Building Product Catalogues on the Semantic Web

J. Beetz
Eindhoven University of Technology, Eindhoven, Netherlands

B. de Vries
Eindhoven University of Technology, Eindhoven, Netherlands

ABSTRACT: In this paper we describe a prototype implementation of an ontology repository that captures the concepts in the Ontology Web Language OWL. We describe how these concepts can be used directly by embedding them in standard HTML pages and thus augmenting traditional product catalogues with semantically rich information by means of RDFa. As an addition to the ISO part 12006-3, where such a mechanism is not specified explicitly, we propose a way to instantiate actual products, their types and attributes through an instance-of relationship. Building upon the rich family of Semantic Web standards such as SPARQL and RDFa, we demonstrate how information in building product catalogues can be made machine-accessible in more efficient and generic ways. Using the Open Source persistency frameworks we demonstrate how real-world products can be linked to 13,000+ concepts with some 44,000 names in different languages in efficient ways.

1 INTRODUCTION

To enable more specific searches for building products within the vast range of available solutions from product manufacturers, traditional searching often does not lead to the desired results. This problem is partly rooted in the string-based nature of the search algorithms of today's World Wide Web. To overcome such shortcomings, new approaches to interweave information are being developed within the Semantic Web initiative. Here, information is being modeled as concepts by means of logical axioms that form lattices which are referred to as ontologies. By referring to these concepts from central ontologies, machine-interpretable semantic meaning can be attached to product specifications across the borders of language, measuring-units and local building regulations.

With the family of standards of the ISO 12006 International Framework for Dictionaries (IFD) (ISO 12006-3 2006) a similar notion of a central repository of building related concepts has been proposed and standardized. In part three of this standard "Framework for object-oriented information", a STEP EXPRESS schema (Schenck & Wilson 1994) has been defined to capture elementary information of building products and relate them amongst each other through objectified relationships such as specialization and decomposition. A reference implementation has been developed in a joint effort of several international partners to store instances of concepts in a central repository. However, the current way of exposing these concepts through a web-service API still requires a considerable amount of hand-crafted code to harness this information in real-world scenarios and use it for concept addition, query, and retrieval. Furthermore, no standardization or accepted best-practices for actual instantiations of this schema are available as of yet. In the later parts of this paper we propose an approach to address this need for a common modeling strategy.

2 RELATED WORK

The need to organize building information models (BIM), (national) classification systems and building product libraries in standardized, interoperable ways has been outlined in many research works and has found its way into many of today's commercially available applications and standards. Rooted in earlier classification systems such as the SfB (Giertz 1982), initiatives like the Dutch Lexicon effort led by Woestenenk (2000, 2002) and the BARBi library by Bell and Björkhaug (2006) have created the need of finding a uniform standard for modeling taxonomies and complex ontologies. The ISO 12006 parts 2 and 3 were defined as an answer to this need. Their modeling choices are discussed in section 3 of this paper.

Aiming at a standardized model, the Industry Foundation Classes (IFC) defined a number of building elements to be captured alongside other aspects such as actors, processes, resources and geometric/topologic representations as an interoperability vehicle to transport information between different domain applications. However, the structure of the model has several shortcomings regarding its practical use. One of these issues was its lack of a flexible and yet semantically rigid extension mechanism. Though its earliest extension mechanism – PropertySet definitions – enable the annotation of existing elements or blank proxies with additional properties, the addition of new element classes, flexible decomposition and the functional- and aspect-independent organization of building components cannot be realized in straight-forward ways. As Ekholm (2004) pointed out, a flexible coupling with external classifications is a necessity. As a result of several years of development and negotiation of the XM-7 working group a new extension mechanism was introduced with the IfcExternalReference and finally the IfcResourceObjectAssociationRelationship in the current 2x4 version of the model. However, these external references – which indeed provide many desirable features such as the possibility to point to specific URIs and keys in external sources – do not solve these problems, since they are only applicable to a limited set of entities on the resource layer of the model: While properties, materials, units, quantities etc. are definable using external modeling constructs, entities on the kernel level cannot be defined and thus external references remain second-class citizens in the model. The paradigm shift that Ekholm suggested still has to be addressed.

In order to overcome some of these modeling limitations, a number of research and development efforts have aimed at other means of coupling several independently defined models to the IFC model. Taking existing taxonomies as a starting point, the eConstruct (Tolman et al 2001), e-COGNOS (Lima et al 2003) and FUNSIEC (Lima et al 2006) line of research initiatives have used XML and Semantic Web Technologies to create controlled vocabularies and ontologies that allow the incorporation of legacy IFC data.

Recently, Shayeganfar et al (2008) have used a Semantic Web version of the IFD model to capture product catalogue data in RDF persistency data bases. However, the specifics of the modeling approaches have not become apparent and the prototype has been focused on a single, well defined and uniform product database with inferences limited to more conveniently formulate queries. The intended contribution of the work presented here is the extension of this approach for heterogeneous information sources distributed over different locations.

The use of ontologies for the description and search of products has been a topic outside of the building and construction industry for both research and development. A number of relevant standardized product classification and description formats such as eCl@ss, UNSPSC, eOTD (Fensel et al 2001, for a content-centered overview see Hepp et al 2005) are in use in other industries and several research projects have aimed at their translation into ontologies using formal languages such as DAML+OIL or OWL (Ciu et al 2003, Lee et al 2006).

3 MODELING OF THE CONCEPT LIBRARY AND ITS INSTANTIATONS

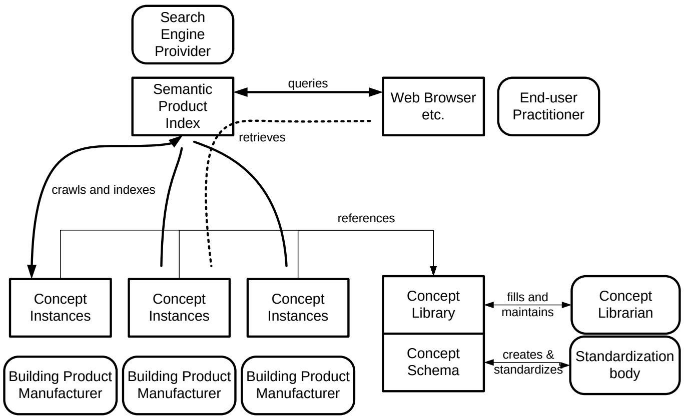

Figure 1 Overview of distributed information resources indexed by a semantic product search engine.

The scope of model defined in the ISO 12006 part 3 is the “the specification of a taxonomy model, which provides the ability to define concepts by means of properties, to group concepts, and to define relationships between concepts. Objects, collections and relationships are the basic entities of the model.” (ISO 12006-3 2006). And although “the model described in this standard is proposed as a bridge between classification systems as described in ISO 12006-2 and product modelling”, unfortunately, no explicit definitions have been included into the standard, that regulate the *use* of taxonomies modelled with this meta-model. In order to harness concepts defined in libraries, additional agreements have to be made that ensure interoperability among target applications such as product catalogues.

We propose the distinction of four hierarchical meta-modelling layers (abbreviated as M#) to achieve product descriptions based on IFD concepts, namely: (1) The *kernel layer* M0, (2) the c*oncept library layer* M1, (3) the *product kernel layer* M2 and the (4) *product instantiation layer* M3. An overview of the architecture and their respective elements is given in Figure 2.

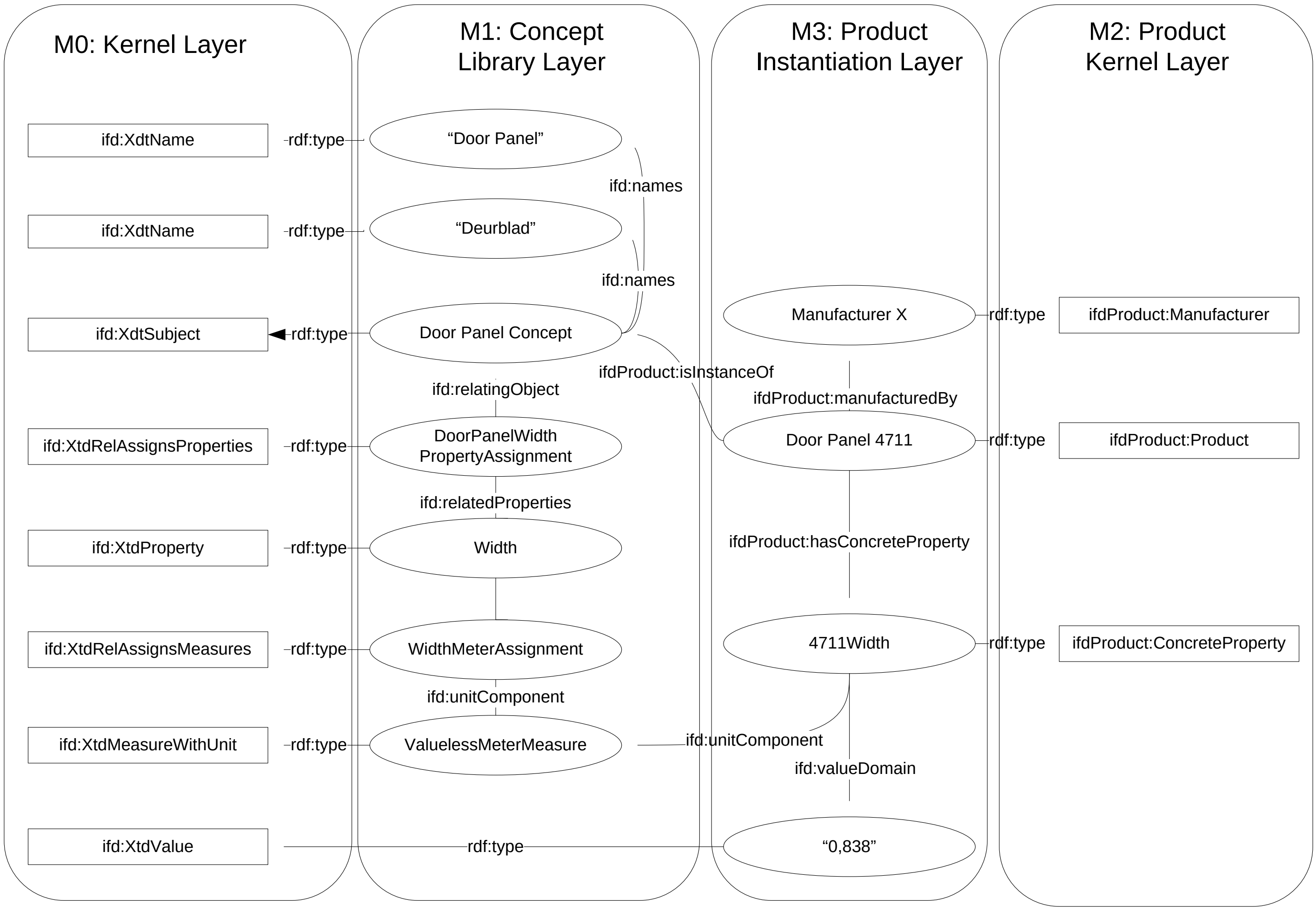

Figure 2 Simplified diagram of the four layers of concrete product instantiations based on ISO 12006-3. Arrows represent owl:ObjectProperties with the appropriate namespace in front of the colon to indicate their provenance. Oval shapes are RDF node instantiations, box shapes OWL Classes.

3.1 *Kernel Layer* M0

This layer consists of a straightforward translation of the EXPRESS schema defined in ISO 12006 part 3 into OWL. This translator has been described in (Beetz et al 2005, Beetz et al 2009).

The XtdRoot class is the superclass of all other essential classes on this layer with the exception of the natural language related classes. The most important feature of this class is the ability to attribute it with a globally unique identifier (GUID) to be unambiguously referred to. This modelling choice was made to overcome the inability of STEP-based modelling technologies to identify entity instances across the local address scope of memory, database-id or '#positiveInteger'-names in ISO 10303 part 21 serializations. Even though the use of an RDF-based form chosen for our architecture renders this feature superfluous (since this functionality is taken over by the URI of RDF nodes) the GUID has been kept for compatibility reasons. Furthermore, every XtdRoot and its subclasses have 1-n names and in freely definable languages (including non-human 'languages' such as 'IFC2x2' found in the current ifd-library.org repository) and an arbitrary number of descriptions. Descending from this root class, three subclasses have been defined:

1. The subclasses of XtdObject allow the description of subjects, properties, actors, activities, measures, units and values.
2. XtdRelationship and its specializations allow the creation of objectified relations to connect XtdObjects among each other. In further subclasses these relationships are specialized as assignments, specializations, decompositions, collections and sequences.
3. XtdCollection and its two descendents bag and nest allow the qualification of collection relations.

3.2 *Concept Library Layer* M1

On this layer, instances of the concepts entities defined on the *kernel layer* M0 are gathered. Currently such a library is available at ifd-library.org and consists of the joint concepts of the Norwegian BARBi library and the Dutch LexiCon initiative which are in the process of harmonization. Other local taxonomies and classification systems, such as the OmniClass, UniFormat and MasterFormat classifications will be added in future. However, the latter are "non-descriptive" classifications systems in that they do not define attributes for the individual nodes in the classification hierarchy, whereas such attributes are present in LexiCon and BARBi. We used a snapshot of the public library that consists of some 13,000+ concepts with more than 44,000 names in different languages, we converted to OWL individuals of the appropriate concept types. These were stored in a database-backed RDF repository using the Open Source frameworks Sesame and Joseki.

For the purpose of the prototype, additional concepts, properties and relations to describe door products were inserted into the library.

3.3 *Product Kernel Layer* M2

On this layer additional concept classes are modelled that enable the instantiation of abstract concepts from the M1 such as "door panel" into a concrete product "Door model 4711 by Manufacturer X". In our approach we have chosen to introduce additional *is-a* relationships that connect concrete product instances to the abstract concepts on the *concept library layer* M1. It might be argued that instead of choosing this additional relation, concrete products, product types and specific parameter configurations of these products could be modelled along the specialization axis provided by XtdRelSpecializes on the *kernel layer* M0. However, in order to keep a strict separation between abstract concepts and their concrete product instantiations these new relations are being introduced. This choice is also rooted in best practice recommendations from the ontology engineering community. The knowledge modelling aspects of a strict separation between concept and instance is also going to be addressed in section 4 of this paper.

3.4 *Product Instantiation Layer* M3

On this layer building product manufacturers themselves or building product catalogue providers instantiate concepts provided by a central repository on the *product library layer*. In order to create the concrete instance of a building product, a "Product" concept – defined on the *product kernel layer* M2 – is instantiated. Secondly, an "isInstanceOf" relation is created, which points to an XtdSubject concept instance in a *concept library* M1 repository. Concrete values of properties that are associated with the XtdSubject are again asserted to a "Product" instance via a specialized relationship "hasConcreteValue" that connects a XtdProperty to an M3-instance of an objectified relationship concept "ConcreteValue" defined on M2. This "ConcreteValue" relationship object functions as a selector to determine e.g. which of the possible XtdMeasureWithUnits "distance in m", "distance in cm" or "distance in mm" that have been assigned to a "DoorWidth" property is actually asserted with a specific value of "838".

3.5 *Summary of modelling mechanisms*

Apart from the four meta-levels introduced in the previous section a fifth level will come into existence once a number of pieces of a specific door will be used in a BIM for a concrete building model and have to be individually identifiable. These then become instances of products which in turn are instances of concepts. Additional layers such as type information that are specialized only in limited aspects (a certain door panel product with three different coating options, and a number of standard sizes) are likely to be required in real-world usage scenarios.

4 THE ADDED VALUE OF A RDFS/OWL FORM

Encoding the concept library M1 itself as well as its instantiations on the other layers using RDF(S)/OWL has several advantages compared to current implementations on STEP basis.

- Instantiations can be stored in W3C certified notations that have a large user base and can be expected to play an even more important role in future systems. They are location independent and can be distributed over arbitrary information resources such as web pages and vendor product databases.
- RESTful query endpoints in exiting RDF query languages such as SPARQL (Perez et al 2006) and SeRQL (Broekstra & Kampman 2003) can be used for much more efficient retrievals of complex graphs than currently possible with the SOAP API in the ifd-library.org reference implementation. Here, several sequential calls and answers have to be made in order to retrieve e.g. all properties and measures and values of the door concept.
- Through the addition of mappings and rules additional explicit facts can be inferred from the implicit information present in concept and product repositories. By keeping the logic independent and separate from the actual content flexible architectures can be composed that allow fine-grained control over the knowledge stakeholders are willing to publicize. In the following sections three examples of such inferences and mapping will be demonstrated.

5 INFERENCE EXAMPLES

5.1 *Cross-language queries*

One of the strong points of the IFD library architecture is the assignment of multi-lingual names and classification information for concepts. This allows the retrieval of concepts and their relating instantiations in form of concrete products in different contexts. Door panels in product catalogues that are exclusively described in British English can be searched by issuing a search query in Dutch: the search string “deurblad” will be found as being one of the names attached to this specific concept. Querying for all products that have been tagged with this concept (using its English name during the creation of the British manufacturer) information resources that have no human-readable descriptions of their products languages other than English cannot be achieved using traditional search strategies. Using SPARQL, this can be done in a single query:

```
SELECT ?product
WHERE  {
        ?concept ifd:names ?conceptname .
        ?conceptname ifd:name ?name .
        ?product product:isInstanceOf ?concept
        FILTER regex(?name, "deurblad", "i")
        }
```

In a similar fashion, users could search on the basis of classification codes (e.g. “21-41 51 13 21” for “exterior doors” in OmniClass)

5.2 *Unit-independent constraint search*

To limit search queries to door panels of certain sizes, or other property specifications, parameter constraints like “width = 83,8 cm” are issued for search queries. In traditional string-based search manufacturer entries like “w: 838 mm” will not be included in the query results for this search. Even when using standardized properties, measures and units, conversion rules are needed to be able to map one to the other. In traditional systems, such rules are often only embeddable by hard-coding them into the system for a number of standard units. A very central idea of the Semantic Web initiative however, is to keep these rules – part of the knowledge to be captured for automation – separate from the code as well. With rule languages such as SWRL (O’Connor et al 2005) that themselves can be formulated as RDF graphs (and hence be distributed over arbitrary locations) transformation such as measure and unit conversions from this simple example can be achieved like this:

```
  relatedMeasures(?doorWidth, ?me-
terMeasure) ^
unitComponent(?mMeasure “m”) ^ va-
lueDomain(?mMeasure ?mValue) ^
swrlb:divide(?mmValue, ?mValue,
1000)
  -> relatedMeasures(?doorWidth,
?mmMeasure) ^ unitComo-
nent(?mmMeasure, “mm”) ^ valueDo-
main(?mmMeasure, ?mmValue)
```

which can be read as “if some property has a measure with a unit “meter” then it follows, that it has a measure with the unit “mm” and a value for that measure of the meter value divided by 1000”.

6 PROTOTYPE IMPLEMENTATION

In the prototype we used to test the architecture described above, we used the Open Source Sesame RDF framework to store a RDF/OWL representation of the 4 meta-layers. The M0 and M1 layers were translated from the ISO 12006-3 Express schema and an instance file converted from a snapshot of the ifd-library.org database. The resulting RDF repository consists of several ten thousand triples and can answer to simple queries illustraated in paragraphs 5.1 and 5.2 within half a second on current average notebook hardware. The additional layers M2 and M3 were created as a proof of concept. The M2 product kernel layer is a lightweight ontology that enables tracking of the provenance data of products by storing their manufacturer, data source and providing relations such as isInstanceOf (see section 3.3). For testing purposes instances of doors with random parameters were created. In a real-world-scenario, this information can be scraped from RDFa (Adida & Bribeck 2008) meta-data that is seamlessly embedded into web pages of manufacturers. A search engine, that visits these pages indexes them and stores the information converted from RDFa or other Semantic-Web micro-formats together with its provenance data in a database. Since full SWRL integration is computationally too expensive to be used as a just-in-time inference engine for datasets discussed here, we have made use of “LET” constructs of the ARQ SPARQL implementation (ARQ 2009) to achieve unit transformations described in 5.2.

7 CONCLUSIONS AND FUTURE WORK

In this paper we have proposed an approach for distributed heterogeneous catalogues of building related products using Semantic Web methods. We have shown how an Ontology Web Language notation of concept definitions and their instantiations based on ISO 12006-3 can be used in addition with a lightweight ontology for concrete products to retrieve product information from distributed resources. We have introduced a four-layered approach and have shown example uses using a prototype implementation based on open standards and technologies.

The observations and experiences made in the context of this research can be categorized into two categories:

1. ISO 12006-3 related: It is imperative for the successful uptake of the standard to define clear guidelines for the use of it. The current version of the standard is too broad and leaves too much room for interpretations on how to structure and use concepts and their instantiations. To overcome this, a set of design guidelines should be developed to be encouraged – if not enforced – for future and current users. It begins with very simple things: Should concepts, that are specializations of another concept inherit the properties of the super-concept, e.g. should all doors have the same height and width attributes assigned to them, or should the assignment be made on a per-concept basis?

2. API-related. After having worked with both the existing SOAP API and the SPARQL endpoint for the concept library, complex searches can be achieved in a more straight-forward manner and needs less calls and exchanged messages between client and server using the proposed architecture compared to current STEP/SOAP implementations. The use of widely used interfaces such as SPARQL query endpoints potentially lowers the threshold of the uptake of the IFD framework in practice.

Future work will include the further addition and testing of different inference rules and the search for the optimal balance between genericity and expressiveness on the one hand and decidability and performance issues on the other. General purpose rule languages such as SWRL bring performance penalties with them and seem to be overkill for most of the uses of the IFD concept libraries and their instantiations. For a more in-depth study we plan to apply the approach introduced here to an extensive product catalogue of a commercial Dutch vendor.

8 REFERENCES

ARQ 2009 ARQ - A SPARQL Processor for Jena. Online http://jena.sourceforge.net/ARQ/, last accessed 12.05.2009

Adida, B., & Birbeck, M. 2008 RDFa Primer. Bridging the Human and Data Webs. On the WWW. http://www.w3.org/TR/xhtml-rdfa-primer/ , last accessed 12.05.2008

Beetz, J., Leeuwen, J.P. van, Vries, B. de. 2005. An Ontology Web Language Notation of the Industry Foundation Classes. *Proceedings of the 22nd CIB W78 Conference on Information Technology in Construction*. pp. 193-198

Beetz, J., van Leeuwen, J., & de Vries, B. (2009). IfcOWL: A Case of Transforming EXPRESS Schemas into Ontologies. *AI EDAM, 23(Special Issue 01)*, pp. 89-101.

Bell, H. & Björkhaug, L. 2006. A buildingSMART ontology. *Ework and Ebusiness in Architecture, Engineering and Construction: Proceedings of the 6th European Conference on Product and Process Modelling*, 13-15 September 2006, Valencia, Spain.

Broekstra, J., & Kampman, A. .2003. SerQL: A second generation rdf query language. In Proc. SWAD-Europe Workshop on Semantic Web Storage and Retrieval. pp. 13-14.

Cui, Z., Shepherdson, J., & Li, Y. (2003). An Ontology-Based Approach to eCatalogue Management. *BT Technology Journal*, *21(4)*, 76-83.

Giertz, L.M. 1982 SFB and Its Development 1950-1980. (CIB/SFB International Bureau on behalf of the author; Foras Forbartha distributor: Dublin, Ireland.

Ekholm, A. 2005. ISO 12006-2 and IFC – Prerequisites for Coordination of Standards for Classification and Interoperability. *ITCon* 10, 275-289.

Fensel, D., Ying Ding, Omelayenko, B., Schulten, E., Botquin, G., Brown, M., et al. (2001). Product data integration in B2B e-commerce. *Intelligent Systems*, *IEEE*, 16(4), 54-59.

Hepp, M., Leukel, J., & Schmitz, V. 2005. A quantitative analysis of eCl@ ss, UNSPSC, eOTD, and RNTD content, coverage, and maintenance. In *Proceedings of the IEEE international conference on e-business engineering 2005* pp. 572-581.

ISO 12006- 3. 2006 Building construction — Organization of information about construction works — Part 3: Framework for object-oriented information

Lee, T., Lee, I., Lee, S., Lee, S., Kim, D., Chun, J., et al. 2006. Building an operational product ontology system. *Electronic Commerce Research and Applications*, 5(1), 16-28..

Lima, C.P., Fiès, B., Lefrancois, G. & Diraby, T.E. 2003. The challenge of using a domain Ontology in KM solutions: the e-COGNOS experience. *10TH International Conference on Concurrent Engineering*, 771—778.

Lima, C., Silva, C., Duc, C., & Zarli, A. 2006. A Framework to Support Interoperability among Semantic Resources. In Interoperability of Enterprise Software and Applications pp. 87-98.

O'Connor, M., Knublauch, H., Tu, S., Grosof, B., Dean, M., Grosso, W., et al. 2005. Supporting Rule System Interoperability on the Semantic Web with SWRL. In *The Semantic Web – ISWC 2005* . pp. 974-986.

Pérez, J., Arenas, M., & Gutierrez, C. 2006. Semantics and Complexity of SPARQL. In *The Semantic Web - ISWC 2006*. pp. 30-43.

Schenck, D. A., & Wilson, P. R. 1994. Information modeling: the EXPRESS way. New York, NY, USA: Oxford University Press, Inc.

Shayeganfar, F., Mahdavi, A., Suter, G. & Anjomshoaa, A. 2008. Implementation of an IFD library using semenatic web technologies: A case study. In: *ECPPM 2008 - eWork and eBusiness in architecture, Engineering and Construction* Sophia Antipolis, France 539 – 544.

Tolman, F., Böhms, M., Lima, C., Van Rees, R., Fleuren, J., & Stephens, J. 2001. Econstruct: expectations, solutions and results. Electronic Journal of Information Technology in Construction, 6, 175–198.

Woestenenk K. 2000 Implementing the LexiCon for practical use, *Construction Information Technology 2000. Taking the construction industry into the 21st century*.; ISBN 9979-9174-3-1; Reykjavik, Iceland, June 28 – 30.

Woestenenk, K. 2002 From LexiCon to XTD. In *EWork and EBusiness in Architecture, Engineering and Construction: Proceedings of the Fourth European Conference on Product and Process Modelling in the Building and Related Industries*, Portorož, Slovenia, 9-11 September 2002.