Making sense of building data and building product data

Pieter Pauwels
Dept. of Architecture and Urban Planning
Ghent University
J. Plateaustraat 22
B-9000 Ghent, Belgium
pipauwel.pauwels@ugent.be

Thomas Krijnen, Jakob Beetz
Dept. Built Environment
Eindhoven University of Technology
P.O. Box 513
NL-5600 MB Eindhoven, The Netherlands
t.f.krijnen@tue.nl;
j.beetz@tue.nl

Abstract

The architectural design and construction domain works with massive amounts of data (building data, engineering data, product manufacturer data, geographical data, regulation data) on a daily basis. More and more of this data is being handled using semantic web technologies. This position paper documents existing initiatives, focusing on the Industry Foundation Classes (IFC) ISO standard and the buildingSMART Data Dictionary (bSDD) and outlines how a multilingual lexicalized semantic network like BabelNet can make a useful contribution to this particular domain.

1 Introduction

Building data is modelled by many stakeholders involved in the building process, including architects, engineers, contractors and owners. Capturing the unambiguous meaning of the many concepts handled in the construction industry is one of longest standing challenges in this domain. The construction of many buildings using IT tools can be compared to the construction of the Tower of Babel, in which the building 'fails' as those working on it could no longer communicate properly[1]. With the advent of information technologies, this is typically also referred to as an interoperability problem.

This interoperability challenge is being addressed since many decades with the tremendous efforts on the production of a standardised data exchange format, Industry Foundation Classes (IFC) (Liebich et al., 2013), which is modelled in the EXPRESS information language (ISO, 2004). IFC is standardised by buildingSMART International (bSI) and the International Organization for Standardization (ISO). The IFC allows to semantically describe a building as a digital building model, including element types (walls, windows, spaces), complex 3D geometry, custom property sets and many more. Capturing such information in digital format is generally referred to as Building Information Modelling (BIM) (Eastman et al., 2008).

In addition to a core data exchange model, bSI produced the BuildingSMART Data Dictionary[2] (bSDD), which can be considered as a hierarchically structured encyclopedia of the different terms and concepts that are available in the international building product manufacturing market. This encyclopedia is instantiated by a concept repository following the ISO 12006 guidelines (ISO 12006, 2005) with an API so that it allows (1) to creating multiple dictionaries, ontologies and other content and (2) to mapping content in these ontologies and dictionaries. The bSDD is multilingual and contains tens of thousands of concepts and relationships representing international building classifications and codes.

Both the IFC data model and the bSDD are being made available as RDF graphs within bSI, allowing its usage outside the restricted and closed construction industry domain. In this position paper, we give a brief overview of these efforts. We furthermore outline how the available sources can be enriched with links to the BabelNet[3] data, finishing with an outline of how this can benefit the construction industry expert.

2 buildingSMART and semantic web technologies

The construction domain is now looking into the usage of semantic web technologies for enabling a decentralised building data management approach and linking building data more effortlessly with

[1] see also http://constructioncode.blogspot.be/2012/07/end-of-babel-ifc-promotional-video.html

[2] http://bsdd.buildingsmart.org/

[3] http://babelnet.org/

data in other domains (product manufacturer data, geographical data, regulation data). This has evolved into the W3C Linked Building Data Community Group[4] and the BuildingSMART Linked Data Working Group[5] (LDWG). One of the results of these efforts is a conversion from EXPRESS to OWL (TBox) and from IFC (an EXPRESS schema) to a corresponding RDF graph (ABox) (Pauwels and Terkaj, 2016; Beetz et al., 2013). As a result, BIM models can now be made available as RDF graphs that comply with the ifcOWL ontology[6]. For reference, the ifcOWL ontology contains 1230 OWL classes, 1578 object properties, 1627 individuals (Pauwels and Terkaj, 2016). A sample repository with open ifcOWL compliant RDF graphs is available[7], but most sample RDF data is not publicly available.

The construction sector also works intensively on a BuildingSMART Data Dictionary (bSDD[8]), which can be considered as a hierarchically structured encyclopedia of the different element types that are used through classification systems in the international building product manufacturing market. Using this multilingual data dictionary, which is designed as a thesaurus for representing classifications in parallel in a common repository, it is not only possible to describe more precisely what types of building products are made available by manufacturers, but it also makes the exchange and (multilingual) interpretation of these data easier. Efforts are also underway to make the bSDD data available as an RDF graph (Beetz, 2014). A sample bSDD dataset[9] is temporarily available, containing a total of 986161 triples. However, a standard procedure for generating the RDF graphs from the bSDD API is not available, nor are there any links with external vocabularies available (like ifcOWL or BabelNet).

One of the lexical concepts available in the bSDD, namely the Calcium silicate board concept[10], is reproduced in Listing 1. As can be seen in this example, this concept contains a definition in English (@en) and Norwegian (@nb-no), it has a GUID that is maintained in the URI of the concept, and version information is included. Furthermore, this concept is typed as a 'subject'. Other concept types used are property, bag, document, classification, measure, unit, value, nest, activity, and so forth. This information is now explicitly contained as strings in the RDF graph, but they could clearly be represented in a multitude of alternative RDF graph representations. For example, properties could be listed as object and data properties if they are available for a particular subject, which is the case for the Calcium silicate board concept.

```
<bsdd:3MyXi0NvmHt00000PR1IRl>
  a owl:Class ;
  rdfs:comment "Bygningsplate basert p
    kalsiumsilikat (sement, kisel og kalk
    ), med armering av cellulosefiber.
    Platene fremstilles ved
    autoklavherding. Benyttes innendrs i
    miljer hvor det stilles krav til
    fuktbestandighet og brannbeskyttelse
    ."@nb-no ;
  rdfs:comment "Building board based on
    calcium silicate (cement, silica and
    mortar), with reinforcement of
    cellulose filament. The boards being
    produced by autoclave curing. Being
    used in environments with demands to
    moisture resistance and fire
    protection."@en ;
  rdfs:label "Calcium silicate board"@en
    , rdfs:label "Kalsiumsilikatplate"@nb
    -no ;
  :conceptType "SUBJECT" ;
  :guid "3MyXi0NvmHt00000PR1IRl" ;
  :status "DRAFT" ;
  :versionDate "2007.09.10" ;
  :versionId "1 2007.09.10" .
```

Listing 1: RDF graph for the Calcium silicate board concept in the bSDD.

3 Combining IFC and bSDD with BabelNet

The bSDD contains lexical concepts that are made available from within the construction industry sector only. By providing the data as RDF graphs, however, the data can easily be enriched with lexical data. As a first step, a number of vocabularies have been semi-automatically pre-aligned (Shvaiko, 2013) in the context of the FP7 DURAARK project. In a second step, these pre-alignment relations between concepts in different vocabularies can be reinforced by experts or crowds. To facilitate this, a user interface has been created[11], of which a screenshot is shown in Fig. 1.

[4] http://w3.org/community/lbd/

[5] http://buildingsmart-tech.org/future/linked-data

[6] https://w3id.org/ifc/IFC4_ADD1#

[7] http://smartlab1.elis.ugent.be:8889/IFC-repo/

[8] http://bsdd.buildingsmart.org/

[9] http://bw-dssv16.bwk.tue.nl:8080/openrdf-workbench/repositories/bsdd/summary

[10] http://bsdd.buildingsmart.org/#concept/details/3MyXi0NvmHt00000PR1IRl

[11] http://bw-dssv19.bwk.tue.nl/interlink/

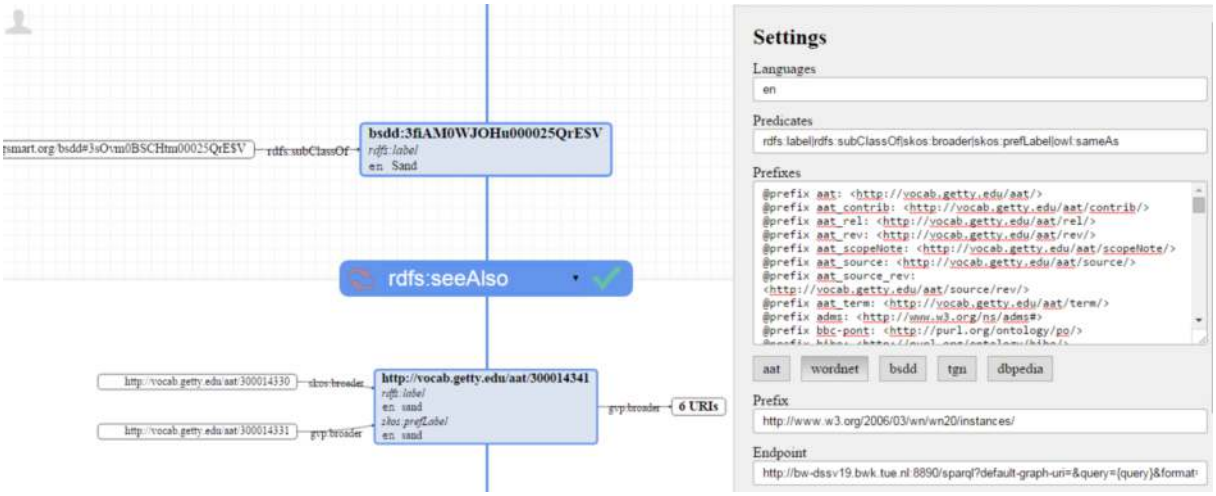

Figure 1: Screenshot of a dedicated concept mapping web interface for the bSDD.

This interface presents the user with a concept in the bSDD and a concept in an external vocabulary, including the AAT Getty Arts & Architecture Thesaurus[12], WordNet[13], the TGN Getty Thesaurus of Geographic Names[14], and DBPedia[15]. The interface then allows to specify the link between the two presented concepts, namely the concept for Sand in the bSDD and in the Getty AAT in the case of Fig. 1.

As a result, the concepts of these diverse thesauri can be combined into a global semantic network of concepts. The same approach could also be followed for enriching the bSDD with lexical concepts available in BabelNet. Such a considerably enriched multilingual lexicalized semantic network that includes bSDD, AAT, WordNet, BabelNet, WordNet, and TGN can be highly useful for the construction industry stakeholders in the sense that it helps construction domain specialists to get used to semantically structuring their data using state of the art semantic technologies. It also allows extending initial data sources specific to the construction industry (bSDD as well as ifcOWL) with multilingual data from third-party vocabularies (e.g. BabelNet).

4 Conclusions

In this position paper, we made a case for:

1. making building data available as structured RDF graphs,
2. allowing links to externally available structured vocabularies such as BabelNet.

This case was presented for the IFC schema and the bSDD data dictionary. An interface was also presented that allows to interactively create links from various concepts in the bSDD to concepts in various outside schema's (AAT, TGN, DBPedia, WordNet). Making this effort can result in a global semantic network of concepts, thus considerably enlarging the set of concepts and descriptions that is currently available in the bSDD. This semantic network of concepts can then be used in addition to existing BIM tools and services to classify building elements and exchanging well-defined information between construction industry stakeholders.

Of course, this is but an initial outline of what could be realised. Further research is highly necessary, more particular for:

1. optimising and finalising the conversion from the information in the bSDD to a usable RDF graph,
2. linking concepts in the bSDD graph to outside concepts, and
3. designing an optimal strategy to maintain links, authority and ownership (see 4 strategies presented in (Beetz, 2014)).

Acknowledgments

For their support, the authors would like to thank the Special Research Fund (BOF) of Ghent University, and the Eindhoven University of Technology.

[12] http://vocab.getty.edu/aat/
[13] http://www.w3.org/2006/03/wn/wn20/instances/
[14] http://vocab.getty.edu/tgn/
[15] http://dbpedia.org/resource/

References

T. Liebich, Y. Adachi, J. Forester, J. Hyvarinen, S. Richter, T. Chipman, M. Weise, and J. Wix. 2013. *Industry Foundation Classes IFC4 Official Release*. available online: `http://www.buildingsmart-tech.org/ifc/IFC4/final/html/index.htm`.

Anders Ekholm 2005. *Iso 12008-2 and Ifc Prerequisites for Coordination of Standards for Classification and Interoperability*. ITCon 10(2005):275-289.

Jakob Beetz, Jos Van Leeuwen, and Bauke de Vries. 2009. IfcOWL: a case of transforming EXPRESS schemas into ontologies. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 23(1):89–101.

Pieter Pauwels and Walter Terkaj. 2016. EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Automation in Construction*, 63:100–133.

ISO International Organization for Standardization. 2006. *ISO 10303-11: Industrial automation systems and integration - Product data representation and exchange - Part 11: Description methods: The EXPRESS language reference manual*. available online: `http://www.iso.org/iso/iso\_catalogue/catalogue\_tc/catalogue\_detail.htm?csnumber=38047`.

ISO International Organization for Standardization. 2005. *12006-3:2006 Building construction Organization of information about construction works Part 3: Framework for object-oriented information*.

Charles M. Eastman, Paul Teicholz, Rafael Sacks, and Kathleen Liston. 2008. *BIM handbook: a guide to building information modeling for owners, managers, architects, engineers, contractors, and fabricators*. John Wiley & Sons, Hoboken, NJ, USA.

Jakob Beetz. 2014. *A scalable network of concept libraries using distributed graph databases*. Proceedings of the 2014 International Conference on Computing in Civil and Building Engineering, Orlando, FL, pp. 569–576.

P. Shvaiko & J.Euzenat 2013. *Ontology Matching: State of the Art and Future Challenges*. IEEE Transactions on Knowledge and Data Engineering 25(1):158-176