conference of International Building Performance Simulation Association, San Francisco, CA, USA.

Wang, M., Lilis, G. N., Mavrokapnidis, D., et al. (2024). A knowledge graph-based framework to automate the generation of building energy models using geometric relation checking and hvac topology establishment. Energy and Buildings, 325:115035.

Wu, Z., Cheng, J. C. P., and Wang, Z. (2024). An ontology-based framework for building energy simulation in the operation phase. In Advances in Information Technology in Civil and Building Engineering: Proceedings of ICCCBE 2022, volume 357 of Lecture Notes in Civil Engineering, pages 351–366. Springer Nature Switzerland.

Education, Policy & Standardisation

Sub-Tracks:

Digital Skills Education & Curriculum Design

Innovative Pedagogies & Immersive Learning Platforms

Continuous Professional Development & Training

Policy, Regulations & Incentives for Digital Construction

Standardisation & Interoperability Frameworks

Ethics, Governance & Data Policy in AEC

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

TOWARDS AN INDUSTRY-ACADEMIA COLLABORATION FRAMEWORK FOR THE DEVELOPMENT OF DIGITAL GEOSPATIAL SURVEYING MICRO-CREDENTIALS

Brian Graham[1], Bryan Hallissey[1], and Shane Murtagh[2]
[1]South East Technological University, Waterford, Ireland
[2]KOREC Group, Dublin, Ireland

Abstract

The geospatial surveying sector is undergoing rapid digital transformation, intensifying demand for advanced digital skills amid widespread workforce shortages and declining education pathways. Micro-credentials have emerged as a flexible solution for targeted, industry-aligned upskilling, yet their effective development requires structured collaboration between Higher Education Institutions and industry. This paper proposes a conceptual framework for designing Digital Geospatial Micro-Credentials through an Industry–Academia Collaboration model, synthesising multidisciplinary literature and preliminary action research. The framework integrates key drivers, collaborative patterns, and outputs, offering a foundation for future validation and enabling responsive approaches to developing digitally skilled geospatial professionals.

Introduction

The Architecture, Engineering & Construction (AEC) sector is currently undergoing a major digital transformation. Known as Construction 4.0, this paradigm shift is seeking to enhance efficiency, productivity, sustainability, safety, and collaboration (Build Up, 2024). In a study of Construction 4.0 adoption in Australia, Siriwardhana et al. (2025) identified the integration of several key technologies including BIM, IoT, digital twins, AI, AR/VR, Blockchain, Robotics, 3D Printing, Big Data and Drones. EU policy initiatives such as public procurement mandates are attempting to overcome the challenges associated with Construction 4.0 which include high initial investment, training needs, and resistance to change (Build Up, 2024). The geospatial surveying sector is also undergoing a profound transformation driven by digitalization. Traditional methods are being replaced by advanced technologies such as LiDAR, UAVs (drones), AR/VR, GNSS, and cloud-based GIS platforms. These tools enable faster data collection, higher accuracy, and real-time processing, allowing surveyors to deliver dynamic, interactive spatial models rather than static maps. To integrate effectively with Construction 4.0, modern geospatial surveyors need strong digital skills, spatial awareness, geodetic knowledge, interdisciplinary collaboration and proficiency in advanced tools like reality capture, GIS software, and data management systems (Trimble, 2025).

Geospatial Skills Shortages

The challenges associated with the rapid digital transformation of the geospatial sector is further exacerbated by a significant skills shortage which has existed for many years, with de Róiste (2014) and Masum et al. (2017) highlighting this issue in Ireland, New Zealand, Australia, and the United States. This problem persists with the most recent report from the Society of Chartered Surveyors in Ireland (SCSI) suggests a continued shortage of land surveyors in Ireland (Murphy, 2023), with both Site Engineers and Setting Out Engineer listed on Irish Government's Critical Skills Occupations List (DoETE, 2024). Martin (2024) suggests that the significant deficit in the number of geospatial professionals in Ireland is due to several factors including a lack of knowledge of geospatial surveying as a discipline, that it is not a protected professional title and the declining number of secondary students entering education programmes. Similarly, the global geospatial industry is facing a critical shortage of skilled surveyors due to an aging workforce and insufficient new entrants (Trimble, 2025). This is further compounded by a decline in surveying education programmes, with the SCSI and the Council of European Geodetic Surveyors as recently as October 2025 expressing *"serious concern over the trend of various universities across Europe closing their educational programmes related to land surveying."* They continue by stating that *"future professionals must have access to high-quality, innovative education and training."*

Higher Education & Micro-Credentials

The higher education sector is also facing growing challenges due to rapid technological change, evolving professional needs in the workplace and the digital transformation of society (Urem et al., 2025). This has expedited the demand for short, targeted, flexible upskilling/reskilling beyond traditional degrees (Walsh Shanahan & Organ, 2022; Islam, 2025). Employers are increasingly seeking immediately applicable digital skills, pushing Higher Education Institutes (HEIs) toward workforce-aligned modular learning (Urem et al., 2025). Micro-credentials, which are defined differently across contexts but share common goals: upskilling and reskilling to meet rapidly changing digital skill demands, provide an opportunity to address such challenges. They

are short, flexible, industry-aligned learning units that can stack toward higher qualifications, supporting lifelong learning through competency certification (Moore et al., 2025). In this regard, both Islam (2025) and Walsh Shanahan & Organ (2022) emphasize the potential for developing micro-credentials in collaboration between academic institutions and industry stakeholders to guarantee workforce alignment, ensure relevance and uptake, meet market needs and foster innovation.

Opportunities in Digital Geospatial Education

Siriwardhana et al. (2025) advocates for the integration of Construction 4.0 concepts such as digital surveying technologies into formal education through collaboration between industry and academia. The potential benefits of industry-academic collaboration (IAC) for geospatial education are also highlighted by de Róiste (2014: 187), who states that it is: *"likely to lead to a greater professionalization and therefore a wider recognition of both the industry and those working in it. The benefits of gaining a qualification (after* employment*) or re-skilling in emerging areas need to be sold to both employees and their organisations."* This is confirmed by Cox (2025) who highlights the shift from traditional classroom-based models to flexible, hybrid, work-integrated learning pathways that align with industry needs. Trimble (2025) also cites the potential of IAC with an emphasis on micro-credentials, and continuous professional development to prepare a digitally skilled geospatial workforce.

Towards a Conceptual Framework?

The rapid digital transformation taking place within geospatial surveying sector, coupled with a significant skills shortage and a decline in formal education programmes, points towards a need for the development of upskilling opportunities for existing professionals within the AEC industry. Through a systematic literature review of 56 peer-reviewed articles on micro-credentials in HE published between 2015 and 2021, Ahsan et al. (2023) found that micro-credentials enhance employability through alignment with market needs, improving graduate employability and bridging skills gaps. They also suggested that micro-credentials be co-designed with industry to ensure relevance and acceptance, thus confirming the importance of IAC. According to Iatrellis et al. (2024), this trend highlights the need for structured, sustainable frameworks to support micro-credential programs and adapt to the changing educational landscape. With micro-credentials gaining traction as a potential solution in delivering enhanced digital skills, this research seeks to establish if an IAC framework for micro-credential design can effectively close the digital geospatial skills gap. This will be achieved by adopting a conceptual research design aimed at synthesizing existing theoretical and scholarly perspectives to create a new Conceptual Framework (CF) for the development of Digital Geospatial Micro-Credentials via IAC, based upon ongoing action research which is being led by South East Technological University (SETU) and KOREC Group, a geospatial surveying services provider in Ireland. In building a CF for multidisciplinary phenomena (i.e. digital geospatial surveying, micro-credentials and IAC) linked to different bodies of knowledge, the approaches proposed by both Jabareen (2009) and Waller (2022) have been adopted in the present study (as outlined in Table 1).

Table 1: Development of a Conceptual Framework

Present Study	Jabareen (2009)	Waller (2022)
Review of Multidiscipline Literature	Mapping sources Extensive reading and categorizing of selected data	Review: literature/themes/ theory
Main Concepts Arising	Identifying, naming, deconstructing & categorizing concepts	Reflect: what are the main concepts/issues?
Relationships and Integration	Integrating concepts	Relationships: what are their relationships?
Reflection & Synthesis into CF Diagram	Synthesis, resynthesis, and sensemaking	Reflect: does the diagram represent it sufficiently?
Validation of CF	Validating the conceptual framework	Review: check it with theory, colleagues, stakeholders, etc.
Revision of CF	Rethinking the conceptual framework	Repeat: review and revise it to see if something better occurs

The intention of this paper and associated conference presentation is to encourage a review, validation and revision of the framework as a work in progress to ascertain *"whether the proposed framework and its concepts make sense not only to the research but also to other scholars and practitioners" (Jabareen (2009: 54).* The paper is structured to reflect the work of both Jabareen (2009) and Waller (2002), with a Literature Review (multidisciplinary literature and main concepts arising), followed by the Development of Conceptual Framework (integration and synthesis into diagram). It is anticipated that further validation and revision of the CF will take place as the research progresses in future. The aim of the CF is to integrate currently disconnected bodies of knowledge—digital geospatial surveying, micro-credential design, and industry–academia collaboration—into a coherent model that can guide HEIs and industry partners in the structured co-development of digital geospatial micro-credentials.

Literature Review

Utilizing relevant scholarly and industry sources, the following review seeks to understand the underlining issues relating to the geospatial surveying profession, the

role of micro-credentials in higher education and the nature of Industry-Academia Collaboration (IAC), leading to the identification of the main concepts arising.

Geospatial Surveying

Geospatial (land) surveying is the science of collecting, analyzing, and presenting spatial information about natural and built environments. Surveyors measure land and sea areas to map boundaries, features, and terrain for purposes such as GIS, construction, and legal disputes. The field is highly technological, using GNSS, LiDAR, drones, and 3D modelling across three main pathways: land, engineering (setting out buildings & infrastructure), and hydrographic (marine) surveying (RICS, 2023). Geospatial professionals are defined by de Róiste (2014) in terms of acquiring, integrating, analyzing, interpreting, presenting, managing, and distributing spatial and temporal information, as well as developing related equipment, software, and services. A major study of geospatial skills in the UK identified "Surveyors" as one of six geospatial occupational subgroups with an emphasis on capabilities including land surveying, field surveys, topographic surveys, use of survey instruments and advanced technologies such as LiDAR (Frontier Economics, 2020). In terms of the Land Surveying profession, CICES (2023) defines the core specialist competences as follows:

1. Undertake and produce topographic surveys using appropriate methodologies and controls
2. Proficient use of surveying instruments (including total stations, GNSS, laser scanners, levels, utility mapping tools, photogrammetry & drones)
3. Understand Geometric Principles
4. Geospatial Data Processing & Deliverables

According to Frontier Economics (2020), typical job titles include Land Surveyor, Utility Surveyor, Hydrographic Surveyor, and Site Engineer, spanning sectors such as Architecture, Engineering and Construction (AEC), geographical information, environmental services, land administration and management, cultural and natural heritage and renewable energy.

In discussing the evolution of the surveying profession, Masum et al. (2017) note the significant impact of technology from the introduction of electronic instruments in the 1960s to GPS in the 1990s and more recent advances with laser scanning and photogrammetric data capture. Geospatial skills encompass a broad range of technologies, including spatial modelling, surveying, Global Positioning Systems (GPS), Geographic Information Systems (GIS) as well as traditional cartography techniques (de Róiste, 2014). Digital transformation of the geospatial surveying sector is continuing at pace, with reality capture (SLAM and 3D laser scanning); cloud-based data management and AI & automation being the key technologies driving change (Trimble, 2025). Recent industry surveys have confirmed that digital skills are in highest demand for geospatial technology professionals, including data capture/collection; data processing & analysis; and data visualization (Frontier Economics, 2020). Siriwardhana et al. (2025) emphasizes that workforce skills are a critical enabler for successful Construction 4.0 adoption, arguing that digital transformation in construction requires a shift from traditional skill sets to advanced competencies such as digital literacy, data analytics, computational and parametric modelling, programming, and AI integration.

Status of Geospatial Education Programs in Ireland

The current offering of geospatial related education in Ireland is dominated by post-graduate programmes (NFQ Level 9) in GIS/Geoinformatics at University College Dublin, University College Cork, Maynooth University and Technological University of Dublin. In terms of full-time undergraduate programmes, there is a 1-year full-time Level 5 Certificate in Geospatial Technology delivered by Cork ETB, while the long-running full-time BSc (Hons) in Geospatial Surveying (Level 8) at TU Dublin was paused in September 2024. At present, TU Dublin is developing a Geospatial Survey Technician apprenticeship, a 2-year Level 6 qualification, in collaboration with the SCSI (Martin, 2024). This shift towards geospatial education programmes at Level 5, 6 and 9 points to a significant gap in educational opportunities at Level 7 and 8. With an ongoing lack of new students entering geospatial education programmes, it is suggested that HEIs focus on providing new upskilling/CPD opportunities for current professionals in industry (Masum et al., 2017; de Róiste, 2014). *"CPD is particularly important in a field that is so deeply influenced by technological innovation. Developments in GNSS, laser scanning, drone-based photogrammetry, and remote sensing continue to reshape workflows and methodologies. Without structured and accessible professional development pathways, there is a risk that skills gaps will widen and sector growth will be constrained" (Cox, 2025: 3).*

The main concepts arising from geospatial literature include skills & competencies, digitalization, key stakeholders, industry needs, skills shortage & gap, opportunities for HEIs, and the potential for IAC.

Micro-Credentials in Higher Education

Micro-credentials present the higher education sector with an opportunity to address the market's demand for flexible upskilling; *"as employers and job seekers become more aware of employment-enhancing micro-credentials programs, demand for such short-term upskilling opportunities is expected to increase" (Yildirim et al., 2025: 538).* A relatively recent innovation in the education landscape, micro-credentials are short, focused, quality assured learning programs that concentrate on specific competencies and skills. They are modular, portable, stackable, and are validated through transparent, reliable assessment and quality assurance (Islam, 2025; Moore et al., 2025; Pawilen et al., 2022). Micro-credentials are offered by diverse providers including universities, commercial entities, and professional bodies. This versatility enables micro-credentials to serve both formal education and workplace learning (Moore et al., 2025).

Table 2: Micro-Credential Categories

	Skill Credential	Micro-Credential Module	Short Learning Programme
Learning Effort/Time	1-25 hours 0 ECTS	25-150 hours 1-5 ECTS	150-1500 hours (stacked)
Award & Assessment	Open Badge Non-Formal Learning	Formal Learning with Assessment	Formal Learning with Assessment
Quality Assurance	None	Often External QA	Always External QA
Learning Outcome	Specific Competence	Acquisition of Learning Outcomes	Career Progression Goals
Pathway	N/A	Often Mapped to Qualification Framework	Always Mapped to Qualification Framework

Both employers and learners place significant value on the validation of micro-credentials for workplace recognition, with traditional academic credentials seen to be more trusted (Ashcroft et al., 2021). Walsh Shanahan and Organ (2022), suggest that the use of ECTS can help to standardize workload and recognition of micro-credentials, recommending a minimum of 5 ECTS for formal recognition. Indeed, Ahsan et al. (2023) found substantial evidence of research which examines integration of micro-credentials into formal curricula. Micro-credentials can function as standalone, credit-bearing courses or be integrated into formal qualifications. Sargent et al. (2023) found that they may appear as optional or compulsory components within traditional programs or as stackable micro-credentials, where multiple related short courses combine to form a recognized award (e.g., Graduate Certificate). Qualifications incorporating micro-credentials include traditional certificates, diplomas, and degrees. This trend reflects growing interest in modular learning pathways that allow learners to accumulate credits toward larger qualifications (Sargent et al., 2023). Micro-credentials are typically categorized in terms of learning effort/time, learning outcomes, assessment methods, quality assurance and pathways to further learning, with 3 categories identified as outlined in Table 2 (Camilleri et al., 2022; Walsh Shanahan and Organ, 2022; Urem et al., 2025). Iatrellis et al. (2024) identified the critical roles and responsibilities of key stakeholders within the micro-credentialing ecosystem as follows:

- **HEIs:** design, develop, and deliver micro-credentials; ensure academic standards, industry relevance, and quality assurance.
- **Industry Partners:** identify workforce skills and co-design micro-credentials; provide practical training opportunities.
- **Regulatory Bodies:** accredit and recognize micro-credentials; ensure consistency, quality and alignment with national frameworks.
- **Students and Learners:** provide feedback to refine relevance and effectiveness; participate in content co-creation for career and educational goals.
- **Quality Assurance Agencies:** evaluate programs against standards; recommend continuous improvement, shaping evaluation criteria.
- **Professional Bodies:** advocate for recognition and validation within industries.

For HEIs, micro-credentials offer several benefits, according to Pawilen et al. (2022): increase visibility and build their brand, attract diverse learners through unique and alternative courses, and appeal to these individuals to enroll in their formal education programs. Ahsan et al. (2023) identified several challenges associated with implementing micro-credentials in higher education including student acceptance, institutional readiness, technology gaps, employer recognition, logistical complexity and quality assurance. Employer awareness of micro-credentials is still low but growing, with traditional academic credentials remaining more trusted. Key factors influencing employer acceptance include industry validation, alignment with job needs, prior experience with credential issuers, and the issuer's reputation, credibility and quality assurance (Ashcroft et al., 2021). This presents a significant opportunity for higher education to lead in this area, with Yildirim et al. (2025: 552) stating that *"this shift could enable more flexible learning pathways, responsive curriculum design, and stronger alignment with lifelong learning and workforce development priorities."*

The main concepts arising from the micro-credentials in higher education literature include nature & categories of micro-credentials, development of skills & competencies, addressing industry needs & rapid digital transformation key stakeholders, course development/curriculum design, opportunities for HEIs, and the potential for greater IAC.

Industry-Academia Collaboration (IAC)

Aligning academic curricula with industry needs has long been a challenge for both educators and practitioners (Mahalingam, 2024), with Ahmed et al. (2022) suggesting that bringing in industry experts to assist in developing the curriculum can help to bridge this gap. Both Karanja and Malone (2022) and Mahalingam (2024) suggest that the co-creation of a curriculum with industry can prove beneficial in terms of improving the quality and relevance of academic programmes, whilst equipping graduates with the skills and knowledge required by industry. The most important outcomes of collaboration with industry were identified by Valiente Bermejo et al. (2022) as bridging the gap between academic education and industry requirements, enhanced graduate employability and strengthened research partnerships. Real, structured partnerships are more effective in driving innovation, with previous research demonstrating that open and networked innovation strengthens the potential of participating organizations (Esangbedo et al., 2024). From a digitalization perspective, Akash and Venuraj (2024) note the importance of IAC in terms of integrating state-of-the-

art technologies into the educational curriculum, an area where HEIs are often found to be lacking.

According to Esangbedo et al. (2024), IAC has been widely studied across different countries, including the UK, USA, and Japan. *"IAC is trending as a strategic framework necessary to bridge the increasing gap between the academic programmes and the emergent requirements of the industry" (Akash and Venuraj, 2024: 3).* A review of literature indicates a variety of frameworks, under various guises, focused on collaboration between academia and industry for curriculum development. These frameworks emerged from a number of different disciplines including business (Mahalingam, 2024); Software Engineering Ahmed et al., 2022); Sustainable Development Goals (Akash and Venuraj, 2024); Manufacturing Engineering (Valiente Bermejo et al., 2022); Information Technology (Junaini et al., 2008); and Industry 4.0 (Esangbedo et al., 2024). Mahalingam (2024) developed the industry-driven curriculum (IDC) framework which focuses on bridging the gap between academic curricula and industry requirements through frequent workshops. Key activities on the framework include identifying industry needs, mapping competencies, gap analysis on learning outcomes, course content, schedule of activities and assessment strategies, and developing a competency-based curriculum. Regular follow-up and updates post-workshop are essential in maintaining engagement with industry, whilst also demonstrating the incorporation of feedback can enhance trust and collaboration throughout the process (Mahalingam, 2024). An Academia-Industry Collaboration Plan (AIC) engineering process model is presented by Ahmed et al. (2022), which suggests that effective communication and collaboration in the early stages of engagement is critical for the transformation of proposals into practical applications, leading to a clearly defined "Finalized Collaborative Agreement." Akash and Venuraj (2024) proposed a framework for IAC which aligns student skill development with real-world requirements. It emphasizes dynamic, iterative collaboration between educational institutions and industry to continuously improve and adapt curricula, incorporate emerging technologies, upskill faculty, thereby enhancing relevance, employability, and responsiveness to evolving market demands. Valiente Bermejo et al. (2022) presents a model for university-industry collaboration in curriculum design and delivery, specifically applied to manufacturing engineering courses. A core element of this model are regular workshops with industry partners in developing, teaching, evaluating and improving the course. An Academic-industry smart synergy model was proposed by Junaini et al. (2008) with industry input in terms of market demand and curriculum development being seen as important to delivering relevant educational programs and a competent workforce. According to Esangbedo et al. (2024) research on IAC generally falls into three categories: Drivers (factors that motivate collaboration), Patterns (how collaborations are structured) and Outputs (results and impacts of collaboration). It is evident that collaboration between industry and academia in curriculum design can lead to significant benefits for both parties. To ensure strategic alignment, the reviewed IAC frameworks suggest that regular communication, collaboration and feedback is essential to ensure the development of competencies, curriculum, content, and assessment which leads to relevant courses that deliver the skilled graduates required by industry.

The main concepts arising from the IAC literature include addressing rapid digital transformation and industry needs, bridging the skills gap, collaboration and communication, key stakeholders, curriculum design and opportunities for HEIs.

Developing the Conceptual Framework

Having identified the main concepts arising from literature, the next stage in the development of a CF involves identifying relationships and integrating concepts, followed by reflection and synthesis inro a CF diagram. It is evident from the main concepts arising that there are a number of overarching concepts that cut across the three domains including digital transformation, skills shortages and educational gaps, micro-credentials for flexible learning and IAC as an enabling mechanism.

Relationships and Integration

The main concepts arising from each of the three literature topics were reviewed and analyzed using a matrix, allowing for identification of relationships between geospatial surveying, micro-credentials and IAC. Jabareen (2009: 54) suggests that *"this phase reduces the number of concepts drastically and allows us to manipulate to a reasonable number of concepts."*

Table 3: Integration of Concepts

Drivers (factors that motivate collaboration)	Patterns (how collaborations are structured)	Outputs (results and impacts of collaboration)
Digital Transformation & Construction 4.0	Industry-Academic Collaboration	Opportunities for HEIs
Decline in Geospatial Education	Active Engagement from Key Stakeholders	Benefits for Industry
Evolving Needs of Geospatial Profession	Key Stages/Activities	Digital Geospatial Micro-Credentials
Changing Landscape for HEIs	Curriculum Design/Course Development	Upskilling of Geospatial Profession

The matrix was developed by listing all identified concepts from each literature strand and grouping them based on conceptual proximity. Concepts were iteratively merged, eliminated, or re-named to achieve parsimony, consistent with Jabareen (2009). Through this analytical process, concepts were refined and integrated using the three categories of IAC research as identified by Esangbedo et al. (2024, namely: Drivers (factors that motivate collaboration), Patterns (how collaborations are

structured) and Outputs (results and impacts of collaboration). These three categories form the basis for integration as they are widely used in IAC studies and enable the structured organization of complex multi-stakeholder processes. This led to the integration and refinement of concepts as evidenced in Table 3 below.

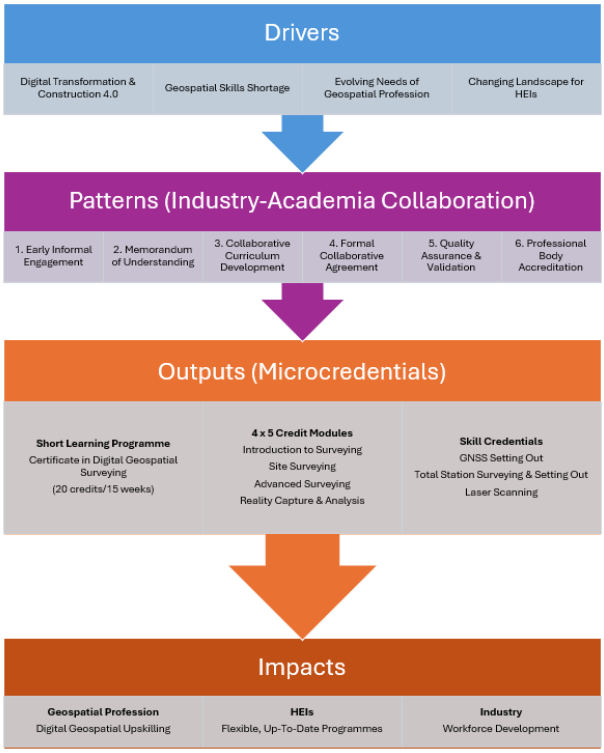
Reflection and Synthesis

Utilizing the information in Table 3, along with reflecting upon a number of Conceptual Frameworks considered as part of the literature review for both IAC and micro-credentials, this stage sought to synthesize this information into a visual representation of the emerging CF. The framework illustrates how sector-level drivers—such as Construction 4.0 digitalisation, geospatial skills shortages, evolving professional expectations, and shifts within HEI create a clear need for structured Industry–Academia Collaboration (IAC). These forces collectively shape a six-stage collaboration pathway that includes:

1. **Early Informal Engagement:** initial conversations between industry partners and the HEI where needs, interests, and emerging skill gaps are explored; builds trust and determines if collaboration is worthwhile.
2. **Memorandum of Understanding:** once mutual interest is established, partners formalise intentions through a non-binding agreement. This clarifies roles, expectations, and broad collaboration objectives without locking in detailed commitments.
3. **Collaborative Curriculum Development:** industry and HEI teams work together to design curriculum content, define competencies, shape assessments, and ensure learning aligns with real-world technologies and workflows. This stage is central to ensuring relevance and authenticity.
4. **Formal Collaborative Agreement:** partners move from intent to commitment through a binding agreement covering governance, responsibilities, IP considerations, resource inputs, scheduling, and shared oversight of teaching and assessment.
5. **Quality Assurance & Validation:** the collaboratively designed programme undergoes internal HEI QA processes; external assessor review, NFQ alignment and institutional approval, ensuring academic integrity and compliance.
6. **Professional Body Accreditation:** to review and endorse the programme or micro-credentials, confirming industry recognition and signalling credibility to learners and employers.

From these collaboration patterns emerge tangible outputs, including a 20-credit Certificate in Digital Geospatial Surveying (NFQ Level 7), four 5-credit modules, and skills-focused micro-credentials (e.g., GNSS, Total Station, Laser Scanning). These outputs impact across three domains: upskilling the geospatial profession, enabling HEIs to remain flexible and current, and strengthening industry workforce development. The proposed micro-credentials will be underpinned by competencies drawn from established geospatial frameworks (e.g., CICES competencies). Assessment is expected to follow practice-based tasks focused on data capture, processing accuracy, and applied problem-solving. A blended delivery model is envisaged, combining online preparation with equipment-based practical sessions co-delivered by KOREC Group staff. Evaluation in later phases of the research will consider indicators such as learner uptake, completion, employer recognition, and evidence of workplace application. These assumptions are preliminary but outline the pedagogical direction supporting the proposed IAC-driven micro-credential design.

Figure 1: Emerging Conceptual Framework



Future Work

The CF presented reflects the first iteration of a dynamic and ongoing collaborative process between SETU and KOREC Group. Currently, the IAC is at the Collaborative Curriculum Development stage, with an outline of the micro-credentials (output) in place. It is anticipated that the IAC CF will act as a catalyst for further development of both the framework and the micro-credentials with additional research involving:

- Validation focus groups with SETU & KOREC staff
- Interviews with surveyors/employers for input
- Cross-checking with existing frameworks

It is anticipated that empirical research will lead to the validation of a more robust IAC framework which might have applicability in other AEC contexts/professions.

Conclusions

This paper addresses the urgent challenge posed by the rapid digital transformation of the geospatial surveying profession and the lack of flexible educational pathways to develop required digital competencies. Existing literature on geospatial skills, micro-credentials, and Industry–Academia Collaboration (IAC) highlights a

persistent misalignment between evolving industry needs and traditional higher-education provision. In response, this study develops a conceptual framework for Digital Geospatial Micro-Credentials, grounded in multidisciplinary evidence and iterative action research between an Irish HEI and a Geospatial Surveying Services Provider. In terms of a novel contribution, the framework synthesises three previously disconnected domains: (1) digital and professional changes in geospatial practice, (2) micro-credentials as agile, workplace-aligned learning mechanisms, and (3) IAC frameworks that emphasise co-design and continuous feedback. Structured around a Drivers–Patterns–Outputs logic, it clarifies the motivations for collaboration, the mechanisms through which it should occur, and the expected impacts on workforce development. This integration contributes a coherent framework that addresses skills shortages, curriculum relevance, and the need for adaptive educational offerings. Limitations include its basis in secondary literature and early-stage partnership activity, requiring further empirical validation, as well as its initial grounding in the Irish geospatial context, which may affect broader applicability. Future work should test the framework in additional sectors, regions, and educational systems to strengthen generalisability. Next steps involve validating and operationalising the framework through co-design workshops, practitioner interviews, focus groups, and pilot micro-credential evaluation. These activities will refine framework components and identify enabling conditions for effective IAC in digital skills development, with potential extensions to national qualification systems, accreditation processes, and European micro-credential initiatives. In conclusion, this study offers a significant first step toward addressing the digital geospatial skills gap through a structured, collaborative, and evidence-based approach to micro-credential design. By integrating previously siloed literatures, the framework provides a foundation for developing relevant, responsive, high-quality upskilling opportunities aligned with the realities of a digitally transformed AEC sector, with further refinement expected as the research progresses. Future work will operationalise and empirically evaluate the pedagogical dimensions of the framework, including competency mapping, assessment design, delivery modality, and indicators of success such as uptake, completion, employer recognition, and workplace impact. These activities will strengthen the framework's validity and ensure its practical applicability in addressing digital geospatial skills shortages, with the potential to address wider challenges relating to the integration of Construction 4.0 technologies into formal education through collaboration between industry and academia.

References

Ahmed, F., Fattani, M.T., Ali, S.R. and Enam, R.N. (2022) 'Strengthening the bridge between academia and industry through the academia–industry collaboration plan design model', *Frontiers in Psychology*, 13, p. 875940.

Ahsan, K., Akbar, S., Kam, B. and Abdulrahman, M.D. (2023) 'Implementation of micro-credentials in higher education: A systematic literature review', *Education and Information Technologies*, 28, pp. 13505–13540.

Akash, R. and Venuraj, S. (2024) 'Bridging the gap between industry needs and student skills for quality education through SDGs: An industry–academia collaboration in curriculum development', *Journal of Lifestyle and SDGs Review*, 4, e03616.

Ashcroft, K., Etmanski, B., Fannon, A.M. and Pretti, T.J. (2021) 'Microcredentials and work-integrated learning', *International Journal of Work-Integrated Learning*, 22(3), pp. 423–432.

BUILD UP (2024) 'The digital transformation of the built environment and Construction 4.0', *European Commission BUILD UP*, 1 February. Available at: https://build-up.ec.europa.eu/en/resources-and-tools/articles/ (Accessed: 30 December 2025).

Camilleri, A.F., Darmanin, M., Kamšek, K. and Poličnik, J. (2022) *Guide to Design, Issue and Recognise Micro-credentials*. European Training Foundation.

CICES (2023) *Specialist Land Surveying Competencies – March 2023 Revision*. Chartered Institution of Civil Engineering Surveyors. Available at: https://www.cices.org (Accessed: 24 October 2025).

Cox, P. (2025) *Empowering the Next Generation of Surveyors: Innovations in Technical Education, Professional Standards, and Sustainable Development in Australia*. FIG Working Week 2025, Brisbane, Australia, 6–10 April. Available at: https://fig.net/resources/proceedings/fig2025 (Accessed: 30 December 2025).

de Róiste, M. (2014) 'Filling the gap: The geospatial skills shortage in New Zealand', *New Zealand Geographer*, 70(3), pp. 179–189.

DoETE (2024) *Critical Skills Occupations List* [online]. Department of Enterprise, Trade and Employment. Available at: https://enterprise.gov.ie (Accessed: 30 December 2025).

Esangbedo, C.O., Zhang, J., Esangbedo, M.O., Kone, S.D. and Xu, L. (2024) 'The role of industry–academia collaboration in enhancing educational opportunities and outcomes under the digital-driven Industry 4.0', *Journal of Infrastructure, Policy and Development*, 8(1), p. 2569.

Frontier Economics (2020) *Demand for Geospatial Skills: Report for the Geospatial Commission*. London: Frontier Economics Ltd. Available at: https://www.gov.uk/government/publications/geospatial-skills-demand-report (Accessed: 24 October 2025).

Iatrellis, O., Samaras, N. and Kokkinos, K. (2024) 'Towards a capability maturity model for micro-credential providers in European higher

education', *Trends in Higher Education*, 3(3), pp. 504–527.

Islam, M.S. (2025) 'The emerging role of microcredentials in higher education: Advancing public health learning and beyond', *Journal of Biosciences and Public Health*, 1(1), pp. 43–54.

Jabareen, Y. (2009) 'Building a conceptual framework: Philosophy, definitions, and procedure', *International Journal of Qualitative Methods*, 8(4), pp. 49–62.

Junaini, S.N., Fadzir, S.F.S., Sidi, J., Khiri, M.J.A. and Othman, R.M. (2008) 'Harnessing university–industry collaboration in Malaysia through industrial training', in *3rd International Conference on Information and Communication Technologies: From Theory to Applications*, Damascus, Syria, pp. 1–5.

Karanja, E. and Malone, L.C. (2022) 'The role of industry and academia partnership in improving project management curriculum and competencies', *Journal of Economic and Administrative Sciences*, 38(4), pp. 667–691.

Mahalingam, T. (2024) 'Bridging the gap between academia and industry: A case study of collaborative curriculum development', *International Journal of Business Performance Management*, 25(4), pp. 589–603.

Martin, A. (2024) *Geospatial Surveying HR Challenges in Ireland*. EuroSDR Workshop: Recruitment and Capacity Development Challenges in the Geospatial Domain, University of Ljubljana, 19 November. Available at: https://www.eurosdr.net/workshops/workshop-recruitment-and-capacity-development-challenges-geospatial-domain (Accessed: 31 December 2025).

Masum, F., Groenendijk, L., Mansberger, R. and Martin, A. (2017) *Enhancing the Role of Surveyors: Bridging the Gap between Demand for and Supply of Professional Education*. FIG Working Week 2017, Helsinki, 29 May–2 June. Available at: https://www.fig.net (Accessed: 10 December 2025).

Moore, R.L., Lee, S.S., Taylor Pate, A. and Wilson, A.J. (2025) 'Systematic review of digital microcredentials: Trends in assessment and delivery', *Distance Education*, 46(1), pp. 8–35.

Murphy, R. (2023) *Employment, Remuneration and Workplace Report 2023*. Dublin: Society of Chartered Surveyors Ireland.

Pawilen, G.T., Lubong, K.S. and Fausto, C.J.R.S. (2022) 'A proposed model in designing curriculum for microcredentials in the Philippines', *International Journal of Curriculum and Instruction*, 15(1), pp. 508–520.

RICS (2023) *All about Surveyors Who Work in Land and Resources*. Royal Institution of Chartered Surveyors, 6 April. Available at: https://www.rics.org/surveyor-careers/surveying/what-surveyors-do/guide-to-a-career-in-the-land-and-resources-sector.

Sargent, J., Rienties, B., Perryman, L-A. and FitzGerald, E. (2023) 'Investigating the views and use of stackable microcredentials within a postgraduate certificate in academic practice', *Journal of Interactive Media in Education*, 2023(1), p. 9.

Siriwardhana, S.D., Moehler, R.C. and Fang, Y. (2025) 'Construction 4.0 in Australia: Evaluating technological implementation, organisational adaptation and skills development', *Journal of Innovation & Knowledge*, 10, Article 100822.

Trimble (2025) *State of the Geospatial Industry 2025*. Sunnyvale, CA: Trimble Inc. Available at: https://go2.trimble.com/state-of-geospatial (Accessed: 30 October 2025).

Urem, F., Kardum Goleš, I. and Pavelić, M. (2025) 'Modernizing computer science education with micro-credentials', in *15th International Conference on Advanced Computer Information Technologies (ACIT)*. IEEE, pp. 1026–1029.

Valiente Bermejo, M., Eynian, M., Malmsköld, L. and Scotti, A. (2022) 'University–industry collaboration in curriculum design and delivery: A model and its application in manufacturing engineering courses', *Industry and Higher Education*, 36(5), pp. 615–622.

Waller, D. (2022) 'How to develop a graphical framework to chart your research', *Times Higher Education*, 31 March. Available at: https://www.timeshighereducation.com (Accessed: 22 January 2026).

Walsh Shanahan, B. and Organ, J. (2022) 'Harnessing the benefits of micro-credentials for Industry 4.0 and 5.0: Skills training and lifelong learning', *IFAC PapersOnLine*, 55(39), pp. 82–87.

Yildirim, H., Kir, Ş. and Bozkurt, A. (2025) 'Towards a future of qualifications: A systematic review of micro-credentials', *Online Learning*, 29(4), pp.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

OBJECT-LEVEL MULTI-DIMENSIONAL ANALYSIS OF LEVEL OF INFORMATION REQUIREMENTS ACROSS PROJECT LIFECYCLE

Pablo Martinez[1] and Marzia Bolpagni[1]
[1]Northumbria University, United Kingdom

Abstract

Optimizing multi-purpose specifications requires understanding information requirement overlap and burden distribution, yet systematic object-level analysis remains absent. This study introduces a hierarchical 12-KPI framework quantifying geometrical, alphanumerical information, and documentation requirements across 80 objects, 22 purposes, and five lifecycle phases using 5,048 specifications. The key contributions include: (1) a validated three-archetype hierarchy with distinct temporal signatures; (2) the identification of universal objects as multi-purpose integration points; (3) the quantification of purpose-specific uniqueness challenging consolidation assumptions; (4) an interactive dashboard enabling evidence-based Exchange Information Requirement optimization. This object-purpose-phase granularity advances systematic BIM specification from aggregate analysis to decision-support frameworks for practitioners.

Introduction

The adoption of the ISO 19650 series as international standards for information management using BIM creates expectations for standardised information delivery across project stakeholders (Abanda et al., 2025). This regulatory push, occurring alongside intensifying BIM mandates, creates substantial pressure for comprehensive, multi-purpose implementation serving diverse information needs throughout building lifecycles (Mitera-Kiełbasa & Zima, 2024). Each construction phase requires information in different forms, and what is required changes with the activities performed. Information models must simultaneously serve design coordination, regulatory compliance, energy analysis, cost estimation, construction planning, and facility management, each potentially requiring different information requirements as specified by stakeholders.

ISO 7817-1:2024 addresses this through the Level of Information Need (LOIN) framework, structuring requirements across three components: geometrical information (detail, dimensionality, location, appearance and parametric behaviour), alphanumerical information (properties and attributes), and documentation (deliverable files). Although "LOIN" is an acronym to be avoided in projects, it is used here for simplicity. The standard also introduces purposes, the specific reasons for information exchange, as the main organising principle for requirement specification. However, implementing LOIN requires decisions about which purposes to include, which building objects to specify for each, and how requirements evolve across phases from design through operation (Lindholm et al., 2025; Liu et al., 2023).

Despite extensive BIM research, systematic understanding of information requirements at object-purpose-phase granularity remains limited. Existing studies analyse requirements at aggregate levels, treating purposes as monolithic without examining variation across object types. Cavka et al. (2017) developed comprehensive owner requirement frameworks but did not differentiate by object. Krijnen and Tamke (2015) analysed implicit knowledge in models using machine learning yet focused on model-level patterns. Patacas et al. (2020) evaluated IFC and COBie for asset management but aggregated across objects. Phase-based studies investigate limited purpose sets or restricted object types (Noardo et al., 2020), and recent infrastructure work acknowledges substantial cross-discipline and cross-phase variation but lacks systematic object-level quantification.

More recent work has begun to operationalise LOIN. Malla et al. (2023) proposed a structured approach aligned with ISO 7817-1 yet remained at the purpose level without disaggregating by object. Lindholm et al. (2025) demonstrated LOIN-driven exchange connected to production planning, confirming applicability but limiting analysis to a single purpose context. EU-funded projects such as BIM4EEB and DigiChecks produced LOIN-based templates for energy retrofitting and regulatory compliance respectively, but address domain-specific purposes without cross-purpose comparison (Rathnasinghe et al., 2022). Quantitative or KPI-based assessment remains scarce: existing metrics focus on model maturity or compliance scoring rather than characterising the structure and distribution of requirements. To date, no comprehensive analysis examines how requirements for specific objects vary across multiple purposes and lifecycle phases.

This gap prevents evidence-based optimisation of multi-purpose specifications. Practitioners selecting purposes for Exchange Information Requirements (EIR) face unanswered questions: Do different purposes require fundamentally different information for the same objects, or do requirements largely overlap, enabling duplicity across models? Which objects accumulate the heaviest specification when multiple purposes converge? Can

shared baseline specifications satisfy multiple purposes, or are requirements genuinely purpose-specific? How do these patterns evolve from preliminary design through facility operation? Without systematic object-level evidence, decisions rely on institutional experience, vendor recommendations, or conservative over-specification, none enabling sustainable data-driven optimisation balancing coverage against practical constraints, overcomplicating BIM implementation and reducing data management efficiencies.

This study addresses this gap through systematic multi-dimensional analysis at object-purpose-phase granularity. This paper examines standardized information delivery purpose specifications across building object types and project phases aligned with RIBA Plan of Work 2020 (PoW), quantifying geometrical, alphanumerical, and documentation requirements for discrete object-purpose-phase sequential combinations. This granular approach enables investigation of requirement variation patterns, multi-purpose object utilization, and consolidation potential not addressable through aggregate analysis. An interactive dashboard complements the analytical framework, supporting practitioners in exploring purpose combinations and assessing marginal contributions for project-specific specification decisions.

This analysis addresses five research questions:

RQ1: How do information requirements vary across purposes, phases, and object types?

RQ2: What is the distribution of geometrical, alphanumerical, and documentation requirements across information delivery purposes?

RQ3: Which building objects exhibit highest information specification burden across multiple purposes?

RQ4: To what extent do individual objects support multiple information delivery purposes across project phases?

RQ5: What is the uniqueness of purpose-specific requirements at object-phase granularity?

Methodology

Data Source

The analysis utilized all 22 standardized information delivery purposes from BIMids, an industry-academic collaboration platform that supports the implementation of ISO 19650 series and ISO 7817-1 standards in construction projects. BIMids is maintained through collaboration between industry practitioners and academic researchers, with specifications developed through expert consensus and aligned directly with ISO 7817-1 LOIN component structures. It represents the largest publicly available multi-purpose LOIN specification dataset, providing systematic coverage across building object types, project phases, and information delivery purposes, supporting its relevance as an empirical basis for cross-purpose analysis. Each purpose specifies information requirements for building object types across three LOIN components (geometrical, alphanumerical, and documentation) and five project phases mapped to RIBA PoW 2020: Phase 1-PD (RIBA 0-2: Strategic Definition through Concept Design), Phase 2-FD/AUT (RIBA 3-4: Spatial Coordination through Technical Design), Phase 3-FP/TEND (RIBA 4-5: Tender through Early Construction), Phase 4-EXE/ASB (RIBA 5: Construction), Phase 5-OPE (RIBA 6-7: Handover through Use).

The dataset encompasses 80 building object types spanning architectural (door, window, wall, etc.), structural (beam, column, slab, etc.), and MEP (duct, pipe, lift, etc.) disciplines. In the end, a dataset is created comprising 5,048 requirements: 2,443 geometrical, 1,866 alphanumerical, 739 documentation (distributed across 2,603 object-purpose-phase combinations).

Interactive Dashboard

To enable practical application, an interactive web dashboard was developed to enable data access and serve as a combinatory playground for practitioners. The tool embeds all 22 purposes with complete object-level data (5,048 requirements) enabling offline browser-based analysis. Implementation uses vanilla JavaScript and Chart.js for universal compatibility.

Users may select purposeful combinations, and the dashboard shows the variability in information requirements based on the delivery purposes. For unselected purposes, the dashboard displays a marginal score (new requirements added) and uniqueness percentage (proportion not already covered by selected purposes), ranking recommendations by contribution value. The intended workflow follows four steps: (1) select initial information delivery purposes reflecting project requirements; (2) review real-time KPI updates showing cumulative coverage, component distribution, and object-level burden for the selected set; (3) evaluate the marginal contribution panel, which ranks unselected purposes by new requirements added and uniqueness percentage; and (4) iterate by adding or removing purposes until the desired coverage-to-overhead balance is achieved.

The dashboard supports exploratory investigation of purpose combinations, helping practitioners make evidence-based decisions about EIR specifications while visualizing information overhead implications.

LOIN-based KPIs

To address the research questions systematically, a hierarchical framework of KPIs operating at three analytical levels is proposed: purpose level (characterizing overall information delivery needs), object level (identifying integration points where multiple purposes converge), and combination level (quantifying granular requirement patterns). This hierarchical structure enables investigation from macro-patterns (which purposes are geometry-dominant?), through meso-patterns (which objects serve most purposes?), to finally micropatterns (how much do purposes overlap for specific objects in specific phases?). Figure 1 illustrates this

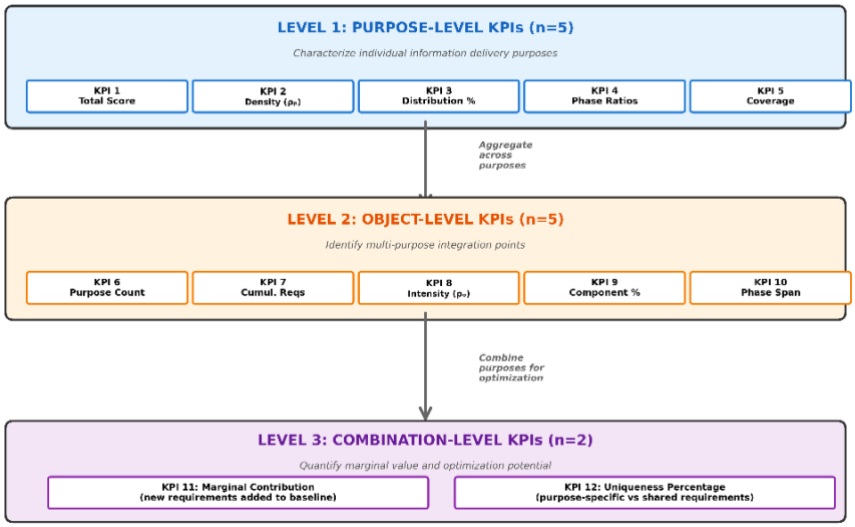


Figure 1. Hierarchical framework of LOIN Key Performance Indicators (KPIs).

hierarchical relationship and the specific LOIN KPIs proposed at each level.

Purpose-Level KPIs. For each purpose (p), five indicators characterizing its overall information requirements are defined.

Total Information Score: it quantifies information requirements burden.

$$S_p = \sum_{o \in O_p}(G_{p,o}w_G + A_{p,o}w_A + D_{p,o}w_D) \quad (1)$$

where (O_p) is the set of objects specified by purpose (p), and ($G_{p,o}$), ($A_{p,o}$), ($D_{p,o}$) represent geometrical, alphanumerical, and documentation requirement counts for object (o) in purpose (p). (w) are weighing coefficients given to each type of information purpose to model different management burden – external file references versus embedded geometrical or alphanumerical data. The default weighting adopted in this study assigns equal weights to geometrical and alphanumerical components and a reduced half weight to documentation. This choice reflects two considerations. First, ISO 7817-1 treats geometrical and alphanumerical information as structurally parallel LOIN components, both requiring embedded model data authored and maintained within the BIM environment, whereas documentation requirements reference external deliverable files with lower authoring overhead per specification. Second, in the absence of empirical evidence quantifying differential management burden across component types, equal weighting represents the most neutral baseline assumption, avoiding a priori bias toward either spatial or property-based specifications. The sensitivity of results to alternative weighting schemes is discussed in Section 5.

Information Density: this is the normalised score by object count, revealing the specification intensity per object.

$$\rho_p = S_p / |O_p| \quad (2)$$

High density purposes impose heavy specifications per object (intensive); low density purposes show broad but shallow coverage (extensive). This metric complements the Total Information Score as purposes can have high scores through either intensive specification of few objects or extensive specification of many objects.

Component Distribution Percentages: this KPI reveals information specialization for each purpose:

$$\%G_p = \frac{\sum_o G_{p,o}}{\sum_o (G_{p,o}+A_{p,o}+D_{p,o})} \quad (3)$$

with analogous definitions for ($\%A_p$) and ($\%D_p$). This KPI enables deterministic approaches to understand how burden is distributed across purposes, i.e. there may be geometrical-dominant purposes if ($\%G_p$)>0.5.

Phase Allocation Ratios: this KPI indicates temporal distribution of each purpose's information requirements across project phases.

$$R_{p,\phi} = \frac{|C_{p,\phi}|}{|C_p|} \quad (4)$$

where ($C_{p,\phi}$) is the set of requirements for purpose (p) in phase (ϕ), and (C_p) are the total requirements for purpose (p). This reveals whether purposes concentrate in design, construction, or operation phases.

Object Coverage: this KPI ($|O_p|$) simply counts the number of objects per purpose, indicating scope breadth versus narrow specialisation.

Object-Level KPIs. Purpose-level indicators aggregate requirements, but obscure which specific objects bear heaviest specification burdens. Object-level KPIs address this by aggregating across all purposes mentioning each object. For each object (o), it is defined:

Purpose Count: this KPI indicates multi-purpose utilization.

$$N_o = |\{p: o \in O_p\}| \quad (5)$$

Objects with a large (N_o) identify critical multi-purpose integration points requiring coordinated information management.

Cumulative Requirements: this KPI sums specifications across all purposes. This reveals absolute specification burden accumulated when multiple purposes converge on single object.

$$R_o = \sum_{p:o\in O_p}(G_{p,o} + A_{p,o} + D_{p,o}) \quad (6)$$

Information Intensity: this KPI provides a symmetric metric to purpose-level density, normalizing cumulative requirements by purpose count.

$$\rho_o = R_o / N_o \quad (7)$$

This reveals average specification burden per purpose using the object. High intensity objects are heavily specified by each purpose, whereas low intensity objects are lightly specified despite appearing in many purposes. This symmetry with (ρ_p) enables comparative analysis: purposes with high density specify many requirements per object; objects with high intensity accumulate many requirements per purpose.

Component Intensity: this KPI analyses cumulative composition of object information, revealing whether object specifications emphasise over geometry (coordination), properties (performance), or documentation (compliance).

$$I_o^G = \frac{\sum_p G_{p,o}}{\sum_p (G_{p,o} + A_{p,o} + D_{p,o})} \quad (8)$$

with analogous definitions for (I_o^A) and (I_o^D).

Phase Coverage: this KPI identifies object lifecycle span.

$$\Phi_o = \left|\{\phi: \exists p, o \in O_p\}\right| \quad (9)$$

distinguishing objects serving only single phases from those persisting through project lifecycle.

Combination-Level KPIs. The most granular level examines individual object-purpose-phase triplets (o, p, ϕ), enabling investigation of requirement overlap and consolidation potential.

Uniqueness: this KPI quantifies purpose-specific requirements versus shared requirements.

$$U_{o,p,\phi} = \frac{1}{|O_p| \times |\Phi_p|} \sum_{o\in O_p, \phi\in\Phi_p} \frac{\left|R_{o,p,\phi} \setminus \cup_{p' \neq p} R_{o,p',\phi}\right|}{|R_{o,p,\phi}|} \quad (10)$$

where ($R_{o,p,\phi}$) is the set of requirements for each triplet of information, averaged across all object-phase combinations where requirements exist. High uniqueness indicates genuine purpose-specific needs requiring dedicated specifications; low uniqueness suggests consolidation opportunities through shared baseline definitions.

Marginal Contribution: this KPI measures the incremental value of each information set when adding a new purpose to a baseline set.

$$M_{p'} = \frac{\left|R_{p'} \setminus \cup_{p\in P_{base}} R_p\right|}{|R_{p'}|} \quad (11)$$

where (R_p) represents all the requirements for purpose (p). This efficiency metric enables evidence-based purpose prioritisation during EIR specification. Purposes with high (M_p) contribute substantial new information, while those with low (M_p) largely duplicate selected purpose information.

Analytical Framework

The analysis employs a progressive validation framework advancing from descriptive pattern identification through multivariate confirmation, structured in five complementary phases. Descriptive analytics will first establish empirical patterns through univariate distributions and bivariate relationships, identifying three purpose archetypes based on component composition (geometrical, alphanumerical, and documentation dominant), universal objects spanning multiple purposes, and temporal requirement concentration patterns. These descriptive findings generate hypotheses requiring statistical validation: whether archetype classifications represent natural groupings versus arbitrary thresholds, whether the proposed 12-KPI framework provides complementary versus redundant measurements, and whether patterns generalize across analytical dimensions.

Multivariate validation is then employed through four methods testing different framework aspects. Correlation analysis (Pearson r across the 12×12 KPI matrix) assesses indicator complementarity, where low-moderate within-level correlations confirm coherence without redundancy. Cronbach's alpha validates internal consistency at each level, where low alpha combined with weak cross-level correlations confirms a multidimensional complementary design rather than a unidimensional redundant scale. Principal component analysis then reveals dataset dimensionality, reducing the 12 KPIs to combined orthogonal components and demonstrating archetype clustering in multivariate space, empirically validating natural groupings. Finally, temporal profile analysis extends validation to the lifecycle dimension, revealing archetype-specific phase concentration patterns that transcend component and object composition differences.

Methodological triangulation across five analytical phases provides convergent validation unavailable through single-method approaches. Patterns identified descriptively (archetype classification, universal objects, phase concentration) receive confirmation through correlation structure, dimensional reduction, and temporal differentiation. This progressive framework—descriptive hypothesis generation followed by multivariate statistical validation—strengthens confidence that findings represent genuine data structure rather than analytical artifacts, supporting framework generalizability beyond the BIMids case study.

Reflecting on the research questions established, RQ1 (requirement variation) is addressed through purpose-level KPIs via descriptive analytics and temporal profiling. RQ2 (component distribution) is addressed through descriptive analytics and PCA archetype validation. RQ3 (object specification burden) is addressed through object-level KPIs via descriptive analytics and correlation analysis. RQ4 (multi-purpose object coverage) is addressed through correlation analysis and PCA. RQ5 (purpose-specific uniqueness) is addressed through combination-level KPIs via descriptive analytics and dashboard implementation.

Results

Descriptive Analysis

This section addresses RQ1–RQ4 through descriptive characterisation of purpose archetypes, object-level burden, and lifecycle patterns. Analysis of LOIN component distribution across 22 purposes reveals three distinct purpose archetypes (see Figure 2). Geometrical-dominant purposes (n=15, 68%) allocate majority requirements to spatial specifications: '2D documents' (100% geometrical, 374 requirements total) and 'Clash detection' (96.5% geometrical, 342 requirements) exemplify purposes where 3D spatial representation is paramount. Alphanumerical-dominant purposes (n=5, 23%) prioritize property data: 'Maintenance' (57.8% alphanumerical, 552 total requirements) focuses on operational parameters including schedules, warranties, and performance thresholds. Documentation-dominant purposes (n=1, 5%) emphasize deliverable files: 'Minimum requirements' (59.4% documentation, 1,130 total) establishes comprehensive baseline spanning all components but emphasizing document deliverables.

Total information requirements vary substantially: 'Minimum requirements' (1,130) and 'Maintenance' (552) impose heaviest specification loads, while specialized purposes like 'Programming' (5) and 'Access management' (43) target narrow scopes with low demand.

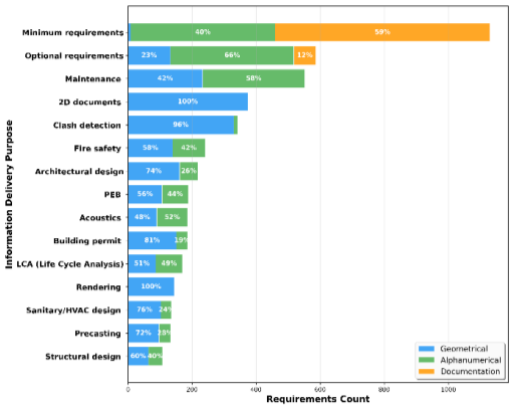


Figure 2. Top 15 component distribution by purpose.

Object-level aggregation identifies universal building elements supporting multiple purposes (see Figure 3). Core architectural objects accumulate highest cumulative requirements: door (172 requirements across 16 purposes), window (148, 16 purposes), ramp (133, 15 purposes), space (125, 17 purposes), slab (122, 14 purposes). These objects function as integration points because they serve multiple purpose archetypes: geometrical-dominant purposes require spatial properties for coordination, alphanumerical-dominant purposes specify performance attributes, and documentation-dominant purposes demand compliance deliverables.

Space achieves highest purpose count (17) spanning spatial programming through facility management; door and window support 16 purposes each. In contrast, specialized MEP equipment serves 1-4 purposes only. Analysis reveals 29 objects (36%) support 10+ purposes each, identifying critical targets for standardization and information quality management.

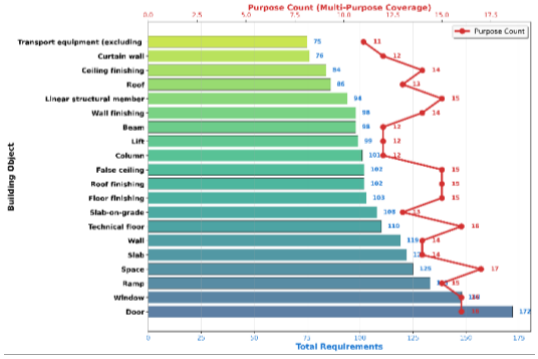


Figure 3. Top 20 object analysis by total requirements and multi-purpose coverage.

Requirement distribution across phases demonstrates progressive intensification (see Figure 4). Phase 1 specifies 660 requirements, escalating to 2486 in Phase 2 (+277%), 3784 in Phase 3 (+52%), peaking at 4750 in Phase 4 (+26%). Phase 5 surprisingly denotes fewer requirements at 3485 (-27%) for operation and use. Note that there is quite a balanced G-A-D distribution across project phases.

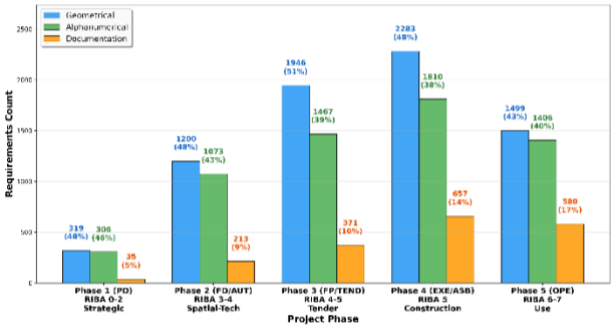


Figure 4. Lifecycle requirement intensification by component.

Correlation Analysis

This section validates KPI complementarity underpinning all five research questions, with relevance to RQ3 and RQ4 (object-level structural relationships). Correlation analysis across hierarchical levels reveals predominantly weak to moderate inter-correlations, supporting complementary nature of the proposed indicators (see Figure 5). However, some can be noted: 1) there is a significant positive correlation ($r > 0.9$) between (N_o) and (R_o) and (ρ_o) and (Φ_o) denoting structural relationships between KPIs (as expected based on their equations); 2) there is a moderate negative correlation ($r = -0.399$) between total information score and uniqueness demonstrating the cumulative nature of information requirements and duplicities existing in practice; and 3) there is a significant positive correlation ($r > 0.9$) also between ($R_{p,\phi}$) and (ρ_o), which may indicate a structural pattern in information requirement distribution across project lifecycle.

When looking at the cross-independence of the KPIs (purpose level vs. object level), there is a minimal

correlation between them (median $|r| = 0.22$). This confirms that the bi-level KPI definition captures different dimensions of LOIN integration and validates the hierarchical approach.

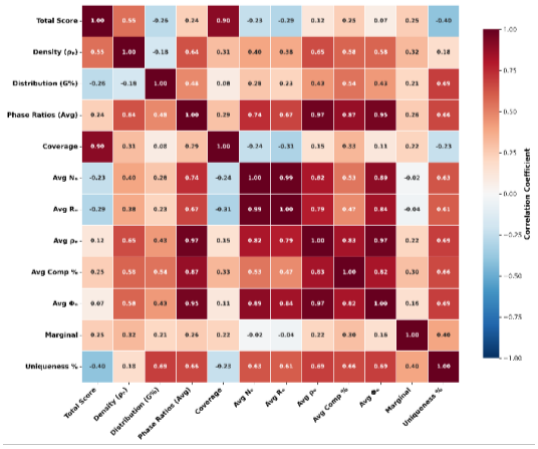


Figure 5. Correlation matrix for the 12 proposed LOIN KPIs.

The Cronbach's alpha results show internal consistency for purpose-level KPIs as they demonstrate low inter-item correlation (α=0.19), confirming multi-dimensional measurement where each indicator captures distinct information aspects. Object-level KPIs show moderate consistency (α=0.56), reflecting structural relationships (e.g., multi-purpose objects accumulate requirements) while maintaining conceptual distinctiveness. Low within-level alpha combined with weak cross-level correlations (median |r|=0.22) validates the framework's multi-dimensional design: indicators provide complementary rather than redundant information, justifying the 12-KPI structure.

Principal Component Analysis

Principal component analysis on all 12 KPIs reveals concentrated variance structure with first three components explaining 87.4% (PC1: 52.3%, PC2: 22.7%, PC3: 12.4%), demonstrating efficient framework dimensionality. PC1 presents existing object-level lifecycle complexity ($\overline{\Phi_o}$: 0.392, $\overline{\rho_o}$: 0.391), separating purposes with sustained cross-phase object engagement from phase-concentrated specifications. PC2 captures purpose scope (Total Score: 0.595, Coverage: 0.540), distinguishing comprehensive from specialized purposes. PC3 emphasizes component archetype (Geom %: 0.599), with purposes clustering in multivariate space according to archetype classifications (see Figure 6). This emergent clustering validates the three-archetype framework: purposes classified by component thresholds demonstrate multivariate similarity within groups and dissimilarity across groups, confirming natural rather than arbitrary categorization. High PC1 variance (52.3%) suggests object-level temporal patterns dominate purpose variation, informing optimization priorities. Framework efficiency—87.4% variance captured by three dimensions from 12 indicators—supports measurement adequacy without redundancy.

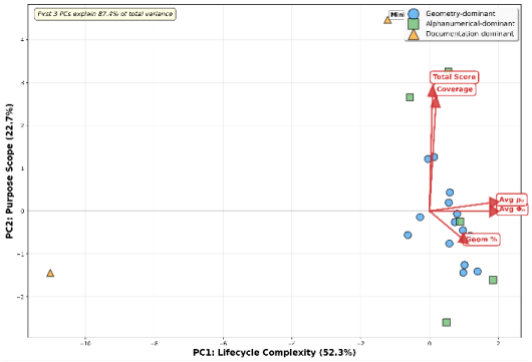


Figure 6. Principal component analysis biplot: purpose clustering and LOIN KPIs loadings.

Temporal Profile Analysis

Temporal profiling completes the validation of RQ1 by demonstrating archetype-specific lifecycle signatures. The temporal requirement profiling reveals that each archetype exhibits a distinct project lifecycle signature (see Figure 7). Geometrical-dominant purposes concentrate in Phase 3 (mean: 32.4%) and decline sharply in Phase 5 (mean: 12.6%). This reflects the spatial coordination emphasis during detailed design that diminishes after the construction stage. In contrast, alphanumerical-dominant purposes distribute more uniformly across phases (concentration coefficient: 0.33) and sustain strong operational presence (Phase 5: 22.5%). This reflects performance monitoring needs throughout the building lifecycle. Documentation-dominant purposes show gradual accumulation without dramatic peaks. Figure 7 left panel compares mean archetype profiles (solid lines) against the overall average (dotted lines). As seen, geometrical purposes peak earlier than average while alphanumerical purposes sustain beyond average into operation. Its right panel shows normalized stacked bars for all 21 individual purposes, revealing systematic clustering by archetype despite variation within groups. Purposes with similar component compositions exhibit similar temporal patterns, independent of their total burden. This temporal differentiation validates archetype classification through an independent dimension beyond component percentages, demonstrating that geometrical-dominant, alphanumerical-dominant, and documentation-dominant purposes differ fundamentally by which project phase requirements concentrate, not merely in what components dominate.

Discussion

The three-archetype classification, validated through PCA and temporal profiling, reveals fundamental differences in how information purposes operate across

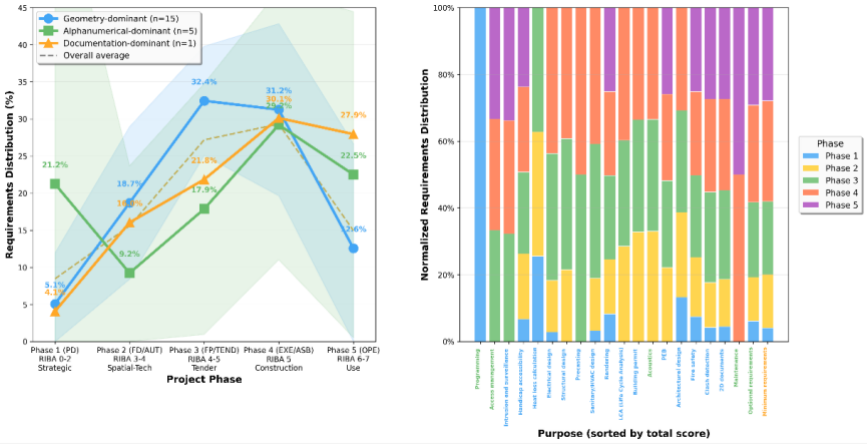


Figure 7. Temporal profiling analysis. Left: average temporal profile of the requirements distribution by project phase. Right: phase distribution per purpose.

building lifecycles. Geometrical-dominant purposes peak during detailed design coordination then declines post-construction, whereas alphanumerical-dominant purposes sustain operational relevance, reflecting performance monitoring needs. This suggests purpose selection should align with project delivery emphasis, consistent with established procurement patterns: design-build contracts typically emphasise early geometry coordination for clash detection, while long-term arrangements demonstrate value in lifecycle information exchange and operational performance monitoring (Ozcan-Deniz & Lokhandwala, 2023), indicating differential emphasis rather than exclusive requirements across procurement routes.

The identification of universal objects supporting 16-17 purposes confirms theoretical expectations from ISO 7817-1 about multi-purpose convergence on core architectural elements. However, the cumulative burden (e.g., door: 172 requirements) raises practical concerns about specification overload and duplicities. High uniqueness scores indicate however limited consolidation potential; available purposes genuinely require distinct information despite object overlap. This validates ISO 7817-1's purpose-driven approach but challenges assumptions about baseline specifications serving multiple purposes efficiently.

The interactive dashboard operationalises findings by quantifying marginal contribution and uniqueness for purpose combinations, letting practitioners identify high-value additions versus redundant purposes. For example, adding 'Maintenance' to geometrical-dominant baselines provides substantial alphanumerical information (57.8%) with moderate uniqueness, whereas adding 'Clash detection' to existing coordination purposes yields diminishing returns.

To illustrate, consider a practitioner preparing an EIR for a design-build project. They select 'Clash detection' and 'Energy analysis' as initial purposes, covering 58 and 41 objects respectively with predominantly geometrical requirements, giving cumulative coverage of 63 unique objects and 684 total requirements. The marginal contribution panel then ranks remaining purposes: 'Maintenance' contributes 552 new requirements (57.8% alphanumerical) with approximately 65% uniqueness, indicating substantial non-overlapping content. In contrast, 'Quantity take-off' contributes only 38% new requirements, largely duplicating geometrical specifications already covered by 'Clash detection'. The practitioner can therefore justify including 'Maintenance' as a high-value addition while deprioritising 'Quantity take-off', reducing overhead without sacrificing coverage.

The 12-KPI framework enables evidence-based EIR optimization, moving beyond institutional practice toward data-driven specification. Purpose-level indicators identify comprehensive versus specialized approaches; object-level metrics highlight multi-purpose integration points requiring quality management; combination-level metrics quantify consolidation feasibility. This hierarchical structure supports decisions at strategic (which purposes?), tactical (which objects?), and operational (which requirements?) levels.

The weighting assumption in Total Information Score treats geometrical and alphanumerical equally ($w_G = w_A = 1$) while assigning reduced weight to documentation ($w_D = 0.5$), potentially oversimplifying data acquisition and management differences between embedded model data and external file references. Preliminary sensitivity analysis varying the weighting ratio w_G:w_A:w_D across three alternative schemes, be that (2:1:0.5), (1:2:0.5), and (1:1:1), indicates that archetype classification remains stable: all 22 purposes retain their original archetype assignment, and the top-10 purpose ranking by Total Information Score changes by at most two positions. This suggests moderate robustness to weighting assumptions for the primary analytical findings. Nevertheless, further weighting scenarios (e.g., $w_D = 2$, reflecting high document management overhead in regulated projects) could alter downstream interpretations, and systematic sensitivity analysis across a wider parameter space, including weighting based on project role or stakeholder positioning, remains warranted.

The empirical findings reflect BIMids specification patterns and may not generalise directly to bespoke employer specifications, non-European regulatory contexts, or different infrastructure typologies (bridges, tunnels, process plants). Nevertheless, BIMids represents

the largest publicly available multi-purpose LOIN specification dataset with expert-consensus development, providing the strongest available empirical basis for cross-purpose analysis. The 12-KPI framework itself is source-agnostic: its formulations depend only on object-purpose-phase triplets with geometrical, alphanumerical, and documentation components, and can be applied to any dataset structured according to ISO 7817-1. Future research should validate archetype classification and universal object patterns through comparative crosswalks against national LOIN implementations (e.g., Dutch, Finnish, or Austrian frameworks) and extend analysis to the remaining LOIN aspects in ISO 7817-1 (detail, dimensionality, parametric behaviour) not covered here.

Conclusions

This study systematically quantified information requirement patterns across object-purpose-phase dimensions using 5,048 specifications from 22 standardized BIM purposes. The hierarchical 12-KPI framework operating at purpose, object, and combination levels revealed three distinct archetypes validated through multivariate analysis: geometrical-dominant purposes (68%) concentrating in design coordination, alphanumerical-dominant purposes (23%) sustaining through operations, and documentation-dominant purposes (5%) establishing baseline deliverables. Universal objects accumulate heaviest specification burden across 16-17 purposes, identifying critical multi-purpose integration points.

Principal component analysis confirmed framework dimensionality (87.4% variance in three components) and archetype clustering, demonstrating natural rather than arbitrary categorization. Temporal profiling validated archetypes through independent lifecycle signatures, transcending component composition differences. Correlation analysis revealed complementary indicator structure (median cross-level r=0.22), supporting measurement adequacy without redundancy.

The interactive dashboard operationalizes findings for practitioners, enabling evidence-based EIR optimization through marginal contribution and uniqueness quantification. High uniqueness scores across purposes challenge consolidation assumptions, validating ISO 7817-1's purpose-specific approach while highlighting ISO 19650 specification overhead implications.

This granular analysis advances BIM research from aggregate requirement studies to object-level multi-dimensional characterization, providing empirical foundation for standardization efforts and practical decision support for project information management. Future research should validate findings across construction typologies, test alternative KPI weighting schemes, and extend temporal analysis beyond RIBA phase structures to international delivery models.

References

Abanda, F. H., Balu, B., Adukpo, S. E., & Akintola, A. (2025). Decoding ISO 19650 Through Process Modelling for Information Management and Stakeholder Communication in BIM. Buildings, 15(3), 431. DOI: 10.3390/buildings15030431.

Cavka, H. B., Staub-French, S., & Poirier, E. A. (2017). Developing owner information requirements for BIM-enabled project delivery and asset management. Automation in Construction, 83, 169-183. DOI: 10.1016/j.autcon.2017.08.006.

Krijnen, T., & Tamke, M. (2015). Assessing implicit knowledge in BIM models with machine learning. In Modelling Behaviour: Design Modelling Symposium 2015 (pp. 397-406). DOI: 10.1007/978-3-319-24208-8_33.

Lindholm, J., Jansson, T., Johansson, P., & Yitmen, I. (2025). Integrating LOIN-driven information exchange connected to construction production planning. In Proceedings of the 2025 EC3 Conference. DOI: 10.35490/EC3.2025.304.

Liu, L., Hagedorn, P., & König, M. (2023). Definition of a container-based machine-readable IDM integrating level of information needs. Proceedings of the 2023 EC3 Conference and the 40th International CIB W78 Conference.
DOI: https://doi.org/10.35490/ec3.2023.221.

Malla, V., Tummalapudi, M., & Delhi, V. S. K. (2024). Perceptions of built-environment professionals on using ISO 19650 standards for information management. Journal of Legal Affairs and Dispute Resolution in Engineering and Construction, 16(1), 04523045. DOI: 10.1061/JLADAH.LADR-1076.

Mitera-Kiełbasa, E., & Zima, K. (2024). BIM Policy Trends in Europe: Insights from a Multi-Stage Analysis. Applied Sciences, 14(11), 4363. DOI: 10.3390/app14114363.

Noardo, F., Harrie, L., Arroyo Ohori, K., Biljecki, F., Ellul, C., Krijnen, T., ... & Stoter, J. (2020). Tools for BIM-GIS integration (IFC georeferencing and conversions): Results from the GeoBIM benchmark 2019. ISPRS International Journal of Geo-information, 9(9), 502. DOI: 10.3390/ijgi9090502.

Ozcan-Deniz, G., & Lokhandwala, Z. (2023). Building Information Modeling (BIM) Implementation in Public–Private Partnership (PPP) Projects. Engineering Proceedings, 53(1), 40. DOI: 10.3390/IOCBD2023-15189.

Patacas, J.; Dawood, N.; Kassem, M. BIM for facilities management: A framework and a common data environment using open standards. Automation in Construction, 2020, vol. 120, p. 103366. DOI: 10.1016/j.autcon.2020.103366.

Rathnasinghe, A. P., Kulatunga, U., Jayasena, H. S., & Wijewickrama, M. K. C. S. (2022). Information flows in a BIM enabled construction project: Developing an information flow model. Intelligent Buildings International, 14(2), 190-206. DOI: 10.1080/17508975.2020.1848783.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

FROM ANALYSIS TO DESIGN: TEACHING ALGORITHM-AIDED DESIGN IN STRUCTURAL ENGINEERING

Marcin Luczkowski[1], Anders Rønnquist[2]
[1]NTNU, Norwegian University of Science and Technology, Trondheim, Norway

Abstract

The increasing integration of computational tools and automated workflows in structural engineering practice calls for corresponding developments in engineering education. This paper presents TKT4198, a Master's-level course designed to introduce algorithm-aided design through a project-based, design-oriented format. Drawing partial inspiration from architectural design studios, the course enables structural engineering students to design structural systems from scratch while progressively integrating parametric modelling, finite element reasoning, and basic algorithmic methods. The paper describes the course structure, key assignments, and pedagogical rationale, and reflects on survey-based observations regarding student perceptions.

Introduction

The Architecture, Engineering, and Construction (AEC) sector is undergoing a rapid transformation driven by the increasing use of computational tools, data-rich models, and automated design workflows. Parametric modelling, algorithm-aided design (AAD), and digital optimisation techniques are now routinely employed in practice to support early-stage design exploration, performance assessment, and interdisciplinary collaboration. Structural engineers are therefore increasingly expected not only to analyse predefined systems, but also to actively engage with design logic, digital models, and evolving computational methods.

Despite these developments, structural engineering education remains largely centred on deterministic analysis methods and established calculation procedures. While these foundations are essential, they often provide limited exposure to computational thinking, parametric reasoning, and exploratory design workflows. As a result, students may graduate with strong analytical skills but limited experience working on design problems where geometry, structural behaviour, and digital tools are tightly coupled and continuously evolving.

In parallel, computational design education has historically been more strongly embedded within architectural curricula, where parametric modelling and visual programming have been used to explore form, spatial organisation, and fabrication logic. Only more recently have civil and structural engineering programmes begun to incorporate algorithmic design approaches, often in the form of advanced electives or software-oriented courses. There remains a need for accessible educational formats that introduce algorithm-aided design as a conceptual engineering skill, rather than as a specialised or tool-dependent competence.

This paper presents TKT4198, a Master's-level course developed at the Norwegian University of Science and Technology that addresses this gap through a project-based, design-oriented approach. Drawing partial inspiration from architectural design studios, the course is structured to allow structural engineering students to design structural systems from scratch while progressively integrating parametric modelling, finite element reasoning, and basic algorithmic methods. Rather than separating design and analysis, the course treats computation as an integral part of structural reasoning and decision-making.

The course emphasises gradual skill development through weekly assignments, live problem solving, and open-ended design tasks. Computational tools are introduced not as isolated software environments, but as instruments for exploring structural behaviour, testing design assumptions, and developing methodological understanding. By doing so, the course aims to prepare students for real-world engineering problems characterised by increasing automation, digital integration, and uncertainty.

The objective of this paper is to describe the pedagogical structure, content, and educational rationale of TKT4198, and to situate it within the broader context of computational design education for structural engineers. By presenting the course as a transferable educational framework rather than a software-driven solution, the paper contributes to ongoing discussions on how algorithm-aided design can be meaningfully integrated into contemporary structural engineering curricula.

In addition, the paper seeks to examine how the integration of parametric modelling, finite element reasoning, and introductory algorithm implementation within a single course framework influences students' perceived understanding of computational design concepts and their ability to engage with design-oriented structural problems. To support this, the study reflects on longitudinal survey data and course observations, addressing the following questions: (1) how students' awareness of parametric and algorithm-aided design

evolves during the course, and (2) to what extent the proposed teaching framework supports the integration of design and analysis in structural engineering education.

Background

Development of FEM Software and Its Integration with CAD and BIM Environments

The development of finite element method (FEM) software has evolved from standalone numerical analysis tools toward increasingly integrated components of digital design workflows. Early FEM applications were primarily verification-oriented, relying on simplified geometric input and manual data transfer, which reinforced a linear separation between design and analysis. Within the broader discourse on computational design, this separation has been critically examined, with authors arguing that designers and engineers should actively construct and adapt their own computational tools and workflows rather than rely solely on predefined software environments (Aish, 2014). This perspective has become increasingly relevant as structural design processes move toward automation, parametric control, and algorithm-aided workflows.

With the widespread adoption of computer-aided design (CAD) systems, FEM software began to accept direct geometric input, reducing the need for model reconstruction and enabling closer alignment between design intent and analytical representation. This shift was further accelerated by the introduction of Building Information Modelling (BIM), which promoted data-rich models combining geometry, material properties, and structural information. BIM platforms such as Autodesk Revit allow the creation of analytical models directly from geometric descriptions, supporting early structural reasoning within the design environment, although such models often require refinement for detailed analysis.

Interoperability between design and analysis tools has been further supported through standardised data exchange formats, most notably IFC. Structural analysis software such as FEM-Design can use IFC-based input, enabling FEM models to be derived from BIM environments while maintaining consistency in geometry and material definitions. In parallel, several FEM platforms have developed stronger links to parametric modelling environments, reflecting the growing importance of algorithm-aided workflows.

Over the last decade, parametric and algorithmic modelling tools have played a key role in automating and integrating structural design processes. Plugins and interfaces connecting FEM software with Grasshopper, such as those developed for FEM-Design and SOFiSTiK, allow procedural generation of analysis models and support iterative design exploration. At the same time, tools such as Karamba3D embed FEM functionality directly within the parametric modelling environment, enabling geometry generation, analysis, and feedback to occur within a single computational framework.

These developments mark a significant shift toward automated and iterative design processes, where FEM is no longer treated solely as a final verification step. Instead, analysis increasingly becomes an active component of design exploration, highlighting the importance of computational workflows that integrate modelling, analysis, and algorithmic control—both in professional practice and in engineering education.

Availability of Parametric Modelling and Algorithm-Aided Design Courses in Higher Education

For many years, parametric modelling and algorithm-aided design (AAD) were taught predominantly within the domain of architectural education. Early courses focused on geometric exploration, form generation, and visual programming, often prioritising conceptual and aesthetic outcomes over structural performance or numerical analysis. Visual programming environments enabled designers to explore complex geometries without extensive programming knowledge, reinforcing the adoption of parametric workflows primarily within architecture and design disciplines.

Over the past decade, this situation has begun to change. The increasing integration of computational tools into engineering practice has motivated civil and structural engineering programmes to introduce parametric modelling and algorithmic workflows into their curricula. These courses often aim to bridge design generation with structural reasoning, though the depth of integration between geometry, analysis, and computation varies significantly across institutions.

Several architecture-led courses incorporate structural performance as part of computational design education. For example, ARCH9110 Code to Production at The University of Sydney integrates parametric modelling with performance-driven design and fabrication workflows, including references to structural behaviour, while remaining primarily rooted in architectural education (The University of Sydney, 2024). Similarly, teaching and research activities at ETH Zurich have long addressed computational structural design through graphic statics and form-finding methods, typically embedded within architectural design studios rather than formal civil engineering courses (ETH Zurich, 2021).

In parallel, several engineering-oriented courses have emerged that explicitly connect parametric modelling with structural analysis. The course Computational Structural Design at the Technical University of Munich combines parametric geometry generation with structural evaluation methods, introducing students to form-finding techniques and finite element analysis within an integrated workflow (Technical University of Munich, 2024). At TU Darmstadt, courses on spatial structures introduce parametric design tools for generating and analysing structural systems, though they tend to focus on tool-based workflows rather than on student-implemented numerical methods (TU Darmstadt, 2024).

At the graduate level, some hybrid courses explicitly address computational optimisation and numerical reasoning. At the Massachusetts Institute of Technology, Computational Structural Design and Optimization integrates finite element methods with computational

design and optimisation strategies, but it assumes strong prior knowledge in structural mechanics and numerical methods and does not focus on introductory programming pedagogy (Massachusetts Institute of Technology, 2024). A comparable approach is found at Carnegie Mellon University, where advanced elective courses explore computational structural design and form-finding, typically targeting students with existing coding and analysis experience (Carnegie Mellon University, 2024).

Other offerings, such as CEE 220C Parametric Design and Optimization at Stanford University, introduce parametric and optimisation-based approaches within civil engineering, with a stronger emphasis on optimisation concepts than on early-stage geometric design or explicit finite element modelling (Stanford University, 2024). Similarly, modelling-oriented courses at Tampere University introduce algorithm-aided workflows within structural engineering contexts, though coding and optimisation are often addressed through predefined software environments rather than through student-implemented algorithms (Tampere University, 2022).

Overall, these examples illustrate a clear expansion of parametric modelling and AAD beyond architectural education into civil and structural engineering programmes. However, most existing courses focus either on parametric geometry generation without deep structural integration, or on advanced computational analysis and optimisation without addressing algorithmic transparency. Few courses systematically combine parametric modelling, finite element reasoning, and introductory algorithm implementation within a single, accessible course framework. This gap highlights the need for pedagogical approaches that treat algorithm-aided design as a conceptual engineering skill rather than as a specialised or software-dependent technique.

Course Context and Objectives

TKT4198 is a 7.5 ECTS course offered at the fifth-year (Master's) level at the Norwegian University of Science and Technology (NTNU, 2025). The course is primarily targeted at structural engineering students but is also open to students from other engineering disciplines, supporting interdisciplinary participation. It introduces algorithm-aided design and computational workflows as a complement to traditional analysis-oriented courses, with focusing on digital modelling, conceptual structural design, and the integration of geometry and analysis within contemporary engineering practice. The course is grounded in a project-based learning approach, where open-ended tasks are used to support the development of engineering judgement in computational design contexts.

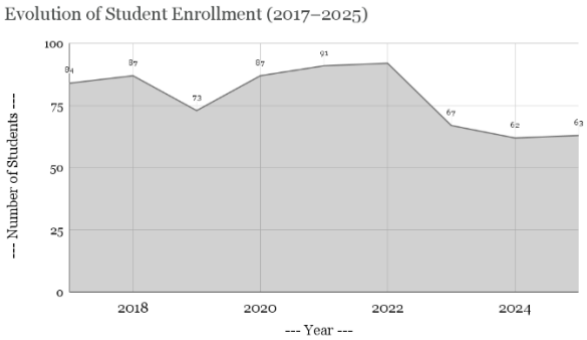


Figure 1: Number of students in the following years

As illustrated in Figure 1, the course has shown consistently high enrolment in recent years. The number of actively participating students in the course has stabilised at around 60 students per iteration, where "active" participation refers to students who regularly submit weekly assignments. In addition, the course typically attracts an additional 10–20 passive participants, who attend lectures and follow the course material without submitting all assignments. Approximately 20% of enrolled students are exchange students, contributing to a diverse cohort with varied academic and cultural backgrounds. This sustained level of participation indicates that computational and algorithm-aided design topics address a clear demand within the student body.

The course assumes prior knowledge consistent with a Master's-level engineering curriculum, including structural mechanics, basic structural analysis, and material behaviour. No prior experience with parametric modelling, visual programming, or programming languages is required. This allows computational methods to be introduced as conceptual design and reasoning tools rather than as advanced software or programming specialisations.

The primary objective of the course is to make students comfortable operating within digital environments commonly used in structural engineering. This includes the creation and interpretation of geometric and finite element (FE) models, as well as the critical assessment of modelling assumptions and analysis results. The course is organised into two main phases: an initial phase focusing on geometric modelling and finite element analysis, followed by a phase addressing conceptual structural design through algorithmic methods, including visual programming and introductory Python scripting. Throughout the course, structural analysis is treated as an integrated part of the design process rather than a post-design verification step, encouraging reflective and critical use of computational tools.

Course Assignments and Learning Progression

The course is organised as a 14-week sequence of weekly assignments, supported by five hours of scheduled teaching per week. In contrast to conventional course formats that separate lecturing and exercises (typically

two hours of lectures and three hours of supervised work), TKT4198 deliberately avoids a strict distinction between lectures and exercises. Each teaching session combines theoretical explanation with live problem solving, where the instructor develops models, analyses, and algorithms in real time while explaining the underlying concepts and design rationale.

This blended format allows students to engage with both theory and application simultaneously. Some students actively follow and reproduce the steps during class, while others observe and revisit the material later. To support different learning paces and backgrounds, all sessions are recorded, and recordings of both teaching and selected assignment solutions are made available through an online video platform. This ensures that students can revisit complex topics and reinforces continuity between weekly assignments and in-class activities.

Assignments constitute the core teaching mechanism of the course and are used to progressively introduce parametric modelling, finite element analysis, and algorithmic thinking. Students must successfully complete at least 80% of the weekly assignments to be eligible for the final examination. Rather than providing introductory software tutorials, the course adopts a learning-by-doing approach, where software proficiency emerges through increasingly complex modelling tasks.

Early assignments focus on fundamental structural modelling concepts, such as beam-based finite element models for investigating different truss topologies. Students are required to compare alternatives and justify structural choices through concise reports. As the semester progresses, assignments introduce parametric modelling, scripting, form-finding, buckling analysis, and optimisation tasks, with increasing freedom for students to propose their own solutions.

The course concludes with a four-hour, on-campus digital examination at NTNU facilities. All students work on identical computer setups with the same software environment. The exam consists of four open-ended design and analysis tasks, closely aligned with the assignments, and serves as a final assessment of the students' ability to integrate modelling, analysis, and algorithmic reasoning within a limited timeframe.

In the following sections, two representative assignments from the course are presented in more detail, as they illustrate the central pedagogical principles and methodological approach of TKT4198.

Parametric Auditorium Assignment

The parametric auditorium assignment is introduced after the third week of the semester, at a point when students have been exposed to basic parametric modelling concepts but have not yet engaged with finite element analysis or structural calculations. The aim of the assignment is to introduce algorithm-aided design through structural reasoning rather than numerical verification, reflecting early-stage design conditions where decisions are often made before detailed analysis is available.

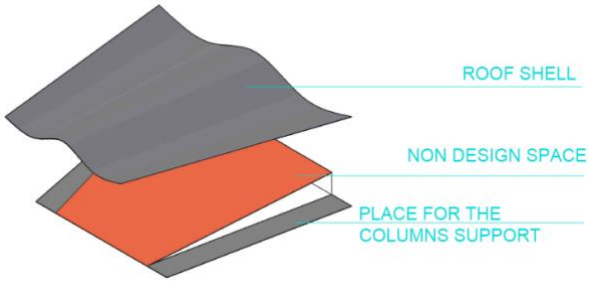


Figure 2: Graphical explanation of assignment task

The task requires students to propose a supporting structural system for an outdoor auditorium roof under clearly defined architectural constraints. As illustrated in Figure 2, the roof geometry is provided as a reference surface that must not be intersected by the structure. A non-design space is defined above the audience area, restricting the placement of structural elements in the central functional zone. Column supports are allowed only in designated areas along the sides of the auditorium. These constraints force students to explicitly plan the load-bearing system, considering issues such as spatial clearance, functional requirements, and constructability, without relying on structural calculations.

Students are required to generate the supporting structure using a fully parametric model, typically exploring systems such as grid shells, frames, or truss-based solutions. Key parameters—including grid spacing, geometric subdivision, and overall dimensions—must be controllable within the model. To demonstrate that the model is genuinely parametric, students are required to present at least three distinct design variants derived from the same algorithmic definition, rather than three separately modelled solutions.

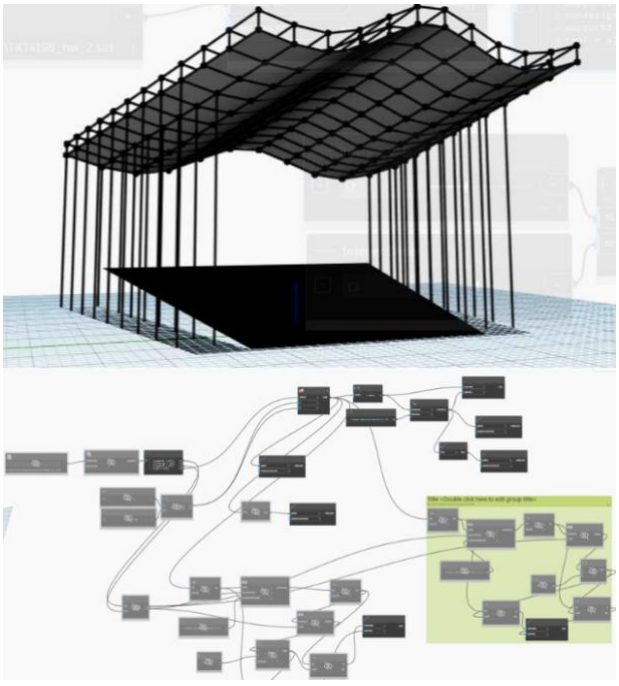


Figure 3: One of the possible solutions sent by student

An example of a student-developed solution is shown in Figure 3, where a grid-based roof structure is generated algorithmically and supported by columns placed outside the non-design space. The figure also illustrates the underlying parametric definition, highlighting how geometric logic, rather than manual modelling commands, governs the structural layout. At this stage, students are not asked to verify structural performance through FEM; instead, the focus is on structural discussion, including the plausibility of the load paths, the efficiency of the support strategy, and the adaptability of the system to changing design requirements.

The assignment is intentionally formulated to discourage the use of ad-hoc or purely geometric modelling commands. Instead, students are encouraged to approach parametric modelling from a methodological perspective, moving from general rules to detailed geometry. By observing how changes in parameters propagate through the model, students experience the practical value of parametric modelling as a design and analysis tool. This approach aims to motivate experimentation with algorithm-aided methods and to establish parametric thinking as a meaningful component of structural design reasoning, rather than as a purely formal or software-driven exercise.

Truss Optimisation with Genetic Algorithms

The truss optimisation assignment is introduced at the end of the semester, after approximately twelve weeks of continuous work with parametric modelling, finite element analysis, and scripting. At this stage, students have already performed FEM analyses of buildings, spatial structures, and bridge-like systems, and have implemented Python code within parametric environments to generate and manipulate structural components. The assignment is scheduled immediately after a dedicated lecture introducing the principles of simple genetic algorithms, including population-based search, mutation, selection, and convergence.

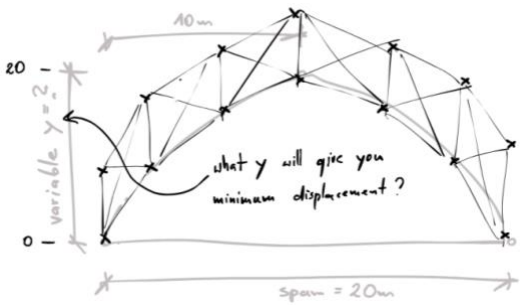


Figure 4: Graphical explanation of the assignment task

The task requires students to integrate parametric geometry generation, finite element modelling, and algorithmic optimisation within a self-written computational workflow. As illustrated in Figure 4, the optimisation problem is formulated using a simple, arc-like truss structure spanning a fixed distance. Selected geometric parameters—most notably the vertical position of nodes—serve as design variables. The objective is defined as a single-objective optimisation, typically minimising structural displacement under a prescribed load case.

Students are explicitly required to implement the genetic algorithm from scratch in Python, without relying on optimisation plugins or external libraries. The optimisation loop integrates three components: parametric geometry creation, linear elastic FEM analysis of a two-dimensional truss system, and fitness evaluation based on structural response. This requirement is intentional and reflects the course's emphasis on algorithmic transparency rather than software proficiency. Students are encouraged to implement and test their code in different computational environments, reinforcing the portability of algorithmic thinking beyond a single platform.

An example student solution is shown in Figure 5, including the resulting optimised geometry and the corresponding structural deformation. Quantitative results, such as convergence of displacement values across generations, are reported alongside visual feedback, allowing students to relate numerical outcomes to geometric changes. At this stage, the goal is not to obtain an optimal engineering solution, but to understand how design decisions propagate through a coupled geometry–analysis–optimisation pipeline.

The simplicity of the optimisation problem is a deliberate pedagogical choice. Based on several years of teaching experience, it has been observed that introducing highly complex numerical models or large variable sets often leads students to reproduce template solutions rather than actively engage with the algorithm. By limiting the number of variables and focusing on a transparent optimisation structure, the assignment encourages experimentation, modification, and critical reflection. In this way, students are exposed to optimisation not as a black-box tool, but as a design method that can be adapted, questioned, and extended—an approach that aligns with the broader aim of positioning algorithm-aided design as an active component of structural engineering practice.

Generation	$y\ [m]$	Displacement $[mm]$
1	7,38	3,03
10	9,40	2,93
20	9,40	2,93

I.e. the overall best arch height is $y = 9{,}4\ m$, yielding a maximum displacement of 2,93 mm.

In Figure 3 below, the green truss arch represents the deformed geometry, scaled with a factor of 200.

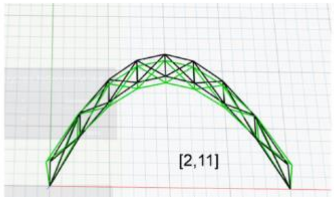


Figure 5: One of the possible solutions send by student

Pedagogical Reflections

This study is based on a longitudinal observation of the course since its introduction in 2017, combining iterative

course development with repeated survey-based data collection. Beginning in 2019, voluntary surveys have been conducted before and after each semester to capture students' perceptions, expectations, and self-assessed competencies related to algorithm-aided design and parametric modelling. Participation in the surveys has consistently exceeded 60% of enrolled students, and all responses are collected anonymously. The surveys include a consistent set of questions repeated annually, combining general course feedback with targeted items addressing parametric modelling awareness, prior exposure to algorithm-aided design, and intended future use of computational tools. Most responses are recorded using a five-point self-assessment scale, complemented by selected categorical questions (e.g. yes/no or multiple choice). The survey instrument consisted of a limited set of recurring items administered before and after the course, focusing on students' background, initial digital literacy, and expectations, as well as their perceived progression in parametric modelling, design skills, and awareness of computational tools. The key survey items used in the analysis are summarised in Table 1.

Table 1. Summary of key pre- and post-course survey items used in the analysis

Stage	Core item	Purpose
Pre	Self-assessed parametric modelling level	Establish baseline
Pre	Self-assessed FEM modelling level	Establish baseline
Pre	BIM/IFC awareness	Assess digital ecosystem familiarity
Pre	Motivation for taking the course	Identify expectations
Post	Self-assessed parametric modelling level	Measure perceived progression
Post	Awareness of new digital technologies	Measure broader learning effect
Post	Perceived improvement in design skills	Measure design-oriented impact
Post	Expected future use of AAD tools	Assess professional relevance
Post	Perceived importance of parametric modelling	Assess value recognition

The motivation for introducing these surveys was twofold. First, the course aims to position itself between academic education and engineering practice, responding to a frequently observed gap between theoretical training and professional workflows. Second, input from industry partners and consulting companies—collected through departmental meetings—has increasingly emphasised the need for greater automation and computational design competence in structural engineering practice. While the surveys are not intended as formal educational assessment instruments, their repeated application over time provides insight into longitudinal trends in student perceptions.

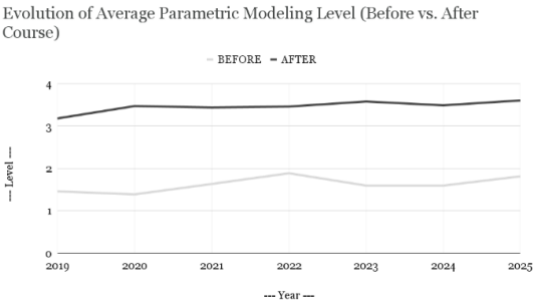


Figure 6: Self-assessed parametric modelling awareness

One consistent observation concerns students' self-assessed understanding of parametric modelling. As shown in Figure 6, students entering the course report a very low level of familiarity with parametric modelling concepts. Across multiple years, the average self-assessed level before the course remains close to 1.5 on a five-point scale, where 0 represents no prior knowledge and 5 represents a high level of understanding. This reflects the limited exposure to parametric and algorithmic design methods within traditional structural engineering curricula. After completing the course, the reported level increases substantially, reaching values close to 4.5, indicating a strong perceived improvement in conceptual understanding.

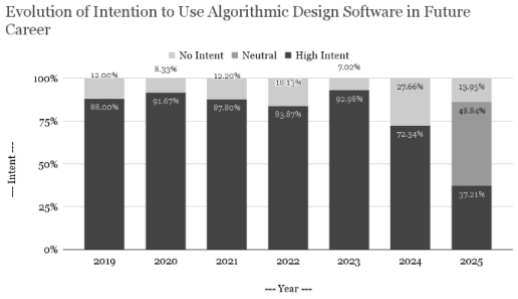


Figure 7: Intended use of AAD software in future careers

In addition to perceived competence, the surveys also address students' attitudes toward the use of algorithmic design tools in future professional practice. As illustrated in Figure 7, the majority of students express a positive intention to use parametric and algorithm-aided design methods in their careers. Until 2024, responses were limited to binary choices (yes/no). In the most recent iteration, a third option ("probably yes") was introduced, allowing for a more nuanced interpretation of student intentions. While a slight downward trend can be observed in recent years, the overall response remains strongly positive. At this stage, it is difficult to attribute this trend to specific factors, and further qualitative investigation would be required to interpret these changes.

Overall, the survey results suggest that the course format supports both increased awareness of parametric modelling concepts and a generally positive attitude toward algorithm-aided workflows. These observations reinforce the pedagogical choice to emphasise gradual skill development through practical assignments and

integrated computational workflows, while acknowledging the limitations of self-reported data.

Implications for Structural Engineering Education

An important implication emerging from the course experience concerns students' awareness of contemporary digital technologies relevant to the AEC sector. As shown in Figure 8, surveys conducted before and after the course indicate that a substantial proportion of students initially report limited or no understanding of concepts such as BIM, IFC, and the potential use of virtual reality (VR), augmented reality (AR), or artificial intelligence (AI) in design processes. While this level of awareness improves after the course, the consistently low baseline observed before the semester can be interpreted as a warning sign regarding how structural engineering students are prepared for current and emerging professional contexts.

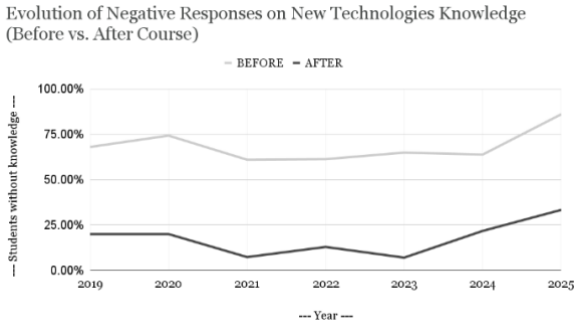


Figure 8: Awareness of new technologies in AEC

It is important to note that many of these technologies do not constitute the core of structural engineering education, where a solid understanding of mechanics, materials, and finite element analysis remains essential. However, limited awareness of digital ecosystems and emerging tools may restrict engineers' ability to participate in innovation-driven design processes. From this perspective, the issue is not the depth of technical mastery of specific technologies, but rather awareness and contextual understanding. If engineers are unfamiliar with the challenges and opportunities associated with new technologies, they are less likely to identify relevant problems or contribute to their solution.

The survey results also raise questions about the assumption that exposure to such technologies can be postponed until professional practice. While industry training plays an important role, relying exclusively on companies to introduce innovation-oriented workflows may be problematic. Early-career engineers often face limited freedom to experiment, while younger students typically show a higher willingness to explore new methods, accept uncertainty, and engage in risk-taking. From an educational standpoint, this suggests that universities play a critical role in creating safe environments for experimentation with emerging design methodologies.

At NTNU, efforts have already been made to strengthen programming competence, for example through the introduction of mandatory Python courses at the undergraduate level and the inclusion of scripting tasks in multiple courses. However, the survey responses suggest that learning programming languages in isolation, without a clear connection to design and engineering workflows, may limit students' ability to perceive their broader relevance. This highlights the importance of project-based courses that situate programming, parametric modelling, and algorithmic reasoning within meaningful engineering contexts.

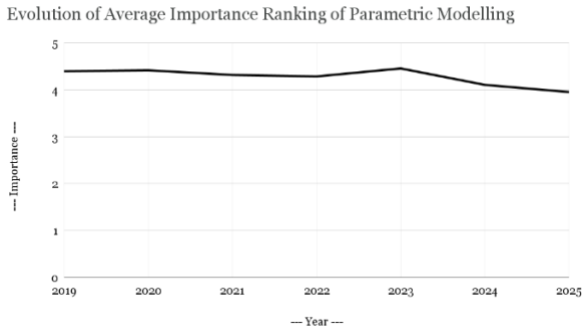


Figure 9: Perceived importance of parametric modelling

Finally, student responses regarding the perceived importance of parametric modelling provide a more optimistic perspective. As shown in Figure 9, students consistently rank the importance of parametric modelling at a high level, with average values exceeding 4 on a five-point scale in recent years. This suggests that, despite initial gaps in awareness, students increasingly recognise the relevance of computational and algorithm-aided approaches. Taken together, these observations support the integration of project-based, computational design courses as a means to bridge foundational engineering knowledge with emerging digital practices in structural engineering education.

Conclusions

This paper presents a Master's-level course that introduces algorithm-aided design into structural engineering education through a project-based, design-oriented format. Drawing inspiration from the practice of conceptual structural design in engineering, the course was developed to enable students to design structural systems from first principles, using computational tools as an integral component of the design process rather than as isolated analysis instruments.

By combining parametric modelling, finite element reasoning, and introductory algorithmic methods, the course positions computational design as a means of supporting structural understanding rather than replacing analytical thinking. The pedagogical approach is based on iterative, open-ended problem-solving tasks, where students are required to construct their own design strategies and explore the relationship between geometry, structural behaviour, and computational logic.

In contrast to traditional structural engineering education, which often focuses on well-defined problems with a single correct solution, the course adopts an open-ended assignment structure. Students are provided with design tasks and computational tools, but without prescribing specific analytical procedures or required verification methods. This shifts the emphasis from reproducing correct solutions toward developing reasoning, evaluation, and decision-making skills within a computational design context.

While the evaluation presented in this study is primarily based on self-reported data, learning effectiveness can also be assessed through observation of student outcomes. A consistent observation is that, both in assignments and in the final examination, students produce a wide range of structurally valid solutions to the same design problem. Within limited time constraints, students are required not only to generate designs, but also to assess their feasibility, including considerations of structural behaviour, constructability, and system logic.

This indicates the development of engineering judgement as a central learning outcome. In a context where computational tools enable the rapid generation of multiple design alternatives, the ability to critically evaluate, rationalise, and select appropriate solutions becomes a key competence. The course therefore aims to support not only technical skill acquisition, but also the ability to engage with computational processes in a reflective and methodologically informed manner.

Furthermore, the diversity and validity of student solutions can be interpreted as an indication of achieved learning outcomes related to design reasoning and structural understanding. Rather than applying predefined workflows, students demonstrate the ability to prototype and evaluate structural systems independently, suggesting a deeper integration of computational skills and engineering knowledge.

Overall, the course represents an initial step toward integrating computational and algorithmic methods into structural engineering education in alignment with contemporary design practice. Future work will focus on developing more systematic methods for assessing learning outcomes, particularly in relation to design reasoning, computational competence, and engineering judgement.

Acknowledgments

The authors would like to thank the students of TKT4198 for their engagement and feedback, which contributed to the continuous development of the course.

References

Norwegian University of Science and Technology (NTNU) (2025). TKT4198 – Algorithm-Aided Design in Structural Engineering. Available at: https://www.ntnu.edu/studies/courses/TKT4198 (Accessed: January 2026).

Aish, R. (2014). First Build Your Tools. In: Brady, R., Bradley, J. and O'Brien, W. (eds.) Computing in Architecture: Re-thinking the Discourse. Chichester: Wiley, pp. 25–34. Available at: https://onlinelibrary.wiley.com/doi/10.1002/9781118653074.ch2 (Accessed: January 2026).

Massachusetts Institute of Technology (MIT) (2024). Computational Structural Design and Optimization. Department of Architecture, MIT. Available at: https://architecture.mit.edu/classes/computational-structural-design-and-optimization-1 (Accessed: January 2026).

Technical University of Munich (TUM) (2024). Computational Structural Design. Chair of Structural Design, TUM. Available at: https://www.arc.ed.tum.de/en/sd/teaching/bachelor/computational-structural-design/ (Accessed: January 2026).

TU Darmstadt (2024). Spatial Structures: Design of Spatial Structures Using Parametric Design Tools. Institute of Structural Mechanics and Design, TU Darmstadt. Available at: https://www.ismd.tu-darmstadt.de/studium_und_lehre_ismd/master_vorlesungen_ismd/raeumliche_stabwerke/raeumliche_stabwerke.en.jsp (Accessed: January 2026).

Carnegie Mellon University (CMU) (2024). Advanced Structural Design: Computational Explorations. School of Architecture, CMU. Available at: https://www.architecture.cmu.edu/courses/fall-2024/advanced-structural-design-computational-explorations (Accessed: January 2026).

Stanford University (2024). CEE 220C: Parametric Design and Optimization. Department of Civil and Environmental Engineering, Stanford University. Available at: https://bulletin.stanford.edu/courses/2155732 (Accessed: January 2026).

The University of Sydney (2024). ARCH9110: Code to Production. Faculty of Architecture, Design and Planning, The University of Sydney. Available at: https://www.sydney.edu.au/units/ARCH9110 (Accessed: January 2026).

ETH Zurich (2021). Computational Structural Design 1 (CSD1). Chair of Structural Design, ETH Zurich. Referenced in: Lee, J. et al. (2021). Computational Graphic Statics for Structural Design. Proceedings of IASS 2021. Available at: https://block.arch.ethz.ch/brg/files/IASS2021_Lee-et-al_1633102829.pdf (Accessed: January 2026).

Tampere University (2022). RAK.TA.230 Modeling of Structures. Faculty of Built Environment, Tampere University. Available at: https://opiskelijanopas.tuni.fi/en/tampere-university/curriculum/course-units/tut-cu-g-47505 (Accessed: January 2026).

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

MACRO ANALYSIS OF COMPUTING IN CONSTRUCTION EDUCATION IN EUROPE

Pablo Martinez[1], Mark Baldwin[2], Niels Bartels[3], Karen Blay[4], Sebastian Esser[5], and Tobias Maile[6]
[1]Northumbria University, United Kingdom
[2]Lucerne University of Applied Sciences and Arts, Switzerland
[3]TH Köln, Germany
[4]Loughborough University, United Kingdom
[5]Technical University of Munich, Germany
[6]TH Augsburg, Germany

Abstract

This study presents the first comprehensive mapping of construction computing education in Europe, analysing 187 validated institutions across 35 countries. Using a three-tier verification methodology, we examine geographic distribution, program characteristics, and curriculum content. A sensitivity analysis estimates that current programs (~13,900 annual graduates) meet 6-49% of workforce demand depending on digitalization assumptions. Key findings include geographic concentration in Western Europe (40% in DACH and British Isles), predominance of master's programs (58.3%), and low ISO 19650 coverage (8.6%). These findings inform educational policy and workforce planning for EU construction digitalization.

Introduction

The construction sector is undergoing rapid digital transformation, driven by Building Information Modelling (BIM) mandates, sustainability requirements, and emerging technologies including digital twins, artificial intelligence, and robotics that showcase an internal transformation to tackle sector challenges. With increased external pressures, such as The European Union's Renovation Wave strategy targeting the renovation of 35 million buildings by 2030, or the revised Energy Performance of Buildings Directive (EPBD) which establishes increasingly stringent digitalization requirements, the sector is due a rather extensive upskilling of their workforce in digital competencies. These policy drivers create substantial demand for construction professionals equipped with advanced digital competencies. The European Construction Industry Federation estimates that more than two million additional workers will be needed in the sector by 2030, with digital skills identified as a critical gap (EFBWW, 2023).

BIM policies have been progressively introduced across European countries, beginning with the United Kingdom's in 2016, followed by Germany in 2021, Ireland in 2024, and additional countries including Spain, France, and Italy have implemented or are developing similar requirements as we speak. The adoption of ISO 19650 as the international standard for BIM information management further standardizes expectations for digital competency across the sector. Skills gaps and the inability to attract appropriately trained talent remain among the leading barriers to technology adoption in construction (Kjellmark & Hjelseth, 2025).

These policy mandates and industry demands translate directly into pressure on higher education institutions to produce graduates with appropriate digital competencies. Universities and technical schools across Europe have responded by developing specialized programs in BIM, digital construction, and construction informatics of various natures (Abdirad & Dossick, 2016). However, this response has been uncoordinated, with institutions independently designing curricula based on local industry needs, available expertise, and national accreditation frameworks. Several systematic reviews have highlighted the lack of common academic guidelines across institutions, noting that effective BIM implementation requires coordinated curriculum modifications and teaching strategies (Besné et al., 2021; Witt & Kähkönen, 2019). This fragmentation is compounded by the absence of a shared competency reference: while the European Digital Competence Framework (DigComp 2.2) and the ESCO classification define digital skills at a generic level, neither provides construction-specific occupational profiles for digitally skilled roles (Vuorikari et al., 2022). Moreover, European workforce intelligence reports have consistently identified construction among the sectors with the widest digital skills gaps, yet without granular data on how educational supply maps to these shortfalls (Cedefop, 2023). It is not surprising that the resulting educational landscape is fragmented: some regions have developed extensive program portfolios while others remain underserved; curriculum content varies significantly in scope and depth; and alignment with international standards such as ISO 19650 is inconsistent. To date, no comprehensive mapping of this provision exists, limiting the ability of policymakers, institutions, and industry to assess whether current educational capacity matches workforce requirements.

This paper addresses three research questions: (1) What is the current landscape of construction computing education across Europe in terms of geographic distribution, program characteristics, and curriculum content? (2) What is the estimated annual graduate output from specialized construction computing programs? (3) How does current educational capacity compare to workforce demand under different digitalization scenarios?

Methodology

To secure a comprehensive provision of the current academic offer across Europe, a sequential process is followed to gather the quantitative evidence. The research methodology comprised four stages: (1) data collection using structured searches and large language model (LLM) assisted extraction, (2) multi-tier verification against primary sources and an interactive dashboard, (3) classification and coding of programs and topics, and (4) sensitivity analysis to estimate capacity gaps.

Data Collection

Data collection followed a systematic multi-stage approach combining manual searches with LLM assisted extraction. Programs are initially identified through mass structured web searches using keywords in English and local languages (BIM, digital construction, construction informatics, computational design combined with country names and institution types). The LLM outputs provide a base dataset that is posteriorly reviewed through the institutions websites to confirm existence and relevance, as well as a review of professional body listings (buildingSMART national chapters, national BIM councils).

For each identified institution, the following data points were extracted: institution name, country, city, geographic coordinates, program names and URLs, language of instruction, delivery mode, duration, annual tuition cost (EUR), estimated graduates per year, and curriculum topics. LLM-assisted extraction was employed to accelerate initial data capture from institution websites and program documentation, with all extracted data subsequently verified against primary sources in the verification stage.

The LLM-assisted extraction used Claude Sonnet 4.5 via structured prompts requesting, for each institutional webpage, the extraction of institution name, programme title, URL, language of instruction, delivery mode, duration, tuition fees, and listed module topics. For non-English webpages, the LLM processed content in the source language with extraction fields specified in English, leveraging its multilingual capabilities; no separate machine translation step was applied. During L2 verification, approximately 60% of LLM-extracted records required correction, with errors concentrated in tuition fee fields (inconsistent currency formatting and missing fees) and graduate estimates (confused with total enrolment figures). The LLM therefore served as a productivity accelerator for initial data capture rather than a definitive extraction tool, with all final dataset values confirmed through manual verification.

Verification Framework

A three-tier verification framework was developed to assess data reliability, supported by an interactive dashboard that enabled systematic review and validation of institution records. Level 1 (L1) confirms program existence through accessible URLs and institution website confirmation. Level 2 (L2) validates core data including curriculum content verified against program documentation/specification, costs confirmed from official sources, and topics coded from syllabus or module descriptors. Level 3 (L3) additionally confirms all previous data, as well as graduate numbers, through direct institutional contact, accreditation reports, or official statistics.

The interactive dashboard served as the primary verification tool, displaying institution records with all associated data fields and enabling reviewers to flag inconsistencies, update verification levels, and track validation progress across the dataset. This systematic approach ensured consistent application of verification criteria across all institutions.

This analysis includes only institutions meeting L2 or L3 verification (n=187), of which 16 achieved L3 verification including confirmed graduate numbers. Eight L1-only institutions were excluded due to insufficient data reliability – this is due to the geopolitical situation in Eastern Europe (Russia, Ukraine, and Belarus institutions are therefore not included in this analysis for now).

Classification

Programs were classified according to ISCED levels: Bachelor studies (ISCED 6, including BEng, BA, and Foundation degrees), Master studies (ISCED 7, including MSc, MEng, MA, Mastère Spécialisé), Doctoral studies (ISCED 8), and Continuing Professional Development (CPD). Integrated programs such as the French Diplôme d'Ingénieur, which span both undergraduate and graduate levels across five years, were counted in both Bachelor's and Master's categories to accurately reflect their educational scope.

Curriculum topics were coded using a controlled vocabulary of 39 terms developed through iterative coding, consolidation of synonymous terms, and validation against ISO 19650 and buildingSMART terminology. The list of terms can be seen in Table 1.

Table 1: List of computational topics used as curriculum drivers.

Topics	
3D Modelling & Scanning	AI & Machine Learning
Architecture	BIM
BIM Software	CAD/CAM
Collaborative Workflows	Construction Engineering
Construction Informatics	Construction Management
Data Management	Digital Construction
Digital Design	Digital Fabrication
Digital Twins	Energy Performance
Facility & Asset Management	GIS & BIM Integration
Heritage BIM	ISO 19650

Infrastructure BIM	Infrastructure Engineering
Innovation & Entrepreneurship	IoT
Lean & Integrated Delivery	Parametric & Generative Design
Prefabrication & Modular Construction	Procurement & Contracts
Programming	Project Management
Risk & Safety Management	Robotics & Automation
Smart Cities & Buildings	Structural Engineering
Sustainability	Timber Construction
Virtual or Augmented Reality	nD BIM
openBIM & IFC	

During the posterior analysis, for specific geographical features, countries were grouped into ten regions based on geographic proximity and construction market characteristics: DACH, British Isles, Iberia, France, Nordic, Italy, Central Europe, Balkans, Benelux, and Baltic. The countries represented on each region and the number of institutions per region are listed in Table 2. A total of 187 institutions from 35 countries are represented in this study.

Table 2: List of regions represented in this study with its associated countries and number of institutions per region.

Region	Countries	Institutions
DACH	Germany, Austria, Switzerland	46
British Isles	UK, Ireland	29
Iberia	Spain, Portugal	25
France	France	23
Nordic	Sweden, Norway, Denmark, Finland, Iceland	12
Italy	Italy	10
Central Europe	Poland, Czech Republic, Slovakia, Hungary, Slovenia	11
Balkans	Greece, Turkey, Serbia, Croatia, Cyprus, Bulgaria, Romania, Bosnia & Herzegovina, North Macedonia, Albania	19
Benelux	Netherlands, Belgium, Luxembourg	8
Baltics	Lithuania, Estonia, Lithuania	4

Sensibility Analysis

The capacity gap analysis employs a parametric model adapted from workforce demand estimation approaches used in labour economics (Willems & De Grip, 1993; BLS, 2024). Annual demand for digitally skilled construction professionals is estimated as:

$$D = W \times P \times (r + g)$$

Where (W) represents total construction workforce, (P) is the digital skills penetration rate (proportion of workforce requiring specialized digital competencies), (r) is the replacement rate (proportion leaving annually due to retirement, career change, or attrition), and (g) is the growth rate (annual increase in digital roles).

The model structure reflects standard workforce planning practice, where total demand comprises both replacement of existing workers and expansion to meet growing requirements. Four scenarios were constructed by varying P across plausible ranges (0.03 to 0.12) and g (0.02 to 0.10), while holding r constant at 0.05 based on typical construction sector turnover rates. The lower bound for P (0.03) reflects the current proportion of explicitly BIM-designated roles in European construction job markets as reported by Cedefop (2024), while the upper bound (0.12) aligns with the EU Pact for Skills construction blueprint target and digital specialist penetration rates observed in more digitised sectors such as manufacturing (Eurostat DESI, 2024). Growth rate bounds (g = 0.02–0.10) span observed annual increases in digital construction job postings across major European markets between 2020 and 2024. Workforce data were drawn from the European Construction Observatory 2024 Skills Report, using the combined construction sector definition (NACE F + M.71) totalling approximately 18 million workers across analysed countries.

To account for institutions not captured in this mapping exercise, an institutional coverage factor (C) of 0.75-0.80 was applied to estimate total European graduate output: $S_{total} = S_{map}/C$. This estimate is a conservative number derived from observed search saturation: during the final two search iterations, fewer than 6% of queries yielded previously unidentified institutions, suggesting diminishing returns approaching dataset completeness. Unmapped institutions in smaller countries, newer programs, and non-English language offerings may contribute additional graduate output. Under this adjustment, estimated total European supply ranges from 16,062 $(C = 0.8)$ to 17,133 $(C = 0.75)$ graduates annually.

Results

Geographic Distribution

The dataset comprises 187 institutions across 35 countries. Germany leads with 32 institutions, followed by the United Kingdom (26), France (23), Spain (20), and Italy (10). Regional aggregation reveals concentration in Western Europe, with DACH countries (Germany, Austria, Switzerland) accounting for 46 institutions (24.6%) and British Isles (UK, Ireland) contributing 29

institutions (15.5%). Eastern and Southern European regions show lower provision density, with the Baltic states contributing to with only 4 institutions (2.1%).

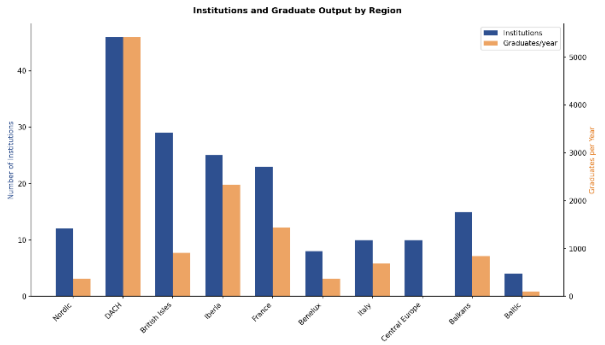


Figure 1: Bar chart of the geographic distribution of institutions and graduates (per year) per region.

Program Characteristics

A total of 372 programs were identified across 187 institutions. Master's-level programs dominate at 232 (58.3%), followed by bachelor's programs at 97 (22.6%), Doctoral programs at 55 (14.8%), and CPD offerings at 16 (4.3%). Twenty-six integrated programs (primarily French Diplôme d'Ingénieur and Double Degrees) are counted in both Bachelor's and Master's categories. Figure 2 represents the distribution of the current offer of programs that integrate construction informatics. Note that CPD offer is the lowest and may be due to bias in the data collection process or an undervalued approach to construction informatics education.

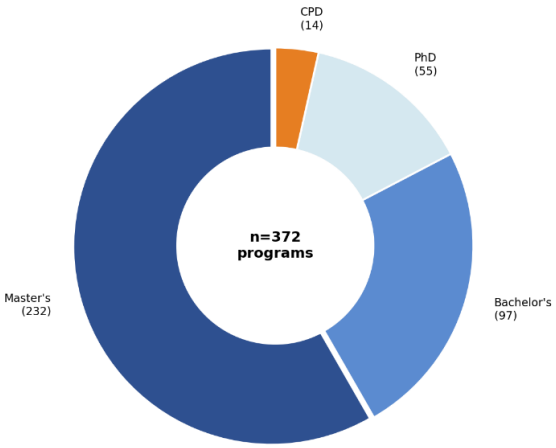


Figure 2: Pie chart of the program type distribution.

English is the most common language of instruction (96 programs), reflecting internationalization strategies particularly at master's level. German (44), French (26), and Spanish (19) serve primarily domestic markets. The predominance of Master's-level programs positions construction computing as a specialization layer atop foundational engineering or architecture education.

The overwhelming majority of construction computing programs in Europe are delivered through traditional presential (on campus) instruction, with 158 institutions (84.5%) operating exclusively in this mode. A further 11 institutions (5.9%) offer hybrid presential/online options, providing some flexibility while maintaining campus-based teaching as the primary delivery mechanism. Fully online provision remains limited, with only 5 institutions (2.7%) offering all programs entirely online, and an additional 5 operating in blended formats that combine synchronous and asynchronous elements. This distribution reflects the hands-on nature of construction computing education, where access to specialist software, hardware laboratories, and collaborative project work traditionally necessitates physical presence. However, the limited online provision may also represent a missed opportunity to reach working professionals seeking to upskill without career interruption, particularly in regions underserved by presential programs.

Full-time study dominates the European construction computing education landscape, with 114 institutions (61.0%) offering only full-time programs. A further 32 institutions (17.1%) provide both full-time and part-time pathways, while 13 (7.0%) cater exclusively to part-time students. Notably, apprenticeship models feature prominently in certain national contexts, with 12 institutions (6.4%) offering apprenticeship-only routes and 8 (4.3%) combining full-time study with apprenticeship options—these are concentrated primarily in France and Germany where dual education systems are well established. The predominance of full-time provision presents a potential barrier to workforce reskilling, as construction professionals seeking digital competencies must often choose between continued employment and formal education. The relatively limited part-time and flexible study options suggest that CPD pathways and micro-credentials may need to play a larger role in addressing the digital skills gap among the existing construction workforces.

Curriculum Content

BIM appears in 87.7% of institutions, confirming its status as the foundational topic. Digital Construction (43.9%), Construction Management (42.8%), and Sustainability (37.4%) follow. Emerging technologies show moderate coverage across programs: Digital Twins (21.4%), Parametric and Generative Design (20.3%), and AI/Machine Learning (14.4%).

Notably, explicit coverage of ISO 19650 standards appears in only 8.6% of institutions, despite its status as the international standard for BIM information management. Similarly, openBIM and IFC coverage stands at 8.6%. This suggests curriculum development has not kept pace with standardization efforts.

Regional specializations are evident. Nordic institutions emphasize Sustainability (67% versus 37% overall) and Digital Twins (58% versus 21%). DACH institutions show strength in Digital Fabrication (17%) and Robotics (11%). British institutions lead in Lean and Integrated Delivery (31%) and ISO 19650 coverage (17%). Iberian institutions focus on Infrastructure BIM (56%) and BIM Software tools (48%). Italian institutions specialize in Heritage BIM (40%) and Parametric Design (60%). This regionalisation can be evidenced in the heatmap provided in Figure 3.

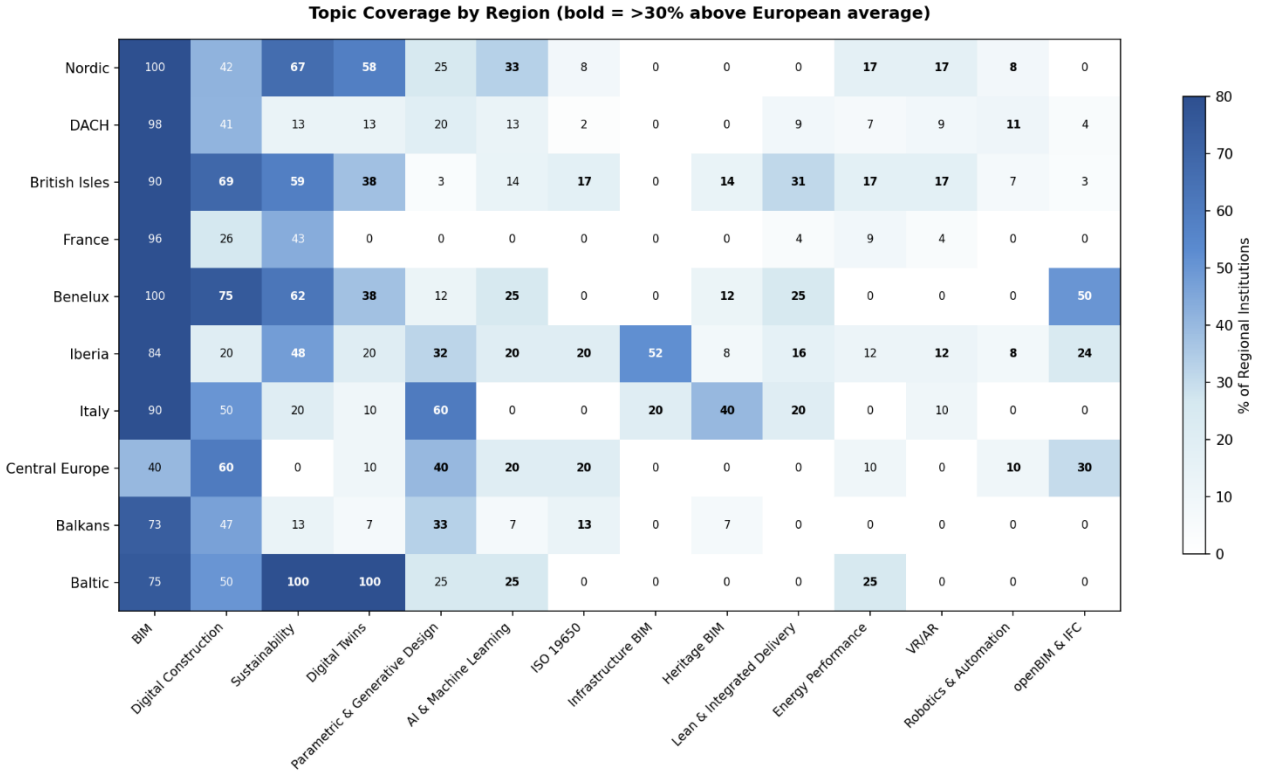


Figure 3: Heatmap of the topic coverage in academic programs across European regions.

Cost Analysis

Tuition costs for construction computing programs vary substantially across European regions, reflecting different national approaches to higher education funding (see Figure 4).

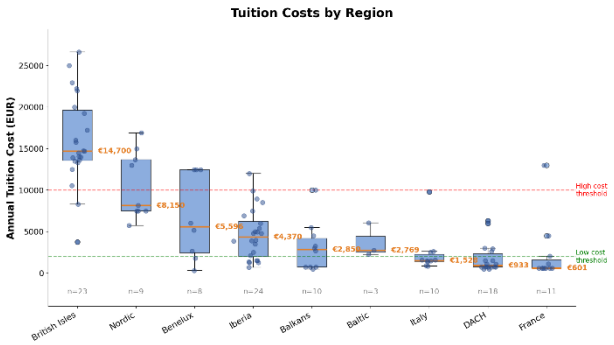


Figure 4: Box plot of the distribution of yearly cost of construction computing education per region.

The British Isles exhibit the highest median costs at €14,700 annually, with considerable variation ranging from approximately €7,000 to over €30,000 depending on institution and residency status. Nordic countries follow at €8,150 median, though this figure is skewed by high fees for non-EU/EEA students; domestic students in Norway, Sweden, and Denmark typically pay no tuition at public institutions. In contrast, DACH countries (Germany, Austria, Switzerland) demonstrate the lowest median costs at €933, benefiting from predominantly public university systems with nominal administrative fees and a predominance of part-time studies for active workers. France shows a bimodal distribution with low-cost public university programs alongside higher-fee Grandes Écoles. Southern European regions—Iberia (€4,370 median) and Italy (€1,528 median)—offer relatively affordable pathways to construction computing qualifications. This cost variation has implications for workforce development: while high-cost regions may produce fewer but potentially better-resourced graduates, low-cost regions could theoretically scale provision more rapidly to meet digital skills demands, though this would require corresponding investment in teaching capacity and infrastructure.

Graduate Capacity and Demand Gap

Graduate estimates from all 187 institutions show a total annual output from mapped institutions at approximately 13,900 graduates. Germany produces the largest volume (4,000), followed by Spain (2,048), France (1,435), Switzerland (975), and the United Kingdom (970).

When normalized against construction workforce size, smaller countries with strong programs show high output ratios: Switzerland produces 21.7 graduates per 10,000 construction workers, Greece 20.6, and Luxembourg 20.0. Larger markets show lower ratios despite higher absolute output: United Kingdom (3.9), France (6.0), and Italy (3.6).

The sensitivity analysis reveals substantial variation in estimated capacity gaps. Table 3 presents coverage percentages across scenarios and coverage assumptions. Under conservative assumptions (P=0.03, only dedicated BIM specialists), mapped programs meet 37% of

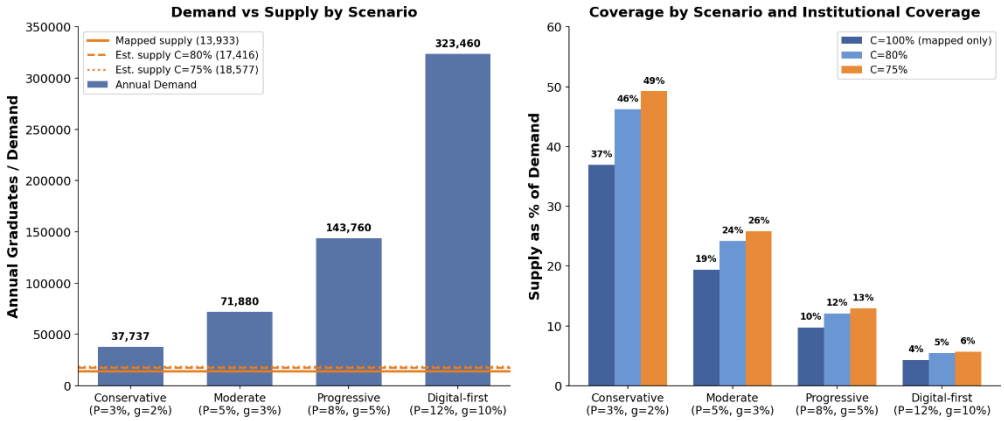


Figure 5: Sensitivity analysis showing demand versus supply by scenario and coverage percentages by institutional coverage assumption.

estimated annual demand; adjusting for institutional coverage (C=0.75-0.80) improves this to 46-49%. Under moderate assumptions (P=0.05, specialists plus power users), coverage drops to 19-26%. Progressive assumptions aligned with EU Renovation Wave ambitions (P=0.08) yield 10-13% coverage, while digital-first scenarios (P=0.12) suggest only 4-6% coverage even with optimistic supply adjustments.

Table 3: Capacity coverage by scenario and institutional coverage assumption.

Scenario	P	g	Annual Demand	C		
				100%	80%	75%
Conservative	0.03	0.02	37,737	36.9%	46.2%	49.2%
Moderate	0.05	0.03	71,880	19.4%	24.2%	25.8%
Progressive	0.08	0.05	143,760	9.7%	12.1%	12.9%
Digital-first	0.12	0.10	323,460	4.3%	5.4%	5.7%

Notes: P = digital skills penetration rate; g = growth rate; r = 0.05 (replacement rate); W = 17.97M workers. C = institutional coverage factor representing proportion of European programs captured in dataset.

Figure 5 illustrates these findings, showing annual demand under each scenario against mapped and estimated supply levels (left panel), and the resulting coverage percentages across coverage assumptions (right panel).

Discussion

Parameter Sensibility

To validate the results obtained, a parameter sensibility is performed by introducing variability and quantifying the impact on the demand. The details of the parameter sensitivity are summarized in Table 4.

Table 4: List of the parameter sensitivity ranked by impact.

Parameter	Range Tested	Impact on Demand	Total Swing
P	3% → 12%	-28,752 to +100,632	129,384
g	2% → 10%	-8,985 to +62,895	71,880
r	3% → 7%	-17,970 to +17,970	35,940
W	15M → 20M	-11,880 to +8,120	20,000

The tornado chart in Figure 6 shows how varying each model parameter individually affects annual demand, starting from a base case (moderate scenario) of 71,880 graduates of annual demand. Parameters are ranked by their impact on demand, with the most influential at the bottom.

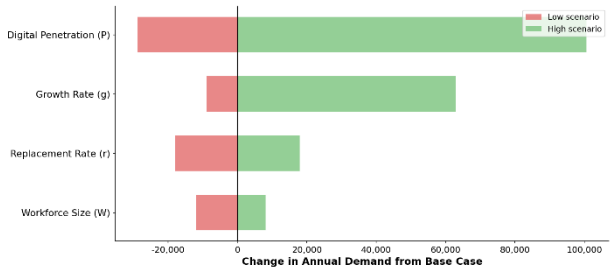


Figure 6: Sensitivity analysis on the impact of parameter variation on graduate demand.

The assumed proportion of the workforce requiring specialized digital skills has by far the largest impact. Varying P from 3% to 12% creates a swing of nearly 130,000 in annual demand, dwarfing all other parameters combined. Growth rate comes in second in impact. Assumptions about how quickly digital roles are expanding (2-10% annually) create a swing of ~72,000. Replacement and workforce size have limited impact. These parameters are relatively well-established from labour statistics and have narrower plausible ranges, limiting their influence on uncertainty. The critical unknown is not demographic (workforce size, turnover) but definitional: what proportion of construction workers

actually need specialized digital competencies? This question requires industry consensus to enable meaningful workforce planning.

Contributions to Academic Understanding

This study makes several contributions to the academic literature on construction computing education. First, it provides the first comprehensive, pan-European dataset of specialized construction computing programs, addressing a gap noted by previous systematic reviews which have highlighted the fragmented nature of curriculum documentation (Abdirad & Dossick, 2016; Besné et al., 2021). The dataset, comprising 187 verified institutions across 35 countries with standardized metadata on programs, topics, costs, and graduate capacity, offers researchers a foundation for comparative studies that was previously unavailable.

Second, the regional classification and topic coding framework enables systematic analysis of curriculum patterns. The identification of distinct regional specializations—Nordic emphasis on sustainability and digital twins, British focus on ISO 19650 and lean delivery, Iberian concentration on infrastructure BIM, Italian strength in heritage applications—provides empirical grounding for discussions of how local industry characteristics shape educational provision. Future research can build on this framework to examine whether regional specializations represent strategic differentiation or path-dependent responses to local market demands.

Third, the sensitivity analysis framework demonstrates a replicable methodology for estimating capacity gaps under uncertainty. Rather than presenting single-point estimates that imply false precision, the parametric approach makes explicit the assumptions underlying workforce demand calculations and allows stakeholders to assess implications under their preferred scenarios. The finding that definitional questions (what proportion of workers need digital skills?) matter far more than demographic parameters (workforce size, turnover rates) redirects attention toward the policy discussions that actually determine perceived shortages.

Fourth, the verification methodology offers a practical approach for maintaining data quality in large-scale educational mapping exercises. The three-tier framework (L1: existence confirmed; L2: curriculum verified; L3: graduate numbers confirmed) provides transparency about data reliability and can be adapted for similar studies in other regions or disciplines.

Implications for Curriculum Development

The findings suggest several directions for curriculum development. The near-universal coverage of BIM (87.7%) indicates that foundational digital modelling skills are well-embedded in European construction education. However, the low coverage of ISO 19650 (8.6%) and openBIM/IFC standards (8.6%) points to a gap between technical tool proficiency and information management competency. As BIM mandates increasingly reference ISO 19650 compliance, institutions may need to strengthen coverage of standardised processes, common data environments, and information delivery planning. Several non-exclusive explanations for this gap warrant investigation: ISO 19650 content may be embedded within general BIM modules without explicit labelling in programme descriptors, leading to systematic underreporting in web-sourced data; curriculum development cycles typically lag behind standardisation timelines, and ISO 19650's relatively recent transposition into many national frameworks may not yet be reflected in programme specifications; and institutions may perceive information management standards as practitioner-level knowledge better addressed through CPD than degree-level teaching. Distinguishing between these explanations would require qualitative curriculum document analysis or structured educator interviews, representing a priority for future research.

The predominance of Master's-level programs (58.3%) positions construction computing as a specialization layer atop foundational engineering or architecture education. While this model produces graduates with deep expertise, it may constrain the scaling of digital competencies across the broader workforce. The relatively limited CPD provision (4.3% of programs) represents a missed opportunity to upskill existing practitioners who cannot undertake full degree programs. Institutions and professional bodies might consider expanding short-course offerings, micro-credentials, and stackable qualifications that allow working professionals to acquire specific competencies incrementally.

Limitations

Several limitations should be noted. The dataset excludes vendor certifications (Autodesk, Bentley, Trimble), non-affiliated training programs or online programs (i.e., buildingSMART certificate), other informal training, and general IT/CS graduates who enter construction, representing significant but unmeasured supply channels. Graduate numbers are estimates for many institutions, with actual output potentially varying by ±30%. The institutional coverage factor (C=0.75–0.80), while supported by search saturation patterns and partial national cross-referencing, could not be empirically calibrated for all 35 countries and may understate or overstate unmapped provision. There is additional bias in the data collection process as the authors do not represent nor know the whole picture of the educational offer in this topic, therefore adding uncertainty and regional bias to the data collection process. The parametric demand model abstracts complex labour market dynamics and should be interpreted as indicative rather than predictive. Finally, the dataset represents a snapshot; given the rapid evolution of construction technology, regular updates would be needed to maintain currency.

Future Research Directions

This study opens several avenues for future research. Longitudinal tracking of program evolution would reveal how curricula respond to changing industry demands and policy requirements. Graduate outcome studies could assess whether regional curriculum specializations translate into differentiated career pathways. Comparative

analysis with vendor certification data would illuminate the relative contributions of academic and industry-based training to workforce development. Finally, extension of the methodology to other regions—North America, Asia-Pacific, Middle East—would enable global comparisons of construction computing education provision and potential strategic best practice.

Conclusions

This study presents the first comprehensive mapping of construction computing education across Europe, identifying 187 validated institutions across 35 countries producing approximately 13,900 graduates annually (estimated 17,400-18,600 total when accounting for search uncertainty). Key findings include: (1) geographic concentration in Western Europe with 40% of institutions in DACH and British Isles regions; (2) predominance of Master's-level programs (58.3%) positioning construction computing as specialization; (3) near-universal BIM coverage (87.7%) but limited ISO 19650 standards coverage (8.6%); (4) strong regional curriculum specializations aligned with local industry characteristics; and (5) current programs meeting between 6% and 49% of estimated demand depending on digitalization and coverage assumptions.

Meeting EU construction digitalization ambitions will likely require both scaling of specialized programs and development of complementary pathways including CPD offerings, vendor certification recognition, and integration of digital competencies at bachelor's level. Policy clarity on role definitions and competency requirements is essential for evidence-based workforce planning.

The authors also hope that the map dashboard developed is of useful to future students to take informed decisions regarding their instruction in the digital construction domain. While we do not aim to transform the created platform into a marketplace, we hope that it serves as an easier method for students to understand the European educational offer in this field, enabling them to succeed in their desired professional ambitions.

Acknowledgments

The authors would like to acknowledge the support of the EC3 Educational Committee members and the other colleagues that have validated the dataset for their own institutions.

References

Abdirad, H., & Dossick, C. S. (2016). BIM curriculum design in architecture, engineering, and construction education: a systematic review. *Journal of Information Technology in Construction (ITcon)*, *21*(17), 250-271. https://www.itcon.org/2016/17.

Besné, A., Pérez, M. Á., Necchi, S., Peña, E., Fonseca, D., Navarro, I., & Redondo, E. (2021). A systematic review of current strategies and methods for BIM implementation in the academic field. *Applied Sciences*, *11*(12), 5530. DOI: 10.3390/app11125530.

BLS. (2024). Estimating occupational replacement needs. U.S. Bureau of Labor Statistics. [Online: https://www.bls.gov/emp/documentation/replacements.htm – Accessed: 09/01/2026]

Cedefop. (2023). Skills in transition: the way to 2035. *Luxembourg: Publications Office of the European Union*. DOI: 10.2801/438491.

EFBWW. (2023). Skills and Quality Jobs in Construction: Executive Summary. European Federation of Building and Woodworkers & ITUC Just Transition Centre. [Online: https://www.efbww.eu/publications/reports-and-studies/skills-and-quality-jobs-in-construction-in-the-framework-of-the/2849-a - Accessed: 09/01/2026]

European Commission. (2024). Report on the State of the Digital Decade 2024. *Brussels: European Commission.* [Online: https://digital-strategy.ec.europa.eu/en/library/report-state-digital-decade-2024 — Accessed: 09/01/2026]

Kjellmark, G., & Hjelseth, E. (2025). Construction 5.0: Exploring Digital Skill Gaps In Construction Education. In *Proceedings of the 2025 European Conference on Computing in Construction*. DOI: 10.35490/EC3.2025.258.

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens. *Luxembourg: Publications Office of the European Union*. DOI: 10.2760/115376.

Willems, E., & De Grip, A. (1993). Forecasting replacement demand by occupation and education. International Journal of Forecasting, 9(2), 173-185. https://doi.org/10.1016/0169-2070(93)90003-6.

Witt, E., & Kähkönen, K. (2019, May). BIM-enabled education: a systematic literature review. In *10th Nordic Conference on Construction Economics and Organization* (pp. 261-269). Emerald Publishing Limited. DOI: 10.1108/S2516-285320190000002042

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

TOWARDS A STREAMLINED DIGITAL PERMITTING PROCESS: A SWEDISH CASE STUDY

Nischal Bhattarai, Annika Moscati , Ulf Seigerroth , Alexandros Tzanetos , Erik Bergström
Jönköping University, Jönköping, Sweden

Abstract

Environmental permitting is a key instrument for enabling sustainable development, but it is often characterized by long processing times, administrative burdens, and uncertainty about outcomes. Although prior research and government inquiries highlight weaknesses in the Swedish permitting process, empirical evidence on how these challenges manifest themselves in practice remains limited. This study investigates the permitting process by triangulating a scoping review, government reports, and an exploratory case study. The findings identify challenges that constrain information flow, coordination, and decision-making, underscoring the need for a digitally enabled process architecture to support a more coordinated, streamlined permitting process.

Introduction

Environmental permitting ensures that infrastructure developments comply with environmental standards. European initiatives, such as the European Green Deal and REPowerEU, emphasize accelerating sustainable projects to achieve decarbonization by 2050 (European Commission, 2019). Achieving these objectives requires accelerating the deployment of clean energy while ensuring environmental protection and addressing regulatory constraints (IEA, 2023).

Sweden is recognized for demonstrating high ambition in environmental policy and climate governance, supported by advanced regulatory structures and strong eco-innovation performance (OECD, 2025). However, challenges remain in translating these ambitions into effective implementation across environmental domains and in coordinating the environmental permitting process (European Commission, 2022). Increasing pressure to accelerate permitting for electrification and infrastructure central to the green transition also exposes the limitations related to regulatory complexity and the lengthy approval process (Phan et al., 2025; Clean Energy for EU Islands Secretariat, 2023).

Although the permitting process aims to balance environmental protection and industrial growth, it remains complex and time-consuming, increasing procedural burdens and uncertainty (Marin et al., 2024). The processes involve multiple interdependent stages, stakeholder interactions, and decision points, affecting coordination within the permitting process (Fauth et al., 2025). In the Swedish environmental permitting process, key stages, such as examining environmental impact assessments, assessing the completeness of the application, consultation, supplementation, and written exchanges, are handled through sequential administrative routines that increase manual coordination and iterative exchanges (Miljötillståndsutredningen, 2024a; Swedish Environmental Protection Agency, 2025). Furthermore, handling and disclosing personal and sensitive data during stakeholder coordination and collaboration also requires a higher level of protection and poses challenges to facilitate such exchanges (Malmaeus et al., 2021).

Digital technologies have been proposed to improve environmental assessments, including GIS-based decision support systems, data-driven analytics, and platforms for real-time information sharing and monitoring (Fothergill and Murphy, 2021). While the implementation of such digital technologies in environmental governance could facilitate green transformation (Zhao et al., 2025), their effectiveness depends on how they are designed, implemented, and integrated into existing institutional and governance contexts (Kloppenburg et al., 2022).

In Sweden, the government inquiry report (SOU 2024:98) highlighted the need for a more coordinated and streamlined digital permitting process, noting that current systems are fragmented, lack coordination, and are insufficient to ensure transparency and predictability (Miljötillståndsutredningen, 2024a). The growing demand for procedural efficiency positions the Swedish permitting system as a valuable case for analyzing both structural challenges and potential reform pathways. Thus, the paper aims to contribute to the understanding of the underlying problems and challenges as a prerequisite to the conceptual design of a streamlined digital permitting process through triangulation of international scientific literature, a case study, and a Swedish government inquiry report addressing the following research question:

RQ: What are the challenges related to the Swedish permitting process that hinder the development of a streamlined digital permitting process?

Background

The Environmental Permitting Process in Sweden is governed by the Swedish Environmental Code, its associated ordinances, along with complementary administrative

and legal regulations (Miljötillståndsutredningen, 2024b). The Environmental Code provides a framework for environmental protection, including requirements for environmental assessments and permits for activities that may have environmental impacts (Cinthio, 2019). Such activities, defined as environmentally hazardous activities due to emissions, discharges, or other disturbances, are subject to a permit or notification obligation.

The environmental permitting process begins with an early investigation and screening consultation with the County Administration Board to determine whether the activity has a significant impact or not (SEI). If SEI is expected, a full environmental impact assessment (EIA) is required; otherwise, a simplified EIA is prepared (Malmaeus et al., 2021). In the next step, the application is formally submitted, together with the EIA and supporting documents, to the competent authority: Land and Environmental Court (activity type "A"), County Administrative Board (activity type "B"), and Municipality (activity type "C") (Swedish Environmental Protection Agency, 2025). After the application is received by the competent authority, it reviews the application completeness, publishes it for public comments, and exchanges. Once the comments are received from stakeholders, the operator addresses them through clarification and, if necessary, supplementation. Finally, the authority decides on the application after receiving clarifications from the operator (Cinthio, 2019). The operator may appeal the decision of the competent authority to higher instances, depending on the authority involved. The highlighted process is a brief overview; however, the actual process often involves sub-activities within each step, multiple supplementation rounds, iterations, and exchanges.

Digitization has been considered a strategic priority for improving decision-making in public governance in Sweden (Regeringskansliet, 2025). Studies have emphasized the role of digital systems in enabling coordination among stakeholders and supporting decision-making (Axelsson et al., 2025; Larsson, 2023). Although digital infrastructure for environmental data and biodiversity monitoring exists in Sweden, these systems are implemented across various authorities, leading to fragmented responsibilities and coordination challenges. The absence of a centralized data platform and the diversity of data formats and standards across systems reduce interoperability and limit the integration of digital systems into unified workflows (Biodiversa+, 2024). Consequently, while digitization in Sweden is relatively mature in terms of data infrastructure and information availability, it remains less developed in supporting integrated, end-to-end process architectures required to coordinate complex regulatory procedures, such as environmental permitting.

Research Method

This study triangulates the scientific literature, an exploratory case study, and SOU 2024:98 to investigate challenges in the permitting process in practice. A scoping review was used to gather information on the normative permitting process in Sweden, while an exploratory case study investigated how complex interactions influence the permitting process in a real project context, where normative process descriptions alone may not fully capture real-world complexity.

Scoping literature review

The PRISMA method for scoping literature reviews was applied (Tricco et al., 2018). The scientific literature was systematically selected and reviewed to analyze the state of the art in environmental permitting and identify the complexity of the process. To obtain a broad overview, relevant literature was searched in Scopus, Web of Science, and Google Scholar. An initial generic search was conducted to identify key terms. The keywords *Permitting process, Environmental impact assessment (EIA), Environmental regulations, Regulatory frameworks, Governance and policies, and Sector/context*, along with their synonyms, were combined using various Boolean operators to achieve the widest possible coverage of relevant studies published between 2020 and 2025. A total of 559 records were identified, exported in .ris format, and imported into the web-based software Rayyan for further screening. Automatic duplicate detection was run, and consequently, 212 records were removed. The remaining 318 articles were screened based on title and abstract, of which 27 were deemed relevant. Following the full-text assessment, 12 articles were selected for a detailed analysis, as illustrated in Table 1. The findings from the selected articles were extracted and organized in a structured format. The data were then categorized into thematic codes following an open coding approach to identify the recurring themes that support the analysis of the study (Strauss and Corbin, 2015).

Table 1: Summary of literature search and screening process

Stage	Records
Initial database search (Scopus, Google Scholar, and Web of Science)	**559**
Duplicates removed	212
Records after duplicates removed	**318**
Title and abstract screening excluded	291
Marked as "possibly relevant" (title/abstract)	27
Full-text studies included in thematic synthesis	**12**

Grey literature review analysis

Understanding of the Swedish permitting process was developed through grey literature, which enabled analysis of official reports, permitting process guidelines, and government inquiries (SOU 2024:98). Some of the grey literature was originally written in Swedish and was accurately translated using Microsoft Word's built-in translation features. The selection of such literature followed the CRAAP test, which evaluated these selections based on five criteria: Currency, Relevance, Authority, Accuracy, and Purpose (Meriam Library, 2010). The inquiry report (SOU 2024:98, part 2) was analyzed as secondary data using a document analysis approach to examine challenges in the Swedish permitting process. The themes thus identified in the investigation report were categorized into codes and reported in this study.

Case Study and focus group

A case study design was adopted to investigate the Swedish permitting process in a real-world context, where such factors are critical for identifying persistent challenges. The study provides analytical generalization by examining a project that followed the full permitting trajectory, including the application and judicial review (Yin, 2018). Thus, this case study provides empirical evidence of the underlying challenges that hinder the digitization of the permitting process. The case serves as the primary data source and is thematically analyzed to identify barriers to process efficiency and digitization. However, analysis is limited to the permitting phase and excludes post-permitting stages.

The case study analyzes the permitting process of one of the established industrial and infrastructure development companies in the Nordic region. The company is involved in the handling of aggregates and quarrying operations, which are classified as environmentally hazardous activities under Swedish Environmental legislation. Depending on the environmental impact and the scale of the project, such activities undergo a permitting process and are subject to environmental permit requirements. Thus, the case company is highly relevant, as it directly interacts with Swedish environmental regulatory frameworks and the permitting process, and serves as the basis for selection.

Furthermore, a focus group discussion was conducted using a purposive sampling of participants directly involved in the permitting process, including administrative practitioners, EIA specialists, an engineer, and a project manager (Creswell, 2009; Patton, 2015). The strategy followed a key informant approach, given their relevance and expertise in the process, and a homogeneous sampling approach due to their participation in the common permitting process (Patton, 2015). The discussion was guided and anchored by a simplified process model developed in accordance with the Swedish Environmental Protection Agency, and practitioners were allowed to express their opinions and reflect on the permitting process. Subsequently, notes were taken and categorized into the themes related to the permitting process in practice. The notes were verified through the member-checking approach and thus served as a tool for contextualizing the case study and interpreting the findings. Even though the purposeful sampling approach can introduce selection bias, several strategies, such as data triangulation, member checking, diverse participants, and focused discussion, helped ensure its credibility and reliability (Creswell, 2009).

Case Study

The case study analyzes a permitting process for the extension of an existing quarry operation, including rock blasting, sorting, asphalt production, landfill operations, and water activities, under the Swedish Environmental Code. The operation, active since 2007, required a new permit application in 2021 as the existing one was set to expire in 2022. Due to the scale and nature of the project, it was classified as Activity type "A" and examined by the District Land and Environment Court (Mark- och miljödomstolen, MMD). It was assessed to have a significant environmental impact, requiring consultation and submission of a full EIA.

During the review, the authorities requested that the application be supplemented and clarified regarding water management and landfill design, and the application was made public by the authority upon completion. Stakeholders raised several concerns about groundwater impacts, blasting noise, and trust in the evaluation of the process. The court proceeded after receiving clarification of the stakeholders' concerns and granted a permit with multiple operating and monitoring conditions.

The applicant appealed parts of the judgment (8, 9, and D3) to the Land and Environmental Court of Appeal (Mark- och miljööverdomstolen, MÖD), requesting clarification and amendments to specific permit provisions: Conditions 8, 9, and D3. The appeal focused on the ambiguous formulations of requirements related to sampling and classification, the restriction on the use of inert waste within the landfill without clarification, and the absence of explicit delegation that allows the supervisory authority to adapt inspection methods. The applicant did not challenge the environmental assessment of the reviewing authority, but argued that rigid and ambiguous conditions could create legal uncertainty. The court reviewed the appeal and, in 2023, amended these provisions while keeping the rest unchanged and legally binding.

The process involved document-based exchanges, including the submission of an application, raising stakeholders' concerns via a contact form, and the examination and interpretation of textual descriptions, permit conditions, and compliance criteria by both the applicant and the reviewing authority. The case study analysis presents an empirical investigation of the permitting process in practice and offers insights into the challenges identified in the literature, serving as a prerequisite for designing an efficient, streamlined, and coordinated digital permitting process.

Discussion and result analysis

Based on the literature review, the case study, and SOU 2024:98, several challenges in the Swedish permitting process are identified and summarized in Table 2. Each of these challenges identified in the study indicates underlying requirements for the conceptual design of a streamlined digital system to enable a transparent, coordinated, and efficient permitting process. The following sections present the challenges observed, the underlying requirements that these challenges produce, and analytical solutions that could potentially address such issues in the permitting process: *Regulatory Ambiguity and Interpretive Inconsistency, Institutional and governance constraints, Procedural sequencing failure, Stakeholder participation and legitimacy Issues, Confidentiality and Competition, Data Fragmentation and Usage Limitation.* The following subsections present the challenges observed in the Swedish permitting process.

Regulatory Ambiguity and Interpretive Inconsistency

One of the challenges identified in this study is the regulatory ambiguity and interpretive inconsistency in the permitting process. The case study demonstrates how ambiguous and unclear permit conditions triggered appeals due to difficulties in interpreting requirements.

Studies show that flexible, loosely harmonized regulatory criteria can lead to diverse interpretations and ambiguous decisions across regulatory authorities, prolonging the permitting process due to uncertainty (Wilson and Allard, 2023; Söderholm et al., 2022). A study by Wretling et al. (2022) highlights that spatial wind power plans in Sweden are applied and interpreted inconsistently across municipalities, leading to variation in their use as a decision-support tool.

Similarly, the SOU inquiry reports that the reviewing authorities are often perceived as not providing a clear position on requirements, such as what supplementary information is needed when assessing the completeness of the application. It highlights that the variation persists in permitting assessment across different authorities, leading to differences in requirements across regulating bodies, and illustrates the interpretative inconsistency of regulatory authorities (Miljötillståndsutredningen, 2024a).

This challenge of diverse regulatory interpretations indicates the need to represent legal requirements, provisions, and permit conditions in a clearer, standardized, and traceable manner. A potential solution could be to translate such regulatory requirements and provisions into machine-readable representations, such as ”rule as code”. This approach may not completely eliminate ambiguity; however, it can make underlying assumptions, thresholds, and relationships between regulatory provisions explicit, transparent, and traceable (Huggins et al., 2024). In addition, a decision support repository for storing prior applications, permit conditions, interpretations, and decision standards could complement it by linking rule representations with case-based evidence through structured metadata. This enables systematic comparison and alignment of interpretations across cases and regulatory bodies, therefore, improving consistency and predictability in the permitting process (Boess et al., 2021).

Institutional and governance constraints

Institutional and governance constraints are also identified as permitting challenges. Environmental assessments often require intergovernmental coordination due to overlapping jurisdictions, resulting in a fragmented, uncoordinated process (Wilson and Allard, 2023). In Sweden, the decentralized, multi-authority structure lacks horizontal coherence, while limited administrative capacity and expertise reduce efficiency (Streit et al., 2025). The increasing regulatory requirements for a broader scope of assessments, project characteristics, and institutional capacity make the permitting process time-consuming and pose a systemic challenge (Johnson et al., 2023; Söderholm et al., 2022).

Similarly, the SOU inquiry highlights unclear investigation requirements and reliance on referral authorities, and the absence of a clear threshold for application completeness. Limited guidance, staff turnover, and resource constraints at the regional and local level further affect consistency and timeliness, while increasing regulatory demands linked to the green transition intensify this pressure (Miljötillståndsutredningen, 2024a).

The case study also highlights decentralized institutional responsibilities and fragmentation in the current permitting process. The reviewing authority’s request for opinions from other regulatory bodies, such as the Building and Environment Administration, the County administration board, and the Geological Survey of Sweden, during application assessment indicates how the reviewing authority’s decisions could be influenced by such referral authorities.

The findings suggest the need for a coordinated, transparent system with common infrastructure, interoperable information exchanges, and clear role designations across authorities. A potential solution could be the design of a coordinated, centralized permitting platform that integrates data, actors, and procedural steps into a unified process architecture. This reduces fragmentation and improves information flow between stakeholders (Miljötillståndsutredningen, 2024a).

Procedural sequencing failure

Unclear procedural sequences and insufficient cross-referencing between regulations create additional challenges in the permitting process. The effectiveness of stakeholder participation is influenced by the sequencing and timing of procedural steps, as unresolved issues deferred to later stages can amplify risks and conflicts (Samec Berghaus et al., 2025; Fauth et al., 2025; Bernauer et al., 2023). Similarly, unclear scope definitions during consultation and redundant information flows across multiple authorities contribute to inefficiencies and delays (Freeman et al., 2022).

Table 2: Summary of the identified challenges in the permitting process and potential analytical solutions

Challenges Identified	SOU 2024:98	Case study/focus group	Literatures	Analytical Solution
Regulatory Ambiguity and Interpretive Inconsistency	X	X	X	Rule-as-code representations, decision-support repository and structured metadata to improve consistency, traceability, and predictability
Institutional and governance constraints	X	X	X	Coordinated centralized permitting platform with common infrastructure, interoperable information exchange, clear role designation across authorities
Procedural sequencing failure	X		X	Workflow orchestration model, rule-based validation, early application-completeness checks, clearly defined stages and information requirements
Stakeholder participation and legitimacy Issues	X	X	X	Digital stakeholder participation with access to project information, traceable feedback and monitoring, transparent review process
Data Fragmentation and Usage Limitation	X	X	X	Interoperable data infrastructure, harmonized metadata, API-based integration
Confidentiality and Competition	X	X		Data-sharing governance model, access control for sensitive information, selective sharing of commercially non-sensitive data

The SOU inquiry also highlights that procedural sequencing and timing, such as assessing completeness of the application and handling of supplementation, are recurring challenges in the Swedish permitting process. It notes the absence of a clear starting point of the deadline that the authority can confirm for the completeness of the application, and when statutory processing times begin. Workflow sequencing varies across authorities: some complete substantial work before formal processing begins, while others proceed only after formal public announcements. Thus, insufficient coordination and sequencing of procedural activities are also identified as contributing to delays in the permitting process (Miljötillståndsutredningen, 2024a).

The identified challenge, as presented above, points to the need for systems with a clear process flow, controlled sequencing of data and activities, and an early validation mechanism to ensure application completeness. A solution to the identified challenge could be implementing a workflow orchestration model in combination with a rule-based validation mechanism. This may reduce the need for iterative supplementation by structuring the process into clearly defined stages and information requirements.

Stakeholder participation and legitimacy Issues

Stakeholder participation and legitimacy are also identified as key challenges affecting the permitting process. While stakeholders play a critical role in shaping permitting outcomes, ineffective participation can undermine the EIA process and increase distrust (Bernauer et al., 2023). Although participation of affected groups, including indigenous communities, has improved, meaningful engagement remains constrained by resource limitations and procedural designs that marginalize concerns (Bernauer et al., 2023; Johnson et al., 2023). Delayed or insufficient early consultation, along with an unclear division of responsibilities, further contribute to conflicts during the review phase (Fauth et al., 2025). Wretling et al. (2022) argues that infrastructure development cannot realistically be achieved without harming the environment, so influencing the local opposition through a narrative of creating local jobs may not be sufficient

The inquiry indicates that the role of authorities in the Swedish permitting process is often perceived as unclear and unpredictable. Supervision and self-monitoring are perceived as person-dependent, raising concerns about consistency and competence. Inconsistent assessments of

similar activities across counties further reflect unclear roles and definitions, sometimes prompting applicants to adjust permit pathways to avoid uncertainty. This lack of procedural clarity leads to repeated supplementation and can prolong the permitting process. The inquiry notes that in some cases, authorities act simultaneously as advisors and supervisors, creating ambiguous roles that influence applicants' strategic choices regarding permit pathways (Miljötillståndsutredningen, 2024a).

The case study also illustrates that local stakeholders lack trust in the process, as they highlighted insufficient consultation and transparency requirements during the written exchange phase of the permitting process. Consequently, the demand for increased monitoring and for addressing perceived risks posed by activities in the vicinity was evident.

The issue of stakeholder participation and legitimacy indicates that transparent, traceable feedback and monitoring, and an inclusive participation mechanism are needed. A solution to address this challenge could be to enable digital stakeholder participation by providing access to project information, facilitating stakeholder engagement, tracking submitted concerns, and ensuring a transparent review process (Boess et al., 2021). Participation is most effective when it combines representation, communication, and delegation of power, and the mechanism ensures meaningful and traceable stakeholder influence in the decision-making process (Newig et al., 2023; Fothergill and Murphy, 2021).

Confidentiality and Competition

Confidentiality and competition are also identified as challenges that could lengthen permitting timelines. Although transparency is required, access to relevant data is constrained by legal considerations, institutional practices, and differences in digital information sharing across stakeholders (Miljötillståndsutredningen, 2024a). This limits the effectiveness of using the data and makes managing it more difficult.

The focus group also revealed that the information disclosure in the permitting process depends upon the sensitivity of the projects. For instance, projects that require handling and use of explosives, a rock quality assessment, and detailed exploration plans may be considered commercially confidential and therefore not shared publicly. This creates conflict between the transparency and the protection of commercial interests. It was emphasized that while early investigations allow applicants to assess and identify strategically viable locations, the public nature of the permitting process increases the risk of exposing competitive information.

It highlights the need for a secure data-sharing mechanism with controlled access to sensitive information, while balancing transparency and confidentiality. A solution to mitigate this issue could be implementing a governance model for data sharing and access control within the permitting process, thereby balancing transparency and competition. In addition, evidence from practice highlights the importance of structured metadata, interoperability standards, and the selective sharing of commercially non-sensitive data to enhance transparency while avoiding disclosure of sensitive information (Boess et al., 2021).

Data Fragmentation and Usage Limitation

Studies suggest that limited data accessibility and use also pose challenges for permitting, especially when evidence-based assessment depends on extensive environmental monitoring data (Streit et al., 2025). For example, in water discharge permitting, a lack of appropriate tools, monitoring technology, and structured emission can hinder evidence-based decision-making (Jensen et al., 2020). The presence of datasets from prior projects could help address stakeholders' recurring concerns and improve consultation efficiency (Freeman et al., 2022). Studies have indicated that the lack of accessible data for environmental permitting investigations can add a burden on applicants to conduct additional studies, which, otherwise, could help address recurring stakeholder concerns and enable evidence-based decision-making (Jensen et al., 2020; Streit et al., 2025; Freeman et al., 2022).

The inquiry report notes that, although County Administrative Boards and the Swedish Environmental Protection Agency initiated the collaboration to publish metadata for environmental assessment cases in digital form, the service may not have covered all relevant aspects and does not meet EU requirements. It emphasizes that the availability of data for measurement and monitoring enables reuse across different cases and can also facilitate automation (Miljötillståndsutredningen, 2024a).

The discussion from the focus group also highlights the challenges posed by the lack of accessibility of detailed data in the permitting process and indicates that, even though initial screening is performed using available information, the existing data lacks sufficient detail for assessment, requiring further studies. Furthermore, it notes that regions vary in digital adoption maturity, and the challenges lie in moving beyond the digital formats.

This challenge indicates that the data are fragmented and distributed across authorities, requiring improved integration, standardization, and interoperability. Due to the institutional independence of regulatory agencies in Sweden, the collection of such data into a centralized system might be difficult; therefore, enabling interoperability and integration between data infrastructures such as the Swedish Biodiversity Data Infrastructure, Species Observation Systems, agencies using shared, harmonized metadata, and application programming interface could be a potential solution (Biodiversa+, 2024).

Conclusions

This study investigated the challenges in the Swedish permitting process through a scoping review of available literature and conducted an exploratory case study. The findings reveal that the existing permitting challenges do not stem primarily from a single factor but rather from several

factors associated with the permitting process, distributed across different stages. Furthermore, it highlights that the permitting process extends beyond a regulatory procedure and involves complex socio-technical aspects, heavily influenced by interactions among legal requirements, stakeholders, procedural complexities, and dependencies. These challenges are not merely procedural but reflect underlying socio-technical constraints related to information flow, coordination, and decision-making across distributed actors. As such, they indicate the need for digitally enabled process architectures that can support structured workflows, interoperable data exchange, and transparent decision-making within the permitting system. Therefore, digitization efforts must go beyond standalone solutions and focus on integrated systems that could facilitate an improved, transparent, and efficient process.

This study also acknowledges several limitations. The analysis is based on a single case study and adds analytically to the understanding of issues within the permitting process; however, it limits generalization to the broader population. Moreover, though the approval process includes application, decision, and appeal, excluding the post-permitting stage limits the study's ability to capture challenges that may arise during implementation and monitoring. In addition, selecting focus group participants through a purposive sampling approach within a single organization may still introduce bias and limit the diversity of perspectives. Lastly, the study identifies digitization requirements at the conceptual level but does not provide empirical testing of the specific technologies to address these challenges. Future research will examine the conditions and requirements for conceptualizing and implementing the digital environmental permitting platform.

Acknowledgments

This work was supported by the project "A streamlined environmental permitting process for construction projects related to critical infrastructure (SMARTE)" [1] and co-funded by Smart Built Environment [2] (project ID: U13-2025-01).

References

Axelsson, J., Stureborg, E., Dochtorowicz, B., Kaaman, M., and Ahlner, E. (2025). Implementation and structure of a contact point for net-zero regulatory sandboxes – implementation-phase proposals. Vinnova - Sweden's Innovation Agency.

Bernauer, W., Hostetler, G., Greene, E., Tester, F., Harris, R., and Tanguay, L. (2023). Undermining assessment: Eia follow-up, stake-holder advisory groups, and extractive industries in nunavut, canada. Impact Assessment and Project Appraisal, 41(2):87 – 101.

Biodiversa+ (2024). Sweden national report biodiversa+ governance sub-pilot. Technical report, European Biodiversity Partnership.

Boess, R., Lyhne, E. ., and Kørnøv, I. . (2021). Digitalisation in Environmental Assessment: International frontrunners. The Danish Center for Environmental Assessment (DCEA), Aalborg University.

Cinthio, K. (2019). Assessment of environmentally hazardous activities. Technical report, County Administrative Board, Skåne.

Clean Energy for EU Islands Secretariat (2023). Clean energy for eu islands regulatory barriers in sweden: findings and recommendations. Technical report, European Commission Directorate-General for Energy.

Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. Sage Publications, Thousand Oaks, CA, 3rd edition.

European Commission (2019). The european green deal. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en. Accessed: 16 January 2026.

European Commission (2022). Environmental implementation review 2022: Country report – sweden. Technical report, European Commission.

Fauth, J., Gade, P. N., Kaiser, S., Raj, K., Pedersen, J. G., Olsson, P. O., Nisbet, N., Ventura, S. M., Hirvensalo, A., Granja, J., Urban, H., Rutešić, S., Verstraeten, R., Raitviir, C. R., Kallinen, A. R., Schranz, C., Stojanov, T., and Tekavec, J. (2025). Investigating building permit processes across europe: characteristics and patterns. Building Research and Information, 53:417–434.

Fothergill, J. and Murphy, J. (2021). The state of digital impact assessment practice. Technical report, International Association for Impact Assessment.

Freeman, M. C., O'Neil, R., Garavelli, L., Hellin, D., and Klure, J. (2022). Case study on the novel permitting and authorization of pacwave south, a us grid-connected wave energy test facility: Development, challenges, and insights. Energy Policy, 168.

Huggins, A., Witt, A., and Burdon, M. (2024). Digital distortions and interpretive choices: A cartographic perspective on encoding regulation. Computer Law & Security Review, 52.

IEA (2023). Net zero roadmap: A global pathway to keep the 1.5 °c goal in reach - 2023 update. Technical report, International Energy Agency.

[1] https://www.smartbuilt.se/in-english/projects/value-chains-and-business-models/smarte/

[2] https://www.smartbuilt.se/

Jensen, D. M. R., Thomsen, A. T. H., Larsen, T., Egemose, S., and Mikkelsen, P. S. (2020). From eu directives to local stormwater discharge permits: A study of regulatory uncertainty and practice gaps in denmark. Sustainability, 12(16).

Johnson, E. L., Ericsson, M., and Löf, A. (2023). The mining permitting process in selected developed economies. Land Use Policy, 131.

Kloppenburg, S., Gupta, A., Kruk, S. R., Makris, S., Bergsvik, R., Korenhof, P., Solman, H., and Toonen, H. M. (2022). Scrutinizing environmental governance in a digital age: New ways of seeing, participating, and intervening. One Earth, 5:232–241.

Larsson, A. (2023). Climate transition in cities – a swedish governance model to accelerate the rate of transition. Technical report.

Malmaeus, M., André, H., Hellsten, S., Lindblom, E., Åsa Romson, and Rydstedt, A. (2021). Miljömålen i miljöbedömning och tillståndsprövning utmaningar och möjligheter [environmental objectives in environmental assessment and permit review challenges and opportunities]. Technical report, Naturvårdsverket.

Marin, D., Pravuljac, M., Tralmaka, I., and Martins, M. (2024). Understanding strategic projects under the crma: A guide to rights and processes. Technical report, ClientEarth.

Meriam Library (2010). Evaluating information – applying the craap test. California State University, Chico.

Miljötillståndsutredningen (2024a). En ny samordnad miljöbedömnings- och tillståndsprövningsprocess, sou 2024:98, del 2 [a new coordinated environmental assessment and permit review process, sou 2024:98, part 2]. Technical report, Regeringskansliet.

Miljötillståndsutredningen (2024b). En ny samordnadmiljöbedömnings- och tillståndsprövningsprocess, sou 2024:98, del 1 [a new coordinated environmental assessment and permit review process, sou 2024:98, part 1]. Technical report, Regeringskansliet.

Newig, J., Jager, N. W., Challies, E., and Kochskämper, E. (2023). Does stakeholder participation improve environmental governance? evidence from a meta-analysis of 305 case studies. Global Environmental Change, 82.

OECD (2025). Oecd environmental performance reviews: Sweden 2025. Technical report, OECD Publishing.

Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice. SAGE Publications, Thousand Oaks, CA, 4th edition.

Phan, N. A., Hellsmark, H., Göransson, L., and Johnsson, F. (2025). Electrifying tensions: Stakeholder narratives to electrification of industry and transport in sweden. Energy Research and Social Science, 126.

Regeringskansliet (2025). Sveriges digitaliseringsstrategi 2025–2030 [sweden's digitization strategy 2025–2030]. Technical report, Regeringskansliet Finansdepartementet.

Samec Berghaus, N., Ferčič, A., and Drnovšek, K. (2025). Slovenia's nuclear energy pathway: Strategic expansion, regulatory hurdles, and future prospects. Journal of Agricultural and Environmental Law, 20(38):123 – 155.

Strauss, A. and Corbin, J. (2015). Basics of Qualitative Research. SAGE Publications, Inc., 4 edition.

Streit, S., Passarella, M., and Tost, M. (2025). The influence of environmental policies on mineral permitting in europe: Challenges and insights from spain and sweden. Extractive Industries and Society, 24.

Swedish Environmental Protection Agency (2025). Så går en tillståndsprövning enligt miljöbalken till [this is how a permit review under the environmental code works].

Söderholm, P., Bergquist, A.-K., Pettersson, M., and Söderholm, K. (2022). The political economy of industrial pollution control: environmental regulation in swedish industry for five decades. Journal of Environmental Planning and Management, 65(6):1056–1087.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., Lewin, S., Godfrey, G., Macdonald, M. T., Langlois, E. V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tunçalp, and Straus, S. E. (2018). Prisma extension for scoping reviews (prisma-scr): Checklist and explanation. Annals of Internal Medicine, 169(7):467–473.

Wilson, G. N. and Allard, C. (2023). Institutional determinants of mining projects in canada and sweden: Insights from the prosperity and kallak cases. Environmental Management, 72(1):53 – 69.

Wretling, V., Balfors, B., and Mörtberg, U. (2022). Balancing wind power deployment and sustainability objectives in swedish planning and permitting. Energy, Sustainability and Society, 12.

Yin, R. K. (2018). Case Study Research and Applications: Design and Methods. SAGE Publications, Thousand Oaks, California, 6th edition.

Zhao, X., Zhang, S., and Lu, F. (2025). The impact of digital environmental governance on green transformation: Theoretical mechanism and empirical test from china. Sustainability, 17.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

IFC5 FOR COLLABORATIVE AND TRACEABLE CIRCULARITY IN CONSTRUCTION

Tong Wang[1] and Thomas Krijnen[2]
[1]Delft University of Technology, Delft, The Netherlands
[2]AECgeeks, Eindhoven, The Netherlands

Abstract

Contemporary practices like evidence-based circularity and material tracking and tracing, are characterized by collaborative processes to layer, recombine, and update information non-destructively. This paper explores how the emerging IFC5, with USD-inspired, layer-based composition, enables such needs. The data model consists of layers, representing geometry, material characteristics, and lifecycle evidence. We explain how the principles help in a social housing circular window reuse scenario using a python-based web application on top of a Postgres database. By structuring information as composable layers, the introduced data model supports evidence-based circularity tracking, allows material and component matchmaking and information management.

Introduction

The transition to circularity in the built environment increasingly depends on the sector's ability to treat existing buildings as future supply systems, i.e. the repositories of products and components that can be recovered, verified, and redeployed. This shifts the feasibility of reuse and remanufacturing away from questions like "can we dismantle this?" toward information questions: what elements exist, where are they, what are their specifications and conditions, and can this evidence travel across project boundaries? This illustrates a critical premise for circular transition: without structured, component-level knowledge of the current stock, reused components remain difficult to specify, price, insure and certify.

The current information management practices in construction are typically fragmented across project phases, proprietary tools, and organizational boundaries, which makes multi-actor updating and validation difficult and leads to "static" records that quickly diverge from reality. This fragmentation is a well-recognized barrier in the circular construction literature, a persistent "information gap" between the data needed to scale reuse and what is available in practice, spanning identification, condition assessment and traceability (Byers et al., 2024).

To overcome circular transition challenges, in practice the use of digital technologies for tracking element reuse is increasing as exemplified in cases such as Durmiseciv (2018) and Hemubo (2026).

However, to scale up this way of managing information for circular transitions, system level stakeholder participation needs to be enabled. This increases the importance of governance, versioning, and responsibility for updates, which are features that are rarely robust in conventional document-based handovers (Çetin et al., 2023). Consequently, advancing circularity requires modular data models and lifecycle information infrastructures that enable multiple stakeholders to contribute, update, and reuse information consistently with traceable information records. This paper aims to answer:

RQ1 What are limitations in contemporary IFC technology in encoding buildings as future supply systems for circular construction?

RQ2 How can IFC5 and its JSON-based serialization format (IFCx) alleviate some of these pain points?

Methods

To answer the research questions, this study follows the following steps.

First, the current IFC technology limitations have been studied in literature as well as the IFC5 core principles to overcome such limitations.

Second, literature review has been conducted to understand the circular transition context in more detail, including the emerging regulations and supporting standards which make circular transition not wishful thinking, but a must take with mandatory data requirements, as illustrated in the emerging regulations and standards section below.

In parallel, we build upon ongoing research projects that support the understanding of challenges in the lifecycle information management of built assets for modular, participatory and traceable circular practice. This enables a solid foundation to showcase the potential that the new development in IFC5 could bring.

To make evaluation more concrete, we design a simple scenario of circular window reuse with the same asset owner.

The last part presents the data model, and the demonstration of the IFC5 at various milestones along the scenario process.

Results

Findings from the literature review as well as prototypical development to assess the applicability of IFC5 in the context of circularity are shown in this section.

Limitations of current IFC

Buildings as future supply systems require a data-oriented information modelling approach and to that extent Building Information Modelling (BIM) and the predominant open standard Industry Foundation Classes (IFC) to encode such information has seen more consistent adoption in the construction industry over time. The latest official edition of IFC is IFC4.3, encoded in the EXPRESS schema language, hereafter referred to as IFC4. There are limitations in IFC4 when it comes to facilitating circular construction. Tomczak et al. (2024) concludes that availability of specific information such as component connections that are vital for understanding demountability, remains a challenge in IFC and BIM-based workflows. This relates to (i) downstream information flows with limited ability for information to travel upwards; detailed manufacturing information might only be present at the time of a construction, but due to export-centric workflows where information flows downstream, authoring tools have little ability to consume this information to maintain a comprehensive record of information; (ii) construction knowledge is distributed across multiple project partners. While there are ample research and initiatives to define and compare information requirements, these only stipulate information contents between given point to point exchanges of information. Within the industry there is little incentive outside of these exchanges to combine knowledge.

Limitations of the current IFC and EXPRESS/SPF (Step Physical File; the predominant text-based serialization format used for IFC) it is based on can be summarized as: (i)Limited support for developer tooling and mindshare of the EXPRESS language and related serialization technologies due to idiosyncrasies and datedness, which does not connect to modern web-based APIs that are quick to deliver value and test assumptions. (ii) Technical serialization-level limitations such as the inability to combine data sets, inability to establish links across information carriers and lack of modularity in the IFC schema. (iii) Lack of uniformity in data due to the extensive set of modelling mechanisms, valuing backwards compatibility over consistency. (iv) High degree of complexity that impedes mapping to other storage technologies such as relational databases. (v)Monolithic exchange structures that do not enable granular composition with other datasets or incremental updates over time (van Berlo et al. 2021). These limitations impede the use of IFC for lifecycle updates of asset data, with multistakeholder collaboration. Building richer retrieval mechanisms on top of IFC serialization and data models has been attempted in Krijnen and Beetz (2018) and Wolf et al. (2025) using SPARQL on top of an efficient binary serialization mechanism or a GraphQL mapping respectively. While that aids discovery of information, it is not a step into the direction of freely recombining information. The semantic web has also been used in disseminating construction information and allows for collaborative authoring due to the nature of linked data, however geometrical information is not always efficiently stored (Pauwel et al. 2017) and there is not a rich ecosystem for working with geometry in the semantic web. The semantic web is most powerful when fully embraced natively in applications to depend on things like inference, which is not the case to date.

These aspects are recognized in buildingSMART. And over time, starting with designing and defining future principles (BuildingSmart,2020), the technical roadmap (BuildingSmart, 2020) and first explorations into a IFC5 resulted in the publication of a standard called IFC5 alpha (BuildingSmart, 2026). It is heavily inspired by Universal Scene Description (USD) (USD, 2026) a standard used in the film industry, a sector with similar needs in collaborative authoring. From USD, IFC5 derives its layer-based composition, data inheritance mechanisms, central placement tree and focus on explicit data for robustness with additional layers of meta-data for semantic richness. With such principles, IFC5 is particularly well aligned with the need for modular, multi-stakeholder lifecycle information management because its development direction explicitly shifts IFC towards modular schemas and extensibility. Information that is important for circular construction, e.g., condition assessment, deconstruction attributes can be added and iteratively refined without forcing every stakeholder to implement the entire schema or exchange the full model.

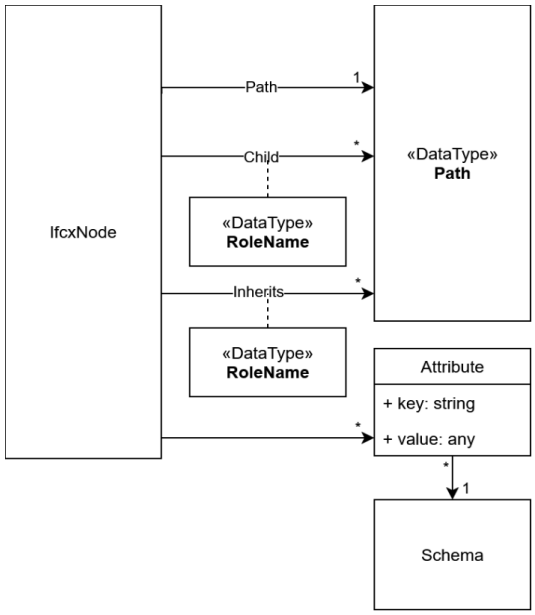


Figure 1: Schematic representation of the data model of nodes in IFCx. Children (for element decomposition and containment) and inheritance (for data reuse) are established as references to paths. The serialization remains flat without nesting. Multiple Paths do not uniquely identify a node. This realizes the collaborative distributed nature by allowing different sources to augment the shared understanding of an element

IFC5 alpha follows an exchange format in JSON with a minimal schema. The exchange structure is called IFCx

and consists of a header with provenance information, imports, schemas and a data array of primitives. Primitives are objects that have a path in the tree, inheritance relationships to other primitives, children relationships to other primitives and attributes as key-value pairs. The schemas in the header govern the validity of these key-value pairs. Imports from the header are essentially prepended to the dataset and can contain schemas and other primitives. Inheritance relationships broadcast all information, attributes and children, along the inheritance edge and are essential to realize re-use of information (see Figure 1).

Emerging regulations & supporting standards

The EU policy context for traceable, reusable construction components is increasingly shaped by a shift toward digitized, interoperable information flows. The Construction Products Regulation (CPR), Regulation (EU) 2024/3110, explicitly motivates the need for "well-functioning information flows" using electronic means and machine-readable formats, so that coherent product performance information is available along the supply chain and supports the digitalization of the construction sector (Eota, 2024). A central mechanism introduced by the CPR for product-level traceability is the Digital Product Passport (DPP). Article 76 requires that a DPP includes core compliance information (notably the declaration of performance and conformity), corresponds to the product type and its unique identification code, and allow designated actors to introduce or update information in the passport. Importantly for reuse workflows, Article 76(3) states that DPP requirements should improve traceability of products along the value chain. CPR also links unique identifiers, data carriers, and the DPP registry mechanisms to the horizontal DPP framework established under the Ecodesign for Sustainable Products Regulation Regulation (EU) 2024/1781, unless CPR delegated acts specify otherwise. For renovation, the Energy Performance of Buildings Directive (EPBD) introduces building-level information pathways via renovation passports and (where they exist) digital building logbooks (DBLs). In Article 12(7)(8) it requires renovation passport information to be uploadable to national energy performance databases, and that the "easiest approach" is to use a machine-readable, standard format that supports seamless integration and querying (European Commission, 2024). Open data schemas such as Industry Foundation Classes (IFC) are explicitly intended to provide vendor-neutral, machine-interpretable representations that enable information exchange across tools and use-cases, forming a backbone onto which circularity modules (e.g., material passports, audit results, condition states, reuse certifications) can be attached (BuildingSmart, 2026).

Lifecycle information management for circularity

Lifecycle information management has long been positioned as an enabler of collaboration in BIM-based delivery and operations, but circularity raises the bar by even requiring data continuity beyond a single project and into end-of-life recovery and second-life deployment. Standards-based guidance associated with ISO 19650 frames the core problem as ensuring the right information is created, shared, and maintained over the whole asset lifecycle through controlled processes and a common data environment (UK BIM Framework, 2021). In circular settings, this lifecycle requirement becomes sharper: component information must remain usable when the original design/construction team is gone and when new actors (deconstruction contractors, remanufacturers, reuse marketplaces, regulators, new designers) need to interpret and trust the data. Policy and practice guidance on pre-demolition or "waste" audits implicitly treat information as the first circular intervention: the European Commission's audit guidance emphasizes qualitative and quantitative assessment of what will arise from demolition/renovation in order to maximize recovery and reuse (European Commission, 2018). Yet, even where audits exist, their outputs are commonly delivered as isolated reports or spreadsheets that are difficult to integrate with design models, asset registers, maintenance systems, and marketplaces. This means that the information is generated but not maintained as an updatable asset. This is the specific information-management failure that circular pilots attempt to mitigate through traceability schemes (e.g., coded material databases in Kerkrade; component-level QR/NFC "DNA") because such mechanisms create a minimum viable link between physical components and persistent records that can be consulted by later stakeholders (Durmisevic, 2018). The same "information-first" premise appears in Amsterdam's circular window-frame practice: embedding an identifier on frames that links to a web page containing provenance, processing, finishing, and expected lifetime data which effectively treats the component as an information object that can be managed and verified across actors and over time (Hemubo, 2026). Together, these applied cases indicate that even though circularity is constrained by recovery logistics (Lu et al., 2023), emerging technologies are making assets into a next use traceable.

However, one key implication is that circularity is not well served by monolithic, one-size-fits-all datasets; instead, it requires modular data models that can (i) represent construction objects consistently, (ii) attach new lifecycle modules (condition monitoring, disassembly history, compliance evidence, ownership transfers), and (iii) support interoperability across heterogeneous systems. Research on digital product passports and material passports has increasingly converged on modular modelling as a response to the complexity and volume of lifecycle data. For example, a modular ontology modelling approach for digital product passports has been proposed specifically to address the challenge of representing, managing, and exchanging circularity-relevant information at scale, arguing that modularity improves extensibility and reuse of data structures as requirements evolve (Kebede et al., 2024). This direction is reinforced by the circular construction ecosystem (e.g., Buildings-as-Material-Banks initiatives), which frames material passports as structured datasets that describe the

characteristics giving materials value for recovery and reuse and positions them as "one-stop" information resources. This is an ambition that is only attainable if information can be structured, updated, and exchanged across actors (BAMB, 2020). The housing-sector evidence base further underlines why modularity matters: social-housing organisations need passport data that covers maintenance, renovation, and demolition use-cases simultaneously, implying that the underlying model must support different stakeholder views and responsibilities without losing semantic consistency (Çetin et al., 2023).

Implementation

To answer the second research question a protypical implementation is conceived to simulate a simplified scenario in a circular construction context. The scenario is developed and validated in the on-going research projects with multiple stakeholders in façade products to rethink future circular constructions (NGF, 2025); the phases and steps described have been discussed and validated as realistic potential to-be scenarios. This scenario is used to link various stakeholders with the modular design principle of IFC5. In this research the focus is not on a demonstration and validation of the entire workflow, but on the technical requirements involved. The scenario is explained as follows: a social housing organization (the client) demolishes a single residential building (Building A) and seeks to reuse selected windows in a new replacement building (Building B). The operational challenge is threefold: first, the demolisher must capture reliable "as-removed" window information during selective strip-out; second, the architect designing the new building must be able to perform matchmaking using data that is integrated into the building information model rather than dispersed across unstructured reports. And lastly the asset owners should be able to provide demolishers with demountable window sets and keep track of the reused windows in the new building.

Demolition phase: demonstrated in updating

Before demolition begins, windows in Building A are represented as identifiable component instances in the client's digital building model. During selective removal, the demolisher updates the model by appending statements to their layer of information, per window instance, the data required for downstream reuse decision-making. In addition, the demolisher records a condition snapshot and handling-related notes relevant to transport and onward processing, reflecting the practical reality that removal method, contamination, and damage significantly affect reuse feasibility. The demolisher's contribution is framed as a specialized dataset: it describes the state of each window at the point of removal and the conditions under which it can plausibly be reused, without requiring the demolisher to author design information for the new building. This separation of responsibilities is essential for scalability as demolishers can provide factual "as-removed" information, while architects can interpret that information within the constraints of the new project.

Design phase: demonstrated in matchmaking

The architect uses the updated window information (stored in a federated set of multiple layers) as a stock list for reuse. Matchmaking is performed by comparing recovered-window descriptors to the dimensional and performance requirements of the new building design. This produces a mapping between the new building's window openings and specific recovered window instances, which supports both procurement planning (reducing the number of new windows that must be purchased) and design development (ensuring that the design accommodates the constraints of the recovered stock, where reuse is intended).

Installation and maintenance phase: demonstrated partially in querying

Once installed in the new building, each reused window is recorded as an "as-installed" component instance, including its installed location, installation date, and any relevant references to associated product information. This paper only demonstrates the querying for windows without linking to the other digital repositories. Where a Digital Product Passport exists for a given window product (for example, if the window has been remanufactured and supplied under a product regime that provides a DPP), the installed component record includes a pointer to that passport. This creates product-level continuity: the same product information can be accessed later for maintenance, renovation, or audit, and updates can be managed by the responsible actors in the passport system rather than being embedded as static attachments. The building-level traceability layer is also established through a building repository approach consistent with the logbook pathway. The owner maintains a record of installed reused windows as part of the building's digital record, enabling future renovation planning. This links product-level traceability and building-level traceability so that reused windows remain discoverable across the building lifecycle, rather than becoming "lost".

Prototype development

Here we illustrate how IFC5 data models and serialization technologies relate to the circular construction context described above. Note that the demonstration is meant to show and validate features on a conceptual level. Due to the limited number of constructs, uniformity of structure and usage of JSON, for which most modern database management systems have native support, this structure is easily serialized to a relational database. For this experiment the following table layout (Figure 2) was chosen. Important to note are the indirections between Project – Layer – Primitive, so that a Project has multiple layers (this is the multi-disciplinary aspect). A primitive has a single attribute blob as a JSON field and inherits children edge pointing to UUIDs. Also note that Primitive UUIDs are not unique (in the IFCx structure they are called "path"), so that multiple definitions pointing to the same path can be retained. This also enables the multi-disciplinary nature so that objects are open to be extended with additional data.

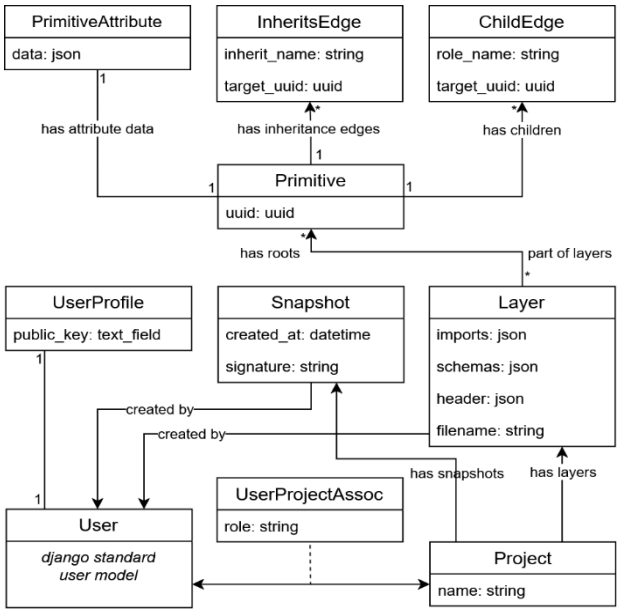


Figure 2: Data model used in the implementation.

Data ingestion follows an append-only nature, meaning that all data is immutable in the database, newer opinions (attributes at the same path) can override pre-existing ones and will come in with a higher strength, but it is possible to go back to a previous state because everything is time-stamped. Snapshot allows stakeholders to take a state of the project and create a digital signature. The database model is generated from python code and implemented in the Django web framework. The queries shown later are actual logged queries from the implemented prototype.

To investigate the implications of circularity tracing and reuse, we have experimented with only storing the layer_id in case of a root primitive (a primitive with no incoming child or inheritance edges). The root primitive is typically the IfcProject entity. Therefore, subsequent children can freely be associated with multiple layers (and therefore physical projects) to see how data can flow between projects in accordance with the physical products when one building is decommissioned and windows are moved to the next building. Listings below illustrate the different purposes enabled by this data model, i.e. querying, updating, and matchmaking.

Querying

As mentioned above, there have been initiatives of query languages on top of reserializations or conversions of IFC models but widely implemented IFC query languages to date are not in use. This section illustrates that by means of the simpler layout of IFC5 data, reliance on industry standard data representation techniques such as JSON and mappings to standard database management solutions, querying techniques are available on top of IFC5 content.

Listing 1 describes Postgres-specific query using JSON-object-subset search. As such we can locate all

PrimitiveAttribute instances of class IfcWindow. Note how the association of a class (or building element type) is like any other data attribute field an arbitrary (but schema-enforced) key-value pair. This is something already observed in van Berlo et al. (2021) where it was called “late bound” type definitions, since distinct building element classes do not relate to specific behavior in software, and mappings from native models to IFC classes need to be fully configurable by the end-user.

```
SELECT *
FROM primitive_attribute INNER JOIN primitive
ON (primitive_attribute.primitive_id = primitive.id)
WHERE primitive_attribute.data @> '{"bsi::ifc::class":
{"code": "IfcWindow"}}';
```

Listing 1. SQL query on top of the data model to demonstrate indexed JSON-object-subset search in contemporary database.

In addition, we want to find (recursively) all elements inheriting from the Primitive with this attribute (but not elements having it as a child) because the class association might be assigned at a re-used archetype (or a Family as sometimes called in BIM authoring tools). Everything that inherits from these primitives is also an IfcWindow (unless the attribute is locally overwritten there). The inheritance aspect can be solved transitively by applying a recursive Common Table Expression (CTE) in Postgres with the results from the previous query, following inwards edges (Listing 2).

```
WITH RECURSIVE edge AS (
  SELECT primitive_id, target_uuid FROM inherits_edge
),
walk AS (
  SELECT
    p.id, p.uuid,
    ARRAY[p.id]::bigint[] AS visited_ids,
    0::int AS depth
  FROM primitive p
  WHERE p.uuid = ANY(ARRAY['2c2d549f-f9fe-4e22-8590-
    562fda81a690'::uuid,'592504dc-469a-44d6-
    9ae8-c801b591679b'::uuid,'25503984-6605-
    43a1-8597-
    eae657ff5bea'::uuid]::uuid[])
  UNION ALL
  SELECT
    p2.id, p2.uuid,
    w.visited_ids || p2.id,
    w.depth + 1
  FROM walk w
  JOIN edge e ON e.target_uuid = w.uuid
  JOIN primitive p2 ON p2.id = e.primitive_id
  WHERE NOT (p2.id = ANY(w.visited_ids))
),
uniq AS (
  SELECT DISTINCT id FROM walk

)
SELECT p.*
FROM primitive p
JOIN uniq u ON u.id = p.id ORDER BY p.created_at, p.id;
```

Listing 2. SQL query using a recursive CTE to find inheriting primitives. This is a verbatim copy of the query. The array of uuids present in the query comes from the Object-Relational Mapping;program code to interact with the database schemas)

This also highlights how modern database solutions can be used to efficiently apply the inheritance and children relationships to the dataset. In the IFC5 project team it has been discussed whether data inheritance is necessary as it has negative implications on data locally and complexity. The experiment here shows that the complexity and performance concerns are not insurmountable.

The query above is for finding primitives that inherit from other primitives. Similarly, a query can be used to do a

forward traversal over inheritance and children edges. This query is used when returning the data to the user to return a full model subset. This query is equivalent, just the edges are formed as the union of inherits_edge and children_edge, and the direction is reversed (concerns the ON clauses in edge and primitive).

This full model subset is valid IFCx again that can be visualized in one of the supported tools (Figure 3).



Figure 3: Query demonstration. Note that the result presented is not a visual selection of relevant models: with the IFCx serialization structure, a query result is a complete and valid model subset that can be used and recombined, for the full set see (IFC5 Development, 2026).

This highlights an important distinction between IFC5 alpha and earlier editions of the spec, which had much more dense graphs requiring a lot of other contexts, such as IfcGeometricRepresentationContext and IfcProject with IfcUnitAssignment. These combinations made extracting and concatenating content much more complicated. In IFC5 because of the flat pre-composition data structure, the literal serialization of the primitives selected provides us with all the relevant contexts.

Updating

When the demolisher receives the window query model, they can go on-site to add inspection information. Using a modular ontology, they create a new layer with only this information. The usage of paths in IFCx, makes sure that information lines up with the pre-existing content. This is to illustrate the applicability of updating information by a specific stakeholder in the process (Listing 3).

```
{
  "path": "23ea7c34-6e5a-47ee-b292-ebedfd585438",
  "attributes": {
    "decommissioningStatus":"removed",
    "currentState": "stored",
    "conditionGrade": "B",
    "damage": "minor_scratch",
    "contamination": "none_detected",
    "disassemblyType": "reversible",
    "reusePotential": "direct_reuse",
    "recommendedAction":
      "clean_and_reinstall"
  }
}
```

Listing 3. IFC5 primitive that is ingested by the demolisher showing the modularity of IFC5 schemas and the collaborative data authoring by composition of data on paths

Matchmaking

At this stage, a new project is created. The architect uses this data model with the dimensions of the windows enriched with the inspection information to select suitable windows to be reused. We illustrate this with just a simple wall with the same void. To this experiment, a wall is created in the open-source BIM authoring tool Bonsai with an opening that matches the to-be-reused window, saved as IFC4 and converted to IFC5. The transfer of the window element from the old model to the new model is manifested by means of creating a new primitive in the old model that inherits from the window to be reused. In this case, a placement transformation is necessary because the demo model used does not have proper relative positioning (Listing 4).

```
{
  "path": "33ea7c34-6e5a-47ee-b292-ebedfd585438",
  "inherits": {
    "circularReuse":  "23ea7c34-6e5a-47ee-b292-
    ebedfd585438"
  },
  "attributes": { "usd::xformop": {
      "transform": [[0.0, -1.0, 0.0, 0.0],
            [1.0, 0.0, 0.0, 0.0],
            [0.0, 0.0, 1.0, 0.0],
            [4.98, 12.97, -0.71, 1.0]]
}}},
```

Listing 4. Primitive that is created in the new model that inherits from and repositions the reused window instance (The term "circularReuse" is a human-readable label to reference the inheritance edge; it has no computer-interpretable semantics.)

Now we place this removed window from another building on the newly created wall (Listing 5, Figure 4).

```
{
  "path": "46b17e20-783e-4bf5-ac74-714b859be0b3",
  "children": {
    "Body": "a03ebe4d-2e40-524e-9144-c03ebec7100e",
    "Window": "33ea7c34-6e5a-47ee-b292-ebedfd585438"
  },
  "inherits": {
    "material":  "8a187a90-3dcf-58cc-b3a6-51a9a407c55a"
  }
}
```

Listing 5: Newly created wall primitive with additional child inserted

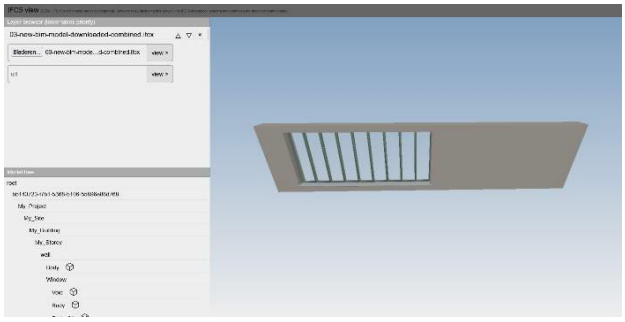


Figure 4: Result of reused window model when recursive query

traverses also into the reused window by means of path references

Discussion

The study can be further improved by testing other user requirements for circular construction, such as updating logistics constraints by transporting parties and warehouse facilities for storing the products and integrating other information sources like IoT for smart homes and linking with digital building logbooks or renovation passports.

It is worth noting that this research does not aim for a complete circular workflow but the technical possibilities of using IFC5 to enable modular, participatory and traceable data exchange. This is used to support circular use cases like querying by the clients for their assets, updating information from the perspective of the demolishers and the matchmaking by the architects using reclaimed windows for a new project. These are based on a simplified circular scenario. Future research should study from the user perspective to what extent this could be adopted and implemented in the inter and intra organizational settings, which will impose a greater need on transport and APIs as this research focused predominantly on the data-model.

In addition to that, the emphasis here has been mostly on the serialization-level aspects of the IFCx data encoding used by IFC5 and how industry-standard solutions such as the Postgres database management system with its support for JSON fields in tables, indexed JSON object subset search and recursive expressions enables a good amount of the core set of functionalities that are needed to deliver a platform to host such functionality. It is expected that other tooling around JSON is equally valid and useful, such as DuckDB or command line tools such as jq. In this implementation there is nothing special about Postgres and that is exactly the advantage of the IFCx JSON-based serialization mechanism.

In IFCx, provenance is tracked by means of the containing layer, but there is still ongoing debate on the balance between primary file-based workflows and more transient datasets created as query results and how this impacts the need for per-component identity or provenance records. The focus on data model in this research left aside most of the queries of transport: how do the various layers get into one (federated) data environment? In this research, the requirements for that have been simplified by simulating a case where the owner of the windows did not change. These are worthwhile, but also mostly orthogonal, questions for future research.

Conclusions

Emerging regulations for circular construction have raised requirements to exchange information among multiple actors in a modular and traceable approach. This paper has described pertinent IFC4 limitations in encoding information for the purpose of circularity and how IFC5 can help different actors in the value chain to deal with the emerging regulations for circular strategies like reuse of built assets using a simple context. The used dataset can be found in the following link: https://zenodo.org/records/18505849.

Currently, re-allocating the data representation of an IFC element to a new project in a fully immutable append-only and traceable data model would have been impossible to accomplish on IFC4 datasets due to the contextual references required to context and objectified relationships and the lack of data-level composition mechanisms.

With the new capability and data model of IFC5, the circular transition can be operationalized and upscaled by the supply chain partners participating in data querying, updating and matchmaking for reusing of assets.

Acknowledgments

Part of this work is funded by Het Nationaal Groeifonds Toekomstbestendige leefomgeving.

References

BAMB. (2020). Materials passports. https://www.bamb2020.eu/topics/materials-passports/

BuildingSMART International. (2026). Industry Foundation Classes (IFC). https://www.buildingsmart.org/standards/bsi-standards/industry-foundation-classes

BuildingSMART International. (2020) Ten principles of a future IFC. https://github.com/buildingSMART/NextGen-IFC/wiki/Ten-principles-for-a-future-IFC

BuildingSMART International. (2020) Technical roadmap.https://www.buildingsmart.org/about/technical-roadmap/

BuildingSmart International (2026) IFC5 development. https://github.com/buildingSMART/IFC5-development/

Byers, B. S., Raghu, D., Olumo, A., De Wolf, C., & Haas,

C. (2024). From research to practice: A review on technologies for addressing the information gap for building material reuse in circular construction. Sustainable Production and Consumption, 45, 177–191. https://doi.org/10.1016/j.spc.2023.12.017

Çetin, S., Raghu, D., Honic, M., Straub, A., & Gruis, V. (2023). Data requirements and availabilities for material passports: A digitally enabled framework for improving the circularity of existing buildings. Sustainable Production and Consumption, 40, 422–437. https://doi.org/10.1016/j.spc.2023.07.011

Çetin, S., Raghu, D., Honic, M., Straub, A., & Gruis, V. (2023). Material passports for social housing stock: A tool. In S. Çetin (Ed.), Towards a circular building industry through digitalisation (pp. 151–186). TU Delft OPEN Publishing. https://pure.tudelft.nl/ws/portalfiles/portal/237912925/Material_passports_for_social_housing_stock.pdf

Ding, L., Wang, T., & Chan, P. W. (2023). Forward and reverse logistics for circular economy in construction: A systematic literature review. Journal of Cleaner Production, 388, 135981. https://doi.org/10.1016/j.jclepro.2023.135981

Durmisevic, E. (2018, November). The Super Circular Estate project: Journal No. 1. Urban Innovative Actions.https://uia.urban-initiative.eu/sites/default/files/2019-02/Kerkrade_Journal.pdf

European Commission. (2025). Renovation passports (Article 12, Annex VIII): Annex 4 to the Commission notice providing guidance on new or substantially modified provisions of the recast Energy Performance of Buildings Directive (EU) 2024/1275 (C(2025) 4132 final).https://energy.ec.europa.eu/document/download/0a4a6 d59-58c0-4e80-a58a a6ee944573ab_en?filename=Renovation+passport+%28Article+12%2C+Annex+VIII%29+-+annex+4.pdf

European Parliament and Council of the European Union. (2024, November 27). Regulation (EU) 2024/3110 laying down harmonised rules for the marketing of construction products and repealing Regulation (EU) No 305/2011 (Text with EEA relevance). Official Journal of the EuropeanUnion.https://www.eota.eu/sites/default/files/files/OJ_L_202 403110_EN_TXT.pdf

European Commission. (2018, September 14). Guidelines for the waste audits before demolition and renovation works of buildings. DocsRoom. https://ec.europa.eu/docsroom/documents/31521/

Hemubo. (2022, December). De Bosleeuw: Circulaire kozijnen.https://www.hemubo.nl/wp-content/uploads/2022/12/ARTIKEL_Bosleeuw-circulaire-kozijnen-DEF.pdf

International Organization for Standardization. (2025). ISO 23387:2025 Building information modelling (BIM)—Data templates for objects used in the life cycle of assets (2nd ed.) https://www.iso.org/standard/85391.html

Kebede, R., Moscati, A., Tan, H., & Johansson, P. (2024). A modular ontology modeling approach to developing digital product passports to promote circular economy in the built environment. Sustainable Production and Consumption, 48, 248–268. https://doi.org/10.1016/j.spc.2024.05.007

Krijnen, T., & Beetz, J. (2018). A SPARQL query engine for binary-formatted IFC building models. Automation in Construction, 95, 46-63.

Ding, L., Wang, T., & Chan, P. W. (2023). Forward and reverse logistics for circular economy in construction: A systematic literature review. Journal of Cleaner Production, 388, Article 135981. https://doi.org/10.1016/j.jclepro.2023.135981

IFC5-Development (2026). BuildingSmart Internationa. https://github.com/buildingSMART/IFC5-development/tree/main/examples/ACCA%20Building

NGF (2025). National Growth Funds for the future proof living environment. https://www.nationaalgroeifonds.nl/overzicht-lopende-projecten/thema-energie-en-duurzame-ontwikkeling/toekomstbestendige-leefomgeving

Pauwels, P., Krijnen, T., Terkaj, W., & Beetz, J. (2017). Enhancing the ifcOWL ontology with an alternative representation for geometric data. Automation in Construction, 80, 77-94.

Tomczak, A., Benghi, C., van Berlo, L., & Hjelseth, E. (2024). Requiring circularity data in BIM with Information Delivery Specification.

Tomczak, A., v Berlo, L., Krijnen, T., Borrmann, A., & Bolpagni, M. (2022, November). A review of methods to specify information requirements in digital construction projects. In IOP Conference Series: Earth and Environmental Science (Vol. 1101, No. 9, p. 092024). IOP Publishing.

USD (2026) Universal Scene Description. https://openusd.org/release/index.html

UK BIM Framework. (2021). Guidance Part 3:Operational phase of the asset life-cycle (Edition 2).https://www.ukbimframework.org/wp-content/uploads/2021/02/Guidance-Part-3_Operational-phase-of-the-asset-life-cycle_Edition-2.pdf

van Berlo, L., Krijnen, T., Tauscher, H., Liebich, T., Van Kranenburg, A., & Paasiala, P. (2021, October). Future of the industry foundation classes: towards IFC 5. In Proc. of the 38th International Conference of CIB (Vol. 78, pp. 11-15).

Wolf, N., Esser, S., & Borrmann, A. (2025). IFC2GRAPHQL: Schema Mapping and Automated Api Construction From IFC to Graphql.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

PROPOSAL FOR A MODERN FOUNDATION FOR A DATA-DRIVEN BUILT ENVIRONMENT

Léon van Berlo[1], Thomas Krijnen[2], Tom van Diggelen[3], Artur Tomczak[4], Evandro Alfieri[4], David de Koning[5], Greg Schleusner[6], Helga Tauscher[7] and Dennis Shelden[8]
[1]Boundless BIM Council, Breda, Netherlands; [2]AECGeeks, Eindhoven, Netherlands; [3]KUBUS, Eindhoven, Netherlands; [4]buildingSMART International, London, UK; [5]Oasys (Arup Canada), Montreal, Canada; [6]HOK, New York, USA; [7]HTW Berlin, Berlin, Germany; [8]RPI, Troy, US

Abstract

The Industry Foundation Classes (IFC) standard underpins BIM interoperability, but its STEP/EXPRESS stack is ill-suited to cloud-native, transactional workflows. Building on buildingSMART community-driven requirements and proposals, review of alternative solutions, and iterative prototyping, we propose the modernisation of IFC. The work includes a proposal for a data schema, JSON serialisation, and exchange mechanisms, while preserving IFC's role as a semantically rich neutral data model. Innovation includes a compositional, layered federation model allowing incremental updates, multi-author collaboration, versioning, and fine-grained, traceable ownership. We compare the proposal to existing standard and discuss the potential, trade-offs and implications.

Introduction

IFC originated in the mid-1990s as a vendor-neutral product model for exchanging building information, initiated by the International Alliance for Interoperability (IAI) and later stewarded by buildingSMART as an open standard for BIM (Building Information Modelling). Through iterative releases in the 1990s–2000s, IFC gravitated towards its current scope. The common IFC use cases span design and construction coordination, design data transfer, asset management, building permitting and documentation archiving. The IFC in its current form has matured into a widely adopted format, implemented in over 400 software tools (buildingSMART Int., 2026b), enabling structured exchange of building and infrastructure data worldwide.

The IFC specification has been extended with each subsequent release, with addition of more entities and properties to describe the broad landscape of architecture, engineering, construction and operations (AECO) concepts. For example, the latest release – the IFC 4.3.2 – introduced classes and properties to more accurately describe roads and rails (buildingSMART Int., 2024). Some updates unified properties, that before could differ in definition depending on the object described.

Aiming to support plurality of industry use cases, the IFC standard grew in complexity, allowing many alternative ways for modelling some aspects. For example, alphanumeric data can be attached to objects as attributes or properties; objects are classified using IFC entities, but also predefined types and classification references; there are forty ways to assign material to an object (Tomczak et al., 2024) and over 170 kinds of geometric primitives - subtypes of *IfcRepresentationItem* and *IfcProfileDef* (buildingSMART, 2024). Noardo et al. (2019) investigated geometric errors and implementation inconsistency. This kind of implementation inconsistency is incredibly detrimental to the necessary trust in a digital language of the construction sector. The combinatorial nature of the geometry definitions results in implementations supporting different subsets of the standard which further inhibits successful interoperability. Lastly, the complexity associated with the immense breadth of geometric paradigms can be seen as a barrier to adoption.

On a specification level, the IFC follows strict compatibility dogmas, so reductions in scope are rarely realised, which results in a gap between specification content and actual usage in practice. While cloud-based data exchange and incremental updates are not new in contemporary BIM tools, the IFC today is primarily used for file-based exchange, where a full dataset needs to be transferred, even for a simple update of one property value. Furthermore, the IFC preserves information about a single author and authoring application per file, limiting traceability.

Related works

An influential attempt at modernizing IFC has been captured by van Berlo et al. (2021). Core ideas there were a re-evaluation of the amount and lack of consistency between relationship types, and introduction of the concept of 'late binding' to IFC. Over time, the community has come to realise that the taxonomy of building elements does not need to map to distinct classes in computer programming languages, and are therefore better kept out of the main schema in a secondary "late bound" layer, more akin to the data format in which property sets are described in the specification. This same research also indicates the desire to move towards 'object-based incremental updates' but does not yet give concrete technical steps towards realizing this, but does list predictable, normalised tree structures in the graph as an enabler. The Technical Roadmap (buildingSMART, 2020) sets the direction and calls to reconsider the technological stack of openBIM standards.

The STEP Physical File (SPF) remains the primary and most widely used serialisation format for IFC, although it is not the only officially supported option. Alternative serialisations include ifcXML and the compressed ifcZIP

formats (buildingSMART, 2022). JSON-based representations, aligned with contemporary web data exchange practices, have been proposed as alternatives to the traditional IFC stack by Afsari et al. (2017) and Brouwer et al. (2020). Despite these efforts, SPF has retained dominance due to its tight coupling with EXPRESS schemas. It enforces strict schema rigidity and ordered, unnamed attribute sequences, allowing SPF to represent complex graph structures concisely. This effective SPF structure has no real equivalent in modern languages and thus limits the industrial adoption of alternative serialisation formats.

Several efforts have explored the application of semantic web technologies for BIM. As part of the IfcOWL project, the IFC has been mapped to OWL, enabling RDF-based representations that can be queried using SPARQL and linked with other web resources (Pauwels et al. 2017). IfcOWL also introduced alternative geometric encodings based on Well-Known Text (WKT) to reduce verbosity and support GeoSPARQL. However, the direct mapping of EXPRESS to OWL introduces many indirections, which negatively affect query performance and reasoning expressiveness. In response, lighter-weight ontologies such as BOT have been proposed to support higher-level building topology without replicating the full IFC schema (Bonduel et al., 2018).

Indirections are the intermediary constructs and steps that must be traversed in order to relate a value to an entity. IFC is known for its many indirections, for example to get from a material to a colour: "*IfcMaterial←IfcMaterialDefinitionRepresentation→IfcStyledRepresentation→IfcStyledItem→IfcStyleAssignmentSelect→IfcPresentationStyleAssignment→IfcPresentationStyleSelect→IfcSurfaceStyle→IfcSurfaceStyleElementSelect→IfcSurfaceStyleShading→IfcColourRgb".* These indirections come at a cost in terms of: (a) performance; data model size; query joins (b) cognitive overhead in understanding (c) lack of ownership clarity when reusing data. In RDF additional indirection occurs by representing ordered lists as binary trees of *<first, rest>*-pairs.

Property graph approaches have been explored in systems such as IFCWebServer (Ismail et al., 2017) and in multiple Neo4j-based mappings (Zhu et al., 2023). These approaches leverage graph traversal and analytics capabilities but typically require bespoke schema mappings and tooling.

Relational database storage of IFC data has also been extensively studied (Li et al., 2016; Solihin et al., 2017), with roots tracing back to early systems such as BLIS and Sable within the IMSrv platform (Beetz et al., 2010). However, mapping EXPRESS constructs - including deep inheritance hierarchies, complex aggregations, inverse attributes, and SELECT types - to relational schemas introduces significant complexity. This results either in large numbers of tables or substantial application-layer logic, a challenge commonly referred to as the object-relational impedance mismatch.

Key-value stores offer a more flexible alternative, as they impose minimal structural constraints and delegate serialisation entirely to the application. BIMserver.org is a prominent example of this approach (Beetz et al., 2010). While efficient and well-suited to complex object graphs, key-value storage is typically binary, non-portable, and not human-readable, limiting its suitability for archival or interoperable exchange scenarios.

Binary formats have also been explored for handling large explicit datasets, particularly meshes and point clouds, using HDF5 (Krijnen & Beetz, 2020). Although technically effective for specific high-volume data scenarios, these approaches address a narrow set of use cases and introduce dependencies on additional complex standards, despite the existence of an official EXPRESS-to-HDF5 mapping in ISO 10303-26. They also do not align well with requirements for incremental exchange or web-based workflows.

More recently, Universal Scene Description (USD), originally developed by Pixar, has gained attention due to its strong support for large-scale collaborative asset composition. Governed by the Alliance for OpenUSD and undergoing standardisation efforts, USD emphasises explicit data, layered composition, and conflict resolution through ordered 'opinions'. Its schema-agnostic composition mechanisms resemble certain IFC concepts, such as TypeObject–Occurrence relationships but operate at the serialisation level rather than within a formal schema. These characteristics make USD conceptually attractive for AEC collaboration, despite its origins outside the domain.

Finally, lightweight mesh-based formats such as glTF support many contemporary visualisation-centric use cases, including coordination and clash detection, while enabling integration with a broad ecosystem of off-the-shelf tools. However, they largely omit semantic richness.

Borkowski (2026) argues that mesh-based formats such as glTF and USD, combined with structured non-geometric data stored for example in JSON, CSV, or SQL, represent a plausible alternative to kernel-dependent IFC. He characterises traditional file-based BIM workflows as slower and more cumbersome to integrate than granular data systems, and suggests that BIM practice may need to shift from a software-centric toward more data-centric.

The breadth of explored serialisation and storage approaches highlights fundamental limitations of the EXPRESS-based ecosystem in supporting contemporary construction industry use cases. The expressive power of EXPRESS, while beneficial for formal schema definition, complicates mapping to modern, web-oriented data frameworks. The requirements driving these investigations are discussed in the following section.

Method

Requirements

The requirements documented here were shaped over years of involvement in the buildingSMART community, and have been refined through in-person meetings and presentations.

1. Distributed ownership and authorship traceability

The construction sector is a multi-disciplinary field with strict responsibility boundaries and liability. It should be possible to combine datasets while strictly retaining the provenance of where information comes from, without inhibiting establishing rich semantic links between concepts from distinct disciplines. To realise accountability, the provenance of information needs to be traceable. IFC to date has an *IfcOwnerHistory* entity, but that tracks only the meta-data of the latest rendition of the object. There is no way to retain deleted information, or a longer provenance trail.

2. Incremental updates

Building models can realistically amount up to gigabytes of data in contemporary engineering projects. It should be possible to provide only the changes to earlier exchanged model.

3. Moderate complexity

Many of the contemporary use cases in which IFC is employed are not that complex, such as visualisation and coordination. There is also an asymmetry in complexity: IFC is often seen as a compromise between many different options of representing elements. Therefore, for an authoring tool it is sufficient to only support one single option and write that out. Conversely, a downstream application, such as a viewer or manufacturing platform must harmonize all different options coming from different tools. This asymmetry needs to be reduced to open BIM data to a wide variety of neighbouring domains and facilitate reuse of information downstream. This could for example mean moving the responsibility of interpreting complex procedural geometry to the producing side, where it is already rendered to a triangle mesh for display anyway.

4. Plurality of representation

Each physical and spatial object can be represented with geometry, but the accuracy, complexity and kind of geometry might differ, depending on the needs. A building at an urban scale might be simplified to a bounding box cube, while at close-up could be highly detailed. An example wall can be modelled as an extruded set of layers with procedurally cut-through openings, or as a triangulated mesh resulted from computation, to speed-up rendering speed. Same wall could be simplified to a guide curve and thickness, or have a point-cloud representation to compare as-built model with reality.

5. Performant

In order to increase the number of iterative design and engineering cycles the import/export/analysis times need to be reduced compared to contemporary IFC workflows. This also means less indirection (intermediary steps) between data, that are problematic in current querying workflows and may obscure ownership semantics.

6. Cloud-friendly, web-native

Common Data Environments (CDE) have become more popular, but true distributed and collaborative workflows are impeded by the monolithic and idiosyncratic nature of SPF-based exchanges.

7. Granular data access

Ability to expose or load only a subset of information. This could be either for performance or confidentiality reasons. For example, an engineer might want to see architect's walls, but not their finishing.

8. Extensibility

To support rapid innovation and multi-disciplinary challenges, extensibility should be possible on a level beyond adding additional key-value pairs as user-defined property sets. Examples include efficient encoding of compressed point clouds (Krijnen & Beetz 2017) or the impediments in realizing modelling circularity related element connectivity (Tomczak et al. 2024). Currently, in EXPRESS-based schemas, there is no realistic way to provide this kind of extensibility as it would require publishing a novel schema version with a unique schema identifier which affects even the software that would ignore the extension because it is outside of their intended domain.

9. More predictable and atomic data patterns.

Currently unclear whether something is a primitive with an identity or an attribute with a value. For example, *IfcMaterial* does not have a Globally Unique Identifier (GUID), and is therefore only a transitive object. Relationships have GUIDs, but tools cannot typically maintain consistency between export cycles.

10. Bridge BIM with GIS, IoT, Digital Twins etc.

Building information is not limited to design data, but fits into larger geospatial context commonly represented with GIS, and can serve during asset operation phase. In case of continuously updated models, often fed with live sensor data, BIM can serve a role of a Digital Twin. Therefore the data exchange standard of the future should not limit to one, but rather allow multiple schemas and representations.

11. Independent and easy to update modules

Up until now, the IFC standard has functioned as a monolithic data model published as one large standard document with almost 500 pages (ISO 16739-1:2024). While this ensured a comprehensive coverage of the built environment, it also resulted in a massive specification that is difficult to maintain and cumbersome for software vendors to implement fully. As the scope of IFC expanded to include new domains such as Rail, Road, and Tunnel, the time required between release cycles increased, often taking years to ratify updates.

Proposed Solution

We propose a novel data exchange solution, that is aimed to resolve common IFC issues, and address the requirements (1-11). The first building-block module of the proposed modern IFC is the schema, enabling further definition of an API module, domain knowledge/semantics modules and technical extensions.

A dataset called the 'Hello Wall' will be used to demonstrate functionalities of the proposed modern IFC, and is available together with other examples, in the *IFC 5 Development* GitHub repository (buildingSMART, 2026a). The dataset is presented in figures 3 and 4.

Layers and independent data storage

We advocate a fundamental dissociation between the business logic data model and its physical storage. A dataset should not be required to be a single monolithic file; instead, it operates as a compositional, layered federation of data. The *layers* mechanism in our proposal addresses requirements for distributed ownership (1) and incremental updates (2) by dissociating the data from physical storage, allowing a single dataset (or 'stream') to be composed of one or multiple federated chunks resolved at runtime.

This architecture enables non-destructive collaboration, where stakeholders can overlay their own data attributes onto a base model without altering the original data, thereby strictly preserving authorship and liability boundaries. Furthermore, this same technical framework handles incremental updates by treating changes as lightweight *delta* streams containing only the proposed modifications, eliminating the need to re-transmit massive monolithic files when only specific data points have evolved.

Composition and paths

For software to make meaning out of disconnected layers or streams, the composition needs to take place. The composition relies on consistent path identifiers like GUIDs, e.g. *'e3035b71-bd9f-4cdc-86fd-b56e2f4605b6'*, or a tree path, e.g. *'project/site/ building/wall_id'*. Because every piece of data is tied to a specific path, information about a single object does not need to reside in the same physical location.

For example, the base geometry of a wall might exist in a *geometry.ifcx* file, while its material properties reside in *materials.ifcx*, and a fire engineer's specific requirements in a yet separate *fire_safety.ifcx* file. When a software tool loads the data by reading these files, it recognises that they all refer to the same object path. It then overlays or *composes* these attributes to form the complete object.

Attributes

The most trivial data to attach to a path is an attribute, which links some set of data to a specific path. Attributes have a name and value and follow a pre-specified schema. Attributes in the proposed IFC are JSON key-value pair objects. For every key a schema needs to be defined either locally in the file or imported. Attribute keys are also namespaced to provide context. Numeric attribute values derive their unit from the quantity kind associated with the schemas. There is no distinction alike IFC4's Attributes and Properties, both are exchanged and validated in the same fashion. Similar situation is presented in figure 1.

```
(…)"data":[
{"path": "c8b…c48",
"attributes": {"name": "Something"}},
{"path": "c8b…c48",
"attributes": {"my::prop": 3}},
```

```
c8b…c48:
- name = "Something"
- my::prop = 3
```

Figure 1: Raw data nodes (top) and resultant composed object with attributes (bottom)

Incremental updates and distributed ownership

The proposed architecture directly supports modern workflows that were difficult to manage with previous IFC versions. Instead of re-sending a massive file when a single property changes, a user can exchange a small *delta* set containing only the changed attributes for specific object paths.

Different disciplines can maintain their own data – their *opinions* – without altering the original author's data. As demonstrated on figure 2, one user can add a '*FireRating: R60*' property in their own layer. If this conflicts with an existing *'FireRating'* of value *'R30'*, the composition engine warns user about the conflict. All while ensuring that ownership and provenance of data are preserved at fine-grained level, allowing to see exactly who authored each information and when.

```
(…)"data":[
{"path":"c8b…c48",
   "attributes":{"FireRating": "R30"}},
{"path":"c8b…c48",
   "attributes":{"FireRating": "R60"}},
```

```
c8b…c48:
- FireRating = ~~"R30"~~ "R60"
```

Figure 2: Demonstration of conflict during composition – the nodes (top) and resultant compiled object (bottom)

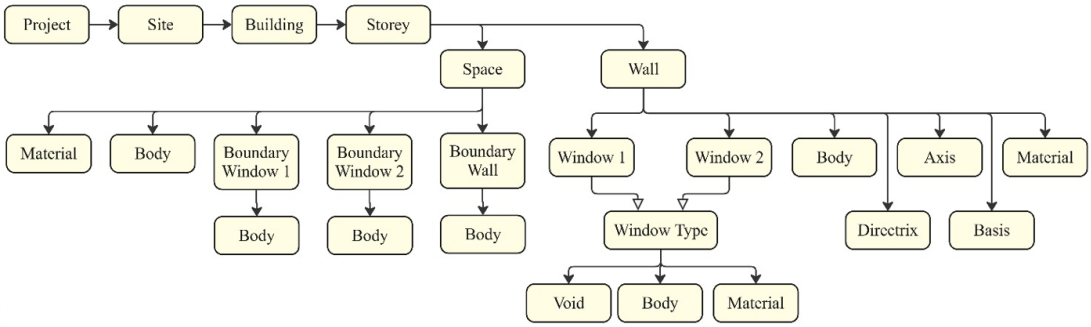


Figure 3: Diagram of the entities in the 'Hello Wall' file

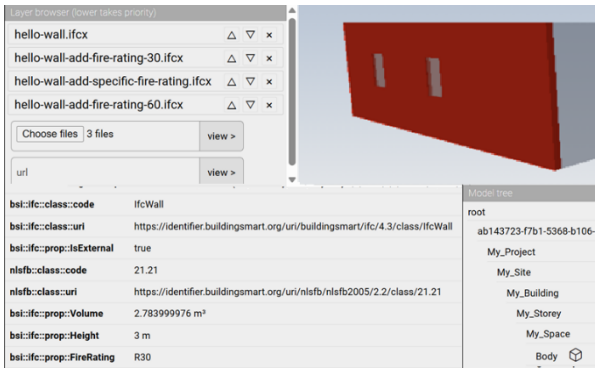


Figure 4: Layered 'Hello Wall' model composed and displayed in a demo IFCX viewer.

Children

Spatial decomposition relationships and local placement known from IFC 4.3.2 are represented in the proposal with *children*. Merging into one concept improves consistency. This, however, only applies to hierarchical IFC relationships like *Decomposes, Nests, AdheresTo, Voids, FillsVoids*. System-like graph relationships will be realised by means of attributes with path references.

Inheritance

We propose a generic, and multi-level information reuse mechanism to avoid complexity of mapped items and textual documentation on how to implement relationship between them. An example application is presented in figure 5, where two windows inherit from the same type defining frame and glazing.

```
(…)"data":[
{"path":"c8b…c48","children":
   {"window1":"e48…834",
   "window2":"caa…15d"}},
{"path":"e48…834","inherits":
   {"windowType":"189…5c3"}},
{"path":"caa…15d","inherits":
   {"windowType":"189…5c3"}},
{"path":"189…5c3","children":
   {"frame":"651…f2b",
   "glazing":"6ed…953"}}]
```

```
c8b…c48
   c8b…c48/window1
      c8b…c48/window1/frame
      c8b…c48/window1/glazing
   c8b…c48/window2
      c8b…c48/window2/frame
      c8b…c48/window2/glazing
```

Figure 5: Demonstration of the children and inheritance – the nodes (top) and compiled hierarchy tree (bottom)

Note that this is not the multiple inheritance that has gotten a negative connotation in certain programming languages with things like the 'diamond problems' but rather a formalised information reusage mechanism.

Modularisation of the schema

To address Requirement 11, our solution introduces a modular setup that streamlines ISO standardisation by decoupling a stable technical core from more dynamic, domain-specific semantics and extensions. Organised into distinct ISO Parts, this strategy replaces the previous monolithic structure to allow for independent versioning and faster industry adoption. This enables software to implement a reliable core quickly while allowing domain experts to expand semantic coverage without disrupting the underlying ecosystem.

Serialisation

The presented solution is serialised entirely in JSON, making it human-readable and compatible with modern web technologies. However, other serialisation forms are not off the table, JSON is just a readable and universally understood logical choice at this phase of the development and facilitates easy dissemination of the ideas around composition and file format mechanisms without being obfuscated by complex serialisation mechanisms. Binary format was discussed and can become a potential solution.

Schema definition

Our solution is defined using TypeSpec, a highly productive API description language that brings a concise, developer-friendly syntax (Microsoft, 2026). By leveraging TypeSpec, the schema benefits from improved readability and modularity, allowing for the automated generation of multi-protocol representations such as JSON Schema, OpenAPI, and GraphQL from a single source of truth. By utilising TypeSpec as the authoritative source for maintaining the standard, the architecture effectively decouples internal schema management from external distribution (JSON Schema, OpenAPI, etc.). This ensures rigor-

ous type safety across disparate data exchanges, ultimately fostering a more agile and interoperable ecosystem for the built environment.

Import flexibility

Our proposed architecture introduces a flexible import mechanism that allows the core schema to be extended with external vocabularies via URIs, preventing the standard from becoming a monolithic 'theory of everything'. This capability enables the composition of data models that mix and match distinct domain semantics while keeping the core lightweight.

In the 'Hello Wall' example, this is demonstrated through the direct import of OpenUSD for geometry, glTF PBR (Physically Based Rendering) for material definitions, and NlSfb and CCI for classification systems, and Open Geospatial Consortium (OGC) standards for geospatial data. By namespacing these attributes, we allow these diverse standards to coexist within a single dataset, facilitating richer interoperability and allowing the industry to leverage existing specialised standards rather than redefining them within IFC.

Similarly, the IFC does not need to reinvent the identifiers like *Global Trade Item Number* (GTIN) or *Global Location Number* (GLN), instead, those can be added to datasets with metadata acknowledging they are originating from GS1 organisation.

Geometry and Representation

Our proposal prioritises the use of explicit geometry to ensure consistent and efficient visualisation across software platforms. While the framework supports computational or procedural geometry definitions as semantic overlays or extensions, a objects should always be accompanied by an explicit geometric representation. This dual approach maximises interoperability by ensuring fast rendering and precise analysis without requiring consuming applications to interpret complex procedural logic. Currently, the alpha examples utilize the USD geometry schema (*usd::usdgeom*) as the rendering core to leverage its robust capabilities in visual composition and incremental updates. However, the architecture remains technology-agnostic regarding specific mesh formats; other explicit geometry definitions, such as those from the OGC and glTF, have also been evaluated to ensure broader compatibility with digital twin and GIS ecosystems.

Discussion and comparative analysis

Schema difference

Differences are visible when comparing equivalent IFC content captured in traditional STEP-based and proposed JSON-based form – see figure 6. The STEP sample has many more first-class objects like Direction, Point or Axis. While the STEP sample lists data in specific and defined entities, like the IfcSpace, the JSON sample defines data through components, inspired by the Entity Component System (ECS) architecture.

```
#1=IFCCARTESIANPOINT((0.,0.,0.));
#2=IFCDIRECTION((0.,0.,1.));
#3=IFCDIRECTION((1.,0.,0.));
#4=IFCAXIS2PLACEMENT3D(#1,#2,#3);
#5=IFCLOCALPLACEMENT($,#4);
#6=IFCCARTESIANPOINT((0.,0.,0.));
#7=IFCDIRECTION((0.,0.,1.));
#8=IFCDIRECTION((1.,0.,0.));
#9=IFCAXIS2PLACEMENT3D(#6,#7,#8);
#10=IFCLOCALPLACEMENT(#5,#9);
#11=IFCCARTESIANPOINT((0.,0.,0.));
#12=IFCDIRECTION((0.,0.,1.));
#13=IFCDIRECTION((1.,0.,0.));
#14=IFCAXIS2PLACEMENT3D(#11,#12,#13);
#15=IFCLOCALPLACEMENT(#10,#14);
#16=IFCCARTESIANPOINT((0.,0.123,0.));
#17=IFCDIRECTION((0.,0.,1.));
#18=IFCDIRECTION((1.,0.,0.));
#19=IFCAXIS2PLACEMENT3D(#16,#17,#18);
#20=IFCLOCALPLACEMENT(#15,#19);
#21=IFCSPACE('x…1',$,$,$,$,#20,#22,$,$,$,$);
```

```
(…)"data":[
{"path": "x...1",
  "attributes": {
    "usd::xformop": {
      "transform": [
        [1,0,0,0],[0,1,0,0],
        [0,0,1,0],[0,0.123,0,1]]}}},
{"path": "x...1",
  "attributes": {
       "bsi::ifc::class": {
         "code": "IfcSpace",
         "uri": "https://.../IfcSpace"}}}]
```

Figure 6: Comparison of a space definition in STEP-based IFC 4.3.2 (top) and proposed JSON syntax (bottom).

Inheritance depth and reification

The traditional IFC is full of reified objects, like a relation or a set of properties. For example, there are five intermediate entities between object and its property value: *"Object←RelDefinesByType→TypeObject→PropertySet→Property→IfcValue (select)→Simple value"*. In contrast, the proposal simplifies the assignment by allowing values to be added directly to objects: *"Object: {Property: Value}"*.

The IFC of today also employs rather deep inheritance of the entities. To read the attributes of an IfcWall shown in figure 7, one needs to traverse six levels up and six down the concept hierarchy. In proposed IFC, the inheritance depth can be three to four times shorter.

```
IfcRoot→…IfcObject→IfcProduct→IfcElement→…
IfcWall
→IfcRelDefinesByType
  →IfcWallType
    →IfcRelAssociatesMaterial
      →IfcMaterialLayerSetUsage
        →IfcMaterialLayerSet
          →IfcMaterialLayer: "mineral wool"
          →IfcMaterialLayer: "concrete"
```

```
Wall
(→WallType)
  →Insulation
```

```
    →Material: "mineral wool"
  →Core
    →Material: "concrete"
```

Figure 7: Comparison of inheritance depth in defining materials of a wall; IFC 4.3.2 (top) our solution (bottom).

On the other hand, the IFC 4.3.2 allows elements to only be typed by one element (type), while proposed IFC allows multiple typical levels to reuse data from. For example, a certain manufacturer's product could inherit from a typical kind of a product, that could also inherit from a more generic type.

Even though using IFC as product catalogues were possible with previous versions, our proposal introduces data ownership that allows to track origins of each property value, ensuring higher trust in the data. For example, a sustainability analyst could spot if Environmental Product Declaration (EPD) values were altered by architect before performing the Life Cycle Assessment (LCA).

Storage size

The mere storage is not a priority these days with fast internet and cheap storage. Still, the smaller files can be exchanged faster, so we performed a comparison of the simple building model represented as both STEP-based IFC and the proposed modernised IFC.

At first glance, the proposed IFC JSON syntax and serialisation does not bring advantages to storage size – the 'Simple scene' model recreated in .ifcx is over a quarter larger (see table 1). However, the JSON format is commonly condensed (minified), which gives our proposed solution a significant advantage over STEP representation – almost 60% reduced size. Additionally, it also compresses twice as efficiently as STEP, which is relevant for archiving.

Table 1: File size comparison on the sample dataset.

Version	Default (pretty-tified)	Minified	Compressed (zipped)
IFC 4.3.2	7.91 Mb	NA (same)	1.04 Mb
Proposed IFC (.ifcx)	10.10 Mb	3.39 Mb	0.52 Mb
Difference	+27.7 %	-57.1 %	- 50 %

The big difference can be observed when we look at a typical exchange process where files are being sent at regular intervals, despite insignificant changes being made. Table 3 shows that whereas in STEP-based IFC each transfer involves export of the full dataset, in the proposed solution only the deltas containing a change are being exchanged. This means significant storage saving throughout project duration.

Next steps

The comprehensive restructuring of the IFC data model into a modular, predictable ecosystem of parts serves as the necessary foundation for the true long-term objective: the creation of a standardised API. As articulated in the development roadmap, the finalisation of the Core data schema is the critical prerequisite that allows the industry to move beyond static file exchanges toward dynamic, cloud-native interactions where online agents or automated services, such as energy analysis tools or (manufacturing, simulation, etc.) bots, can interact directly with data. The current technical shifts toward JSON serialisation and ECS-like components are essentially the predecessor work required to standardize the openAPI interface itself – covering authentication, endpoints, and transactional calls. Binary data streaming is also not ruled out as part of the final solution. This will ultimately enable the real-time, granular data access that modern digital twins and connected ecosystems require.

Table 2: File size comparison on the sample dataset with two iterations of property value changes.

Version	Default (Mb)	1st change (Mb)	2nd change (Mb)	Total growth (%)
IFC 4.3.2	7.91	7.91	7.91	+200
Proposed IFC (.ifcx)	10.10	0.0004	0.0004	+0.008

Conclusions

The proposed modernisation of IFC represents a necessary shift for the construction industry, moving away from the monolithic legacy of previous standards toward a modular, high-performance data architecture. By adopting ECS-like components serialised in JSON, our proposal addresses the long-standing industry demand for files that are not only human-readable and web-friendly but also capable of supporting modern workflows such as incremental updates, version control, and multi-user collaboration. This technological leap is designed to drastically lower the barrier to entry for software vendors, prioritizing implementation speed and predictability to foster a more robust and reliable openBIM ecosystem.

The introduction of a modularisation strategy is central to this vision. By decoupling the stable technical Core from dynamic Domain Semantics and Industry Practices, the standard solves the historical challenge of painfully slow-release cycles. This separation of concerns allows the Core to serve as a reliable, domain-agnostic kernel for identifying, grouping, and positioning objects, while allowing specific disciplines, such as Rail, Tunnel, or Structural, to evolve their semantic definitions independently and rapidly.

While we introduce a breaking change regarding file formats, it ensures semantic continuity by leveraging existing IFC 4.x definitions and integrating with broader industry standards like USD, 3D Tiles, glTF, GS1, OGC or classification systems like ETIM, ECLASS, CCI or UNICLASS. Ultimately, the goal of IFC modernisation is to achieve fully reliable data usage. By reducing the complexity of the core schema and eliminating ambiguous exceptions, our work positions itself as a predictable and reliable foundation for the next generation of automation, digital twins, and connected built environment data.

Acknowledgments

The paper represents the result of the incubation phase of IFC 5 project at buildingSMART during 2021 to 2025. The final version will be further shaped by the buildingSMART and ISO processes. Authors would like to thank all the people involved in the initiative over the years, the members and community of buildingSMART, participants in recurring Implementers Assemblies and independent contributors.

References

Afsari, K., Eastman, C.M. & Castro-Lacouture, D. (2017) JavaScript Object Notation (JSON) data serialization for IFC schema in web-based BIM data exchange. *Automation in Construction*, 77, p.pp.24–51. https://doi.org/10.1016/j.autcon.2017.01.011

Beetz, J., de Laat, R., van Berlo, L. A. H. M., & van den Helm, P. (2010). Towards an open building information model server: report on the progress of an open IFC framework. In *10th International Conference on Design and Decision Support Systems in Architecture and Urban Planning, DDSS 2010*. Eindhoven University of Technology.

van Berlo, L., Krijnen, T., Tauscher, H., Liebich, T., van Kranenburg, A. & Paasiala, P. (2021) Future of the Industry Foundation Classes: towards IFC 5. In: *Proceedings of the 38th International Conference of CIB W78. Luxemburg*, 11 October 2021. p.pp.123–137.

Bonduel, M., Oraskari, J., Pauwels, P., Vergauwen, M., & Klein, R. (2018) The IFC to Linked Building Data Converter-Current Status. *Proceedings of 6th Linked Data in Architecture and Construction Workshop*.

Borkowski, A. S. (2026) A critical analysis of the direction of BIM development: From CAD through file-based approaches and SaaS to granular data. *Journal of Information Technology in Construction*, 31, 380-393. https://doi.org/10.36680/j.itcon.2026.016

Brouwer, J., Pauwels, P., Sparks, D. (2020) ifcJSON Repository. Available at: https://github.com/buildingsmart-community/ifcJSON (Accessed: 23 Jan 2026).

buildingSMART International (2020) Technical Roadmap buildingSMART: Getting ready for the future. Version: 30 April 2020. Available at: https://www.buildingsmart.org/wp-content/uploads/2020/09/20200430_buildingSMART_Technical_Roadmap.pdf (Accessed: 23 January 2026).

buildingSMART International (2022) IFC Formats. Available at: https://technical.buildingsmart.org/standards/ifc/ifc-formats/ (Accessed: 23 Jan 2026).

buildingSMART International (2024) IFC 4.3.2 documentation. Available at: https://ifc43-docs.standards.buildingsmart.org/ (Accessed: 23 Jan 2026).

buildingSMART International (2025) Global openBIM Mandates. Available at: https://www.buildingsmart.org/wp-content/uploads/2025/03/IFC-Mandate_2025.pdf (Accessed: 23 January 2026).

buildingSMART International (2026a) IFC5 Development Repository. Available at: https://github.com/buildingSMART/IFC5-development/ (Accessed: 23 Jan 2026).

buildingSMART International (2026b) Software Implementations. Available at: https://technical.buildingsmart.org/resources/software-implementations/ (Accessed: 23 Jan 2026).

Ismail, A., Nahar, A., & Scherer, R. (2017) Application of graph databases and graph theory concepts for advanced analysing of BIM models based on IFC standard. In: Conference: 24th International Workshop on Intelligent Computing in Engineering (EG-ICE), Nottingham, UK.

ISO 16739-1:2024 (2024) Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries – Part 1: Data schema. Available at: https://www.iso.org/standard/84123.html

Krijnen, T., & Beetz, J. (2020). An efficient binary storage format for IFC building models using HDF5 hierarchical data format. *Automation in Construction*, *113*.

Li, H., Liu, H., Liu, Y., & Wang, Y. (2016). An object-relational IFC storage model based on oracle database. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *41*, 625-631.

Microsoft (2026). What is TypeSpec? Available at: https://learn.microsoft.com/en-us/azure/developer/typespec/overview (Accessed: 23 Jan 2026).

Noardo, F., Arroyo Ohori, K., Biljecki, F., Krijnen, T., Ellul, C., Harrie, L., & Stoter, J. (2019). GeoBIM benchmark 2019: design and initial results. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives* (No. 2/W13, pp. 1339-1346). ISPRS.

Pauwels, P., Krijnen, T., Terkaj, W., Beetz, J. (2017) Enhancing the ifcOWL ontology with an alternative representation for geometric data https://doi.org/10.1016/j.autcon.2017.03.001

Solihin, W., Eastman, C., Lee, Y. C., & Yang, D. H. (2017). A simplified relational database schema for transformation of BIM data into a query-efficient and spatially enabled database. *Automation in Construction*, *84*, 367-383.

Tomczak, A., Benghi, C., van Berlo, L., Hjelseth, E. (2024) Requiring Circularity Data in BIM With Information Delivery Specification. https://doi.org/10.55845/REJY5239

Zhu, J., Wu, P., Lei, X. (2023) IFC-graph for facilitating building information access and query, https://doi.org/10.1016/j.autcon.2023.104778

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

Teaching BIM as a Computational, Queryable, and Analyzable Model: A Project-Based Framework Integrating APIs, Programming, and Early Performance Reasoning

Tobias Maile, Burcu Güldür Erkal
Technical University of Applied Sciences Augsburg, Germany

Abstract

Building Information Modeling (BIM) education in architecture, engineering, and construction (AEC) remains largely tool-centric, limiting students' understanding of data quality, automation, and performance reasoning. This paper presents a project-based educational framework that treats BIM models as computational, data-driven artifacts. Through programming, API-based model access, simplified performance analyses, and human-in-the-loop large language model (LLM) support, students develop computational BIM literacy through learning-by-doing. Student feedback indicates improved awareness of modeling consequences, data quality, and early performance assessment. The framework offers a transferable approach for computationally oriented AEC curricula and reframes BIM education as data quality education.

Introduction

Building Information Modeling (BIM) has become a central methodology in architecture, engineering, and construction (AEC), serving as a shared digital representation for design, analysis, construction, and operation. While BIM is gaining momentum in professional practice, its integration into higher education remains uneven. Many curricula still emphasize software operation and predefined workflows rather than conceptual understanding of BIM models as structured, queryable, data-driven representations (Succar, 2009; Eastman et al., 2018).

A recurring challenge in BIM education is the limited awareness of **data quality and data consistency**. Students frequently interact with BIM models as geometric artifacts, without fully understanding how model structure, parameter completeness, or semantic consistency influence downstream use. As a result, modeling errors or implicit assumptions often remain unnoticed until they manifest as problems in later project stages, such as performance analysis, quantity take-off, or interdisciplinary coordination. This delayed feedback makes it difficult for students to develop a causal understanding of how early modeling decisions propagate through digital workflows.

Automation and computational workflows represent another critical gap. Scripting and model-based automation are often introduced as advanced or optional topics, reinforcing the perception of computational methods as expert-only skills. This contributes to black-box tool usage and reliance on default software settings, limiting transparency and reproducibility of model-based reasoning (Borrmann et al., 2018).

Furthermore, BIM education is commonly dominated by **lecture-driven teaching formats**. While lectures convey conceptual knowledge, they are often insufficient for developing experiential understanding of how digital models behave under computational use. Multidisciplinary analyses, such as energy and structural assessment, are typically taught as isolated workflows, obscuring their shared reliance on a common underlying model.

Multidisciplinary performance analysis is frequently introduced as a set of isolated tools or domain-specific workflows. Structural, energy, and environmental analyses are often treated independently, each relying on separate model abstractions and software environments. This fragmentation obscures the fact that all analyses ultimately depend on the same underlying building model, albeit represented through different abstractions such as analytical models for structural reasoning or simplified thermal representations for energy assessment.

Summarizing the current challenges in education:

- Limited awareness of data quality and data consistency
- Insufficient understanding of how modeling decisions propagate errors
- Automation perceived as expert-only or tool-specific
- Predominance of front-loaded, lecture-based teaching
- Fragmentation of multidisciplinary performance analysis
- Lack of model transparency and explainability
- Over-reliance on software defaults and black-box workflows

Together, these challenges point to a fundamental educational gap: students are rarely taught to understand BIM models as **computational, data-driven artifacts** that can support automation, reasoning, and early performance exploration across multiple domains. Addressing this gap requires a shift from tool-oriented instruction toward **computational BIM literacy**, where modeling, data quality, automation, and performance reasoning are treated as interconnected aspects of a single workflow.

While prior BIM education research emphasizes collaboration, interoperability, and process integration, the ex-

plicit integration of programming, API-based data access, and performance reasoning into core BIM teaching remains largely unaddressed.

This paper investigates the following research question: How can BIM education be redesigned so that students develop computational BIM literacy - understanding BIM models as data-driven, queryable, and analyzable artifacts - rather than learning BIM primarily as a geometric authoring process?

To address this question, we designed, implemented, and evaluated a project-based educational framework that integrates programming, API-based model access, and simplified multidisciplinary performance reasoning within a single workflow.

The core pedagogical mechanism is that simplified performance models foster a learning-by-doing environment that exposes BIM data quality issues during computation.

In this work, the term *framework* refers to a transferable, tool-agnostic structure that integrates (1) BIM data access via APIs, (2) computational logic implemented in Python, and (3) simplified multidisciplinary performance models. The framework has been applied successfully with Revit, ArchiCAD, and Vectorworks, demonstrating that it is not dependent on any specific authoring tool. While the competencies addressed-data structuring, scripting, API querying, and performance reasoning-are most relevant for architecture and engineering students.

Recent developments in BIM education research highlight that these challenges are not only technical but pedagogical. Studies emphasize the importance of constructivist and project-based learning environments for developing digital competencies, particularly when students interact with computational models and simplified analyses Ao et al. (2025b); Nguyen and Adhikari (2025). Positioning BIM teaching within such learning theories reinforces the need for hands-on, exploratory workflows that connect model semantics, assumptions, and analytical consequences. This perspective motivates the project-based computational framework presented in this work.

This paper further positions the framework in recent educational research, describes the course's overall concept and learning objectives, before detailing the computational BIM workflow. It explains the project-based learning and describes the early performance reasoning for the energy and static topics. We refer to the LLM-supported learning, present evaluation results, and finally conclude the paper with a discussion and outlook.

Positioning within Existing BIM Education Research

Recent educational approaches further emphasize the need to move beyond tool-centric BIM teaching toward model-centric, computational, and interdisciplinary learning environments. The concept of "Next Generation BIM" frames BIM models as active carriers of data, logic, and assumptions that support automation, early performance reasoning, and life-cycle thinking across disciplines. This perspective explicitly calls for integrating programming, data access, and computational workflows into BIM education rather than treating them as advanced or optional skills (Wimmer et al., 2025).

Research on BIM education consistently highlights a gap between industry demands and academic training. Early BIM frameworks focused on maturity levels, collaboration, and process integration, while noting that education often lags behind professional practice (Succar, 2009). More recent studies confirm that BIM is frequently taught as software proficiency rather than as an integrated methodological paradigm (Eastman et al., 2018).

Recent global reviews highlight persistent gaps in BIM education, particularly in understanding student learning pathways, interdisciplinary curriculum design, and assessment practices Ao et al. (2025a). Moreover, contemporary studies show that computational thinking, programming, and automation remain underrepresented in most BIM programs despite their increasing relevance in industry. Evidence from AI-supported classroom environments demonstrates that students frequently struggle with debugging and data schema interpretation, underscoring the need for computational literacy as a core educational objective Sahraoui et al. (2025).

Large-scale empirical work further reveals a perception–practice gap: students appreciate the value of BIM for sustainability and analytical workflows but often lack the practical ability to apply these concepts independently Nguyen and Adhikari (2025). This finding aligns with our observations that students rarely develop a deep awareness of model structure, data consistency, and semantic assumptions when BIM is taught only through tool-oriented modeling.

Pedagogically, our approach aligns with constructivist and project-based learning principles. Recent BIM education research highlights the importance of learning motivation, behavioral intention, and authentic project engagement in developing digital competencies Ao et al. (2025b). Global curriculum reports also indicate a shift towards experiential learning models that integrate AI, automation, and computational reasoning Li (2025); Sofiane (2024).

Model-centric and openBIM-oriented teaching approaches have improved interdisciplinary collaboration and awareness of data exchange. However, computational literacy and automation remain underrepresented in many curricula. Prior work on life-cycle-oriented BIM education emphasizes interoperability and collaboration but often relies on predefined workflows and tool-specific processes (Maile et al., 2023). Conceptual frameworks for open BIM education similarly identify limited integration of programming and computational reasoning (Preidel et al., 2024).

The framework presented in this paper builds on this body of work by explicitly integrating programming and API-based model access into core BIM education. Rather than treating computation as an add-on, BIM models are framed as computational artifacts that can be queried,

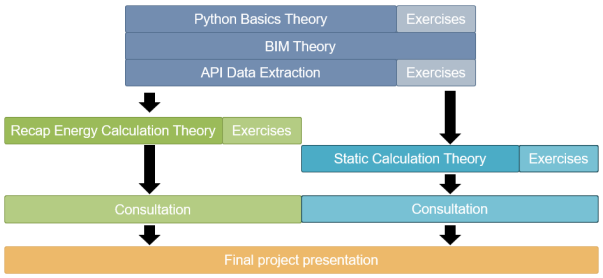


Figure 1: Phased course structure combining short theoretical inputs with exercises, consultations, and iterative project work. Programming, BIM theory, and performance reasoning are introduced incrementally and integrated into a single project-based workflow.

transformed, and evaluated across multiple performance domains.

Course Concept and Learning Objectives

The framework was implemented as a CAX module in the master program Energy Efficiency Design – E2D at the Technical University of Applied Sciences, Augsburg. Programming experience ranged from none to intermediate, and the course was explicitly designed to accommodate this diversity. Learning objectives focused on developing computational BIM literacy rather than tool proficiency. Students were expected to

1. understand BIM models as structured, data-centric representations,
2. access and query models via authoring APIs,
3. formulate explicit modeling and analysis assumptions,
4. implement transparent computational logic, and
5. interpret early-stage performance results critically.

Python serves as a lightweight orchestration layer, supporting analytical reasoning rather than being a primary goal of the work.

Course Structure. Teaching followed a phased progression, combining theoretical inputs with exercises and eventually project-based application as outlined in Figure 1. Programming fundamentals, BIM theory, and API-based data access for the first part of the course. Concepts of energy and structural performance reasoning were then introduced through accompanying exercises. The final part of the course was performing the project from modeling to code writing to result interpretation. In this phase, the students are supported through multiple consultation sessions to answer technical and scope questions and guide them towards a successful project. Consultation sessions provide targeted support and enable reflection on both modeling and computational decisions. Finally, the project was presented by each group in front of the class.

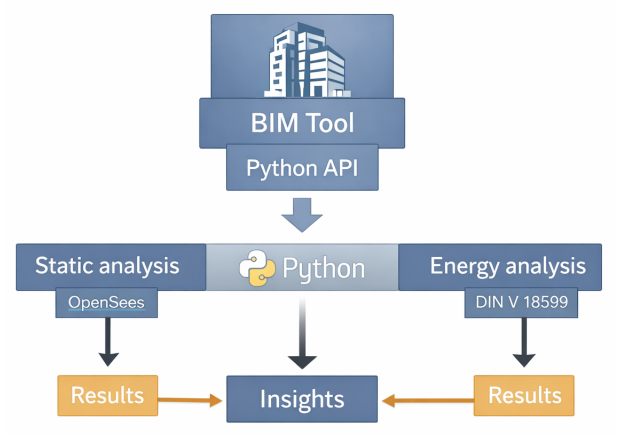


Figure 2: Computational BIM workflow used in the course. BIM models are accessed via authoring APIs and processed through Python-based computational logic to support parallel energy and structural reasoning and the generation of integrated insights.

Methods: Course Structure and Implementation

The course was conducted over 14 weeks (5 ECTS) and included students from the Energy Efficiency Design master program. The weekly structure consisted of typically 90-120 minutes of conceptual input (Python basics, BIM theory, energy and structural fundamentals) followed by guided hands-on sessions on API data access, code debugging, and computational workflow development.

Students were provided with Python templates, API documentation, example extraction routines, and simplified analytical formulations. Their final project required them to: (1) model a small building; (2) extract geometric and semantic data through the authoring tool's API; (3) compute simplified energy balances and generate structural models in OpenSees; and (4) present their assumptions, computational decisions, and debugging steps. Emphasis was placed on transparency, reproducibility, and reasoning rather than numerical accuracy.

Computational BIM Workflow

Students created BIM models using standard authoring tools. While most groups worked with Autodesk Revit, individual groups experimented with Vectorworks and ArchiCAD. The framework was intentionally designed to remain independent of specific authoring environments, focusing on API-based access to model data rather than tool-specific workflows.

Figure 2 summarizes the computational workflow used throughout the course. BIM geometry and metadata are accessed via authoring APIs and processed using Python scripts. Relevant information is mapped into simplified analytical representations for energy and structural reasoning, allowing students to explore performance implications of modeling decisions early in the design process.

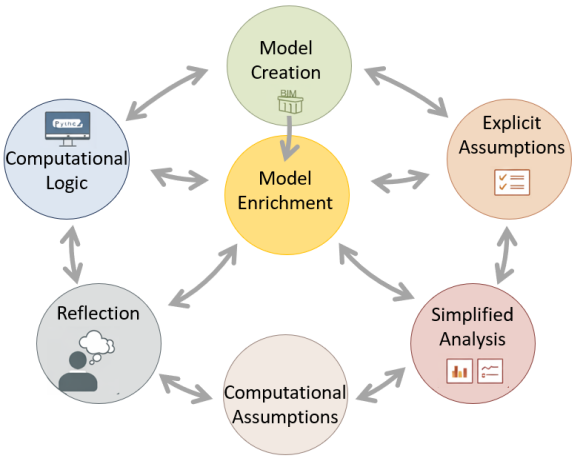


Figure 3: Learning-by-doing and reasoning loop adopted in the course. Students iteratively create BIM models, formulate explicit assumptions, implement computational logic, analyze simplified results, and reflect on outcomes to guide subsequent design iterations.

Object-Oriented Structuring as Didactic Principle

To support transparent and meaningful automation, students were encouraged to structure their scripts using basic object-oriented principles. Computational tasks were organized into domain-specific components, such as energy-related calculations, structural abstractions, and unit handling, rather than monolithic scripts.

This structuring was introduced as a didactic scaffold rather than formal software engineering training. Aligning code structure with domain concepts helped students clarify responsibilities, make assumptions explicit, improve the traceability of computational logic, and ease debugging and testing. In earlier course iterations, debugging within the Python API environment proved challenging for students and motivated the introduction of a more structured coding approach. This is why the object-oriented approach naturally emerged as a way to organize code and provide enough structure so that error checking and debugging were improved.

Human-in-the-loop computational assistance

Large language models were used as human-in-the-loop assistants to support syntax, debugging, and API exploration. Automated solution generation with LLMs was explicitly discouraged. Students remained responsible for validating logic and interpreting results. Recent work suggests that such assistive use of LLMs can effectively support exploratory reasoning in AEC contexts (Maile and Güldür Erkal, 2025).

Project-Based Learning Setup

The course followed an open-ended project-based learning approach without predefined solutions. Students were responsible for model creation, scripting, assumption definition, and interpretation of results.

Figure 3 illustrates the iterative learning process. Students repeatedly move between modeling, computation, and interpretation, enabling them to understand how design decisions, data quality, and assumptions influence analytical outcomes.

Early Performance Reasoning

The complementary expertise of the instructors enabled the integration of early performance reasoning in both energy and structural domains.

Energy Reasoning

Energy reasoning was introduced using a simplified monthly steady-state balance formulation that explicitly distinguishes between transmission losses, ventilation losses, internal gains, solar gains, and heating. This builds on prior coursework in construction physics in the related bachelor's program, where students were introduced to these principles.

This energy balance abstraction, used in the course, is based on the German Standard DIN V 18599 and is further simplified to omit storage effects and system efficiencies. Instead of aiming for regulatory compliance or high-fidelity simulation, the approach emphasizes transparency and causal understanding. Students explicitly define all assumptions and directly experience how changes in geometry, envelope properties, or usage profiles affect heating demand. For the various energy balance terms, the following information was needed from the BIM model:

- ventilation losses:
  - building volume
- transmission losses (for all building elements)
  - external surface area
  - u-value
- solar gains (for all windows)
  - g-value/solar heat gain coefficient (SHGC)
  - external surface area
  - orientation
- internal gains
  - usable area inside the building

Static Reasoning

Structural reasoning was introduced as a simplified approach to support early understanding of structural behavior within a CAX workflow. Rather than performing detailed structural design, the goal was to help students understand how structural layout and geometry influence global structural response.

The following key elements were obtained from the BIM model:

- nodal coordinates derived from element geometry,
- connectivity between structural elements,
- element types such as columns, beams, walls, and slabs,
- idealized material and cross-section properties,

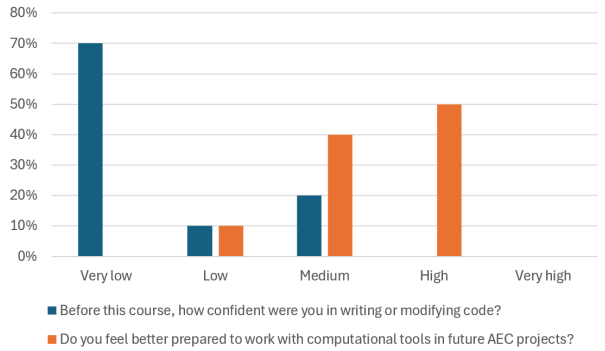


Figure 4: Self-assessed student confidence before and after the course. Responses indicate low initial confidence in writing or modifying code and a substantial increase in perceived preparedness to work with computational tools in future AEC projects.

- support conditions and boundary constraints.

Starting from the BIM model, students used authoring APIs to extract structural information and translate it into simplified analytical representations. Based on this extracted information, students automatically generated a code-based structural model in `Tcl` format for use in OpenSees (McKenna et al., 2010). Reduced assumptions appropriate for early design stages were applied to maintain transparency and ease of interpretation. This allowed students to directly observe how changes in BIM geometry or structural layout affected the generated OpenSees model and its global structural response.

The limitations of this simplified approach were explicitly addressed by emphasizing that the generated models represent idealized structural systems and provide insight into overall behavior rather than detailed effects.

Evaluation and Observed Learning Effects

Evaluation was based on qualitative and self-assessed feedback collected through an anonymous questionnaire at the end of the course, combined with evidence from student project presentations and submitted computational artifacts. In total, 23 students participated in the course, of which n=10 completed the questionnaire.

Questionnaire results. Students were asked to reflect on their prior experience with programming and their perceived preparedness to work with computational tools in future AEC projects. Figure 5 illustrates the distribution of responses. The results show that most students entered the course with low confidence in writing or modifying code, confirming the heterogeneous and largely non-programming-oriented background typical of AEC cohorts. Despite this, students reported a clear shift toward higher confidence in using computational tools after completing the course.

Students frequently emphasized that simplified models, explicit assumptions, and hands-on experimentation were more valuable for understanding performance relationships than black-box tools or lecture-only formats. Rather than focusing on numerical accuracy, students highlighted an improved ability to reason about cause–effect relationships and to explain why results changed.

Evidence from student work. Beyond self-assessment, student project outcomes provided concrete evidence of learning effects. Across all teams, students demonstrated mastery in three interconnected domains:

1. BIM modeling quality,
2. computational data extraction and analysis, and
3. integrated energy and structural reasoning.

Students developed a strong awareness that reliable downstream analyses depend on precise geometry, correct units, and well-structured parameter data. Several teams reported discovering unit inconsistencies (e.g. feet-to-meter conversions), missing thermal parameters, and incorrect element assignments only when attempting to compute energy balances or generate structural models from their BIM data. These experiences directly reinforced the importance of semantically consistent modeling practices. In multiple cases, students had to revise their BIM models because computational scripts exposed missing or inconsistent parameters.

All teams successfully implemented Python-based workflows that accessed BIM data via authoring APIs to compute simplified energy balances and generate analytical structural models for OpenSees. Students produced reproducible outputs such as CSV files, TCL scripts, and graphical plots derived directly from their BIM models. This demonstrates that students were not only able to model buildings but to compute with their models across multiple domains.

Several groups explicitly recognized interdependencies between energy performance and structural modeling. They observed how geometric decisions influenced both thermal behavior and structural response, for example how envelope area affects transmission losses while structural layout influences mass and connectivity. The use of a single parametric BIM model feeding both energy and structural analyses strengthened this integrated design understanding.

Typical outputs produced by the groups included CSV files listing extracted surface areas, thermal parameters, and orientations; Python-generated plots of monthly heating demand; and automatically generated Tcl files for simplified OpenSees models. Several teams performed gravity load analyses and modal evaluations, demonstrating how geometric changes in the BIM model directly influenced the analytical model.

Common errors included inconsistent units (e.g., feet-to-meters), missing or misassigned U-values, disconnected analytical nodes in slabs, incorrect window orientation extraction, and Python path issues in the Revit environment. These issues only became visible during computation, reinforcing the pedagogical principle that data quality awareness emerges most effectively when students compute with their own model data.

Table 1: Summary of lessons learned across all student teams.

Competency Area	Lessons Learned (Evidence)
BIM Model Quality	Students emphasized precise geometry, correct units, and complete thermal and structural parameters. Errors such as incorrect unit conversions reinforced the need for careful model structuring.
Data Extraction and API Usage	Students extracted areas, volumes, U-values, materials, and orientations using authoring APIs. They automated energy balance calculations and implemented robust parameter detection.
Energy Calculation Concepts	Teams implemented simplified energy balances including transmission and ventilation losses, internal and solar gains, and heating demand, recognizing how geometry and assumptions influence results.
Structural Modeling and Analysis	Students exported analytical models and executed OpenSees simulations. They learned the importance of node alignment, connectivity, and consistent element definitions.
Integrated Design Thinking	Students identified synergies and trade-offs between energy demand and structural utilization. A single BIM model fed both energy and structural workflows, enabling informed design decisions.
Workflow and Project Management	Students emphasized consistent naming, version control, file organization, and avoiding special-character issues in BIM workflows.
Debugging and Problem Solving	Students encountered and resolved API syntax errors, incorrect file paths, misassigned properties, and parameter inconsistencies, reinforcing reflective learning.

Students also reported meta-level workflow lessons, including the importance of consistent naming conventions, file organization, API error handling, and validation of modeling assumptions. These reflections indicate that the course fostered realistic digital engineering competencies beyond isolated tool usage.

Table 1 summarizes the lessons learned across all student teams.

Observed learning effects. Three dominant learning effects were observed across participants (as shown in Table 2):

- *Increased programming confidence.* Students reported improved understanding of basic programming concepts such as variables, loops, and functions, indicating a reduced computational entry barrier.
- *Shift in BIM understanding.* Several students described a conceptual shift from geometry-focused modeling toward working with structured, queryable data.
- *Improved performance reasoning.* Implementing simplified energy and structural calculations improved students' confidence in interpreting results rather than accepting outputs from black-box tools.

The role of large language models was also highlighted by many students. LLM tools were primarily used for syntax support, debugging, and API exploration, enabling students to focus on conceptual reasoning.

Overall, the triangulated evidence from questionnaires, student artifacts, and project presentations indicates that treating BIM models as computational artifacts - combined with learning-by-doing and transparent simplifications - can foster meaningful computational BIM literacy even among students with limited prior coding experience.

Implementation Challenges

Integrating programming into BIM education required a substantial time investment, particularly for students with limited coding experience. Approximately one-third of weekly contact time was dedicated to debugging assistance and API exploration. These observations mirror findings from recent studies showing that students in AI-assisted BIM classrooms struggle with syntax, rule formulation, and data-schema interpretation Sahraoui et al. (2025). Instructor expertise in both BIM semantics and computational methods was essential to ensure that the cognitive load remained manageable within a project-based learning environment.

Table 2: Summary of qualitative student feedback (self-reported).

Dimension	Dominant Observation
Prior programming experience	Predominantly low or very low
Perceived programming understanding after course	Improved or partially improved
BIM understanding shift	From geometry-focused to data-oriented (for most students)
Usefulness of LLM assistance	High for syntax, debugging, and API exploration
Performance reasoning confidence	Improved understanding and interpretability
Ability to extend scripts independently	Possible with guidance; limited full independence

Limitations

The evaluation is qualitative in nature and based on a single course iteration with a limited number of participants. No control group was used, and results reflect perceived learning outcomes rather than measured performance gains. Future work will include longitudinal and comparative studies.

Discussion and Outlook

The presented framework addresses several persistent shortcomings in BIM education, including tool-centric teaching, black-box workflows, and fragmented multidisciplinary analysis. By treating BIM models as computational artifacts rather than geometric authoring results, the course enabled students to directly experience how data quality, modeling assumptions, and semantic consistency influence analytical outcomes.

A key insight emerging from this framework is that simplified performance models act as a didactic mirror. When students attempt to compute energy demand or generate structural models from their BIM data, inconsistencies, missing parameters, unit errors, and modeling assumptions become immediately visible. Rather than discovering such issues in later project phases or external tools, students encounter them during the act of computation. This immediate feedback loop fosters a causal understanding of how modeling decisions propagate through digital workflows and strengthens awareness of data quality and consistency.

This approach differs fundamentally from conventional BIM teaching practices. Conventional teaching often follows a sequence of modeling, exporting, and processing models in domain-specific tools. In contrast, the presented framework establishes a direct connection between the BIM model and computational reasoning through API access and scripting. This shifts the educational focus, as shown in Table 3:

Table 3: Comparison between traditional BIM education and the presented computational framework.

Traditional BIM education	Presented framework
Geometry-focused modeling	Data- and structure-focused modeling
Tool proficiency	Computational BIM literacy
Predefined workflows	Querying and transforming model data
Black-box simulations	Transparent simplified reasoning
Fragmented domain analyses	Multiple analyses from the same model

The results indicate that even students with little or no programming experience can develop meaningful computational literacy when programming is framed as a reasoning tool rather than an end in itself. The use of Python as a lightweight orchestration language, combined with object-oriented structuring aligned to domain concepts, allowed students to organize computational logic in a way that was traceable and understandable.

The integration of large language models as human-in-the-loop assistants, enabling students to make and implement their own decisions, further supported this process. Students reported that LLM tools helped them overcome syntactic barriers and explore APIs, allowing them to focus on conceptual reasoning. Importantly, responsibility for validating logic and interpreting results remained with the students, reinforcing critical engagement rather than automated solution generation.

Although the evaluation is qualitative and limited to a single course iteration, observed learning effects and student feedback indicate a shift in how BIM is perceived - from a geometric modeling task to working with structured data that can be queried, transformed, and analyzed. Students reported improved confidence in programming, a changed understanding of BIM, and an increased ability to reason about performance outcomes rather than accept results from black-box tools.

A central strength of the framework lies in its transferability. The approach does not depend on specific authoring software but on the availability of an API and basic scripting capabilities. This makes the framework adaptable to

different BIM environments and educational contexts. Future investigations into open-source BIM platforms and API-rich environments may further enhance its accessibility.

Future work will extend the framework to additional performance domains such as embodied carbon assessment and cost analysis, enabling students to explore further dimensions of model-based reasoning. Longitudinal studies and comparative evaluations with traditional BIM courses are planned to assess learning effects quantitatively. The integration of agent-based and LLM-supported workflows also represents a promising direction for further research into computational BIM education. Future work will explore integrating the Model Context Protocol (MCP) to enable LLMs to access and query BIM models through standardized, auditable tool interfaces (MCP, 2026; Anthropic, 2024). First preliminary investigations of this approach promise more transparent, reproducible, and educationally effective natural-language interactions with BIM data by explicitly linking user intent to authoring-tool API operations.

Ultimately, this work suggests that BIM education can move beyond teaching how to model buildings toward teaching how to compute with building models. By embedding programming, API access, and simplified performance reasoning into core BIM teaching, students can develop the computational literacy necessary to understand, question, and extend digital workflows in contemporary AEC practice.

Beyond its immediate didactic benefits, the framework contributes to broader discussions in BIM education. Recent reviews call for integrated, interdisciplinary, and computationally grounded learning models Ao et al. (2025a); Li (2025). Our findings support these directions by demonstrating that constructivist, project-based environments, combined with transparent computational workflows, can reduce the perception–practice gap documented in current educational research Nguyen and Adhikari (2025). By embedding programming and performance reasoning directly into students' modeling workflows, the course advances emerging pedagogical aspirations for computational BIM literacy.

Acknowledgments

We want to thank the students of the CAX module for their participation in the course and for their feedback.

References

Anthropic (2024). Introducing the model context protocol. `https://www.anthropic.com/news/model-context-protocol`. Accessed Feb 8, 2026.

Ao, Y., Wang, Y., Peng, P., and Zhao, J. (2025a). Bim adoption: A review of global trends, research progress, and future directions. In Building Information Modeling (BIM) Adoption, pages 1–12. Springer.

Ao, Y., Wang, Y., Peng, P., and Zhao, J. (2025b). Conclusion and future research: Bridging bim education, engagement, and industry readiness. In Building Information Modeling (BIM) Adoption, pages 163–169. Springer.

Borrmann, A., König, M., Koch, C., and Beetz, J., editors (2018). Building Information Modeling: Technology Foundations and Industry Practice. Springer.

Eastman, C., Teicholz, P., Sacks, R., and Lee, G. (2018). BIM Handbook. Wiley, 3 edition.

Li, K. (2025). Bim education – global 2025 update report. Technical report, NATSPEC. Version 12.0.

Maile, T., Bartels, N., and Wimmer, R. (2023). Integrated life-cycle oriented teaching of the big-open-bim method. In Proceedings of the 2023 European Conference on Computing in Construction (EC3) and CIB W78 Conference.

Maile, T. and Güldür Erkal, B. (2025). Ai-driven insights in aec: Literature discovery and practical applications. In Proceedings of the 2025 European Conference on Computing in Construction (EC3) and CIB W78 Conference, Porto, Portugal.

McKenna, F., Fenves, G. L., and Scott, M. H. (2010). OpenSees: Open system for earthquake engineering simulation. `https://opensees.berkeley.edu`.

MCP (2026). Model context protocol — official documentation. `https://modelcontextprotocol.io/`. Accessed Feb 8, 2026.

Nguyen, T. D. and Adhikari, S. (2025). Bridging the gap: Enhancing bim education for sustainable design through integrated curriculum and student perception analysis. Computers, 14(11):463.

Preidel, C., Bartels, N., Maile, T., and Wimmer, R. (2024). An open platform for building information modeling education: Conceptual framework and potential impact. In Proceedings of the CIB W78 Conference.

Sahraoui, I., Kim, K., Gao, L., Din, Z., and Senouci, A. (2025). Integrating generative ai in bim education: Insights from classroom implementation. arXiv.

Sofiane, M. (2024). The 2024 natspec report: A detailed overview of bim education and maturity.

Succar, B. (2009). Building information modelling framework: A research and delivery foundation for industry stakeholders. Automation in Construction, 18(3):357–375.

Wimmer, R., Bartels, N., and Maile, T., editors (2025). Next Generation BIM: Für Praxis und Lehre. bSD Verlag.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

NARRATIVES OF DIGITAL TWINS: HOW FRAMINGS SCRIPT GOVERNANCE

Johan Ninan and Ranjith K. Soman
Delft University of Technology, the Netherlands

Abstract

Infrastructure digital twins are transformative tools for managing complex systems. This study examines how infrastructure digital twins are narratively constructed and how these narratives shape governance. Drawing on narrative theory, we analyze 75 academic publications and their framing of digital twins to identify recurring conceptual patterns. The findings indicate that digital twins are beyond mere technical artefacts, rather are embedded in narratives that implicitly script technological function, institutional roles, and governance alignments. There is a need to treat digital twin adoption as a technical innovation and also as an institutional design challenge shaping how infrastructure decisions will be made.

Introduction

Digital twins are increasingly promoted as transformative tools for the planning, operation, and governance of infrastructure systems (Keskin et al., 2022). Across transport networks, water management, and energy systems, digital twins are expected to improve performance monitoring, predictive maintenance, and long-term asset management (Pan & Zhang, 2021). In this sense, they are often presented as key enablers of data-driven infrastructure governance. Yet research on digitalization in the construction sector has consistently shown that the success of digital technologies depends not only on technical capability but also on their integration into organizational practices, professional roles, and institutional decision processes (Miraj et al., 2025). Studies of building information modelling and related technologies demonstrate that implementation outcomes are shaped by how tools are interpreted, embedded, and governed within organizational life, rather than by their technical sophistication alone (Datta et al., 2020). This suggests that the way emerging technologies are conceptualized may influence how they are specified, procured, and ultimately used (Broo & Schooling, 2023).

The rapid expansion of digital twin research has resulted in a focus on defining what a digital twin is, identifying technical architecture, or assessing potential performance benefits (Jiang et al., 2021). Much less attention has been paid to how digital twins are conceptually framed in practice. In particular, little empirical work examines how infrastructure digital twins are described in academic discourse and what these descriptions imply about users, responsibilities, and decision processes. This is a significant gap, because the language used to characterize new technologies often shapes how problems are understood and how solutions are designed (Ninan et al., 2022). The paper therefore asks: How are infrastructure digital twins conceptually framed in academic discourse, and what institutional assumptions do these frames imply for their governance?

The aim is to contribute to an understanding of digitalization in infrastructure sectors by highlighting the role of conceptual framing in shaping technological expectations and institutional integration. In the literature review section, we review literature on digital twins in infrastructure and the role of narratives in technologies. Following this, we detail the methodological choices relating to the use of academic literature, search criteria and highlight how data was collected and analyzed. In the findings section, the dataset is summarized and in the discussion section the conceptual categories are argued. The conclusion section summarizes the findings from this research and emphasizes the theoretical and practical contributions before highlighting the limitations and scope for future research.

Literature review

Digital twins in infrastructure projects

Digital twins have rapidly emerged as a central concept in the digital transformation of infrastructure projects and support predictive maintenance, optimization, and long-term planning (Ham & Kim, 2020). They have potential to improve operational efficiency, sustainability, and system reliability by linking physical assets with continuously updated digital representations (Kaewunruen & Lian, 2019). Digital twin research in the built environment spans multiple strands, including framework development, technical integration, and case-based applications (Broo & Schooling, 2023). Much of the literature categorizes studies into technical modelling approaches, implementation architectures, or application domains such as asset management and smart cities (Keskin et al., 2022). Similarly, work focusing specifically on civil infrastructure emphasizes issues such as interoperability, data integration, and lifecycle modelling, reflecting the strong engineering orientation of the field (Jiang et al., 2021). Across these studies, digital twins are typically treated as technical systems that integrate IoT data, analytics, and simulation capabilities.

Olde Scholtenhuis & Soman (2025) challenge this techno-centric framing, showing that many construction digital twins function effectively despite operating with varying latency, fidelity, and levels of automation, suggesting that usefulness depends more on contextual decision needs than on technological sophistication. While existing focus has produced valuable advances in system architecture and modelling techniques, however, the conceptual and institutional dimensions of digital twin adoption remain comparatively under-explored.

While early digital twin research emphasized technical architectures, more recent work acknowledges that digital twins operate within complex sociotechnical environments (Miraj et al., 2025). Scholars increasingly recognize that digital twins affect organizational coordination, professional roles, and governance arrangements, rather than simply providing new computational capabilities (Agrawal et al., 2023). Research in the AEC sector has long shown that digital technologies reshape work practices and decision processes (Ernstsen et al., 2021). Studies of BIM adoption demonstrate that implementation outcomes depend on how digital tools are interpreted, negotiated, and embedded in organizational routines (Papadonikolaki et al., 2019). Similar insights are now emerging in digital twin research, where scholars highlight organizational adaptation, data governance, and institutional coordination as key barriers to implementation (Broo & Schooling, 2023). Digital twin deployment often encounters conceptual ambiguity and difficulties in creating shared understanding among stakeholders (Tripathi et al., 2024). Practitioners and other stakeholders struggle not only with technical integration but also with defining what a digital twin actually is and how it should be used within organizational processes (Kober et al., 2024). Thus, digital twins should be understood as sociotechnical systems whose performance depends on institutional integration and shared interpretive frameworks (Nochta et al., 2021).

A related strand of research focuses on the role of digital twins in supporting decision-making, policy and governance processes (Solman et al., 2022). In smart city and infrastructure contexts, digital twins are increasingly presented as tools for scenario analysis, policy evaluation, and long-term system planning (Ham & Kim, 2020). Digital twins afford multiple possibilities to support public-sector decision-making by modelling complex urban dynamics and enabling scenario-based planning (O'Dwyer et al., 2020). There is a need to understand digital twins as human-technology interfaces rather than purely analytical systems; for instance, studies of construction control rooms show how digital twins can function as collaborative decision environments that enhance coordination and judgement rather than simply automate analysis (Soman et al., 2025). Similarly, research on digital twin ontologies and standards emphasizes the need to align digital twin development with principles such as trustworthiness, public value, and governance transparency (Lavikka et al., 2018; Deng et al., 2021). Thus, digital twins must support technical performance as well as accountability and institutional legitimacy.

Despite these emerging concerns, digital twins are primarily conceptualized as information systems or modelling platforms rather than as environments for structured institutional decision-making (Korenhof et al., 2021). Researchers and practitioners use the term to refer to a wide range of systems, from data visualization dashboards to high-fidelity simulation environments (Nochta et al., 2021). To resolve this ambiguity, typologies, architectures, or maturity models are proposed (Uhlenkamp et al., 2022). Yet few studies examine how digital twins are actually described in discourse or how conceptual metaphors shape expectations about their functions.

Narratives of technology

The role of narratives and framing in shaping technological trajectories are emphasized in construction innovation, particularly on how narratives shape innovation (Ninan et al., 2022). Technologies diffuse based on technical superiority as well as how they are interpreted, justified, and legitimized through narratives that make certain futures appear plausible and desirable (Williams, 2006). Within infrastructure sectors, narratives help manage tensions between innovation and risk, often shaping how stakeholders interpret responsibility, value, and uncertainty (Sergeeva and Ninan, 2023). Conceptual framings therefore influence how technologies are discussed and how they are designed, procured, and governed (Davidson, 2002).

From this perspective, analyzing how digital twins are described can reveal the institutional assumptions embedded in their conceptualization. Language such as "platform," "service," or "dashboard" invokes particular models of interaction and responsibility that originate in software ecosystems rather than infrastructure governance. Although existing research acknowledges the sociotechnical nature of digital twins and the importance of governance considerations, little empirical work examines how digital twins are conceptually framed in academic discourse. Most studies focus on technical architectures, implementation challenges, or application benefits. Few investigate how language constructs expectations about digital twins, what kinds of actors and decisions are implied, or how responsibility and uncertainty are represented. This gap is particularly important in infrastructure contexts, where digital twins must support long-lived systems involving multiple organizations, regulatory constraints, and public accountability. Conceptual framings that emphasize software metaphors may obscure these institutional realities and shape design priorities in ways that misalign with governance needs.

Research setting and method

The present study adopts a qualitative interpretive design to examine how infrastructure digital twins are conceptually framed (Gephart, 2004). Conceptual framing is understood here as a sociotechnical process

through which technologies are described, interpreted, and rendered meaningful. Such processes are not directly observable in technical artefacts themselves but are expressed through language, metaphors, and descriptive formulations (Goatly, 2004). A qualitative approach is therefore appropriate because it allows the study of how technologies are imagined, positioned, and justified through discourse rather than through measurable performance outcomes (Dyer & Wilkins, 1991).

Academic publications are treated in this study as a discursive arena in which technological expectations and norms are constructed and reproduced. They play an important role in shaping technological trajectories because they influence funding priorities, procurement practices, standards development, and professional education. The way technologies are described in the literature therefore contributes to how they are specified and implemented in practice. The empirical material consists of peer-reviewed publications indexed in Scopus. The dataset was constructed using the search query: TITLE-ABS-KEY ("digital twin as" AND infrastructure) till 31 December 2025. This query was chosen because "digital twin as …" constructions explicitly reveal conceptual framings rather than technical functions. Such phrases tend to appear in definitional passages where authors articulate what a digital twin is understood to be. The search returned 75 publications spanning infrastructure domains including transport, energy systems, water management, and urban infrastructure. The papers represent a heterogeneous set of disciplines including engineering, information systems, urban planning, and construction management. Together, they provide a cross-disciplinary corpus capturing how digital twins are framed in infrastructure research and constitute as a naturally occurring data (Ninan, 2020). The analysis therefore focuses not on the intentions or beliefs of the authors, but rather on how digital twins are represented and what institutional assumptions those representations imply.

The study employs qualitative content analysis with iterative coding (Miles & Huberman, 1994). The analysis proceeded in three stages. Stage 1 involves the identification of framing constructions. All publications were read to identify passages where digital twins were defined, described, or metaphorically framed. Particular attention was given to sentences using the structure "digital twin as …", and these passages were extracted and compiled into a corpus of framing statements. Stage 2 involved coding of conceptual framings. Framing statements were coded inductively to identify recurring conceptual categories. Through iterative comparison, seven dominant framings emerged such as digital twin as product, digital twin as platform, digital twin as service, digital twin as dashboard/interface, digital twin as predictive or optimization system, digital twin as organizational capability, digital twin as decision environment. Each framing category was treated as an interpretive lens through which authors implicitly define the system's purpose, users, and scope. Stage 3 involved coding of institutional implications. The themes were grounded in existing literature for validation.

Findings

The analysis of 75 infrastructure digital twin publications reveals three empirical patterns such as dominance of software metaphors, marginalization of decision processes, and absence of governance-centered framings. These are described below.

Dominance of software metaphors

Across the entire period from 2017 to 2025, digital twins are overwhelmingly framed using software- and computation-oriented metaphors. Early publications establish this pattern clearly, presenting digital twins as virtual replicas, models, or services that mirror physical systems. For example, Borodulin et al. (2017) describe manufacturing processes as being "accompanied by virtual models – Digital Twins," emphasizing that these models should be implemented as "independent computational service[s]." In this framing, the digital twin is conceptualized primarily as a software artefact that represents and simulates the physical world. This metaphorical emphasis persists in subsequent years. Kudriashov and Markov (2019) frame digital twins as analytical instruments designed to analyze the behavior of complex technical systems under various operating conditions, reinforcing the idea of the digital twin as a computational lens rather than an organizational or political construct. Even as application domains expand, the core metaphor remains rooted in modelling, simulation, and software execution.

From 2020 onwards, software metaphors become increasingly elaborate and infrastructural. Zibuschka and Ruff (2020) describe digital twins as systems that aggregate data from multiple sources, implicitly positioning them as data platforms. Deutschmann et al. (2021) similarly emphasize integration, noting that digital twins must combine technical, operational, and social components within what they describe as "industrial-social mobility hubs." Despite this broader scope, the language remains grounded in systems engineering and software integration.

Subsequently, software metaphors are further reinforced through references to real-time operation, automation, and scenario generation. Michalec (2025), while critical of prevailing framings, notes that digital twins are commonly promised as systems "able to automatically operate in real-time and generate countless scenarios to advise with planning of new infrastructures." Even here, where the article explicitly interrogates governance, the dominant discourse it critiques remains firmly anchored in computational imaginaries.

Marginalization of decision processes

A second major finding concerns how decision-making is treated in the publications. While many papers state that digital twins support decisions, the decision processes themselves are rarely described, theorized, or problematized. Instead, decision-making is presented as a

downstream activity that benefits automatically from improved models and data. In early industrial literature, this marginalization is particularly pronounced. Borodulin et al. (2017) focus extensively on simulation orchestration and scalability but make no reference to who uses these simulations or how decisions are made on their basis. Similarly, Kudriashov and Markov (2019) emphasize system analysis without specifying decision contexts, actors, or institutional settings. As digital twins are increasingly linked to complex systems, decision-making begins to appear more frequently, yet still in a limited form. Deutschmann et al. (2021) acknowledge the socio-technical complexity of airports but do not elaborate on how conflicting objectives or interests are negotiated through digital twin systems. Decision-making is implied rather than examined.

By 2022 and 2023, several publications explicitly reference uncertainty and coordination, which implicitly presuppose decision-making under conditions of risk. Barat and Kulkarni (2022) frame digital twins as responses to "environmental uncertainties and hyperconnectivity," suggesting that they play a role in navigating complex choices. However, the publications do not specify decision rules, accountability structures, or the distribution of authority.

More explicit references to decision-making appear in the most recent publications, yet even here the process remains under-specified. Fett and Wilking (2025) state that digital twins offer support of decision-making processes, but do not clarify whose decisions are supported, how conflicting values are addressed, or how responsibility is allocated. Decision-making is treated as a technical function rather than a social or political practice. The marginalization of decision processes is particularly notable given the increasing scope and influence attributed to digital twins. As these systems are positioned as tools for planning infrastructure, managing energy systems, or shaping urban environments, the lack of attention to how decisions are made, contested, and legitimized becomes a significant analytical gap.

Absence of governance-centered framings

The third and most striking finding is the persistent absence of governance-centered framings across the dataset. Although many papers deal with domains that are inherently governance-intensive, such as energy systems, urban infrastructure, and sustainability, governance is rarely named explicitly and even more rarely analyzed.

In early publications between 2017–2019, governance is entirely absent. Digital twins are framed as internal tools for firms or engineers, with no reference to regulation, public accountability, or collective decision-making. This absence reflects the technical orientation of early digital twin research communities. From 2020 onwards, governance-relevant issues begin to surface indirectly. Zibuschka et al. (2020) discuss data aggregation, while Deutschmann et al. (2021) describe industrial-social mobility hubs, implicitly raising questions about coordination and oversight. However, these issues are not framed in governance terms.

The clearest engagement with governance appears in a small subset of recent publications. Michalec (2025) explicitly critiques the dominant framing of digital twins-as-models, arguing that this framing conceals the politics of data sharing. The paper highlights that debate on ethics and politics remain limited to modelling concerns such as "bias, accuracy or explainability," while broader governance questions are sidelined. Michalec (2025) argues that energy governance should instead focus on "anticipating the reconfiguration of the political and economic relationships enabled by new data sharing infrastructures." Despite this exception, governance-centered framings remain rare. Even sustainability-oriented studies, such as López-Alfaro et al. (2025), frame digital twins as tools for "reducing light pollution and promoting sustainable lighting strategies" without addressing regulatory authority, public participation, or democratic accountability. Sustainability is treated as a technical outcome rather than a governance challenge.

Overall, the findings demonstrate that digital twin discourse has expanded in scope but not in governance awareness. While digital twins are increasingly positioned as central to infrastructure planning, sustainability transitions, and public decision-making, the papers largely avoid explicit engagement with governance concepts. These findings show a consistent pattern across time. Digital twins are predominantly framed through software metaphors, decision-making is treated as a secondary and largely unexamined activity, and governance-centered perspectives are largely absent. This combination suggests that digital twin discourse remains technically sophisticated but institutionally and politically underdeveloped. The persistence of these patterns across multiple publications indicates that they are not incidental but structural features of how digital twins are currently conceptualized.

Discussion

This study sets out to examine how infrastructure digital twins are narratively framed in academic discourse and what institutional assumptions these framings imply for governance. The findings indicate that digital twins are beyond mere technical artefacts, rather are embedded in narratives that implicitly script technological function, institutional roles, and governance alignments. Drawing on narrative theory, we interpret these framings as institutional scripts that shape how technologies are expected to function within socio-technical systems.

Narratives of technological function

Across the literature, infrastructure digital twins are predominantly narrated through metaphors of platforms, monitoring systems, optimization engines, and predictive tools. These framings define the technology primarily in terms of functional capability rather than institutional purpose. Narrative theory suggests that such descriptions do more than communicate features; they establish what counts as the legitimate problem-solving role of technology within organizational life.

By emphasizing simulation, optimization, and real-time monitoring, these narratives position digital twins as instruments for enhancing technical performance. In doing so, they foreground engineering rationality while backgrounding institutional negotiation. This aligns with broader observations in science and technology studies that technical narratives often stabilize innovation trajectories by narrowing attention to calculable functions rather than contested purposes (Bowker, 2008). Within infrastructure contexts, this functional scripting matters because it frames digital twins as solutions to operational inefficiencies rather than as tools embedded in political or governance processes. As a result, literature implicitly constructs digital twins as technologies of optimization rather than technologies of institutional coordination.

Narratives of institutional roles

A second narrative pattern concerns the institutional roles implied by digital twin framings. Actors are frequently described generically as "users," "operators," or "stakeholders," while decision processes remain largely implicit. From a narrative perspective, this abstraction is consequential: it removes the need to specify authority structures, decision rights, or accountability mechanisms. Narratives that present digital twin outputs as automatically actionable insights reinforce what can be termed a technocratic authority script, in which the legitimacy of decisions derives from data visibility rather than institutional deliberation. This pattern resonates with research on socio-technical imaginaries showing how technological visions can naturalize particular governance assumptions without explicitly debating them (Jasanoff, 2015).

In the context of infrastructure systems, this narrative shift is particularly important. Infrastructure decisions typically involve trade-offs between public value, cost, environmental impact, and political priorities (Foerster et al., 2015). Yet when digital twins are framed as providing objective, real-time knowledge, the decision process itself becomes narratively compressed. Responsibility appears to follow from information access rather than institutional authority. Thus, literature implicitly constructs governance as a downstream effect of data rather than as a structured process of collective judgement.

Narratives of governance alignment

The third narrative dimension concerns how digital twins are imagined aligning with institutional arrangements. Much of the literature situates digital twins within procurement cycles, lifecycle management frameworks, or operational integration processes. Governance is therefore narrated as an implementation issue rather than as a domain of institutional design.

Narrative theory suggests that such framings matter because they define the "setting" in which technologies are expected to operate (Coeckelbergh & Reijers, 2016). By embedding digital twins within performance management and efficiency narratives, literature positions them as tools that reinforce existing institutional logics rather than as technologies that might reconfigure them. These echoes work on technological narratives showing how innovations are often framed as extensions of current systems rather than as catalysts for institutional change (Bijker & Law, 1994). The absence of governance-centered farming is therefore not merely an omission but a narrative effect. When digital twins are described primarily through operational and performance lenses, alternative imaginaries, such as digital twins as platforms for public deliberation, regulatory coordination, or cross-sector negotiation, remain outside the discursive horizon.

Taken together, these narrative patterns show that academic discourse does more than describe digital twins; it constructs institutional expectations about them. Literature tends to script digital twins as technologies for optimization rather than coordination, sources of actionable insight rather than inputs to deliberation, and tools embedded in existing institutional structures rather than catalysts for institutional redesign. This interpretation contributes to narrative theory by demonstrating how technological discourse can function as a form of institutional scripting. It also extends infrastructure governance research by showing that governance assumptions are often embedded in technical descriptions long before systems are implemented. In this sense, digital twin narratives shape not only how technologies are understood but also how their governance becomes thinkable.

Conclusion

This study examined how infrastructure digital twins are narratively framed and what institutional assumptions these framings imply for governance. The analysis shows that infrastructure digital twins are predominantly framed using software-oriented concepts such as platforms, dashboards, and services. While these framings may be useful for describing technical components, they risk obscuring the role of digital twins as socio-technical decision environments embedded in organizational processes and governance structures.

Three key insights emerge. First, digital twins are predominantly framed through narratives of technological function that emphasize optimization, monitoring, and predictive capability. These narratives foreground engineering rationality while backgrounding institutional negotiation. Second, narratives of institutional roles tend to abstract decision-makers and compress governance processes, implying that actionable knowledge flows automatically from data visibility. Third, narratives of governance alignment position digital twins within existing procurement and performance structures, thereby reinforcing current institutional logic rather than questioning them. Together, these findings suggest that digital twin discourse performs an anticipatory governance function: it defines how the technology is expected to operate within institutions before those arrangements are explicitly designed. This insight contributes to narrative theory by showing how technological descriptions act as institutional scripts that shape future governance possibilities. It also contributes to infrastructure studies by highlighting that digital

transformation is not only a technical process but also a narrative one, in which assumptions about authority, responsibility, and coordination are constructed through narratives and discourses. Practically, the study suggests that policymakers and infrastructure organizations should treat digital twin adoption not only as a technical investment but as an institutional design challenge. If digital twins are to support strategic decision-making, cross-agency coordination, and public accountability, governance arrangements must be explicitly articulated rather than assumed to follow from technical capability.

The study has several limitations. It focuses on academic discourse rather than implementation contexts, meaning that it captures how digital twins are imagined rather than how they are enacted. Future research could compare these narratives with practitioner discourse, policy frameworks, or project case studies to examine how narrative assumptions translate into institutional practice. Comparative studies across infrastructure sectors or national contexts could also reveal whether governance assumptions differ depending on regulatory environments or political traditions.

Acknowledgments

We would like to acknowledge funding from NWO, DST, and FAPESP.

References

Agrawal, A., Thiel, R., Jain, P., Singh, V., & Fischer, M. (2023). Digital Twin: Where do humans fit in?. Automation in construction, 148, 104749.

Barat, S., Kulkarni, V., Clark, T., & Barn, B. (2022, October). Digital twin as risk-free experimentation aid for techno-socio-economic systems. In Proceedings of the 25th international conference on model driven engineering languages and systems (pp. 66-75).

Borodulin, K., Radchenko, G., Shestakov, A., Sokolinsky, L., Tchernykh, A., & Prodan, R. (2017, December). Towards digital twins cloud platform: Microservices and computational workflows to rule a smart factory. In Proceedings of the 10th IEEE/ACM International Conference on Utility and Cloud Computing (pp. 209-210).

Bowker, G. C. (2008). Memory practices in the sciences. MIT Press.

Bijker, W. E., & Law, J. (Eds.). (1994). Shaping technology/building society: Studies in sociotechnical change. MIT press.

Broo, D.G., and Schooling, J. (2023). Digital twins in infrastructure: definitions, current practices, challenges and strategies. International journal of construction management, 23(7), 1254–1263.

Coeckelbergh, M., & Reijers, W. (2016). Narrative technologies: A philosophical investigation of the narrative capacities of technologies by using Ricoeur's narrative theory. Human Studies, 39(3), 325-346.

Datta, A., Ninan, J., & Sankaran, S. (2020). 4D visualization to bridge the knowing-doing gap in megaprojects: An Australian case study. Construction economics and building, 20(4), 25-41.

Davidson, E. J. (2002). Technology frames and framing: A socio-cognitive investigation of requirements determination. MIS quarterly, 329-358.

Deng, M., Menassa, C.C., and Kamat, V.R. (2021). From BIM to digital twins: a systematic review of the evolution of intelligent building representations in the AEC-FM industry. Journal of information technology in construction, 26 (5), 58–83.

Deutschmann, A., Schier-Morgenthal, S., & Kaltenhäuser, S. (2021). Digitaler Airport Zwilling–Herausforderungen bei der digitalen Transformation von Flughäfen im Kontext des Airportmanagements. Automatisierungstechnik, 69(12).

Dyer Jr, W. G., & Wilkins, A. L. (1991). Better stories, not better constructs, to generate better theory: A rejoinder to Eisenhardt. Academy of management review, 16(3), 613-619.

Ernstsen, S. N., Whyte, J., Thuesen, C., & Maier, A. (2021). How innovation champions frame the future: Three visions for digital transformation of construction. Journal of construction engineering and management, 147(1), 05020022.

Fett, M., Wilking, F., Goetz, S., Wartzack, S., & Kirchner, E. (2025). Utilization of RFLP in V-Model methodology for the interdisciplinary development of Digital Twins. Forschung im Ingenieurwesen, 89(1), 28.

Foerster, A., Macintosh, A., & McDonald, J. (2015). Trade-offs in adaptation planning: protecting public interest environmental values. Journal of Environmental Law, 27(3), 459-487.

Gephart, R. P. (2004). Qualitative research and the Academy of Management Journal. Academy of Management Journal, 47(4), 454 – 462.

Goatly, A. (1997). The language of metaphors. Routledge.

Ham, Y., and Kim, J. (2020). Participatory sensing and digital twin city: updating virtual city models for enhanced riskinformed decision-making. Journal of management in engineering, 36 (3), 04020005.

Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power, 1-33.

Jiang, F., Ma, L., Broyd, T., & Chen, K. (2021). Digital twin and its implementations in the civil engineering sector. Automation in Construction, 130, 103838.

Kaewunruen, S., and Lian, Q. (2019). Digital twin aided sustainability-based lifecycle management for railway

turnout systems. Journal of cleaner production, 228, 1537–1551.

Keskin, B., Salman, B., and Koseoglu, O. (2022). Architecting a BIM-based digital twin platform for airport asset management: a model-based system engineering with SysML approach. Journal of construction engineering and management, 148 (5), 04022020.

Kober, C., Medina, F. G., Benfer, M., Wulfsberg, J. P., Martinez, V., & Lanza, G. (2024). Digital twin stakeholder communication: characteristics, challenges, and best practices. Computers in Industry, 161, 104135.

Korenhof, P., Blok, V., and Kloppenburg, S. (2021). Steering representations—towards a critical understanding of digital twins. Philosophy & technology, 34 (4), 1751–1773.

Kudriashov, N. S., Markov, S. A., & Potekhin, V. V. (2019). Adaptive Control System Synthesis Methods for Complex Manufacturing Objects. Annals of DAAAM & Proceedings, 30.

Lavikka, R., Kallio, J., Casey, T., & Airaksinen, M. (2018). Digital disruption of the AEC industry: Technology-oriented scenarios for possible future development paths. Construction management and economics, 36(11), 635-650.

López-Alfaro, C. V., Dincel, S., Nasar, W., Styve, A., Besenecker, U., & da Silva Torres, R. (2025). Urban Digital Twins as Tools for Sustainable Lighting Design: The Assessment of NorDark-DT. Buildings, 15(24), 4425.

Michalec, O. (2025). Models vs infrastructures? On the role of digital twins' hype in anticipating the governance of the UK energy industry. Environmental Science & Policy, 168, 104041.

Miles, M.B. & Huberman, A.M. (1994). Qualitative Data Analysis: An Expanded Sourcebook. Sage, Thousand Oaks.

Miraj, P., Wang, T., Koutamanis, A., & Chan, P. (2025). Organising digital twin in the built environment: a systematic review and research directions on the missing links of use and user perspectives of digital twin in Architecture, Engineering and Construction (AEC) sector. Construction Management and Economics, 43(6), 465-481.

Ninan, J., (2020). Online Naturalistic Inquiry in Project Management Research: Directions for Research. Project Leadership and Society. 1(1), 1-10

Ninan, J., Sergeeva, N., & Winch, G. (2022). Narrative shapes innovation: a study on multiple innovations in the UK construction industry. Construction management and economics, 40(11-12), 884-902.

Nochta, T., Wan, L., Schooling, J. M., & Parlikad, A. K. (2021). A socio-technical perspective on urban analytics: The case of city-scale digital twins. Journal of urban technology, 28(1-2), 263-287.

O'Dwyer, E., Pan, I., Charlesworth, R., Butler, S., & Shah, N. (2020). Integration of an energy management tool and digital twin for coordination and control of multi-vector smart energy systems. Sustainable Cities and Society, 62, 102412.

olde Scholtenhuis, L., & Soman, R. K. (2025, September). Trapped in technocentric thinking? Revisiting Digital Twins through a pragmatic framework. In Proceedings of the 41st Annual ARCOM Conference (Vol. 1, p. 3).

Papadonikolaki, E., van Oel, C., & Kagioglou, M. (2019). Organising and Managing boundaries: A structurational view of collaboration with Building Information Modelling (BIM). International journal of project management, 37(3), 378-394.

Pan, Y., and Zhang, L., (2021). A BIM-data mining integrated digital twin framework for advanced project management. Automation in construction, 124, 103564.

Sergeeva, N. & Ninan, J. (2023). Comparisons as a discursive tool: Shaping megaproject narratives in the United Kingdom. Policy and Society, 42(2), 197–211.

Solman, H., Kirkegaard, J. K., Smits, M., Van Vliet, B., & Bush, S. (2022). Digital twinning as an act of governance in the wind energy sector. Environmental science & policy, 127, 272-279.

Soman, R. K., Farghaly, K., Mills, G., & Whyte, J. (2025). Digital twin construction with a focus on human twin interfaces. Automation in Construction, 170, 105924.

Tripathi, N., Hietala, H., Xu, Y., & Liyanage, R. (2024). Stakeholders collaborations, challenges and emerging concepts in digital twin ecosystems. Information and Software Technology, 169, 107424.

Uhlenkamp, J. F., Hauge, J. B., Broda, E., Lütjen, M., Freitag, M., & Thoben, K. D. (2022). Digital twins: A maturity model for their classification and evaluation. IEEE Access, 10, 69605-69635.

Williams, R. (2006). Compressed foresight and narrative bias: Pitfalls in assessing high technology futures. Science as Culture, 15(4), 327-348.

Zibuschka, J., Ruff, C., Horch, A., & Roßnagel, H. (2020). A human digital twin as building block of open identity management for the internet of things. In Open Identity Summit 2020 (pp. 133-142). Gesellschaft für Informatik eV.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

BUILDING ENERGY SIMULATION MODULE: THE DESIGN, IMPLEMENTATION AND EVALUATION OF A DUAL-ASSESSMENT MASTERS MODULE

Divyanshu Sood[1], Donal Finn[1], James OḊonnell[2]
[1] School of Mechanical and Materials Engineering, University College Dublin, Ireland
[2] School of Engineering and Bernal Institute, University of Limerick, Limerick, Ireland

Abstract

Building energy simulation is a key competence in energy systems engineering education. Students are often challenged to appropriately apply core concepts in modelling workflows. This paper presents a novel curriculum design for masters level building energy module delivered through a combined lecture and project format. Cohort evidence includes pre and post comparison of project outcomes. Project outcomes show similar central tendency pre versus post periods, with reduced spread in the post period, consistent with more structured milestone-driven delivery. The paper provides a transferable module structure, workflow, and assessment rubric for building energy education using professional modelling toolchains.

Context and motivation

Energy use in buildings is central to the climate agenda and building professionals are increasingly expected to evaluate retrofit pathways using performance simulation modelling (Sood et al., 2025). Within tertiary education, a key challenge is that students need to need to acquire competence in building energy system concepts, yet they often struggle when these concepts need to be applied in a modelling workflow where outcomes depend on assumptions and controls. This topic directly aligns with five of the seventeen United Nations Sustainable Development Goals (SDGs): 7) Renewable Energy, 8) Good Jobs and Economic Growth, 11) Sustainable Cities and Communities, 12) Responsible Consumption and 13) Climate Action (Kartal et al., 2024). It is therefore of paramount importance that graduate students understand fundamental building physics, as well as building energy simulation know-how, so they are equipped with the required skills and experiences to make informed decisions during their professional careers.

A recurring concern, observed in project-based delivery, is that a subgroup of students engage late or inconsistently and therefore do not reach the workload threshold required for deeper analysis. This pattern aligns with reported observations in similar project-led modules. The curriculum design described here aims to reduce this failure mode through milestone-driven structure and explicit assessment criteria.

The module is delivered to masters level mechanical and energy systems engineering students in the final year of their programme. The module combines theory, worked examples, and a simulation retrofit project. The design intent is that students learn concepts in lectures and then use them in the project to build and defend a building retrofit recommendation. The summative weighting was 40% project and 60% examination prior to 2021, and was changed to 50% project and 50% examination from 2021 onward to strengthen sustained engagement with project-based learning. Module assessment comprises two components: an end of trimester exam (worth 50% of the the overall grade) and a project, an example of which is shown in Figure 1, for groups of two or three students (worth 50% of the overall grade). The same grade is awarded to all group members for the project component.

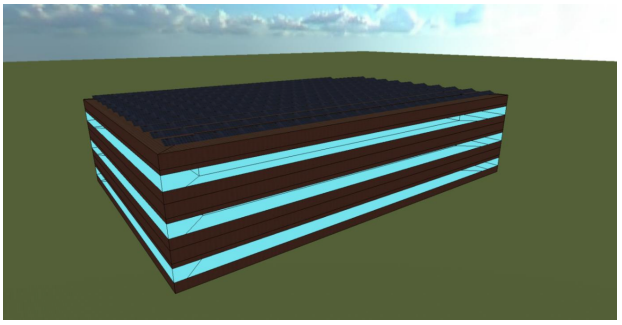


Figure 1: Example building project used for the group retrofit project.

Curriculum design approach

The curriculum design goal is to deliver deep pedagogical outcome through a balanced combination of lectures, problem sets, and project work (Gilchrist, 1999). The project is designed as a top-down building performance analysis and retrofit appraisal of an existing commercial building. The core threshold concept for the module is holistic building performance analysis through systems thinking, where building physics, HVAC systems operation, comfort, and economics are treated as linked rather than isolated topics.

The approach follows integrated course design principles, focusing on: learning goals, assessment and feedback, and aligned teaching activities. The design also draws on threshold concept theory (Timmermans and Meyer, 2019) and project based learning principles (Savery, 2006). A practical requirement is that the assessment

structure must encourage sustained engagement across the trimester rather than late engagement with the project component. This module has prerequisites and expects a considerable level of prior learning in the areas of thermodynamics and heat transfer. The key issue is that buildings can be viewed rather simplistically as static assets over which occupants have very little control but in reality buildings are dynamic and fulfil a range of functions that vary significantly over time. The lecture material, combined with related problem sets and exercises contribute to much of the formative learning within the module and in turn contribute to the group project that applies, analyses, evaluates and synthesises the course content. The end of trimester summative assessment also examines a deep understanding of course material but in a different way to the group project. Figure 2 is used to frame the design from both teacher and student perspectives.

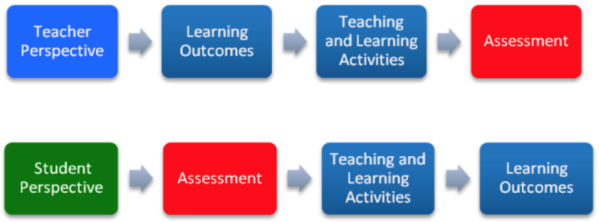


Figure 2: Curriculum design from teacher and student perspectives (adapted from (Biggs, 2003)).

Threshold concepts and learning objectives

The module is organised around a key number of threshold concepts that students must master to progress:

1. Key principles of building physics and heat transfer in buildings.
2. Psychrometrics and understanding of air-conditioning processes.
3. HVAC systems and control.
4. Building economic performance analysis.

The learning objectives and key concept map used to structure lecture content and project tasks are shown in Figure 3. The concept map is used as a reference throughout the trimester to reduce content fragmentation and to make linkages between topics explicit. The map is organised around a central question: how to ensure deep learning throughout the module. Learning objectives branch into four threshold concept areas: fundamental building physics, HVAC systems operation, building energy performance simulation, and the role and nature of occupancy in buildings. Prior learning in fluid mechanics, heat transfer, and thermodynamics is a stated prerequisite. Two summative assessment components emerge from the learning objectives: the end-of-year exam and the EBL group project. Both connect to programme outcomes and professional accreditation. The group project is designed for a variety of learning styles and is enabled by formative learning and in-class tutorials. It is assessed through presentations and links to potential industry and academic employers as external stakeholders. The exam ensures authenticity and transparency of individual learning. Assessment methods collectively facilitate skill development and transparency. The design is grounded in focused problem-based learning exercises and informed by relevant legislation, standards, and guides. Feedback continuously improves the design. Node colours identify type: green for the key question, orange for module-specific content, blue for programme-specific content, and yellow for external nodes.

Table 1 maps each intended learning outcome to the teaching activities through which it is developed and to the assessment component through which it is evaluated. The rubric criteria in Table 2 correspond directly to these outcomes: assumption traceability and plausibility checking map to outcomes 1 and 2; comfort interpretation maps to outcome 2; economic logic maps to outcome 4; and the oral defence maps to outcome 5.

Table 1: Constructive alignment: intended learning outcomes, teaching activities, and assessment methods.

Learning outcome	Teaching activity	Assessment
Building physics and heat transfer	Lectures wks 1–4; problem sets	Exam; baseline model
Psychrometrics and A/C processes	Lectures wks 5–7; chart exercises	Exam; comfort evaluation
HVAC systems and control	Lectures wks 8–10; tutorials 4–6	Exam; retrofit options
Economic performance (NPV)	Lectures wks 10–11; discounting exercise	Exam; economic analysis
Defend modelling decisions holistically	Mid-term gate; reasoning exercise	Oral defence (individual)

Lecture–project integration and milestone staging

The module is delivered through aligned lectures and project sessions over a 12 week teaching trimester. Lectures introduce concepts and methods. Project sessions require students to apply these concepts to the project case study through staged tasks. The staging is intentional. Without milestones, students tend to allocate effort toward software tasks and output generation rather than interpretation and diagnosis.

A student-centric overview of the module structure, including lecture content, project tutorials, and assessment milestones, is shown in Figure 8 in Appendix. The figure is intended to communicate what is expected, by when, and why each milestone exists. This covers 12 teaching weeks plus an exam period across six columns: week, topic, lecture topic, reading material, learning milestone, progress milestone, and assessment detail. Weeks 1 and 2 cover buildings and climate, the internal environment, and

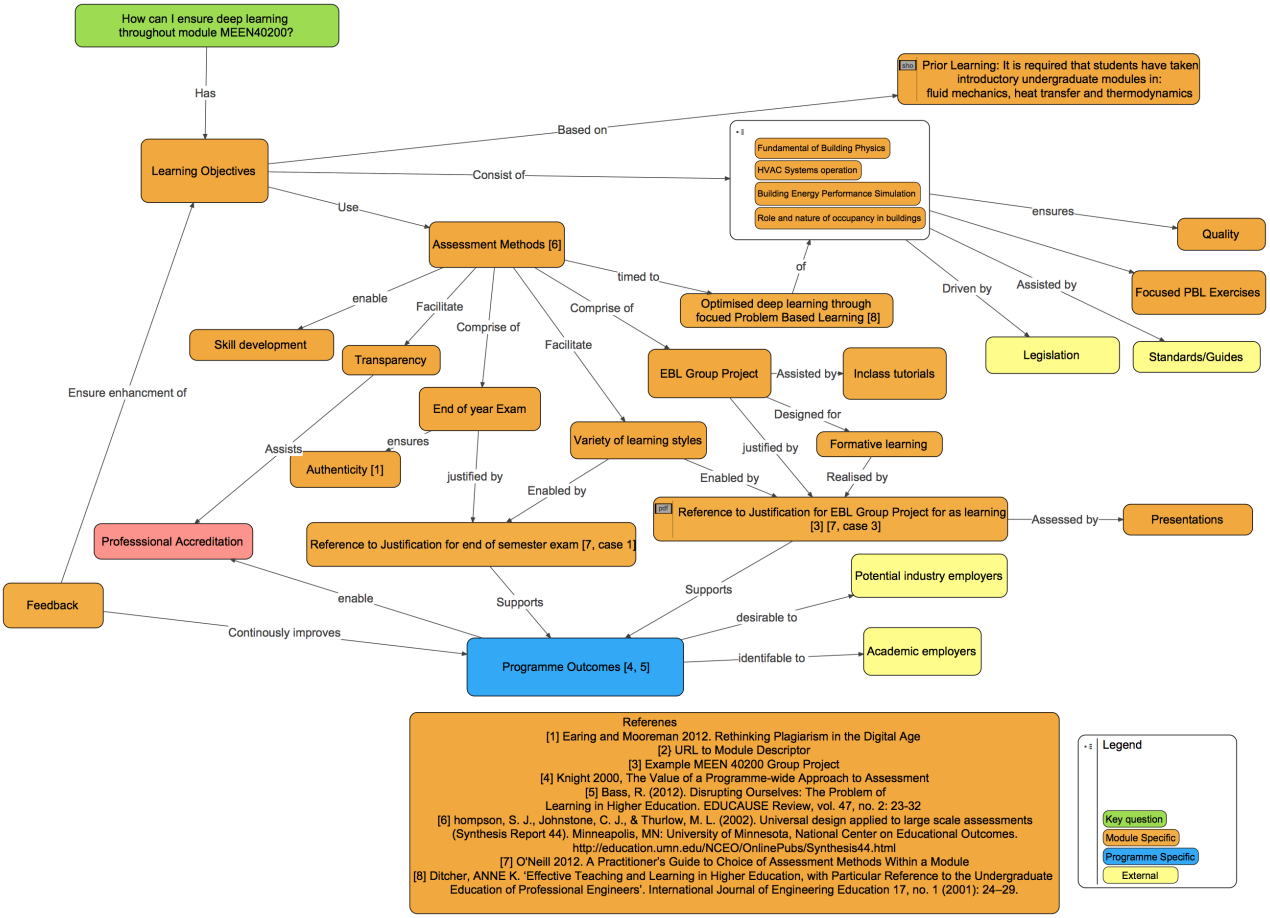


Figure 3: Learning objectives and key concepts used to structure the module.

building energy balance; the first project overview session establishes teams and the initial site. Week 3 introduces energy simulation and heat transfer; building fabric investigations are expected to be complete by its end. Week 4 covers internal gains, thermal mass, and thermal bridges; a formative self-assessment checkpoint occurs here, shown in orange. Week 5 covers daylighting and introduces psychrometrics; the key learning milestone is to understand the key principles of building physics. Week 6 covers natural ventilation and a formative problem set is released to students. Week 7 completes psychrometrics in practice; the base case model must be finished and the initial mid-term report submitted by the end of this week, which is the quality gate. After two study weeks, weeks 8–10 cover demand-side equipment, plant heating, heat pumps, and cooling; HVAC design alternatives are progressively completed during this period. Weeks 11 and 12 are reserved for make-up sessions and final project submission and presentation. The summative exam falls in the exam period. Row colours identify content type: white for lectures, blue for project tutorials, green for prerecorded lectures, light grey for reading material, yellow for key learning outcomes, purple for progress milestones, and orange for assessment components.

A week-by-week schedule is provided to students to support planning and to reduce drift in the project. An example is shown in Table **??** in Appendix.

Assessment design and feedback

Assessment components

The assessment model combines two summative components:

- An individual written examination to assess conceptual understanding and analytical competence under time constraints.
- A group retrofit project to assess applied modelling judgement and decision justification through a report and oral defence.

Formative feedback and quality gates

The project includes structured feedback loops. A mid-term formative deliverable requires groups to present baseline model behaviour, identify key drivers of performance, and propose retrofit pathways for the second half of the module. The deliverable is returned quickly to support iteration and project progression. This mid term checkpoint functions as a quality gate. It reduces late rework by identifying issues in schedules, internal gains,

controls, and zoning logic before option comparisons are finalised.

Oral defence

The final presentation includes technical assessment across individual group members. The questioning targets: assumption selection, plausibility checks, interpretation of comfort outputs, and the causal link between retrofit measures and changes in energy, comfort, and cost. Questioning is directed at each group member individually in turn. A student who has not engaged with the project in depth will most likely answer incomplete or vague to these questions at the required level, and the responses inform the overall project assessment alongside the written report. This is the primary mechanism through which individual competence is differentiated within the group grade.

Project workflow and evidence of operationalisation

Workflow overview

The project workflow follows five staged steps. In the first stage, baseline model establishment, the group configures and validates the building model against reference benchmarks to confirm that model behaviour is physically reasonable before any retrofit analysis begins. In the second stage, baseline diagnosis, energy drivers, comfort conditions, and operating patterns are identified and documented alongside the assumptions that underpin them. In the third stage, option definition, retrofit measures are selected and their technical feasibility assessed under the constraints of the building and its occupancy. In the fourth stage, option comparison, the selected measures are evaluated against the baseline on energy, comfort, and economic criteria using net present value analysis. In the fifth stage, final justification, the group defends the chosen retrofit pathway and explains the reasoning behind the assumptions, results, and conclusions. The workflow is designed so that lecture content is operationalised at each stage. For example, psychrometrics and HVAC concepts appear in control intent and comfort interpretation, and discounting appears in NPV-based comparison rather than as a stand-alone calculation.

The simulation software used for the project is IES Virtual Environment (IES VE) (IES, 2026). Prior to IES VE, Simergy was used. The module design is not dependent on either tool specifically. Any professional-grade building energy simulation environment that supports multi-zone thermal and HVAC modelling would support the same workflow.

The module structure is transferable to other institutional contexts. Three factors are relevant when adapting the design. The first is class size, which affects how group supervision and oral defence logistics are managed. Larger cohorts may require more structured oral defence scheduling and more explicit supervision records to support individual contribution monitoring. The second is software availability; the workflow does not depend on IES VE specifically. The third is curriculum prerequisites: the module assumes prior learning in thermodynamics and heat transfer.

Example of holistic building performance analysis

The module leverages the concept of Problem Based Learning (PBL) and this requires three key elements (Lloyd-Jones et al., 1998): the students must be presented with a problem or initiating trigger at the start of the process, learning takes place through the research of issues identified in the initial tutorial, and knowledge is used to further the understanding of the trigger situation. This approach is used within individual concepts but overarching threshold concept of holistic building performance analysis through systems of systems thinking is achieved through a **project** based learning approach for the group project.

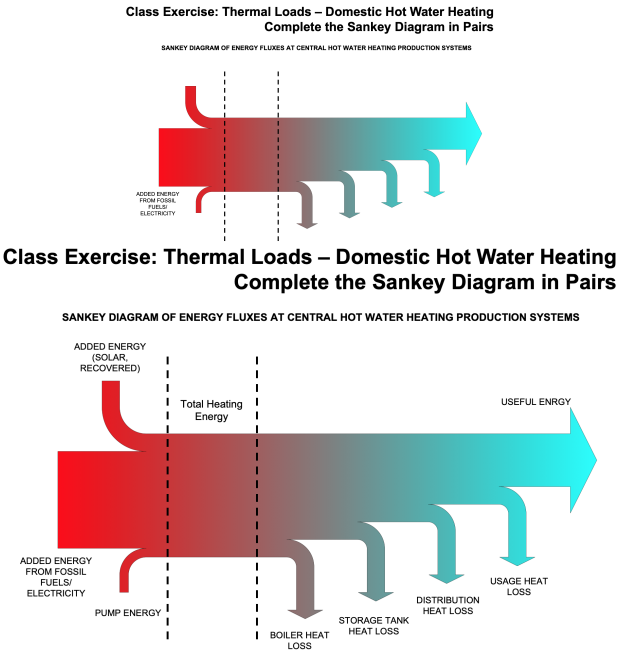


Figure 4: Example exercise used to develop holistic performance reasoning: (top) student attempt, (bottom) reference solution.

Figure 4 illustrates an in-class exercise used to develop the threshold concept of holistic building performance analysis. Students begin with an incomplete reasoning chain and must account for direct and indirect changes to energy flows and building physics under a trigger event. The exercise is used to prepare students to interpret simulation results as a system response rather than as isolated outputs.

Transferable rubric structure for the project

To make expectations explicit and to support consistent grading, the project assessment can be structured around decision-quality criteria. Table 2 provides a compact rubric structure consistent with the milestone design and

the intended learning outcomes. The intent is to reward assumption transparency, diagnosis, and interpretation rather than output volume.

Table 2: Rubric structure for the retrofit project report and defence.

Criterion (evidence in report, model, and defence)	Weight (%)
Baseline model definition and traceable assumptions	20
Plausibility checks and diagnosis (benchmarks and drivers)	20
Retrofit option definition and feasibility under constraints	15
Comfort evaluation and interpretation against stated criteria	15
Economic analysis (NPV structure and consistency checks)	15
Communication and oral defence (technical justification)	15

Supporting cohort evidence

Data sources and methodology

The evaluation draws on two data sources: project grade records spanning eleven years and institutional module feedback surveys covering four delivery years. The analysis is descriptive. The intent is to characterise outcome patterns across the pre-change and post-change periods. Two summary measures are reported: central tendency, represented by the median of the midpoint-proxy distribution within each period, and spread, represented by the interquartile range. These measures are appropriate for ordinal data and are robust to the unequal band widths in the institutional grade scale. No parametric tests are applied because the pre-period data consist of grade-frequency counts rather than individual records, and because the comparison is across cohorts of different sizes rather than matched individuals. The comparison is therefore descriptive and is intended to characterise whether the pattern of outcomes changed between periods in a way that is consistent with the intended effects of the module redesign. The evaluation is a design-led case study. It covers one module at one institution over one programme. No control group is available because the module redesign was applied to the full cohort in each year. The pre-post comparison conflates the effect of the assessment redesign with other changes over the same period, including changes in cohort composition, programme structure, and the shift to online delivery during 2020/21. These confounds are acknowledged and the cohort evidence is presented as supporting rather than confirmatory.

Project outcomes

To provide supporting evidence for the assessment redesign, project outcomes were compared across two periods: 2014–2021 (pre-change) and 2022–2025 (post-change). The pre-period records were available as grade-frequency counts from archived spreadsheets, while post-period records include numeric project marks for 2022–2024 and letter grades for 2025. Where numeric marks were available, they were converted to letter grades using the institutional grading scheme. To enable a single visual comparison across the full cohort history, each letter grade was then assigned the midpoint of its numeric band as a 0–100 proxy (UCD, 2026). The resulting scale is ordinal and descriptive: the midpoint values are used only to characterise the distribution shape and spread and are not used to make proportional claims about performance differences between grade levels. The comparison is therefore descriptive and is intended to characterise outcome patterns rather than to estimate causal learning effects.

Figure 5 shows that the central tendency of project outcomes is similar across periods, while the spread is lower in the post period. The reduced spread is consistent with a more structured project delivery and milestone-driven pacing, where most groups reach an expected standard of analysis and reporting (Figure 6).

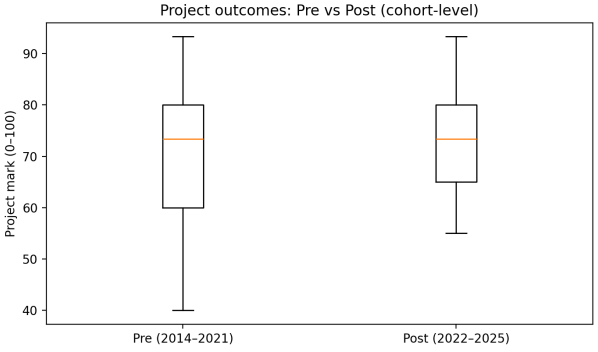


Figure 5: Project outcomes pre (2014–2021) versus post (2022–2025). Pre-period values are derived from grade-frequency records mapped to a 0–100 proxy for visual comparability.

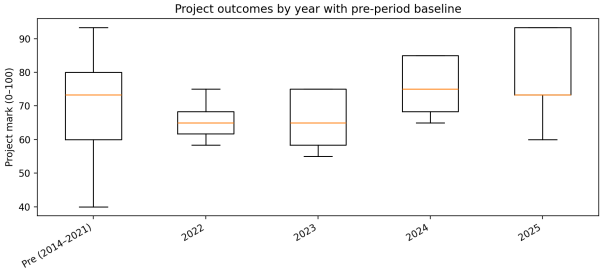


Figure 6: Project outcomes by year for 2022–2025 shown against the aggregated pre-period baseline (2014–2021).

Student module feedback

Module feedback surveys were administered to students across all delivery years using the institutional survey instrument. The survey covers five core questions on a Likert scale from one to five: Q1 subject understanding, Q2 assessment relevance, Q3 learning outcomes achieved,

Q4 teaching supported learning, and Q5 overall satisfaction. Table 3 reports the mean scores for MEEN40200 across four trimester cohorts alongside the school-level MEEN benchmark for comparison. Figure 7 shows the year-by-year trend.

Table 3: Mean Likert scores (1–5) for MEEN40200 across four delivery years and MEEN school benchmark (mean of Q1–Q5 per year).

Year	Q1	Q2	Q3	Q4	Q5	n
2020/21	4.50	4.00	3.75	4.00	3.50	4
2021/22	5.00	5.00	5.00	5.00	5.00	2
2022/23	4.67	4.33	4.33	5.00	4.33	3
2023/24	5.00	5.00	4.00	4.00	4.00	1
MEEN avg	4.07	4.07	3.86	3.83	3.73	–

Across all four years, module scores on Q1 to Q5 are at or above the MEEN school average. Q4 (teaching supported learning) and Q1 (improved subject understanding) are consistently the strongest results. Response rates were lowacross all years, ranging from 4% to 17%, which is a known limitation of voluntary institutional survey systems. The feedback data are therefore treated as corroborating evidence rather than a rigorous evaluation of student experience or learning outcomes.

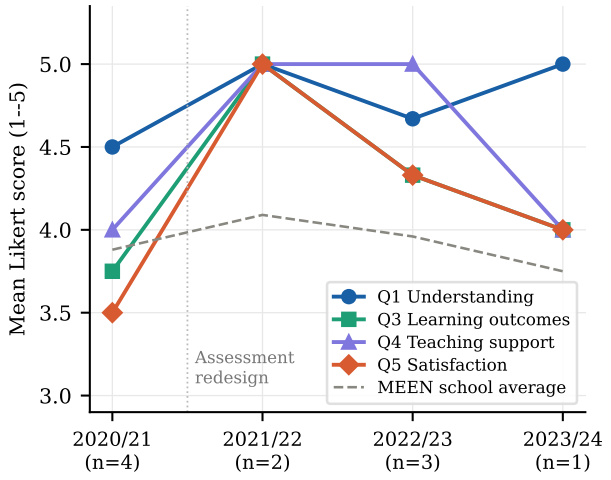


Figure 7: Mean student Likert responses for MEEN40200 across four delivery years compared against the MEEN school average. Response counts range from $n = 1$ to $n = 4$; results should be interpreted as corroborating evidence only.

Open-ended responses from 2020/21 provide qualitative context for one of the key design decisions described in this paper. Students identified the building energy simulation software as a source of significant friction, reporting that a large amount of time in the first weeks was spent on model construction rather than on analysis. This feedback informed the decision to provide a pre-built working model from the outset of the project and to shift early milestone tasks toward diagnosis and interpretation, as noted in the Workflow overview section above.

Implementation considerations

Three implementation considerations have shaped the module design. First, students can spend time on low-value tasks that do not improve the validity of conclusions, particularly when early modelling tasks are overly open-ended. This has been addressed by providing a working base model and by using milestones that prioritise baseline behaviour, diagnosis, and interpretation. Second, software issues can shift attention away from learning objectives. This has been mitigated by reducing the amount of custom model construction required early in the trimester and by placing emphasis on analysis and justification in assessment criteria. Third, students differ in study patterns. The milestone design and mid-term checkpoint reduce the risk that project work is deferred until late in the semester.

Implications for education and assessment

The lecture - project structure supports an explicit pathway from theory to application. The design makes key concepts reappear as modelling decisions and report arguments, and uses milestones to support iteration. For competency assurance, the dual assessment structure provides two complementary views: conceptual reasoning under time constraints and applied modelling judgement evidenced through a report and oral defence. If minimum competence is required in both domains, a practical policy option is to require a minimum pass standard in both components. If improved grading resolution is required within the project, the rubric weight can be shifted further toward assumption traceability, plausibility checking, and interpretation. Another option could be the application of peer review within each project group.

Conclusions

This paper presented a curriculum design for a masters level building energy systems module delivered through a coupled lecture-project format and assessed through an individual written examination and a group retrofit simulation project. The design is organised around threshold concepts and staged milestones, including a mid-term feedback and a final oral defence and reporting. The intent is that lecture content is not treated as separate theory, but is applied in the project through explicit assumptions, plausibility checks, and reporting that links model outputs to energy, comfort, and economic reasoning.

The contribution to the teaching domain is an explicit, transferable design pattern for simulation-based building energy education. The paper specifies how a lecture sequence can be aligned to a project workflow from the start of the semester, how milestones can be used to stabilise baseline models before retrofit comparison, and how questioning can be used to assess the reasoning behind modelling choices and interpretation. This addresses a recurring difficulty in simulation teaching: students may be able to generate outputs, yet struggle to explain why results change, whether outputs are plausible, and how con-

clusions follow from assumptions and controls.

The cohort evidence provides practical insight into assessment behaviour under this design. A pre and post comparison of project outcomes indicates similar central tendency across periods, with reduced spread in the post period. This is consistent with a more structured, milestone-driven delivery in which most groups reach the expected standard of analysis and understanding. At the same time, examination outcomes provide a stronger basis for individual differentiation, which supports the rationale for dual assessment when programme requirements include both conceptual mastery and applied modelling judgement. The implication is that project and examination components should be treated as complementary measures rather than substitutes, and that overall grades alone are not sufficient to infer competence because weighting policies and component dispersion affect how final outcomes are formed.

For future teaching practices, the results support three specific recommendations. First, define project marking around decision quality criteria. Second, retain an early quality gate with timely feedback to reduce late-stage rework and protect time for interpretation and justification. Third, where minimum competence is required in both domains, consider minimum thresholds for each component to avoid systematic compensation. Together, these recommendations frame building energy simulation education as a professional reasoning task supported by a structured project workflow and an assessment model that makes competence observable.

Acknowledgments

This publication has emanated from research supported partly by a grant from the Sustainable Energy Authority of Ireland (SEAI) under the Research, Development and Demonstration Funding Programme Grant No. 22/RDD/822. The opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the SEAI. The authors acknowledge teaching assistant support and the student cohort whose anonymised outcomes informed the cohort summary.

References

Biggs, J. (2003). Teaching for Quality Learning at University.

Gilchrist, M. D. (1999). Manufacturing & design with engineering polymers: Educational aspects of a specialist dissertation in a part-time elective postgraduate mechanical engineering degree course for industry-based students.

IES (2026). Integrated Environmental Solutions | IES.

Kartal, M. T., Santosh, M., Ulussever, T., Pata, U. K., and Kılıç Depren, S. (2024). A comprehensive review of progress in sustainable development goals from energy and environment perspectives. Energy Strategy Reviews, 56:101550.

Lloyd-Jones, Margetson, and Bligh (1998). Problem-based learning: a coat of many colours. Medical Education, 32(5):492–494.

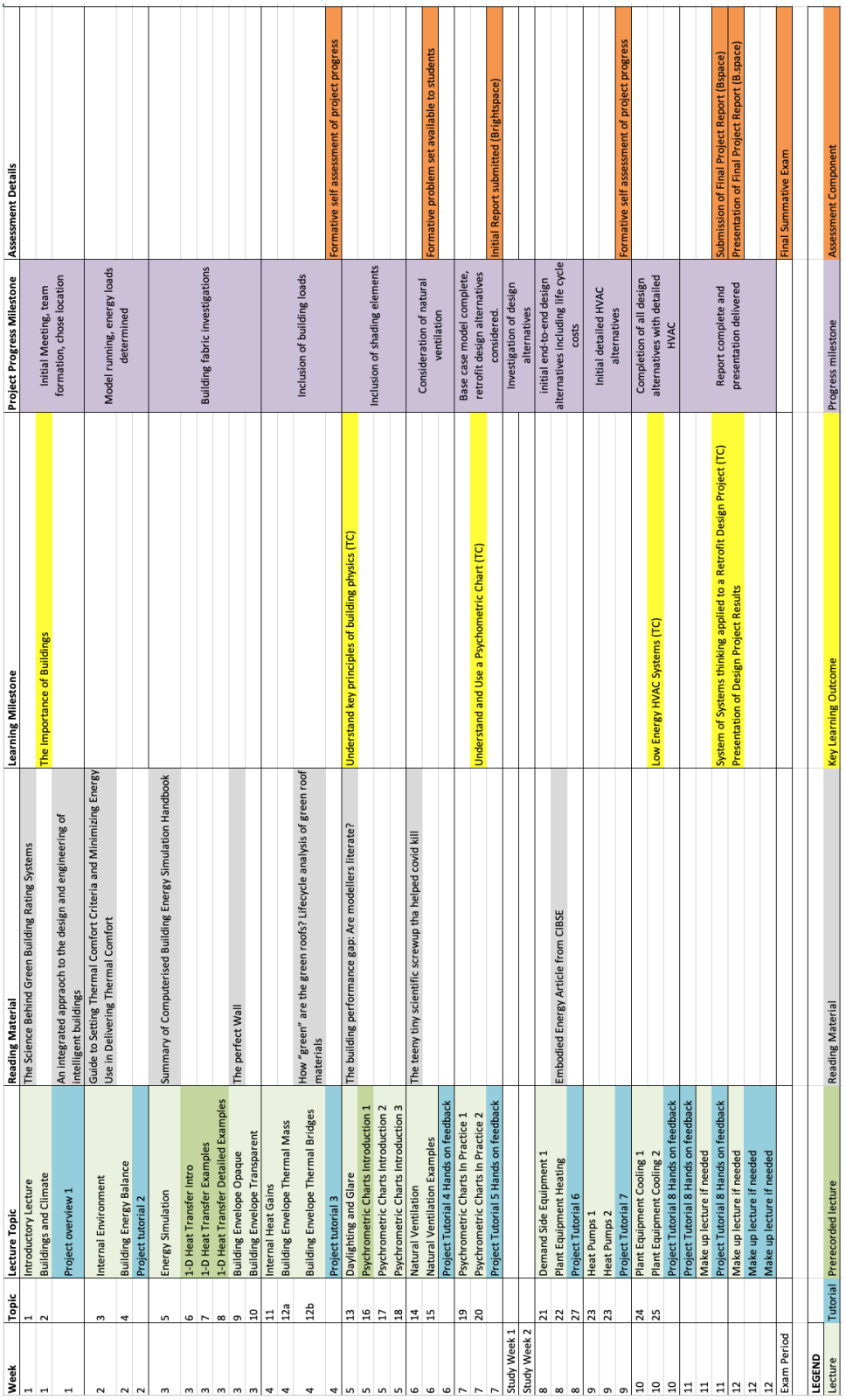
Savery, J. (2006). Overview of Problem-based Learning: Definitions and Distinctions. Interdisciplinary Journal of Problem-Based Learning, 1(1).

Sood, D., Alhindawi, I., Ali, U., Finn, D., McGrath, J. A., Byrne, M. A., and O'Donnell, J. (2025). Machine learning based multi-objective optimisation of energy consumption, thermal comfort and CO2 concentration in energy-efficient naturally ventilated residential dwellings. Building and Environment, 267:112255.

Timmermans, J. A. and Meyer, J. H. F. (2019). A framework for working with university teachers to create and embed 'Integrated Threshold Concept Knowledge' (ITCK) in their practice. International Journal for Academic Development, 24(4):354–368.

UCD (2026). Understanding Grades - UCD Current Students.

Appendix



Week	Topic	Lecture Topic	Reading Material	Learning Milestone	Project Progress Milestone	Assessment Details
1	1	Introductory Lecture	The Science Behind Green Building Rating Systems	The Importance of Buildings	Initial Meeting, team formation, chose location	
1	2	Buildings and Climate				
1		Project overview 1	An integrated appraoch to the design and engineering of intelligent buildings			
2	3	Internal Environment	Guide to Setting Thermal Comfort Criteria and Minimizing Energy Use in Delivering Thermal Comfort		Model running, energy loads determined	
2	4	Building Energy Balance				
2		Project tutorial 2				
3	5	Energy Simulation	Summary of Computerised Building Energy Simulation Handbook		Building fabric investigations	
3	6	1-D Heat Transfer Intro				
3	7	1-D Heat Transfer Examples				
3	8	1-D Heat Transfer Detailed Examples				
3	9	Building Envelope Opaque	The perfect Wall			
3	10	Building Envelope Transparent				
4	11	Internal Heat Gains				
4	12a	Building Envelope Thermal Mass				
4	12b	Building Envelope Thermal Bridges	How "green" are the green roofs? Lifecycle analysis of green roof materials		Inclusion of building loads	
4		Project tutorial 3				Formative self assessment of project progress
5	13	Daylighting and Glare	The building performance gap: Are modellers literate?	Understand key principles of building physics (TC)		
5	16	Psychrometric Charts Introduction 1			Inclusion of shading elements	
5	17	Psychrometric Charts Introduction 2				
5	18	Psychrometric Charts Introduction 3				
6	14	Natural Ventilation	The teeny tiny scientific screwup tha helped covid kill		Consideration of natural ventilation	Formative problem set available to students
6	15	Natural Ventilation Examples				
6		Project Tutorial 4 Hands on feedback				
7	19	Psychrometric Charts In Practice 1			Base case model complete, retrofit design alternatives considered.	
7	20	Psychrometric Charts In Practice 2		Understand and Use a Psychometric Chart (TC)		Initial Report submitted (Brightspace)
7		Project Tutorial 5 Hands on feedback				
Study Week 1					Investigation of design alternatives	
Study Week 2						
8	21	Demand Side Equipment 1	Embodied Energy Article from CIBSE		initial end-to-end design alternatives including life cycle costs	
8	22	Plant Equipment Heating				
8	27	Project Tutorial 6				
9	23	Heat Pumps 1			Initial detailed HVAC alternatives	
9	23	Heat Pumps 2				
9		Project Tutorial 7				Formative self assessment of project progress
10	24	Plant Equipment Cooling 1		Low Energy HVAC Systems (TC)	Completion of all design alternatives with detailed HVAC	
10	25	Plant Equipment Cooling 2				
10		Project Tutorial 8 Hands on feedback				
11		Project Tutorial 8 Hands on feedback				
11		Make up lecture if needed				
11		Project Tutorial 8 Hands on feedback		System of Systems thinking applied to a Retrofit Design Project (TC)	Report complete and presentation delivered	Submission of Final Project Report (Bspace)
12		Make up lecture if needed		Presentation of Design Project Results		Presentation of Final Project Report (B.space)
12		Make up lecture if needed				
12		Make up lecture if needed				
Exam Period						Final Summative Exam

LEGEND						
Lecture	Tutorial	Prerecorded lecture	Reading Material	Key Learning Outcome	Progress milestone	Assessment Component

Figure 8: Student-centric view of module structure including lecture content, project tutorials, milestones, and assessments.

Extended Abstracts: Poster Presentations

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

ANALYSING GREEN AND DIGITAL SKILLS IN THE TWIN TRANSITION

Farina Koller, Niels Bartels, and Claudia Ziller
TH Köln, University of Applied Sciences, Cologne, Germany

Abstract

The building sector is undergoing a twin transition that combines digital and sustainable transformation. This process requires professionals with integrated green and digital skills. However, their systematic integration into education and professional practice remains limited. This paper addresses this gap through a qualitative case study based on 53 expert interviews and a validation workshop in the Cologne metropolitan region. The findings reveal significant competency gaps and form the basis for a framework supporting the integrated development of green and digital skills.

Introduction

The building sector accounts for approximately 35 % of energy consumption and 30 % of CO2 emissions in Germany (United Nations, 2024).

To address these challenges, the sector is undergoing a twin transition, that combines digitalisation and sustainability (Hofmann et al., 2023). This transformation requires integrated green and digital skills, yet their implementation in education and practice remains insufficiently understood (Ziegler et al., 2023).

This study analyses existing and required skills and explores their integration into higher education.

Design and method

A qualitative research design combining semi-structured expert interviews and a validation workshop was applied. Stakeholders from planning, construction, operation, education, and regulation were included. In total, 53 interviews were analysed using qualitative content analysis (Kuckartz, 2018) and discussed in a workshop with 30 participants from academia and industry. In this study, green skills are defined as job-related skills supporting sustainable construction, while digital skills comprise the effective use of digital technologies and data throughout the building lifecycle, as shown in Figure 1.

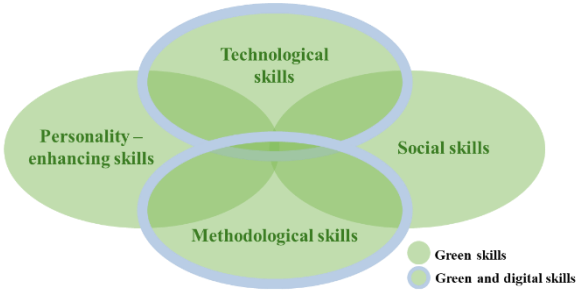


Figure 1: Used definition of green and digital skills

Results

The findings reveal a mismatch between existing and required skills. While basic technical green skills are generally present, they remain fragmented and are rarely integrated into lifecycle-oriented approaches.

A major deficit concerns methodological skils. Experts highlighted limited capabilities in applying lifecycle assessment, lifecycle costing, and data-based sustainability evaluations. Analytical and evidence-based decision-making skills are often lacking.

These deficits are closely linked to insufficient digital skills. Although digital are increasingly available, their potential for supporting sustainability goals remains underutilised. Skills related to digital planning tools, data interpretation, and simulation-based analyses are frequently underdeveloped and rarely connected to sustainability objectives.

The analysis also identified significant gaps in interdisciplinary skills. Strong discipline-specific expertise exists, but knowledge exchange across professions and lifecycle phases remains limited. As a result, coordination challenges emerge and integrated solutions are difficult to implement. Communication and collaboration skills were identified as critical requirements for successful transformation processes.

The analysis highlights the important role of higher education in preparing professionals for the twin transition but reveals significant shortcomings. Sustainability and digitalisation are often treated as isolated topics rather than integrated skills. In addition, interdisciplinary and practice-oriented learning formats remain limited, and teaching staff frequently lack up-to-date expertise. As a result, current education systems are not yet fully aligned with the competency requirements of the twin transition.

The findings indicate that the prevailing education and training systems are not yet fully aligned with the skills required for the twin transition, particularly with regard to the integration of green and digital skills as interconnected domains.

Conclusion and recommendations for action

The study demonstrates that the twin transition is constrained by a mismatch between existing and required skills. Green and digital skills should not be considered separate domains; rather, digitalisation acts as an important enabler of sustainability.

To address current deficits, green and digital skills should be embedded as cross-cutting skills throughout higher education and professional training. Key measures include:

- integration of sustainability and digitalisation across curricula
- strengthening data-driven and methodological skills
- expanding interdisciplinary and practice-oriented learning formats
- fostering close collaboration between academia and industry

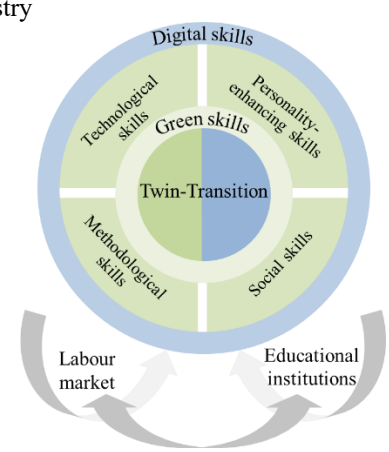


Figure 2: Framework for the integrated development of green and digital skills

The conceptual underpinnings of this integration are delineated by the framework depicted in Figure 2. It is recommended that future research efforts concentrate on the implementation and evaluation of integrated competence models.

Acknowledgments

This study is part of a project funded by the RheinEnergie Foundation in cooperation with the University of Applied Sciences Cologne and the University of Cologne.

References

Hofmann, J., Ricci, C., Kleinewefers, C., & Laurenzano, A. (2023). Doppelte Transformation. https://doi.org/10.11586/2023001

Kuckartz, U. (2018). Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung (4th, revised edition). Beltz.

United Nations. (2024). 2023 Global Status Report for Buildings and Construction: Beyond foundations - Mainstreaming sustainable solutions to cut emissions from the buildings sector. United Nations Environment Programme. https://doi.org/10.59117/20.500.11822/45095

Ziegler, P., Eder, A., & Wöhl, W. (2023). Green Skills im Aufwind? Zur Bedeutung von "grünen" Kompetenzen und Qualifikationen für die Ausbildung an Berufsbildenden Höheren Schulen, Universitäten und Fachhochschulen in Österreich. AMS Info(645/646).

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

BLOCKCHAIN-ENABLED DIGITAL TWINS FOR ENHANCED LIFECYCLE MANAGEMENT IN BUILDING HANDOVER

Yutian Yi[1], Davide Lombardi[1] and Theodoros Dounas[2]
[1]Xi'an Jiaotong-Liverpool University, Suzhou, China
[2]Heriot-Watt University, Edinburgh, Scotland

Abstract

Building handover is a critical stage in the building lifecycle, yet project information is often fragmented, difficult to verify, and hard to reuse in later operation and maintenance. Digital Twins (DTs) can improve lifecycle visibility, while BC can strengthen trust and traceability(Dounas et al., 2020). However, current studies rarely combine these with semantic knowledge representation. This work proposes a Blockchain-enabled semantic DT framework for building handover, using a five-layer architecture and a knowledge graph to structure building information. A lightweight prototype demonstrates trusted data admission and auditable event recording. The contribution is a practical and reproducible framework for improving lifecycle continuity beyond handover.

Introduction

Building handover is the point where responsibility moves from design and construction teams to owners, operators, and facility managers (Adu-Amankwa et al., 2023). At this stage, large amounts of information must be transferred, including room data, asset details, commissioning evidence, sensor information, and maintenance-related records. In practice, this transfer is often incomplete, inconsistent, or difficult to verify. As a result, valuable information is lost or underused after project delivery, which weakens later operation and maintenance (Boje et al., 2020).

DTs offer a way to organise and connect lifecycle information in a living digital environment (Opoku et al., 2021). Blockchain (BC) can complement DTs by providing tamper-evident records and stronger trust in shared data. Yet many current DT–BC studies focus mainly on data exchange or on-chain records, with limited attention to semantic structure (Dounas et al., 2023). Without semantic enrichment, handover information remains difficult to interpret across systems and lifecycle stages.

This poster proposes a BC-enabled semantic DT framework for building handover. The framework combines a knowledge graph based on BOT, BRICK, and SAREF with BC-supported governance to improve trusted information transfer at the handover stage. The system is organised into five layers: data acquisition, data transmission, semantic, BC, and application.

Framework overview

The framework shown as fig. 1 is designed to connect physical building information, semantic context, and trusted digital services across the handover process.

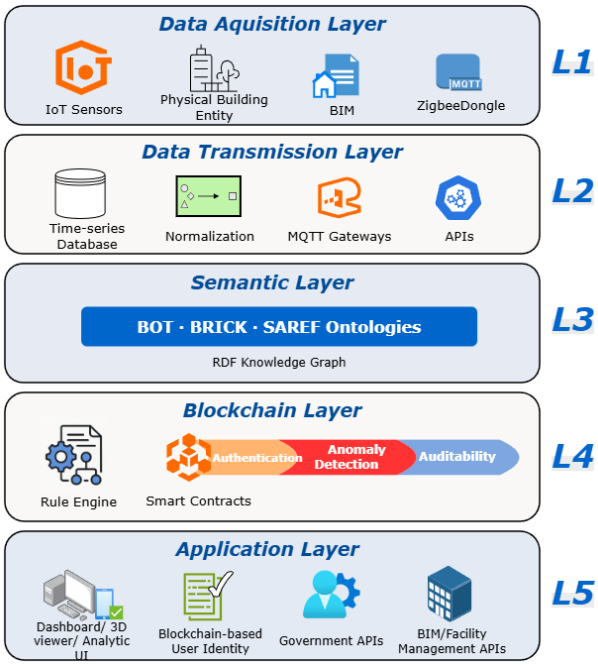


Figure 1: Methodology flowchart

The data acquisition layer gathers building and asset information from physical entities, BIM-related data, and IoT devices. The data transmission layer moves this information through standardised workflows and prepares it for processing. The semantic layer uses a knowledge graph to connect spaces, devices, and measured properties in a machine-readable form. The BC layer adds trust by verifying data sources and recording important lifecycle events. The application layer supports dashboards, analytics, and service integration for later building use. This structure allows data to move from fragmented project delivery information toward a more connected and trustworthy handover process.

Methodology

The proposed methodology integrates DT and BC through a structured three-stage workflow: data acquisition, BC access, and system enrichment, as illustrated in Fig. 2.

First, real-time data are collected from physical assets and IoT sensors, including temperature, energy, and occupancy. These data are transmitted, pre-processed, and partially handled at the edge to reduce latency and ensure a consistent format.

Second, the processed data enter the BC access stage, where key information is cryptographically hashed and anchored to the BC. Smart contracts are used to automate permissions, validation, and transaction logic, while an immutable ledger records critical events to ensure traceability and trust.

Finally, in the system enrichment stage, validated data are integrated into the Digital Twin model, enabling synchronised updates of building states. This enriched DT environment supports advanced analytics, simulation, and user-oriented visualisation. Through this workflow, the methodology ensures that building data are not only collected and processed efficiently but also verified and made trustworthy for lifecycle management.

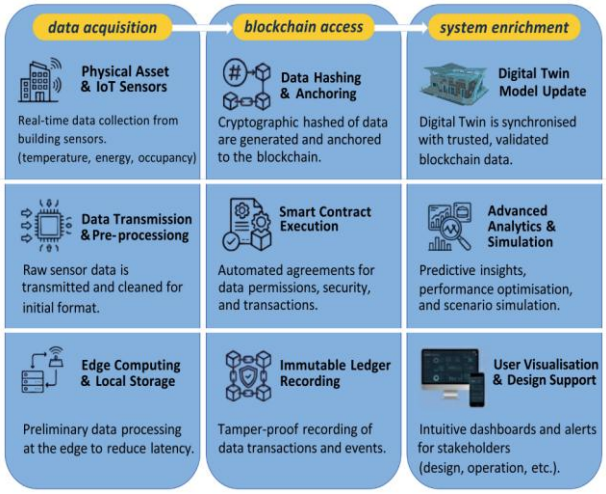


Figure 2: Methodology flowchart

Simulation and experiment

A lightweight prototype was developed in a containerised environment to demonstrate the feasibility of the framework. The prototype combines a knowledge graph, a time-series database, a workflow engine, and a local BC service. Rather than trying to represent all handover documents, the prototype focuses on one key principle: trusted entry of lifecycle information into the DT.

A semantic knowledge graph was prepared to represent rooms, devices, and measured properties. BC services were used to verify whether an information source was authorised before its data could enter the DT workflow.

Conclusion

This study presents a BC-enabled semantic DT framework for building handover. The proposed five-layer structure integrates semantic enrichment and BC-supported trust into one lightweight and reproducible system. The proof of concept shows that authorised information sources can be admitted while unknown sources can be blocked before entering the DT workflow.

The framework contributes a practical way to strengthen lifecycle continuity at handover by improving both interoperability and trust. Future work will extend the approach to richer handover data types, broader stakeholder workflows, and AI-supported lifecycle services.

References

Adu-Amankwa, N. A. N., Rahimian, F. P., Dawood, N., & Park, C. (2023). Digital Twins and BC technologies for building lifecycle management. Automation in Construction, 155, 105064.

Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic Construction Digital Twin: Directions for future research. Automation in Construction, 114. Scopus. https://doi.org/10.1016/j.autcon.2020.103179

Dounas, T., Jabi, W., & Lombardi, D. (2020). Smart contracts for decentralised building information modelling. https://rgu-repository.worktribe.com/output/968307

Dounas, T., Lombardi, D., Vele, J., & Prokop, Š. (2023). A CRYPTO-TWIN FRAMEWORK FOR THE AEC INDUSTRY - ENABLING DIGITAL TWINS WITH BLOCKCHAIN TECHNOLOGIES. Architecture and Planning Journal (APJ), 28. https://doi.org/10.54729/2789-8547.1215

Opoku, D., Perera, S., Osei-Kyei, R., & Rashidi, M. (2021). Digital twin application in the construction industry: A literature review. JOURNAL OF BUILDING ENGINEERING, 40. https://doi.org/10.1016/j.jobe.2021.102726

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

FROM CAPTURE TO CO-CONSTRUCTION: A SCOT- INFORMED ANALYSIS OF DIGITAL TWIN PRACTICES IN CEMETERY SCANNING

Ekpenyong Ekpo, Nikolic Dragana, and Green Martin
University of Reading, United Kingdom

Abstract

Although digital twins (DT) have been promoted as technical solutions in facility management, their adoption is a social process. This study addresses the gap in the socio-technical understanding of how DTs are enacted in practice. To understand how practices emerge, this study uses Social Construction of Technology (SCOT) framework and its interpretative flexibility component to present the development of a DT of a cemetery. Using vignette-based ethnography, the findings illustrate how the DT creation process is shaped by user interpretations, as well as social and environmental conditions. Overall, this study contributes an empirically grounded sociotechnical account of a DT creation.

Introduction

Digital twins have attracted growing attention in facility management because of their data centric pairing of digital models to physical assets. However, much of the current discourse largely focuses on technical capacities such as data integration, visualization and system performance (Asare Kofi et al., 2024). While these technology-focused approaches are helpful, adoption of digital twins in practice, as of any novel technologies, is rarely straightforward (Nikolić and Ewart, 2025, Schweber and Harty, 2010). Hence, processes and practices for developing and using digital twins are often fluid and emergent before they become routines.

To understand how these processes and practices unfold, this study draws on the SCOT framework, and using vignette-based evidence, presents moments were meanings, decisions, and practices were negotiated.

Aim and objectives

To explore how practices emerge, a pilot case study was conducted to create a digital twin for a local cemetery. Specifically, the study aimed to:

1. Examine how social constructs shape the DT creation process.
2. Identify moments of negotiation and adaptation that occur during the scanning process.

Understanding how these practices emerge is critical for moving beyond a deterministic view of DT implementation towards frameworks that support their situated adoption.

Methodology

This study examines the DT creation as a sociotechnical process and using SCOT as a lens, presents the scanning process of a local cemetery as a negotiation between technology, site conditions, environment and users. The cemetery as a space rich in history and meaning offers a pertinent case study to explore social constructs such as purpose, privacy concerns, and interactions between diverse stakeholders.

The study adopts interpretive reflexivity as a key component of SCOT (Bijker, 2015), both as a theoretical position and a methodological strategy. Ethnographic vignettes, which represent short stories that describe situations (Azman and Mahadhir, 2017), are adopted to foreground moments where technology's meaning and use were negotiated.

Results

The findings present reflexive ethnographic vignettes that highlight the emergent nature of practices surrounding the creation of the digital twin of a local cemetery.

Scan density

Vignette 1: Trees resisting technology (Figures 1, 2, and 3)

In the densely wooded section, tree branches swayed continuously in the wind, repeatedly disrupting scan alignment. The scanner was repositioned multiple times and additional scans were required to be able to align scans properly.

This represents moments of interpretive flexibility in how the technology is understood and used. The correct way to capture the space became dependent on environmental conditions rather than technical protocols. The repeated repositioning of the scanner and increased scan density reflects a process of negotiation between the user, the technology and the material environment, highlighting SCOT's emphasis on context and social factors shaping technological practices.

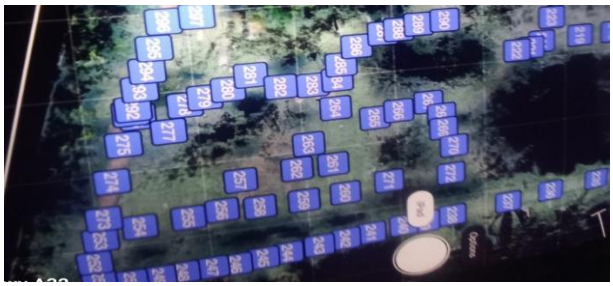


Figure 1: Scan density at dense portion of the cemetery



Figure 2: Image of dense portion of the cemetery



Figure 3: Scan density at open terrain portion of the cemetery

Negotiating Privacy

Vignette: Negotiating Privacy (Figures 4&5)

A group of visitors arrived and gathered around a particular monument while scanning was underway. Sensing their discomfort with my presence and the proximity of the scanner, I relocated the equipment to another part of the cemetery to afford them privacy.

At this moment, the digital twin did not passively capture space; it was actively shaped through ethical judgement embedded in practice. The DT produced is co-constituted through social responsibility, showing where technological operation becomes inseparable from situated ethical awareness.

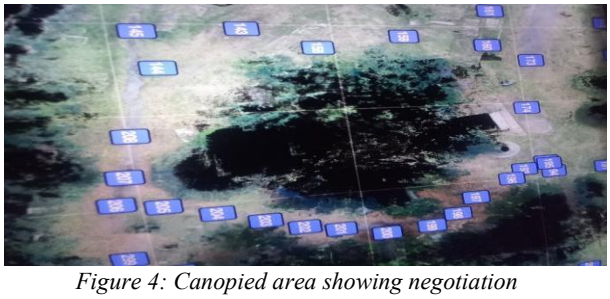


Figure 4: Canopied area showing negotiation



Figure 5: Non serial scanning points showing pause due to presence of visitors

Discussion

The aim of this study was to explore the socio-technical processes involved in the creation of a DT. The findings reveal that the emerging practices surrounding DT creation are closely linked to the process of interpretive flexibility, where the meaning of the technology is not fixed but continuously evolving.

Conclusion

This study demonstrates that DT adoption is rarely a purely technical endeavour, but rather grounded in sociotechnical negotiation. The findings show that data capture for DT creation is shaped by social interactions. DTs can be seen not as a static product, but as a dynamic, socially mediated, and negotiated outcome.

Hence, it recommends an approach to DT adoption that recognises co-production through the interaction of technological, human, social and environmental factors.

References

Asare K, A. B., Liu, R., Anumba C, J. & Issa R,R.A. 2024. Facilities Management Professionals? Perceptions of Digital Twins as Intelligent Realities. Construction Research Congress 2024.

Azman, H. & Mahadhir, M. 2017. Application of the vignette technique in a qualitative paradigm. GEMA Online Journal of Language Studies, 17, 27–44.

Bijker, W. E. 2015. Technology, Social Construction of. In: WRIGHT, J. D. (ed.) International Encyclopedia of the Social & Behavioral Sciences (Second Edition). Oxford: Elsevier.

Nikolić, D. & Ewart, I. 2025. Going beyond net zero: digital twins for achieving socio-ecological sustainability in the built environment. Smart and Sustainable Built Environment.

Schweber, L. & Harty, C. 2010. Actors and objects: a socio-technical networks approach to technology uptake in the construction sector. Construction Management and Economics, 28, 657–674.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

PARSING CONSTRUCTION STANDARDS INTO MACHINE-PROCESSABLE FORMAT

Mohamed Helmy Ammar and Merna Emara
Technische Universität Dresden, Dresden, Germany

Abstract

Manual compliance checking in construction remains challenging due to the complexity and unstructured nature of regulatory documents. This research presents an NLP-based pipeline that automatically extracts structured Information Requirements (IRs) from German construction standards (DIN 4108-2, -3, -4). A fine-tuned T5 transformer model is trained on 1,007 annotated sentences, achieving an F1-score of 92.86% on test data and 88.03% on unseen regulations. The extracted rules are converted into SPARQL queries and integrated into a BIM-based compliance checking workflow, enabling automated verification of IFC models.

Introduction

Construction compliance checking is still largely manual, requiring engineers to interpret lengthy and complex regulatory documents. Although BIM improves design coordination, the integration of regulatory requirements into automated workflows remains limited. Existing approaches often rely on manually formalized rules, restricting scalability and adaptability.

This research proposes an end-to-end NLP-based pipeline that transforms unstructured regulatory text into structured, machine-processable representations and directly integrates them into BIM-based compliance checking. The approach combines transformer-based information extraction with semantic rule execution, enabling automated and scalable regulatory validation.

Literature review

Previous research in automated compliance checking has explored rule-based, ontology-driven, and NLP-based approaches. While rule-based methods provide interpretability, they require extensive manual modeling. Recent transformer-based models improve information extraction but are often limited to text-level evaluation without integration into executable BIM workflows.

Methodology

The proposed pipeline extracts structured Information Requirements (IRs) from regulatory text using a supervised learning approach. German construction standards (DIN 4108-2, -3, and -4) are translated into English, and 1,007 sentences are manually annotated using four components: entity, property, relation, and value.

The dataset is split into training (80%), validation (10%), and testing (10%) sets. A T5-base transformer model is fine-tuned to convert raw sentences into structured outputs. T5 is selected due to its strong performance in sequence-to-sequence tasks and prior success in regulatory text parsing.

The extracted IRs are transformed into SPARQL queries for automated rule execution. BIM models in IFC format are converted into RDF to enable semantic querying, and compliance checks are executed using Apache Jena Fuseki, with results visualized in BIMvision. The overall pipeline connects regulation parsing, structured representation, and BIM-based compliance checking in a unified workflow. The overall workflow is illustrated in Figure 1.

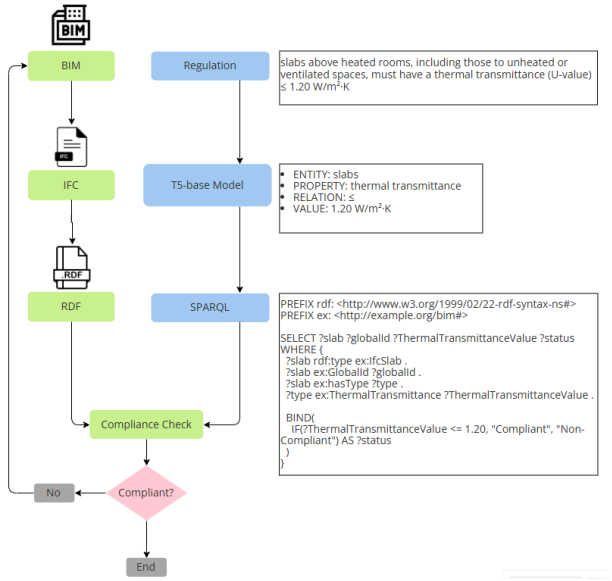


Figure 1: Overall Workflow of the Proposed Compliance Checking Process, (Adapted from [(Chen et al., 2024)])

Implementation and results

A total of 1,155 sentences are extracted from DIN 4108 standards, of which 1,007 are used for training and evaluation. Sentences are manually annotated in Excel to capture semantic components required for rule generation.

The dataset includes various building elements and thermal performance requirements, ensuring coverage of energy efficiency regulations. Preprocessing includes

normalization and cleaning, while model training is conducted using the Hugging Face Transformers library.

The fine-tuned T5 model achieves strong performance across all evaluation metrics. On the test dataset, the model reaches an F1-score of 92.86%, with high precision (92.51%) and recall (93.22%).

To evaluate generalization, the model is applied to an unseen dataset of 150 sentences, achieving an F1-score of 88.03%. These results demonstrate the model's ability to accurately extract structured Information Requirements from new regulatory content. Errors primarily occur in sentences with complex structures or multiple conditions. Additionally, transformer-based models may introduce hallucinations, generating incomplete or incorrect outputs in rare cases.

Rule compliance checking case study

A case study is conducted using a BIM model of a school building in Germany to evaluate the practical applicability of the pipeline. Extracted IRs are converted into SPARQL queries and executed on an RDF representation of the IFC model.

The system automatically identifies non-compliant elements based on thermal performance requirements. Results are visualized in BIMvision, where non-compliant building elements are automatically highlighted, as shown in Figure 2. This visual representation enables designers and reviewers to quickly identify violations and understand their spatial context within the building model, supporting efficient decision-making during the design review process.

A comparison with Solibri Model Checker confirms that the generated results are consistent with established compliance checking tools, demonstrating the reliability and practical applicability of the proposed approach.

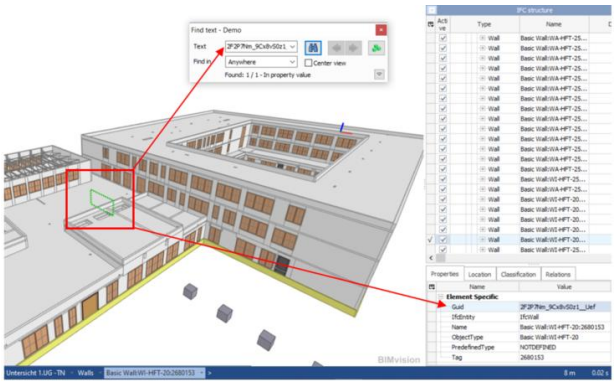


Figure 2: Visualization of query results in BIMvision

Conclusions

This research presents an NLP-based pipeline for transforming unstructured construction regulations into machine-processable rules and integrating them into BIM-based compliance checking. The results demonstrate high extraction accuracy and successful application in a real-world case study.

The approach reduces manual effort and improves the scalability of compliance checking workflows. Future work will focus on extending the method to additional regulatory domains, improving robustness for complex sentences, and reducing reliance on manual annotation.

References

Chen, N., Lin, X., Jiang, H., & An, Y. (2024). Automated Building Information Modeling Compliance Check through a Large Language Model Combined with Deep Learning and Ontology. 14(7), 1983. https://doi.org/10.3390/buildings14071983

Durmus, D., Rožanec, J., Carbonari, A., Giretti, A., Turk, Ž., & Grobelnik, M. (2024b). Potential of Large Language Models for Ontology Development in Construction Domain. Proceedings of the Creative Construction Conference (2024). http://dx.doi.org/10.3311/CCC2024-062

Preidel, C., & Borrmann, A. (2018). BIM-based Code Compliance Checking. Building Information Modeling: Technology Foundations and Industry Practice, 367–381. https://doi.org/10.1007/978-3-319-92862-3_22

Zhang, J., & El-Gohary, N. M. (2016). Extending Building Information Models Semiautomatically Using Semantic Natural Language Processing Techniques. Journal of Computing in Civil Engineering, 30(5), 1–52.https://doi.org/10.1061/(ASCE)CP.1943-5487.0000536

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A WORKFLOW FOR FEM GENERATION FROM BIM MODELS BASED ON GEOMETRIC VERIFICATION

Villem Vaktskjold, Marcin Luczkowski, and Lars Olav Toppe
Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Abstract

This extended abstract addresses the persistent gap between geometric BIM exchange and analysis-ready structural models. Although IFC supports analytical entities, exchanged models often contain geometry without the grids, axes, and connectivity definitions required for finite element analysis (FEA). A geometry-driven workflow is presented for reconstructing local axes, inferring structural grids, correcting alignment, and generating FEM models directly from IFC-derived geometry. A multistorey case study demonstrates that the reconstructed analytical model can be used for numerical analysis and produces structurally meaningful response fields. The contribution is a practical abstraction workflow that improves BIM–FEA interoperability without requiring manually authored analytical models.

Introduction

BIM and the Industry Foundation Classes (IFC) (International Organization for Standardization, 2018) standard are widely used for information exchange in the AEC sector. In structural engineering, however, exchanged IFC models frequently provide detailed geometry while lacking the analytical abstractions required for finite element analysis, including reference axes, structural grids, and consistent connectivity (Dean, 2010; Fedorik et al., 2016; Gomes et al., 2022). This forces engineers to reconstruct the analytical model manually before numerical assessment can begin.

This problem has been discussed in earlier work on IFC interoperability in the structural domain. While IFC aims to support object-based exchange, the reliable transfer of analytically relevant information remains limited, particularly across heterogeneous software workflows (Dean, 2010; Fedorik et al., 2016). More broadly, BIM models are often shaped by coordination and visualisation requirements rather than analytical needs, which can further reduce their suitability for structural assessment (Holzer, 2011). As a result, geometric completeness should not be confused with analytical readiness. The objective of this work is therefore to present a workflow that reconstructs minimal analytical reference systems directly from geometric information and uses them to generate a finite element model.

Related work

Previous BIM–FEA integration approaches have primarily focused on direct mapping between IFC entities and structural analysis software, or on proprietary workflows within specific software environments. These approaches can transfer geometry and selected metadata but typically depend on either predefined analytical models or substantial manual preprocessing (Fedorik et al., 2016; Gomes et al., 2022). Earlier studies have also shown that missing or inconsistent object mappings, unsupported attributes, and incomplete analytical entities remain recurring issues in IFC-based structural exchange (Dean, 2010; Ciccone et al., 2023).

Although IFC includes concepts for analytical representations, these are inconsistently implemented in practice. Consequently, current workflows do not adequately address the common case in which only geometric information can be assumed to be reliably available. The present work differs from direct conversion approaches by treating analytical abstraction itself as the central problem, rather than assuming it has already been solved upstream.

Method

The proposed workflow consists of four steps (A–D), as illustrated in Figure 1.

Step A: Element identification. Structural elements relevant for analysis (columns, beams, slabs) are extracted from the IFC model. Their geometric representations are used as the primary input, independent of analytical entities.

Step B: Axis estimation and grid point generation. Local axes are estimated from geometry using principal directions and element extents. Based on these axes, candidate grid points representing potential analytical node lo-cations are generated.

Step C: Grid inference and alignment. The generated points are grouped using geometric regularity, dominant directions, and spacing consistency to infer a global structural grid. Structural elements are then projected onto this grid to enforce consistent alignment and connectivity.

Step D: FEM model generation. The corrected analytical representation is converted into an FEM model with aligned axes and consistent connectivity, enabling numerical structural analysis.

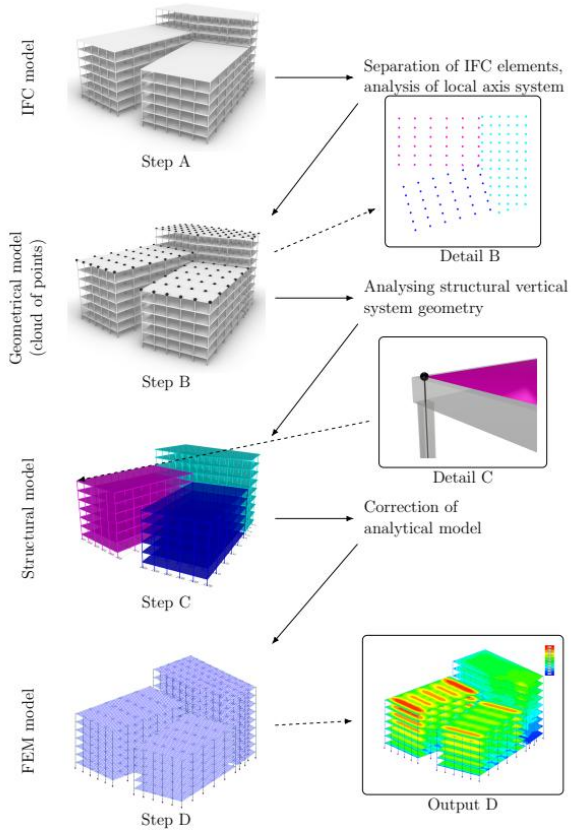


Figure 1: Geometry-driven workflow for FEM generation from IFC data. (A) Extraction of structural elements. (B) Estimation of local axes and generation of grid points. (C) Inference and correction using a global structural grid. (D) Generation of an FEM model and structural response.

Results and discussion

The workflow was applied to a multistorey building model represented in IFC. The input model contained geometrically detailed structural components but did not provide a directly usable analytical grid or consistent axis system. The geometry-based procedure successfully identified the primary structural components, estimated their orientations, and reconstructed a coherent structural grid.

Using this reconstructed grid as a global reference reduced ambiguity in member alignment and improved consistency at intended connections between beams, slabs, and columns. The corrected analytical representation was then used to generate a finite element model that could be analysed without requiring manual redefinition of the global structural layout. Representative analysis output showed stable and physically meaningful structural response.

The main benefit of the workflow is that it does not depend on manually authored analytical entities in the exchanged BIM model. This is important because many practical BIM workflows prioritise coordination and geometric detail rather than structural abstraction. At the same time, the approach has limitations. The current workflow is best suited to structural systems in which geometric regularity is sufficient to infer dominant reference directions. Future development will focus on extending the approach to highly irregular layouts, improving the treatment of sup-port conditions, and incorporating structural walls.

Conclusions

This extended abstract has presented a geometry-driven workflow for reconstructing analytical reference systems and generating FEM models from IFC data. By explicitly recovering local axes, structural grids, and alignment from geometry, the method supports structural model generation even when exchanged BIM models are not analysis-ready. The results indicate that analytical abstraction can be treated as a recoverable step in the workflow, improving BIM–FEA interoperability while remaining compatible with existing modelling practice.

References

Ciccone, A., Ciotta, V., and Asprone, D. (2023). Integration of structural information within a BIM-based environment for seismic structural e-permits. Journal of Civil Engineering and Management, 29(2):171–193.

Dean, E. T. (2010). Interoperability and the structural domain. In Structures congress 2010, pages 1652–1659, Reston, VA. American Society of Civil Engineers.

Fedorik, F., Makkonen, T., and Heikkilä, R. (2016). Integration of BIM and FEA in automation of building and bridge engineering design. In Proceedings of the 33rd international symposium on automation and robotics in construction (ISARC 2016), pages 1–6.

Gomes, A. M., Azevedo, G., Sampaio, A. Z., and Sánchez Lite, A. (2022). BIM in structural project: Interoperability analyses and data management. Applied Sciences, 12(17):8814.

Holzer, D. (2011). BIM's Seven Deadly Sins. International Journal of Architectural Computing, 9(4):463–480.

International Organization for Standardization (2018). ISO 16739-1:2018 – Industry Foundation Classes (IFC) for data sharing in the construction and facility manage-ment industries. ISO, Geneva.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A BIM-BASED DIAGNOSTIC DIGITAL TWIN FOR PREVENTIVE SEISMIC RETROFIT OF OCCUPIED SCHOOL BUILDINGS

Lavinia Chiara Tagliabue[1], Alessandro Zichi[2], Silvia Meschini[1], Stefano Rinaldi[3], and Hira Ovais[1]
[1]University of Turin, Turin, Italy
[2]Zichi Engineering, Milan, Italy
[3]University of Brescia, Brescia, Italy

Abstract

Preventive seismic retrofit of school buildings requires diagnostic-driven uncertainty reduction and planning compatible with continuous use. BIM and Digital Twin often address isolated tasks (geometry, monitoring, retrofit), while integrated workflows linking diagnostics, seismic verification, and construction safety in occupied schools remain limited. The work presents a BIM-based Diagnostic DT combining survey, in-situ testing, seismic modelling, and phased retrofit planning. A Milan school case demonstrates how diagnostic evidence can be mapped to BIM objects and used to calibrate key modelling assumptions for seismic verification. The workflow is replicable across school portfolios and supports tender-evaluable deliverables for safe execution and as-built consolidation.

Introduction

School buildings are critical public infrastructure with vulnerable users and high occupancy. In Italy, a large share of the school stock predates modern seismic provisions, motivating preventive assessment and retrofit. Unlike other building typologies, schools cannot be easily closed for long periods; therefore, feasibility depends on operational constraints such as segregation of work areas, protected circulation and emergency routes, and limited time windows for disruptive tasks. While BIM enables structured integration of geometry and semantics (Volk et al., 2014) and Digital Twin (DT) research increasingly focuses on data-driven representations (Liu et al., 2024), diagnostic evidence and execution records are often managed as stand-alone reports, limiting traceability across design revisions and as-built handover. This work proposes a BIM-centred Diagnostic DT linking survey, in-situ diagnostics, seismic assessment and phased retrofit planning for an occupied school, consistent with existing-building verification practice (MIT, 2018; CEN, 2005).

Methodology and case study

The proposed BIM-based Diagnostic DT integrates four components: (i) BIM-based geometric and semantic modelling, (ii) integration of diagnostic investigations, (iii) seismic assessment and intervention design, and (iv) construction-phase organisation and safety management. The overall Diagnostic DT workflow follows a structured progression from geometric survey and in-situ diagnostics to seismic verification, retrofit design, construction-phase phasing, and final traceability (Figure 1). It is conceived as iterative: diagnostic and execution evidence feed back into model calibration, ensuring transparency between observed conditions, modelling assumptions, and implemented interventions.

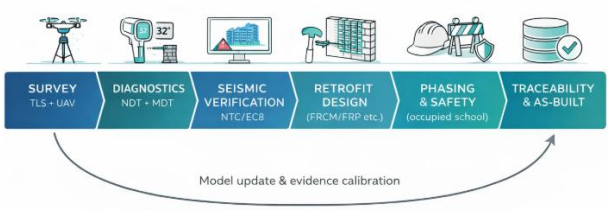


Figure 1. Diagnostic DT workflow

Survey and BIM modelling rely on terrestrial laser scanning and drone photogrammetry to support BIM reconstruction and to locate structural discontinuities (Figure 2). Diagnostic investigations (thermography, endoscopy, sonic tests, and single/double flat-jack tests for masonry, plus targeted tests on reinforced concrete members) reduce epistemic uncertainty on materials, stress state, and hidden stratigraphy (Sousa et al., 2014). Test outputs are mapped to BIM objects using consistent identifiers and stored as structured properties to ensure object-level traceability.

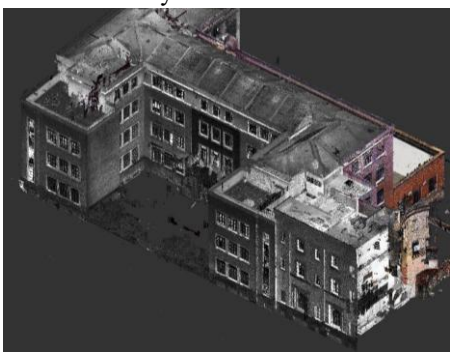


Figure 2. Terrestrial laser scanning point cloud supporting BIM reconstruction

A structural model derived from BIM supports seismic verification within an existing-building workflow; diagnostics inform material parameters and detailing assumptions used in checks (MIT, 2018; CEN, 2005). Phased scenarios are defined to minimise interference with educational activities by modelling safety zones and protected routes in BIM. Finally, the proposed award-

criteria matrix (Table 1) translates these requirements into tender-evaluable deliverables. Weighting prioritises life-safety and retrofit effectiveness while recognising phasing and safety governance as decisive constraints in occupied schools (Italian Republic, 2023; MIMS, 2021; Meschini et al., 2022). The framework is demonstrated on the listed Istituto Spinelli school in Milan, which remained operational during diagnostics and design.

Table 1. Proposed award-criteria matrix for procurement of seismic retrofit works in occupied schools

Award criterion	Weight (%)	Why it matters (occupied school)	Example evidence - deliverables
Structural effectiveness & compatibility	30	Primary objective: risk reduction; compatibility with existing systems and constraints	Retrofit design report; compatibility sheets
Construction phasing & continuity plan	20	Keeps school operational; reduces interference and downtime	Phasing plan; area-use schedule; route maps
Safety management (users & workers)	20	Segregation, protected egress, dust/noise/vibration control	Safety zoning; emergency route plan; mitigation measures
Digital deliverables & traceability	15	Verifiable progress, object-level records, as-built consolidation	BIM/DT schema; mobile checklists; update protocol
Environmental management (CAM-oriented)	10	Controlled site practices and documentation where CAM applies	Waste/material tracking; CAM records
Stakeholder communication plan	5	Coordination with school community reduces conflicts and delays	Communication protocol; notices; escalation plan

Survey and diagnostic data supported BIM reconstruction, seismic verification and the identification of critical masonry and reinforced-concrete elements (MIT, 2018; CEN, 2005). Retrofit and phasing measures were then coordinated in BIM to ensure compatibility, limited invasiveness and safety-oriented execution with protected routes, safety zones and temporary access constraints (Figure 3).

Replicability and conclusions

The Diagnostic DT is conceived for portfolio-scale use through standardised diagnostic protocols, reusable data structures and object-level identifiers, enabling comparison across school buildings and supporting preventive seismic planning. In the Milan case study, survey, diagnostics, seismic verification and phased execution are linked within a single traceable information environment. The main contribution is to make diagnostic evidence auditable and actionable, connecting uncertainty reduction to modelling assumptions, retrofit choices and safe execution in occupied schools. Future work will test transferability across additional typologies and quantify uncertainty reduction, execution traceability and portfolio-level prioritisation.

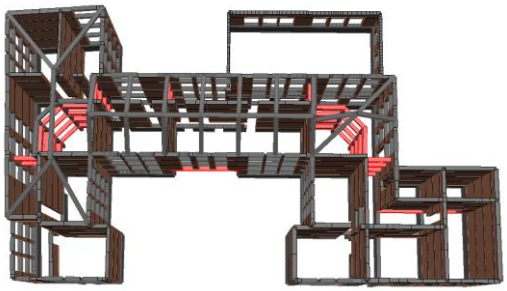


Figure 3. BIM-derived structural model with critical elements targeted by seismic verification and retrofit

Acknowledgments

The authors acknowledge Fondazione Milano and the Politecnico di Milano laboratory for support during surveys and investigations.

References

CEN (2005) Eurocode 8: Design of structures for earthquake resistance - Part 3: Assessment and retrofitting of buildings. Brussels: European Committee for Standardization.

Liu, W., Lv, Y., Wang, Q., Sun, B. and Han, D. (2024) A systematic review of the digital twin technology in buildings, landscape and urban environment from 2018 to 2024. Buildings, 14(11), 3475.

Meschini, S., et al. (2022) An innovative digital twin prototypes system to enhance tender evaluation in green public procurement. Proceedings of the European Conference on Computing in Construction (EC3).

Italian Republic (2023) Legislative Decree 31 March 2023, n. 36 (Public Contracts Code), Art. 43: Digital information management methods and tools for construction works. Rome, Italy.

MIMS (2021) Decree of the Ministry of Infrastructure and Sustainable Mobility (DM) 312/2021 on the adoption of digital methods and tools for construction information management. Rome, Italy.

MIT (2018) Decreto 17 gennaio 2018: Aggiornamento delle Norme Tecniche per le Costruzioni (NTC 2018). Gazzetta Ufficiale della Repubblica Italiana, n.42.

Sousa, H.S., Lourenço, P.B. and Almeida, M.G. (2014) Masonry wall characterization using flat-jack tests. Construction and Building Materials, 72, 102–113.

Volk, R., Stengel, J. and Schultmann, F. (2014) Building Information Modeling (BIM) for existing buildings—Literature review and future needs. Automation in Construction, 38, 109–127.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

A LARGE LANGUAGE MODEL-BASED MULTI-AGENT SYSTEM FOR ENHANCED INFORMATION RETRIEVAL IN CIVIL ENGINEERING

Sebastian Kunz, Benedikt Justin Kandler, and Uwe Rüppel
Technical University of Darmstadt, Darmstadt, Germany

Introduction

The construction industry is characterized by heterogeneous data sources (Cisterna & Haghsheno, 2024), including Building Information Modeling (BIM) models, technical standards, and various project-related documents. Efficiently retrieving information from these disparate sources remains a major challenge. Recent advances in Large Language Models (LLMs), particularly through Retrieval-Augmented Generation (RAG) (Lewis et al., 2020), offer promising opportunities to address this issue. This work presents an LLM-based multi-agent system (MAS) designed to enhance information retrieval across heterogeneous data environments in civil engineering.

The objective of the proposed system is to process complex user queries, decompose them into subtasks, and retrieve relevant information from multiple data sources using specialized agents and tools. The system follows a hierarchical architecture centered around a host agent. Upon receiving a query, the host agent interprets the request, determines the required data sources, and delegates subtasks to appropriate sub-agents or tools. The retrieved information is then aggregated into a coherent and context-aware response.

To evaluate the feasibility of the approach, BIM models and technical standards are selected as exemplary data sources due to their fundamentally different characteristics. BIM models are structured, dynamic, and project-specific, whereas technical standards are typically unstructured or semi-structured documents, often provided as PDFs and updated less frequently (Borrmann et al., 2021; Wang et al., 2025). This contrast provides a suitable testbed for assessing the system's ability to integrate heterogeneous data types.

Retrieval methods

For technical standards, a classical RAG approach is implemented. Documents are transformed into vector embeddings, enabling similarity-based retrieval of relevant text passages. The retrieved content is then used by the LLM to generate responses. (Lewis et al., 2020; Gao et al., 2023) To improve contextual understanding, document chunks are enriched with hierarchical information derived from tables of contents. This approach balances implementation simplicity with retrieval performance while ensuring traceability of results.

In contrast, BIM models require a different retrieval strategy due to their complex internal structures. The system employs LLM agents that translate natural language queries into tool calls. These tools are provided via the Model Context Protocol (MCP), a standardized interface that enables LLMs to access external functionalities (Anthropic PBC, 2024; Hou et al., 2025; Singh et al., 2025; Ray, 2025). Specifically, MCP servers such as revit-mcp (Revit MCP, 2025) and Bonsai-mcp (Rodriguez, 2025) allow interaction with BIM models in Autodesk Revit and IFC-based environments. The LLM-based model sub-agent autonomously selects and invokes these tools to extract model information.

Multi-agent system architecture

The MAS architecture consists of host agent, standards retrieval tool, and BIM sub-agent (Figure 1). The host agent orchestrates the workflow, while the standards tool performs retrieval from technical documents. The model sub-agent is responsible for BIM-related queries and interacts with MCP-based tools. This modular design enables the integration of additional data sources and retrieval methods, supporting system extensibility.

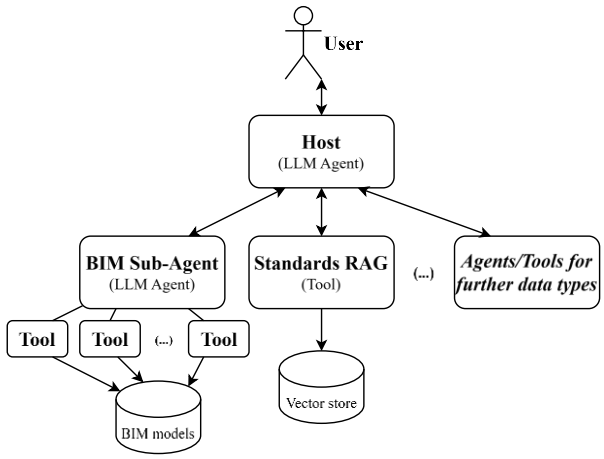


Figure 1: Simplified system architecture

Experiment

As a proof of concept, experiments are conducted using a BIM model in both RVT and IFC formats and selected

standards stored in a vector database. The LLM employed is GPT-4.1, and a zero-shot prompting strategy (Brown et al., 2020; Wang et al., 2021) is used for all agents.

Explorative test cases show that the system can answer complex queries that require combining information from both BIM models and standards. For example, the system successfully determines whether a fatigue analysis is required for a selected structural component by integrating model-specific data with relevant standard provisions.

However, the results also reveal limitations. In some cases, the system produces only partially correct answers due to suboptimal reasoning or incomplete information retrieval. For instance, when identifying applicable standards for model elements, the system fails to infer material properties correctly, highlighting challenges in query decomposition and decision-making.

Discussion, result analysis, and conclusion

While the proposed MAS demonstrates strong potential, its performance depends on numerous design choices, including chunking strategies, prompting techniques, and tool configurations. The use of zero-shot prompting contributes to high agent autonomy but can also lead to inconsistent results. Furthermore, the lack of systematic evaluation limits the ability to quantify performance and compare alternative approaches.

In conclusion, the study demonstrates that LLM-based multi-agent systems can serve as a viable approach for information retrieval across heterogeneous data sources in civil engineering. Even relatively simple system architectures can integrate multiple retrieval methods and provide unified responses. Nevertheless, further research is required to improve reliability, optimize system components, and evaluate performance systematically. Future work should also explore the integration of additional data sources and the combination of multiple retrieval methods per data type. The adoption of standardized protocols such as MCP may play a key role in enabling scalable and interoperable solutions for advanced digital workflows in the construction industry.

References

Anthropic PBC (2024) Introducing the Model Context Protocol. Available from: https://www.anthropic.com/news/model-context-protocol [Accessed 3 February 2026].

Borrmann, A., König, M., Koch, C. & Beetz, J. (2021) Die BIM-Methode im Überblick. In: Borrmann, A., König, M., Koch, C. & Beetz, J. (Eds.) Building Information Modeling. Springer Fachmedien Wiesbaden: Wiesbaden, pp. 1–31.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D. & Dhariwal, P. et al. (2020) Language Models are Few-Shot Learners. In: H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan & H. Lin (Eds.) Advances in Neural Information Processing Systems, 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver. Curran Associates, Inc, pp. 1877–1901.

Cisterna, D. & Haghsheno, S. (2024) Akzeptanz und Marktdurchdringung von KI in der Bauwirtschaft. In: Haghsheno, S., Satzger, G., Lauble, S. & Vössing, M. (Eds.) Künstliche Intelligenz im Bauwesen. Springer Fachmedien Wiesbaden: Wiesbaden, pp. 63–79.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J. & Bi, Y. et al. (2023) Retrieval-Augmented Generation for Large Language Models: A Survey. Available from: http://arxiv.org/pdf/2312.10997v5 [Accessed 3 February 2026].

Hou, X., Zhao, Y., Wang, S. & Wang, H. (2025) Model Context Protocol (MCP): Landscape, Security Threats, and Future Research Directions. Available from: http://arxiv.org/pdf/2503.23278v2 [Accessed 3 February 2026].

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V. & Goyal, N. et al. (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. Available from: http://arxiv.org/pdf/2005.11401v4 [Accessed 3 February 2026].

Ray, P.P. (2025) A Survey on Model Context Protocol: Architecture, State-of-the-art, Challenges and Future Directions. Available from: https://doi.org/10.36227/techrxiv.174495492.22752319/v1.

Revit MCP (2025) revit-mcp (GitHub repository). Available from: https://github.com/revit-mcp [Accessed 3 February 2026].

Rodriguez, J. (2025) Bonsai-mcp - Model Context Protocol Integration for IFC through IfcOpenShell and Blender (GitHub repository). Available from: https://github.com/JotaDeRodriguez/Bonsai_mcp [Accessed 3 February 2026].

Singh, A., Ehtesham, A., Kumar, S. & Khoei, T.T. (2025) A Survey of the Model Context Protocol (MCP): Standardizing Context to Enhance Large Language Models (LLMs). Available from: https://doi.org/10.20944/preprints202504.0245.v1.

Wang, Y., Yao, Q., Kwok, J.T. & Ni, L.M. (2021) Generalizing from a Few Examples: A Survey on Few-shot Learning. ACM Computing Surveys, 53(3), 1–34. Available from: https://doi.org/10.1145/3386252.

Wang, Z., Liu, Z., Lu, W. & Jia, L. (2025) Improving knowledge management in building engineering with hybrid retrieval-augmented generation framework. Journal of Building Engineering, 103, 112189. Available from: https://doi.org/10.1016/j.jobe.2025.112189.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

FROM CULTURAL DATA TO POLICY NARRATIVE: A SEMANTIC FRAMEWORK FOR URBAN TRANSFORMATION

Hira Ovais, Lavinia Chiara Tagliabue, Silvia Meschini, and Vincenzo Lombardo
Department of Computer Science, University of Turin, Italy

Abstract

Data-driven approaches increasingly inform urban policy, yet the integration of intangible cultural heritage remains limited due to object-centric documentation and the separation of tangible and intangible domains. Existing frameworks conceptualize data as static records, inadequately supporting the semantic articulation of cultural meaning within urban contexts. This paper applies a semantic narrative approach based on a triadic model linking physical elements, procedural knowledge, and social practices. The framework structures heterogeneous cultural data into interpretable relational sequences. Through ecclesiastical heritage in Karachi, Pakistan, the study shows improved coherence between cultural data and policy-oriented interpretation and contributes to knowledge applicable to urban transformation.

Introduction

Approaches to cultural heritage digitization increasingly seek to transform collected information into accessible and meaningful resources by integrating both tangible and intangible heritage dimensions (Lupo, et al., 2023). In urban contexts, heritage is no longer perceived as a static collection of relics but as a dynamic system embedded in community identity and development processes (Ripp & Clifford, 2025). Cultural heritage reflects a synchronized relationship between society, norms, and values, where tangible elements act as expressions of underlying meanings. Within this relationship, intangible heritage provides the broader framework through which tangible heritage acquires significance (Bouchenaki, 2003).

Despite this recognition, urban heritage policies continue to privilege physically measurable assets due to their ease of regulation and conservation within planning systems (Sarihan, 2021). This imbalance is reflected in contemporary documentation practices, where intangible dimensions such as ritual practices, symbolic meanings, and social use of space remain underrepresented (Labrador, 2022).

This limitation is particularly evident in Karachi, Pakistan, where heritage governance has traditionally focused on architectural conservation and monument listing (Naeem, 2013). Existing measures prioritize material condition and stylistic value, offering limited scope to document living traditions and symbolic meanings (Afridi, et al., 2025). Ecclesiastical architecture, as a minority typology, inherently embodies interdependencies between architectural form, ritual practice, and collective memory (Adeboye, 2015).

The paper proposes a semantic, data-driven storytelling approach that models these interdependencies, structuring cultural data into narrative layers aligned with UNESCO ICH domains and semantic heritage standards to support conservation and urban transformation processes (Schmitt, 2008).

Methodology

The study adopts a semantic methodology grounded in the principles of the CIDOC CRM and its architectural extension CRMba, which enable modelling cultural heritage as networks of entities and relationships across time and space (Bekiari, et al., 2021). Building on this foundation, a triadic model is introduced, linking physical elements, procedural knowledge, and social practices as the core analytical unit as shown in Figure 1.

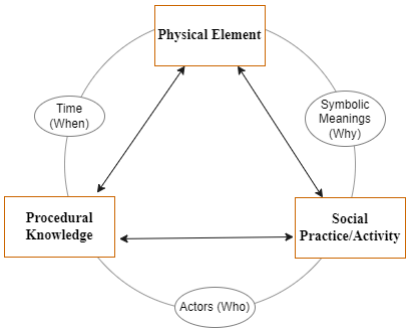


Figure 1: Triadic Semantic Model

The methodological process involves: (1) identifying tangible architectural elements, (2) documenting associated procedural knowledge such as craft techniques, and (3) identifying recurring social practices linked to these elements. These components are semantically connected using CIDOC CRM relationships, forming triadic units that represent tangible–intangible interdependencies. Cultural practices are further categorized according to UNESCO Intangible Cultural Heritage domains.

The structured data is implemented within a digital environment using Omeka-S, enabling navigation across elements, practices, actors, and spatial contexts. The methodology is applied to three colonial-era churches in

Table 1: Relational Matrix of Tangible-Intangible Interdependency

Tangible Element (E24 / CRMba)	Procedural Knowledge (E29)	Social Practice (E7)	Actors (E39)	Spatial Context	ICH Domain
Stained glass window	Glass-cutting and glazing technique	Interpretive viewing during worship	Clergy, visitors	Bay Window	Traditional Craftsmanship / Knowledge & Practices
Hymn boards	Display procedure	Congregational hymn recitation	Priest, congregation	Nave Columns	Oral Traditions

Karachi constructed between 1852 and 1900. Primary data collection included field surveys, photographic documentation, and observational analysis of spatial use, ritual practices, and symbolic interpretation. Among these, Holy Trinity Cathedral was selected as the pilot case due to its continuity of use and concentration of cultural practices.

Results and Discussions

During the investigation at the Holy Trinity Cathedral, tangible architectural features were examined in relation to the practices and procedures through which they acquire cultural significance. As outlined in Table 1, elements such as stained glass windows, hymn boards, the pipe organ, baptismal font, memorials, and carved stonework are consistently associated with recurring activities rather than existing as isolated objects.

Furthermore, the semantic structuring process revealed that tangible–intangible interdependencies may correspond to either a single UNESCO ICH domain or span multiple domains simultaneously, depending on how cultural meaning is enacted and interpreted. The stained glass window exemplifies this multi-domain structuring. It is modelled as an instance of E24 Physical Human-Made Thing, produced through specialized craft procedures (E29 Design or Procedure) aligned with the Traditional Craftsmanship domain. At the same time, its ongoing engagement through theological interpretation and commemorative reading is modelled as E7 Activity, situating the same element within the Knowledge & Practices domain.

In contrast, hymn boards demonstrate a more contained semantic alignment. While also represented as E24, their primary significance lies in facilitating congregational recitation, structured through a display procedure (E29) and enacted as a recurring oral practice (E7), corresponding predominantly to the Oral Traditions & Expressions domain.

The application of this framework demonstrates that semantically structured data can support relational knowledge systems beyond geometric modelling. The implementation in Omeka-S enables data traversal through relationships, supporting interoperability with digital environments and knowledge representation across systems.

Conclusions

A redefined understanding of cultural heritage is advanced by positioning architectural elements within interconnected systems of practice and knowledge through which significance is constituted. The proposed semantic structuring enables cultural data to be interpreted beyond static representation, supporting governance and digitally mediated heritage management. While grounded in a single case and interpretive analysis, the framework establishes a basis for scalable integration within computational environments and broader urban contexts.

References

Adeboye, A. B., 2015. Exploration of the Predominant Determinants of Ecclesiastical Architecture in Nigeria. *International Journal of Education and Research,* III(2), pp. 509-518.

Afridi, H. S., Gul, R. & Virk, A. A., 2025. The Challenges in Protecting Intangible Cultural Heritage: A Comparative Analysisof Pakistani Laws with WIPO Policy. *The Journal of Research Review,* II(4), pp. 31-44.

Bekiari, C. et al., 2021. *Volume A: definition of the CIDOC conceptual reference model,* s.l.: CIDOC CRM Special Interest Group.

Bouchenaki, M., 2003. *The Interdependency of Tangible and Intangible Heritage.* Paris, ICOMOS.

Labrador, A. M., 2022. Integrating ICH and education: A review of converging theories and methods. *International Journal of Intangible Heritage,* Volume 17, pp. 18-36.

Lupo, E. et al., 2023. *Digital for Heritage and Museums: Design-Driven Changes and Challenges.* s.l., s.n.

Naeem, A., 2013. The Conflict of Ideologies and Ambiguities in Conservation Policy: A Legacy of Shared Built Heritage in Pakistan. In: *Asian Heritage Management: Contexts, Concerns and Prospects.* s.l.:Routledge, pp. 87-104.

Ripp, M. & Clifford, J., 2025. Heritage-Based Urban Development. *Encyclopedia,* 5(2).

Sarihan, E., 2021. Visibility Model of Tangible Heritage. Visualization of the Urban Heritage Environment with Spatial Analysis Methods. *Heritage,* IV(3), pp. 2163-2182.

Schmitt, T., 2008. The UNESCO Concept of Safeguarding Intangible Cultural Heritage: Its Background and Marrakchi Roots. *International Journal of Heritage Studies,* 14(2), pp. 95-111.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

PREDICTIVE FRAMEWORK FOR DOMINANT DELAY RISKS IN METALLURGICAL PROJECTS

Shivam Pandey, Sahil Garg
Indian Institute of Technology Delhi, Delhi, India

Abstract

Indian steel and aluminium megaprojects frequently experience multi-year schedule overruns despite detailed planning and risk registers. This paper presents a predictive delay-risk framework that learns from 20 completed projects using contextual similarity and interaction-aware modelling of 39 delay events. Projects are encoded through execution context, while events are represented using normalised delay impacts and a structured dependency network derived from empirical co-occurrence and expert judgement. A prototype tool generates probabilistic completion envelopes and ranked, interaction-aware risk lists for new projects. Initial results align with observed completion behaviour and highlight a small set of high-leverage risks for targeted mitigation.

Introduction

Industrial projects in the steel and aluminium sector continue to experience persistent schedule overruns, often extending several years beyond baseline plans. Although prior research identifies numerous delay causes such as approvals, procurement disruptions, contractor performance, and labour issues, these are typically analysed as independent factors or through isolated case studies.

In practice, project teams rely on schedules, Earned Value indicators, and qualitative risk registers, which do not indicate which risks will dominate or how combinations of risks propagate delays. This limits proactive decision-making and leads to reactive control.

This study addresses this gap by proposing a predictive framework that identifies dominant delay risks and models their interaction using limited but structured historical data.

Empirical dataset and delay structure

The framework is developed using an empirical dataset of 20 completed steel and aluminium projects, including greenfield plants and brownfield expansions. Observed schedule overruns range from moderate (6–12 months) to multi-year delays.

A structured library of 39 delay events is defined, covering approvals, procurement, logistics, execution, and governance-related disruptions. Each event is characterised by its phase and delay impact using triangular estimates. Projects are linked to events through a project event matrix, enabling consistent modelling of delay behaviour.

Methodology

Contextual Similarity Learning

Each project is represented by a context vector including plant type, project nature, capacity, region, climate, and contracting model. Similarity between a new project and historical projects is computed using weighted categorical matching and Gaussian similarity for numerical attributes.

Similarity scores are transformed into normalized weights, ensuring that contextually similar projects exert greater influence on delay estimation.

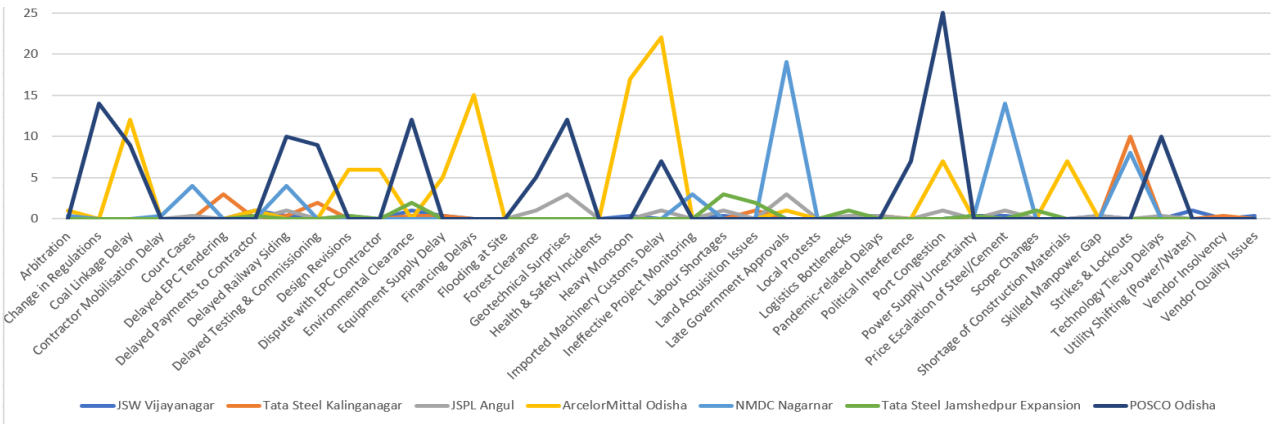


Figure 1: Multi-project delay profile for dominant events showing mean schedule impact (months)

Dominance Estimation of delay events

Delay events are treated as contributors to total delay rather than independent occurrences. For each project, event impacts are normalised to represent their relative contribution to overall delay. For a new project, dominance scores are computed by aggregating these contributions across similar projects.

Table 1: Top delay events ranked by probability, contribution, and mean impact.

Delay events	Probability	Contribution (%)	Mean impact (month)
Price Escalation of Steel/Cement	0.21	8.8	1.04
Shortage of Construction Materials	0.39	8.0	0.94
Delayed Payments to Contractor	0.61	6.6	0.61
Local Protests	0.39	6.2	0.39

Interaction and Amplification Modelling

Delay events are often interdependent. To capture this, a dependency network is defined representing trigger response relationships between events. Amplification factors, derived from empirical co-occurrence and expert judgement, adjust probabilities and impacts of dependent events during simulation.

This enables modelling of cascading effects, where upstream risks propagate across project phases and create clustered delay scenarios.

Predictive Simulation

The model is evaluated using Monte Carlo simulation. Event impacts are sampled from triangular distributions, modified by dependency relationships, and aggregated to estimate total project delay.

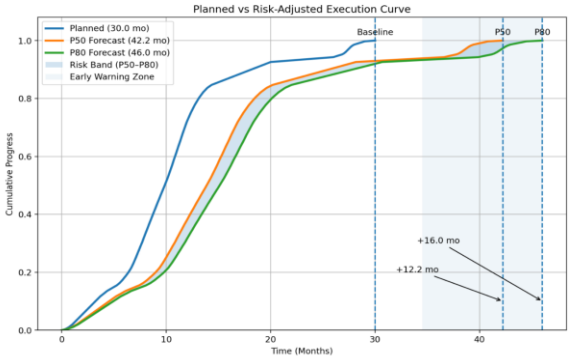


Figure 2: Baseline vs probabilistic completion curves (P50, P80) with risk band

Results and Implementation

The predicted P50 and P80 completion times fall between baseline and actual values, indicating realistic estimation of delay magnitude and uncertainty. Results show that a small subset of risks dominates total delay, while interaction modelling captures cascading effects that produce heavy-tailed outcomes.

The framework is implemented as a Python/Streamlit prototype, generating completion curves, dominant risk rankings, and phase-wise delay insights to support targeted mitigation.

Conclusions

This study presents a predictive framework for delay behaviour in metallurgical projects using contextual similarity, dominance estimation, and interaction-aware simulation. The results show that delay is driven by a small set of dominant and interacting risks, and that modelling these interactions improves realism. The framework enables early identification of high-impact risks, supporting proactive project control.

References

MOSPI, “Flash Report on Programme Implementation,” Ministry of Statistics and Programme Implementation, Government of India, New Delhi, India, 2024.

Reza Azimi, SangHyun Lee, Simaan M. AbouRizk, and Amin Alvanchi, “Framework for an automated and integrated project monitoring and control system,” Automation in Construction, 2013.

A. Shibani & K. Arumugam, “Avoiding Cost Overruns in Construction Projects in India,” Procedia Engineering, 2015.

PMI (India), “Study on Project Schedule and Cost Overruns,” Project Management Institute – India Research Whitepaper, 2017.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

A SYSTEMATIC REVIEW OF DISTRIBUTED REINFORCEMENT LEARNING FOR BUILDING-TO-DISTRICT ENERGY NETWORKS

Katharina Steib[1], Thomas Ohlson Timoudas[2], and Qian Wang[1]
[1]Department of Civil and Architectural Engineering, KTH Royal Institute of Technology, Sweden
[2]RISE Research Institutes of Sweden, Kista, Sweden

Abstract

Distributed Reinforcement Learning (RL) enables coordinated HVAC control in district energy systems but existing approaches vary in architecture, addressed challenges, and practical feasibility. This review classifies recent studies by distribution level and problem dimensions and uses a research map to identify trends, research gaps, and future directions for practical, privacy-aware, and safe distributed control.

Framework definition for distribution levels

We introduce a framework that categorizes RL approaches in Building Energy System (BES) by distribution level:

- *L0 Centralized Single-Agent RL* [Zhou et al. (2022), Bao and Xu (2023), Ojand and Dagdougui (2022), Ding et al. (2022)]
- *L1 Partially distributed RL with central Coordination* [Qin et al. (2022), Ahrarinouri et al. (2022), Ruan et al. (2023), Huo et al. (2024), Tomin et al. (2022)]
- *L2 Coordinated distributed RL* [Zhang et al. (2025), Miao et al. (2024), Shen et al. (2022), Bo et al. (2023), Zhu et al. (2024), Tomin et al. (2024), Nweye et al. (2025), Rezazadeh and Bartzoudis (2022)]
- *L3 Fully distributed RL* [Pei et al. (2025), Sun et al. (2024), Zhang et al. (2022b), Tariq et al. (2025), Zhang et al. (2022a)]

Problem categorization

We define system-level core problem dimensions, covering challenges from classical HVAC control, limitations addressed by RL, and new challenges introduced by distributed RL: P1 Scalability, system size and complexity; P2 Coordination and communication; P3 Privacy and Data access; P4 Safety, Stability and Constraint satisfaction; P5 Training efficiency, adaptability and generalization.

Comparison using a research map

Table 1 presents a two-dimensional research map of the 22 reviewed studies, organized by distribution architecture and problem focus. Primary and secondary problems are indicated by filled and unfilled circles, respectively.

With the research matrix we identified the following gaps: 1) limited attention to P3–P5, particularly at higher distribution levels; 2) except for Miao et al. (2024), all reviewed studies rely solely on simulation; 3) few L3 studies and no work addressing all five problems simultaneously.

Table 1: Distribution matrix

	P1	P2	P3	P4	P5
L0	○	○	○	○	○
L1	●	●	○	○	●
L2	●	●	○	○	○
L3	●	●	●	○	○

● primary focus ○ secondary focus

Discussion and conclusion

Rezazadeh and Bartzoudis (2022) combines RL and federated learning for privacy-preserving multi-owner districts. Huo et al. (2024) and Sun et al. (2024) propose safe RL for safety-critical HVAC systems. Ruan et al. (2023) uses offline training on historical data to reduce online exploration.

This work reviewed distributed RL for HVAC control in district energy systems. The key takeaways are: 1) Distributed RL enables scalable and coordinated multi-building control; 2) Major challenges remain: privacy, safety, training efficiency, and real-world deployability; 3) studies address only a subset of problems, highlighting the need for holistic, system-level approaches; 4) Future work should focus on integrated architectures and real-world validation under multi-owner constraints.

Acknowledgements

This study is funded by the Swedish Energy Agency (P2023-01521, TwinVista), and European Union's Horizon Europe programme (101096789, HYSTORE).

References

Ahrarinouri, M., Rastegar, M., Karami, K., and Seifi, A. R. (2022). Distributed reinforcement learning energy management approach in multiple residential energy hubs. Sustainable Energy, Grids and Networks, 32:100795.

Bao, G. and Xu, R. (2023). A Data-Driven Energy Management Strategy Based on Deep Reinforcement Learning for Microgrid Systems. Cognitive Computation, 15(2):739–750.

Bo, W., Wang, X., Jing, G., Xu, H., and Jia, L. (2023). Comfort and energy management of multi-zone

HVAC system based on Multi-Agent Deep Reinforcement Learning. In 2023 IEEE 18th Conference on Industrial Electronics and Applications (ICIEA), pages 1641–1646, Ningbo, China. IEEE.

Ding, Z., Fu, Q., Chen, J., Wu, H., Lu, Y., and Hu, F. (2022). Multi-zone Residential HVAC Control with Satisfying Occupants' Thermal Comfort Requirements and Saving Energy via Reinforcement Learning. In Shen, H., Sang, Y., Zhang, Y., Xiao, N., Arabnia, H. R., Fox, G., Gupta, A., and Malek, M., editors, Parallel and Distributed Computing, Applications and Technologies, volume 13148, pages 441–451. Springer International Publishing, Cham. Series Title: Lecture Notes in Computer Science.

Huo, X., Liu, B., Dong, J., Lian, J., and Liu, M. (2024). Optimal Management of Grid-Interactive Efficient Buildings via Safe Reinforcement Learning. In IECON 2024 - 50th Annual Conference of the IEEE Industrial Electronics Society, pages 1–6, Chicago, IL, USA. IEEE.

Miao, C., Cui, Y., Li, H., and Wu, X. (2024). Efficient multi-agent reinforcement learning HVAC power consumption optimization. Energy Reports, 12:5420–5431.

Nweye, K., Kaspar, K., Buscemi, G., Fonseca, T., Pinto, G., Ghose, D., Duddukuru, S., Pratapa, P., Li, H., Mohammadi, J., Lino Ferreira, L., Hong, T., Ouf, M., Capozzoli, A., and Nagy, Z. (2025). CityLearn v2: energy-flexible, resilient, occupant-centric, and carbon-aware management of grid-interactive communities. Journal of Building Performance Simulation, 18(1):17–38.

Ojand, K. and Dagdougui, H. (2022). Q -Learning-Based Model Predictive Control for Energy Management in Residential Aggregator. IEEE Transactions on Automation Science and Engineering, 19(1):70–81.

Pei, Y., Yao, Y., Zhao, J., Hao, J., Ding, F., and Wang, J. (2025). Multi-Agent Hierarchical Deep Reinforcement Learning for HVAC Control With Flexible DERs. IEEE Transactions on Smart Grid, 16(6):5589–5601.

Qin, Y., Ke, J., Wang, B., and Filaretov, G. F. (2022). Energy optimization for regional buildings based on distributed reinforcement learning. Sustainable Cities and Society, 78:103625.

Rezazadeh, F. and Bartzoudis, N. (2022). A Federated DRL Approach for Smart Micro-Grid Energy Control with Distributed Energy Resources. In 2022 IEEE 27th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), pages 108–114, Paris, France. IEEE.

Ruan, Y., Liang, Z., Qian, F., Meng, H., and Gao, Y. (2023). Operation strategy optimization of combined cooling, heating, and power systems with energy storage and renewable energy based on deep reinforcement learning. Journal of Building Engineering, 65:105682.

Shen, R., Zhong, S., Wen, X., An, Q., Zheng, R., Li, Y., and Zhao, J. (2022). Multi-agent deep reinforcement learning optimization framework for building energy system with renewable energy. Applied Energy, 312:118724.

Sun, Y., Zhang, S., Liu, M., Zheng, R., and Dong, S. (2024). Energy management based on safe multi-agent reinforcement learning for smart buildings in distribution networks. Energy and Buildings, 318:114410.

Tariq, S., Ali, U., Kim, S., and Yoo, C. (2025). Multi-agent distributed reinforcement learning for energy-efficient thermal comfort control in multi-zone buildings with diverse occupancy patterns. Energy, 332:137082.

Tomin, N., Kolosok, I., Kurbatsky, V., and Korlina, E. (2024). A Demand-Response Approach for HVAC Systems Using Internet of Energy Concept. In Sari, M. and Kulachinskaya, A., editors, Digital Transformation: What are the Smart Cities Today?, volume 846, pages 275–293. Springer Nature Switzerland, Cham. Series Title: Lecture Notes in Networks and Systems.

Tomin, N., Shakirov, V., Kozlov, A., Sidorov, D., Kurbatsky, V., Rehtanz, C., and Lora, E. E. (2022). Design and optimal energy management of community microgrids with flexible renewable energy sources. Renewable Energy, 183:903–921.

Zhang, B., Hu, W., Ghias, A. M., Xu, X., and Chen, Z. (2022a). Multi-agent deep reinforcement learning-based coordination control for grid-aware multi-buildings. Applied Energy, 328:120215.

Zhang, G., Hu, W., Cao, D., Zhang, Z., Huang, Q., Chen, Z., and Blaabjerg, F. (2022b). A multi-agent deep reinforcement learning approach enabled distributed energy management schedule for the coordinate control of multi-energy hub with gas, electricity, and freshwater. Energy Conversion and Management, 255:115340.

Zhang, Y., Zhao, Y., Zhang, C., and Feng, C. (2025). Multi-agent reinforcement learning-based method for demand response of building HVAC systems. Journal of Building Engineering, 108:112734.

Zhou, Y., Ma, Z., Zhang, J., and Zou, S. (2022). Data-driven stochastic energy management of multi energy system using deep reinforcement learning. Energy, 261:125187.

Zhu, S., Lu, W., Feng, Y., and Sun, C. (2024). Energy Optimization for Building Energy Management with Thermal Storage: A Multi-Agent Deep Reinforcement Learning Approach. In 2024 43rd Chinese Control Conference (CCC), pages 2805–2812, Kunming, China. IEEE.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

TOWARDS AUTOMATED BUILDING MAINTENANCE AND RENOVATION: A MULTI-AGENT LLM-BASED APPROACH

Jiaxin Ling[1] and Weiwei Chen[1,2]
[1]University College London, London, United Kingdom
[2]University of Cambridge, Cambridge, United Kingdom

Abstract

This study presents a multi-agent AI framework for automating building maintenance and renovation recommendations. Given inputs such as building location, type, and reported issues, a task identification agent classifies cases as maintenance or renovation. A knowledge agent retrieves and interprets relevant building codes, standards, and guidelines, while a solution generation agent produces maintenance recommendations, renovation strategies, or compliance assessments. By integrating task classification, regulation interpretation, and recommendation generation through large language models (LLMs), the framework automates key decision-making processes traditionally performed by human experts.

Introduction

Europe's building stock is ageing, with over 40% constructed before the 1960s, and the sector accounts for more than 40% of final energy consumption and 36% of greenhouse gas emissions (European Parliament, 2016; Vendel et al., 2025). As a result, the Repair, Maintenance and Improvement (RM&I) sector is identified as a key growth area (Gov.UK, 2025). Building maintenance and renovation are knowledge-intensive processes requiring integration of fragmented information from multiple sources. While digital technologies such as BIM and AR improve data accessibility, they remain largely data-driven and lack the ability to capture and reason over regulations, standards, and expert knowledge.

To address this gap, this study proposes a multi-agent LLM-based framework for maintenance and renovation decision-making. The framework decomposes complex tasks into specialised agents, including task identification, knowledge retrieval, and solution generation. By coordinating these agents, it enables context-aware and structured decision workflows, supporting more intelligent and efficient maintenance and renovation processes.

Methodology

This study proposes a multi-agent LLM-based framework for automated maintenance and renovation decision-making in buildings. Figure 1 presents the workflow of the proposed multi-agent framework for enhanced building maintenance and renovation decision-making.

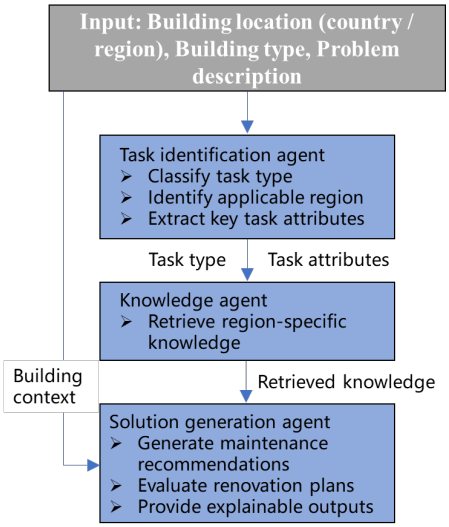


Figure 1: Workflow of the proposed multi-agent LLMs

Case study

The case study focuses on a heritage building located in Poland, originally constructed in 1928. As a historically protected structure, the building presents typical challenges associated with ageing construction, including material deterioration and constraints imposed by conservation requirements.

LLM agent development

The proposed multi-agent framework is implemented using the GPT-4o large language model, which serves as the core reasoning engine for all agents. GPT-4o was selected due to its strong capability in understanding technical text, handling contextual information, and generating coherent and explainable outputs (Dai et al., 2025).

Results

Maintenance recommendation

Table 1 summarizes the outputs of the task identification agent and the solution generation agent for the façade crack maintenance scenario.

Based on these outputs, the solution generation agent produces corresponding maintenance recommendations that prioritise minimal intervention and material compatibility.

Table 1: Maintenance recommendation for façade cracks (heritage building)

Agent	Results
Task identification agent	Task type: Maintenance (façade condition) Context: Heritage building in Poland (likely protected/subject to conservator oversight) Observed issue: Surface cracks + local material degradation on external wall/facade (image-supported) Decision constraint: Minimal intervention, reversible where possible, retain historic fabric; ensure traceability of actions and rationale (documentation)
Solution generation agent	(1) Immediate, low-intrusion actions Condition recording and baseline survey: photo log, crack mapping (location, length, width), and date-stamped documentation before any intervention. Traceability basis: conservation actions should be documented; maintaining best possible condition requires evidence-based monitoring. (2) Heritage-compatible repair measures Localized plaster/finish repair using compatible materials (e.g., lime-based where historically appropriate; avoid hard cement-rich patches that can trap moisture and accelerate decay). If cracks indicate structural instability: prioritize stabilization measures designed to minimize loss of historic fabric (e.g., local reinforcement only where necessary, reversible techniques where possible).

Limitations and Future Directions

Despite the promising results, this study has several limitations. First, the proposed framework relies on the reasoning capabilities of LLMs. Although it's effective for interpreting unstructured regulatory text and contextual information, it does not replace formal regulatory approval or expert judgement. The generated maintenance recommendations and renovation compliance assessments are intended to support early-stage decision-making rather than serve as definitive or legally binding conclusions. In addition, the evaluation is based on a single heritage building case study, and broader validation across different building types, regulatory contexts, and levels of data availability is required to further assess generalisability.

Future work will focus on extending the framework in several directions. Additional case studies across multiple regions and building typologies will be conducted to evaluate scalability and robustness. Moreover, future research will explore tighter coupling between LLM-based agents and quantitative analysis modules, such as structural assessment using numerical simulation or energy performance evaluation, as well as the incorporation of human feedback to support iterative and collaborative decision-making in maintenance and renovation workflows.

Conclusions

This paper presents a multi-agent LLM-based framework to support building maintenance and renovation decision-making, with a focus on heritage buildings. By decomposing complex tasks into specialised agents for task identification, knowledge handling, and solution generation, the approach enables structured and explainable decision support, addressing the challenge of integrating domain knowledge into current workflows.

Acknowledgement

This study was supported by the EU-funded INHERIT project (Project ID: 101123326): Next Generation Solutions for Sustainable, Inclusive, Resource-efficient and Resilient Cultural Heritage.

References

European Parliament. Energy efficiency of buildings - European Parliament. On-line: https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/582022/EPRS_BRI(2016)582022_EN.pdf, Accessed: 09/01/2026.

GOV.UK. Construction building materials: commentary September 2025. On-line: https://www.gov.uk/government/statistics/building-materials-and-components-statistics-september-2025/construction-building-materials-commentary-september-2025, Accessed: 09/01/2026.

Vendel, Kädi-Riin, Ergo Pikas, Müge Tetik, Lauri Koskela, and Jarek Kurnitski. Industrialized district-level renovation of similar buildings in proximity: Environmental, economic, and renovation production implications. Energy and Buildings: 116790, 2025.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

AN ARCHITECTURAL WORKFLOW DATABASE FOR GENERATING EVIDENCE-BASED HEALTHCARE FACILITIES' DESIGN VIA TRANSFER LEARNING

Julia Zieleniewska[1] and Dimosthenis Kifokeris[2]
[1]Poznan University of Technology, Poznan, Poland
[2]Chalmers University of Technology, Gothenburg, Sweden

Abstract

Evidence-based design (EBD) principles can help overcome legislative shortcomings and long-standing challenges in designing healthcare units. Leveraging these principles through transfer learning can utilize pre-trained knowledge to generate optimal spatial arrangements, yet the dataset-formulating factors remain scarcely investigated. Mixed methods (including workflow analysis and behavioral mapping in an emergency care unit) are employed to generate a highly contextual EBD database. Using four actor flow scenarios, dead zones, nodes, collision points, approximate nursing distances, and zone mapping were identified as key contextual factors. This study contributes to informing design considerations and the contextual training of transfer learning in generative healthcare architecture.

Introduction

The emergency department spatial organization affects patients and personnel. ERs play a key role in healthcare systems (Zieleniewska et al., 2026). Comparing Swedish and Polish ERs reveals certain design factors, which mostly originate from evidence-based design (EBD) (Gawlak and Stankiewicz, 2022).

The information base for architectural factors remains fragmented and difficult to transfer between case studies. Machine learning and data-driven methods were proposed for synthesizing a knowledge base. Study aims to propose an architectural approach including combination of workflow, plan analysis and behavior mapping (Grzywiński et al., 2020; Mahmeen et al., 2024) of selected ER, as a method of generating the basis for transfer learning-assisted design space generation. Transfer learning (TL) can be promising for extracting operational and spatial patterns from existing ER layouts.

Research method and materials

Analyzing ERs enables identification of design patterns and transferable data extraction. Employing mixed methods, a literature review (Pruszewicz-Sipińska et al., 2023) and walkthrough analyses (field observations, staff interviews) (Niezabitowska, 2014) were conducted in four Polish and one Swedish hospital case studies. The case selected in this extended abstract comprises the currently under-construction ER extension of a provincial hospital in western Poland, with a daily admission of 200-300 patients and a capacity of 24-30 beds, which features a combined hospital and ER registration impacting the flow of personnel and patients. Due to confidentiality, the unit's name and location are not disclosed. Workflow and behavior mapping were conducted using ArchiCAD29 on plans and via staff interviews. The 2D plans contributed to flow diagrams and critical point analysis. Combined with focus interviews, this served for behavior mapping, revealing scenarios linking rooms and inter-zone flow, establishing action sequencing. For pattern identification, functional zones (registration, medical, urgent, staff), key users, their roles and behaviors were defined. Study was limited to main patient scenarios (Figure 1): short stay, admission, emergency pathways and personnel maintenance. The next step includes creating a visual pathway for users.

Results

The patient flow is mostly linear, while the staff's is mainly circular. Results stemming from the workflow and behavior mapping were identified (Figure 1):

1. Dead zones:
   a. Niches in the foyer - need to verify actual usage through post occupancy evaluation (POE).
   b. The break room is located close, but outside ER; distances may discourage from regular breaks.
2. Nodes:
   a. Corridor bend connecting the urgent track with the rest of the hospital.
   b. Corridor bend connecting the main corridor and the urgent track, limiting distances and clarifying orientation. Possible paths collisions.
3. Collision points:
   a. One large nurse station; better supervision, may generate staff crowding.
   b. Operational disturbance between procurement staff and transported patients
4. Bottlenecks: Registration, triage, transfer corridor

Mapping clarifies context, revealing processes, scalable factors related in circulation, zoning and visibility:

1. Waiting area for both scheduled and emergency patients (risks: chaos, paths collisions)
2. Larger isolation room without access to exterior area: possible problems in the event of pandemic
3. Single observation area: increased supervision

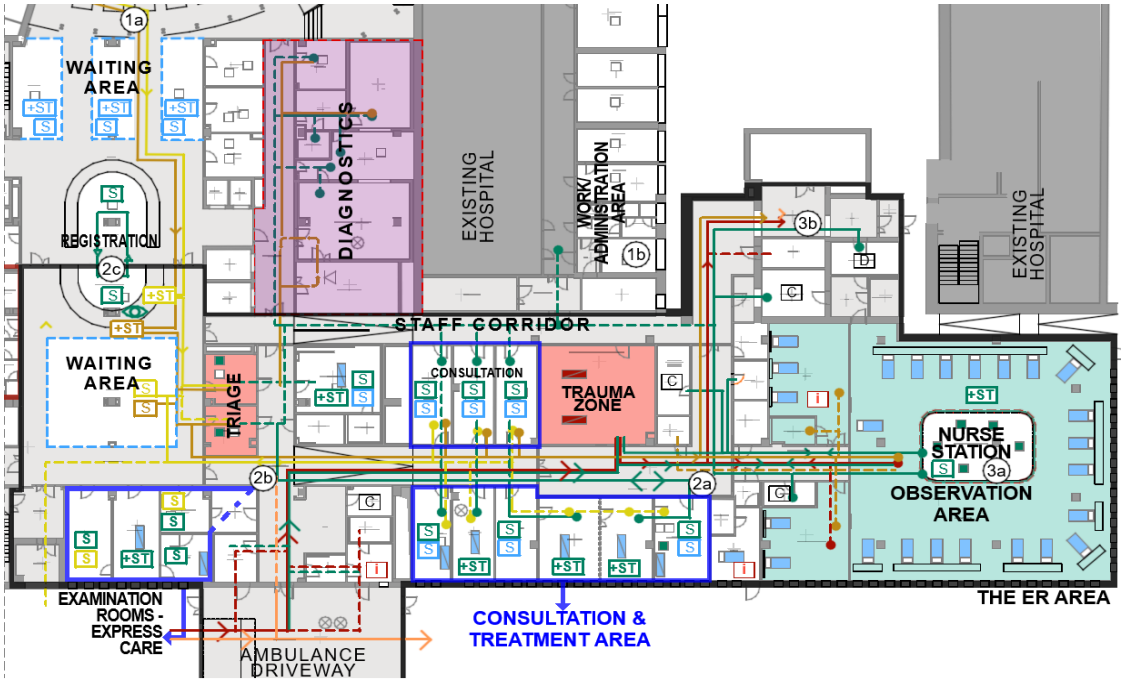


Figure 1: Workflow and behavior mapping (own work). Blue: patients in general; red: urgent patients; yellow: few-hours patients; brown: mild cases; green: personnel; orange: paramedics; navy blue: zones; i: isolation; c: clean storage; d: dirty storage; ST: standing; S: sitting; Alternatives: dashed. lines.

4. Trauma zone nearby observation: quick transfer.
5. Diagnostics located nearby reduce distances
6. Separate express care eliminates unnecessary patient influx into medical zone.
7. Nursing distances: ~38m (staff corridor), ~41m (main corridor, urgent part: ~35m), ~44m (triage to observation room). Communication covers ~26% of the area.

Discussion – transfer learning

Transfer learning may address the limited generalizability of ER layouts, using workflow mapping data as structured input for relational patterns (Khodaee et al., 2024) and decision-making. Mapping methods, coupled with normative knowledge can form a hybrid training dataset for transfer learning: relational determinants compatible with graph-based models, sequence learning, and hybrid-symbolic systems (Rahbar et al., 2022), supporting transferability and replicability.

Conclusions

ERs' spatial organization shapes clinical performance and patient experience. Workflow reveals spatial issues and supports structured data for TL-assisted design.

The main limitation is that ER is under construction. Future studies should include POE using live tracking. Kernel density analysis may support further investigation of contextual factors not examined at this stage.

References

Gawlak, A. & Stankiewicz, M. (2022) Specific Needs of Patients and Staff Reflected in the Design of an Orthopaedic and Rehabilitation Hospital—Design Recommendations Based on a Case Study (Poland). International Journal of Environmental Research and Public Health, 19(22), 15388.

Khodaee, P., Viktor, H.L. & Michalowski, W. (2024) Knowledge transfer in lifelong machine learning: a systematic literature review. Artificial Intelligence Review, 57, 217.

Mahmeen, M., Mehdi, S.A., Friebe, M., Pech, M. & Haider, S. (2024) Mapping and deep analysis of hospital radiology department to identify workflow challenges and their potential digital solutions. Journal of Health Management, 26(4), 581-593.

Niezabitowska, E. (2014) Metody i techniki badawcze w architekturze (In English: Research methods and techniques in architecture). Wydawnictwo Politechniki Śląskiej. [Accessed 08 February 2026]

Pruszewicz-Sipińska, E. & Matuszewska, M. Zieleniewska, J. (2023) Kształtowanie przestrzeni pracy w szpitalu jako jedna z propozycji rozwiązań kryzysu kadrowego w sektorze zdrowia w Polsce. Zeszyty Naukowe Politechniki Poznańskiej. Architektura, Urbanistyka, Architektura Wnętrz, 17 Wydanie Specjalne.

Rahbar, M., Mahdavinejad, M., Markazi, A.H.D & Bemanian, M. (2022) Architectural layout design through deep learning and agent-based modeling: A hybrid approach. Journal of Building Engineering, 47, 103822.

Zieleniewska, J., Matuszewska, M., & Pruszewicz-Sipińska, E. (2026) Future of Polish Hospital Emergency Departments: Architectural Strategies for Technological and Socio-Demographic Change in the Post-Pandemic Era. Buildings, 16(4), 800.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

RGB-D BASED WORK AREA RECOGNITION MODEL FOR FIREPROOFING SPRAY ROBOT

Sangmin Lee, Kangmin Bae, Minseung Cha, Sebeen Yoon, and Taehoon Kim
Seoul National University of Science and Technology, Seoul, Republic of Korea

Abstract

Fireproofing spray on steel structures is essential for fire resistance, yet it is still manual, exposing workers to hazardous and dusty conditions. Robots have been developed to address these issues, but some systems depend on work area map, which limits real time adaptation. In addition, object level recognition cannot provide surface level work areas. Therefore, we propose an RGB D work area recognition model that classifies work area surfaces on a steel structure beam. On experiment, the model achieves a macro average F1 score of 0.8682. Finally, the proposed model provides surface level work area for fireproofing spray robot.

Introduction

Fireproofing spray on steel structures is a required process to secure fire resistance. Because it is frequently performed at height and under dusty conditions, workers can be exposed to safety risks, and spraying quality can vary with worker's skill level.

Fireproofing spray robots have been developed to address these issues. However, some systems depend on offline preparation, where spraying work files are generated from structural design data in advance and the robot operates only after a working area map is created beforehand (Ikeda et al., 2020). Because the work area must be set up before each spraying task and updated when site conditions change, productivity can drop and real time adaptation during spraying remains limited.

To improve real-time adaptation, recent studies use vision sensing to segment steel structures, but results remain at the object level(Zhu et al.,2022). Such object level recognition does not provide surface level work areas for robot spraying.

Therefore, unlike prior approaches that depend on offline work area map preparation or remain at the object level, this study proposes an RGB-D based work area recognition model that classifies five work area surfaces on a steel structure beam for fireproofing spray robot.

RGB-D based Work Area Recognition Model

This study proposes an RGB-D based work area model for fireproofing spray robot(Figure 1).

Data Preparation

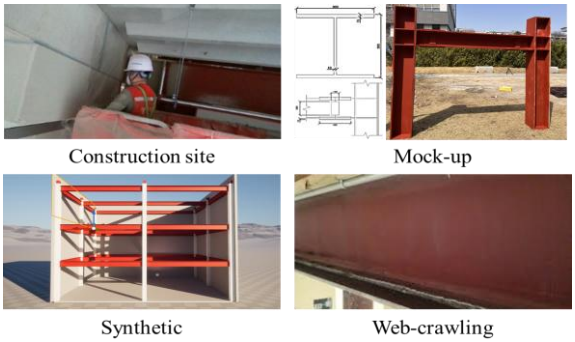


Figure 2: Four types of train dataset

For data preparation, RGB-D videos are recorded during fireproofing spray using an RGB-D camera, and RGB images and depth maps are extracted at 30 fps. Also, a YOLOv11 segmentation model is trained for steel structure beam segmentation using four types of train dataset, with a total of 480 images(Figure 2).

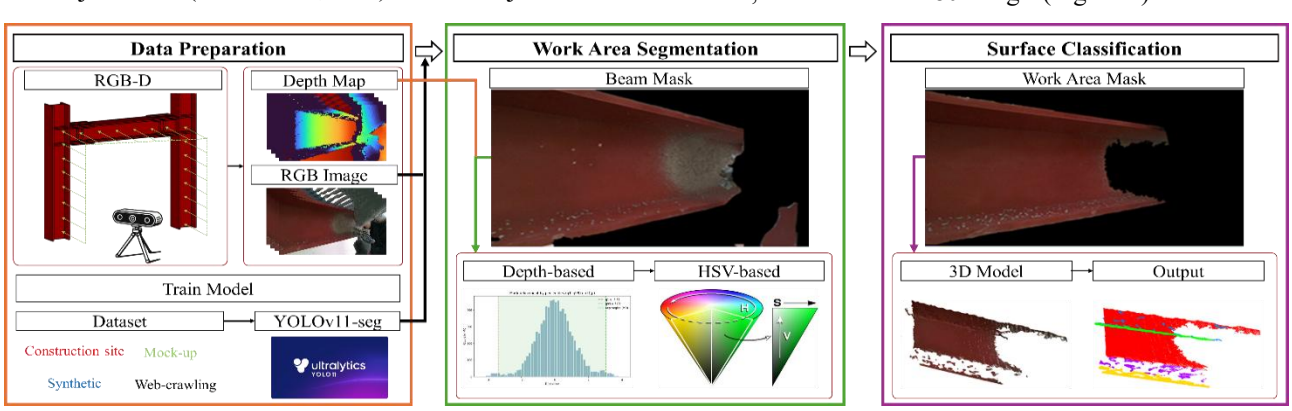


Figure 1: RGB-D work area recognition model framework

Work Area Segmentation

A trained YOLOv11 segmentation model generates an initial mask of the steel structure beam from the RGB image. The mask is refined using the depth map by expanding the boundary and keeping only pixels within the 1st–99th percentile depth range inside the mask. Within the refined region, HSV-based rules separate unworked and worked areas: unworked area is selected by red Hue ranges (0–20 or 160–180) with sufficient saturation and value, while worked area is excluded using low saturation and high value without constraining Hue. The final work area mask is obtained by removing worked area pixels.

Surface Classification

A 3D point cloud is reconstructed from the final work area mask and the depth map. RANdom Sample Consensus(RANSAC; web: thresh = 0.01 m, minPoints = 5000, maxIteration = 4000; flange: thresh = 0.008 m, minPoints = 1500, maxIteration = 3000) iteratively extracts planar surfaces by selecting the plane with the largest inlier set, removing its inliers, and repeating on the remaining points. Density-based spatial clustering of applications with noise(DBSCAN; eps = 0.02, min_points = 10) then clusters inliers on each plane into spatially consistent groups, yielding five classified work area surfaces on the steel structure beam.

Experiment

Experiment was conducted using the setup shown in Table 1, with 51 pairs of RGB image and depth map separately collected for evaluation.

Table 1: Experimental setup

Subject	Medium-density fibreboard((MDF) mock-up beam(H-300×300×10×15)	
Data Acquisition	Sensor	Intel RealSense D455
	View Direction	Same as the spray direction
	Distance	300~600mm
Time of Recording	During fireproofing spray operation	
Data	Resolution	RGB Image 1280×720, Depth map 848×480
	Quantity	51 Pairs of RGB Image&Depth map

Result Analysis

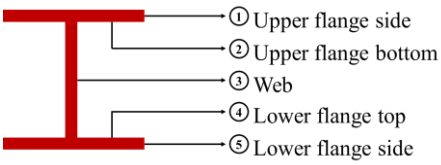


Figure 3: Steel structure beam surfaces

We defined five work surface classes on the steel structure beam, as illustrated in Figure 3. The five surface classification performance was evaluated using precision, recall, and F1-score computed on the 3D point cloud by aggregating outcomes for each class. Overall performance was summarized using the macro average across the five surfaces. Each surface results are reported in Table 2.

Table 2. Experiment results

Surface	Precision	Recall	F1-score	Macro Average
①	0.9992	0.9958	0.9771	0.8682
②	0.8398	0.4995	0.6264	
③	0.9794	0.9238	0.9508	
④	0.9469	0.7385	0.8298	
⑤	0.9758	0.9385	0.9567	

Surface ② showed the lowest recall, likely because one inner flange surface was captured at an unfavorable viewing angle under the current setup. As the camera was positioned approximately perpendicular to the web, the opposite inner flange surface was less visible, resulting in sparse depth points and more difficult surface separation.

Conclusion

This study proposed RGB-D based work area recognition model for a fireproofing spray robot that classifies five work area surfaces on a steel structure beam. It outputs surface level work areas and achieved a macro average of 0.8682 during spraying. Therefore, the proposed model provided surface level work area information for fireproofing spray robot.

However, the experiment was conducted in a controlled environment and did not fully reflect varying site conditions such as illumination changes, dust, and occlusions. Future work will include an ablation study to evaluate the contribution of each pipeline component, improve the model, and validate robot integration for progress monitoring and spray path generation.

Acknowledgments

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant No. RS-2025-11802969).

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant No. RS-2022-00143493).

References

Ikeda, Y., Segawa, H., & Yabuki, N. (2020). Development and application of a fire-resistive covering spraying robot for steel structures. Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC 2020).

Khanam, R., & Hussain, M. (2024). *YOLOv11: An overview of the key architectural enhancements*. arXiv preprint arXiv:2410.17725.

Zhu, M., Zhang, G., Zhang, L., Han, W., Shi, Z., & Lv, X. (2022). Object Segmentation by Spraying Robot Based on Multi-Layer Perceptron. Energies, 16, 232.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

HUMAN-DATA INTERACTION (HDI) TECHNICAL COMMITTEE

Diego Calvetti[1,2], Mahir Msawil[3], Xiang Xie[4], Jennifer Li[5], Dimosthenis Kifokeris[6], Gulben Calis[7], Peter Demian[8], Nick Nisbet[9], Dragana Nikolic[10], Kay Rogage[3], May Shayboun[11], Sujesh Sujan[12], Alexandre Mugnai[13], Ranjith K. Soman[5], Mustapha Munir[14].

[1]Pontificia Universidad Católica de Chile, Chile; [2]University of Porto, Portugal; [3]Northumbria University, UK; [4]Newcastle University, UK; [5]Technical University of Delft, The Netherlands; [6]Chalmers University of Technology, Sweden; [7]Ege University, Türkiye; [8]Loughborough University, UK; [9]University College London, UK; [10]University of Reading, UK; [11]Halmstad University, Sweden; [12]Norwegian University of Science and Technology, Norway; [13]ESTECO, Italy; [14]University of Salford, UK.

Introduction

Human-Data Interaction (HDI) is a transformative research agenda that also represents a critical gap in current digital transformation efforts. Such transformation efforts have produced complex digital ecosystems that raise questions of, *e.g.,* trust, agency, and ethics, but the predominantly technocentric approaches tend to prioritise system performance over human-centric aspects. The Built Environment (BE) is data-rich but insufficiently human-centred in how data are accessed, interpreted, governed, and acted upon.

The HDI Committee, a permanent technical committee of the European Council on Computing in Construction (EC^3), is dedicated to exploring the intricate relationship between humans and data in the BE. It investigates a wide range of topics, including user experience, change enablement, ethical data practices, and technological advancements that enhance our interaction with data. The Committee focuses on fostering dialogue and promoting leadership practices in HDI. It seeks to prioritise HDI-related research and educational initiatives, promoting practical applications across key areas.

Theoretical Foundations

As illustrated in Figure 1, throughout the Built Asset Lifecycle (BAL), humans serve as enablers, decision-makers, and users across various processes. The adoption of advanced digital technologies (*e.g.,* AI-based solutions, robotics, digital twins) has improved the efficiency and effectiveness of work processes while generating vast streams of digital data (Atuahene et al., 2023). In increasingly data-driven systems, it is not only the role of technology that matters, but the ways in which humans engage with the data these systems generate, transform, and act upon.

HDI addresses this gap by highlighting how users interact with digital data, focusing on access, interpretation, action, and ethical management in everyday tasks. Mortier et al.'s (2014) seminal paper identified the three foundational pillars of HDI (*legibility*, *agency*, *negotiability*) within the wider context of information systems. With the aim of adapting these pillars to the context of the BE's digitalisation, recent scholarly contributions (Li et al., 2024) have unpacked them to derive operational dimensions, including *trust, human empowerment, safety, understanding, and privacy*. These dimensions, grounded in BE research, can be linked to these pillars to establish overarching principles for HDI tailored to digitalisation efforts across BAL. Taken together, this theoretical direction aims to reinforce HDI's relevance in settings where digital solutions are technically developed (*e.g.,* requirement gathering and analysis) and/or implemented to inform individual decisions and behaviours across BAL.

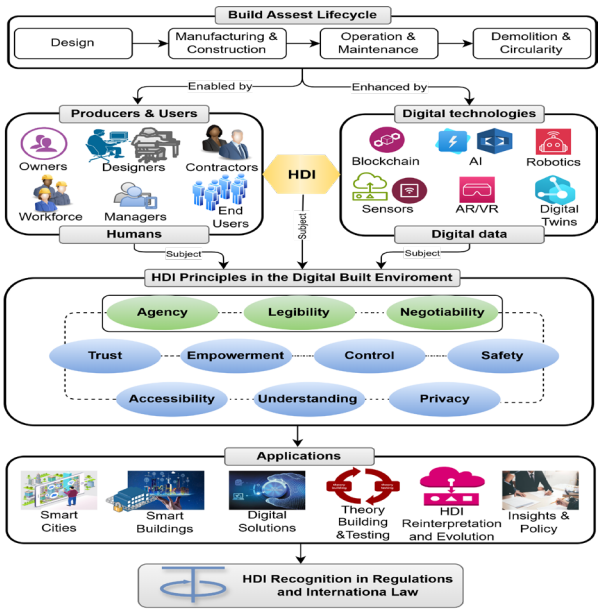


Figure 1: HDI Principles in the BE

From an enabler perspective, HDI emphasises that those managing change, through platform configurations, standards, and workflows, enhance data's findability, interoperability, and trustworthiness, shaping conditions for effective and responsible use. From a decision-maker perspective, HDI highlights the importance of timely, clear, and ethically governed data for evaluating trade-offs and making lifecycle decisions.

The overarching aim of HDI in the BE is to develop theoretical foundations while extending its applicability across diverse contexts. Contexts include smart cities, global policy development, and the reinterpretation of HDI principles in relation to cognition-driven intelligent systems like generative AI and autonomous DTw.

Committee's Contributions

HDI is a vital framework for understanding user-data relationships throughout the BAL within the context of computing in construction. Committee members contribute to knowledge dissemination through multiple scientific papers and outreach activities through engaging with and supporting all EC[3] conferences, delivering knowledge through presentations, track sessions, and discussions on computing in construction.

In September 2024, the Committee launched its premier webinar introducing HDI in the BE, featuring presentations on HDI theory, generative AI, intelligent environments, and industry-academia collaboration. The Committee's White Paper (Kassem and Kifokeris, 2024) serves as a *Call to Action*, establishing HDI as a central research domain that unifies various interaction models under a human-centred approach. This *Call to Action* illustrates the shift from data as a neutral concept to understanding how data is constructed, interpreted and contested across stakeholders.

Committee members have collectively advanced theory and application through research. Li et al. (2024) developed an HDI theory for construction contexts incorporating trust, empowerment, safety, accessibility, and understanding, alongside the three pillars of HDI. Li et al. (2023) examined the implications of HDI for blockchain implementation, highlighting how stakeholder interactions shape immutability, transparency, and governance. Calvetti et al. (2023) proposed maturity levels for HDI in Digital Twins (DTw), also emphasising HDI's role in electronic performance monitoring at sensored construction sites, addressing data ownership and GDPR compliance (Calvetti et al., 2021, 2024). Rogage et al. (2025) highlighted the need for responsible data practices to manage the data carbon footprint in construction. Kifokeris (2025) explored HDI's potential to transform regenerative building production processes. Saif et al. (2025) advocated for sociotechnical approaches in DTw that balance technology with stakeholder engagement. Lastly, Munir et al. (2025) identified gaps in HDI strategies concerning sustainability and social good to drive purposeful digital implementations. Collectively, these contributions provide a foundation for future research to further integrate HDI into scalable, responsible, and human-centred digitalisation across the BE.

References

Atuahene, B.T., Kanjanabootra, S. and Gajendran, T. (2023) 'Transformative role of big data through enabling capability recognition in construction', *Construction Management and Economics*, 41(3), pp.208-231.

Calvetti, D., Kifokeris, D., Mêda, P., and Hjelseth, E. (2023) 'Human-data interaction in incremental digital twin construction', *2023 European Conference on Computing in Construction and the 40th International CIB W78 Conference,* Crete, Greece, 10-12 July.

Calvetti, D., Kifokeris, D., Mêda, P., and Sousa, H. (2024) 'Human-Data Interaction as a critical enabler of Electronic Performance Monitoring at construction sites', *Journal of Information Technology in Construction*, 29, pp. 722–749.

Calvetti, D., Mêda, P., Sousa, H., & Chichorro Gonçalves, M. (2021) 'Human Data Interaction in Sensored Sites, Challenges of the Craft Workforce Dimension', *2021 European Conference on Computing in Construction*, (online), 26-28 July.

Kassem, M. and Kifokeris, D. (2024) Emerging Human-Data Interactions within Construction and the BE: A White Paper by the Human-Data Interaction (HDI) Committee of the EC[3].

Kifokeris, D. (2025) 'Initiating the Conceptualisation of Human-Data Interactions in Regenerative Building Production', *2025 European Conference on Computing in Construction*, Porto, Portugal, 14-17 July.

Li, J., Kifokeris, D., Barati, M., Calis, G., Hall, D., Hunhevicz, J., Kassem, M., Msawil, M., and Srećković, M. (2023) 'Human-Data Interaction (HDI) and blockchain: An exploration of the open research challenges for the construction community', *2023 European Conference on Computing in Construction and the 40th International CIB W78 Conference*, Crete, Greece, 10-12 July.

Li, J., Shayboun, M., Calis, G., Calvetti, D., Demian, P., Kassem, M., Msawil, M., and Soman, R. (2024) 'Toward a working theory for Human-Data Interaction (HDI) in the Built Environment (BE)', *2024 European Conference on Computing in Construction*, Crete, Greece 14-17 July.

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., and Crowcroft, J. (2014). Human-Data Interaction: The Human Face of the Data-Driven Society. SSRN Electronic Journal.

Munir, M., Sujan, S., and Hjelseth, E. (2025) 'Exploring Digital Transformation by the IKIGAI Philosophy', *2025 European Conference on Computing in Construction*, Porto, Portugal, 14-17 July.

Rogage, K., Gledson, B., Jones, B., and Rodriguez, P. (2025) 'A Framework to Measure the Digital Carbon Footprint of Design Workflows', *2025 European Conference on Computing in Construction*, Porto, Portugal, 14-17 July.

Saif, W., Li, J., Doukari, O., and Kassem, M. (2025) 'Reframing Digital Twins in Construction: A Socio-Technical Lifecycle Perspective', *2025 European Conference on Computing in Construction*, Porto, Portugal, 14-17 July.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

GENERATING IFC-BASED THERMAL MODELS FOR INDUSTRIAL FACILITIES

Adolfo Sánchez, Eduardo Adán, and Antonio Adán
Department of Electrical Engineering, Electronics, Automation and Communications
University of Castilla La Mancha. Spain

Abstract

Recent LiDAR-based technologies provide accurate geometric data of indoor scenes and facility interiors (e.g., factories, halls, or large rooms), which can be processed to obtain standard BIM models in various formats, resulting in a basic digital twin of the environment. A step beyond conventional digital twins is the creation of scene models incorporating dense thermal information at a given time. This study proposes a methodology to generate BIM models of building interiors in IFC format, embedding temperature as a property of each architectural or functional element.

Introduction

Although multiple studies have reported the generation of 3D thermal maps using digitization systems that include one or more infrared cameras (Ramón, Adán and Javier Castilla, 2022), existing work on indoor environments has so far been limited to floor-level representations (essentially multiple rooms connected by corridors) in the resulting thermal point clouds (Collins et al., 2024). Autonomous robotic systems (Adán, López-Rey and Ramón, 2023) and UAV-based platforms (Zhang et al., 2023) capable of thermally digitizing parts of building interiors and exteriors have also been proposed; however, most of these systems fail to generate fully omnidirectional thermal models covering the entire visible surface of the environment.

Research on thermal digital twins is still very limited, particularly at the scale of an entire building floor. Studies addressing the generation of standard IFC formats enriched with temperature properties did not appear until 2026, as reported in (Ramón-Constantí et al., 2026), which can be considered an initial contribution in this field. In that work, a process was designed to generate an open IFC file in which average surface temperatures were embedded as native properties of architectural elements of a curtain wall.

Methodology and experimentation

The proposed methodology in our study is structured into three phases, as shown in Figure 1(a). In the first phase, an omnidirectional thermal point cloud of the environment is acquired using a LiDAR-based sensing platform referred to as MoPAD. In the second phase, an engineer or architect models the interior space using an appropriate modelling tool. Finally, in the third phase, the results of the two previous stages are integrated by associating a temperature value with each structural element, all within an open standardized format widely used in the AEC domain, namely IFC (Industry Foundation Classes). This poster is specifically focused on this third integration phase. In turn, the creation of the Thermal-IFC model comprises three main steps, which are summarized as follows. 1) Extraction of the coordinates of the centroids and vertices of each geometric element contained in the IFC file; 2)

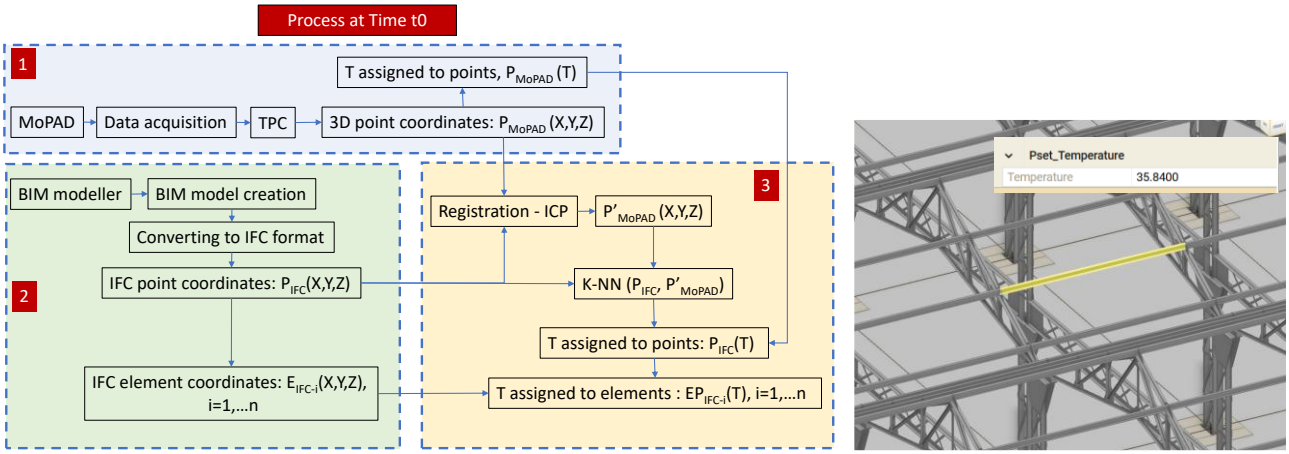


Figure 1: a) Flowchart of the proposed methodology, illustrating its three main steps. b) Steel load-bearing structure. Detail of a steel bar with a highlighted region showing the measured temperature

Assignment of temperature values to the elements through automatic registration between the element vertices and the thermal point cloud; 3) Generation of the IFC file by integrating temperature attributes into the corresponding elements. The two basic algorithms used in the step 2) were the ICP (Iterative Closest Point) algorithm for point registration between those captured by the digitising platform and the vertices of the elements in the IFC file, and the k-NN (k-nearest neighbours). The ICP algorithm yielded a registration RMSE below 2 cm.

This methodology has been successfully tested in various indoor environments (buildings, industrial halls, boiler rooms, etc.). As a representative high-complexity case study, the poster presents results from a large industrial hall (39 × 88 × 10 m) containing 10,377 IFC elements, each assigned a temperature value. The resulting thermal IFC model was successfully visualised in BIM platforms, enabling thermal behaviour analysis. Temperature calibration relied on a contact sensor (0.5 °C accuracy), ensuring a final temperature error below this threshold. Figure 1(b) illustrates the generated thermal IFC model.

Conclusions and contributions

This work presents a practical and scalable methodology for embedding dense thermal information into IFC-based BIM models by automatically associating temperature data with architectural elements, enabling enriched thermal digital twins. Validation in large-scale indoor environments demonstrates its applicability to real-world scenarios and its potential for energy analysis and building performance assessment in the AEC domain.

Since temperature in IFC models is inherently temporal, as it corresponds to a specific time instant, the proposed approach can be understood as a temporal digital model suitable for applications such as thermal monitoring and indoor thermal simulation.

The main contributions of this work can be summarized as follows:

- This study advances the integration of thermal models within the construction context, improving the efficiency of previous approaches.
- The proposed integration method is automatic and accessible to AEC professionals, without requiring specialized computer or programming skills.
- The resulting thermal IFC models can be directly used by AEC professionals to analyse building thermal behaviour, supporting energy efficiency and occupant comfort assessment.

Acknowledgments

This publication is as part of the PID2022-141275OB-C21, CIN/AEI/10.13039/501100011033/ FEDER, UE and SBPLY23180225000113 projects co-funded by FEDER (European Regional Development Fund).

References

Adán, A., López-Rey, A. and Ramón, A. (2023) 'Robot for thermal monitoring of buildings', Automation in Construction, 154, p. 105009. doi:10.1016/J.AUTCON.2023.105009.

Collins, F.C. et al. (2024) '360-Degree Tri-Modal Scanning: Engineering a Modular Multi-Sensor Platform for Semantic Enrichment of BIM Models', ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-1–2024(1), pp. 41–48. doi:10.5194/isprs-annals-X-1-2024-41-2024.

Ramón-Constantí, A. et al. (2026) 'Integrating thermal point clouds into BIM-GIS environments: workflow proposal for multi-layer digital twins', Frontiers in Built Environment, 11, p. 1715347. doi:10.3389/fbuil.2025.1715347.

Ramón, A., Adán, A. and Javier Castilla, F. (2022) 'Thermal point clouds of buildings: A review', Energy and Buildings, 274, p. 112425. doi:10.1016/J.ENBUILD.2022.112425.

Zhang, C. et al. (2023) 'Thermal-textured BIM generation for building energy audit with UAV image fusion and histogram-based enhancement', Energy and Buildings, 301, p. 113710. doi:10.1016/J.ENBUILD.2023.113710.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

A Self-Verification Framework Toward Reliable Text-to-BIM Generation

Tobias Sesterhenn[1], Bharathi Kannan Nithyanantham[2], Stefan Lüdtke[2], and Christian Bartelt[1]
[1]Institute for Software and Systems Engineering, TU Clausthal, Germany
[2]Institute for Visual and Analytic Computing, University of Rostock, Germany

Abstract

Recent advances in generative AI have enabled Text-to-BIM systems that generate building models from natural-language prompts; however, these systems typically lack automated verification, requiring manual inspection and iterative error correction. To improve the reliability of LLM-generated BIM outputs, we propose a Text-to-BIM workflow with integrated self-verification. For each prompt, the system derives two complementary validations: an Information Delivery Specification (IDS) for rule-based checking, and LLM-driven code-based verification for requirements beyond IDS. The resulting feedback is used to refine the IFC model in a closed loop. The framework is implemented and evaluated through a case study, demonstrating improvements in model quality.

Introduction

Text-to-BIM systems generate Building Information Models (BIM) from natural language instructions (Deng et al., 2025), but despite their capabilities, current approaches lack a task-specific self-verification component that ensures the final BIM model actually meets the request of the human user. For example, Text2BIM (Du et al., 2026) includes a rule-based model checker, but all checks are predefined and may lose task-specific requirements. Consequently, users must manually inspect IFC models and iteratively correct deviations, limiting reliability and scalability of such systems.

To address this, inspired by self-verification (Gou et al., 2024), we propose a closed-loop, LLM-assisted Specification-Modification-Verification framework. For each prompt, a *Specifier* agent derives two complementary specifications: (i) an automatically checkable Information Delivery Specification (IDS) encoding all formally verifiable, schema-level constraints, and (ii) a higher-level specification capturing semantic and contextual requirements that are evaluated by a *Verifier* agent. This allows a *Modifier* agent creating a BIM model to iteratively receive automated feedback by the *Verifier*, to finally output a model that meets the given specification. Our contributions are as follows:

- A fully automated Text-to-BIM framework with integrated self-verification.
- A novel Text-to-IDS module that turns natural-language requirements into valid IDS constraints.
- A simple case study demonstrating the framework.

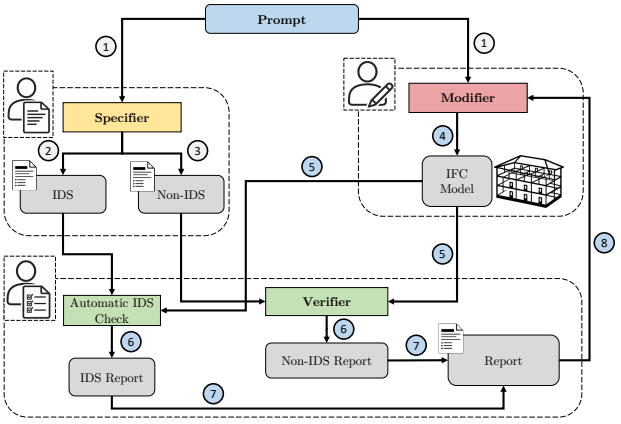


Figure 1: The Modifier-Specifier-Verifier Framework. Steps 4-8 are colored blue to indicate they are repeated in each iteration of the Modifier-Verifier Loop. The Specifier is only required in the first iteration to produce the specification documents.

Framework

As shown in Figure 1, the framework coordinates three agents – the *Specifier*, *Modifier*, and *Verifier* – within a closed-loop architecture to translate a user's design prompt into an iteratively refined IFC model. Given a human prompt (①), the *Specifier* Agent derives two complementary artifacts: (i) an IDS document encoding all formally expressible, schema-level requirements (e.g., properties, simple relations) (②), and (ii) a less formal non-IDS specification capturing constraints that cannot be represented in IDS form, such as geometric, topological, or complex relational conditions (③).

The *Modifier* Agent generates and iteratively updates an IFC model based on the prompt and – in following iterations – subsequent feedback (④). In each iteration, validation on the current IFC model is then performed at two levels (⑤). An automated IDS checker evaluates the model against the IDS document, producing an IDS Report (⑥). In parallel, the *Verifier* Agent operationalizes the non-IDS specification by generating and executing custom verification code (i.e., programmatic checks over the IFC structure), also producing a report. Both reports are consolidated into a unified feedback report (⑦), which is returned to the Modifier Agent to guide model revision (⑧). The loop continues until the IFC Model satisfies both the formal IDS constraints and the non-IDS requirements, or until

a predefined iteration limit is reached.

Case Study

To demonstrate the framework, we conduct a case study using the prompt: *"Create a simple but realistic house, 4 rooms per level, 2 levels, made out of wood. The roof should be a gable roof."*[1]

We use GPT-5.2 for five iterative refinement cycles. The *Modifier* and *Verifier* agents are equipped with a toolbox based on MCP4IFC (Nithyanantham et al., 2025), enabling iterative programmatic IFC model construction and querying. Concretely, both agents can interact with a python interpreter and run code using either available functions from the toolbox (e.g. `create_wall()`) or by writing own `ifcopenshell-python`[2] code. In each iteration the agent is allowed to make multiple tool calls and to incorporate the results into its reasoning process, as formalized by the ReAct framework (Yao et al., 2022). Automated IDS validation is performed using the open-source `ifctester`[2] implementation.

As illustrated in Figure 2, the system progressively resolves detected violations. For instance, the IDS requirement of eight `IFCSpace` entities fails in Iteration 1 and is satisfied by Iteration 3. Similarly, the geometric verification of the gable roof initially fails and is corrected by Iteration 3, together with the assignment of the material attribute "*Timber*" to relevant wall and slab elements. By Iteration 5, the roof geometry and overall building envelope are further refined.

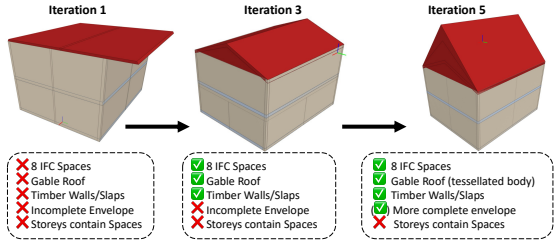


Figure 2: Iterative refinement in the case study. The boxes indicate compliance with individual IDS and non-IDS requirements at each iteration of the self-verification loop.

Discussion

The case study provides a proof of concept for the proposed framework; however, a comprehensive evaluation is required to assess its robustness and generalizability. In the presented experiment, several limitations became apparent. For instance, the system exhibited difficulties in reliably deleting existing elements and failed to resolve certain constraint violations across iterations, such as the missing relationship between the eight `IFCSpace` entities and their corresponding `IFCBuildingStorey` elements. Moreover, complete coverage of all formal and non-IDS requirements during specification generation cannot be guaranteed, and any errors or omissions at this stage directly impact subsequent validation and refinement.

Conclusion & Future Work

We presented an agent-based Text-to-BIM framework centered on self-verification. For each prompt, the system derives checkable constraints and leverages automated feedback to iteratively revise the IFC model until detected violations are minimized or resolved. The simple case study suggests that this specification-modification-verification loop improves model quality and reliability, while also highlighting the need for further evaluation and refinement to ensure robustness and broader applicability. Future work should therefore focus on systematic evaluation and strengthening of the individual agents. The *Specifier* agent could be assessed through a dedicated Text-to-IDS benchmark to measure completeness and correctness of constraint extraction. The *Verifier* agent may benefit from integrating deterministic geometry and topology checks to reduce reliance on purely reasoning-based validation. In addition, hybrid validation strategies should be explored, combining LLM-generated specifications with predefined or human-authored rule sets to improve reliability.

Acknowledgments

This research was partially funded by the Federal Ministry of Research, Technology and Space (BMFTR) and the Federal Ministry for Economic Affairs and Energy (BMWE) of Germany.

References

Deng, Z., Du, C., Nousias, S., and Borrmann, A. (2025). Bimgent: Towards autonomous building modeling via computer-use agents. arXiv preprint arXiv:2506.07217.

Du, C., Esser, S., Nousias, S., and Borrmann, A. (2026). Text2bim: Generating building models using a large language model-based multiagent framework. Journal of Computing in Civil Engineering, 40(2):04025142.

Gou, Z., Shao, Z., Gong, Y., yelong shen, Yang, Y., Duan, N., and Chen, W. (2024). CRITIC: Large language models can self-correct with tool-interactive critiquing. In The Twelfth International Conference on Learning Representations.

Nithyanantham, B. K., Sesterhenn, T., Nedungadi, A., Peral Garijo, S., Zenkner, J., Bartelt, C., and Lüdtke, S. (2025). MCP4IFC: IFC-based building design using large language models. arXiv preprint arXiv:2511.05533.

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R., and Cao, Y. (2022). React: Synergizing reasoning and acting in language models. In The eleventh international conference on learning representations.

[1] Our code is publicly available at github.com/Tsesterh/Text2BIM-Self-Verification.

[2] `https://docs.ifcopenshell.org`

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

INTEGRATING PARAMETRIC MODELING, COMPUTER VISION, AND MIXED REALITY: REAL-TIME QUALITY CONTROL IN CRAFTSMANSHIP

Simon K. Hoeng, Christoph J. Holzfurtner, Anne-Sophie Saffert,
Florian Weininger, Thomas Linner and Mathias Obergriesser
Ostbayerische Technische Hochschule Regensburg, Regensburg, Germany

Abstract

European craftsmanship suffers from shrinking apprenticeships and skill shortages while project demands continue to grow. High tech, hands on workflows can both modernize existing practices and renew interest among potential future workers. We present and evaluate an end-to-end framework that couples Parametric Modeling, Artificial Intelligence (AI) in the form of Computer Vision (CV), and Mixed Reality (MR) to support real time assistive quality control in stone paving, serving as a representative, under-digitized craft domain. The work also demonstrates how the core CV pipeline can be embedded in a course format for students or craftspeople to cultivate practical AI skills, highlighting the demand for high-tech integration in craftsmanship and related education.

Background and Motivation

As the European construction sector progresses toward Construction 4.0, it faces severe labor shortages driven by an aging workforce and a demographic decline in available apprentices (Brucker Juricic et al., 2021). Concurrently, the industry struggles to attract and retain young professionals, often due to a disconnect between modern digital ideals and traditional working conditions. Recent research identifies educational reform and innovative teaching methods as the strongest factors in improving the sector's career attractiveness (McDonough, 2026) In parallel, automated construction inspection and progress monitoring has been widely identified as a promising means improve reliability in quality control, with computer vision emerging as a core enabling technology (Samsami, 2024). Prior construction research further shows that AR/MR systems are well suited for contextual, task-level guidance at the point of work (Asmar et al., 2021). Together, these findings motivate the integration of CV-based inspection with mixed reality feedback in this work, enabling quality-related information to be communicated directly in the worker's field of view under controlled conditions. Consequently, embedding assistive technologies like AI and MR into manual trades is not merely a technical upgrade, but a strategic way to modernize curricula and secure a future-ready workforce.

Method: Interconnected Technologies

This work demonstrates a closed loop framework for quality control in paving that links the digital model and real-world status: (1) a parametric as planned model encodes per stone geometry and rotations; (2) a real time instance segmentation pipeline infers as built stone states from a monocular video stream; (3) QR based registration maps pixel detections into model coordinates via homography; (4) deviation metrics (position/rotation) are computed against tolerances and the model is updated; (5) MR overlays communicate per stone status and deviation to the worker. The framework is validated in a controlled experiment setting and embedded in an academic course to generate labeled data, train models, and test the pipeline.

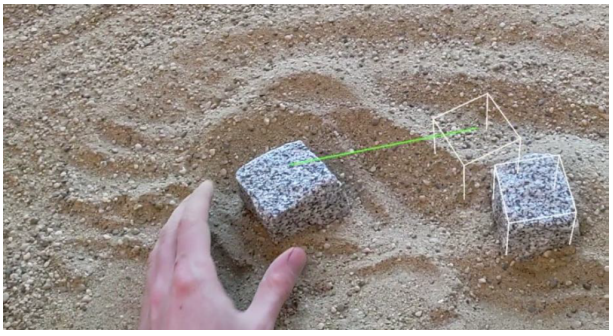


Figure 1: View through MR device view with AI-based guidance

Experimental Setup and Workflow

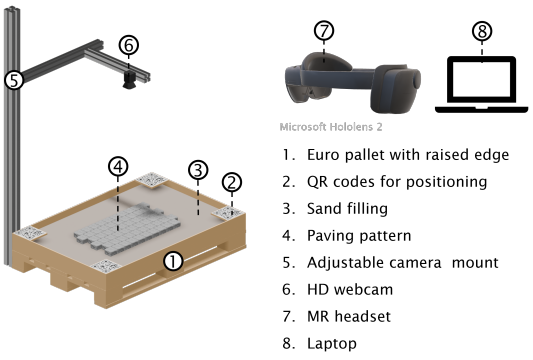


Figure 2: Schematic laboratory setup

The experimental setup consists of a portable paving box with sand bedding mounted on a Euro pallet, a consumer grade HD webcam, and a standard work laptop that hosts both the AI and the parametric pipeline; MR output is streamed to a MR-Device (HoloLens 2) over WLAN. The planned paving pattern is authored in Rhino 8 with Grasshopper as a data driven parametric model in which each stone is assigned an ID, Cartesian coor-

dinates (X, Y, Z), and rotations (Rx, Ry, Rz). Within this environment, Grasshopper components structure the input and output, perform perspective rectification, compute tolerances, and drive MR materials and state logic so that feedback appears immediately in the headset. For execution, a YOLOv11n seg model tracks stones (centroid, Z rotation) via a custom Python script, streaming results to Grasshopper. Four corner QR markers establish a 2D homography, allowing the system to map these video detections to model coordinates and compute deviations against the planned pattern. Finally, Fologram on HoloLens 2 visualizes this feedback, overlaying per stone status (e.g., position/rotation error) and providing interactive controls for the worker to adjust tolerances or reset the workflow.

Educational Integration

The framework was embedded in an academic course in which master's-level civil engineering student teams produced labeled datasets and trained vision models for the demonstrator. The course exposed students to the full CV pipeline, including data acquisition, labeling, training/validation, and deployment. By experimenting with different labeling strategies, including synthetic data, students generated multiple individual vision models as well as a shared dataset enabling more robust combined model training. This approach enables scalable training data generation while illustrating the practical steps required to apply CV methods in a traditionally low-tech craft.

Demo-Test and Results

The integrated setup was tested in a laboratory setting by a small group of experienced paving professionals using the full technology stack, including live CV inference and MR feedback. Under these controlled conditions, the pipeline achieved interactive update rates, providing stable 2D position feedback within centimeter-level tolerances for most stones. The configuration was considered sufficiently realistic to serve as a viable demonstrator for high-tech assistive tools. The combined use of parametric modeling, AI-based instance segmentation, and mixed reality was considered suitable for addressing a central challenge in paving: placing stones precisely while maintaining the intended pattern. A more comprehensive study with rigorous scientific evaluation is planned.

Discussion and Conclusions

The prototype demonstrates that real time 2D position control is feasible with low cost, readily available hardware; coupling a parametric model with AI in Grasshopper provided a transparent logic, and MR made tolerances legible directly at the point of work. At the same time, rotation estimates were not yet robust, QR markers were sensitive to dirt and glare, and tracking stability degraded under occlusions, which are limitations that could surface on real construction sites. Planned refinements include more reliable rotation cues, ruggedized relief markers, multi camera views, and the addition of depth sensing (e.g., LiDAR) to progress toward full 3D quality assurance and a more formal evaluation in follow up studies. This work contributes (i) an end-to-end, reproducible Parametric-CV-MR pipeline for craft scale quality control built on widely accessible tools, (ii) a didactic blueprint that can be adapted from a semester course to compact workshops and transferred beyond paving to related crafts, and (iii) an experimental setup to further investigate empirical evidence to clarify current strengths and gaps, providing a concrete roadmap toward field ready systems in under-digitized crafts.



Figure 3: Laboratory testing of demo setting

Acknowledgements

Supported by the EU H2020 program, project MAIA (GA-No. 873077), and the Interreg project WALZ 4.0.

References

Asmar, P. G. E., Chalhoub, J., Ayer, S. K., and Abdallah, A. S. (2021). Contextualizing benefits and limitations reported for augmented reality in construction research. 26(39):720–738.

Brucker Juricic, B., Galic, M., and Marenjak, S. (2021). Review of the Construction Labour Demand and Shortages in the EU. Buildings, 11(1):17.

McDonough, A. F. (2026). Exploring the Prospects of the Construction Industry for Youth: The Role of Technology and Educational Reform. International Journal of Information and Education Technology, 16(2):393–403.

Samsami, R. (2024). A Systematic Review of Automated Construction Inspection and Progress Monitoring (ACIPM): Applications, Challenges, and Future Directions. CivilEng, 5(1):265–287.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

CATEGORIZING RECURRING TASKS FOR FASTER AND CONSISTENT EMBODIED CARBON ASSESSMENTS

Eleni Alexandraki[1], Giulia Scagliotti[1], Calvin Kam[1, 2], Joy Li[2], and Martin Fischer[1]
[1]Stanford University, Stanford, Unites States of America
[2]Strategic Building Innovation (SBI), Palo Alto, United States of America

Abstract

Embodied carbon (EC) assessments are slowed by fragmented information handover, requiring extensive information-retrieval effort. This study decomposes EC workflows for windows across four building projects by categorizing recurring life cycle assessment (LCA) tasks. Results show that 71-81% of tasks involve data extraction and data wrangling rather than requiring EC expertise. Standardized EC property sets and consolidating data into one information source could reduce EC workflow tasks by 41-67%, improving assessment speed and consistency.

Introduction

The absence of standardized information handover slows embodied carbon (EC) assessments, despite rising demand driven by EC regulations (One Click LCA, 2022).

To complete an EC assessment, Life Cycle Assessment (LCA) specialists refine the Bill of Quantities (BoQ) and compile additional LCA-relevant information (Tam et al., 2022). This involves substantial effort in information-retrieval tasks such as data extraction (Matillion, 2024) and data wrangling (Wickham, 2014), often requiring consultation of multiple sources, including Building Information Models (BIM), 2D drawings, and project stakeholders (Scagliotti & Fischer, 2025).

Existing standards provide high-level guidance on BIM-based information management (ISO, 2018) and building LCA (Papakosta & Sturgis, 2017). Nevertheless, no standards developed from the LCA specialist's perspective define the LCA-relevant information that architecture, engineering and construction (AEC) stakeholders should provide for each building component (Abdelaal & Guo, 2022). Furthermore, the success of BIM to LCA data exchange through standardized file formats, such as Industry Foundation Classes, depends on whether the models have been properly enriched with LCA-relevant information (Theißen et al., 2020).

Given these insufficient requirements, it is necessary to determine what information should be included in the BIM (Guntermann et al., 2025). This study decomposes the EC workflow LCA specialists follow by categorizing the tasks completed, measuring their recurrence, and deriving LCA-relevant information in the case of windows. Establishing this building component information as a BIM requirement for AEC stakeholders would improve information handover and enable faster, more consistent EC assessments.

Methodology

To better understand where LCA specialists expend the most effort and why, we conducted cradle-to-gate EC assessments for windows in four residential projects, ranging from single-family houses to low- and high-rise multifamily buildings. We selected windows as a representative component because they are a carbon intensive element of the building envelope and are present in all residential buildings.

We systematically decomposed EC assessment workflows by analyzing recurrent LCA tasks and grouping them into three categories: (1) data extraction; (2) data wrangling; and (3) impact factor search and assignment. Each category includes several tasks typically performed by LCA specialists, leading to a total of 17 (A-Q), as shown in Figure 1.

Figure 1 shows the LCA tasks required to complete an EC assessment for windows in four projects. For example, in Case Study 1, out of 30 tasks, the LCA specialist requested five clarifications from the project team (data extraction) and performed six simple calculations (data wrangling). We hypothesize that many such tasks could be avoided through standardized information handover, enabling LCA specialists to focus on core activities, such as assigning environmental product declarations (EPDs).

Case Study Results and Discussion

Decomposing the EC assessment workflow for windows revealed that most LCA tasks revolve around creating or enriching a BoQ and do not require EC expertise. As shown in Figure 1, data extraction accounts for 24-30% of total tasks, while data wrangling represents 43-56%.

In contrast, tasks requiring genuine EC expertise, such as searching and assigning impact factors, account for only 9%-29% of total tasks. Therefore, AEC stakeholders should improve BoQ information handover, so LCA specialists can focus on tasks requiring their expertise.

Decomposing the EC workflow also revealed the LCA information requirements for windows. Multiple sources (e.g., BIM, 2D drawings) were used to compile the geometric (e.g., dimensions, area, volume) and non-geometric (e.g., number of panes, glass type) properties required for the EC assessment. The reliance on multiple

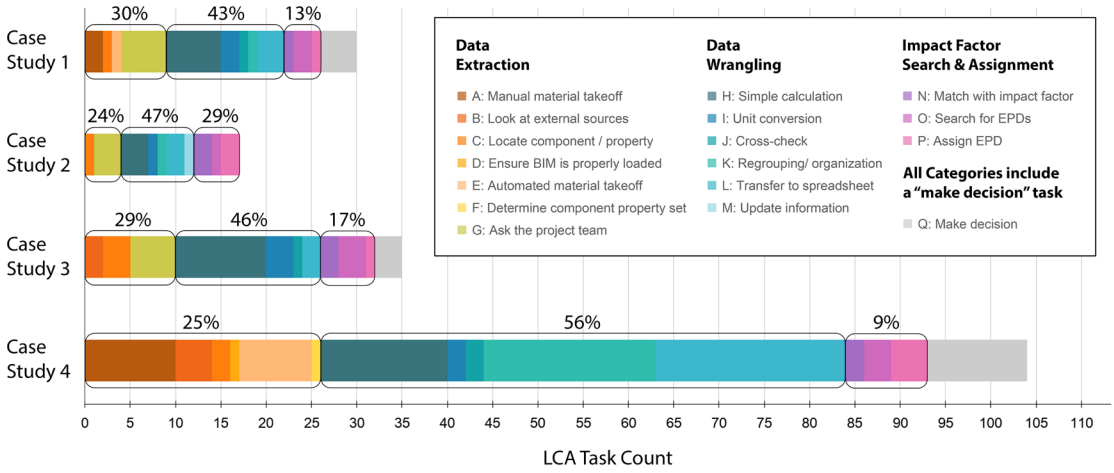


Figure 1: Recurring LCA tasks required to complete EC assessments for windows. Tableau color palette based on Arnold.

sources to retrieve LCA-relevant properties contribute to slower and inconsistent EC assessments.

Improving information handover can make EC workflows faster and more consistent. This can be achieved if project teams (1) have a standardized LCA-specific property set for each building component and (2) consolidate all LCA-relevant information into a single BIM. Building on existing BIM property set definitions (Guntermann et al., 2025), we defined an EC property set for windows. Revisiting the EC workflows for windows, we demonstrated that data extraction and wrangling tasks could be largely eliminated. This approach could reduce total task counts by 41%-67% across the four case studies, significantly streamlining EC assessments.

Conclusions

To make EC assessment faster and consistent, we identified and categorized recurring LCA tasks using windows as a representative building component from four case studies. Future work should extend this method to other components and building types to confirm the effort in data extraction and wrangling for completing EC assessments.

Our results demonstrate that providing completed and standardized EC property sets and consolidating data into a single BIM can reduce LCA task counts by roughly half. By shifting effort from data extraction and data wrangling to impact factor search and assignment, LCA specialists can focus on tasks requiring genuine EC expertise.

References

Arnold, J. Tableau Color Palettes (discrete). Available at: https://jrnold.github.io/ggthemes/reference/tableau_color_pal.html (Accessed: 22 Feb. 2026).

Abdelaal, F. & Guo, B.H., 2022. Stakeholders' perspectives on BIM and LCA for green buildings. Journal of Building Engineering, 48, p.103931.

Guntermann, L., Hagedorn, P. and König, M., 2025, July. Enhancing IFC Models for Circular Design: Validation of Required Information through an IDS Framework. In Proceedings of the 42nd International Symposium on Automation and Robotics in Construction (ISARC 2025), Montreal, QC, Canada, pp. 28-31.

International Organization for Standardization (ISO), 2018. ISO 19650: Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) – Information management using building information modelling. Geneva: ISO.

LCA, One Click, 2022. Construction carbon regulations in Europe. Rev. Best Pract.

Matillion (2024) Guide to Data Extraction: Definition, how it works & examples. Available at: https://www.matillion.com/blog/what-is-data-extraction-everything-you-need-to-know (Accessed: 4 April 2026).

Papakosta, A. & Sturgis, S., 2017. Whole life carbon assessment for the built environment. London, UK: Royal Institution of Chartered Surveyors, RICS Professional Standards and Guidance.

Scagliotti, G. & Fischer, M., 2025. Practical challenges in assessing embodied carbon and cost trade-offs in the design phase: Insights from a residential building project. In: CIB Conferences, 1(1), p.40.

Tam, V.W., Zhou, Y., Illankoon, C. & Le, K.N., 2022. A critical review on BIM and LCA integration using the ISO 14040 framework. Building and Environment, 213, p.108865.

Theißen, S., Höper, J., Drzymalla, J., Wimmer, R., Markova, S., Meins-Becker, A. & Lambertz, M., 2020. Using open BIM and IFC to enable a comprehensive consideration of building services within a whole-building LCA. Sustainability, 12(14), p.5644.

Wickham, H., 2014. Tidy data. Journal of Statistical Software, 59, pp.1–23.

AI Use: Based on a version written by the authors, the title and 100-word abstract were iterated upon with ChatGPT. The authors post-edited the text and take full responsibility for it.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

THICKNESS SIMULATION WITH KINEMATIC CONSTRAINTS FOR ROBOTIC FIREPROOFING SPRAY

Kangmin Bae[1], Sangmin Lee[1], Minseung Cha[1], Sebeen Yoon[1] and Taehoon Kim[1]
[1]Seoul National University of Science and Technology, Seoul, Republic of Korea

Abstract

Achieving precise thickness is essential for fireproofing performance. Insufficient thickness fails to secure the required fire resistance, whereas excessive thickness causes material waste and rework. This study proposes a simulation framework that combines a Gaussian accumulation model with a kinematic nozzle motion model. As a result of tests on a fixed raster path, kinematic constraints suppressed excess accumulation, increasing the within range area from 87.55% to 91.16% and reducing the over thickness area from 6.83% to 0%. These results show that spray quality should be evaluated using CV with under/over coverage, within range area to help ensure the required thickness.

Introduction

In fireproofing spray, keeping the thickness within the target range is important for both fireproofing performance and material efficiency. If the thickness is insufficient, the required fire resistance cannot be secured and additional repair or rework may be needed. In contrast, excessive thickness causes unnecessary consumption of spray material.

Recent automated spraying research has focused on trajectory generation and velocity optimization for complex free form surfaces (Verduyn et al., 2023; Nieto Bastida and Lin, 2023). However, unlike thin film paint, heavy construction materials such as wet mix shotcrete and fireproofing form thickness differently, and this process is influenced by spray dynamics (Ning et al., 2024). At corners, robot kinematic constraints such as acceleration and deceleration limits can reduce nozzle speed and increase local thickness, but this effect is not directly addressed in macroscale 3D trajectory planning.

This study presents a simulation framework for robotic fireproofing spray and evaluates how kinematic limit scaling near rounded corners affects thickness maps and the area with thickness within the target range.

Methodology

This study develops a thickness simulation with kinematic constraints for robotic fireproofing spray on a planar target surface. A fixed L shaped raster path with rounded corners is used, and the region of interest is evaluated on a 5 mm grid. A single elliptical Gaussian footprint is used for accumulation. The path is resampled at 1 mm intervals for kinematic profiling and 20 mm for accumulation.

This dual resampling balances motion precision with computational efficiency. According to Eq. (1), local thickness is accumulated along the path, with each accumulation contribution determined by the effective flow rate, the Gaussian footprint, and the inverse of the local travel speed.

$$T(x,y) = \int_s \frac{Q\eta}{v(s)} G(x,y,s)ds \quad (1)$$

Two motion settings are compared. OFF uses the nominal kinematic limits over the full path. ON applies corner speed, acceleration, and deceleration scaling of 1.5 near rounded corners. The predicted thickness map is evaluated using mean thickness, CV, under coverage below 17 mm, within range area for 17 to 19 mm, and over coverage above 19 mm.

Table 1: Baseline simulation settings

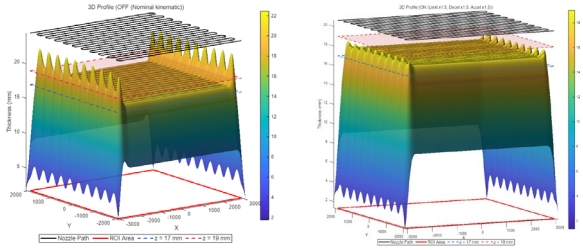
Parameter	Value	Description
Q	36 L/min	Volumetric flow rate
η	0.90	Transfer efficiency
h	300 mm	Standoff distance
α	30°	Nozzle half-angle
k_σ	0.70	Spread factor
Δ	5 mm	Grid resolution
ROI	6000x4000 mm	Spray region
s_{row}	200 mm	Row spacing
r_{turn}	150 mm	Turn radius
v_{min}, v_{max}	100 ~ 150 mm/s	Speed bounds
$a_{tan,max}$	200 mm/s^2	Tangential acceleration limit
$a_{lat,max}$	200 mm/s^2	Lateral acceleration limit
ds_{kin}, ds_{dep}	1 mm, 20 mm	Sampling steps
Footprint	285.51 mm^2	Calculated spray area

OFF	-	Normal limit
ON	1.5	Kinematic limit scaling factor
Mean	18 mm	Target thickness mean

Results

Figure 1 compares OFF and ON at 150 mm/s, the upper bound of the tested speed range reached in both settings. In the OFF case, ridge like peaks appear near the turn side boundary. In the ON case, these local peaks are suppressed and the boundary profile becomes flatter.

Compared with OFF, ON reduced mean thickness from 17.94 mm to 17.60 mm and increased CV from 5.46% to 6.88%, while over coverage decreased from 6.83% to 0.00% and within range area increased from 87.55% to 91.16%.



(a) (b)

Figure 1: 3D Profile OFF(a), ON(b)

Discussion

OFF shows a lower CV than ON, but ON shows lower over coverage and higher within range area. In ON, over coverage decreases from 6.83% to 0.00%, while under coverage increases from 5.62% to 8.84%, which lowers mean thickness and increases CV. Therefore, CV should be interpreted together with under coverage, within range area, and over coverage, especially near rounded corners.

Table 2: Sensitivity Analysis (ON setting)

Parameter	Variation	CV (%)	Target Range Compliance (%)
Baseline	-	6.88	91.16
k_σ	±10%	6.60 ~ 7.31	91.64 ~ 92.25
η	±5%	6.88 ~ 6.91	91.27 ~ 92.18
Kinematic limit	±20%	6.88 ~ 6.89	91.16 ~ 91.16
Δ	1, 10mm	6.74 / 7.06	91.26 / 90.86

Table 2 summarizes the results obtained by varying k_σ, η, kinematic limit scaling factor and Δ around their baseline values. For the tested variations, within range area remained between 90.86% and 92.25%, and CV remained between 6.60% and 7.31%. Notably, within the speed range specified in this study, no significant changes due to variations in the kinematic limits were observed.

Conclusions

This study presented a thickness simulation with kinematic constraints for robotic fireproofing spray on a planar target surface. The main finding is that kinematic limit scaling near rounded corners reduced over coverage and increased within range area, even though CV increased. This shows that CV alone is not sufficient for evaluating spray quality. Instead, it should be interpreted together with under coverage, within range area, and over coverage. These evaluation metrics can be used to contribute to securing the required fireproofing thickness. Future work will extend the present framework beyond the fixed assumptions used in this study by considering additional flow rates, standoff distances, and model and kinematic parameters. Experimental validation of the predicted thickness maps will also be conducted under practical fireproofing conditions.

Acknowledgments

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant No. RS-2025-11802969).

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant No. RS-2022-00143493).

References

ASTM International (2023) ASTM E605/E605M-19(2023) Standard Test Methods for Thickness and Density of Sprayed Fire-Resistive Material (SFRM) Applied to Structural Members.

Conner, D.C., Greenfield, A., Atkar, P.N., Rizzi, A.A. & Choset, H. (2005) Paint Deposition Modeling for Trajectory Planning on Automotive Surfaces. IEEE Transactions on Automation Science and Engineering, 2(4), pp.381–392. doi: 10.1109/TASE.2005.851631.

Verduyn, A., De Schutter, J., Decré, W. & Vochten, M. (2023) Shape Based Path Adaptation and Simulation Based Velocity Optimization of Initial Tool Trajectories for Robotic Spray Painting. In: 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE). doi: 10.1109/CASE56687.2023.10260519.

Nieto Bastida, S. & Lin, C.Y. (2023) Autonomous Trajectory Planning for Spray Painting on Complex Surfaces Based on a Point Cloud Model. Sensors, 23(24), 9634. doi: 10.3390/s23249634.

Ning, F.W., Sui, W., Bai, Y. & Chen, B. (2024) Mass Distribution Characteristics of Wet Mix Shotcrete During Spraying Process and Adhesion Process. Construction and Building Materials, 433, 136485. doi: 10.1016/j.conbuildmat.2024.136485.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

RGB-ONLY WORK-AREA RECOGNITION FOR FIREPROOFING SPRAY ROBOT

Minseung Cha, Kangmin Bae, Sangmin Lee, Sebeen Yoon, and Taehoon Kim
Seoul National University of Science and Technology, Seoul, Republic of Korea

Abstract

Deep learning–based segmentation is widely used for construction work-area recognition. But it may lead to mask boundary misalignment, degrading work quality in processes requiring precise work-area recognition, such as fireproofing spray. To address this limitation, we present a pipeline for accurate work-area recognition of steel structures. It uses deep learning for segmentation and edge detection, followed by computer vision for line extraction and work-area segmentation. Experiments show mean Intersection over Union (mIoU) gains of 4.47 pp before spraying and 4.51 pp during spraying over the segmentation model. These results indicate our pipeline enables more accurate work-area segmentation.

Introduction

Fireproofing spray is essential for delaying the loss of load-bearing capacity of steel structural members under fire exposure. However, it is still performed manually at height on many construction sites, posing substantial risks due to airborne dust and fall hazards. Accordingly, interest in fireproofing spray robots has increased (Ikeda et al., 2020). For reliable operation, the robot typically needs to know where to spray, which requires accurate work-area recognition. In construction robotics, work-area recognition is often treated as a deep learning-based segmentation task in which predicted masks are used directly for localization and downstream processing (Dung et al., 2019; Wu et al., 2024). However, in fireproofing spray, small boundary misalignment in these masks can degrade work quality. To improve boundary reliability, some approaches combine depth sensing and 3D processing. However, these pipelines typically require additional sensors such as RGB-D or LiDAR, increasing hardware cost (Schmidt et al., 2024). Moreover, a fireproofing spray robot must recognize each surface separately because the spray position and robot pose differ across surfaces such as the web and flanges. Therefore, this study proposes an RGB-only work-area recognition pipeline for a fireproofing spray robot in construction robotics, enabling separate work-area segmentation for each surface of steel structures.

Methodology

Figure 1 illustrates the overall framework of the proposed RGB-only work-area recognition pipeline. First, an instance segmentation model based on YOLOv11m-seg segments a steel structure into surface classes and outputs binary masks (Khanam et al., 2024). Next, for each class, we generate an edge map using a Lightweight Dense CNN for Edge Detection (LDC) model (Soria et al., 2022). Since the extracted edge map is produced in grayscale, we eliminate noise through binarization and contour-length–based filtering. From the refined edge maps, boundary lines are extracted through Random Sample Consensus (RANSAC) based line detection. Because the edge detector responds on both sides of the same physical boundary, RANSAC-based line detection produces a pair of nearby lines for one internal boundary. Therefore, we apply midpoint correction by averaging the endpoints of the neighboring line pair to obtain a single representative boundary line. Finally, the outer boundary lines and the corrected internal boundary lines are used to segment the

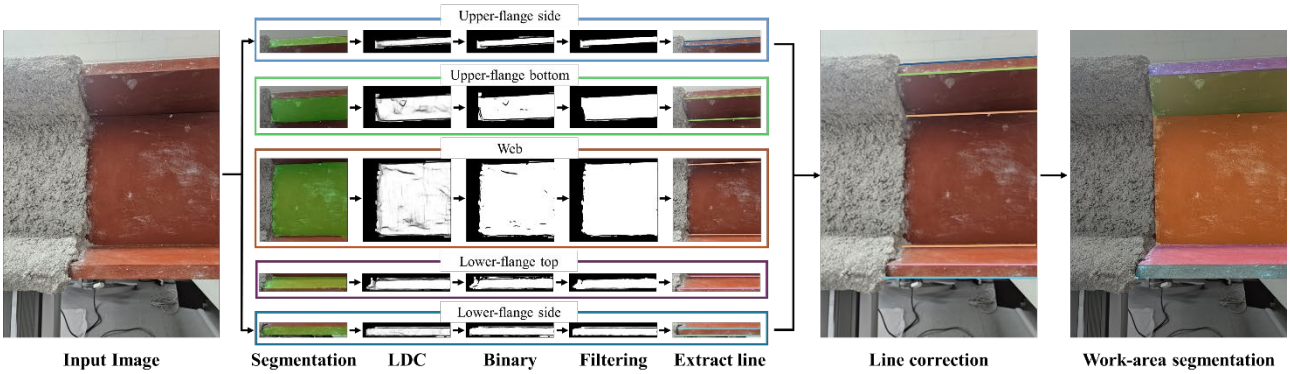
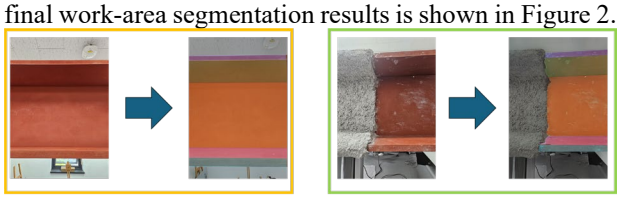


Figure 1: Framework

steel structures into surface work areas. An example of the final work-area segmentation results is shown in Figure 2.



Before spray During spray

Figure 2: Example of the final outputs

Experiment

Preparing Dataset

To evaluate the proposed pipeline, we fabricated a Medium Density Fiberboard (MDF) H-beam mock-up and collected RGB images from viewpoints relevant to robotic operation. A total of 250 images were used for model development, and an additional 20 images were used for quantitative evaluation. The evaluation set included both before-spraying and during-spraying conditions.

Train Segmentation Model

The YOLOv11m-seg model was trained using a 7:2:1 split for training/validation/test. During training, the segmentation classes were defined as five surfaces: Upper-flange side, Upper-flange bottom, Web, Lower-flange top, and Lower-flange side.

Evaluation metrics

Performance was measured using Intersection over Union (IoU) as in (1) and mean IoU (mIoU) as in (2).

$$IoU = \frac{|Prediction \cap Ground\ Truth|}{|Prediction \cup Ground\ Truth|} \quad (1)$$

$$mIoU = \frac{1}{N}\sum_{i=1}^{N} IoU_i \quad (2)$$

Results and Discussion

Table 1 summarizes the quantitative results by comparing the proposed pipeline with the segmentation model. It reports IoU for each of the five surfaces and mIoU averaged across the five surfaces, both before spraying and during spraying.

Table 1: Quantitative results

	Before spraying		During spraying	
	Proposed	Seg.model	Proposed	Seg.model
Upper-flange top	0.8877	0.9125	0.9509	0.9285
Upper-flange bottom	0.9846	0.8950	0.9718	0.9190
Web	0.9931	0.9210	0.9891	0.9274
Lower-flange top	0.9512	0.8988	0.9609	0.9216
Lower-flange bottom	0.9264	0.8923	0.9694	0.9200
mIoU	**0.9486**	**0.9039**	**0.9684**	**0.9233**

The quantitative results show that the proposed pipeline improves work-area recognition compared with the segmentation model. Before spraying, mIoU increases from 0.9039 to 0.9486, an improvement of 4.47 percentage points. During spraying, mIoU increases from 0.9233 to 0.9684, an improvement of 4.51 percentage points. These results indicate that the proposed pipeline enables more accurate separation of surface work areas on steel structures. However, the evaluation was conducted on an MDF H-beam mock-up rather than actual steel structures, and only 20 images were used for evaluation. In addition, the pipeline was sensitive to outdoor shadows, which were sometimes mistaken for structural edges and caused incorrect work-area segmentation.

Conclusions

This study presents an RGB-only work-area recognition pipeline for robotic fireproofing. The proposed pipeline uses deep learning for segmentation and edge detection, followed by line extraction and midpoint-based boundary refinement to derive final surface work areas on H-beams. Experiments on MDF mock-up showed improvements over the segmentation model, achieving mIoU 0.9486 before spraying and 0.9684 during spraying. These results suggest that the proposed method can provide more precise work-area recognition than using the segmentation model alone. Future work will validate the pipeline on actual steel structures under diverse field conditions and improve robustness to edge noise caused by shadows.

Acknowledgments

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant No. RS-2025-11802969).
This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant No. RS-2022-00143493).

References

Ikeda, Y., Segawa, H. & Yabuki, N. (2020) Development and application of a fire resistive covering spraying robot to building construction site. Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC 2020), pp.1464–1471.

Dung, C. V. & Le, D. A. (2019) Autonomous concrete crack detection using deep fully convolutional neural network. Automation in Construction, 99, pp.52–58.

Wu, D., Xu, X., Li, R., Peng, X., Gong, X., Lee, C.-H., Pan, P. & Jiang, S. (2024) Curtain wall frame segmentation using a dual-flow aggregation network: Application to robot pose estimation. Automation in Construction, 168, p.105816.

Schmidt, P., Katsatos, D., Alexiou, D., Kostavelis, I., Giakoumis, D., Tzovaras, D. & Nalpantidis, L. (2024) Towards autonomous shotcrete construction: semantic 3D reconstruction for concrete deposition using stereo vision and deep learning. Proceedings of the 41st International Symposium on Automation and Robotics in Construction (ISARC 2024), pp.896–903.

Soria, X., Pomboza-Junez, G. & Sappa, A. (2022) LDC: Lightweight dense CNN for edge detection. IEEE Access, 10, pp.68281–68290.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

ANOMALY DETECTION IN INDOOR ENVIRONMENTAL SENSOR DATA BASED ON A DIGITAL TWIN

Arezoo Fathollahizenooz, Cornelius Preidel, and Simon Vilgertshofer
HM Munich University of Applied Sciences, Munich, Germany

Abstract

Anomaly detection in building sensor networks often relies on static thresholds that fail to capture temporal dependencies and system changes. This work combines a baseline model of expected indoor environmental behavior with anomaly detection to form a data-driven Digital Twin. In this framework, a regression model learns expected behavior from historical data, while deviations are quantified as multivariate residuals and analyzed using unsupervised and semi-supervised methods. Preliminary analysis on a representative campus sensor shows detection of anomalous behavior and associated distributional shifts, with sensor relocation serving as an illustrative system change.

Introduction

Sensor networks are used in modern buildings to monitor indoor environmental quality (IEQ). However, due to the scale of the data, manual inspection is insufficient to detect anomalies. Consequently, research has shifted toward data-driven anomaly detection. Unsupervised and semi-supervised methods are used to detect IEQ deviations without the need for large labeled datasets (Chen et al., 2023; Bi et al., 2024). Although methods such as Long Short-Term Memory can capture complex temporal dynamics, they often lack an integrated analysis of normal system operation (Wei et al., 2022; Rajabi and McArthur, 2023). To address this gap, this work investigates whether DT-based anomaly detection can identify environmental irregularities in indoor sensor data. Residual-based anomaly detection is integrated with a Digital Twin (DT) representing expected indoor environmental behavior, while anomalies are identified as deviations between physical observations and digital expectations, enabling detection of patterns suggestive of HVAC malfunctions or sustained humidity conditions associated with increased mold risk.

Dataset

The proposed framework incorporates indoor environmental data from a campus-wide sensor infrastructure that offers ongoing time-series data from various rooms in university buildings. Temperature, humidity, CO_2 concentration, illumination, and motion are included in the dataset. The possibility of sensor relocation, which adds uncertainty about the spatial context, is an important feature of this dataset.

Digital Twin

The physical layer in DT represents the building space, while sensors measuring indoor environmental variables serve as the interface between the physical asset and its digital representation. The digital layer consists of the recorded time-series data that describe the state of the system. A Random Forest regression model (Breiman, 2001) serves as the analysis component of the Digital Twin, producing predictions $\hat{y}_t = f_{\mathrm{RF}}(\mathrm{x}_t)$, where x_t comprises time-based and lagged features capturing temporal dependencies and non-linear relationships. Hyperparameters (n_estimators = 300, max_depth = 10, min_samples_leaf = 5) are selected with GridSearchCV with 5-fold time-series cross-validation, and the predictive performance is reported using MAE and R^2 on the final 20% of baseline data. This coupled physical-digital representation enables detection of distributional shifts indicative of structural changes, such as sensor relocation, while providing a learned baseline that adapts as system conditions evolve.

Residual-Based Anomaly Detection

Anomaly detection operates on residuals $r_t = y_t - \hat{y}_t$, where y_t are the observed measurements and $\hat{y}_t$ is the DT prediction at time t. Multivariate residuals represent these deviations, enabling identification of abnormal behavior relative to learned expectations. Anomaly detection is implemented using two methods: Unsupervised approaches based on Isolation Forest (Liu et al., 2008) are applied to identify deviations from learned temporal patterns in multivariate sensor data without requiring labeled anomalies. In addition, a semi-supervised setting using One-Class Support Vector Machines (OCSVM) (Schölkopf et al., 2001) is explored. For OCSVM training, domain knowledge determines inlier conditions (20–25 °C temperature, 20–60% humidity), whereas observations outside these ranges are used to assess anomaly-detection behavior.

Results and Discussion

Period segmentation is applied in the initial analysis of a particular sensor (S12) from June to October 2023. Gaps longer than six hours, which are unlikely under normal operations, are used as indicators of potential system disruption and divide the data into baseline (Jun–Jul; 56,738 measurements), relocation/sensor outage (July–October; 5,510 measurements), and post-relocation (October; 3,884 measurements) as shown in figure 1. The trained multivariate DT baseline model exhibits substantial increases

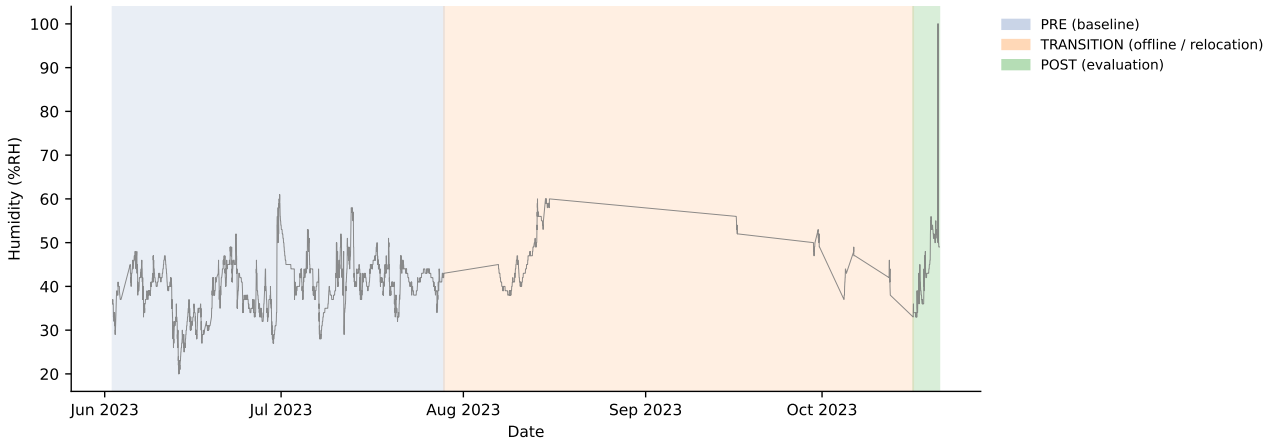


Figure 1: Period segmentation based on sensor gaps

in prediction error in all variables (temperature, humidity, and CO_2) when applied to post-relocation data, confirming a significant shift in distribution. Figure 2 shows this quantitative degradation by a clear shift in the residual distributions across the variables. During the first unsupervised phase of this framework, an Isolation Forest model identified every post-relocation observation as anomalous, indicating a distributional shift. Next, semi-supervised learning methods will be used to enhance the classification of these deviations. Although a representative, operationally stable sensor (S12) is selected from the network of 82 sensors to illustrate the framework, extension to the full sensor network is part of ongoing work.

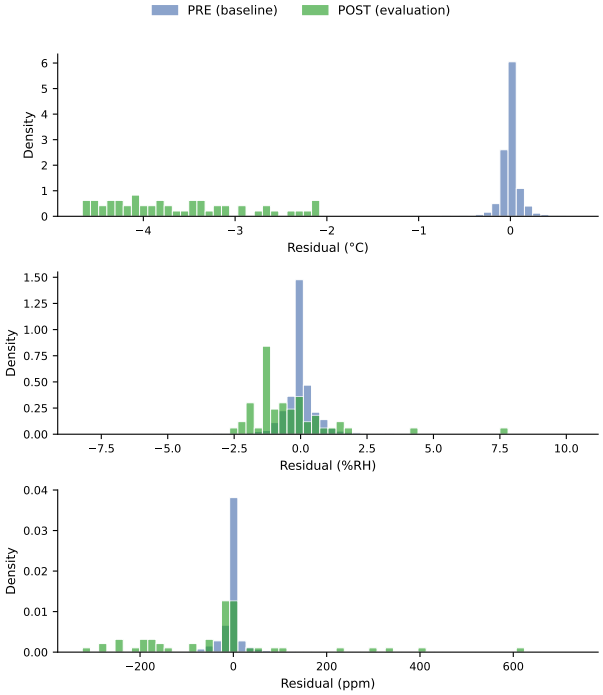


Figure 2: Residual distribution shift across variables

Summary

This research presents a DT-based framework for anomaly detection in indoor environmental sensor data. A Random Forest baseline learns expected behavior from historical data, and residual analysis detects distributional shifts with Isolation Forest. Semi-supervised OCSVM for environmental anomaly detection and adaptive DT updating are the next steps in ongoing work.

References

Bi, J., Wang, H., Yan, E., Wang, C., Yan, K., Jiang, L., and Yang, B. (2024). AI in HVAC fault detection and diagnosis: A systematic review. Energy Review, 3.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1):5–32.

Chen, Z., O'Neill, Z., Wen, J., Pradhan, O., Yang, T., lu, X., Lin, G., Miyata, S., Lee, S., Shen, C., Chiosa, R., Piscitelli, M., Capozzoli, A., Hengel, F., Kührer, A., Pritoni, M., Liu, W., Clauß, J., Chen, Y., and Herr, T. (2023). A review of data-driven fault detection and diagnostics for building HVAC systems. Applied Energy, 339:121030.

Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation Forest. In 2008 Eighth IEEE International Conference on Data Mining.

Rajabi, F. and McArthur, J. J. (2023). Unsupervised Learning for Fault Detection of HVAC Systems: An optics -based approach for terminal air handling units.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. Neural Computation.

Wei, Y., Jang-Jaccard, J., Xu, W., Sabrina, F., Camtepe, S., and Boulic, M. (2022). LSTM-Autoencoder based Anomaly Detection for Indoor Air Quality Time Series Data.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

BIMCHAINED: A BLOCKCHAIN SUPPORTED DIGITAL FRAMEWORK FOR REAL-TIME CONSTRUCTION SITE MONITORING

Eleonora Vendetti, Giuseppe Piras
Sapienza University of Rome, DIEE Dpt-Lean Construction for the digital design of buildings and environment

Abstract

The construction industry uses BIM and Common Data Environments (CDE) for coordination, nevertheless struggles with data integrity, auditability, and trusted verification for safety compliance amid fragmented processes and BIM-blockchain oracle issues. This study proposes a hybrid framework integrating BIM, Dynamo-Python automation for parameter extraction and data validation, cryptographic hashing, and QR-codes linking physical assets to blockchain-ready records. In a Rome heritage refurbishment case study on mobile scaffolding, it produced verifiable inspection registers with compliance status and traceable hashes. The key contribution is a scalable prototype enhancing BIM safety monitoring traceability, enabling blockchain anchoring and openBIM-aligned digital twin provenance.

Introduction and background

The construction industry is currently undergoing a digital transformation to address systemic inefficiencies, fragmented information flows, and high accident rates. While BIM has established a foundational infrastructure for project coordination and clash detection, it still mostly depends on manually updated models and struggles to guarantee data integrity and auditability (Ng et al., 2025). In multi-stakeholder environments, manual verification and paper-based certifications persist, while digital records remain vulnerable to unauthorized modifications, which is particularly critical for safety and regulatory compliance.

CDE facilitate centralized repositories but lack immutable audit trails to link digital models reliably with physical on-site outputs. Integration with blockchain technology offers a distributed paradigm for information management, providing tamper-resistant records and verifiable transaction histories through cryptographic verification. Recent research explores hybrid architectures where large geometric datasets remain in BIM repositories while blockchain infrastructures secure metadata and hash-based fingerprints to help reduce risks of data inconsistency (Wang & Wang, 2025).

Despite this potential, significant gaps remain in the empirical implementation of these technologies within a cohesive, governed pipeline for safety workflows. A primary technical barrier is the oracle problem, the difficulty of ensuring off-chain site data accurately reflects physical conditions before blockchain anchoring (Khuc et al., 2024). To face this, the study employs Dynamo-based computational workflows to support the traceability of physical inspections through cryptographic validation; while visual programming is well-established in BIM design, its systematic application to safety compliance verification remains underexplored. The study targets quality improvement of H&S site inspections, which are labour-intensive, inherently exposed to human error, selecting mobile scaffolding as proof-of-concept, given its role among the leading causes of workplace injuries in construction making systematic and automated compliance verification both technically relevant and practically necessary.

Methodology

The methodology consists in an enhanced integration of BIM environments, Python-based Dynamo scripts, and cryptographic algorithm techniques. The primary objective is to improve transparency and traceability by creating structured, immutable compliance records for subsequent blockchain infrastructure addressing the need for data provenance in collaborative environments where shared datasets risk being revised, overwritten, or lost.

The initial stage focuses on the data modelling layer, defining BIM requirements and safety goals through regulatory analysis, to establish key design principles for parameter selection and rule-based validation logic.

The core of the operational methodology is a BIM–Dynamo pipeline designed for automated parameter extraction and compliance checking, while further developments of the Digital Twin (DT) paradigm could enable continuous synchronization with IoT sensors and actuators to support safety monitoring and data integration (Wu & Li, 2025). Safety-related parameters are extracted directly from the digital model and processed through a Python script integrated into the visual programming environment. In the case study conducted on a heritage masonry building refurbishment project in Rome, five safety-critical parameters were systematically extracted from the BIM model and evaluated against the requirements of UNI EN 1004:2021: configured height, load class, installation context, stabilizer configuration, and guardrail status. The pipeline produced structured inspection registers in which each scaffolding configuration is associated with a compliance label, a cryptographic hash functioning as a unique digital

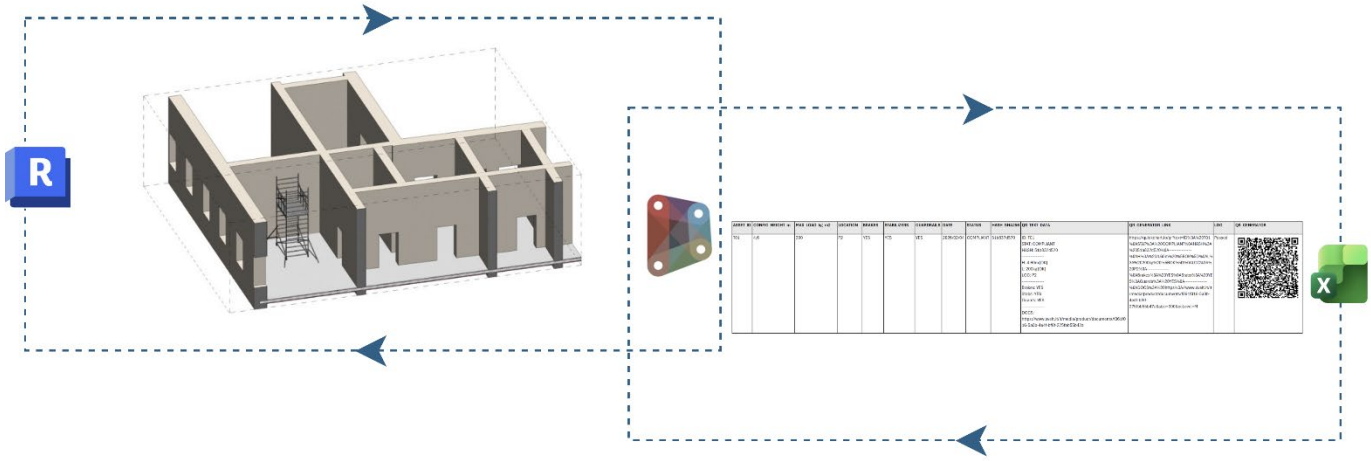


Figure 1: BIM–Dynamo workflow from Revit model to QR-coded inspection register in Excel for mobile scaffolding safety validation.

fingerprint, and a QR code enabling traceability between the physical asset and its digital record within the CDE.

This output represents a measurable improvement over conventional inspection documentation methods, reducing reliance on paper-based or informally managed records while ensuring that only validated, standardized data is prepared for subsequent blockchain checking stage.

Conclusion and discussion

The implementation of blockchain-enabled validation within BIM-based safety workflows contributes to the enhancement of data integrity and compliance documentation. By embedding validation logic directly into BIM platforms via Dynamo and Python, the system supports automated extraction of safety parameters and related compliance assessment. This configuration strengthens the reliability of inspection datasets while supporting structured documentation in multi-stakeholder environments. However, several systemic constraints persist, notably the *oracle problem*, which continues to pose challenges in establishing trusted links between physical site activities and digital verification systems. Additionally, interoperability limitations, scalability concerns, and organizational resistance remain significant barriers to sectoral adoption. Future research will focus on aligning this framework with openBIM specifications and developing more mature smart contract integration scenarios, guided by LLM capabilities integration.

References

Khuc, T. Q., Nguyen, V. T., & Do, S. T. (2024). Barriers to the adoption of blockchain technology in the construction industry: a total interpretive structural modeling (T ISM) and DEMAT EL approach. *Construction Innovation: Information, Process, Management*, *26*(1), 260–285. https://doi.org/10.1108/CI-12-2023-0307

Ng, R. C. W., Cheng, J. C. P., Cheng, G. C. W., Fung, K. H., & Fong, C. W. (2025). Performance-based payment mechanism for common data environment (CDE) adoption in construction projects. *Automation in Construction*, *174*, 106089. https://doi.org/10.1016/j.autcon.2025.106089

Wang, Y., & Wang, X. (2025). Values, application scenarios and challenges of blockchain technology: A perspective from construction 4.0. *KSCE Journal of Civil Engineering*, *29*(8), 100281. https://doi.org/10.1016/J.KSCEJ.2025.100281

Wu, Y., & Li, Z. (2025). Digital twin-enabled BIM-blockchain integration for automated and transparent construction progress payments. *Developments in the Built Environment*, *24*, 100781. https://doi.org/10.1016/J.DIBE.2025.100781

2026 European Conference on Computing in Construction
Corfu, Greece
July 12–15, 2026

A CROSS-DOMAIN FRAMEWORK FOR EXPLAINABLE STRUCTURAL INSPECTION

Vedant Girish Dalvi, Cornelius Preidel, and Simon Vilgertshofer
HM Munich University of Applied Sciences, Munich, Germany

Abstract

The digital transformation of structural inspection has led to the widespread use of image-based artificial intelligence for automated damage detection. Deep learning models, particularly convolutional neural networks, have demonstrated strong capability in identifying surface-level defects such as cracks from photographic data. However, although these systems provide pixel-level predictions, they lack mechanisms for semantic interpretation, cross-domain data integration, and structured knowledge representation. For deployment in civil engineering workflows, AI-based inspection must move beyond detection toward contextual reasoning and interoperable knowledge capturing.

Introduction

This paper presents a conceptual cross-domain framework that connects four layers of structural damage intelligence: (1) damage detection and quantification, (2) semantic damage captioning, (3) ontology-based data integration, and (4) knowledge graph population, as shown in Fig. 1. The proposed framework aims to bridge geometric, semantic, and relational representations of damage within digital construction environments.

Methodology

The proposed framework facilitates not only the detection and quantification of pixel-level damages and the explanation of contextual relationships in overlapping damage scenarios, but also the standardization of these damages through the modification of an existing ontology. This enables the population of a knowledge graph and, ultimately, the development of a digital twin, as discussed in the following sections.

Damage Detection and Quantification

The first layer focuses on multi-label damage detection from inspection images belonging to three different damage domains: concrete, steel, and glass. These domains differ in terms of damage type, causality, severity, complexity, and relationships. Custom multi-label semantic segmentation models are developed to detect damages across these domains, with domain-specific models later combined via deterministic orchestration or AI agents. Currently, models have been trained and successfully inferred on pre-existing concrete damage datasets like the

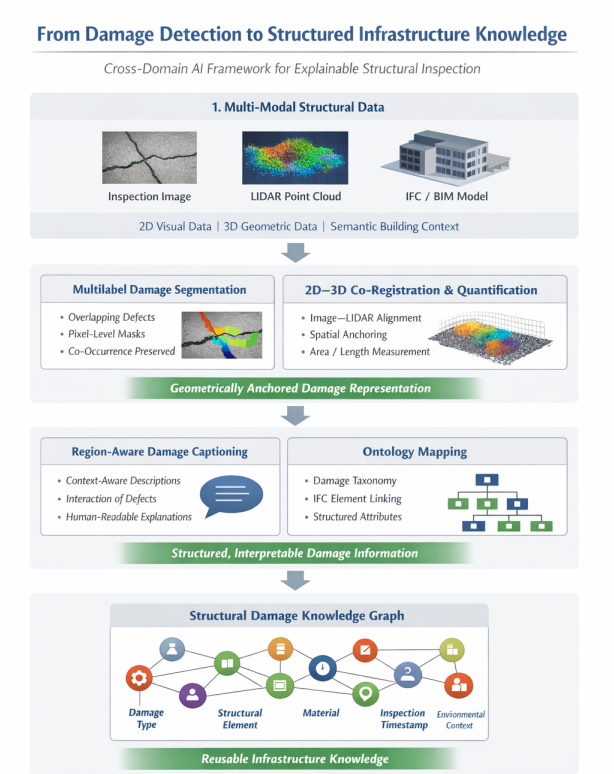


Figure 1: Proposed Methodology

DACL10k dataset introduced in Flotzinger et al. (2024) and new benchmark dataset created for this purpose, as described by Krefft et al. (2026).

Beyond 2D image analysis, the framework incorporates co-registration between image data and LiDAR-based geometric models. By aligning segmented damage regions with three-dimensional point clouds, damage localization can be anchored in spatially accurate structural coordinates. This integration enables the quantification of defect dimensions, and spatial orientation, as demonstrated by Krefft and Hoegner (2025). The result is a damage mask and parametric damage metadata, which can then be linked to a 3D model of the structure.

Semantic Damage Captioning

While detection provides spatial precision and quantification, engineering interpretation requires contextualization. The second layer introduces vision-language modeling for damage captioning. These captions can explain the over-

lapping patterns, and relative pixel positions among defects. Visual Language Models (VLMs) have been fine-tuned on a curated image captioning dataset to explain the context in overlapping concrete damages, and initial results have been presented by Dalvi et al. (2025). This layer enhances interoperability by transforming geometric predictions into human-readable contextual descriptions.

Ontology-based Data Integration

The third layer addresses the challenge of interoperability. AI outputs, whether masks or parametric data, must be aligned with construction-specific information standards to be usable. The proposed framework, therefore, incorporates ontology-based structuring of detected damages. The existing Damage Topology Ontology (DOT), proposed in Hamdan et al. (2019), is modified, and the detected damage categories are mapped to standardized taxonomies via ontology-aligned JSON templates and APIs and then linked to structural elements within digital building models, as described by Dalvi et al. (2026).

By structuring semantic outputs into ontologies, damage information becomes machine-readable and interoperable. This step enables alignment with Industry Foundation Classes (IFC)-based models and facilitates integration into a digital twin, a continuously coupled virtual representation of a physical structure that integrates geometry, damage observations, and structured knowledge to support inspection, reasoning, and long-term maintenance. The workflow assumes availability of a building model (e.g., IFC), but it can work with alternative structured data, such as classified point clouds or other spatial representations.

Knowledge Graph Population

The RDF (Resource Description Framework) graph explicitly models multiple relation types, representing information as a network of interconnected nodes and edges. These include spatial relationships (e.g., adjacency, overlap, containment), inspection timestamps, and hierarchical associations between damages and structural elements. Temporal evolution across inspections is represented by connecting damage instances over time, enabling tracking of progression, propagation, or mitigation effects.

In addition, the graph incorporates expert-defined causal relations that link observed damage patterns to potential causes, including environmental exposure, loading conditions, and material degradation mechanisms. Through this relational modeling, inspection data, digital building model information, and domain expertise converge into a unified structure. The resulting knowledge graph supports querying, and cross-project knowledge capturing, enabling analysis of recurring defect patterns, condition evolution trends, and intervention effectiveness.

Conclusion

This approach aims to transform visual damage detection into structured, explainable, and reusable infrastructure knowledge within digital construction workflows. Future work will focus on AI model orchestration, empirical validation of the integrated pipeline, refinement of ontology mappings, exploration of reasoning capabilities within the constructed knowledge graph, and development of a front-end application for visualizing all the detected damages and attributes.

Acknowledgments

The authors gratefully acknowledge the support of the Nemetschek Innovation Foundation as part of the AInspect2Twin project, which enabled our research and its practical application.

References

Dalvi, V. G., Martin, J., and Weilbach-Eyüboglu, T. (2025). Image captioning for building damage using deep learning: Explaining context in overlapping structural defects. In Proceedings of the 36th Forum Bauinformatik 2025, pages 349–357, Aachen, Germany.

Dalvi, V. G., Preidel, C., Fröhlich, D., and Dauberschmidt, C. (2026). A semantic workflow linking image-based concrete damage segmentation to ontology-driven knowledge graphs for digital twin integration and automated structural diagnostics. Manuscript in preparation.

Flotzinger, J., Rösch, P., Benz, C., Ahmad, M., Cankaya, M., Mayer, H., Rodehorst, V., Oswald, N., and Braml, T. (2024). DACL challenge: Semantic segmentation during visual bridge inspections. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, pages 716–725.

Hamdan, A.-H., Bonduel, M., and Scherer, R. (2019). An ontological model for the representation of damage to constructions. In Proceedings of the 7th Linked Data in Architecture and Construction Workshop (LDAC 2019), volume 2389 of CEUR Workshop Proceedings, pages 64–77, Lisbon, Portugal.

Krefft, L., Fröhlich, D., Hoegner, L., and Dauberschmidt, C. (2026). Crackquant: Multimodal dataset for the detection and quantification of cracks in concrete using image and spatial data. Manuscript in preparation.

Krefft, L. and Hoegner, L. (2025). A method for crack detection and quantification in masonry using neural network-based image analysis. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, X-4/W6-2025:113–119.

2026 European Conference on Computing in Construction
Corfu, Greece
July 12-15, 2026

EVALUATING LOCAL SMALL LANGUAGE MODELS FOR ARCHITECTURAL MEETING TRANSCRIPTION SUMMARIZATION AND ACTION EXTRACTION

Sean Bayly[1,2], Sneha Verma[1], Mohamad Kassem[1] and Huizhi Liang[1]
[1]Newcastle University, [2]Ryder Architecture

Abstract

This paper presents a locally hosted solution, for LLM driven summarization and action extraction for project-based meetings, that addresses the privacy, confidentiality, and cost concerns of externally hosted solutions. Developed for architectural project coordination, this system reduces the administrative burden of documenting multidisciplinary design meetings, tracking actions, and maintaining project records across distributed teams. Qwen3.6 27B was used as a judge to evaluate meeting transcript summarization and action extraction by Qwen3.5 4B and 9B SLMs in BF16 and AWQ INT4 formats. The 9B models achieved the best quality-speed tradeoff, producing accurate actions and concise, grounded summaries with lower wall-clock times.

Introduction and Related Research

Coordination between teams in the built environment involves frequent meetings with manual tracking of actions being onerous and prone to error (Ghimire et al., 2024). Many existing summarization workflows rely on external AI providers (e.g., Fireflies.ai and Otter.AI), requiring meeting transcripts and associated project discussions to be transmitted beyond the organisation's primary collaboration platform. In architectural practice, where meetings may contain commercially sensitive, contractual, or personally identifiable information, this can introduce additional privacy, data governance, and compliance concerns (Yan et al., 2024).

Table 1: Comparative Analysis of Meeting Intelligence Systems

Systems	Automatic Capture	Multi-Speaker Structuring	Project-Level Integration	Role-Based Access Control
Otter.ai	Yes	Strong	Limited	Platform-level
Fileflies.ai	Yes	Strong	Limited	Platform-level
Meeting Intelligence System	Yes (Graph Webhooks)	Structured Transcript Chunk	Deep relational embedding	Azure AD hierarchical role

This risk can be mitigated through the local deployment of LLM based models, which limits the egress of transcript data to third-party services during inference. The primary research challenge, therefore, is whether locally hosted SLMs can generate meeting summaries and action items of sufficient quality for use in architectural project coordination. In this paper, we investigated 4B and 9B parameter models from the Qwen3.5 family (Qwen Team, 2026). The impact of AWQ quantization on output quality is also evaluated.

Prototype: Development and Testing

Developed solution utilises three pipelines (Figure 1) to ingest and transform data atomically, from discovery through transcript ingestion and eventual summarisation.

30 transcripts were sampled, from project related

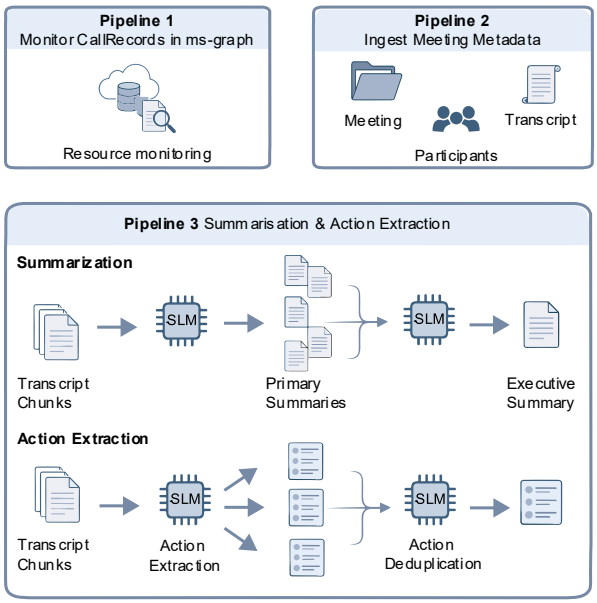


Figure 1: Architectural overview. The first two stages are concerned with data discovery and ingestion, with the third focused on processing transcripts into actions and summaries

meetings in an architectural practice, to form the evaluation dataset. They vary from 2000-30000 tokens and are stratified into three groups of 10 by token length at boundaries 8000 and 15000 tokens. The summarisation process is evaluated via LLM as a Judge (Zheng et al., 2023) in which a Qwen3.6-27B (Qwen Team et al. 2026) is tasked with scoring the summarisations on a scale of 0-5 (Li et al., 2026) on two categories: Faithfulness, and Completeness (Zheng et al., 2023). These metrics correspond to tendency to hallucinate, and inclusion of all available facts, respectively. For action extraction, the judge model produces its own actions and compares them to the actions of the candidate models. Finally, F1 score was calculated to represent agreement between the candidate and the judge model.

Result and Analysis: Action extraction F1 score (Figure 2) across all conditions was consistently above 0.85, indicating that the SLMs and the Judge were mostly in agreement. Though, the quantized models show larger variation in standard error. Upon inspection, this was caused by the AWQ models occasionally failing completely and generating zero actions, upsetting the

mean and reported standard error. In practice, a system can detect and retry under such circumstances. Scores were broadly flat across meeting lengths, validating the small window, high overlap and dedupe approach.

Faithfulness (Figure 2) for immediate summarizations of transcript windows are consistently strong across models with quantization having a minimal impact. Executive summaries exhibit a steep decline, likely due to the high compression and propagation of hallucinations from tier

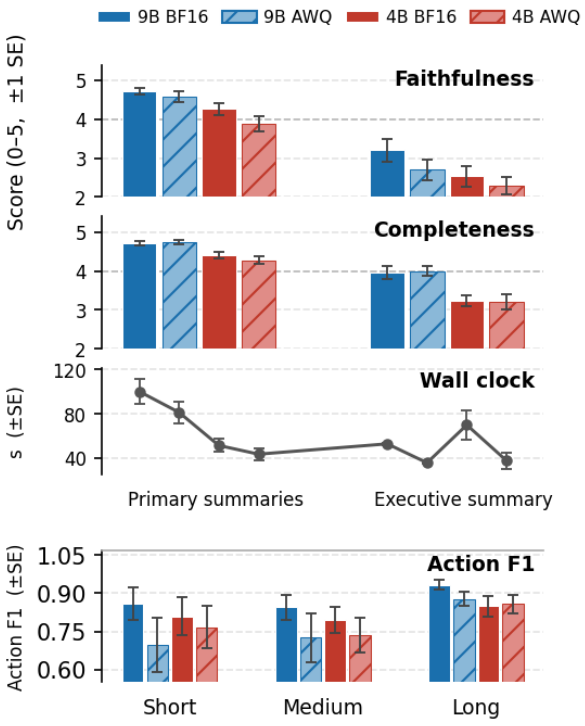


Figure 2: LLM Judge scores on Faithfulness and Completeness, Action F1 for 9B/4B variants of Qwen3.5 in BF16 and AWQ INT4. Wall clock time represents the mean average time taken for completion of a summary

one summarizations.

The 9B parameter models clearly outperform the 4B parameter models in Completeness (Figure 2) indicating that they tend to preserve more of the available information from the transcript.

Wall clock time (Figure 2) decreases steadily with reduction in model size, this is expected as LLM decode is memory bandwidth bound, smaller models are faster. However, for executive summaries, the 4B models break that trend, taking more time to produce summaries. This is because the smaller models are consistently producing longer outputs at that stage. Alongside a decline in Faithfulness, and a steeper decline in Completeness, this indicates that the outputs are not only longer, but less valuable.

Conclusion

This work presented a locally deployed LLM based summarisation system for architectural meetings, offering a privacy-conscious alternative to cloud-based documentation tools. The 9B model in quantization (INT4 AWQ) was identified as the optimal configuration, delivering acceptable summarisation quality with significant decode speedup over the BF16 variant, while producing more concise and consistent outputs than the 4B model. These results demonstrate the feasibility of locally deployed LLMs for supporting privacy-conscious architectural workflows and reducing manual meeting documentation overhead. Although, Meetings organised by external tenants cannot be automatically accessed through MS Graph due to tenant ownership restrictions, and users within meeting rooms cannot be directly identified. Additionally, the asynchronous architecture introduces eventual consistency delays. Future work will explore broader evaluation strategies using both open-source and frontier judge models to reduce model-family bias, alongside extending experiments to next-generation SLM families such as Gemma4.

Database and Source Code Availability: All source code, datasets, and supplementary implementation details related to this work are publicly available at the following repository.

Acknowledgments

Innovate UK (Grant 10113484) for facilitation of a Knowledge Transfer Partnership (KTP)
ReferencesQwen Team. (2026) Qwen3.5: Towards Native Multimodal Agents, Available at: https://qwen.ai/blog?id=qwen3.5

Qwen Team. (2026) Qwen3.6-27B: Flagship-Level Coding in a {27B} Dense Model, Available at: https://qwen.ai/blog?id=qwen3.6-27b

Ghimire, P., Kim, K. & Acharya, M. (2024) Opportunities and Challenges of Generative AI in Construction Industry: Focusing on Adoption of Text-Based Models. Buildings, 14(1), p.220.

Yan, B., Li, K., Xu, M., Dong, Y., Zhang, Y., Ren, Z. & Cheng, X. (2024) On Protecting the Data Privacy of Large Language Models (LLMs): A Survey. arXiv preprint arXiv:2403.05156

Li, W., Zhao, M., Dong, W., Cai, J., Wei, Y. et al. (2026) Grading Scale Impact on LLM-as-a-Judge: Human-LLM Alignment Is Highest on 0-5 Grading Scale. arXiv preprint arXiv:2601.03444

Zheng, L., Chiang, W., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E.P., Zhang, H., Gonzalez, J.E. & Stoica, I. (2023) Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. Advances in Neural Information Processing Systems, 3