The Role of the Semantic Web for Knowledge Management in the Construction Industry

Igor Svetel
Innovation Center, Faculty of Mechanical Engineering
Belgrade, Serbia
E-mail: isvetel@mas.bg.ac.rs

Milica Pejanović
Faculty of Architecture, University of Belgrade
Belgrade, Serbia
E-mail: pmilica@arh.bg.ac.rs

Keywords: semantic web, building information modelling, BIM, industry foundation classes, IFC, knowledge management

Received: October 31, 2009

The architecture, engineering and construction (AEC) industry is knowledge intensive field. Significant heterogeneity of the forms of knowledge mobilized in the construction industry prevented adoption of IT based knowledge management in the field. Recently, a large international initiative is launched to provide extensive IT support that will enable model-based interoperability among all professions in the AEC industry. Resulting standards coupled with Semantic Web technologies have potential to serve as the foundation for the knowledge management in the AEC field. The paper gives an overview of the both technologies and depicts ways in which they can provide knowledge management support for the AEC industry.

Povzetek: Predstavljena je vloga semantičnega spleta pri upravljanju znanja v industriji.

1 Introduction

The architecture, engineering and construction (AEC) industry faces with continued social pressure to improve quality, responsiveness, reliability and efficiency in its business. Time delays, unforeseen work and costs, and resulting lawsuits among participants become routine events and result in exaggerated prices of buildings. Worldwide reports give an account that approximately 5% of total annual turnover in the AEC industry is lost because of the inadequate interoperability among participants in the industry and a lack of standardization in technology adoption among stakeholders [1] [2] .

Recently, a large international initiative is launched to provide extensive Information Technology (IT) support that will enable model-based interoperability among all professions in the AEC industry. Open standards developed as the part of the initiative have an extended potential to serve as the foundation for the knowledge management in the AEC field, especially in connection with Semantic Web technologies.

The paper gives an overview of the standards developed to provide interoperability in the AEC industry and standards developed under Semantic Web initiative. It also depicts ways in which both can be combined to provide knowledge management support for the AEC industry.

2 Knowledge in AEC

The AEC industry is knowledge intensive field. It encompasses heterogeneous expertise from multiple fields and diverse occupational groups. The knowledge ranges from tacit knowledge of architects on the ways of combining spaces to accommodate various social needs to the practical knowledge of builders on the use of tools and materials to construct physical spaces.

Research on the knowledge management in the AEC industry identified many specific features of the process [3] . The process of aligning the symbolic and the material is essential feature of the knowledge in the AEC industry. Up to the phase of the physical assembly of the building, the whole process is based on the symbolic representations in the form of plans, calculations and descriptive texts. Still, in minds of the participants those representations are constantly linked to the real physical entities. This feature was main reason for the rejection of IT systems in AEC community because those systems supported only symbolic processing.

Another significant feature is the importance of talk and communication in the process of knowledge sharing. This aspect is often ignored and consequently IT solutions support only formal representations. However, observations and field research demonstrate that formal representations serve as the stage around which series of informal verbal discussions occur that are essential for the enlargement of the knowledge about design [4] .

The importance of external influences is also a feature that lacks support in traditional IT systems for the AEC. The large amount of the knowledge that constitutes one AEC project is produced elsewhere and is continuously attended and integrated into the existing procedures and activities. Many new products that constitute finished building are permanently produced, as well as procedures and regulations that account for their proper use. AEC project is an ongoing effort at aligning a variety of heterogeneous resources and practices.

3 Open standards for AEC

Based on years of research on a general data model for AEC [5] [6] the term Building Information Modelling (BIM) denotes a process of using IT to model and manage data encompassing the whole facility lifecycle [7] . The BIM concept means to build a facility virtually, prior to building it physically, in order to work out problems and simulate and analyse potential impacts. It is easier to fix a problem by moving element with a mouse than to demolish and rebuild elements on a construction site. The commercially developed BIM applications support creation of the computer-based facility model using parametric three-dimensional (3D) components with attached descriptive parameters that are necessary to identify particular elements. Still, those applications typically use proprietary data formats to represent facility models thus keeping all information locked in distinct software.

The need to establish interoperability among applications dealing with different phases of the facility lifecycle, such as architectural design, civil engineering, HVAC design, building construction, and facility management (FM), was met with the development of the Industry Foundation Classes (IFC) standard [8] . The currently available model is Version 2, Revision 3 and is registered as the ISO/PAS 16739:2005 standard. IFC is a neutral and open model whose development is conducted by the International Alliance for Interoperability (AIA), which has 550 member organizations in 24 countries. The standard provides the following basic functionality:

- Data interchange without information loss among all AEC and facility management (FM) applications.
- Unified model-based description of all building components.
- Information on the graphical representation of components.
- Description of relationships with other components and their location in the whole structure.
- Link to property and classification data and access to external libraries.

The open specification of the IFC data model allows commercial software developers to write interfaces for their software that enable exchange and sharing of the same data in the same format with other software applications, regardless of the internal data structure of any individual software application. All leading software companies like Autodesk, Bentley System, Graphisoft, Nemetschek, Data Design System, Solibri, Tekla, Archimen Group, Vector–Works, etc. support IFC in their applications.

Being an object-oriented data model, the IFC standard is comprised of class definitions representing all things and events occurring in the facility lifecycle. At the top of the hierarchy is a domain layer that describes classes related to basic functional units: building controls, plumbing and fire protection, structural elements, structural analysis, heating, ventilation and air conditioning, electrical circuits, architecture, construction and facilities management. Below that layer rests the interoperability layer that defines all classes essential for connection and cooperation among disciplines. Next is the core layer, containing basic model classes depicting controls, products, and processes. The resource layer is at the bottom, embodying classes that represent all building elements. Elements encompass not only physical components, as traditional models, but also actors and their roles, time, price, approval, etc.

The IFC standard does not produce one monolithic data model encompassing the whole lifecycle. Instead, many separate models are generated. In the context of IFC, a View is a defined as a subset of the IFC Object Model that a number of implementers have agreed to support in their implementations. The software certification process is conducted according to IFC Views. Depending on agreement, many IFC Views can exist with partially overlapping content or with entirely different contents. The data exchange between applications should occur within the scope of a specific View. The entire facility lifecycle is represented across multiple Views.

The IFC standard relates to the representation of a particular instance of the facility, its components, properties, and relationships. Using the vocabulary of object-oriented modelling, it can be said that it deals with object instances. It does not allow representation of the object classes and their relationships (i.e., that part is covered by the International Framework for Dictionaries (IFD), registered as the ISO 12006-3:2007 standard [9]).

IFD is the classification system for all information in the AEC/FM field. It is an object-oriented framework that defines objects, collections and their relationships. It is intended to work as the overarching structure that will provide support for the development of the unified AEC/FM vocabularies at the national, regional or domain levels. Since all share the same structure it will be possible to translate terms between languages and domains, preferably using automated software agents. IFD identifies each object in the model and this provides the capability to define context within which a concept is going to be used. Each object can have multiple names providing for the definition of synonyms or usage in different languages. An object is related to a formal classification system using references. The standard supports the following types of objects: Subjects, Activities, Actors, Units, Measures with Units, and Properties. Relationships are divided into Association, Collection, Specialization, Composition, Involvement, Property Assignment, Sequence and Measure

Assignment. Employing these mechanisms, the user can create a model-based definition of all concepts in AEC/FM including facts about classification systems, information models, object models, and process models. In other words, IFD functions as the IFC metamodel. In addition, IFD provides a unique global reference for any AEC/FM concept. The mechanism that relates IFC and IFD standards is scheduled to be published in the IFC 2x4 standard revision. The actual realization of IFD is the IFD Library, an international initiative currently run by four nations: Canada, Netherlands, Norway and USA. The purpose of the library is to provide semantic knowledge to the construction industry in a global and uniform way.

In addition to the above-described standards, a second type of interoperability formats has been developed based on another open standard - eXtensible Markup Language (XML). XML is a general-purpose specification capable of describing published data. The data description mechanism is based on the insertion of tags in the traditional text and the user can choose any term to define a particular tag. The language permits representation of arbitrary data arranged as an hierarchical tree with one element serving as the tree root [10] XML enables the structured representation of any kind of information but does not provide any mechanism to infer the meaning of the terms used in tags. One approach to the definition of a tag's meaning is the XML schema. It is a language that provides a description of a type of XML document, usually articulated in terms of constraints on the structure and content of related XML documents. Many schemas have been developed for the AEC/FM field. The gbXML (Green Building XML) schema is used for describing data relating to the building energy efficiency of the facility and its impact on the environment. The aecXML schema is used for depicting all building data in design, engineering and construction disciplines, and the CityGML schema is used for geo-spatial data representation. In addition, IFC data can be represented with the ifcXML schema. Since the IFD is an EXPRESS model, the EXPRESS to XML Schema Converter [11] can be used to obtain the XML schema for IFD.

Open standards developed for the AEC/FM industry relate to the problem of interoperability, since this is the most obvious obstacle in the industry. These standards enable the highly structured representation of information about buildings but do not consider the problem of information reuse outside of the context of a particular facility lifecycle or the automatic creation of new information for later reuse. Since all described standards have suitable XML schema representations, extension of the knowledge management capabilities can be achieved with technologies developed under the Semantic Web initiative.

4 Semantic Web, ontology and knowledge management

The recent advent of the Semantic Web has given a new impulse to the old knowledge management research field. The goal of the endeavour is to build a unified information medium that is both understandable for people and computers and that can be used for the automatic deduction of meaningful inferences [12] . To function effectively, the Semantic Web should be built on structured collections of information and sets of inference rules. Research on knowledge representation conducted as the part of the long time effort to build artificial intelligence systems has already developed many useful technologies. Unfortunately, those systems are centralized, relying on limited sets of rules to describe narrowly defined domains making the reuse of rules in new domains impractical. Similarly to the hypertext concept, when existing knowledge representation concepts were coupled with the global information system, the full potential of the technology was realized and this spurred a new wave of interest in the knowledge management field.

The new attempt to create universal knowledge representations is based on the layered structure of representation standards. The upper layer exploits functionality of lower layers and provides greater semantic expressiveness. At the bottom level resides XML. Meaning is expressed in the next layer containing the Resource Description Framework (RDF), a data model for representing information about entities in the Web [13] . In the Semantic Web standards, an entity or thing is referred as the resource. RDF achieves its functionality by using triples, a structure consisting of subject, predicate and object. This formation states that a particular thing (subject) has a property (predicate) with a particular value (object). The Universal Resource Identifier (URI) identifies subject and predicate and their value is either URI or literal. URIs ensure that concepts are not just bare terms devised by someone, but are connected to unique definitions on the Web. When multiple triples point to the same resource, they start to form a network of information about related things. That way information that defines a single entity is not held in one place but is spread over the Web forming a distributed web of data. However, RDF has no mechanism for determining that two or more dissimilar terms point to the same concept.

The next level of the semantic expressiveness is achieved with ontology. In the Semantic Web domain, ontology is identified as the formal representation that defines relationships among terms. The first level of ontological functionality is achieved with the RDF Schema (RDFS) [14] . Like other schema languages, RDFS provides information about basic RDF structures. It accomplishes this task by supplying constructs that allow the declaration of classes, subclasses, property, and subproperty relationships among resources. Constructs domain and range describe the relationship between properties and classes. These definitions are expressed using RDF triples. RDFS provides a limited set of inference rules that are restricted to the transitive closure of the hierarchies.

The Web Ontology Language (OWL) currently provides the highest level of ontological functionality among Semantic Web technologies. It is a family of languages based on two semantics. OWL Lite and OWL DL are based on Description Logic semantics that guarantee completeness of reasoning. OWL Lite is a restricted version of OWL DL and is intended as a quick migration path for thesauri and other taxonomies. OWL Full provides maximum expressiveness and the syntactic freedom of RDF, but does not support complete or efficient reasoning. The language provides constructs like class, property, property restrictions, Boolean combinations, enumerations and instances. A wide range of services like reasoners and editing tools enable users to express and test knowledge using this formalism leading to the widespread acceptance of this technology [15] .

The level of expressiveness and functionality realized in the Semantic Web development surpasses previous attempts to model computer-based knowledge management systems. It is the idea of an open community, essential to the notion of the Web, which attracts so many people to the field and generates so many results. Anyone can use open standards to develop personal systems, use open source software to express his/her knowledge, or can engage in the development of standards. This openness of the development process has resulted in the remarkable range and richness of topics covered by the Semantic Web. Moreover, the organic development that grows from the interests and energy of the supporting community persuades an increasing number of researchers to adopt both Web standards and the open development principle as the foundations for development in their professional domains.

5 Semantic Web based knowledge management in AEC

Even if the IFC standard was not developed with the knowledge management in mind, it contains many supporting features. First, it represents a standardised object-oriented model of the building around which all professions in the AEC industry can focus their collaboration. In addition, since the globally unique identifier (GUID) identifies every object in the model, the IFC model provides unique definition of the IFC objects on the Web. If the GUID is connected with the physical building element by attaching an RFID tag to the element, it provides unique identification of the elements in the real world. That way a "symbol grounding problem" [16] that arises when a meaning is assigned to a symbol system can be solved to the level of real physical entities in the AEC field.

Multi-disciplinary professionals are involved in AEC projects, with various viewpoints, goals, priorities, and backgrounds. Diverse terms are employed to represent similar concepts or a single term for different concepts. By providing a unique global reference for any concept in the AEC field, the IFD supports communication among participants. If any dispute arises around denotation of some term, the IFD serves as the central authority. In addition, it provides mapping mechanism between usage of terms in different AEC professions and occupations enabling participants to understand model elements from their point of view.

Still, full knowledge management support can be obtained only by extending these standards with Semantic Web technologies. The IFC can describe only particular object instances, and IFD provides domain ontology, the structure of concepts and the relationships between them. They lack power to express classes, aggregations and rules.

The Semantic Web technologies provide all functionality necessary to support importance of external influences in AEC knowledge management.

The search for information on required materials and building elements represents around 18% of time spent on the building's design [17] . The Semantic Web can reduce this time by automatically acquiring links to relevant resources on the Web providing information about needed products that matches projects requirements. This process will use both Semantic Web technologies to acquire meaning of the information published on the web, and IFD to connect that information to the terms used in the IFC model. The automatic discovery and invocation of building product information additionally improves design process by enabling retrieval of cost and performance criterions, regulatory standards or component availability or delivery schedule that enables designer to give more precise predictions about building performances and construction schedules to his clients.

The scale of the waste from reinvention in design firms is also around 18% of time [17] . The Semantic Web technologies coupled with IFC and IFD standards provide a mechanism that will keep all information on previous projects in the form that will enable automatic retrieval of required information in future projects. The information on the project will serve not only particular firm, but also if published on the Web it will become a global reference available to all interested parties.

So far, few prototype systems that link Semantic Web and IFC standard are developed. The easier technique to extend existing AEC standard formats and enable knowledge management functionality is to add semantic annotations using RDF. The method is demonstrated in the system for conformity checking in construction [18] . The norms are extracted from the electronic regulations and formalized as SPARQL queries in terms of the IFC model. The conformity checking process is based on matching an RDF representation of a project to a SPARQL conformity query. The project's RDF representation is extracted from the ifcXML schema and later manually enriched with domain knowledge.

More projects are using OWL to add knowledge management functionality. One notable example is the Sydney Opera House facility management model [19] . The basic IFC model is transformed using an IFC-OWL converter [20] Existing tools are used to manually enrich this OWL representation with ontology and rules. The OWL representation is associated with the IFC model enabling the publication of performance by selecting spaces in the 3D model.

The InteliGrid project [21] also uses a custom developed IFC-OWL transformer [22] to enable the representation of expertise in addition to the basic representation of building elements and services.

The SWOP project [23] uses a custom developed Product Modelling Ontology (PMO) that is sufficiently rich to represent product ontology for any parametric product type. PMO is a layer on top of the RDF, RDFS and OWL hierarchy and models the product from the ontology point of view, meaning that an IFC model or any other product model can be obtained automatically as the result of the modelling process.

Today many open source converters for Semantic Web formats can be found. XSD2OWL provides transformation of an XML Schema into an OWL ontology and XML2RDF enables transformation of XML into RDF [24] . These tools together with the availability of many open source editors and development environments that support Semantic Web standards provide an opportunity to add meaning and enrich open AEC standards.

6 Conclusion

Integration of knowledge management capabilities with BIM methodology has great potential, especially since the current open standards approach shares the same technological background as the recent Semantic Web initiative. The openness of both activities motivates many researchers and software developers to join this initiative and make their contributions. This process guaranties the best adaptation of the technologies to the users' needs and their widest support and endorsement. However, the approach also has its drawbacks. The number of technologies and their variants is massive. Moreover, the pace of change is so rapid that systems based on today's most recent technology become obsolete over a two to three-year period, requiring developers to update constantly their products to take advantage of the latest version of the applied technology.

Acknowledgement

The Ministry of Science and Technological Development - Republic of Serbia supported this work under grant TP-16025. It is a part of the project 'Pilot Project: Application of the IFC (ISO/PAS 16739:2005) Standard in Drywall Systems.' The project director is Dr Igor Svetel.

References

[1] Ercoskun, K., and Kanoglu A. (2003). Bridging The Gap Between Design and Use Processes: Sector-Based Problems of a CRM Oriented Approach. *9th EuropIA International Conference: EIA9: E-Activities and Intelligent Support in Design and the Built Environment*, Istanbul, Turkey, pp. 25-30.

[2] Gallaher, M.P., et al. (2004). *Cost Analysis of Inadequate Interoperability in the U.S. Capital Facilities Industry.* NIST GCR 04-867 Report.

[3] Styhre A. (2009). *Managing Knowledge in the Construction Industry*. Taylor & Francis, Oxon.

[4] Lawson B. (2004). *What Designers Know*. Architectural Press, Oxford.

[5] Eastman, C.M., Bond, A. and S. Chase (1991). A Data Model for Engineering Design Databases. *Artificial Intelligence in Design '91.* (ed. J.S. Gero), Butterworth-Heinemann, pp. 339-365.

[6] Björk, B.-C. (1989). Basic structure of a proposed building product model. *Computer-Aided Design.* vol. 21, no. 2, pp. 71-78.

[7] Lee, G., Sacks, R., and C.M., Eastman (2006); Specifying parametric building object behavior (BOB) for a building information modeling system; *Automation in Construction*, vol. 15, no. 6, pp. 758-776.

[8] Liebich, T. et al. (eds.) (2007). *IFC2x Edition 3 TC1.* International Alliance for Interoperability. Available from http://www.iai-tech.org/ifc/IFC2x3/TC1/html/index.htm

[9] IFD (2008). *IFD Library White Paper.* Available from http://www.ifd-library.org/images/IFD_Library_White_Paper_2008-04-10_I_.pdf

[10] Harold, E.R., Means, W.S. (2004). *XML in a Nutshell.* Third Edition, O'Reilly Media, Inc., Sebastopol, CA.

[11] EXC (2002). *EXPRESS to XML Schema Converter.* Available from http://cic.vtt.fi/projects/ifcsvr/index_exc.html (May 2009)

[12] Berners-Lee, T., Hendler, J., and Lassila O. (2001). The Semantic Web. *Scientific American,* Vol. 284, No. 5, pp. 34-43.

[13] Klyne, G. and J.J., Carroll (eds.) (2004). *Resource Description Framework (RDF): Concepts and Abstract Syntax.* W3C Recommendation 10 February 2004, Available from http://www.w3.org/TR/rdf-concepts/

[14] Allemang D. and Hendler J.A. (2008). *Semantic web for the working ontologist modeling in RDF, RDFS and OWL.* Morgan Kaufmann, Burlington.

[15] McGuinness D.L., van Harmelen F., (eds.) (2004). *OWL Web Ontology Language Overview.* W3C Recommendation 10 February 2004, Available from http://www.w3.org/TR/owl-features/.

[16] Harnad, S. (1990). The Symbol Grounding Problem. *Physica D 42.* pp. 335-346.

[17] Bartholomew D. (2008). *Building on Knowledge*. John Wiley & Sons Ltd, Chichester.

[18] Yurchyshyna, A., et al. (2008). Towards the Knowledge Capitalisation and Organisation in the Model of Conformity-Checking Process in Construction. *Knowledge-Based Intelligent Information and Engineering Systems. KES 2008.* Part I, vol. 5177 (eds. I. Lovrek, R.J. Howlett, and L.C. Jain), Springer, pp. 341–348.

[19] Schevers, H., et al. (2007). Towards Digital Facility Modelling for Sydney Opera House Using IFC and Semantic Web Technology. *Itcon*, Vol. 12, pp. 347-362.

[20] Schevers, H. and R. Drogemuller (2005). Converting IFC Data to the Web Ontology

Language. *Proc. International Conference on Semantics, Knowledge and Grid.* Beijing, China pp. 73-73.

[21] Dolenc, M., et al. (2007). The InteliGrid Platform for Virtual Organisations Interoperability. *Itcon*, Vol. 12, .pp. 459- 477.

[22] Beetz, J., van Leeuwen, J. P., and B. de Vries (2005). An Ontology Web Language Notation of the Industry Foundation Classes. *Proceedings of the 22nd CIB W78 Conference on Information Technology in Construction.* (eds. R.J., Scherer, P., Katranuschkov and S.-E. Schapke) CIB publication no. 304, Dresden, pp. 193-198.

[23] Böhms, M. et al. (2008). *The SWOP Semantic Product Modelling Approach.* STRP NMP2-CT-2005-016972 report. Available from http://www.swop-project.eu/swop-solutions/semantic-product-modelling-approach/SWOP_D23_WP2_T2300_TNO_2008-04-15_v12.pdf (May 2009)

[24] ReDeFer (2009). *ReDeFer - compendium of RDF-aware utilities.* Available from http://rhizomik.net/redefer (May 2009)