

## ONTOLOGIES AND LINKED DATA FOR STRUCTURING PUBLISHED BIM ARTICLES

Karim Farghaly

Oxford Brookes University, Oxford, United Kingdom

### Abstract

In the last decade, Building Information Modeling (BIM) has been of broad and great interest in both the academia and industry. Massive amounts of BIM researches have been published in academic journals, books and conference proceedings. However, more than half of the published researches are not cited, and some researches with similar questions and outputs are published. In this research, the Linked Data paradigm is utilized with ontologies to provide a structured database of the different articles of BIM and mapped to different BIM aspects in a semantically meaningful format. The proposed ontologies and Linked Data can accelerate and enhance systematic reviews and metadata analysis for future BIM studies. Also, it can assist young researchers to define the gaps that need to be filled in BIM studies and avoid repetitive researches.

### Introduction

Since the idea of Building Information Modeling (BIM) has been coined in 1978 (Eastman et al, 2011), BIM has been increasingly implemented in the Architecture, Engineering, and Construction (AEC) sectors and recently in the Operation and Maintenance (O&M) sectors (Farghaly et al. 2018). BIM is evolving as an approach driven by technology for generating, managing and exchanging a well-structured facility's data throughout its lifecycle (Succar & Sher, 2014). In the last decade, BIM has been the new international benchmark for better efficiency and collaboration in the AEC and O&M industry (Arayici, 2015). Several public and private sectors have mandated BIM for all their projects. This leads to the publication of public policies and standards, and also academic publications to improve the implementation of BIM in the AEC and O&M sectors.

At present, there is a massive amount of BIM researches in academic journals, conferences, books and master theses. Based on the Web of Science (WOS) database, 1836 journal articles contain the keyword "BIM." Figure 1 represents the number of articles that have been published in five of the main journals in the AEC fields during the last ten years; namely, Automation in Construction, Advanced Engineering Informatics, Journal of Computing in Civil Engineering, Journal of Construction, Engineering and Management, and Journal of Information Technology in Construction. It also shows

how exponentially the number of articles is increasing.

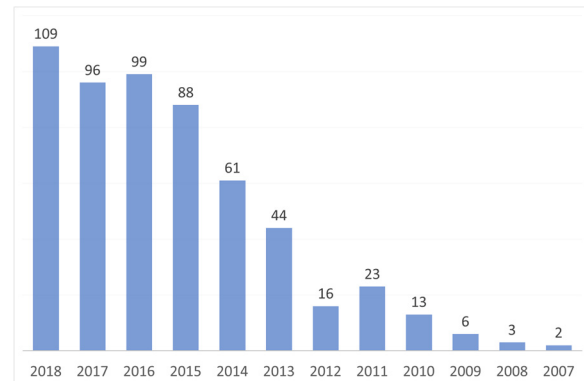


Figure 1: Number of articles published related to BIM in the main five journals in AEC domains.

Due to this massive amount of literature, BIM is reaching maturity in some aspects, and the quality and complexity of recent papers are significantly high (Santos et al, 2017). However, it is hard to establish which BIM aspect has been covered and which still has gaps to be filled. This leads to similar and repeated research questions, objectives and outputs. Moreover, several researchers do not provide much literature on the identified problem and do not evaluate it critically. Instead, they provide general literature and rush to the formulation of research objectives and methods.

Providing a structured database of the different articles of BIM could assist researchers to identify and select the appropriate articles for their research problem and objectives. BIM researches can be classified based on several aspects, such as BIM tri-axial framework (Succar, 2009), project phases, project types, BIM uses, and BIM dimensions (Charef et al, 2018; Succar, 2009). To overcome the barrier of the different classifications, it has been argued that Linked Data (LD) and ontologies could improve interdisciplinary research through linking between different domain-specific knowledge bases (Schmachtenberg et al, 2014). LD aims to define a process to publish and share machine-readable interlinked data on the web, based on a set of design principles. The implementation of LD in the AEC and O&M sectors is still in its infancy (Radulovic et al, 2015). Pauwels et al. (2015) categorized the work of LD in the AEC sector to three main groups based on aims and barriers to overcome; namely, interoperability, linking across domains, and logical inference and proofs. Several publications implemented LD with BIM for better

building operation and maintenance (Curry et al, 2013), efficient management of building energy activities (O'Donnell et al, 2013), sharing construction defect information (Lee et al, 2016), and linking Industry Foundation Classes (IFC) objects with the facilities management work information (Kim et al, 2018).

So far, as per the author's knowledge, there is no available work that has attempted to explore using LD and ontologies to publish structured data on the web related to all BIM articles. Therefore, this research proposes the different ontologies required to create reporting standards for BIM articles. It also recommends a framework of mapping the articles to the proposed ontologies and publishing them in the Web using LD.

### **Linked Data**

Linked Data is the coding of data with semantic meaning for human and machine use (Dutta, 2017). LD works in the same way as relational data in conventional databases; they are stored, processed and queried by computers, not by humans (Bizer et al., 2009). The term Semantic Web was coined before Linked Data by Sir Tim Berners-Lee, "the inventor of the Web" at WWW (World Wide Web) in 1994 and documented in a scientific American article in 2001 (Berners-Lee et al., 2001). The semantic web is an extension of the current web where information (data and documents) is well defined to ensure better cooperation between computers and humans. The purpose of the semantic web is to achieve data universality and data linking with any other data.

However, Bernes-Lee (2006) noticed that some semantic data published on the web are not linked to other outside semantic data. Therefore, Bernes-Lee outlined four principles that need to be adopted to truly obtain LD. The four principles are (1) use URIs (Uniform Resource Identifiers) as names of things, (2) use HTTP URIs so that people can look up those names, (3) when someone looks up a URL, provide useful information using the standards, and (4) including links to other relevant URIs, so they can discover more things. In 2010, Bernes-Lee suggested a five-star deployment schema for Linked Data based on the four Linked Data principles. The five-star deployment schema is accumulative, it starts with one star, and the data earn more stars when proprietary formats are detached, and more links are added.

The web of Semantic Data, unlike the document web, requires standards to ensure a highly interconnected network where the huge amount of heterogeneous data has been given a well-defined meaning. Therefore, the World Wide Web Consortium (W3C) team has developed an ecosystem of standards named Semantic Web Stack, to support LD. The stack consists of several layers, and the first two layers are known as the

standard Internet technologies for providing data on the web. The combination of these two layers with Cryptography, which verifies the reliability of inputs such as the digital signature of the source's origin, forms the structure of the normal web. Resource Description Framework (RDF) layer is a semantic web core representation for data in graphical format. The RDF is a generic and flexible language that permits the presentation and integration of information from different knowledge domains. RDF is a framework based on triples subject-predicate-object that forms a graph of data. Each concept in an RDF graph where it is subject, object or predicate has its own URI (Uniform Resource Identifier). The RDF Schema (RDFS) contains the most basic elements in describing the ontologies built by the RDF triples, such as the description of classes, subclasses, comments, relationships and data types. Ontologies represent knowledge in specific domains and enable semantic interoperability by linking to other external data sources (Guarino, 1998). More rich and complex elements and greater machine interpretability for expressing ontologies and RDF statements are found within Web Ontology Language (OWL). For querying and manipulating RDF data as well as RDFS and OWL ontologies, Simple Protocol and RDF Query Language (SPARQL) is commonly used. It provides SQL-like operations the ability to retrieve knowledge from OWL. RDF(S), OWL and SPARQL provide the foundation to allow working with rules and proofs. Relying on rules and proofs, several applications can be built and could reach a precise level of trust.

Due to LD capabilities, it has enjoyed great popularity in different domains including biology, medical records, digital libraries, cultural heritage, finance, accounting, and social media (Schmachtenberg et al, 2014). There are three main benefits of utilizing LD in the research lifecycle of science; namely, to facilitate the serendipitous discovery of information resources and novel links, enhance the systematic review of the existing body of knowledge, and provide metadata openness and sharing (Alemu et al, 2012; Zaveri & Ertaylan, 2017).

### **Ontologies Development**

The process of the development of an ontology consists of seven main steps as shown in Figure 2 (Noy & McGuinness, 2001). The first step is to define the covered domain and the scope. As mentioned before, this research concentrates on the articles in the BIM domain and present the different ontologies for BIM articles. The second step is to consider reusing existing ontologies. Several classifications have been taken in consideration as ontologies in this research such as BIM fields, and BIM stages (Succar et al, 2007) and n-dimensions developed by the University of Salford (Lee et al, 2005). Also, the RIBA plan of work stages is adapted to develop building lifecycle phases.

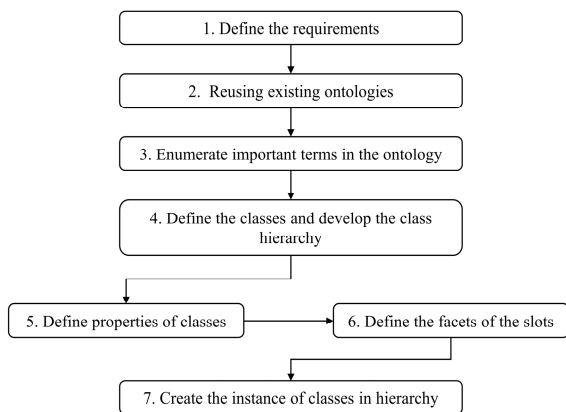


Figure 2: The process of ontology development

The third step is to enumerate important terms in the ontology. In this step, terms are extracted to form a list of concepts (classes, relationships, and slots) from the data schema regardless of any overlap between the concepts they represent. The names of the selected terms have to follow a specific strategy as specified in define resources-naming strategy task. In this stage, all the classes and their related instances are identified. Some of the terms are selected from existing taxonomies or classifications, and the author proposes the others. For example, Succar (2007) classified BIM fields in three interlocking nodes; namely, policy, process, and technology. That classification has been adapted to develop the class of BIM field and its related instances. On the other hand, the author has noticed the engagement of several technologies in the implementation of BIM, such as Geographic Information Systems (GIS), Virtual Reality (VR), Augmented Reality (AR), Linked Data and Big Data. Therefore, it is crucial to develop an ontology for the different technologies that can be integrated with BIM. The name of the class is “BIM with other technologies,” and it contains several instances representing the different technologies. Also, another ontology is identified with its own classes, instances,

and relationships for the articles. It consists of the main information required to classify the articles, such as the source, the year, the author and the country.

The fourth step is to define the classes and develop the class hierarchy. Several approaches can be used for developing a class hierarchy; namely, top-down, bottom-up and combination. Most of the ontologies are developed based on the top-down approach, which starts from an abstraction of a domain and continues to a concrete level. However, it has been argued that the bottom-up approach is more effective as the domain modeling is based on raw and evidential data instead of theoretical conceptualization (Zaveri & Ertaylan, 2017). One of the pros of integrating LD with ontologies is that it allows utilizing the bottom-up approach for the ontology development (Dutta, 2017). In this research, the top-down approach is used for the reusable ontologies and concepts, while the bottom-up approach is selected for the development of the new ontologies. For example, the BIM use and the BIM benefit classes developed starting with identifying the instances. The instances were identified based on the review of 100 of articles and classifying them based on BIM usage and BIM benefits.

The fifth and sixth steps are closely interconnected and are usually done together. The fifth step is to define the properties of classes (slots); while the sixth step is to define the facets of the slots. The values of slots are described in different facets such as; value type, allowed values, cardinality, and other facet features. The value type facet can be described in different value types, such as string, number, Boolean and enumerated. The allowed value facets define the range of slot and the cardinality facets define how many values the slot can have. In this research, the string value type is used for defining most of the slots for the classes’ properties. Finally, the seventh step is to create the instance of classes in the hierarchy. This step consists of three tasks respectively, which are – choosing the class, creating the instance of that class

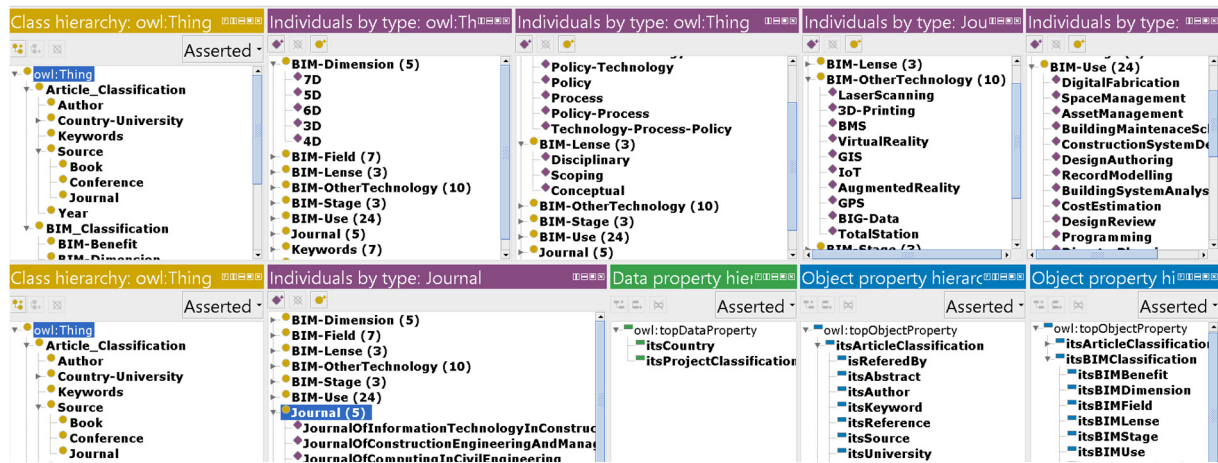


Figure 3: Screenshot for the developed classes and instances using protégé.

and finally filling in the slots values. Figure 3 represents a screenshot for the developed ontologies using Protégé 5.5.0.

In this research, instances are used to represent the different attached individuals to a specific class such as BIM technology attached to the BIM field class, while individuals are utilized for the BIM articles. After the development of the different ontologies, a logical base approach is used to evaluate the ontologies syntactically. Syntactic evaluation aims to ensure that all defined concepts are viable from a technical viewpoint (Tomašević et al, 2015). Several SPARQL queries are executed to observe the credibility of obtained results and Hermit reasoner was used to validate the consistency with the used ontologies. Hermit is a reasoner for ontologies written using the Web Ontology Language (OWL), and it is a preinstalled Protégé plug-in.

### Linked Data Development

A main step in developing and publishing data in the web is to provide crosslinking between the instances of an ontological source and the data ‘out there’ based on the background knowledge. LD mapping relies on setting RDF links between URI aliases in order to be able to track the different information referring to the same article. This stage aims to provide visible indicators that have not already been harvested such as the interconnections, incoming and outgoing links between the different BIM classifications. Figure 4 illustrates the link between the different ontological sources and a specific BIM article. The different relationships are color-coded and depicted in the bottom left of Figure 4.

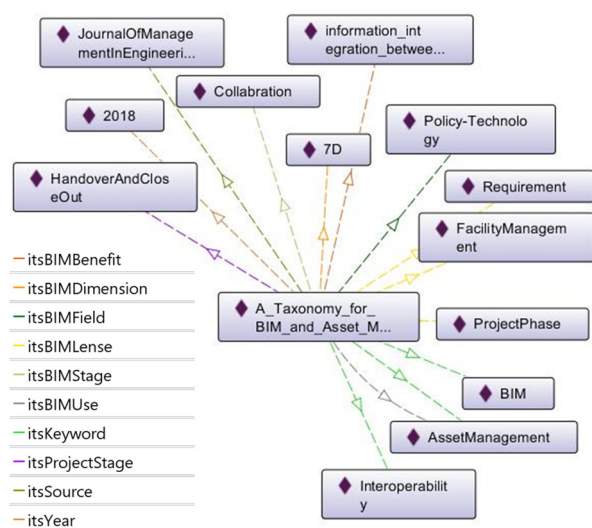


Figure 4: An example of mapping all the BIM aspects to a journal article using Protégé OntoGraf plug-in

The selected journal paper, “A Taxonomy for BIM and Asset Management Semantic Interoperability” has

been mapped to predefined ontologies. Ten main relationships were identified using object properties to map between the ontological instances and the selected article. For example, the relationship “*itsBIMLenses*” links the instances related to the ontology of BIM lenses to the article. Since the BIM lenses class consists of three main subclasses – disciplinary, scoping and conceptual, three different instances were linked using the *itsBIMLenses*. Another relationship was developed using data properties such as *itsAbstract* and *itsIssueNo* (Data properties are not shown in Figure 4). Data properties describe relationships between individual and data values. For example, *itsIssueNO* relationship is assigned to the subclass *Journal* to identify the issue number where the article has been published.

The developed ontologies and the defined resource naming strategy can be utilized in the transformation process of the BIM articles from the different sources into RDF format. Figure 5 illustrates the proposed framework for data transformation of the different BIM articles. The first step is selecting and accessing the data sources. In the ongoing research, the articles were retrieved from the WoS core collection using two keywords: BIM and Building Information Modeling. The University subscriptions have provided the accessibility for the different journals sources. 1836 research papers were selected, accessed and analyzed at the end of this stage.





Data Sources	 WEB OF SCIENCE	1- Select Data Sources. 2- Access Data Sources.
Data Transformation		1- Select RDF Serialization. 2- Select tool to transfer. 3- Use selected tool. 4- Evaluate RDF datasets.
Data Integration		1- Mapping the different BIM ontologies with the articles.
Data Mining		1- Systematic reviews. 2- Metadata Analysis. 3- Novel Links.

Figure 5: Proposed framework for the flow of data

The second step is to transform the datasets to RDF format. The data are exported from the WoS website in HTML (HyperText Markup Language) format and then converted to xls format. The exported data for the BIM articles were arranged in seven main columns; which are, research title, abstract, source, keywords, year, authors and issue no. OpenRefine with the RDF extension has been utilized to transform the xls data into RDF. To achieve the transformation, a mapping between the data and the ontology has to be defined

taking into consideration the defined resource-naming strategy. This has been achieved in several tasks; firstly, the preliminary transformations to the data made in order to correct errors. Secondly, mappings between the columns and rows in the table and the ontology and specifying the pattern for naming instances according to the resource-naming strategy (Radulovic et al, 2015). Consequently, the RDF syntax is chosen and the datasets are generated and evaluated semantically and syntactically. The Turtle serialization was selected as the research RDF serialization. More mapping is required between the datasets and the ontologies related to the BIM aspects. The author has manually mapped the different articles to their related instances in the different BIM ontologies. However, he believes that the authors of the selected articles can classify their work more accurately. Therefore, future work includes approaching the authors to classify their researches based on the developed ontologies.

Once all the data are transformed and integrated with the ontologies, the discovery of links between the data sets and the retrieving of required data are now possible using SPARQL queries. Finally, the proposed framework can provide outputs which are characterized by maximally findable, accessible, interoperable and reusable (FAIR) data (Zaveri & Ertaylan, 2017).

## Conclusion

In both the academia and industry sectors, BIM has been the AEC evolving topic. Massive amounts of BIM researches have been published in academic journals, books and conference proceedings. Meanwhile, more students are enrolled in the BIM master's degree and required to conduct a research for their dissertations. The massive amounts of articles available mislead students to find related literature for their research questions, which leads to repetitive outputs and less contribution in the field of BIM.

Linked Data (LD) and ontologies can enhance interdisciplinary research by enriching knowledge through linking among multiple domain-specific knowledge bases. In this research, ontologies are utilized to organize and describe the BIM articles based on the different BIM concepts such as the tri-axial BIM framework, BIM dimensions and BIM uses, BIM benefits and technologies adapted with BIM. While LD is implemented to expose data in different integrated ways and enhance the Web through the addition of structured data. This paper is a part of an ongoing research; more outputs related to the datasets transformation will be presented in the future. More ontologies and relationships are going to be added to future works, and also a systematic review for the implementation of BIM in the Asset Management (AM) sector and meta-analysis of the articles based on the building phase where the implementation of BIM occurred.

## References

- Alemu, G., Stevens, B., Ross, P. & Chandler, J. (2012) Linked Data for libraries: Benefits of a conceptual shift from library-specific record structures to RDF-based data models. *New Library World*, 113(11/12), 549-570.
- Arayici, Y. (2015) *Building Information Modelling*, 1st edition. London, UK: Bookboon Publisher.
- Charef, R., Alaka, H. & 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.
- Curry, E., O'Donnell, J., Corry, E., Hasan, S., Keane, M. & 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.
- Dutta, B. (2017) Examining the interrelatedness between ontologies and linked data. *Library Hi Tech*, 35(2), 312-331.
- Eastman, C., Teicholz, P., Sacks, R. & Liston, K. (2011) *BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors* John Wiley & Sons.
- Guarino, N. (1998) Formal ontology and information systems, *Proceedings of FOIS*.
- Kagioglou, M. (2003) *Developing a vision of nD-enabled construction* Construct IT Centre of Excellence, University of Salford.
- Kim, K., Kim, H., Kim, W., Kim, C., Kim, J. & Yu, J. (2018) Integration of ifc objects and facility management work information using Semantic Web. *Automation in Construction*, 87, 173-187.
- Lee, A., Wu, S., Marshall-Ponting, A., Aouad, G., Cooper, R., Tah, J., Abbott, C. & Barrett, P. (2005) nD modelling road map: A vision for nD-enabled construction.
- Lee, Y.-C., Eastman, C. M. & Solihin, W. (2016) An ontology-based approach for developing data exchange requirements and model views of building information modeling. *Advanced Engineering Informatics*, 30(3), 354-367.
- Noy, N. F. & McGuinness, D. L. (2001) *Ontology development 101: A guide to creating your first ontology*. Stanford knowledge systems laboratory

- technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA.
- O'Donnell, J., Corry, E., Hasan, S., Keane, M. & Curry, E. (2013) Building performance optimization using cross-domain scenario modeling, linked data, and complex event processing. *Building and Environment*, 62, 102-111.
- Pauwels, P., Törmä, S., Beetz, J., Weise, M. & Liebich, T. (2015) Linked Data in Architecture and Construction. *Special Issue Automation in Construction, erscheint Juni*.
- Radulovic, F., Poveda-Villalón, M., Vila-Suero, D., Rodríguez-Doncel, V., García-Castro, R. & Gómez-Pérez, A. (2015) Guidelines for Linked Data generation and publication: An example in building energy consumption. *Automation in Construction*, 57, 178-187.
- Santos, R., Costa, A. A. & Grilo, A. (2017) Bibliometric analysis and review of Building Information Modelling literature published between 2005 and 2015. *Automation in Construction*, 80, 118-136.
- Schmachtenberg, M., Bizer, C. & Paulheim, H. (2014) Adoption of the linked data best practices in different topical domains, *International Semantic Web Conference*. Springer.
- Succar, B. (2009) Building information modelling framework: A research and delivery foundation for industry stakeholders. *Automation in construction*, 18(3), 357-375.
- Succar, B. & Sher, W. (2014) A competency knowledge-base for BIM learning, *Australasian Journal of Construction Economics and Building-Conference Series*.
- Succar, B., Sher, W., Aranda-Mena, G. & Williams, T. (2007) A proposed framework to investigate building information modelling through knowledge elicitation and visual models. *Melbourne, Australia, Australasian Universities Building Education (AUBEA2007)*.
- Tomašević, N. M., Batić, M. Č., Blanes, L. M., Keane, M. M. & Vraneš, S. (2015) Ontology-based facility data model for energy management. *Advanced Engineering Informatics*, 29(4), 971-984.
- Zaveri, A. & Ertaylan, G. (2017) Linked Data for Life Sciences. *Algorithms*, 10(4), 126.