Enabling automation of BIM-based cost estimation by semantic web technology

David Fürstenberg, david.furstenberg@ntnu.no
Norwegian University of Science and Technology (NTNU), Trondheim, Norway and COWI, Bergen, Norway

Lars Wikström, lars.wikstrom@triona.se
Triona, Falun, Sweden

Eilif Hjelseth, eilif.hjelseth@ntnu.no
Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Ola Lædre, ola.ladre@ntnu.no
Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Abstract
While research on semantic web technology within the Architecture, Engineering and Construction (AEC) industry has significantly increased, there are still few practical and multidisciplinary results. Picking low hanging fruits might be one possible solution to make AEC practitioners interested in semantic web technology. Cost estimation has been a manual and repetitive task prone to human error. We investigate how semantic web technology can support BIM-based automated cost estimation and what challenges that can appear. Standardized specifications of work for Norwegian road projects (with process codes and process descriptions) were linked to an IFC (Industry Foundation Classes) property set by semantic web technology. Future projects that use the standardized specifications of work can adopt the established setup and thereby automate cost estimation with minimal efforts. This practical approach will accelerate the application of semantic web technology in the AEC industry to automate services.

Keywords: cost estimation, automation, semantic web technology, ontology, interoperability.

1 Introduction

Cost estimation has traditionally been a manual and repetitive task prone to human error (Sacks et al 2018). Digital means of working introduced to the AEC (Architecture, Engineering and Construction) industry have enabled automation of cost estimation. In the literature especially two means are stressed, Building Information Modeling (BIM) and semantic web technology.

Elghaish et al (2020), Sacks et al (2018), Wu et al (2014) and Olatunji et al (2010) reported on the possibilities of BIM for automating cost estimation. A recent study (Fürstenberg et al in print) showed how BIM-based cost estimation can be automated in a real-life road project using commercial software. Staub-French et al (2021), Liu et al (2016), Niknam & Karshenas (2015) and Lee et al (2014) investigated the use of semantic web technology for the automation of BIM-based cost estimation.

While research on semantic web technology within the AEC industry has significantly increased (Zhong et al 2019), there are still few practical and multidisciplinary results available (Im et al 2021). Picking low hanging fruits like combining semantic web technology with well-known frameworks might be one possible solution to make AEC practitioners interested in semantic web technology.

Therefore, this paper reports from a novel testing of automated BIM-based cost estimation in a Norwegian road project by answering the following research questions:

RQ1: How can semantic web technology support automated cost estimation?
RQ2: What are the challenges when using semantic web technology?

This study covers some of the many steps needed for cost estimation. It investigates standard bridge-related specifications of work mandatory for projects executed by The Norwegian Public Roads Administration (NPRA). Bridge-related specifications of work were chosen because they follow a specific Norwegian bridge classification system, V440 (Norwegian Public Roads Administration 2009). Mapping specific specifications of work to specific instances of the bridge classification system (for example columns) in a model is complicated, because one object often requires several work processes. There is a 1:n relationship between an object and work processes. In the case of columns specifications of both formwork, concreting, rebar and protecting and hardening measures are necessary for a complete cost estimate of the columns.

2 Theoretical background

2.1 Digital transformation of Norwegian road projects

A common strategy for digital transformation of Norwegian road projects, is to transform traditional drawing-based processes into model-based processes (Fürstenberg & Lædre 2019). In short, drawings are replaced by models. While this enables automation of recurring processes, for example cost estimation, it also implies a change in the way information is provided and processed. A common way to provide all the information traditionally conveyed in drawings is to enrich models with properties. However, not all information should be presented as properties to prevent information overflow (Hjelseth 2011). Only information necessary to map specific specifications of work to specific model instances needs to be provided as properties. All other information should rather be linked from external systems. The links can either point to documents saved on proprietary systems on the web like Google Drive, Microsoft SharePoint or Bentley ProjectWise or they can point to data saved on the semantic web.

2.2 Industry Foundation Classes (IFC)

The Industry Foundation Classes (IFC) are the most prominent vendor-neutral data schema in the AEC industry. With IFC both geometric and semantic information can be exchanged between different stakeholders and software. The data schema is often serialized as a step file (.ifc), but it is also available as an ontology (ifcOWL) presented in Pauwels & Terkaj (2016). The data schema consists of several hundred entities describing both products (e.g. columns or windows) and processes (e.g work schedules or cost items). In the context of cost estimation, the entity IfcCostItem is especially relevant. It is a non-geometric entity consisting of a cost value with or without a quantity (buildingSMART 2021). A cost value can represent either a unit cost or a total cost. IfcCostItems can be classified and they can be nested to create cost assemblies.

2.3 Classification systems, taxonomies, and ontologies

Globally, classification systems in AEC projects are used to classify either products or processes. Examples for product classification systems are Uniclass, Omniclass, CoClass or NPRA's bridge classification system V440. Product classification systems are intended to classify objects and do not contain units or specifications. Product classification systems are well suited for object-based systems like BIM. Whereas process classification systems are intended to classify work processes and do contain units and specifications. Examples of process classification systems are the Norwegian "process code", the Swedish AMA or the German DIN 276. Like mentioned by Fürstenberg et al (in print), process classification systems are often adapted to drawing-based processes and therefore not well suited to classify model objects.

When describing classification systems in the context of automation, the existence of an underlying logic is important to consider (El-Diraby et al 2005). The classes can either be based on generally accepted characteristics, e.g. what things are or what they do, or they can be based on numerical values. In the first case, the meaning of the classes builds up a logical hierarchy of the whole system in the form of a taxonomy. It is self-evident what the classes contain even if the user is not completely familiar with the content. In the second case, the hierarchy of the classes is only based on numerical values, without any underlying logic. The process code classification system falls into that category. While machines can read data independent of an underlying logic, they can only interpret the data if an underlying logic exists (European Committee for Standardization (CEN) 2020).

If domain knowledge is structured according to a taxonomy with defined relationships between the classes, ontologies are created (Beetz 2018). According to Chandrasekaran & Josephson (1997), the benefits of ontologies include clarifying the structure of knowledge and enabling knowledge sharing. Semantic web technology is well-suited to link knowledge stored in different domains (Pauwels et al 2017). During the last years, ontologies have gained importance in the AEC industry (Zhong et al 2019). This trend is also notable in Norway, where the NPRA recently created ontologies for two of their bridge-related handbooks (V440 and V441).

2.4 The Norwegian process code (R761/R762)

The main project delivery method for Norwegian road projects is still design-bid-build and the project costs are estimated based on standard specifications of work necessary to finish the project. The standard specifications of work are defined in the General Specifications which build up a process classification system known as "process code". The process code is a hierarchical system based on numerical values and consists of 9 main processes (Figure 1), divided into three parts. One part is main process 9 which is only used internally in the NPRA. A second part contains the main processes 1-7 which describe all processes necessary in a road project except bridge-related processes. Main processes 1-7 are published in handbook R761 (Norwegian Public Roads Administration 2018a). The third part is main process 8 which contains only bridge-related processes and is published in the handbook R762 (Norwegian Public Roads Administration 2018b). Unique to all processes belonging to main process 8 is that they need to be assembled according to the NPRA's bridge classification system V440. Both handbooks are available in PDF and XML format. A specification text in these handbooks is called a process. A process consists of a numeric code, a title, a specific unit and in most cases also detailed requirements.

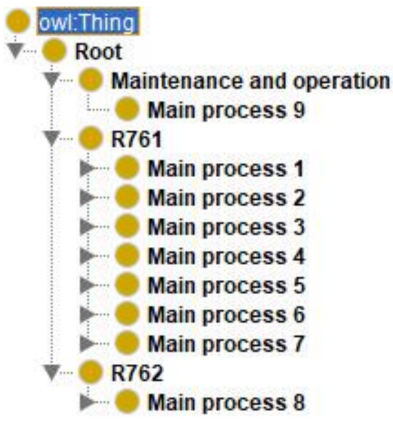

Figure 1. Structure of the Norwegian process code (R761/R762), a standard specification of work for roads

3 Method

Design Science Research "creates and evaluates IT artifacts intended to solve identified organizational problems" (Hevner et al 2004). A recent study (Fürstenberg 2021) confirmed that Design Science Research is often used for research on information management in AEC projects. The research project reported on here developed and tested an artifact in real-world conditions. The artifact was an ontology serialized in Turtle format (ttl) and published on the web through the standard http protocol.

The development and testing followed the six steps in Design Science Research described by Peffers et al (2007):

1. problem identification and motivation,
2. definition of the objectives for a solution,
3. design and development,
4. demonstration,
5. evaluation, and
6. communication.

In 2020 and 2021, the second author led two joint workshops for the NPRA and buildingSMART called "openLAB: Interoperate" (Wikström 2020). The goal of these workshops was to create machine-readable versions of two of NPRA's handbooks using semantic web technology. The NPRA intended to enable automation in both design, construction, and maintenance of their road projects. After 1. Problem identification and motivation and 2. Definition of the objectives for a solution in these workshops, the authors carried out 3. Design and development, 4 Demonstration (in a road project, i.e. real-world conditions), 5. Evaluation and finally 6. Communication.

4 Findings

Data that are not machine-readable (the process codes), and stored in different domains (the handbooks R761, R762 and V440), hinder the automation of cost estimation. Therefore, we suggest using semantic web technology to cope with this issue. We will now describe how semantic web technology supports automation of cost estimation and what the challenges are.

4.1 Automated cost estimation supported by semantic web technology

The process code classification system is already serialized in an open source format, namely Extended Markup Language (XML), providing structure. The XML file was converted into a spreadsheet and complemented with additional data necessary for defining an ontology. Since there was no XML Schema Definition (XSD) containing the relationships between the classes we had to define them in the spreadsheet. No additional means of data modeling were necessary. We used a setup that enabled an automated conversion by a script. We used the process title without spaces, e.g. *Leveling_with_concrete_on_soil* as the Uniform Resource Identifier (URI). Since there are no official English versions of the handbooks available, we used the Norwegian titles but removed all Norwegian letters (å, æ, ø).

We used Protégé (2021) and TopBraid Composer (Top Quadrant 2021) to create the ontology based on Resource Description Framework (RDF), Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL). The hierarchical structure of the process code was re-created by using *"rdfs:subClassOf"*. The Norwegian process titles and the requirement titles were described as *"rdfs:label"*. The text of the requirements, the unit and the code were described as annotations. The ontology was published on the web through the standard web http protocol. Unfortunately, we could not test the ontology in a triple store. We only published it locally. GraphDB (Ontotext 2021) was used to provide access to the ontology as a knowledge graph through the standard Simple Protocol and RDF Query Language (SPARQL) (World Wide Web Consortium (W3C) 2013).

Due to the limited scope of this paper we focused on bridge-related processes (main process 8). To be more precise we focused on process *84.4 Concreting* and created an ontology with all attributes (numeric code, title, unit and requirements) for the subordinate 43 processes. For the other processes only a lightweight ontology including the numeric code, title, and unit was created. Therefore, the depth of the developed prototype varies from three to seven levels.

Like mentioned earlier, semantic web technology is well-suited to link knowledge stored in different domains (Pauwels et al 2017). Therefore, we tested to link knowledge from our developed ontology to two other ontologies. We chose ifcOWL, presented in Pauwels & Terkaj (2016), because it represents one of the most prominent data schema for the interoperability in the AEC industry globally. We also chose a Norwegian ontology developed by the NPRA, namely the V440 ontology, presented in Wikström (2020). The V440 ontology was chosen because it represents a bridge classification which is mandatory to use in cost estimation of Norwegian road projects.

The three ontologies intersect at the entity level, illustrated in Figure 2. IFC entities can among others hold classifications and both nested and unnested cost items. In our investigated example, the IFC entity "column" is classified by the V440 code "C23" and has a nested IfcCostItem. This nested cost item consists of sub cost items which are classified according to the process code classification system. Thereby, an automatic data re-classification was provided, based on semantic reasoning. Such an ontology alignment enables means for linking different "classification islands" together, increasing the potential use and value of the information in various applications.

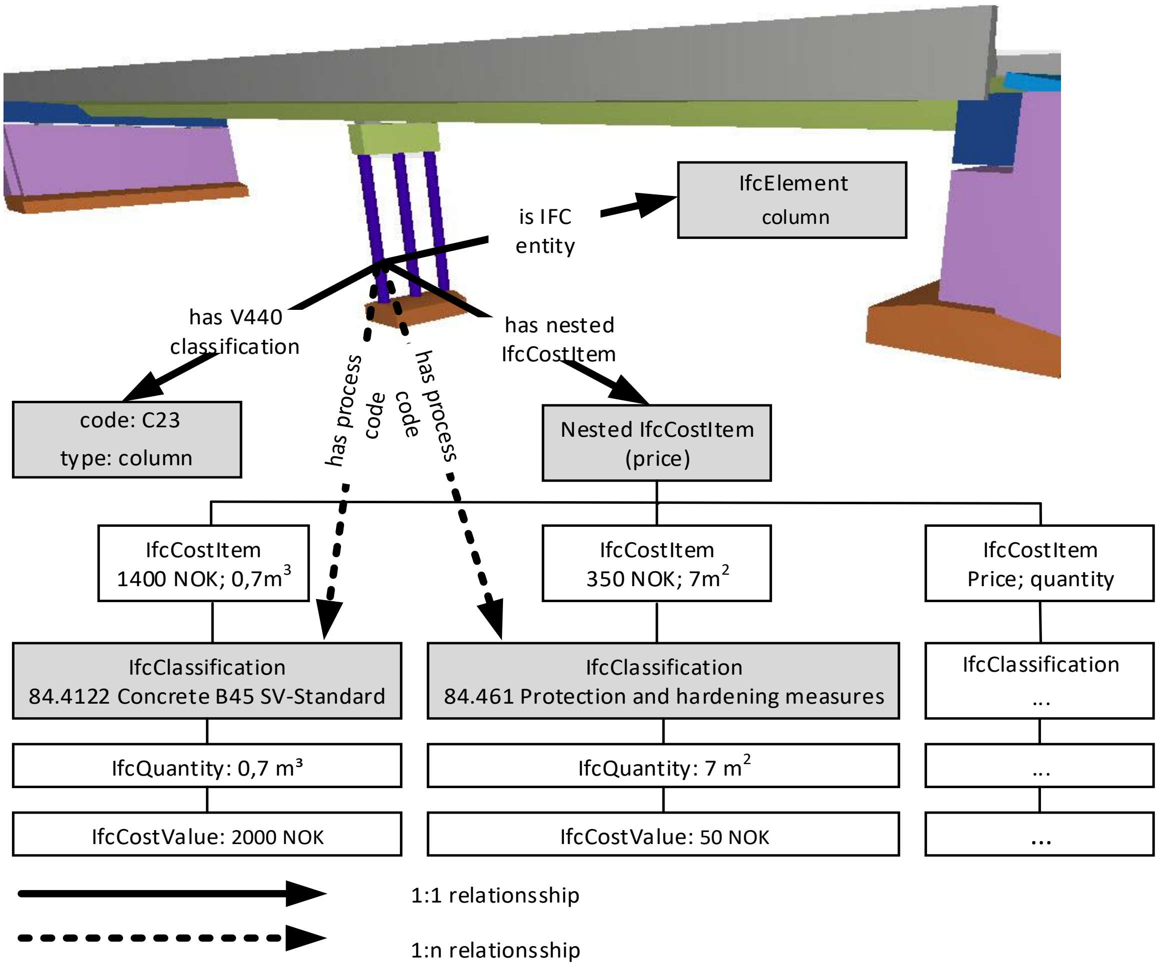

Figure 2. Linking process code to ifcOWL and V440

Finally, we tested our developed ontology under real-life conditions. The ontology was implemented into an IFC step file of a bridge, visualized in Figure 2. The IFC file was earlier produced for a Norwegian road project and made available to all participants of the openLAB: Interoperate workshop. Since IfcCostItems are not visible in the user interface of most commercially available IFC viewers, we created a custom property set and attached it to the IfcColumn entity. Figure 3 shows an example of the custom property set. Only the codes (84.4122 and 84.461), title ("Concrete B45 SV-Standard" and "Protection and hardening measures for formwork surfaces") and its URI are included as properties. Thereby, the number of properties within the model was reduced while still providing all information from the handbook R762. Code and title/label were presented in cleartext and intended for humans. Code, title/label and all other information from R762 (namely unit and requirements) were provided by the URI, both for humans and machines. For the convenience of the readers the Norwegian titles were translated by the authors.

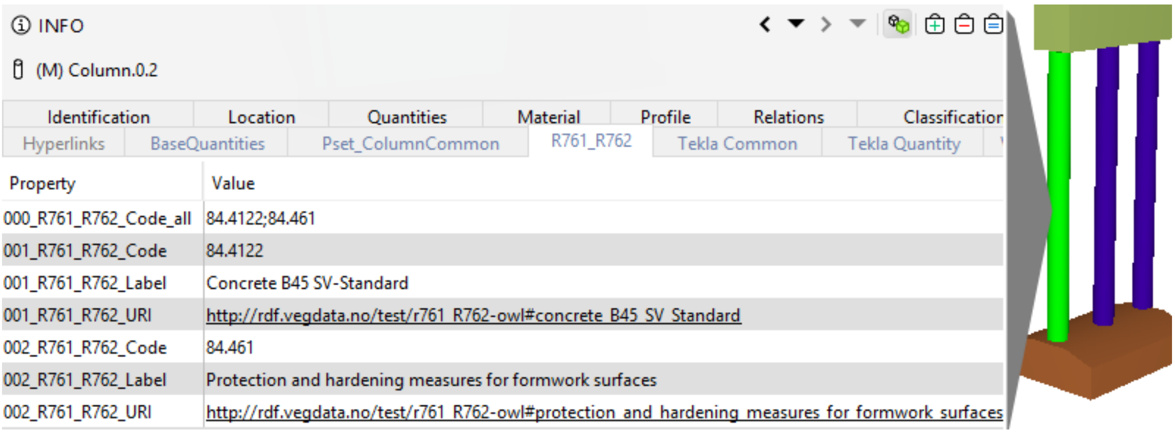

Figure 3. Custom property set with code, label, and URI of the developed ontology

4.2 Challenges when using semantic web technology

We have identified two challenges when using semantic web technology for automating cost estimation.

Firstly, the created ontology does not have an underlying logic in the form of a taxonomy. It is converted from a number-based classification system. While this enabled semantic interoperability with little effort, the generalization of the ontology is somewhat limited. The setup of the process code is well-known to Norwegian domain experts because of their work experience. They know the system behind the codes, e.g. codes starting with 2 involve earthworks or codes starting with 8 are bridge-related. Machines can be used to read and interpret the process codes but only in the given context, namely Norwegian road construction. The process code describes what things are, what things are made of and how things are measured. Therefore, several ontologies are necessary to make the process code machine-interpretable also outside of the specific Norwegian context. The necessary ontologies are, 1) an ontology for *what things are* (e.g. constructions, roads, pipes), 2) an ontology for *what things are made of* (e.g. concrete, asphalt, steel) and 3) an ontology for *how things are measured* (e.g. net area, net volume, net length). Only then, machines can interpret the data autonomously.

Secondly, there is a 1:n relationship between model entities and work processes. Therefore, it is not possible to directly re-classify all entities. While the V440 ontology is only product-based, and the process code ontology is only process-based, IfcOWL is both product- and process-based. While a column is always only an "IfcColumn" in ifcOWL or always "C23" in V440, it could require one or several work processes to be finished (e.g. concreting, rebar, formwork). Therefore, it is only possible to define that a column has "at least" one process code from the subclasses *84 Concrete* or *85 Steel*. However, this challenge could be solved by creating a nested IfcCostItem. This nested IfcCostItem shown in Figure 2 consists of several sub cost items that have different process codes and possibly different quantity units (e.g. m^2 or m^3) but the same monetary unit (NOK). Therefore, the total cost can be summed up but not the total quantity.

5 Discussion

This study is based on a multi-disciplinary and pragmatic approach which has contributed to automation and interoperability. It was hard to find similar types of studies with this combination applicable to AEC projects.

We created an "ontology light" with classes converted from a hierarchical, number-based classification system. The meaning of the classes is well-known and logic in the context of road construction in Norway but there is no underlying logic in the form of a taxonomy. This hinders full interoperability to other, especially international, domain knowledge. However, with a pragmatic approach based on a proxy (nested IfcCostItems) we could link the created classes to both a national (V440) and an international (ifcOWL) ontology.

To test our developed ontology in a real-life project we linked it to an IFC property set. Thereby only necessary information (process code and title) were inserted as properties. All other information is linked through an URI. While it is easily possible to use proprietary web-based solutions like Google Drive, Microsoft SharePoint or Bentley ProjectWise we used semantic web technology. This provided an open, standardized, and machine-readable solution which can be further expanded to other uses cases than cost estimation.

While the actual work on semantic web technology done in this study may seem little, it shows that it does not require much effort to create automation and interoperability, leading up to a more productive AEC industry. This practical bottom-up approach is different to the predominant top-down approach when applying semantic web technology in other fields. While a top-down approach enables more robust solutions, it may hinder practical implementations. Semantic web technology may seem overwhelming to practitioners of the AEC industry. Practical solutions that combine semantic web technology with well-known frameworks and well-known technologies are necessary to motivate AEC practitioners to start using semantic web technology. Additionally, public bottom-up initiatives like openLAB: Interoperate are a good arena to engage AEC practitioners and disseminate lessons learned.

6 Conclusion

This study has presented how semantic web technology can be used to develop automated solutions. BIM-based cost estimation was used as case in this study, but it is important to be aware that this approach can be implemented for multiple other purposes.

We used Design Science Research and followed the six steps defined by Peffers et al (2007) in this paper. Step 1 “problem identification and motivation” and step 2 “definition of the objectives for a solution” were presented in the introduction chapter. Step 3 “design and development” and step 4 “demonstration” were presented in section 4.1 answering research question 1. Step 5 “evaluation” was presented in section 4.2 answering research question 2. This paper is the final step 6 “communication” to the academic community.

When it comes to how semantic web technology can support automated cost estimation, we demonstrated how standardized specifications of work for Norwegian road projects (with process codes and process descriptions) were linked to an IFC property set by semantic web technology. By using the ifcOWL ontology as basis, we mapped both 1:1 and 1:n relationships between different ontologies. As a result, only necessary information from the specifications of work was presented in a property set readable by both humans and machines. Thereby, semantic interoperability between a product-based system (V440) and a process-based standard specification of work for Norwegian road projects (R761/R762) was established. Future projects that use the standardized specifications of work can adopt the established setup and thereby automate cost estimation with minimal efforts.

There are two challenges when automating BIM-based cost estimation using semantic web technology. First, when creating ontologies from number-based classification systems data are only machine-interpretable within the context of the classification. Machines require an underlying taxonomy to really interpret the data. Only then full semantic interoperability to other domains is established. Second, there can be challenges when establishing 1:n relationships between product-based and process-based ontologies.

While introducing new technology to the AEC industry might seem like an overwhelming task, picking low hanging fruits might be one possible solution to make AEC practitioners interested in semantic web technology. If standardized specifications of work for road projects (with process codes and process descriptions) are linked to IFC properties by semantic web technology, automation of cost estimation can be achieved with minimal efforts.

Recommendations for future work consists of two phases. First, the developed ontology needs to be published in a triple store and tested by software developers through Application Programming Interfaces (API). Second, a full taxonomy-based ontology needs to be developed.

Acknowledgements

The authors would like to thank the Norwegian Public Roads Administration (NPRA) for their openLAB: Interoperate initiative. This research was funded by The Norwegian Research Council, grant number 290676 and COWI AS.

References

Beetz, J. (2018), "Structured Vocabularies in Construction: Classifications, Taxonomies and Ontologies", in Borrmann, A., König, M., Koch, C. and Beetz, J. (Eds.), *Building Information Modeling: Technology Foundations and Industry Practice*, Springer International Publishing, Cham, pp. 155-165.

buildingSMART. (2021), "IfcCostItem", available at: https://standards.buildingsmart.org/IFC/DEV/IFC4_3/RC2/HTML/ (accessed 2021-03-26).

Chandrasekaran, B. & Josephson, J.R. (1997). The ontology of tasks and methods. in *AAAI 97 Spring Symposium on Ontological Engineering*, Stanford University, CA California, USA.

El-Diraby, T.A., Lima, C. & Feis, B. (2005). Domain Taxonomy for Construction Concepts: Toward a Formal Ontology for Construction Knowledge. *Journal of Computing in Civil Engineering.* 19 (4). pp. 394-406.

Elghaish, F., Abrishami, S., Hosseini, M.R. & Abu-Samra, S. (2020). Revolutionising cost structure for integrated project delivery: a BIM-based solution. *Engineering, Construction and Architectural Management.*

European Committee for Standardization (CEN). (2020), "Building Information Modelling (BIM) – Semantic Modelling and Linking Standard (SMLS)", in.

Fürstenberg, D. (2021). Information Management in AEC Projects: A Study of Applied Research Approaches. Cham, Springer International Publishing, pp. 272-284.

Fürstenberg, D., Gulichsen, T., Lædre, O. & Hjelseth, E. (in print). BIM-based cost estimation in a road project – proof of concept and practice. paper presented at *13th European Conference on Product & Process Modeling (ECPPM)*.

Fürstenberg, D. & Lædre, O. (2019). Application of BIM Design Manuals: A Case Study. in *Proc. 27th Annual Conference of the International Group for Lean Construction (IGLC)*, Dublin, Ireland, 2019/07/03, Dublin, Ireland, pp. 145-156.

Hevner, A.R., March, S.T., Park, J. & Ram, S. (2004). Design Science in Information Systems Research. *MIS quarterly.* 28 (1). pp. 75-105.

Hjelseth, E. (2011). Modular BIM Guidelines. in *Proceedings of the 28th International Conference of CIB W78*, Sophia Antipolis, France, 26-28 October.

Im, H., Ha, M., Kim, D. & Choi, J. (2021). Development of an Ontological Cost Estimating Knowledge Framework for EPC Projects. *Ksce Journal of Civil Engineering.* 25 (5). pp. 1578-1591.

Lee, S.-K., Kim, K.-R. & Yu, J.-H. (2014). BIM and ontology-based approach for building cost estimation. *Automation in Construction.* 41). pp. 96-105.

Liu, H., Lu, M. & 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.

Niknam, M. & Karshenas, S. (2015). Integrating distributed sources of information for construction cost estimating using Semantic Web and Semantic Web Service technologies. *Automation in Construction.* 57). pp. 222-238.

Norwegian Public Roads Administration. (2009). *V440, Bruregistrering. Håndbøker i Statens vegvesen,* Directorate of Public Roads and Transport Department. Oslo.

Norwegian Public Roads Administration. (2018a). *R761; Prosesskode 1; Standard beskrivelse for vegkontrakter; Hovedprosess 1-7. Håndbøker i Statens vegvesen,* Directorate of Public Roads and Transport Department. Oslo.

Norwegian Public Roads Administration. (2018b). *R762; Prosesskode 2; Standard beskrivelse for bruer og kaier; Hovedprosess 8. Norwegian Public Roads Administration Manuals,* Directorate of Public Roads and Transport Department. Oslo.

Olatunji, O., Sher, W. & Gu, N. (2010). Building Information Modelling and Quantity Surveying Practice. *Emirates Journal for Engineering Research.* 15 (1). pp. 67-70.

Ontotext. (2021), "Graph DB," available at: https://www.ontotext.com/products/graphdb/ (accessed 2021-02-11).

Pauwels, P. & Terkaj, W. (2016). EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Automation in Construction.* 63). pp. 100-133.

Pauwels, P., Zhang, S. & Lee, Y.-C. (2017). Semantic web technologies in AEC industry: A literature overview. *Automation in Construction.* 73). pp. 145-165.

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.

Protégé. (2021), "Protege.stamford.edu", available at: https://protege.stanford.edu/ (accessed 2021-02-11).

Sacks, R., Eastman, C., Lee, G. & Teicholz, P. (2018). *BIM Handbook: A Guide to Building Information Modeling for Owners, Designers, Engineers, Contractors, and Facility Managers.* 3rd ed., John Wiley & Sons. Hoboken.

Staub-French, S., Fischer, M., Kunz, J., Paulson, B. & Ishii, K. (2021). An Ontology for Relating Features of Building Product Models with Construction Activities to Support Cost Estimating.

Top Quadrant. (2021), "Topbraid Composer", available at: https://www.topquadrant.com/products/topbraid-composer/ (accessed 2021-02-11).

Wikström, L. (2020), "Machine Readable Norwegian Classification Manual for Bridge Registrations - V440", available at: https://awards.buildingsmart.org/gallery/eMXBbjzL/vwAbLmDp?search=0dbf5ddcb220aee7-1 (accessed 2021-01-26).

World Wide Web Consortium (W3C). (2013), "SPARQL 1.1 Protocol", available at: https://www.w3.org/TR/sparql11-protocol/ (accessed 2021-02-01).

Wu, S., Wood, G., Ginige, K. & Jong, S.W. (2014). A technical review of BIM based cost estimating in UK quantity surveying practice, standards and tools. *Journal of Information Technology in Construction (ITcon).* 19). pp. 534-562.

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). pp. 17-31.