A Scalable Network of Concept Libraries Using Distributed Graph Databases

J.Beetz[1]

[1]Eindhoven University of Technology, Department of the Built Environment, Den Dolech2, 5800 MB Eindhoven, The Netherlands; email: j.beetz@tue.nl

ABSTRACT

Currently, work on enabling the use of building and construction-related concept libraries is being carried out in a number of disparate initiatives. The Norwegian reference implementation of the buildingSMART data dictionary (bSDD), the Dutch initiative for a concept library for the infrastructural sector (CB-NL), the Concept Model Ontology (CMO) and the generation of meta-data for long-term archival of BIM data in the context of the EU-project “Durable Architectural Knowledge” (DURAARK) share a great number of common aims and goals. These cover the areas of the general conceptual architecture, governance strategies for the content, accessibility and interfaces as well as technical aspects of implementations and others. In this paper the ongoing collaborative work on a harmonized framework for networked, distributed concept libraries is introduced and discussed. It is shown how such systems may be used to expose the concepts for semantic mash-up scenarios in the context of the Linked Open Data (LOD) initiative. A number of use-cases that demonstrate the potential applications of such systems in various interoperability scenarios are illustrated and discussed.

INTRODUCTION

The buildingSMART Data Dictionary (bSDD) is the evolving collaborative effort to provide an open, extendable and reliable repository of concepts in the building industry. It is based on ISO 12006 framework consisting of the conceptual modeling guidelines in part 2 and a concrete data model defined in part 3. This framework, has triggered a number of individual implementations of useful information repositories over time. At present, a number of parties under the umbrella of the buildingSMART organization is deploying this structure on a production-ready platform.

Content ownership, trust and reliability issues can be identified as some of the main obstacles for the large-scale adaption 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. Furthermore, interested users are reluctant to contribute new content which might be interfered with by others, is not transparently communicated and might need introduce costs to be retrieved and used. Practical adoption problems might also be traced to the current approach to adapt to data structures and software interfaces that are only used within the building and construction domain.

As a possible future solution approach, the decentralization of “the” bsDD into a set of individual but connected and synchronized concept repositories following widely accepted standards from the Linked Open Data community is outlined in this paper.

EVOLUTION OF STRUCTURED VOCABULARIES AND ONTOLOGIES

Currently, most of the content found in the bSDD has been added and edited in a semi-structured manner. Main contributions have been received from individual and national initiatives such as BARBi (Norway), Lexicon (The Netherlands), Omniclass (US), IFC4 PSets (buildingSMART International) and have been translated, merged and mapped in ‘ad hoc’ processes and episodic efforts. This was enabled by the close formal and informal collaboration of the individual initiatives and driven by enthusiastic and passionate stakeholders. However, such ad-hoc management is limited with regards to its scalability and sustainability: As more and more stakeholders wish to model their own domain, national classification systems and organizational structures, a number of issues will have to be addressed allow such growth. These include the prevention of ‘pollution’ of the library by duplicate concepts and their relations as well as avoiding and managing contradictions and inconsistencies between different items in the vocabulary. Other principal issues affect the necessity of a quality assurance mechanism including consistency and integrity checking as well as versioning control including the archival of past versions that will have to remain valid. Issues on lower technical levels such as transaction safety including roll-back capabilities are depending on the underlying implementation of such vocabulary systems that are likely to be resolved on lower technical levels.

Other knowledge-intensive domains (and particularly medical and biological sciences) are facing similar issues, and a considerable community of researchers, developers and software vendors have devised a number of methods, technologies and best-practices that should be investigated and harnessed by the building and construction domain and standardization organizations such as the buildingSMART organization.

In particular, the controlled, user-driven evolution of structured vocabularies, taxonomies and full-fledged description logic ontologies has been the subject of many research investigations. Among the available conceptual frameworks, the evolution strategies devised by (Klein, 2004; Mizoguchi and Kozaki, 2009a; Noy et al., 2006; L. Stojanovic et al., 2002) have received wide-spread attention. Recent overviews of the topics involved can be found in (Flouris et al., 2008; Hartung et al., 2011; Mizoguchi and Kozaki, 2009b).

A general strategy of ontology evolution according to (L. Stojanovic et al., 2002) is shown in figure 1, lower right. Applied to the building and construction domain within the scope of the CB-NL and bSDD initiatives, these steps include

0) Analyze and formulate the requirements for a particular use case (e.g. the modeling and description of bridges). Discover and retrieve existing concepts already present in the concept library/libraries. Identify missing concepts or facets
1) Organize groups of domain experts to gather terms, formalize, model and capture the knowledge of the bridge-domain including its main concepts.
2) Represent the modifications (e.g. additions) in the library using the agreed upon modeling guidelines, preferably in a local 'shadow copy' or another sandbox environment
3) Describe and detect what changes will have to be implemented (deletion, addition, property modification, etc)
4) Implement the changes in the vocabulary
5) Propagate the changes by e.g. applying them to existing individuals or instances referring to the affected parts of the model
6) Validate the changes and the consequences of the propagation
7) Re-iterate and refine if necessary

Depending on the semantic rigidness (from simple dictionaries of unrelated terms to Description Logic based axiomatic theorems) even subtle changes may have severe impacts and side-effects on connected concepts. A step not fully applicable to the buidlingSMART / CB-NL context is the propagation of changes (step 5): Since no editor of the model should be able to affect referring (instance) models, concept states being published for reference should be preserved and not change their meaning or scope.

To address the needs identified in the first Pilot phase of the CB-NL project, an environment should be created that allows the independent creation, use and maintenance of concept repositories on four cascading tiers: 1) International level 2) Local level, 3) Organizational level and 4) Project level. These tiers should have an increasing level of generality with the international level being the most generic and the project level being the most specific. Lower tiers should re-use and refer higher-level concepts where possible and refine them according to the local requirements, e.g. local building regulations, national classifications or in-house organization project structure templates.

Three main approaches can be identified that potentially allow to create, use and maintain such multi-tiered concept repositories.

Single centralized concept repository. In its current implementation the bSDD, as illustrated in Figure 1, upper left, is a centralized repository: Multiple users with different roles have access to a repository via a single address. Within the repository compartments of the overall concept model are created using a 'context' mechanism which is also discussed later in this paper. Governance and maintenance, including access right management, safekeeping and integrity checking is done by a restricted group of managers and service providers. Concepts can be queried and modified within the model via an API exposed via SOAP and REST protocols. Reference clients such as browsers and editors are provided. All changes to the repository are done on the production server.

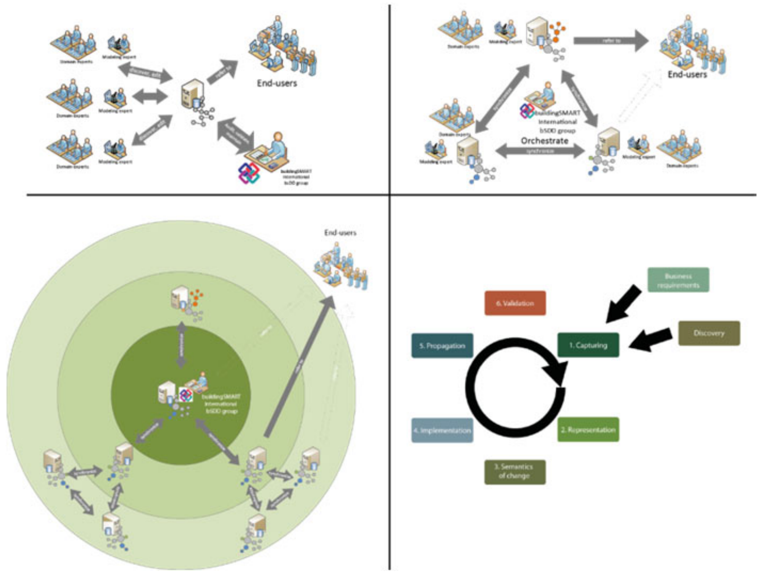

Figure 1. Approaches to organize concept libraries: centralized (upper left), peer-to-peer (upper right), hybrid (lower left). Vocabulary evolution mode according to Stojanovic (2009)

Distributed peer-to-peer concept repositories In this completely decentralized approach, no repository has a dominate authoritative nature by structural means, even though 'trusted' concept clusters will eventually evolve over time. Here, the agreement on a light set of interfaces and common modeling guidelines serves as a minimal technical umbrella that allows the publication of vocabularies by organizations independently from a central authority. The reference example of such loosely organized vocabularies are the "5 star rules" of the Linked Open Data community.

Hybrid network architectures In this approach the centralized and peer-to-peer approaches are combined into a multi-tiered, cascading network of linked individual repositories The mixture of the distributed and centralized approaches has the following characteristics: Repositories are organized into the four tiers identified at the beginning of this section

1. The central bSI bSDD super node collects publically available additions using e.g. a push/pull mechanism. A reviewing board in the bSDD working group decides whether or not to accept additions from local chapters and merges them into to the repository.
2. Local budildingSMART chapters review and edit local – e.g. national – content and approve them for public reference and/or promotion on the first tier

3. On the organizational tier, working groups for domain extensions, organizational structures and regional specifications pull the national repository from the national tier extend them and send a request for inclusion to the regional chapter
4. Project-specific repositories are derived from the third tier and are only visible and used within a specific context of e.g. a project. Best practices, details added concept types etc. for the specific project can eventually be promoted to the upper tier.

USAGE SCENARIOS

In this section a number of brief use case scenarios are provided that illustrate how such decentralized structures are used.

Discovery and usage of concepts. Concepts in the repositories can be referred to from within domain-specific applications including BIM, GIS, CAD, SE and other modeling tools to semantically enrich objects. Another scenario is the use of concepts in specification documents of building products and components. References to concepts identify the concept itself (e.g. by its GUID) as well as its provenance, e.g. the context of the concept and the version of the concept at the time of its reference.
Modular, re-usable reference clients to access and refer concepts should be provided by buildingSMART to jump-start the implementation on various platforms. They should behave similarly across applications but can be customized by software vendors

Parts of the end user Graphical User Interface specifications have been defined in the context of the CB-NL project. Once discovered and attached to an information artifact, end-users should be guaranteed to find the referred state of the concept(s) without having to ‘upgrade’ along with the evolving library.

Creating concepts for a domain or local context. Clusters of concepts describing a specific domain (e.g. roads, bridges, dormers, doorknobs) or local context (national building regulations, classification systems etc.) are modeled by domain experts supported by information modelers. To minimize the modeling effort it should be possible to reuse existing dictionaries, classification systems and information models and to map and transform them into concepts and their relations ranging from shallow lists of terms, elaborate “is-a” hierarchies to full-fledged ontological concepts in roles. For this, individual domain communities create, govern and maintain their domain independently and occasionally release major revisions of the domain concept model to the general public. Such activities could be coordinated by initiatives like OpenINFRA. Localization should enabled on various levels including natural language translations and the addition of local regulations and classifications. The collaborative editing of the domain should be supported by annotation, change request management, voting and revision mechanisms. The re-use of existing concepts is encouraged and desirable and could be facilitated through layered upper- and meta-level concept hierarchies and design patterns as well as best practices which e.g. could be provided by buildingSMART.

Creating concepts for an organization To meet the specific requirements e.g. of large organizations, in-house concept clusters for larger should be created by customizing, additionally specifying and restricting publicly available concepts from higher tiers. Such mechanism should allow project-specific assemblies of concepts which are compiled from templates and project-specific assemblies of concepts reuse and specialize and/or instantiate concepts from higher-level repositories

Governance A reviewing and quality assurance mechanism allows the cascading promotion and merging of decentralized content into the main reference repository. The semi-automatic detection of duplications, clashes and contradictions of concepts are supported by formalized business rules The accumulation of changes from different sources could be bundled into periodic releases. The governing institution should issues change-logs to provide human- and machine-readable deltas/diffs. Back-issues of releases are kept and served for several years and are sustainably archived for later retrieval

TECHNICAL ASPECTS

The IFD / bSDD model based the ISO 12006 part 3 data model currently has a number of shortcomings that make the implementation of the suggested distributed vocabulary repositories difficult. For example, it relies on contexts residing within the same database. It is also restricted to objectified relationships: in the current model, context does not cover the attachment of e.g. natural language translations of individual terms. If somebody wants to add the Dutch term ‘gebouw’ as an ifdName for the ‘Building’ concept he/she has to have write-capabilities to the concept that may reside in a concext owned by another bSDD user.

Semantic Web. Some of the above issues can be addressed using Semantic Web methods and technologies. For example, reification describes the notion of “making statements about statements”. It is an essential mechanism in the semantic web stack to help realize the concept of ‘trust’: In an OWA environment that supports the “Anyone can say Anything about Anything” (AAA) principle. Reifications enables to encode the provenance a statement allowing expressions like “RWS sees an ‘office building’ as a specializations of a ‘building’”. However the general usefulness of the principle is widely accepted in the semantic web community, the most practicable way of encoding this is still under debate: Three main approaches can be identified
Pure RDF Reification RDF reification using RDF descriptions are the default, yet somewhat cumbersome solution illustrate in a small example in Table 1.

Table 1 RDF reification "User 1 claims that 'Building' is an instance of ifdSubject”

```
:Building-type-ifdSubject
          rdf:type rdf:Statement ;
          rdf:object :ifdSubject ;
          rdf:predicate rdf:type ;
          rdf:subject :Building ;
          rdfs:isDefinedBy :ifdUser_1 .
```

Named Graphs The clear advantage of this approach is the capability to store and arbitrary number of statements into a graph. This way, a single GUID could identify a whole cluster of concepts including individual values that can be assigned to an object (e.g. a IfcProxy object in an IFC file) with a single name, e.g. a GUID. Another approach is to group a number of triplets into a graph. Another serialization is to store a context alongside every triple statement. The N-Quads syntax (see table 3) is also natively supported by a number of persistency back ends, referred to as “quad stores”

Table 2 Several statements clustered into a named graph context in TriG syntax

```
:RWS { :Building rdf:type ifd:ifdSubject .
       :OfficeBuilding rdf:type ifd:Subject .
       :_ rdf:type ifd:ifdRelSpecializes .
       :_ ifd:relating_object :Building .
       :_ ifd:related_objects :OfficeBuilding
}
```

OWL 2 Axiom annotations OWL 2 allows the addition of annotations to individual axioms. This can be seen and used as a kind of reification for the purposes of meta-modeling to e.g. encode provenance data. This approach can be considered a higher-level approach since it is not based on the underlying RDF model and its various forms of serialization and storage but

Table 3 OWL 2 annotation expressing the Building->OfficeBuilding specialization

```
:OfficeBuilding rdfs:subClassOf :Building .
 []  rdf:type       owl:Axiom ;
     owl:subject    :OfficeBuilding;
     owl:predicate  rdfs:subClassOf ;
     owl:object     :Building;
     ifd:ifdUser    MickBaggen .
```

CONCLUSION

In this paper a number of approaches, requirements and technical solution options have been introduced and discussed that potentially allow the creation of a distributed network of concept libraries. A number of beneficial aspects for the use of semantic web and graph structures have been discussed.

At present, the goal of flexible, granular and cascading concept libraries for the building industry is still in its early conceptual stages. The detailed specification, prototypical implementation, evaluation and testing of some of these approaches are currently carried out. Their finidings and results will be published at later stages and any feedback from the CIB W78 and other R&D communities is very much appreciated and needed to make such endeavors successful.

Parts of the work documented in this paper has been funded by the Dutch “Concept Library for The Netherlands – CB-NL” initiative under the umbrella of the national Building Information Council (BIR), the 3DSDI project from the Next Generation Infrastructures initiative and the FP7 project DURAARK (Grant 600908). Many of the ideas documented in this paper have been developed and discussed alongside the activities of the ICT advisory group of this initiative as well as the with the IFD/bSDD technical group of the buildingSMART organization. All members of these communities should be named here, but cannot.

REFERENCES

bSDD, 2013 http://www.ifd-library.org, last accessed 2014-01-02

CB-NL, 2013, http://www.cbnl.nl/ last accessed 2014-01-02

Beetz, J., Dietze, S., Berndt, R. & Tamke, M. (2013). Towards the long-term preservation of building information models. Proceedings CIB W78 2013, 9-12 October 2013, Beijing, China, (pp. 209-217). Beijing, China.

Flouris, G., Manakanatas, D., Kondylakis, H., Plexousakis, D., Antoniou, G., 2008. Ontology change: Classification and survey. Knowl. Eng. Rev. 23, 117–152.

Hartung, M., Terwilliger, J., Rahm, E., 2011. Recent Advances in Schema and Ontology Evolution, in: Bellahsene, Z., Bonifati, A., Rahm, E. (Eds.), Schema Matching and Mapping, Data-Centric Systems and Applications. Springer Berlin Heidelberg, pp. 149–190.

Klein, M.C.A., 2004. Change management for distributed ontologies.

Mizoguchi, R., Kozaki, K., 2009a. Ontology Engineering Environments, in: Staab, S., Studer, R. (Eds.), Handbook on Ontologies, International Handbooks on Information Systems. Springer Berlin Heidelberg, pp. 315–336.

Mizoguchi, R., Kozaki, K., 2009b. Ontology engineering environments, in: Handbook on Ontologies. Springer, pp. 315–336.

Noy, N.F., Chugh, A., Liu, W., Musen, M.A., 2006. A framework for ontology evolution in collaborative environments, in: The Semantic Web-ISWC 2006. Springer, pp. 544–558.

Palma, R., Corcho, O., Gómez-Pérez, A., Haase, P., 2011. A holistic approach to collaborative ontology development based on change management. Web Semant. Sci. Serv. Agents World Wide Web 9, 299–314.

Stojanovic, L., Maedche, A., Motik, B., Stojanovic, N., 2002. User-Driven Ontology Evolution Management, in: Gómez-Pérez, A., Benjamins, V.R. (Eds.), Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 285–300.