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

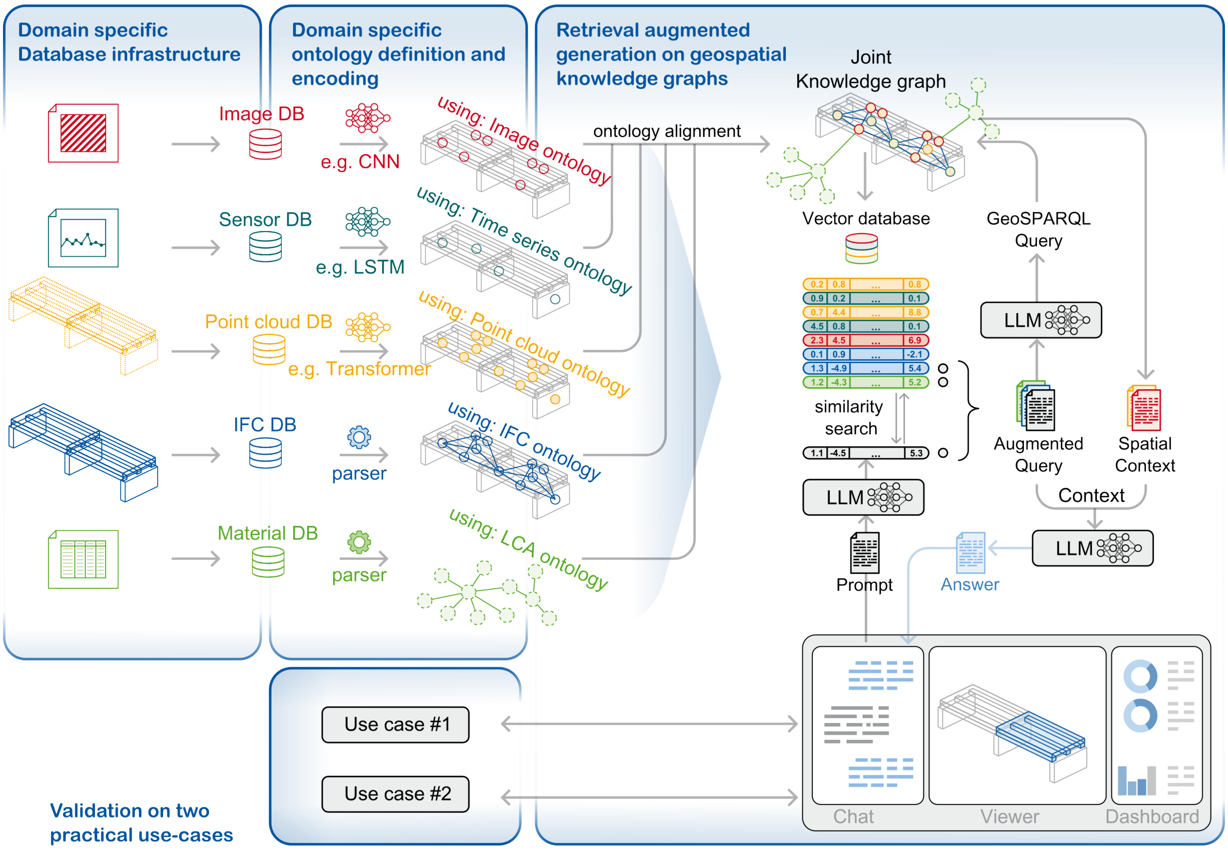

Figure 1: Schematic overview of the CARE Scheme, a hybrid data integration pipeline for construction intelligence.

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/

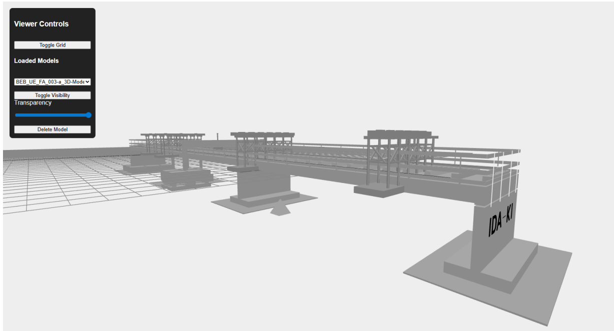

Figure 2: Image of the OpenLAB Bridge in the Three.js based web-viewer

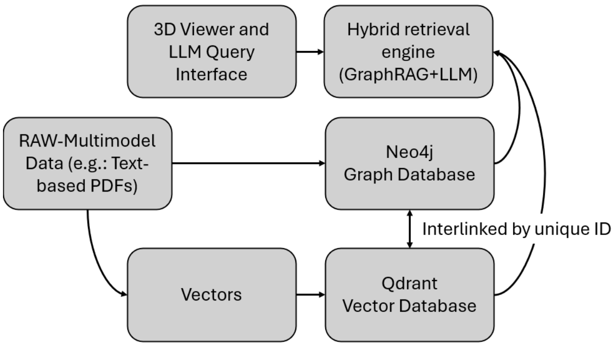

Figure 3: Overview of the current Retrieval and Generation Pipeline

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 retrieval 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.