Review

Semantic web technologies in AEC industry: A literature overview

Pieter Pauwels\textsuperscript{a,}\textsuperscript{*}, Sijie Zhang\textsuperscript{b}, Yong-Cheol Lee\textsuperscript{c}

\textsuperscript{a}Department of Architecture and Urban Planning, Ghent University, J. Plateaustraat 22, Ghent B-9000, Belgium
\textsuperscript{b}Chevron ETC, Houston, TX, United States
\textsuperscript{c}Louisiana State University, Baton Rouge, LA, United States

\textbf{ABSTRACT}

Over the recent years, the usage of semantic web technologies has notably increased in the domains of architecture, engineering and construction (AEC). These technologies are typically considered as complementary to existing and often used Building Information Modelling (BIM) software. The usage of these technologies in the AEC domains is thereby motivated by (1) a desire to overcome the interoperability issue among software tools used in diverse disciplines, or at least improve information exchange processes; (2) a desire to connect to various domains of application that have opportunities to identify untapped valuable resources closely linked to the information already obtained in the AEC domains; and/or (3) a desire to exploit the logical basis of these technologies, which is currently undisclosed in the AEC domains. Through an extensive literature study and survey, this article investigates the development and application progress of semantic web technologies in the AEC domains in accordance with these three primary perspectives. These examinations and analyses provide a complete strategical map that can serve as a robust stepping stone for future research regarding the application of semantic web technologies in the AEC domains. Results show that semantic web technologies have a key role to play in logic-based applications and applications that require information from multiple application areas (e.g. BIM + Infra + GIS + Energy). Notwithstanding fast developments and hard work, challenging research opportunities are situated in (1) the creation and maintenance of the links between the various data sets and in (2) devising beneficial implementation approaches that rely on appropriate combinations of declarative and procedural programming techniques, semantic and legacy data formats, user input, and automated procedures.

\textsuperscript{*} Corresponding author.

\textsuperscript{a} E-mail addresses: pipauwel.pauwels@ugent.be (P. Pauwels), sijie.zhang@chevron.com (S. Zhang), yclee@lsu.edu (Y. Lee).

1. Introduction

1.1. Building Information Modeling (BIM) and beyond

For two decades now, the concept of Building Information Modelling (BIM) \cite{1} has had a tremendous impact on the architectural, engineering, and construction (AEC) industries, resulting in the generation and broad employment of BIM authoring and application tools. This emerging trend has led to a paradigm change of the industries in ways to define, tailor, and manage the semantics of product models closely linked to geometry. As a result, industry domains and software developers have become more interested in organizing and sharing the ‘semantics’ of a building. This interest is developed for the entire building life-cycle, including not only design and construction, but also facility management (FM), operational building management, building engineering, HVAC design, simulation, renovation, and demolition. Rather than just adopting software applications, which simply display geometric perspectives and views of a building, or lengthy textual descriptions and spreadsheets of unstructured data, the industries have made significant progress towards the development of a robust semantic structure and a well-organized semantic connectivity map.

The semantics advanced by BIM technology has also led to a significant shift in research and development in the AEC industries. A number of the more recent outlook and review articles give an indication of the latest research directions and themes in BIM research. For example, Yalcinkaya and Singh \cite{2} provide a list of 12 research themes, carefully obtained through a latent semantic analysis (LSA) study (which is a Natural Language Processing (NLP) technique) of papers with BIM as a topic. The following 12 research themes are outlined, giving an indication of what is the main interest in current BIM research.

1. Implementation and adoption
2. Energy performance and simulation
3. Academy and industry training

http://dx.doi.org/10.1016/j.autcon.2016.10.003
0926-5805/© 2016 Elsevier B.V. All rights reserved.

Please cite this article as: P. Pauwels et al., Semantic web technologies in AEC industry: A literature overview, Automation in Construction (2016), http://dx.doi.org/10.1016/j.autcon.2016.10.003
4. Information exchange and interoperability
5. Safety management
6. Urban/building space design and analysis
7. Construction and project management
8. Design codes and code compliance
9. As-is, as-built data
10. Promotion and technology development
11. Maintaining and managing facilities
12. Architectural design process

This list clearly shows how interests in BIM research expands towards the entire building life-cycle, including areas like energy performance and simulation, safety management, urban space design and analysis, design code compliance checking, facility management, and architectural design. As BIM research in itself has a lot of focus on ‘information’, research in the areas listed above tends to stress on the ways in which information can be made available for addressing the core research challenges in any of the twelve given research areas. In addition, this significant focus on information has led to increased attention on efficient usage and smooth exchange of data and information, across the various application areas in the building life-cycle.

As an example, Dave et al. [3] gives an insight in the technology requirements for construction management, which is the 7th research domain in the above list. They point towards the Internet of Things (IoT) as a possible means to improve lean construction management, which is typically based on many ad hoc decisions and methods. Clearly, there is a high focus on information exchange and flows in this study. IoT standards have been proposed to allow and improve communication between the multiple devices and systems available in the construction sites and offices, regardless of the system or application features.

Some common concrete research challenges in the other areas of research are given below:

- Enable vendor-neutral model exchange
- Combine different information representations
- Support use case based information exchange
- Manage and share information
- On-site visualization of building information
- Combine product manufacturer data with building data
- Support building performance analysis and optimization
- Generate BIM models from point cloud models
- Model change management (versioning)
- Efficient combination of multiple models
- Connect BIM and GIS
- Enable automated regulation compliance checking
- Check model consistency and completeness
- Logical inference for building energy performance, construction safety, cost estimation, home automation, etc.

1.2. The advent of semantic web technologies in the AEC domain

Each of the research challenges presented above requires the presence of building information in some form. Various information sources (e.g. BIM and GIS) need to be combined and federated for improving the availability and efficiency of information. Considering this high focus on combinations of information and data, it comes as no surprise that there has been an increasing interest in the use of semantic web technologies and linked data technologies. Semantic web technologies namely allow to represent information in structured graphs and efficiently integrate building information of an entirely different nature (e.g. GIS data, FM data, city data, material repositories, regulation data, cadaster data). As a result, the development of software applications that rely on multiple information sources is within reach.

Researchers started to propose the use of semantic web technologies in the AEC industries in the early 2000s. One of the earliest proposals for applying semantic web technologies in the AEC industries is outlined in Pan et al. [4] and Elghamrawy and Boukamp [5]. Early articles focusing on the added value of semantic web technologies similarly see these technologies as one of the diverse sets of web technologies that can bring improvements to information exchange in the construction industry. For example, Aziz et al. [6,7] consider semantic web technologies together with web services and multi-agent systems.

Secondly, semantic web technologies were found useful to increase the value of BIM by enabling data integration and complex search queries across several data sources. An interesting viewpoint on the added value of semantic web technologies to the construction industry can be found in Shen and Chua [8]. They see semantic web technologies as one of three web technologies (semantic search, cloud computing and mobile computing) that are not commonly used in the construction sector, but that could provide considerable value in addition to the already existing technologies (such as BIM).

With the increase of the application of sensing technology in the construction site, the third added value brought by semantic web technology is to incorporate sensing technology to manage construction document information in the field. Elghamrawy and Boukamp [9] first incorporate sensing technology into semantic web technology and present a use case in the field for managing construction document information using RFID-based semantic contexts.

Furthermore, Rezgui et al. [10] and El-Diraby [11] present invaluable discussions and overviews on the reasons to shift from a model-centric approach towards a more distributed semantic approach. As Rezgui et al. [10] indicates, a shift towards the usage of semantic web technologies implies that we need to “try to interpret, accommodate and model what is, rather than trying to change reality to fit a single model. This inevitably results in different ontologies for different communities, but the challenge then is to find ways to allow these communities to collaborate effectively with one another whilst maintaining their existing, efficient, effective separate world views.”

This tendency of using semantic web technologies is recently also embraced in the technical roadmap of BuildingsSMART, which is displayed in Fig. 1. This figure illustrates the three long-standing levels of the technical roadmap, and this is supplemented by a fourth level to the right with ‘semantic search in the cloud’ and a ‘cloud library’. To realize this part of the technical roadmap, the Linked Data Working Group (LDWG) [12] has been launched, aiming to support the usage of semantic web technologies in the construction industry, such as the Resource Description Framework (RDF [13]) and the Web Ontology Language (OWL [14]).

1.3. Promises and expectancies

The primary question this article investigates is what has been and can be obtained by adopting semantic web or linked data technologies for the AEC industries. In this investigation, we consider three main topics that are often used in arguing for the usage of semantic web technologies in the design and construction industry.

1. Interoperability: The usage of semantic web technologies has been considered as an opportunity to improve interoperability in the AEC industries [16–21], thus resulting in an integrated and successful data exchange environment. Namely, semantic web technologies appear to provide a way to describe information in a computer-understandable manner. To rephrase Berners-Lee et al. [22]: “the Semantic Web will enable machines to comprehend semantic documents and data”. So, from this hypothesis, one can assume that it should be possible to apply these technologies in the construction industry and thus enable computers in this industry to understand the
We investigated the most recent articles pertaining to semantic web technologies in the AEC industry in order to identify the development and application progress of semantic web technologies in this industry according to the three topical axes outlined above.

1.4. Methodology

We focused on articles in SCI-indexed journals in this domain, in particular:

- Automation in Construction
- Advanced Engineering Informatics
- Journal of Information Technology in Construction
- Journal of Computing in Civil Engineering
- Expert Systems with Applications
- Journal of Construction Engineering and Management
- Computer-Aided Design

Instead of making a purely quantitative analysis of articles, as was for example done for BIM research in Yalcinkaya and Singh [2], we have made a qualitative assessment of the information in the surveyed papers and critically analyzed to what extent contributions are made to validating any of the three outlined topical axes. Where needed, we have included reference articles that were cited in the core set of articles. Table 1 briefly illustrates the identified use cases of semantic web technologies adopted in the AEC industries.

Similar to the aim of this paper, Abanda et al. [26] reviewed several studies pertaining to semantic web technologies and their applications on built environment domains. Abanda et al. [26] pursues two objectives: (1) grasping overall trends of semantic web applications used in the built environment domain and (2) demonstrating the different aspects of applied technology in the management of built environment information. In addition, the paper illustrates findings with regards to the emerging progress of semantic web technologies. As there have been considerable developments in this domain over the past few years, however, our work will go in a considerable more depth, both in the technical overview and the overall analysis and discussion.

The results of our assessment are discussed in Sections 3 to 5. In Section 6, we give a qualitative overview of our conclusions from the literature study and we outline recommended future directions for research and industrial applications. But, first, the following section gives a brief overview of the main concepts and current status of semantic web technologies.

2. Semantic web technologies

2.1. The RDF core

At the core of the semantic web stands a flexible and generic language that allows to easily represent and combine information from diverse knowledge domains, namely RDF [13]. The semantic web approach.

Please cite this article as: P. Pauwels et al., Semantic web technologies in AEC industry: A literature overview, Automation in Construction (2016), http://dx.doi.org/10.1016/j.autcon.2016.10.003

![Technical roadmap for product support](image-url)

Table 1. The technical roadmap for product support by BuildingSMART, with web technologies represented in the right side of the graph (original figure in Ref. [15]).
web thus becomes a semantic network in which information is represented as directed labeled graphs (RDF graphs). Each node in such a graph represents a concept or object in the world, identified with a Unique Resource Identifier (URI). By describing all information as such in interlinked directed labeled graphs, a uniform representation of information is achieved, making information reusable by both humans and computer applications.

An RDF graph can be serialized using various syntaxes including RDF/XML, N-Triples, Turtle, and Notation-3. RDF graphs can be given an improved semantic structure using RDF vocabularies or ontologies. The most basic elements describing such ontologies are contained in the RDF Schema (RDFS) vocabulary, which consists of the specifications of classes, subclasses, comments, and data types. An RDFS interpreter is able to infer extra RDF statements that are implicitly available via the RDFS constructs. More expressive elements to describe ontologies are available within OWL. In short, OWL further enhances the RDFS concepts to allow making more complex RDF statements, such as cardinality restrictions, type restrictions, and complex class expressions. The RDF graphs constructed with OWL concepts are called OWL ontologies.

OWL(S) and OWL provide the basis to allow working with rules and proofs. By relying on rules and proofs, it is possible to build applications that reach particular levels of trust, precisely because of the way in which they deploy their rules and build their proofs. This idea of building semantic web applications is nicely displayed in ‘the semantic web stack’, as it was originally presented by Berners-Lee. There have been many versions of this semantic web stack. An indication of the diversity and the kinds of discussions that have been circling around this semantic web stack can be found in Horrocks et al.

### 2.2. OWL semantics and OWL profiles

The semantic expressiveness of the OWL language defines what can be represented in an OWL ontology. This semantic expressiveness is specified in multiple W3C hosted specification documents. The first W3C Recommendation for OWL dates from 2004. This version is now superseded by the OWL2 language specification issued in 2012. All relevant references to the exact semantics of OWL2 can be found in Ref. Fig. 2 provides an overview picture that we will use to explain the basics of OWL profiles.

As pointed out in Ref. [33], “the Direct Semantics assigns meaning directly to ontology structures, resulting in a semantics compatible with the model theoretic semantics of the SROIQ description logic – a fragment of first order logic with useful computational properties”. This leads to a semantic expressiveness for OWL2 that is properly grounded in a particular description logic, namely SROIQ. This semantic expressiveness is graphically displayed as the outer ellipse in Fig. 2 (OWL2 Full). However, “some conditions must be placed on ontology structures in order to ensure that they can be translated into a SROIQ knowledge base” [33]. For instance, transitive properties cannot be used in number restrictions. Whenever an OWL2 ontology satisfies these conditions, the expressiveness of the ontology is in the smaller outer ellipse in Fig. 2, namely OWL2 DL. An OWL ontology should remain within this boundary if it is to be used by SROIQ-based tools, which are the tools typically supplied by the semantic web community.

As in the case of OWL, also OWL2 has a number of so-called profiles, namely OWL2 EL, OWL2 QL and OWL2 RL [35]. Fig. 2 displays the relationships between these three key profiles. As outlined in Motik et al., an OWL2 profile “is a trimmed down version of OWL2 that trades some expressive power for the efficiency of reasoning”. In short, in each of the given OWL2 profiles, a number of statements that can be used in OWL2 DL is not allowed. By not allowing these statements, and thus sacrificing some expressiveness, important improvements can be made in terms of performance. Namely, inference engines do not need to check a number of restrictions as they are not allowed (and thus not considered) in particular profiles. More information about the expressiveness of each of the profiles can be found in Motik et al.

#### 2.3. Closed World Assumption (CWA) vs. Open World Assumption (OWA)

Two distinct approaches to knowledge representation are relevant when dealing with traditional technologies (e.g. BIM) versus semantic web technologies: Closed World Assumption (CWA) and Open World Assumption (OWA). According to CWA, any statement that is not known to be true, must be considered as false. When applied to a BIM model, one can conclude that whatever something is not specified in the model, it is most definitely not there. On the other hand, according to OWA, a statement that is not known to be true, is not necessarily false, nor true, but unknown. In other words, it might be true or false in the future, when more information is supplied, but no conclusion can be drawn until then.

Many traditional software applications adopt a CWA, including BIM tools and common database systems. Semantic web technologies, however, generally rely on an OWA because the technologies are supposed to be used on the Web, which is a system with incomplete information. One cannot conclude that something...
is not true simply because no one specified it on the web. Hence, an OWA needs to be adopted. The difference between CWA and OWA plays a key role when an ontology is used to represent a BIM model, because if something is not specified, then one cannot conclude much, except that it might still be true or false. A whole different kind of information usage and inference becomes available.

Mapping information representations in CWA to information representations in OWA is not that hard; the main difference lies in the usage of the information that is presented in both. Furthermore, it is even possible to run a CWA-based validation of an OWL ontology [36,37]. However, the OWA of semantic web technologies is still something different from the traditional CWA features in current software applications. In many cases, both types of assumptions have their value (e.g. Terkaj and Sojic [38]). If adopted properly, the usage of semantic web technologies is a fruitful addition to (and not replacement of) existing technologies such as BIM authoring tools.

2.4. Linked data vs. semantic web

Other terms that can regularly be found in relation to semantic web technologies are linked data, web of data, and semantic web (see also Abanda et al. [39]). Web of Data is commonly associated to linked data, whereas semantic web is commonly considered as quite a different thing. The term ‘semantic web’ was coined by Tim Berners-Lee in 2001 [22] and was quite visionary as it included all features in the semantic web stack (OWL, rules, proofs, truth). The term ‘linked data’, on the other hand, was coined in 2006, also by Tim Berners-Lee [40], in response to the finding that quite some data was being published on the web, seemingly following the semantic web idea but actually never linking to outside data, and thus in fact not realizing the initial core idea behind the semantic web, which is linking data [22]. Therefore, Berners-Lee [40] laid out four rules that need to be followed in order to truly obtain linked data. These have by now evolved into the five stars of linked data [41].

Ontologies, rules and proofs are clearly not discussed in this linked data proposal by Berners-Lee [40]. As a result, the linked data or web of data field emerging out of this proposal typically leaves aside most of the other elements in the semantic web, including OWL, rules, proofs and so forth. A key result of this research domain is the Linked Open Data (LOD) cloud, for which statistics and an overview are available in Jentzsch et al. [42], Auer et al. [43], and Schmaachtenberg et al. [44].

3. Aim 1: interoperability

With the key principles and concepts of semantic web technologies explained in Section 2, the current section (interoperability) is the first of three sections documenting how semantic web technologies can be deployed for added value in the AEC domains. In terms of interoperability, one can distinguish between syntactic and semantic interoperability [45]. Reaching full semantic interoperability is one of the long-standing desires in the architectural design and construction industry. Hannus et al. [46] illustrates this desire using the ‘islands’ image depicted in Fig. 3. If one island is interoperable with another island, it would imply that a building model can be serialized to a common data model so that it can automatically be parsed into a different data structure (e.g. from BIM model to simulation model), and back. Different representations of the same object would thus be available (cfr. note the difference between interoperability and linking across domains outlined in Section 1). This would allow product data models to perform well in diverse domains, processes, and modeling platforms while preserving geometries, semantics, relationships, and properties.
usage practice because heterogeneous IFC translation and binding processes of each BIM authoring tool could result in unintended geometric transformations and semantic errors [48]. These poorly implemented as well as poorly used import and export features will only result in distorted and incomplete IFC models, not interoperable IFC models.

2. Adaptability: Another challenge is the limitation in quick adaptation of the schema. Indeed, various industry domains and software vendors require a tremendous amount of work to agree upon and generate the compliant IFC schema and associated specifications. It thus requires a significant amount of time to reach consensus on its content, representation and definition scope, so that the schema typically does not encompass state-of-the-art technology and newly-launched construction methods.

3. Extensibility: Somewhat related to the last point, the IFC schema is also not easily extensible for relevant people who are not highly familiar with the EXPRESS language. If one wants to express information that is not available as such in the IFC schema (and he does not know how to do this in EXPRESS), IfcProxy concepts and custom IFC property sets can be used. However, these are semantically very loose additions, in the sense that anybody who wishes to use this information will need to manually interpret strings in their application logics.

3.2. A semantic alternative

In response to the earlier outlined challenges, it was therefore investigated whether semantic web technologies might provide an alternative technical means to address the interoperability issue [16–21,50]. Making an agreed ifcOWL ontology available is a first required step in being able to make this test. The BuildingsSMART LDWG has therefore produced an ifcOWL ontology that can serve as a domain ontology for the construction industry. The LDWG working group has heavily relied on earlier proposals to convert the IFC schema into an OWL ontology and to convert IFC STEP Physical Files (SPF) into RDF graphs that follow the ifcOWL ontology [51–58]. One of the most detailed and recent accounts of the conversion from EXPRESS to OWL for the construction industry (ifcOWL) is available in Pauwels and Terkaj [59]. The ifcOWL ontology has now evolved into a recommendable status. It closely resembles the ifcOWL ontology as it was proposed in Pauwels and Terkaj [59]. It is maintained in Ref. [12], and a free IFC-to-RDF conversion service is provided and maintained in Ref. [60]. This conversion service allows to upload IFC files, which are then returned to the user as RDF graphs that follow the ifcOWL ontology.

Further modifications and extensions to the ifcOWL ontology are now proposed, based on critical ontological analysis and tested alternative suggestions. For example, Terkaj and Sojic [38] have proposed an extension to the ifcOWL ontology, which encapsulates the EXPRESS WHERE rules in OWL class expressions that can be combined with the standard ifcOWL ontology. Furthermore, it is suggested in Borgo et al. [61] and de Farias et al. [62] that the way in which type information is now included in IFC can be improved, especially within an ifcOWL ontology. This proposal has been further developed by de Farias et al. [62] into an ifcWOD ontology, which formally extends the ifcOWL ontology. In addition, a fuzzy extension to the ifcOWL ontology has been proposed in Gómez-Romero et al. [63]. Another issue that is typically considered to be open for change, is the way in which LIST data types in EXPRESS are translated into OWL class expressions. This is an issue that has been briefly investigated in Pauwels et al. [58] and de Farias et al. [62]. This is closely related to the question whether geometric information, which carries little semantic meaning, should actually also be converted into an RDF graph. As is already indicated in Beetz et al. [64], an RDF representation of this geometric information is rather inefficient and does not add much additional value as long as it is not used in a logical inference process.

3.3. Binding parallel representations

Of course, adopting semantic web technologies (ifcOWL) cannot address bad implementation and usage practices. However, (1) as they provide a single data model (RDF) for representing any kind of information; (2) as they allow adding a logical DL basis to this representation using OWL; and (3) as they focus intensively on linking diverse graphs of information together in a web-like fashion, semantic web technologies might be the ideal technical means to provide interoperability while also allowing to flexibly handle new semantic structures (see ‘extensibility’ and ‘adaptability’ above).

For example, a semantic web approach has been suggested to improve the interoperability of CAD information by Abdul-Ghafour et al. [17]. The authors indicate how semantic web technologies allow the combination of information from several different knowledge domains, enabling a seamless coupling of 3D information with non-geometric information, such as design intent and domain-specific product features. Also, García [65] proposed to capture core CAD file formats (e.g. DXF) in OWL ontologies in order to make CAD data available as a knowledge base. OWL is hereby proposed as a CAD data exchanger. This short article only looks into one CAD standard, but proposes to do the same for other CAD data formats (cfr. interoperability of CAD data). Similar arguments and conclusions were made in Pauwels et al. [18] and Argüello et al. [66].

So, semantic web technologies seem to provide interoperability by allowing the co-existence and linking of multiple ontologies, often to represent the same physical elements. Indeed, semantic web technologies allow to combine different representations of information, e.g. a box-like wall in IFC, X3D and STL [67]. However, how such parallel representations of information are to be combined and used is entirely left open for the developer’s and user’s choice. This investigation is related to the ‘binding’ challenge listed in Section 3.1. An IFC, X3D and STL representation of a box can all be interlinked in diverse ways, but if one of these representations changes, the same changes should also be made for the other two representations in an interoperable system. We found several proposals in literature, which can be classified in two distinct approaches: (1) link sets and (2) mapping schemas in formal rules.

---

Fig. 4. IFC is available in an EXPRESS (native), XSD and OWL format on a schema level (right). The IFC SPF, ifcXML, and ifcRDF then follow the corresponding schema an represent instance files.

Source: Adapted from original figure in Ref. [49].
3.3.1. Link sets

One interesting solution proposed in El-Gohary and El-Diraby [68] is an ontology integrator (Onto-Integrator) for facilitating ontology interoperability within the AEC domains. The Onto-Integrator offers a heuristic for ontology merging, including the merging of concept taxonomies, relations, and axioms. Furthermore, Törmä [69] argues for the need for instance-level interoperability in addition to ontology-level or schema-level interoperability. This is particularly important when actual exchange of partial models (requirement model, architectural model, MEP model) takes place, in which the key challenge is to find out which elements in the diverse partial elements are actually identical (e.g. diverse representations for one and the same wall, window, slab). Törmä [69] points out, indeed, that “the difficult part, the use of the exchanged information in the receiving tasks remains largely unsolved: it requires human interpretation and manual work”.

There is no mechanism or suggestion made in the semantic web domain to properly deal with this situation. This leaves a high risk of redundancy and inconsistency of information (see also Törmä et al. [70] and Scherer et al. [19]). This situation is investigated in Pauwels [71], leading to the schema displayed in Fig. 5. It shows a decentralized web of linked data in the center, with many diverse interlinked RDF graphs. The blue arrows require a transition mechanism such as mapping or conversion, which can in theory be implemented in many ways, but which in practice typically requires human interpretation. For example, Scherer et al. [19] and Törmä [69] propose to implement this transition mechanism with linked data technologies, which results in ‘linksets’: sets of links that represent the relationships between partial models (i.e. the blue circular arrows in Fig. 5). These linksets still need to be managed through human intervention though. In other words, the interoperability problem is moved to the data level (creation and management of links), but it is still undoubtedly there.

If a manual approach is adopted for creating and managing the links between multiple ontologies and datasets, automation support is available from ontology alignment tools. Ontology alignment typically occurs by the “automated comparison and mapping based on the ontology structures and the linguistic similarity between concepts” (Cheng et al. [72]). As a result, the result of an automated ontology alignment procedure should always be checked for misinterpretations caused by erroneous linguistic associations (hyponyms, homonyms, synonyms). As is outlined in Cheng et al. [72], however, they can prove to be of invaluable help in mapping several ontologies or semantical data sets.

3.3.2. Mapping schemas in formal rules

An alternative automated method for dealing with the mentioned interface points and providing interoperability was suggested in Refs. [62,67,73]. Assuming that all RDF graphs in the central web of data follow a specific OWL ontology, and thus have a solid formal structure, it should be possible to devise a mapping schema between specific pairs of schemas. By representing that mapping schema in formal rules, one can use an inference engine to automatically infer data in alternative ontologies, starting from data in a master ontology or central ontology [67]. This inference process can occur on demand: one geometric description (e.g. IFC) is available in an RDF graph and descriptions following a different schema (e.g X3D, STL) are generated on demand by a rule engine and a set of inference rules.

In this regard, Beetz et al. [74] proposed to use Semantic Web Rule Language (SWRL) rules and SPARQL queries to convert on-demand IFC geometry into alternative geometric representations, e.g. topological space-centered geometry used by energy simulation software. de Farias et al. [73] propose to use SWRL rules to generate a simplified version of the IfcOWL ontology. As SWRL rules match particular patterns in an IFC/RDF graph, alternative representations of that graph are produced on demand. In the case of de Farias et al. [62], the focus is entirely on the adaptation of IfcRelationShip instances into simpler constructs. Note however, that such a rule-based approach to managing the interface points in Fig. 5 also requires implementers to nominate a master ontology and devise a solid set of mapping rules, which will again require a considerable amount of human interpretation, and which is in itself not that different from a traditional IFC approach (Fig. 3).

In theory, this approach of mapping schemas in formal rules can also be used in the context of Model View Definitions (MVDs). An MVD is a subset of the building product model schema (the IFC schema) that provides a complete representation of BIM exchange data needed for a particular domain of the AEC industries. MVDs are currently captured in mvdXML files. The information that is currently captured in an mvdXML file can also be captured using semantic web rules (SWRL, N3Logic, and other) or SPARQL CONSTRUCT queries, as is proposed in Beetz et al. [64] and Weise and Pauwels [75]. In principle, this would allow to ‘query’ the complete IFC/RDF file and automatically output the required subset on demand. One advantage of this approach is that the process can be implemented using regular semantic web technologies, including out-of-the-box inference engines, triple stores and query interpreters. The greatest advantage of this approach, however, is likely that it allows a far more flexible mechanism to generate subsets. Both semantic web rules and SPARQL construct queries namely allow to output information that does not have to follow the IFC schema (as opposed to a regular MVD subset). As such, output might be generated that automatically matches the semantic information structure of a target program, hence supporting interoperability in an alternative fashion (mapping schemas in formal rules).

The approaches that were suggested by de Farias et al. [62] Pauwels et al. [67], and de Farias et al. [73] can be considered the extreme examples of this semantic subset selection and publication method, as they all start from IFC data and output it partially to an alternative schema. de Farias et al. [76] moves this further and documents a federated ontology architecture (POWL) that is based on ontology alignments, logical rules, and inference mechanisms.
The two main components in this architecture, the Federal Controller (FC) and the Federal Descriptor (FD) are placed in-between a knowledge base and a user interface (Fig. 6). These two components take into account the ontologies used in the query coming from the user interface and translate the query via ontology alignments and inference mechanisms to the ontology structures used in the knowledge base. A technical overview of the query translation process and SWRL rule selection process is given in de Farias et al. [77]. Hence, data and custom data structure become available in an on demand kind of fashion. Yet, also in this case, we are not close to a practical implementation in construction industry.

4. Aim 2: linking across domains

An aim that might be easier to reach using semantic web technologies involves linking (instead of mapping) information stemming from diverse domains (e.g. BIM, GIS, heritage, sensor data, simulation data, smart cities) into one web of linked building data. This purpose has been the main driver behind a myriad of individual research initiatives. Some of these approaches are more closely affiliated to linked data rather than semantic web technologies (see Section 2.4), although ontologies equally often play a crucial role as well. When specifically relying on ontologies, the key research question is often related to the creation of domain ontologies, which aim at providing a shared representation for the concepts within a domain of knowledge [11], and how they should be linked together and still remain useful (see also Tőrő [69,78]). A highly recommended set of guidelines has been proposed in this regard by Radulovic et al. [79], focusing specifically on construction industry use cases.

4.1. Collaborative information management

The building industry is divided in many specialized disciplines to design, construct, and operate a building. Hence, construction projects are characterized by a strong need for close collaboration between AEC team members and require an effective approach to the management of information from diverse sources [6,21,80,81]. One of the earliest examples of ontology-based information management was proposed by Lima et al. [82–84] (e-Cognos project) in their proposal to build a domain ontology for the construction domain, which describes four key elements in construction, namely actors, resources, processes and products. In addition, Katranuschkov et al. [85] can be counted among the early explorers of ontology-based collaborative information management. Many of the lessons learned from the e-COGNOS project are documented in El-Diraby [11], which presents a domain ontology for construction knowledge (DOCK 1.0), starting from the earlier work spanning 2005 to 2013 in building a domain taxonomy for the e-COGNOS platform [86–92] and building a beta DOCK ontology for construction products and processes [93–95]. The objective of this ontology is “not to exhaustively list concepts, but rather to build a conceptual architecture of key terms in construction, their relationships, and behavior” [11]. In this article, it is argued that the main contribution of building an ontology is an improved understanding of the actual domain of discourse (conceptualization over mere formalization). Indeed, one is typically more critically challenged when modeling an ontology from scratch rather than formalizing it in a new syntax. One could follow this idea to argue for an ontology design for IFC (or other legacy data models) starting from scratch, in order to avoid inconsistencies in the ontology because of the transition from an alternative syntax to the OWL syntax [20,61].

Of the other efforts in building domain ontologies for improved collaborative information management, Ruikar et al. [80] proposed an extensible set of modular ontologies (design-process ontology and team profile ontology), which are then deployed in an ontology-based knowledge-sharing environment (OntoShare) for usage by various stakeholders in the construction industry. Later, Anumba et al. [81,96] discussed the role of ontologies in information and knowledge management in construction and developed a framework for semantic web-based information management (SWIMS) for effective collaborative information management. Zhang and El-Diraby [97] proposed a social semantic web portal (SSWP) for effective social communication in construction. Zhang and El-Diraby [98] furthermore proposed a construction information and knowledge portal (CIKP) as an information and knowledge management system that addresses the challenges in information exchange and knowledge sharing. A similar social networking website is proposed by Niknam and Karshenas [99] for improving collaboration in AEC projects. Both Niknam and Karshenas [99] and Zhang and El-Diraby [97] take a service-oriented approach in the implementation. Furthermore, applications targeting the support of collaboration also typically look at the entire life-cycle of a building, aiming at providing a holistic view of the building and the building process. For instance, Scherer et al. [19] uses semantic web technologies to allow an integrated lifecycle energy management. This work is closely related to the work in Scherer and Schapke [100], in which a multi-model driven construction management system is proposed with a layered ontology framework at the center. The ontology framework includes a Project Collaboration Ontology (PCO), which is composed of five sub-ontologies: a Construction Core Ontology (CCO), a multimedia visualization ontology, a software service ontology, an organization ontology, and an information process ontology [100]. The system aims at sharing distributed, yet interrelated, engineering and management application models in so called multi-models. So, also this approach focuses on a linking data implementation plan rather than a centralized information model plan. Management of the links between the models is done using separately managed link models (LMs). As another example, Le and Jeong [21] aim at supporting, with linked data techniques, decision making and asset management for the preventive treatment selection process for flexible pavements, which requires the comprehensive analysis of multiple factors in many domains in all phases of the entire life-cycle of that particular pavement. In this case, all native data are wrapped into a ‘unified RDF layer’, which can then be queried for decision support (see Fig. 7).

The goal of supporting collaboration through a common platform using semantic web technologies is clearly not that much different.
from the interoperability aim discussed in Section 3. Hence, very similar issues and challenges reside in the research initiatives focusing on this kind of inter-domain linking:

- How to keep parallel representations of the same thing consistent and complete [20,71];
- How to keep the links between partial models effectively manageable [70];

Indeed, creating RDF graphs with construction data, linking and using them for querying and reasoning is often explained, but the way in which the links between data sets (the blue circular ‘human interpretation’ arrows in Fig. 5) need to be created and managed, typically remains unexplained. This challenge is also specifically mentioned by Kiviniemi [101].

4.2 Product manufacturer data

Another typical example in direct support of the construction industry is the combination of product manufacturer data with building data. In a number of pilot cases, product manufacturer data are represented in building product catalogs using semantic web technologies, making it possible to integrate these data directly with building data in RDF (e.g. ifcOWL). A key semantic repository of building material data is the BauDataWeb Materials Database [102]. Another example is provided by the BAUKOM repository [103–105]. Furthermore makes an integration with regulation data with the aim of enabling building product recommendations based on the building model and applicable local building regulations. [106] designed and implemented an open knowledge base to capture, distill, analyze and share information on building sustainability among the stakeholders. As another example, [107,108] illustrate how the semantic connection of a building model with product data and cost data can be used to enable automatic construction cost estimation. Niknam and Karshenas [108] specifically use an approach that relies on semantic web services. Such semantic web services are also used for building a social networking website for AEC projects [99] and for building an energy analysis application [109]. Also Staub-French et al. [110] focused on construction cost estimating, as early as 2003, and proposed a feature ontology that is designed as an enrichment of the IFC schema for those features that are relevant for construction cost estimating. Although it is not mentioned whether this feature ontology is designed using (semantic) web technologies, the principles explained in Staub-French et al. [110] are close to the idea of using semantic web technologies.

As soon as product data catalogs are to be used on more international scales, the required formal representation of product manufacturer data rather quickly leads to the need for classification in building elements as part of large concept libraries. Building such an international concept library is one of the main goals of the BuildingSMART Data Dictionary (bSDD). Such a multilingual dictionary of terms or concept library can be used to clarify what products are made available. As is explained in Ref. [111], however, “content ownership, trust and reliability issues can be identified as some of the main obstacles for the large-scale adoption of the bSDD/concept libraries: Users and implementers alike hesitate to rely on a single location and provider for the access of content due to issues with overwhelming number of concepts, the dependency of a centralized system and the fact that the vocabulary is subject to constant evolution”. As a remedy, Beetz [111] proposes to rely instead on networked, distributed/decentralized concept libraries relying on semantic web technologies for the publication of concept libraries. Three approaches are outlined in this regard: a single centralized concept repository, distributed peer-to-peer concept repositories, and a hybrid network architecture. Decentralizing the bSDD using semantic web technologies could then result in a multi-tiered and cascading management of concept libraries [111].

This management architecture would contain a centrally managed international reference concept library (bSDD) that is first surrounded by and directly linked with national concept libraries, which are in turn surrounded and directly linked with more local concept libraries. The content in the more local dictionaries can be more specialized and relaxed, hence responding to local needs. Yet, as they are linked to the central curated bSDD, the curators of the bSDD can be more easily informed of desirable changes. Also here, a critical open issue remains the way in which links between concepts and concept libraries are actually to be created and managed over time. This open issue is again closely affiliated to the required human interpretation steps indicated with the blue circular arrows in Fig. 5.

4.3 Building performance analysis

Many research initiatives developed applications to support building performance analysis and optimization leveraging semantic web technologies. These building performance analysis use cases typically focus intensively on the design stage of a building, as well as the operational stage of a building (monitoring). For instance, Curry et al. [112] combined scenario modeling and linked data to support building design decisions in both stages. Curry et al. [113] and O’Donnell et al. [114] further extended the usage of linked data for combining diverse cross-domain building data in support of operational building management support. This work focuses a lot on the objectives (events) and scenario’s in which the building data is meant to be used. Corry et al. [115] discussed using semantic web technologies to aid the integration of ‘soft’ AEC data (qualitative social media messages) into an existing building performance measurement framework for evaluating building performance in the operational phase. Further work towards the development of a full performance assessment application, relying on a performance assessment ontology, is documented in Corry et al. [116]. Also Tomasevic et al. [117] focuses on the operational phase and discusses how the same ontology-based approach to building performance analysis can be relied upon to provide direct and useful feedback to the facility manager. Furthermore, Dibley et al. [118,119] proposed an OntoFM system, in which real-time building monitoring is supported through the combination of a multi-agent system with access to semantic building data. The building data follows a building ontology based on IFC, a sensors ontology that relies on the OntoSensor ontology [120], and a general purpose ontology —
Suggested Upper Merged Ontology (SUMO [121]), which captures domain independent concepts. The system also relies on rules and specifically indicates that computational performance can vary significantly depending on which reasoning approach is taken (amount and expressiveness of input data and reasoning strategy), as is also one of the key points in Pauwels et al. [122].

These energy performance analysis studies typically attempt to make a combination of core building data (e.g. IFC) and energy simulation data. One primary example of energy simulation data is the SimModel, which was devised as an interoperable data model for exchange of simulation data between energy simulation tools. This model is made available in an OWL ontology [123], which can be used to generate RDF graphs of these model data [124]. By porting the data into an RDF data model, they can be more easily combined with other RDF data. However, how IFC and SimModel graphs are actually going to have to be combined and managed is open for discussion.

Abanda et al. [125] developed and proposed an OWL ontology for photovoltaic (PV) technology. The system developed on top of this ontology (PV-TONS) then allows to size and select PV-system components for different types of buildings. This work relied on the work initially done in Tah and Abanda [126]. More precisely, Tah and Abanda [126] resulted in a PV ontology, which needed to be further enriched for this study with external factors (sunny availability, tree shades) to allow the considered practical use case implementation. In yet another energy-focused research initiative [127,128], a ThinkHome OWL ontology was developed, including concepts related to resources (white and brown goods), building (layout, spaces, material), actions (schedules, preferences, context), energy (environmental impact, energy providers), comfort (thermal and visual), and exterior influences (weather, climate). This ThinkHome ontology is combined with an OWL ontology that captures gbXML building geometry and data (instead of IFC). An agent-based system then interacts through SPARQL queries and DL inference with the data (Fig. 8) in order to autonomously control the smart home in an energy-efficient and comfort-oriented manner [128]. The ThinkHome ontology and project relies heavily on the data coming from household appliances. This is inspired to some extent by the ontology-based household device models in DomoML [129] and the Domotic OSGi Gateway (Dog) Ontology (DogOnt) [130]. The DOG ontology is also used for energy management purposes in Rosselló-Busquet et al. [131], in combination with a set of SWRL rules, the Jess rule engine [132], and the SWRL-JessBridge [133] that allow to automatically apply energy saving strategies.

Schevers et al. [134] provides an example focused on the combination of IFC data with FM data that is available in relational databases for the Sydney Opera House. By doing so, a digital facility model is made available that gives transparent and integrated access to the available information. In this work, a well-balanced system is devised that combines the best of relational database, OWL ontologies, a geometric visualization engine and existing software. The COMANCHE system presented in Meshkova et al. [135] deals with home devices and appliances as well, but it focuses its ontology more on the complex relations between services, software, users, and providers. Furthermore, Ricquebourg et al. [136] proposes to implement smart home services using a combination of RDF (ABox), OWL (TBox) and SWRL (RBox). This would allow to automatically process incoming sensor data using SWRL rules and make actions in the smart home accordingly, e.g. switch on the kitchen lights when it is sensed to be occupied. SESAME is a similar system that is proposed in Tomic et al. [137], aiming at the integration of a smart metering system with a building automation system through a middleware semantic layer. This approach relies heavily on logical inference (SWRL) and semantic web services, while putting a full context model of the building (smart metering + building automation) at the center. Also Shah et al. [50] proposes an ontology for

![Fig. 8. Overview of the ThinkHome system (original figure in Reinisch et al. [127]).](image-url)
Corrado et al. [148], “semantic technologies have been used to create models of urban energy systems able to assess the energy performance of an urban area”. This system relies on the semantic combination of census data, land registry data, energy data, building system data, and climate data (see Fig. 9). One of the main drawbacks outlined by the authors [148] is the fact that the described data keeps being modified, now and in the future, which will require continuous updates and maintenance efforts. The main issue is thus again situated in the creation and management of the links. However, it seems that in many of these approaches, it is considered a useful approach to take this semantic approach and rely on engineers for link management.

4.4. Regulation compliance checking

As for construction-specific applications, researchers have looked into regulation compliance checking as well (see also Section 5). One of the most commonly known industrial cases in automated regulation compliance checking, although not relying on semantic web technologies, is the ePlanCheck system in Singapore [149]. According to Eastman et al. [150], rule-based systems are understood as systems that “apply rules, constraints or conditions to a proposed design, with results such as ‘pass’, ‘fail’ or ‘warning’, or ‘unknown’ “. In the context of regulation compliance checking [149], a number of logical frameworks has been considered that could produce these outcomes. For example, Kerrigan and Law [151] indicated the usefulness of a logical basis in regulation compliance checking, as early as 2003. At that moment, this process was implemented as an addition to plain XML. Alternatively, the usage of conceptual graphs, which is based on First Order Logic (FOL), has been proposed for the formalization of the regulation compliance checking process by Solihin and Eastman [152]. Furthermore, a number of open standard rule engines have been proposed for the representation of regulatory knowledge for compliance audit purposes. In particular, Drools and its Drools Rule Language (DRL) has been suggested and used by Beach et al. [153] and Solihin and Eastman [152].

As semantic web technologies typically rely on the decidable subsets of FOL, including DL, we will focus on these proposals. Yurchyshyna and Zarli [154] proposed a formal ontology-based approach for the formalization and semantic organization of conformance requirements based on SPARQL. Zhang et al. [155] demonstrated a use case on BIM-based hazard identification and mitigation showing detailed ontology development, instance generation, a combination of SWRL rules and a rule engine, and finally visualization of checking results in a BIM authoring tool. Zhong et al. [156,157] used a domain ontology for assisting construction quality compliance checking and supporting the plan definition and verification process in pit excavation. Furthermore, Pauwels et al. [158] aimed at implementing acoustic regulation compliance checking for BIM models.

Note however, that the outlined research initiatives typically conclude that only about 70 to 80 % (estimated) of the regulatory knowledge in a building regulation can be explicitly and unambiguously be formalized (see also Nawari [159] for an outline of the difficulties). The other 20 to 30 % is too vague or qualitative, and therefore requires human interpretation in the regulation compliance checking process. This will likely remain a key issue in this domain. Note here that choosing more expressive (rule) languages allows to express more of the rules in a regulation, but using these languages on the other hand also results in more complex (often undecidable) representations and lower run-time performance (it takes longer to process the rules).

Besides the expressiveness of the rule language (see also Section 5), also the conversion from a natural language regulation text to a formal representation (or reverse) remains a critical research issue. A recent research effort specifically aiming at this challenge is the Requirement, Applicability, Selection, Exception (RASE) tagging mechanism [160,161]. This RASE system suggests to add XML tags to regulation texts, which can then be automatically interpreted in order to obtain a formal representation of the regulation (in SWRL, DRL, IfcConstraint Instances, SPARQL queries or whichever). An alternative approach is to rely on Natural Language Processing (NLP) techniques, as proposed in Salama and El-Gohary [162], Zhang and El-Gohary [163], and Yeung et al. [164].

4.5. Geographical and infrastructure data

Semantic web technologies also benefit the integration of data in the AEC domain with data that is typically outside this domain. For example, Geographical Information Systems (GIS) are applied throughout the different phases of any civil infrastructure project. GIS data standards (e.g. CityGML) are managed by the Open Geospatial Consortium (OGC). This community has been turning towards the usage of web technologies and semantic web technologies over the last 10 to 15 years. This has resulted in a higher availability of GIS data on the web, allowing better and more usability of the data [165,166]. Also when remaining in the geographical domain, the issue of managing multiple models containing the same information (interoperability) remains a key issue, as is explained in Métral et al. [165] by investigating a combination of a CityGML ontology, an ontology for transportation systems (OTS), and an ontology of urban planning process (UOPP). Métral et al. [167] furthermore combines this approach with archaeological urban data, maintaining CityGML as the central ontology.

The integration of GIS, BIM and CAD using web technologies has been a topic in the AEC industries as well as in the OGC. This is turning towards the adoption of semantic web technologies in more recent work. For example, Akinci et al. [168] propose a web-service based approach that enables semantic interoperability between CAD and GIS platforms. El-Mekawy and Östman [169] propose the usage of an intermediate reference ontology, the Unified Building Model (UBM) to implement the bi-directional mapping between IFC and CityGML. Mignonard et al. [170] propose a SIGA3D system that extends BIM information with geographical information to allow urban facility management. Irizarry et al. [171] integrate BIM and GIS into a system for improving visual monitoring of construction supply chain management. Karan et al. [172] and Karan and Irizarry [173] further utilized semantic web technologies to facilitate such integration and interoperability of BIM and GIS. Note that all these approaches focus a lot on semantic interoperability of geographical data and urban models, overlapping with many of the concepts and conclusions outlined in Section 3.
The combination of geographical and building data is most commonly required in large infrastructure projects. A number of proposals and pilot applications can be found. These use cases tend to rely primarily on their own domain ontologies, as is also indicated in Le and Jeong [21]. In terms of urban infrastructure products, El-Gohary and El-Diraby [93] developed a domain ontology describing the multi-stakeholder project development process to support knowledge-enabled process management and coordination across various stakeholders, disciplines, and projects for urban infrastructure. Similarly, El-Diraby and Osman [174] developed a domain ontology for construction concepts in urban infrastructure products to provide a conceptualization for knowledge in civil infrastructure. Also Montenegro et al. [175] developed an LBCS OWL ontology based on the Land Based Classification Standards (LBCS). This ontology is proposed as a basic structure to allow city information modelling (CIM).

Ideally, however, infrastructure projects that include GIS and BIM data also rely on the data models and ontologies provided in those two domains. To achieve this, it has been attempted within BuildingSMART to extend the available EXPRESS schemas with infrastructural schema extensions (leaving out GIS schema extensions out of scope for now). As a result, the Infrastructure Room of BuildingSMART is developing an IFC Bridge, IFC Road, and IFC Alignment extension in the EXPRESS information modeling language. However, as already indicated in Section 3.1, the usage of EXPRESS would again limit extensibility and adaptability, not to mention the difficulties in binding to the diverse available commercial software solutions. An alternative approach was therefore proposed by Beetz et al. [176] towards enriching existing BIM systems with GIS and infrastructure data. This is demonstrated for quay walls in Rotterdam [176]. This approach is closely associated to the way in which a semantic web version of the bSDD is proposed to be used (see Section 4.2). Namely, existing BIM environments remain to be used for modeling building information, but schema extensions or relations to information outside the schema, like bSDD URIs, Infra domain ontologies, quay wall ontologies, GIS ontologies, are made by incorporating a property URI and a value or object URI in the name and value of an IfcPropertySingleValue instance. If properly published, these URIs then provide access to formally represented information that is made available outside of the BIM environment (material data, bSDD data, Infra data, GIS data). A schema of this approach is provided in Fig. 10, with the legacy IFC model on the left and the external RDF graphs on the right.

Another approach towards the combination of BIM, GIS, and Infra data is provided in the COINS project [177]. In this approach, diverse legacy files can be combined in one container, allowing the continued usage of standard BIM, CAD and GIS systems. An additional OWL file in the container then describes how the diverse elements in all legacy files are related or linked to each other [178]. If all data is available in RDF graphs, however, making and managing links between such files would likely become a lot more convenient (see slide 26 in Ref. [179]).

5. Aim 3: logical inference and proofs

A third and last topic that is often used in arguing for the adoption of semantic web technologies in the architectural design and construction industry, relates to the logical foundations available in semantic web technologies. Using generic inference engines, extra information can be inferred from the information in RDF and OWL through simple DL principles. Moreover, it is possible to represent IF-THEN rules, for example using SWRL [141], thus allowing reasoning within FOL. When chaining these rules and combining them with original building data and a reasoning engine, a considerably powerful inference process can be realized, an approach which is schematically represented in Fig. 11.

In principle, the formal logical basis of semantic web languages allows to do more than simple inference. This basis also allows to automatically generate the proofs for what is inferred in a reasoning process. These proofs can be used by semantic web applications.
OWL ontologies and completeness’ [181]. This is clearly of high relevance in checking is most commonly interpreted as checking for consistency and completeness once, tends to differ among implementation plans and backgrounds. Checking is understood, however, to build trust around their results (see Section 2.1). At this moment, most semantic web applications are limited to the usage of RDF(S), OWL, and rules, especially in the architectural design and construction industry.

The inference process with semantic web technologies can take a quite diverse number of forms. A full technical overview of those forms is clearly out of scope here, but initial outlines are given elsewhere [122,180]. Moreover, Pauwels et al. [122] provide a quantitative assessment of some available rule-checking procedures in theform of a performance benchmark in the AEC industries. Pauwels and Zhang [180] distinguish between hard-coded rule checking after querying for information (SPARQL SELECT), rule checking by querying (SPARQL CONSTRUCT), and semantic rule checking with dedicated rule languages (SWRL).

5.1. Regulation compliance checking

The logical basis supplied by semantic web technologies is most commonly relevant to rule checking use cases such as regulation compliance checking (see Section 4.4). The way in which the inference process and the actual ‘rule checking’ is understood, however, tends to differ among implementation plans and backgrounds. ‘Checking’ is most commonly interpreted as ‘checking for consistency and completeness’ [181]. This is clearly of high relevance in combination with MVD information exchange: one needs to check one needs to check whether the resulting model views actually also contain the required information [182]. However, the kind of checking required for MVD typically benefits from a CWA. In an OWA, the fact that information is not there cannot lead to a fail message (see Section 2.3). This is a crucial context that needs to be taken into account in any discussion in the AEC domains on the topic of rule checking. Although it is possible to perform a CWA-based inference process using semantic web technologies, it is not the main virtue of this set of technologies. Instead, the inference process should mainly be used to assert additionally derived information, both in the case of OWL reasoning and reasoning with dedicated rule languages (see Section 2.1).

If limiting to OWA-based regulation compliance checking, one of the earliest semantic approaches in this domain relies extensively on ontologies and SPARQL rather than considering dedicated semantic rule languages [see also Ref. [180]]. Namely, the approach presented by Yurchyshyna et al. [183] and Bouzidi et al. [184] relies entirely on SPARQL SELECT and CONSTRUCT queries for regulation compliance checking. This approach of querying a model or ontology with a query language is still regularly used for regulation compliance checking. A similar proposal has been proposed by Dimyadi et al. [185] for regulation-compliance audits in general. A second example of an OWA-based reasoning process for regulation compliance checking is the effort by Pauwels et al. [158], which aims at accommodating acoustic regulation compliance checking for BIM models using N3Logic rules.

A more recent and probably also one of the most elaborate approaches is proposed by Beach et al. [186]. With the semantic approach for automated regulatory compliance checking in the construction sector, Beach et al. [186] primarily aim to allow domain experts to formally represent and maintain their own rules in a rule-checking system (as opposed to the closed rule sets in proprietary platforms). In addition, an increased understanding would be obtained of what the regulatory compliance system is actually checking [186]. The schema of the proposed system and how it can be used for regulation compliance checking is repeated in Fig. 12. The system uses five ontologies: an abstract regulation ontology, a core domain ontology, a data format ontology, a regulation ontology, and a regulation mapping ontology. These ontologies allow to semantically represent building data on the one hand and regulation data on the other hand, both using their own terms and vocabularies. Indeed, the vocabulary used in a regulation is typically quite different (see also Refs. [180,187,188]) from the vocabulary used in a BIM model (core ontology), let alone in one of the data models used to capture that building model (data format ontology, like ifcOWL).

The rules in the regulation ontology of Beach et al. [186] are built semi-automatically from the regulation texts by first annotating the texts using the RASE approach [160,161] and then parsing and converting from these RASE annotation tags. Other approaches proposed to use natural language processing (NLP) techniques for this step [162–164]. A mapping ontology then allows to link the regulation and domain ontologies, making both compatible. In the case of the BREEAM regulations studied in Beach et al. [186], it resulted in 180 procedures to map 854 data items in the IFC ontology to the BREEAM ontology. In conclusion on this topic, the linking between the diverse domains is being done here using programming code. Alternative approaches have already been discussed, including a mapping through logical rules [158], a pure engineering approach [147,148], and a linkset engineering approach [69]. Whichever approach is taken, it must be clear that the key research challenge in this domain is again linking one ontology (domain ontology) with another ontology (regulation ontology), which leads to the same challenges as in Section 3.

5.2. Interoperability and model handling

Obviously, rules and reasoning engines are not only relevant for regulation compliance checking. As an RBox, rules form a very important third element next to the TBox (ontology) and ABox (instances). This has become clear from the number of times that rules and rule engines were mentioned in the last two sections (Sections 3 and 4). The examples given regarding the on-demand application of rules for query rewriting [77], subset selection and subset publication [62,67,73] (see end of Section 3.3.2), including the FOWL architecture [76], are excellent examples of how the logical basis provided via semantic web technologies can be used at its maximum. The examples used in de Farias et al. [73] particularly propose to simplify ifcRelationship instances following the ifcOWL ontology, and to simplify the representation of external versus internal walls. Automatically generating these parallel representations allows end users to make far simpler and more intuitive queries.

The example rule-checking implementation that is documented in Lee et al. [189] might be considered similar to these last proposals, in the sense that it also targets the inference of information in one representation (as required for cost estimation) based entirely on information in another representation (IFC), which relates to the same interoperability challenge. Lee et al. [189] propose two small OWL ontologies, a work item ontology and a work condition ontology, which are engineered with the purpose of helping in the building cost estimation process. The authors propose to extract
information from an ifcXML file and parse this information as RDF instances of the work item and work condition ontologies. Using these RDF instances and the OWL ontologies, an OWL reasoning engine (in this case the RETE-based Bossam reasoner) is able to infer additional properties and class memberships. A user interface then allows a user to query a SPARQL endpoint holding the resulting graph. The query results can be used more intuitively in a cost estimation application. This example relies not on explicit IF–THEN rules but rather on a combination of SPARQL queries and OWL class expressions.

5.3. Inference processes within regular use cases

As indicated, inference processes regularly take place within ‘common’ use cases as well. The cases that we consider here overlap with a considerable number of the cases outlined in Section 4. In these cases, the choice for a semantic web-based inference process is most commonly motivated by the desire to implement a part of an application in a declarative rather than a procedural way. As soon as data is available, part of the conclusions can be generated by a generic inference engine in terms of logical declarative assertions, thus extending the data set that is available at the source. Procedural programming can then be limited more to straight-forward implementation tasks (like a GUI and finding datasets), and to those kinds of assertions that can be expressed in a formal language (like the 20–30% of regulation text that cannot be formalized).

An example showing the way in which rules can be deployed for construction industry and building information management is provided by Kadolsky et al. [143] and Baumgärtel et al. [144]. In this example, the authors propose to represent a building in an ifcOWL ontology, after which rules can be used to retrieve information that is relevant for building energy performance. Baumgärtel et al. [144] specifically shows how rules can be used to allow a thermal insulation check: the right-hand side of one of the SWRL rules includes the statement ?summ eeBIM:definition "Thermal insulation check failed". For resolving inconsistency issues during design coordination, Kim and Grobler [190] describe how potential conflicts can be detected through an ontological consistency checking system, after which a repair or modification of the inconsistencies can be achieved.

In the area of Health and Safety (HS) measures, the “Job Hazard Analysis” (JHA) application proposed by Zhang et al. [155,191] provides another excellent example of a semantic rule-checking process. The authors propose to combine an RDF representation of the building model, contained in Tekla Structures, with a number of ontologies and SWRL rules that allow the analysis of the construction project in terms of jobs, tasks, safety procedures, and the resources that are required to allow the safe execution of these job steps. Abanda et al. [192] relies heavily on SWRL rules to compute labor cost in construction projects. In the use cases documented by Abanda et al. [192], a lot of manual modelling occurred to allow the combination of RBox (SWRL rules) with TBox (OWL ontologies) and ABox (RDF instances), dealing mostly with challenges in merging semantically close concepts like labor productivity and labor efficiency. Furthermore, a proposal is made in Benevolenskiy et al. [193] to rely on formal ontologies and rules to model construction processes, so that they can be better organized and managed, which is much in line with the proposal made by Zhang et al. [155,191]. This work on process modelling relies heavily on earlier work in ontology engineering for business process objects (BPOs) [194,195]. Other use cases that often include reasoning engines and semantic web services can be found in the home automation domain. The ThinkHome system mentioned in Section 4 is one such an example [127,128], as it relies on agents that interpret information from home sensors and devices and then react in terms of home automation.

6. Conclusion

6.1. Application areas for semantic web technologies

In this article, we have identified three key application areas of semantic web technologies and articulated the opportunities and challenges of their current practices. A large number of relevant papers has been taken into account to make a contribution along
these three axes. An overview of the listed papers is given in Tables 2 to 4, indicating which use cases they respond to, what advantages semantic web technologies could offer for them, and what the current status of research is.

1. Interoperability: Semantic web technologies are often associated with the long-standing interoperability challenge in the architectural design and construction industry. Even though the term interoperability entails diverse connotations and interpretations, this paper pinpointed that no solid proposal exists so far for fully resolving interoperability issues in these disciplines using semantic web technologies. At least, there is no proposal that solves it any better than existing approaches. In that regard, most domain professionals currently aim first at providing semantic data exchange rather than full interoperability. Such an approach leaves room for implementers and engineers to interpret and restructure incoming and outgoing information, thus embracing interpretations, heterogeneous standards, and versatile adaptation. Semantic web technologies have capabilities to extend and add functions, features and possibilities towards an improved information exchange, as the usage of the RDF data model implies a common data model across islands of information. But, in order to reach interoperability, which is more associated with consensus and agreements on conceptual data models and data exchanges among stakeholders, it is more important that domain professionals make agreements about how their conceptual data models are related. This should ideally start in the earliest phase of data exchange specifications, which is in fact the purpose of the IDM and MVD specifications processes. An IDM allows to define and maintain distinct data exchanges and their specifications, hence this is the best medium to employ the technology for generating formalized domain semantics and offering a robust base for integrating diverse disciplines.

2. Linking across domains: As was indicated already in Section 2.4 (linked data vs. semantic web), there is a difference between semantic web and linked data applications. Linked data was hereby explained as a response to the finding that quite some data was being published on the web, seemingly following the semantic web idea but actually never linking to outside data [40], and thus in fact not realizing the initial core idea behind the semantic web at all, which is linking data. When taking a linked data approach, one uses only a subset of the stack of available semantic web technologies. For many researchers and developers, linked data is thus the ‘fast track’ forward. Not surprisingly, applications and use cases that focus entirely on linking data are also the most rapidly emerging applications and use cases. Because of the limited number of involved technologies and the simpler approaches, they are easier to conceive and implement. Within the construction industry, the highest number of valuable use cases lies in this axis. Undoubtedly, this will be a fruitful area for future research initiatives and application development initiatives.

3. Logical inference and proofs: Applications and use cases that focus on the logical basis of OWL and SWRL are less common than the use cases focusing on linking across domains. In fact, the exact opposite situation can be perceived here. OWL and SWRL, but also inference engines and proof engines, are technologies that are not commonly present in linked data-inspired applications and use cases. Since these technologies are the upper parts of the semantic web technology stack, it takes considerably more effort to implement and use them effectively in practical use cases. Yet, for an important research area in the construction industry as rule-checking and regulation-compliance checking (Section 5.1), it is totally worthwhile to make this effort. Even more so, assuming that linked data applications mature and become more common, also other logic-based applications as documented in Section 5.3 will undoubtedly become more mainstream.

6.2. Key research challenges

By making a qualitative assessment of the diverse proposals made in these domains, we have also been able to retrieve where the

<table>
<thead>
<tr>
<th>Use cases</th>
<th>Semantic web applicability and advantages</th>
<th>Research status and challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable vendor-neutral model exchange [16-21,50]</td>
<td>• Provides a single data model for representing any kind of information</td>
<td>• Human intervention and manual work required for ontology alignment and interface</td>
</tr>
<tr>
<td>Combine different information representations [19,62,68,69,71-74]</td>
<td>• Allows adding a logical basis to this representation using OWL and rule languages</td>
<td></td>
</tr>
<tr>
<td>Support use case based information exchange [20,60,62,67,73,76,77]</td>
<td>• Focuses intensively on linking diverse graphs of information together in a web-like fashion</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Allow to combine different representations of information</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Exchange partial models (requirement model, architectural model, MEP model)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Use cases</th>
<th>Semantic web applicability and advantages</th>
<th>Research status and challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ontology-based information management and sharing [11,80-99]</td>
<td>• Support collaboration through a common platform</td>
<td></td>
</tr>
<tr>
<td>Combine product manufacturer data with building data [99-111]</td>
<td>• Enable network-based publication of concept libraries</td>
<td></td>
</tr>
<tr>
<td>Support building performance analysis and optimization [50,112-131,134-148]</td>
<td>• Description logic basis</td>
<td></td>
</tr>
<tr>
<td>Connect BIM and GIS [21,165-179]</td>
<td>• How to keep parallel representations of the same thing consistent and complete [20,71]</td>
<td></td>
</tr>
<tr>
<td>Enable automated regulation compliance checking [149-158,160-164]</td>
<td>• How to keep the links between partial models effectively manageable [70]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Links between concepts and concept libraries need to be created and managed over time</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• At least 20% to 30% of regulation codes are too vague or qualitative, and therefore requires human interpretation in the regulation compliance checking process</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Conversion from a natural language regulation text to a formal representation (or reverse) remains a critical research issue</td>
<td></td>
</tr>
</tbody>
</table>

Please cite this article as: P. Pauwels et al., Semantic web technologies in AEC industry: A literature overview, Automation in Construction (2016), http://dx.doi.org/10.1016/j.autcon.2016.10.003
key research challenges lie in the usage of semantic web technologies. These research challenges are very closely related to practical engineering challenges in construction industry.

1. How to define and manage semantic links? In all three application axes outlined above, the most prevalent problem revealed throughout the literature review, is the open question of how to create and manage the links between distinct submodels, partial models, rule sets, and so forth. This critical question typically requires human input (cfr. blue circular interface points in Fig. 5). This is considered to be a key reason why semantic web technologies will likely not solve the interoperability issue, but will at best improve information exchange. Semantic web technologies provide a semantic or formal better grounding to information, but agreements between ontologies and schemas still need to be made.

2. Where is the optimal balance between procedural and declarative implementation efforts? Less prevalent in the considered articles, but closely associated with the question on how to create and manage links, is the challenge of finding a balance between the diverse semantic web and other technologies that are available to effectively build an application in support of a practical use case in the construction industry. This is a question that drives many researchers and implementers in this area, but that is not so often elicited in the articles. Some of the considered articles propose building a central ontology and build everything around that; others aim to manage data in a fully decentralized manner, almost entirely using semantic web technologies (ABox, TBox, RBox); others make an effective combination of semantic web technologies in the data layer, but rather quickly switch to procedural programming languages as soon as integration of data is required; others rely very intensively on semantic web services in a service-oriented architecture (SOA); and yet others focus on minimal use cases in an otherwise traditional software environment. It is impossible to make out in this article what the best approach is. The best answer is likely that a good evaluation and assessment needs to be done of the targeted use case and application, before the actual implementation is being done, and a solid conceptual use case and implementation plan needs to be conceived that takes the best of the available technologies, which is a pure software engineering challenge.

3. How to effectively bring in input by a human user? A last research challenge that can be pinpointed is the need to build a good knowledge base while allowing the user to provide additional input. It is clear from the considered applications that all data eventually comes from an end user. Ontologies can be conceived to grasp and structure data so that computers can interpret it. However, there will always be a portion of the user data that cannot be represented within the existing ontology (or even within the data model). This data needs to be given a place as well in the targeted software applications. This is especially true for regulation compliance checking. Typically, only about a portion of a natural language regulation text (70–80% estimated) can be captured in a formal rule set, implying that the user needs to be able to manually input also that other 20% to 30%. But also regular ontology-based applications should provide an appropriate place for an end user to provide additional information that is not easily contained in an ontology.

To conclude, one of the harder future research directions likely focuses on the realization of an environment for efficient information and data management and exchange in the AEC domains. Such a research direction should focus heavily on the combination of the existing MVD and IDM approaches and methods in combination with semantic web technologies. Currently, diverse industrial partners and academic institutes have invested and participated in developing data exchange specifications for BIM modeling. However, since no approach for defining semantics and sharing requirements exists, exchange specifications have been separately defined and executed in different ways, which results in a lack of consistency. Thus, as an effort to formalize exchange specifications and organize them in a well-structured classification, semantic web technologies could be employed for defining and sharing IDM and MVD requirements.

Acknowledgements

This research was made possible by the funding support of the Special Research Fund (BOF) of Ghent University.

References


Please cite this article as: P. Pauwels et al., Semantic web technologies in AEC industry: A literature overview, Automation in Construction (2016), http://dx.doi.org/10.1016/j.autcon.2016.10.003.
Please cite this article as: P. Pauwels et al., Semantic web technologies in AEC industry: A literature overview, Automation in Construction (2016), http://dx.doi.org/10.1016/j.autcon.2016.10.003.


