Controlled Vocabularies in the European Construction Sector: Evolution, Current Developments, and Future Trends

Celson Lima[a,1], Alain Zarli[a], Graham Storer[b]

[a] CSTB, Centre Scientifique et Technique du Bâtiment, Sophia Antipolis, France.
[b] GSC, 18 Amersham Hill Drive, High Wycombe, HP13 6QY, United Kingdom.

Abstract. In the last 40 years, the development of Controlled Vocabularies (CVs), such as dictionaries, classifications, taxonomies, and of course the "appealing" ontologies, has been the focus of many research projects around the world targeting the Construction sector. Being involved in several pan-European initiatives, the authors of this paper show milestones on the path of evolution (what has happened so far), the current situation (where we are now) in terms of development and adoption of results, the main problems found regarding both development and adoption of the CVs, and finally, present some speculative and provocative ideas about the future of CVs in the European Construction sector.

Keywords. Controlled Vocabularies, Ontologies, Taxonomies, thesaurus, Construction.

1 Introduction

In the last 40 years, the development of Controlled Vocabularies such as dictionaries, classifications, taxonomies, and of course the appealing ontologies, has been the focus of many research projects in Europe. A non-exhaustive list of well known efforts in this are is the following: ISO12006 parts 2 and 3, LexiCon (the Netherlands), Barbi (Norway), bcBuildingDefinitions taxonomy (e-Construct Project), ICONDA terminology (Fraunhofer IRB), BS6100 and UNICLASS (British Standards), e-COGNOS ontology (e-COGNOS project), Standard Dictionary for Construction in France (SDC). It is worth recalling that in other continents similar efforts were also conducted, such as the SI/SfB, Masterformat, Omniclass, and the Canadian Thesaurus, just to name a few.

Even a brief review of the above listed projects/initiatives allows us to imagine how much effort has been devoted to this area around Europe[2], (likely) guided by a

[1] CSTB, Centre Scientifique et Technique du Bâtiment, Route des Lucioles, BP 209, Sophia Antipolis CEDEX, France. Tel: +33 4 93956722; Fax: +33 4 93956733; Email: celson.lima@cstb.fr.

single aim: to put the Construction sector at the leading edge for the best advances of semantic-related ICT resources. Preliminary thoughts were about developing useful e-Commerce/e-Business related tools and resources to help construction companies publish their own catalogues using their own languages and, at the same time, become actors in the European arena.

Based on the results achieve so far, however, a quite pragmatic question arises: after one decade (or more) investing on this topic, what is the reality in Europe nowadays? Are Controlled Vocabularies really used in a daily basis by Construction actors or are they still remaining a piece of art contemplated and admired behind the security protection provided by the 'research walls'?

If Controlled Vocabularies are fully adopted and used on a daily basis, what might we do with them in the future, and what are the trends currently observed in the research area? What are the domain(s) of work where Controlled Vocabularies will likely play an important role? What is the future of e-business and e-commerce related activities in Europe, a very fragmented market where the national (sometimes regional) norms and regulations impose a strict control on products and services construction-oriented? And a more ambitious question, what Construction can do in order to place itself properly regarding the business exploitation of the Semantic Web?

This paper discusses the questions raised above, based on the experience gathered by the authors through their involvement in several European initiatives related to the subject. Section 2 presents the main reasons behind the development of CVs in Construction (why). Section 3 discusses very briefly a selected set of European/International initiatives on this area. Section 4 draws a picture about where we are now, preparing the ground for the speculative and provocative discussion in Section 5, about where we are going *versus* where we could/should go. Finally, some conclusions close the paper.

2 Major reasons behind the development of Controlled Vocabularies in Construction

Well, why develop controlled vocabularies for any field of activity? Here are some unquestionable reasons for this necessary activity:

- **Vocabularies give names to things that have meaning at a certain level of detail.** Vocabularies provide a convenient shorthand for exchanging information. For example, "dog" means "a domestic carnivorous animal with four legs that typically has a long muzzle, pointed ears, a fur coat, a long fur-covered tail, and whose characteristic call is a bark". It is certainly different to "elephant" or "bicycle". So, if somebody says, "Where is my dog?" we know the kind of thing to look for. But there are many dogs. We could add adjectives like "small", "long", "short-legged", "drooping eared", "German" (which adjectives must have agreed meaning in the dog

[2] Not only Europe, but for the context of this paper the discussion will remain inside European borders.

context) or we could simply use another name “dachshund” or “sausage dog”. These need to have agreed meaning, not least because to the English or French the word dachshund is foreign and the other is a descriptive nick-name. The deeper we need to go with meaning to add detail or to differentiate, the more control there needs to be in the use of the language. Between specialists in one discipline there can be quite precise understanding of words (in this case zoologists who might even use Latin names) but between experts and non-experts and different kinds of expert there can be misunderstanding. To change to a construction example, what is the difference between a “brick pillar” and a short length of thick wall made from brick? A bricklayer and a cost estimator might use different terms. The answer (in UK at least) is that the difference is defined by rules related to the dimensions.

Figure 1. About the meaning of words (from Alice in Wonderland)

- **Vocabularies are important to conveying human thought in a concise way and with precision in a given working context.** Vocabularies must be controlled to achieve this or we have the Humpty Dumpty situation (from Alice in Wonderland) pictured alongside. There must be as much preciseness as possible although in human exchanges we sometimes say that something is like something else e.g. the dog is like a dachshund but with longer legs. We can then ask questions to refine meaning and (perhaps finally) identify the breed of dog.
- **Controlled vocabularies are even more important to electronic information exchange in any form.** Whilst humans can ask clarifying questions based on their experience and knowledge, computers do not have that as a general capability (though in limited contexts artificial intelligence may enable that). So there needs to be precision built into the language of computer communication used. There is much less possibility for confusion if an object is referred to by say its catalogue/part number and as a buyer I use that to describe my need to the supplier. But not everything can be conveyed that simply. Architectural details, a building frame, a plumbing

system etc. are usually designed to result in requirements that facilitate choice of components to satisfy the need. So generic types like wall, pump, foundation are used that are then specialised according to properties (such as dimensions, material, colour, strength etc) which themselves must have precise (i.e. agreed) meanings. Although codes could be used to identify components and systems, it is far more convenient that the codes take the form of the names we humans use. “Pump” not A254GHT7 unless we are buying from a catalogue!

- **With control of vocabulary (including between natural languages) being so important, who should control it for the Construction Industry**? The answer is the practitioners and ICT providers through standardisation processes of one kind or another. The struggle of the past decade or more has been to put in place the national, European, international and cross discipline organisation for that to happen. The Construction Industry is hugely fragmented in terms of disciplines involved, their locations and skills and often undertakes major projects as international consortia. Such consortia are temporary alliances for the duration of a project, perhaps no more than a year or so in a design phase. Project lead times can be exceedingly short with teams from different disciplines (and within disciplines) put together at quite short notice. The need for standardised, controlled vocabularies in such a work environment is crucial. But where has the resource come from to cover the cost of vocabulary development and control? The industry is highly cost competitive and not noted for its research investment by virtue of the small size of the vast majority (80%+) of companies. Progress has largely been made through efforts arising from collaborative research projects, initiatives and developments such as those listed below. Finance for effort has been contributed to by funding programmes (national and EU) and the participating organisations themselves.

3 State of the Art on CVs for Construction in Europe

As previously indicated, a large effort has been devoted to the creation and use of CVs around the world. This section briefly summarises a suite of relevant research projects, and pan-European & international initiatives (Figure 2). It is worth noting that this broad scene is not presented as exaustive or complete. It simply provides inputs to the brainstorming and discussion presented in the next section.

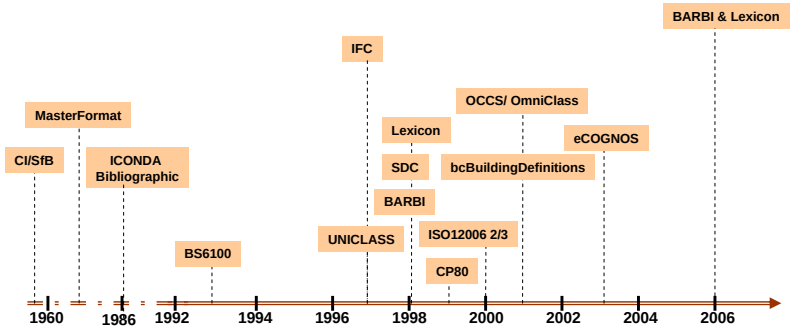

Figure 2. Some examples of CV-focused initiatives in Europe and worldwide

We start with the CI/SfB (Construction Index/*Samarbetskommitten for Byggnadsfragor*), a Scandinavian system of classification originally set up in 1959 and specially designed for the construction sector. It is claimed to be now in use worldwide for any technical and trade literature in the broad construction area. The CI/SfB was used in North America as the basis of the MasterFormat™, which is the specification-writing standard for most commercial building design and construction projects.

MasterFormat is a master list of numbers and titles intended for use in the organizing of specifications, and contracting and procurement requirements initially started with 16 *divisions* coded with 5 digits. In order to cope with the changes required by the modern Construction industry, in 2004 MasterFormat was heavily updated; new sections were added (the initial 16 were extended to 50) and the number codes are composed now of 8 digits (instead of the initial 5). MasterFormat targets the standardised communication of projects for all actors involved.

MasterFormat works together with Uniformat. Uniformat is an arrangement of construction information based on physical parts of a facility called systems and assemblies, aiming to: (I)achieve consistency in economic evaluation of projects; (ii) enhance reporting of design program information; and (iii) promote consistency in filing information for facility management, drawing details and construction market data. Masterformat tells **what** the construction item is, whilst Uniformat says **where** the construction item is.

Beginning in the 90's there is the British Standard 6100 (BS6100, the pioneer in UK); this is a glossary of the terminology used in the UK Construction sector, aiming to provide *a comprehensive list of terms that will promote better understanding between various sections of the construction industry, facilitate trade and provide better tools for improving handling of information.*

The use of **BS6100** was combined with the Unified Classification for the Construction Industry (UNICLASS, published in 1997 as a substitute for the widely accepted but increasingly out of date CI/SfB), which is a Construction-specific information classification system that covers information generated from all phases of a construction project. Uniclass is structured with a faceted classification system rather than a hierarchical one.

The **ISO 12006**[3] family (part 2 and part 3) came from another level of concern: the International Organisation for Standardisation. ISO was also targeting the development of standard CVs for the Construction sector in a world-wide scale. On one hand, ISO12006-2 targeted the definition of a model for classification systems (it is not a classification system in itself); rather it sets out an approach whereby particular classification systems that meet regional or national requirements can be developed according to a common international approach. On the other hand, the ISO 12006-3 defines a schema for a taxonomy model, providing 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.

The ISO foundation work was adopted and used by some institutions around the world. Among them, we can cite Stabu (Netherlands), Edibatec (France), and the Norwegian construction industry, which respectively started their own implementations of ISO-based tools, namely the **LexiCon, SCD, and BARBI**. In other words, the three of them are independent implementations of dictionaries that are compliant with the specification given in ISO 12006-3.

Next we talk about the International Alliance for Interoperability (IAI) and its Industry Foundation Classes (IFC). The IFC model has been progressively developed by the IAI since 1995 through several releases implemented in software for data exchange and sharing across applications. Since the IFC.2x release (October 2000), a core part of the model has been protected against change and formally accepted as ISO PAS 16739 in November 2002 under the external "harvesting" procedures of ISO TC184/SC4. IFC is the IAI vehicle aiming to promote and support the implementation of the concept of a Building Information Model (BIM) to increase the productiveness of design, construction, and maintenance operations within the life cycle of buildings.

The IFC model is rooted in approaches initially developed within the work of ISO TC184/SC4; most particularly in the development of the ISO 10303 series of standards (STEP q.v.). In particular, IFC has adopted and/or adapted certain parts of the STEP standards including: formal specification of IFC is in the EXPRESS language from ISO 10303 part 11; encoding of files for data exchange is undertaken using ISO 10303 part 21; and the IFC model uses schema that have been adopted from the resource standards within ISO 10303, particularly parts 41, 42, 43 and 46. Despite the fact that from a "semantic perspective", the IFC model per se cannot be considered ontology/taxonomy, part of it has been used to support reasoning and to exchange meaningful pieces of information among different software tools.

Many others European projects (research-oriented, standards-biased, etc.) were performed after this. A brief xlist includes: eConstruct [1], e-COGNOS [2], CEN/ISSS eConstruction series of Workshops [3], FUNSIEC [4], and the on-going CONNIE [5], and SEAMLESS [6] projects.

[3] It is worth noticing that both standards were officially relased by 2001, after the normal "evolution" through the standardisation channel (PAS-Publicly Available Specification, DIS-Draft International Standard), altough they have started been used before that.

The eConstruct project developed the *Building and Construction eXtensible mark-up Language* (*bcXML*), which supports the eBusiness communication process needed between clients, architects and engineers, suppliers and contractors for the (e)procurement of products, components, and services. The *bcBuildingDefinitions*, the taxonomy developed by eConstruct to show the power of bcXML, contains nearly 3000 terms specifically related to *doors*, expressed in six languages[4]. Such taxonomy can be instantiated to create catalogue contents or the actual requirements (queries) and solutions (answers) messages.

The e-Cognos developed a KM-oriented software infrastructure enabled by a semantic pillar: an ontology server (and its respective ontology). Such ontology focuses on construction concepts as they were related to e-COGNOS main objective: **consistent knowledge representation of construction knowledge items**. The e-COGNOS ontology is composed of two taxonomies (concepts and relations). The taxonomy of concepts is grounded on the IFC model, which was used to form its highest levels, according to the following motto: *In the context of a Project, a group of Actors uses a set of Resources to produce a set of Products following certain Processes within a work environment (Related Domains) and according to certain conditions (Technical Topics)*.

The CEN/ISSS eConstruction series of workshops worked towards the standardisation (or as close as possible since CEN means *European Committee for Standardisation*) in which the required semantic themes were also formulated. This initiative recognised that it is not possible to propose standardised *Semantic Resources* (SRs – i.e., ontologies, taxonomies, dictionaries, thesauri, and the related resources) for the construction sector but that it was possible to recommend what organisations could do after deciding to use SRs to support their business activities. Additionally, this initiative emphasised the need to take into account two key parameters, namely *purpose* and *application* areas when considering development and/or use of SRs.

The FUNSIEC project worked with the following question: is it possible to establish semantic links (mappings) between different SRs? And if yes, is it possible to evaluate how good are they? The first answer is yes, it is possible, which was demonstrated through the results of such projects: the OSIECS Kernel, and both OSIECS meta-model/model. The former is a software tool built to identify and propose semantic mappings between two SRs. OSIECS meta-model/model are the mapping tables produced by the OSIECS Kernel.

The CONNIE project tackled the problem of exploitation of multi-lingual content representing norms and regulations for the European Construction sector. It produced a software infrastructure to help organise, index, classify and use (in a pan-European way), the contents (regulation/norms) currently available within the CONNIE environment. This infrastructure strongly relies on the use of CVs in order to index and share the use of multi-lingual contents in an efficient way.

The SEAMLESS project targets the deployment of a seamless infrastructure to help SMEs to participate more easily in the e-business world (i.e. providing e-services to support business needs, such as procurement, production follow-up,

[4] : English, French, Dutch, German, Norwegian, and Greeklish (Greek language written with Latin characters). Additional information about e-Construct can be found at (Lima 2003).

etc.). The SEAMLESS infrastructure has been developed in a sector-independent way, but in order to demonstrate its potential two vertical sectors were selected: Textile and Construction. The knowledge-related side of SEAMLESS is based on a hierarchy of ontologies covering three levels of representation, namely: the global level (the whole SEAMLESS environment), the mediator level (the intermediate level providing a mapping between the global level and the SMEs), and the local levels (the lowest level where the SMEs are placed with their small CVs). In order to support the operation of the SEAMLESS environment, a sector-specific hierarchy of ontologies is being developed.

Currently, North America is attempting to bring classification within a single, multi-facetted approach called Omniclass, which started under the name of Overall Construction Classification System (OCCS) in 2001 and was renamed to Omniclass in 2002. It is based on ISO12006-2 as a framework and it uses MasterFormat for work results, UniFormat for elements, and Electronic Product Information Cooperation (EPIC) for structuring products. First version of Omniclass 1.0 was officially released by March 2006.

Last but not least, we complete the picture by referencing the ICONDA Terminology and the Canadian thesaurus. The former is the CV supporting the operation of the whole ICONDA@ family of products (e.g. the ICONDA database holding technical information on Construction problems). The latter is a bi-lingual thesaurus specifically created to represent construction terms in English and French. The enrichment of this thesaurus has been re-launched and new developments and improvements are expected in the near future.

4 Where we are now and where we could/should go

The work and initiatives described in the previous section allows us to say that we are in a good position but we have not achieved what we have been looking for. Companies are not yet sufficiently capitalising on the results provided by the research world and standardisation still needs to find its place in this arena. However, we should not be pessimistic: very good work has been produced, and solid results are now available. Education is the key word behind what needs to be done in order to push things forward.

The assessment of the results produced by FUNSIEC emphasised the importance of **education** (in the large sense) of the practitioners from Construction regarding the use of semantic resources. CEN/ISSS eConstruction workshop suggested the same approach. Education here means providing good practice examples to the final users showing how they can benefit from the use of CVs in their daily business, how they can expand their capabilities and potentialities in terms of market, what are the tangible benefits/improvements CVs can offer to them.

The authors keep working on this field and the latest experiences show that although good and powerful contributions are already in place, every time new research starts, people want to develop new things and to break with the past is a way to show innovation. This is a very natural human behaviour and, to the best of

our knowledge, quite difficult to change. For instance, in a new IST project[5], where CVs are required, the development team finds arguments to justify the development of 'yet another' ontology editor and a new tool to produce semantic mappings. What is behind this behaviour? Is it the need to differentiate the from previous ones? The need to leave 'fingerprints' on this area? Authors do not know; what they know is that recommendations from standards-related initiatives, like the CEN/ISSS eConstruction workshops, are not really being taken into account. These recommendations talked about 'analysing what is available', 'reusing current results', etc., which is definitely relegated to a second place. What matters is to propose new ideas, develop new things and try to be innovative & revolutionary, even when only keeping pace and going nowhere!

Business initiatives, even supported by less advanced solutions, are pushing things forward. IFCs have catalysed the adoption of the BIM concept and, considering the fact that part of the IFC model can be considered as an ontology, new experiments on the area have been launched and we can wait for solid results very soon. Other more modest initiative but also very useful (e.g. ICONDA family, CSTB products) are making money using very dedicated CVs, which for some people is more than enough. But it is not for some others. ICONDA is trying to push their 'semantic side' to something more modern and supported by new technologies and CVs. For instance, the ICONDA agency has started agreements with several countries around the world in order to enrich its terminology; CSTB has started an internal project to extend the capabilities of dictionaries and taxonomies supporting the search process of content-based products, such as their CD-REEF and I-REEF), and so on. Both examples are also looking very closely to the standard-related initiatives aiming at capitalising on them.

5 Conclusions

Communication is about exchanging signs. Humans are able to use words, body gestures, images, etc.. Jargon is often used inside a given community and those not belonging to this community *will* have problems to communicate. If we are to be clear and unambiguous, we must 'control' the vocabulary we are using in communication. Only parties knowing the words and their meanings are equipped to engage in communication free from misunderstanding. When it comes to computer-based communication, this is even more crucial since computers cannot establish dialogues in order to know elucidate precisely 'what is meant by that'. The conceptual approach to handle this situation often relies on the adoption of formal CVs (as much as possible) which help define the universe of discourse (the working context) of those involved in the communication process.

Several examples can be found around the world, coming from very different initiatives ranging from industrial support to feasibility projects funded by research programs. Results are emerging; education is gaining a new status in the European scene for several reasons, including European policies, businesses profit, and natural evolution of the area. LexiCon and BARBI (two implementations of ISO

[5] Authors are intentionally avoiding to identify the initiative, for obvious reasons.

12006) have joined forces; IFCs are becoming the standard supporting the inevitable BIM concept; IFD is attracting worldwide attention, and governments have published policies that directly or indirectly enforce the adoption of shared CVs and semantic-related resources. This is the future path we are called to – with no acceptable excuses and no other choice in moving forward.

Recalling McGuiness (and adapting her sayings to our context), *an ontology (or CV) is required when there is a need to communicate/exchange (transfer and/or share) various sorts of information where the meaning is fundamental. Ontology (CV) is also useful when the reuse of existing knowledge is required. From a non-exhaustive list of uses, an ontology (CV) can be used for simple kinds of consistency checking, interoperability support, validation and verification testing, configuration support, help to perform structured, comparative and customised search as well as to exploit generalisation/specialisation of information* [7]. This means whenever we must communicate precisely, our vocabulary must be controlled, our jargon must be shared and meaningful, and our semantics must be refined for the sake of the communication process. This is the mission behind the development and use of CVs in the Construction sector. This is the justification for proposing, developing, and assessing CVs. This is the quest that keeps the authors of this paper involved in this field. Results are still in their infancy, but they are promising and exciting, and hold the key to the future.

6 References

[1] Lima C, Stephens J, Böhms M. The bcXML: Supporting eCommerce and Knowledge Management in the construction industry. Itcon Journal, v. 8, p. 293-308, 2003.

[2] El-Diraby T, Lima C, Fiès B. Domain Taxonomy for Construction Concepts: Toward a Formal Ontology for Construction Knowledge, Journal of Computing in Civil Engineering, Vol. 19, No. 4, October 2005, pp. 394-406.

[3] Böhms M, Lima C, Storer G, Wix J. Framework for Future Construction ICT, International Journal of Design, Sciences & Technology, Volume 11, Number 2, p. 153-162, editor: Dr. Reza Beheshti.

[4] Lima C, Silva C, Sousa P; Pimentão JP, Duc CL. Interoperability among Semantic Resources in Construction: Is it Feasible? In proceedings of CIB / W78 22nd Conference on Information Technology in Construction, p. 285-292, ISBN 3-86005-478-3, CIB Publication No. 304, Dresden, Germany, July 2005.

[5] Cerovsek T, Gudnason G, Lima C. CONNIE - Construction News and Information Electronically. Joint International Conference on Computing and Decision Making in Civil and Building Engineering, Montreal, Canada, 14 - 16 June 2006.,

[6] Lima C, Bonfatti F, Sancho S, Yurchyshyna A. Towards an Ontology-enabled Approach Helping SMEs to Access the Single European Electronic Market, In proceedings of the 13th ISPE International Conference on Concurrent Engineering: Research and Applications, 18 - 22 September, 2006, Nice, France.

[7] Mcguiness D. Ontologies Come of Age, In Dieter Fensel, J im Hendler, Henry Lieberman, and Wolfgang Wahlster, editors. Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential. MIT Press, 2002.