JIS

*Article*

# The INNOVance Lexicon: Organization of Terms and Concepts about Construction Products

Journal of Information Science
1–12

DOI: 10.1177/0165551510000000
jis.sagepub.com
SAGE

**Maria Teresa Guaglianone**
Construction Technologies Institute of the National Research Council, Italy

**Giovanna Aracri**
Construction Technologies Institute of the National Research Council, Italy

**Elisabetta Oliveri**
Construction Technologies Institute of the National Research Council, Italy

**Abstract**
The construction sector is also a knowledge-intensive domain, in which effective and unambiguous communication and knowledge sharing are, at the same time, both essential yet difficult to accomplish. This is primarily due to the several professionals interacting and facing situations involving diverse resources, processes and activities. Each of them brings a different background and perspective, often generating poorly integrated information. Knowledge Organization Systems (KOSs) are crucial for ensuring completeness, consistency and quality of information. Despite the international trend to encourage the development and use of controlled vocabularies, especially classification systems, until recently in Italy the national coordination policy has not been effective enough. This paper describes the first national attempt made, the INNOVance Lexicon that collects and organizes knowledge about construction products. It combines taxonomic, terminological and semantic aspects of knowledge and it is a reference language to support information exchange and sharing in collaborative context.

**Keywords**
Knowledge Management; Knowledge Organization Systems; Construction domain; INNOVance Lexicon.

## 1. Introduction

Language is both the major means of communication, because it reflects the current terminology use, and one of the main problems in communication and information exchange, because of its ambiguity: it is made up of words that may have different meanings derived from the context of use (surrounding context) and the same concept can be specified through a set of different synonyms. These issues, inherent to natural language usage, can implicate difficult to manage consequences above all in specialized domains, such as data loss, ambiguity and lack of consensus on meaning in information exchange. In order to ensure proper communication, it is important to develop controlled vocabularies, a limited set of standardized descriptors, subjected to terminological and semantic control, which mediate concept comprehension and use.

The new digital environments have enhanced the role of these tools, properly named Knowledge Organization Systems (KOSs) [1], because they are efficient and fundamental not only for document indexing and retrieval but also on account of the fact that they have become essential for the entire information life cycle (creation, organization, preservation, diffusion and reuse).

In the construction domain, as well as in other manufacturing sectors, the community of professionals and the institutions are working towards digitising the process and redefining the information flow, through the integration of

**Corresponding author:**
Maria Teresa Guaglianone, Construction Technologies Institute of the National Research Council, 20098 San Giuliano Milanese (MI), Italy.
**Email:** guaglianone@itc.cnr.it

tools able to support the management of the entire building process and the information generated throughout the various phases [2]. Taking into account the multiple resources, processes and activities available in this specific field, the opportunity to have a common and coherent data management and solutions could reduce or eliminate most of the current issues in the domain, such as longer production times and the use of additional off-budget investments. In 2004, the National Institute of Standards and Technology (NIST) estimated the costs of inefficiency and loss of opportunities as consequences of inadequate interoperability. Nowadays, interoperability is a challenge, it relates to both data exchange and transfer and information comprehension and reuse [3-4].

So, the well-being of the construction industry depends on the process streamlining, the integration of different skills and knowledge, the saving of time and costs, the quality assurance, the consistency and reliability of information exchange.

Already in 1940, issues about information standardization and procedure rearrangement were perceived as essential within the domain to support communication, information access and reuse and to promote cooperation among the stakeholders involved in the construction process. Indeed, that same year the CI/SfB Classification System was created [5], arising from the Universal Decimal Classification (UDC)[1] and the CIB Master List [6]. The scheme was mainly used for cataloguing publications and organizing technical information about the initial planning and design phase and the construction of facilities. This initiative was followed by many others.

Despite the international trend to promote semantic interoperability in digital contexts, in Italy there has not been an adequate national coordination policy for the implementation of the methodological and technological solutions resulting from the research activities of universities and research institutes, in cooperation with trade associations and the world of production [7].The use of information technologies to support information exchange is not effective, since most of the interactions are paper-based. In fact, the awareness concerning the importance of using suitable tools has grown with the approval of the new Italian Public Procurement Code, which enhances the possibility to use tools and interoperable platforms to optimize the design and the check process.

This paper presents one of these initiatives through the description of the INNOVance Lexicon, implemented in the framework of the project INNOVance[2] that provides a Common Data Environment (CDE) [8] used by several stakeholders to manage all project information [9]. After a brief overview of some related works, developed internationally, the INNOVance Lexicon is described along with the methodological choices which led to its construction.

## 2. Related Work

Using KOSs for managing terminology and organizing knowledge in a specialized domain is an extremely active subject of debate. Issues concerning knowledge organization and management have become preeminent also in the construction domain. Therefore, the need to include KOSs in Information Technology (IT) tools is clear, in order to support information access, retrieval and sharing and to improve interoperability.

Before exploring some of these terminological and semantic resources, it is important to mention the work started in 1999 by the ISO/TC 59/SC 13 - Organization of information Working Group aimed at defining a framework to ensure terminology harmonization and knowledge organization in digital data sharing and exchange. The results of this work are represented by:

- ISO 12006:2015 – Building construction – Organization of information about construction works - Part 2: Framework for classification. It explains how to create a classification system and proposes the relevant elements that should be considered for implementation. The model takes advantage from the faceted classification, therefore, the most important construction entities are organized in tables, representative of specific and persistent features of the domain. The model intends to support experts in developing coherent KOSs and terminologies able to express the variety of language.
- ISO 12006:2007 – Building construction – Organization of information about construction works - Part 3: Framework for object-oriented information exchange. It proposes a data model, declared in EXPRESS [10], to specify objects, collections and properties. EXPRESS is used to develop vocabularies and give information about the building life cycle.

The ISO 12006 tries to enhance interoperability in the domain and suggests a flexible and intuitive approach that consists in structuring objects involved in the complete life cycle of construction works, by combining multiple

dimensions, considered as jointly exhaustive categories. This method, particularly suitable for digital environments, allows for a more efficient information management and communication.

As disclosed in Lima et al. [11] and summarized in Biscaya [12], different existing KOSs have been developed on the basis of these guidelines. The main scheme for the implementation of the principles of the ISO 12006 is OmniClass™ Construction Classification System (OCCS) which offers a comprehensive semantic representation of the domain. Its development started in 2000 and it follows the OCCS Development Committee recommendations [13]. OmniClass™ provides a classification scheme about the North American Architectural, Engineering and Construction (AEC) Industry by organizing concepts related to the whole construction process. It is made up of 15 tables in which entities are hierarchically organized: tables and entities correspond to categories (facets) and concepts, respectively, according to a multi-dimensional logic. Entities can be considered on their own to describe simple information or in combination with others from different tables to identify complex information. The number of hierarchical levels is not the same in all tables, since it depends on the principle of specialization applied and on the subject area. For example, information concerning construction entities are organized according to both the function (Table 11 - Construction Entities by Functions) and the form (Table 12 - Construction Entities by Form), while information about construction products are classified considering only their function in a built unit (Table 23 - Products).

OmniClass™ uses a notation system that allows for an alphanumerical representation of concepts and ensures efficiency in information exchange and sharing. Moreover, the alphanumerical notation points out the collocation of each concept in a specific category and its semantic relationships with concepts belonging to other categories. The result is a list of classes and subclasses with the related number (Table 1).

**Table 1.** Example of hierarchical organization from OmniClass™ (Table 23 – Products).

| Numbers | Level 1Title | Level 2 Title | Level 3 Title | Level 4 Title |
|---|---|---|---|---|
| 23-17 00 00 | Openings, Passages, and | | | |
| 23-17 13 00 | | Windows | | |
| 23-17 13 11 | | | Window Components | |
| 23-17 13 11 11 | | | | Window Sections |
| 23-17 13 11 13 | | | | Window Linings and Boards |
| 23-17 13 11 15 | | | | Window Vents |
| 23-17 13 11 17 | | | | Window Frames |

In order to foster terminology consistency and an increasingly wider semantic coverage, OmniClass™ includes other existing KOSs, such as Uniclass (Unified Classification for the Construction Industry)[3], MasterFormat®[4], UniFormat™[5] Electronic Product Information Cooperation (EPIC) [13].

Uniclass is the first European response to the ISO 12006. It is a classification system published in 1997 and now available in the updated version Uniclass 2015[6], which is part of the UK government actions to promote innovation in the domain and the use of state-of-the-art methodologies, such as Building Information Modeling (BIM) [14]. Actually, Uniclass 2015 is the semantic hub of the BIM Toolkit projects, because it ensures terminological consistency and accuracy in managing information during the life cycle of construction works[7].

Considering the architecture, as well as OmniClass™, Uniclass 2015 consists of tables (classes) linked following a top-down approach: the largest class includes all aspects related to a project described in overall terms (Complexes) and it can be broken down into other tables, concerning the project and its implementation (Entities, Activities, Spaces/Locations, Elements, Systems, Products). The structure is expressed by a code that identifies tables, groups, sub-groups, sections and objects and shows how the concepts are organized (Table 2).

**Table 2.** Example of hierarchical organization from Uniclass 2015 (Table Ss – Systems).

| Table | Group | Sub-group | Section | Object |
|---|---|---|---|---|
| Ss_25 | Wall and barrier systems | | | |
| Ss_25_30 | | Door and window | | |
| Ss_25_30_95 | | | Window systems | |
| Ss_25_30_95_96 | | | | Window walling system |

This type of organization is inclusive, therefore the future additions do not modify the general structure and the relationships defined between pairs of concepts belonging to different tables.

MasterFormat® differs from the aforementioned classification systems, since it is used for most commercial building design and construction projects in North America to organize project manuals and cost information, and to relate drawing notations to specifications. It provides a list of numbers and titles classified by work results for structuring and managing data about construction requirements, products, and activities [15]. It is constantly updated according to the stakeholders' needs and the evolution of the built environment.

Another important KOS included in OmniClass™ is EPIC, which provides a classification about construction products. It was developed after an agreement among European countries to support the market evolution and the diffusion of electronic commerce. More specifically, it is a standard used to store and exchange information between several heterogeneous databases. The approach at the basis of EPIC is interesting because it explicitly considers the need of saving costs which may lead to their long-term reduction. The content of EPIC is part of the OmniClass™ Table 23 - Products and it has supported the improvement of some others such as, Table 41 - Materials and Table 49 - Properties. Actually, EPIC was an independent system for a short period of time (from 1994 to 1999) subsequently included in OmniClass™ and since then has never been individually updated.

As well as the previous KOS, also UNIFORMAT II [16] (the evolution of UniFormat™) implements classification principles for supporting the decision making process in each phase of the building life cycle on the basis of consistent information. UNIFORMAT II organizes elements as part of a building mainly considering their functional characteristics, suitable for some applications such as cost analysis and estimation, project description and evaluation, etc. Compared to the previous version, UNIFORMAT II not only deals with information concerning civil building, but also with civil infrastructures and other non-building features. In fact, UNIFORMAT II intends to offer a general overview about a building process by encouraging greater collaboration and sharing and by providing a detailed information description through the hierarchy granularity.

So far, we have illustrated some studies whose main purpose is to organize terms and concepts using classification principles by defining broad classes and identifying within them sub-classes according to common or dissimilar characteristics. Their target is to offer a systematized representation of the building domain for project teamwork (engineering, architects, designer, etc.) and less for external stakeholders, including customers. Instead, LexiCon [17], created and maintained by STABU, and the terminology developed within the CDCON Project [18], combine classification needs to those related to information dissemination by providing normalized descriptors, definitions, which explain the meaning of each concept considered within its hierarchical structure (i.e. Door: "Construction for closing an opening, intended primarily for access" in LexiCon), and relationships between terms.

With regard to the Italian construction industry, efforts concerning knowledge and information management are still not comparable to those made in other countries. The most relevant initiative is the UNI 11337:2009 Edilizia e Opere di Ingegneria Civile - Criteri di codificazione di opere, attività e risorse - Identificazione, descrizione e interoperabilità, issued by the Italian Organization for Standardization (UNI) to promote terminological harmonization, to define technical specifications for enhancing this specific domain, to encourage the production process. It represents the first attempt in defining criteria to support the identification and the description of individuals, objects and activities involved in the building process.

The purpose is to favour information interoperability throughout the entire building life cycle among people involved in the process. All entities (individuals, objects and activities), taking part in the constant improvement of the building process are identified by means of a Complex Name that is unique and unambiguous.

Currently, the only effort, which follows this tendency concerning the diffusion of best practices for information communication and sharing by means of controlled vocabularies is the INNOVance Project and, in particular, the INNOVance Lexicon as a project result.

Compared to most of systems presented, INNOVance Lexicon examines more specifically semantic aspects by offering at the same time a conceptual organization and further relevant information by explicitly expressing semantic relationships. Considering this aspect, it is quite similar to LexiCon, since both combine classification criteria and terminological principles about term control and normalization. This approach is more efficient, since human-based exchange of knowledge cannot exclude the semantic dimension in information structuring. Indeed, also according to Soergel [19], the need for classification has been perceived in other fields as well, going beyond the traditional application in libraries, in which the awareness that it is not just a problem of classification but also of terminology [19] has been steadily increasing. So, except LexiCon, other systems give information about the hierarchical structure and not about the concept explicit organization and meaning, which are essential to comprehend the conceptualization of a domain and to have harmonisation and consensus on terms and concepts: this is the novelty, as far as the national

framework is concerned, that the INNOVance Lexicon intends to bring to the reference domain, by clarifying the different types of semantic relationships which may exist and by specifying definitions and a set of information about the surrounding context, which are important to allow knowledge sharing.

Considering the general methodological approach, the INNOVance Lexicon is in line with the most relevant KOSs developed for other specialized domains (i.e. AGROVOC[8], EuroVoc[9], etc.) and it is compliant with the most current standard recommendations for controlled vocabulary construction and interoperability with other vocabularies.

## 3. INNOVance Lexicon: characteristics and methodology

The use of common controlled vocabularies is extremely important to create and have explicit and shareable knowledge representations in collaborative environments [20].

INNOVance Lexicon provides the shared terminology about construction products, most of which are under EU Regulation 305/2011, and at the same time, it proposes a systematic representation of a part of the reference domain. Its objective is to facilitate and ensure collaboration and comprehension during different types of interaction among professionals throughout the different stages of the building life cycle.

Thanks to the definition of semantic relationships between terms, the Lexicon can be defined as structured, since the semantic network in which terms are located is clear, while generally a lexicon has a linear presentation and the terms are alphabetically ordered. Its construction required the evaluation of the purpose and the scope, by taking into account the information environment and the potential users, and the preliminary analysis of the domain, by identifying and studying agreement upon concepts and relationships between concepts [21].

The methodology for the construction and the maintenance of the Lexicon is essentially based on manual intellectual analysis, even if it is an intensive and costly process. According to Aitchison et al. (2005) [22], such method is recommended when the knowledge organization activity deals with the terminology, because of the language dynamism and complexity [21]. Similarly to a thesaurus, the Lexicon, is fundamentally linguistic and conceptual in nature [23] and it is characterized by structural, semantic and terminological aspects. This approach is a combination of textual source analysis and direct interaction with domain experts who review the theme, suggest potential candidate terms and propose class arrangements [24].

The Lexicon construction required the collection of a set of candidate terms within the area of construction products, the terminological and semantic control and normalization of the identified terms, and the creation of term classes and relationships between terms. Therefore, the starting point of the methodology was the creation of a reference text corpus for term extraction, followed by the definition of criteria for the choice of preferred terms, identifying specific concepts. Immediately afterwards, the semantic relationships between terms were defined, with the consequential recognition of synonyms, and terminological records were created.

As known, a text corpus is a collection of documents created for supporting the analysis and the exploration of a certain language or a linguistic variant. The corpus contains different types of current, complete and authoritative documents about the construction sector (such as harmonized standards, trade association manuals, manufacturers manuals, etc.). Therefore, it is a homogeneous set, also called specialized (special purpose) corpus, since it is referred to a specific domain and based on a sectorial language.

Such a corpus (533 documents) is the source for the term extraction which is carried out prior to the selection of a set of candidate terms. The aim is to identify preferred terms among them, as focal points which collect all the information about the concept, and non-preferred terms, denoting terms equivalent to the preferred ones (variants, synonyms, etc.). Non-preferred terms provide an entry point from which the user may be directed to the appropriate preferred term [21].

The selection of candidate terms takes into account both a list of terms specified by domain experts and the set of terms manually extracted from the corpus, considering term frequency and its presence in more than one source. Terminology collected should be validated on the basis of different knowledge sources [21], and not only on the expert's conceptual view of the domain, not always representative enough if considered on its own.

The domain is largely supported by a set of standards and technical regulations, especially those focused on construction products under CE Marking[10]: the EU Regulation 305/2011 provides the rules that all European countries have to use for construction products marketing.

Sources of law have been fundamental, since in most cases they provide a hierarchical classification, useful for term selection and structuring. Therefore, an order of priority is defined among chosen sources and the degree of preference of a term is determined according to it. So, a term retrieved in sources of law is considered preferred over a synonym attested in other sources.

At the end of the term extraction phase, candidate terms are normalized, by attributing multiple terms (spelling variants, acronyms, abbreviations, etc.) to a single term, in order to increase the consistency of the search results, to reduce semantic ambiguity and to ensure accuracy of communication in the specific domain. Regarding the choice between singular and plural form, the former is preferred in order to meet the habits of the generic user in employing a dictionary. This preference is also due to the inclusion of the definition, aimed at favouring the usage of the Lexicon by both expert and non-expert users. Furthermore, the inclusion of the numerous compound or multiword terms widely used to identify construction products is part of the terminological control.

The normalized term set represents the input for the following phase of specification of semantic relationships between pairs of concepts. The definition of the hierarchical structure results from a combined approach, since it has been established in most cases on a broader-to-narrower basis and, in other cases, on a narrower-to-broader basis [21]. So, the tendency is to first identify the terms that represent the broadest classes, and then to allocate other terms to these classes on the basis of their logical relationships.

Terms are structured according to the internationally acknowledged principles laid down by the ISO 25964-1:2011 Information and documentation - Thesauri and interoperability with other vocabularies - Part1: Thesauri for information retrieval. Therefore, to ensure information access and reuse, the semantic relationships between terms are those typically used for thesauri (equivalence relationship, hierarchical relationship, associative relationship).

The equivalence relationship regards synonymy management to provide access to information through a wider and richer terminology. The similarity among terms needs to be supervised in the controlled vocabularies, in order to guarantee the terminology appropriateness without disregarding users' need for information. When dealing with several synonyms, it is necessary to analyse them and to select the preferred term (descriptor) to identify the concept, while the others will be considered non-preferred terms. The equivalence relationship is established only between the preferred term and the non-preferred ones, thus there are no relationships between pairs of non-preferred terms [25]. In the INNOVance Lexicon the equivalence relationship is specified through the label Synonyms, i.e. "finestra ad anta-ribalta" (tilt-and-turn window) and "finestra ad anta e ribalta" (tilt turn window).

As mentioned, INNOVance Lexicon offers a classification of concepts, arising from the definition of hierarchical relationships, which express the degree of subordination or superordination between terms belonging to the same hierarchical tree [26]. In fact, this relationship is established between a general concept, identified by the label Hypernym, and a specific concept, identified by the label Hyponym, i.e. "porta" (door) and "porta a battente" (swing door). The hierarchical semantic-based structure allows for specification of the concept meaning in the reference context, thanks to its subordination to a broader concept, which designates the semantic framework. Besides, users have the possibility to browse through the structure, choosing the level of detail to satisfy search needs.

The associative relationship is used to specify relationships between terms that share a strong semantic connection and [27] that are neither hierarchical nor equivalent. It is employed to link terms on the basis of the meaning and the context of use. This relationship is symmetrical [28] and it is specified by the label Related Terms, i.e. "porta resistente al fuoco" (fire resistant door) and "resistenza al fuoco" (fire resistance). In our case, it includes terms related to performance characteristics of the construction products, distinguishing the essential ones as described in Annex ZA of the harmonized standards. Vice versa, regarding the performance characteristics, the related terms include the construction products to which they can refer.

This organization considers two complementary levels of representation: a vertical structure (classificatory and taxonomic), based on the hierarchical relationships (genus-species and whole-part type), which link together semantically higher-level concepts with semantically subordinate concepts; a horizontal structure, essentially based on the implicit relationship that links two terms that share the same hypernyms, on the equivalence relationship between different variants or synonyms and on the associative relationship which mainly includes links of a thematic nature.

As mentioned previously, a terminology record has been defined for each term in the Lexicon which, in addition to semantic relationships, where possible, contains further relevant information that include:

- the domain to which the term refers;
- the source in which the term is documented;
- the definition of the term, if present in the source of reference;
- the correspondence of the term in other languages, if present in the standards;
- the annotations and additional information about term definition or use.

For an illustration, some examples of terminological records are described below.

The first type of information given by the record, as presented in Figure 1, is the fact that the term is considered preferred (TP) or non-preferred (NTP). In this case, the label TP identifies the suitable term to refer to a construction product, while the label TNP specifies synonyms, linked to the descriptor. A dedicated terminological record is created, irrespective of whether the term is preferred or not.


**Figure 1.** Example of equivalence relationship for the concept "outward opening casement window".

As shown in Figure 2, information specified in the record provide a systematization of construction products in categories and sub-categories: such a hierarchical structure is expressed by the organization of terms in hypernyms and hyponyms, which classify construction products by defining the family products and the different product types for each family, respectively.

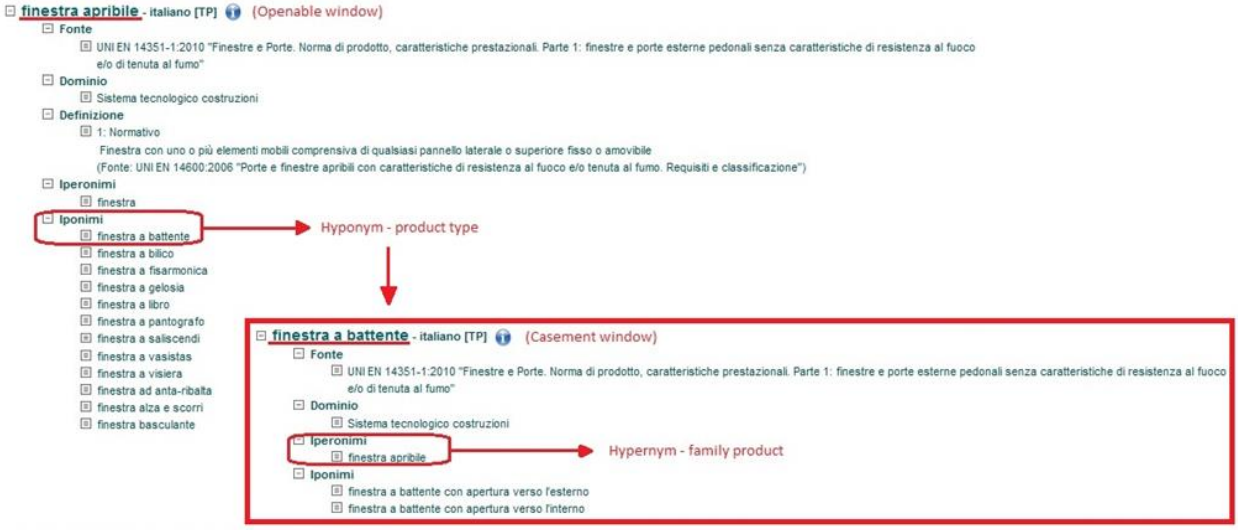


**Figure 2.** Example of hierarchical relationship between the terms "openable window" (BT) and "casement window" (NT).

Concerning the associative relationship, as mentioned, in the case of construction products under CE Marking, the performance and essential characteristics have been specified for each product. Since this kind of relationship is symmetrical, the terminological records about the performance characteristics include, among the related terms, the construction products to which they refer. Figure 3 shows an example of this kind of semantic relationship.

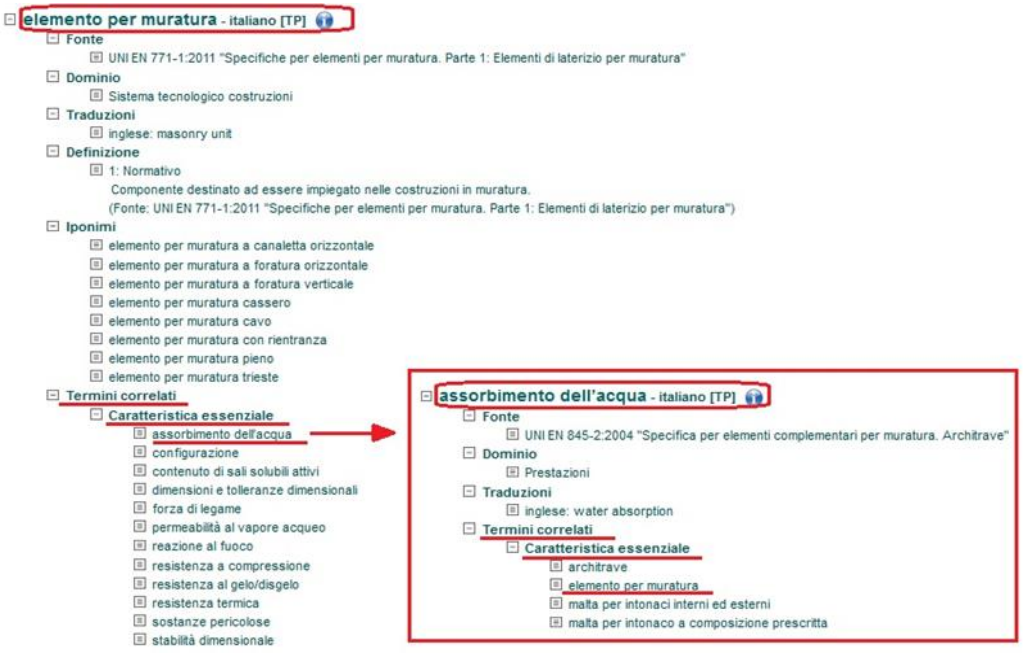


**Figure 3.** Example of associative relationship between the terms "masonry unit" and "water absorption".

Further, the terminological records we have illustrated provide information about the domain to which the term refers, the source in which the term is documented, the definition and the translation.

Figure 4 shows the distribution within the Lexicon of terms, distinguishing between Preferred and Non-Preferred Terms, and semantic relationships, specifying the number of relationships by type (hierarchical, equivalence and associative relationships).

| Total terms | | | |
|---|---|---|---|
| 921 | | | |
| Construction Product_terms | Performance_terms | Preferred terms | Non-preferred terms |
| 691 | 230 | 877 | 44 |
| Total relationships | | | |
| 1270 | | | |
| Hierarchical relationships | Equivalence relationships | Associative relationships | |
| 594 | 44 | 632 | |

**Figure 4.** Distribution of terms and relationships within the Lexicon.

The management of the Lexicon is supported by a web application based on open-source solutions (MySQL and PHP), created and maintained by ITC-CNR. Multiple users, once logged in, can simultaneously add, edit or delete terminological records and participate in vocabulary improvement by providing feedback. The collaborative nature of the system has made it necessary to develop a special section dedicated to user registration and profiling. So, each user

has a role which determines the access level or permissions in using system functions. All the operations are stored in the system and they are monitored by the supervisor who may decide to validate or reject proposals.

The web application allows to specify the status (validated, in progress, waiting for validation, to be deleted) of each term and its visibility on the web (yes/no). Only the user with the role of supervisor can give the final approval of terminological records, thus ensuring content quality and integrity [29].

The terms and the terminological records, can be displayed on the project portal[11] either by browsing the alphabetical and the structured lists or through the search mode, according to the user's level of specialization and knowledge about the sector.

As expected, the information need is expressed by means of the most familiar term and, in the case of a non-preferred term, it refers to the one designated as preferred, retrieving the information needed. This feature allows the lexicon to achieve terminological control and language harmonization, intrinsic functions of controlled vocabularies. The aim is to promote the coherent and unambiguous use of terminology, to enhance semantic interoperability and to avoid communication problems, while respecting and taking into account the common language used in the industry.

At present, the INNOVance Lexicon has not yet been used for indexing and retrieval. Since the most known measures for thesaurus quality assessment consider the success factors in relation to its usage, they could not be reasonably applied [30]. However, the conceptual organization, the level of detail and subject coverage, etc. of the Lexicon has been evaluated and improved thanks to the expert analysis; the presentation and layout have been revised and refined on the basis of the expert and non-expert feedback on usability and desired design features. Moreover, its characteristics have been compared with international standards, in order to guarantee its intrinsic reliability and reusability [31].

## 4. Discussion

In brief, by displaying the terminological record it is possible to: access the vocabulary also by means of a non-preferred term, linked to the preferred one; display and browse, for each term, the links to other terms; acquire a set of additional information. Therefore, such a terminological record allows for keeping track of the context in which a term is located, favouring disambiguation thanks to the conceptual organization offered by the semantic network.

The strength of the methodology employed, according to us, consists in the fact that the Lexicon is a hybrid (combined) resource, which presents both the advantages offered by a thesaurus and those offered by a lexicon and which is suitable for expert and non-expert users, as aforementioned and illustrated. On the one hand, the standardized method used for vocabulary construction and semantic relationship definition meets the international standards and gives the Lexicon the characteristics of terminological control and interoperability with other resources. On the other hand, the choice to use expressive and intelligible labels for specifying semantic relationships, the presence of term definition and the specification of additional relevant information give the Lexicon the features of effective usefulness and usability. Professionals may find important information about construction products and their normative requirements (i.e. to obtain CE Marking); at the same time, generic users, such as customers, may have at their disposal an informative and explanatory tool about construction products they are going to use or purchase.

An original aspect of the present work is the application context, since Italy has never before experienced such a terminological and semantic tool for the construction sector. The industry, until now, has been more focused on the product innovation than on the process innovation also based on the information as a key element for improving efficiency. The Lexicon, supported by a set of IT tools, could work as a reference tool and a source of documentation, useful to support several applications in knowledge management as part of the INNOVance portal, including semantic indexing, information retrieval and knowledge-based collaborative project development and management.

Although the INNOVance project has been completed, the work on the Lexicon continues. At present, the terms belong to 2 subject categories, Construction Products and Performances. The intent is to include new terminology and to increase the number of topic categories concerning spaces, resources, activities, etc. In addition, a recently undertaken activity concerns the improvement of interoperability by means of the application of semantic web best practices, which allow for interactions within the domain and across the different domains. In this process, the identification of a persistent namespace is necessary in order to make the Lexicon uniquely identified and its features clearly represented on the web. This way, each concept will be recognized by a Uniform Resource Identifier (URI) and will be understandable by humans and machines. This process considers both technical and semantic interoperability issues by using common formats and standards and by following the recommendations concerning the establishment and the maintenance of the relationships with other KOS, in accordance with ISO 25964-2:2013 Information and Documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies.

The specific aim is to improve the semantic dimension of interoperability, which is the major challenge for supporting information quality and consistency by matching coherent information.

First of all, the specification of external mappings has concerned the alignment of the terminological record structure to the Simple Knowledge Organization System (SKOS) Core model, the format used to publish KOSs on the web in a machine-readable and machine-understandable form. At present, the elements involved in this process are more specifically those that carry out semantic functions: synonyms, hyponyms, hypernyms, related terms and definitions. The result of the comparison of the two models is shown in the following Figure 5. The correspondence between the two schemas is evident, though SKOS provides a machine-understandable format, useful to enable the publishing, the access and the re-use of such KOSs via web by anyone and everywhere.

| INNOVance Lexicon | SKOS Core model |
|---|---|
| Synonyms | skos:altLabel |
| Hypernyms | skos:broader |
| Hyponyms | skos:narrower |
| Related Terms | skos:related |
| Definition | skos:definition |

**Figure 5.** Mapping between the terminological record structure and the SKOS Core model.

To improve the interoperability of the Lexicon with other general or specific KOSs, it is necessary that its original schema is compliant with SKOS and with the ISO 25964 conventions. The procedure of adaptation to the Semantic Web standards and principles is at the initial stage and the most suitable long-term strategy to support the formal representation of the Lexicon is under evaluation. The best approach to the alignment to external KOSs needs to be evaluated, even if it could be convenient to begin the mapping with resources available in Italian and to extend the external matching in the future to other general or domain-specific KOSs both in Italian and in English. The aim is to ensure the communication also in an international context, in part supported by the current specification of the English translation.

## 5. Conclusions and future perspectives

In conclusion, the present work is an important effort at the national level in Italy to develop a controlled vocabulary and to facilitate the achievement of semantic interoperability in the construction domain. This challenge can be tackled by encouraging the inclusion of such terminological resources in everyday use, especially in combination with IT tools, to overcome difficulties in managing and organizing information. The INNOVance Lexicon allows to organize terminology and preserve knowledge, as well as to discover implicit relationships between concepts. Its structure can be equalized to that of a thesaurus, so it represents the conceptual basis for knowledge-based systems and for data definition and object hierarchies in software systems [19]. Moreover, it works as a semantic road map to the construction products field and it provides for relationships within the domain and with other domains: it maps the space of a concept, relates the concepts to the terms and gives definitions, providing guidance as a reference tool [19]. The purpose of the present work is to manage the lexical richness of the sectorial language and to achieve the semantic uniqueness in the technical communication. It implies that in the interaction among professionals and among professionals and generic users there is agreement upon term and concept meaning.

In the Lexicon both terminological and semantic aspects coexist.

> A thesaurus may be defined as a lexico-semantical model of a conceptual reality or its constituent, which is expressed in the form of a system of terms and their relations [23 p489].

So, it appears that the process of thesaurus construction is a simulation in a lexical form of realities and concepts or of a part of them, and of hierarchical and associative relationships between these realities and concepts [23].

As mentioned, INNOVance Lexicon evolves continuously. In fact, the future perspective is to increase the semantic structure and to include other knowledge areas about the complex construction domain. At the same time, the adaptation to the semantic web standards and the alignment with other terminological and semantic resources [6] is currently underway and will continue with a view to making it useful for further scopes. This upgrade is functional to digital asset management and knowledge integration, since it supports the realization of the digitalization process through systemic and integrated approaches and suitable strategies for information organization, preservation and exchange.

## Notes

1. http://www.udcc.org
2. http://www.innovance.it
3. https://www.thenbs.com/knowledge/the-new-uniclass-work-sections-table
4. http://www.csinet.org/masterformat
5. http://www.csinet.org/uniformat
6. http://www.cpic.org.uk/uniclass2
7. https://toolkit.thenbs.com/articles/classification
8. http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus
9. http://eurovoc.europa.eu/drupal/?q=it/node
10. http://ec.europa.eu/growth/single-market/ce-marking/index_en.htm
11. http://www.innovance.it/lessico

## Acknowledgements

This research is part of the project INNOVance, funded by Ministry of Economic Development within the Industrial Innovation Program – Industria 2015.

## References

[1] Zeng ML. Knowledge Organization Systems (KOS). Knowledge Organization 2008; 35: 160-182.
[2] Carrara G, Fioravanti A, Loffreda G, Trento A. Conoscere collaborare progettare. Teorie e tecniche e applicazioni per la collaborazione in architettura. Roma: Gangemi Editore, 2014, p. 287.
[3] Gallaher, M P, O'Connor A C, Dettbarn J L and Gilday L T. 'Cost analysis of inadequate interoperability in the U.S. Ccpital facilities industry, http://fire.nist.gov/bfrlpubs/build04/PDF/b04022.pdf (2004, consulted October 2016).
[4] Zhou Z, Goh Y M and Shen L. Overview and analysis of ontology studies supporting development of the construction industry. Journal of Computing in Civil Engineering 2016; 30:1-14.
[5] Vetriani G and Marolda G. Piano di classificazione PC/SfB, 1th ed. Milano: ITEC Editrice, 1983, p. 87.
[6] Ekholm A. Principles for classification of properties of construction objects. In: Agger K, Christiansson P and Howard R. (ed.) Distributing Knowledge in Building. Salford: CIB, 2002.
[7] Guaglianone MT, Aracri G and Oliveri E. Gestione della terminologia e organizzazione della conoscenza nella digitalizzazione dei processi nella AEC Industry. AIDAInformazioni 2016; 1-2: 181-188.
[8] Designing Buildings Wiki the construction industry knowledge base, 'Common data environment CDE', http://www.designingbuildings.co.uk/wiki/Common_data_environment_CDE (2016, consulted May 2016).
[9] Pavan A. 'INNOVance un progetto di filiera per la digitalizzazione (BIM) del settore costruzioni', Nuovi Processi 43, http://dedalo.assimpredilance.it/argomento/innovance-un-progetto-di-filiera-per-la-digitalizzazione-bim-del-settore-costruzioni (2016, consulted May 2016)
[10] International Organization for Standardization (ISO). Industrial automation systems and integration - Product data representation and exchange - Part 11: Description methods: The EXPRESS language reference manual. Geneve: ISO, 2004, p. 255.
[11] Lima C, Zarli A, Storer G. Controlled Vocabularies in the European Construction Sector: Evolution, Current Developments, and Future Trends. In: Loureiro G and Curran R. (ed.) Complex Systems Concurrent Engineering. London: Springer, 2007, pp. 565-574.
[12] Vieira Nobre Biscaya S, 'Coordination and management of information for construction design projects. A framework for Portugal', https://core.ac.uk/download/files/130/12800033.pdf (2012, consulted November 2016).
[13] Omniclass Construction Classification System (OCCS) Development Committee, 'OmniClassTM A strategy for classifying the built environment Introduction and user's guide', http://www.omniclass.org/pdf.asp?id=1&table=Introduction (2006, consulted May 2016).
[14] Ciribini ALC. L' information modeling e il settore delle costruzioni: IIM e BIM. Santarcangelo di Romagna: Maggioli Editore, 2013, p. 458.
[15] Construction Specifications Istitute (CSI) and Construction Specifications Canada (CSC), 'MasterFormat ® Numbers & Titles', http://csinet.org/numbersandtitles (2016, consulted May 2016).

[16] Charette RP and Marshall HE. UNIFORMAT II Elemental Classification for Building Specifications, Cost Estimating, and Cost Analysis. Gaithersburg: National Institute of Standards and Technology, 1999, p. 93.
[17] Woestenenk K. 'The LexiCon: structuring semantics', in Proceedings of CIB W78 Conference on Distributing Knowledge in Building, Aarhus, Denmark: Aarhus School of Architecture, 2002, pp. 241-247.
[18] Amorim S R, Peixoto L, Madeira LCM, Nunes . 'Establishing the basis for systems interoperability: the terminology challenge', in Proceedings of CIB W78 Conference on Construction Informatics Digital Library, Aarhus, Denmark: Aarhus School of Architecture, 2002.
[19] Soergel D. The rise of ontologies or the reinvention of classification. Journal of the American Society for Information Science 1999; 50(12): 1119-1120.
[20] Pauleen DJ. Cross-cultural Perspectives on Knowledge Management. Westport: Libraries Unlimited, 2007, p. 259.
[21] Schneider JW. Verification of bibliometric methods' applicability for thesaurus construction. Aalborg: Department of Information Studies, Royal School of Library and Information Science, 2004, p. 356.
[22] Aitchison J, Bawden D, Gilchrist A. Thesaurus construction and use: a practical manual. 4th ed. London: Aslib, 2005, 240.
[23] Miller U. Thesaurus construction: problems and their roots. Information processing and management 1997; 33(4): 481-493.
[24] Lancaster FW. Indexing and abstracting in theory and practice. 3th ed. London: Facet Publishing, 2003, p. 451.
[25] Broughton V. Costruire thesauri. Milano: Editrice Bibliografica, 2008, p. 351.
[26] Spinelli S, 'Introduzione ai thesauri', http://www.dfll.univr.it/documenti/OccorrenzaIns/matdid/matdid385115.pdf (consulted May 2016).
[27] International Organization for Standardization (ISO). Information and Documentation – Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval. Geneve: ISO, 2011, p. 152.
[28] Green R. Relationships in Knowledge Organization. Knowledge Organization 2008; 35: 150-159.
[29] Aracri G, Artese MT, Folino A, Gagliardi I, Oliveri E. Un glossario per la divulgazione del sapere nel fotovoltaico. In: Vellutino D and Zanola MT (ed.) Comunicare in Europa. Lessici istituzionali e terminologie specialistiche. Milano: EDUCatt, 2015, p. 115-134.
[30] Kless D and Milton S. Towards Quality Measures for Evaluating Thesauri. Communications in Computer and Information Science 2010; 108: 312–319.
[31] Owens L A and Atherton Cochrane P. Thesaurus Evaluation. Cataloging & Classification Quarterly 2004; 37: 87-102.