# Ontological Representation of Knowledge Related to Building Energy-efficiency

German Nemirovski
Business and Computer Science
Albstadt-Sigmaringen-University of Applied Sciences
Albstadt, Germany
nemirovskij@hs-albsig.de

Álvaro Sicilia, Fàtima Galán, Marco Massetti
Leandro Madrazo
ARC Enginyeria i Arquitectura La Salle
Universitat Ramon Llull
Barcelona, Spain
[asicilia, fatima, mmassetti, madrazo]@salleurl.edu

*Abstract* — **Within the research project RÉPENER, co-financed by the R+D+i Spanish National Plan, an information system to capture the energy-related data throughout the whole building life cycle is being developed. The purpose of the system is to provide improved quality information to the various stakeholders participating at the different stages of the building life cycle. This higher-quality information is derived from interlinking disparate data sources – proprietary and open – and from the application of mining techniques to the semantically modelled data. This paper describes the design and the most important features of the RÉPENER global ontology, which is the core component of the information system is being developed. The ontology embraces knowledge originated from three realms: canonical domain knowledge, praxis-related usage cases and energy-related data stemming from various sources. The ontological design process – which includes the acquisition, unification, extension, formal specification and evaluation of the knowledge – is presented as a case study on knowledge discovery and engineering.**

***Keywords-semantic web; ontology; taxonomy; information system; energy-efficiency; energy model.***

## I. INTRODUCTION

Due to rapid technological development and the imminent shortage of fossil energy resources required in nearly all technological areas, crucial decisions must be made in regards to the reduction of energy consumption. In recent years, particularly in the area of building construction, a great deal of meaningful data has been collected, the analysis of which can help in the decision-making process related to this domain. On aggregate, this data appears to be a real treasury for data mining and visualization methods, which might help to improve the energy performance of buildings. However, the available data is located in different data sources, heterogeneously structured and formatted. Thus, access to data in the right format and at the right time remains a substantial challenge for those who will develop services that help to improve the energy-efficiency of existing and planned buildings.

Two critical questions must be answered on the way towards the development of such services: 1) how to enable efficient querying over the entire space of distributed data, e.g., for the purposes of data mining; and 2) how to make the portfolio of all available data transparent to actors operating at each phase of the building life cycle – from the design to construction and operation. Subsequently, an information system addressing these questions and tasks should provide lookup, browsing and data-transformation facilities which operate over the entire distributed data space.

The purpose of the RÉPENER project [2] is to develop an ontology-based information system which supports decision-making processes and knowledge discovery by actors concerned with the energy management of buildings. In recent years, studies on data integration using ontologies have delivered substantial results. The main example which proves the feasibility of tasks solutions is the Linked Open Data project [1]. By September 2011, it had integrated 31 billion data records specified in the RDF format, which is the most popular language for the specification of ontology-related information.

This paper presents the design of the ontology which is a core component in the RÉPENER information system. A comprehensive project description can be found in [2]. An important feature of this ontological design is the conceptualization of the domain knowledge determined from three different perspectives: first, the perspective of actors expressed through use case specifications such as energy consumption analysis and prediction; secondly, the perspective of canonical domain expertise expressed through standardization approaches in the field of energy performance of buildings; and thirdly, the perspective of data access expressed through models of data sources (e.g., entity relationship models).

The balance of the paper is structured as follows: Section II is dedicated to the description of background and methodology; in Section III, the process of knowledge acquisition is explained; the implementation and the ontology coding details are described in Section IV; Section V focuses on the goals and method of the ontological evolution; and lastly, in Section VI, some conclusions are summarized.

## II. METHODOLOGY

### A. Role of ontologies in data integration

The term "ontology" has been used in computer science since the early 1990s. One widely acknowledged definition was given by Gruber [3]. Ontology is an explicit

 ISBN: 978-1-61208-240-0 

conceptualization of a knowledge domain whereby the basic ontology element, a concept, represents a term and its relationship to other terms from the vocabulary used in this domain. Therefore, ontological concepts are interrelated, e.g., "house" is a sub-concept of "building." Such relationships can be defined in the form of axioms, conceptual properties connecting concepts to each other or attributes connecting concepts to value domains, such as "integer" or "literal". A subsumption hierarchy of concepts interrelated by specialization/generalization or sub-concept/super-concept relationship lies at the core of the ontology. This is called taxonomy. Ontologies are formally specified using Description Logic formalisms, and are coded in machine-readable languages like OWL. These features make ontologies essential for the specification of vocabularies in such fields as natural language processing [4] and Semantic Web [5].

The general idea of using ontologies for the interlinking and querying of distributed data is based upon the property of ontological concepts to be represented by their instances. For example, the two records "residence of Nobel laureate in chemistry Carl Bosch" and "house, located in Schloß-Wolfsbrunnenweg 33, 69118 Heidelberg" can be specified as instances of one single concept titled "Villa Bosch". In this case, the semantic equality of these records becomes evident, not only for humans but also for artificial agents performing ontology-based information retrieval. Even if these records are stored in two different sources using different formats and data models, an artificial agent searching for occurrences of the concept "Villa Bosch" will be able to identify the concept/instance relations and retrieve both of them. Thus, the interoperability of heterogeneously structured data can be achieved by establishing references between data chunks and ontological concepts or, in other words, by revealing data semantics. This fact has made semantic modelling one of the most efficient technologies for the integration of distributed heterogeneously structured data. The Linked Open Data Project mentioned in the introduction follows a decentralized modelling approach based on this principle, and it uses shared identifiers (URIs) to interlink data distributed over the linked sources. Therefore, most open-link data sources are represented by an ontology and a single access point able to process queries formulated in a standard query language, e.g., SPARQL with respect to this ontology. However, the Linked Open Data approach faces two obstacles: 1) the structure of single data sources, i.e., the architecture of the corresponding ontologies, is usually unknown and, therefore, combining data stored in different sources requires discovering all sources where data may be located; 2) to discover data sources, one needs to interact with multiple endpoints offering a data querying interface [6]. In terms of openness and flexibility, such an approach works well. However, for the sake of efficiency and the completeness of the information thus retrieved, a centralized approach is preferable.

According to the centralized modelling approach, a single ontology is used as the main reference for all distributed data. The data of a single source either refers to concepts of the central ontology or to concepts of dedicated source ontologies univocally mapped to the central ontology, by which each concept of the source ontology corresponds to one of the central ontology. In such a system, agents query data sources in interaction with a single, central end point, whereby all queries use the vocabulary of the central ontology. The process of indexing and looking up the entire distributed data space constitutes an integrated service of the information system. In this context, Calvanese [7] described an information integration scenario in which source models are mapped onto a central enterprise model specifying the entire knowledge over the distributed knowledge space. This approach was followed by Doerr [8], using the term "core ontology" to refer to an integrative ontology similar to the enterprise model. Uschold [9] defined the global ontology as either an intersection of local ontologies -- given that it encompasses concepts, properties and axioms shared by local ontologies -- or as a *union of elements from all local ontologies in the case of an intended application of the global ontology as one which would reference the entire space of terms.* Calvanese [10] introduced a formal framework which facilitates the efficient querying of integrated data corpus in a centralized manner.

The RÉPENER project follows the centralized approach of ontology-based data integration, adopting the terminology of Uschold [9]. Accordingly, ontologies which specify the data located in single sources are called local ontologies, while the central ontology serving as a target for the mapping of local ontologies, being defined as the union of their elements, is called global ontology.

### *B. Related approaches for energy data*

Semantic technologies have already been applied to model energy information. However, and according to Keistead [11], "*there is not yet one widely used conceptualization for energy systems*".

However, there are ontologies developed in specific domains, such as in building usage and operation. Shah and Chao [12] created an electrical home appliance ontology which facilitates the occupant's awareness regarding energy consumption in the house. For the same purpose, a smart home knowledge base has been developed using semantic web standards [13]. Additionally, ontologies have been used in the process of designing a device platform to integrate different device standard models [14]. In this respect, semantic technologies have been applied for the purpose of ensuring the interoperability among device industry standards such as BACnet, KNX, LON, or EnOcean [15]. Ontology inference processes have been used to enhance a building management system based on ontology modelling [16]. More recently, Wagner proposed the semantic web as a foundation for the Smart Grid communication architecture [17].

Applications of semantic technologies to specific domains related to energy-efficiency in buildings – operation, interoperability, smart grid – are present in the literature, but they do not model the energy data generated by different applications throughout the building's life cycle. To our knowledge, one of the first attempts to model these data was carried out during the IntUBE project [18].

 ISBN: 978-1-61208-240-0 

### C. *General design strategies: collaboration and modularity*

The design of formally specified ontologies has been an object of research since the early 1990s. Two significant works in this regard were carried out by Gruber [3] and Uschold and King [19]. The former defines the properties of ontological knowledge representation for the purposes of the engineering sciences. The latter deals with the design process of ontologies, being described as consisting of four phases: identifying ontology purposes, building the ontology, evaluating and documenting. In turn, the phase of ontology building is subdivided into three steps: 1) ontology capture, namely, definition, naming and description of key concepts and relationships between them 2) ontology coding, that is, using one of the formal languages or tools and 3) integrating existing ontologies. This approach has been further elaborated in work on this topic. A survey of up-to-date methodologies for ontological design was provided by Contreras and Martinez-Comenche [20].

Already in the 1990s, it became obvious that ontologies designed for practical industrial or medical application could be large and complex. Therefore, to overcome the complexity of ontology management, two approaches have emerged: collaborative ontological design supported by dedicated multi-user environments, as shown in Swartout [21], Sure [22], or Tudorache [23]; and reusing ontology elements, e.g., design patterns, as shown in Presutti [24] and Gangemi [25], or ontology modules as discussed by Cuenca Grau [26].

In contrast to the classic procedure described by Uschold and King [19], the design process of the RÉPENER global ontology can be depicted as a sequence of iterations encompassing knowledge capture (conceptualization, concept naming, and description), and ontology coding using OWL 2 specification language and evaluation (Figure 1). Our approach followed this scenario. After each iteration, the ontology became more and more comprehensive.

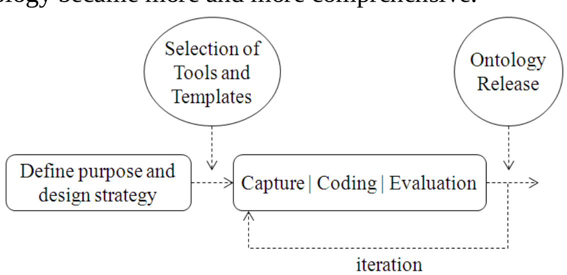


Figure 1. Design process of RÉPENER global ontology.

The design process involved energy-domain experts and ontology engineers working at different locations in Germany and Spain. In diligence style [20], different ideas and proposals were generated in a distributed way through tools like Adobe Connect, Skype and Google Docs used as a platform for the project development. An Excel document was used as an instrument to capture the domain knowledge from the different realms in order unify the terms and identify relationships between them. The resulting structure was the base of the ontological design process (see Section III).

Based upon the approach of modular ontological design, RÉPENER global ontology is built on certain selected modules of an upper-level ontology. In this way, each concept of the global ontology subsumes the concepts of the Suggested Upper Merged Ontology (SUMO). In this way, the foundational relationships and axioms which are valid for SUMO concepts remain valid for those of the RÉPENER global ontology. Hence, the philosophical, engineering and linguistic issues incorporated by the SUMO ontology have been inherited by the RÉPENER global ontology.

## III. Information Capture

### A. *Vocabulary acquisition*

In each design iteration, the knowledge capture was carried out by: a) keeping in mind the purpose of the RÉPENER global ontology, i.e., data management; b) taking into account the services to be performed by an information system for the energy-efficiency of buildings; and c) referring to the canonical knowledge structure of the domain of interest. This paradigm is reflected in the three-dimensional architecture of the term space (Figure 2), which became part of an informal knowledge specification aiming at determining the ontology vocabulary, including terms, relations data types and units of measure. One of the challenges in the vocabulary acquisition process is to avoid redundancy and terminology mismatching, which usually occur in the aggregation of heterogeneous information. To avoid this, a maximum number common terms for each dimension was identified.

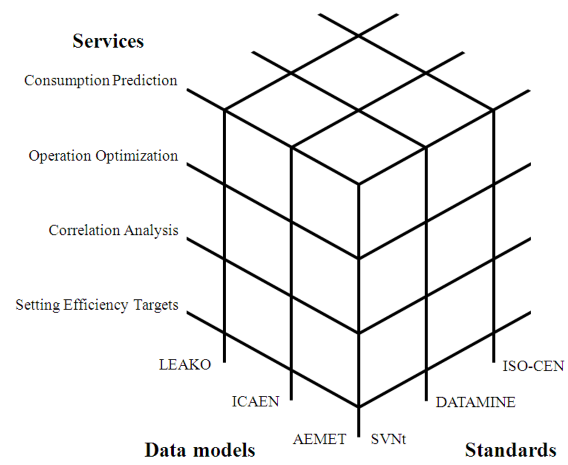


Figure 2. Three-dimensional architecture of the RÉPENER term space.

The three dimensions of the term space mentioned above are illustrated in Figure 2. The first dimension comprises data sources containing two kinds of energy information: building information (building systems, energy consumption, energy demand, etc.) and contextual data (economic context, demographic context, climatic context, etc.). In the initial project phase, data from three sources was used: a) a

 ISBN: 978-1-61208-240-0 

database of LEAKO, a Basque company handling installation, distribution, and HVAC control. The database contains consumption data for thermal (kwh) consumption for heating, hot water, gas and water consumption and indoor conditions, e.g., air temperature in several monitored residential buildings; b) a database of ICAEN, an organization of the Catalan government which gathers the energy certificates of newly planned buildings, including their simulated performance; and c) AEMET climate data from the Spanish Meteorological Agency. In this last case, the terms, relationships, data, and units of measure were extracted from the entity relationship models specifying the data sources.

The second dimension was built on the basis of standards and key parameters classifications, used to manage energy performance of buildings. The energy certification of buildings defined by DATAMINE project [34], the ISO CEN standards following the European Directive 2002/91/EC (e.g., ISO 13790:2008) and the Standard Network Variable Types from LonWorks (SNVTs), were utilized in the first two years of the RÉPENER project. The terms were extracted out of document texts and tables.

The third dimension comprises services addressing support to stakeholders in the realms of their decision-making processes (design, maintenance). The first group of prototypically developed services consists of: a) a prediction service launched in the design phase, whose goal is to provide qualified information regarding the consumption and demand of a building construction; b) an operation optimization service for building managers to optimize the building's behaviour based on the reference data obtained from other buildings; c) a correlation analysis service to identify the key factors influencing energy consumption; and d) a service for setting the energy targets to be reached in the refurbishment of the existing buildings.

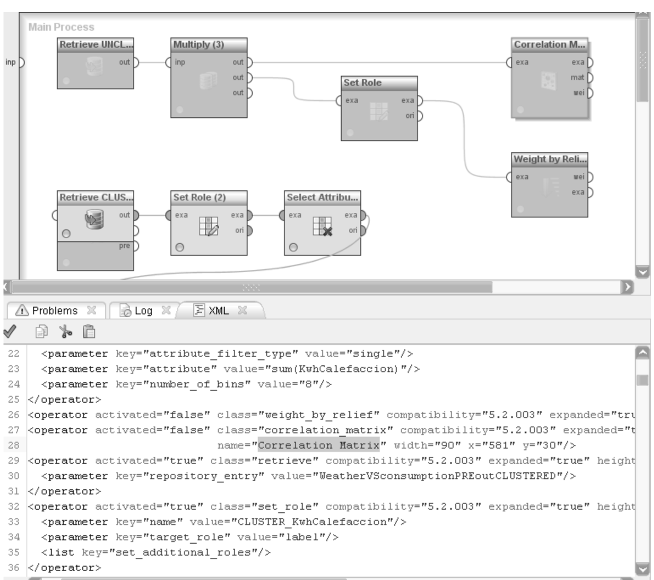


Figure 3. Data-mining process specification in RapidMiner software.

In this case, terms were extracted from the data-mining process specifications that were defined using RapidMiner software. In this software, processes are specified in XML and presented in a graphical editor, as shown in Figure 3. For obtaining terms for the energy model, a simulation of the above-described services took place by specifying the corresponding RapidMiner processes for propositionalized data from LEAKO and ICAEN databases. The terms were then extracted from the process specification (Figure 4).

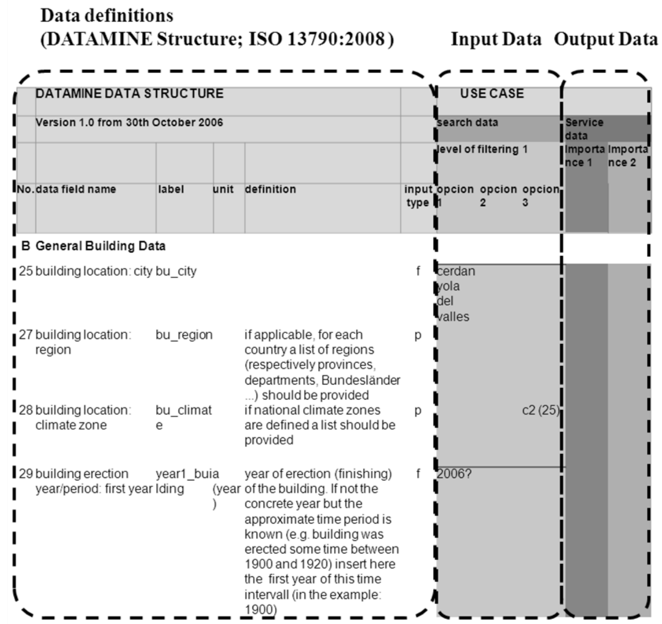


Figure 4. Mapping DATAMINE terms onto the input/output parameters of services.

The result of the vocabulary acquisition has been documented in a series of Excel tables implementing relationships within the three-dimensional term space, being transparent for all participants of the collaborative knowledge-capture process. The DATAMINE data structure, which includes energy certificate data, general data of the building, building envelope data, energy demand and/or energy consumption has been used as the primary source of the terms. Figure 4 shows how DATAMINE field names (in the right part of the figure) are mapped onto the input and output terms of the data-analysis services (titled as "use cases" in the right part of the table). Three tables of this type are required for mapping the three dimensions in succession.

## B. *Hierarchy of terms*

In Section III.A, it has been shown how the terms, which originated at different realms of the three-dimensional term space, are mapped onto each other for the purpose of identifying a common vocabulary. Such dimensional mapping represents part of the energy model, which is the first step in the process of creating a formal ontology.

The other part of the energy model is a hierarchy of terms unified by the mappings. Such a hierarchy has been specified by means of the relationships *contains/part of*. The top level of the hierarchy is made of the domain names,

 ISBN: 978-1-61208-240-0 

while the second level contains terms specifying sub-domains. This partitioning is extended up to the last hierarchy level, which contains terms associated with basic parameters such as envelope properties or heat-transfer coefficient.

Figure 5 shows the hierarchy of terms in a simplified form. The most important parts of the building energy domain, which is the core domain of the energy model, can be defined as follows:

- General project data: parameters which identify the project and define its generic characteristics such as location, use, project execution data, and site description;
- Performance: building performance indicators regarding energy use (energy demands, consumption of different energy carriers, e.g., gas or electricity, and different uses, e.g., heating, cooling, hot water, electricity and appliances), CO2 emissions and indoor conditions (e.g., temperature and humidity);
- Building properties: geometric characteristics, construction systems and building services;
- Outdoor environment: climate characteristics and conditions of the physical environment which determine the building's performance: outdoor temperature, wind speed and direction, and solar radiation;
- Operation: usage and management of the building and its facilities for maintaining comfort levels (e.g., solar protection and thermostat regulation). It also includes the effects of the occupant activity in the indoor environment, such as thermal loads produced by occupants, lighting and appliances;
- Certification: information associated with building energy certificates. It includes indicators to qualify a building based on performance, e.g., according to a conventional scale as (A, B, C, etc.). It also includes the certification-process methodology.

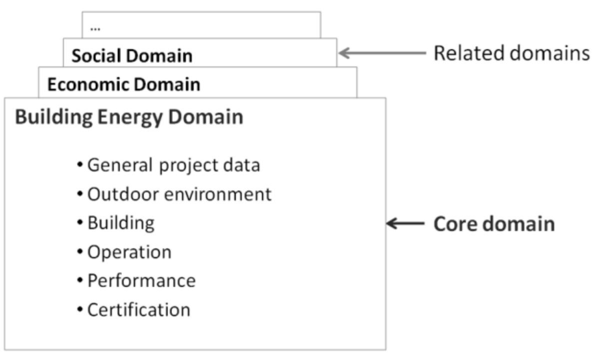


Figure 5. Energy model domains

Studies [27] have shown that the energy consumption of building correlates to socio-economic factors like real estate prices or the income levels of the inhabitants. To take this fact into consideration, we included the economic/social domain into the building's energy model along with the building energy domain.

## IV. ARCHITECTURE AND CODING

### A. Ontology Architecture

As stated in Section II.C, the RÉPENER global ontology uses the Suggested Upper Merged Ontology (SUMO) at the upper level. The selection of SUMO for this role was made after comparing it to other foundational ontologies, such as DOLCE, PROTON, General Formal Ontology (GFO) and Basic Formal Ontology (BFO). SUMO scored well in such fields as simplicity of understanding, applicability for reasoning and inference purposes, and potential reuse in the Building Energy Domain, for instance, reusing concepts for specifying units of measure defined by the SI system (meter, watt, joule, etc.).

Some of the SUMO concepts subsume concepts of the RÉPENER ontology. For example, the concept *Building* is subsumed by SUMO's *StationaryArtifact* and SUMO's *Attribute* subsumes *BuildingProperty*, which in turn subsumes *BuildingGeometry:*

$$\texttt{BuildingGeometry} \sqsubseteq \texttt{BuildingProperty} \sqsubseteq \texttt{Attribute}$$

The resulting RÉPENER global ontology is a combination of two hierarchies: one of them is the taxonomy based on the concept of subsumption, where the upper level of the taxonomy is represented by generic SUMO concepts. The second hierarchy consists of the terms described in Section III.B, whereby building elements of this hierarchy are aggregative *has* or *includes* properties such as the property *hasGeometry* (Figure 6). The former hierarchy (Figure 5) is required for the formal reasoning, while the latter one (Figure 6) represents the knowledge from the perspective of the domain experts and users.

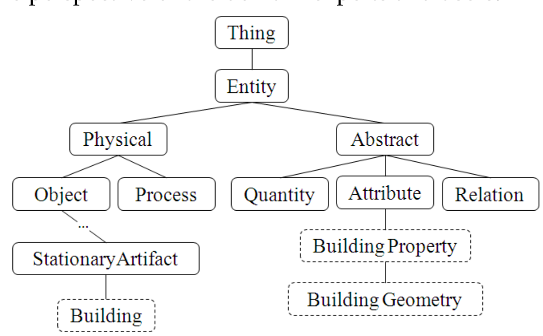


Figure 6. hierarchies as the basis structures of the RÉPENER global ontology.

### B. Coding

OWL 2 has, in recent years, become a sort of default standard for ontology coding. As shown in Calvanese [28], the use of this specification language in its full version may be disadvantageous in terms of the computability of particular reasoning tasks, particularly those which require

 ISBN: 978-1-61208-240-0 

conjunctive queries of large data volumes. Poggi [29] suggested a somewhat restricted *$DL\text{-}Lite_A$* formalism, which helps to overcome this obstacle. This approach was adopted by the RÉPENER global ontology.

A detailed description of *$DL\text{-}Lite_A$* formalism is out of scope of this paper. Nevertheless, it is important to mention two of the most important features of an OWL-dialect that implements *$DL\text{-}Lite_A$*: 1) domain and range of properties can be specified only for functional data properties; and 2) definition of an object property connecting two OWL classes with each other, has to be modelled by means of axioms and not by specifying property's domain and range. For example, two following axioms in DL notation use subsumption ($\sqsubseteq$), existence quantification ($\exists$) and inversion ($^{-}$ ) to express that the class BuildingGeometry relates to the class Building via the *hasGeometry* property.

$$Building \sqsubseteq \exists hasGeometry$$
$$\exists hasGeometry^{-} \sqsubseteq BuildingGeometry$$

In OWL the same is specified as follows:

```
<SubClassOf>
   <Class IRI="http://www.owl-
   ontologies.com/SUMO155.owl#Building"/>
  <ObjectSomeValuesFrom>
   <ObjectProperty IRI="#hasGeometry"/>
   <Class abbreviatedIRI=":Thing"/>
  </ObjectSomeValuesFrom>
</SubClassOf>
```

and

```
<SubClassOf>
  <ObjectSomeValuesFrom>
   <ObjectInverseOf>
     <ObjectProperty IRI="#hasGeometry"/>
   </ObjectInverseOf>
   <Class abbreviatedIRI=":Thing"/>
  </ObjectSomeValuesFrom>
  <Class IRI="#BuildingGeometry"/>
</SubClassOf>
```

Although domains and ranges of properties are not explicitly specified in the code, if an ontology specification is valid, they can be inferred by reasoner software to be then visualized and viewed by the user.

## V. EVALUATION

Apart of the already mentioned work of Gruber [3], different views on essential ontology properties are described by Gómez-Pérez, [30], Obrst [31], and Gangemi [32]. After a comparative analysis of these approaches, we found that the following three criteria are of primary priority for the RÉPENER global ontology:

- Completeness: in the RÉPENER context this means that all terms and relations of the three-dimensional space of terms are explicitly specified in the ontology code or can be inferred by reasoning.
- Intelligibility: the ability of actors using the ontology and ontology-based applications in their decision-making process to understand the ontology structure.
- Computational integrity and efficiency: the ability of the ontology to support reasoning tasks such as conjunctive querying on high efficiency level, i.e., with a comparatively short response time.

Brank [33] described four types of evaluation approaches: 1. comparing ontologies with a "golden standard", e.g., another ontology; 2. comparing ontologies with source data; 3.evaluating ontology application; and 4. evaluation by humans. In the RÉPENER project, we have followed three of these approaches: we compared our ontology with source data, evaluated it by humans and evaluated it through the application of reasoners.

1. Comparing ontologies with source data to evaluate ontology completeness: a set of randomly selected items, such as fields and table names from databases of LEAKO and ICAEN or terms from the DATAMINE classification, are (manually) mapped by testers onto the current version of the ontology. If the mapping result for one item corresponds to the mapping in the energy model (Figure 4), the resulting coefficient, initially nullified, will be incremented by one. Success is measured as a percentage where 100% corresponds to the number of preselected items. We have carried out ten evaluations of this kind, selecting twenty different terms from the above-mentioned sources. Six of the twenty terms could be identified in the ontology. Three evaluations ended with the score 18, and one ended with a score of 17. Therefore, the total completeness of the ontology was rated at 95.5%, resulting from the following calculation: $(20 \cdot 6+18 \cdot 3+17) \cdot 100 / (10 \cdot 20)$. However, taking into account the fact that two of nine missing terms were intentionally omitted from the ontology, the completeness would be 96.5%.

2. Evaluation by humans aiming at the quantification of intelligibility: independent testers (who did not participate in the design process) are given the task of navigating the ontology or, in other words, finding a concept. The navigation is carried out in an ontology viewer developed for this purpose. The shortest navigation path from the top of the concept hierarchy (depending on the task, it can be the energy model or the concept taxonomy) is calculated in advance. The result of evaluation is measured as a percentage, where 100% corresponds to the number navigation steps equal to the number of edges in the shortest path minus one and 0% to this number plus 30, i.e., if a tester needed 30 clicks above the required minimum, his score was set to 0. The evaluation was carried out by two groups of testers. One group contained eight computer science students, and the other group contained five experts in the field of building energy. Each tester was offered three terms to find in the ontology. The surprising result of this evaluation was that the average score of these two groups did not differ a lot. The intelligibility of the ontology for

 ISBN: 978-1-61208-240-0 

domain experts was 97.30%, while this metric for computer science students achieved the value 91.20%.

3. Evaluation of ontology application with the focus on computational integrity and efficiency: as stated above, in an ontology developed on the basis of $DL\text{-}Lite_A$, the domains and ranges of properties specified using axioms can be inferred by reasoners. However, this method does not provide a measure for the quality of the ontology. Instead, it demonstrates the coding or conceptualization errors which have to be treated immediately.

However, for practical reasons the time required to complete the reasoning tasks is an important matter of consideration. This time strongly depends on i) the expressivity of the DL-Language used to specify the ontology; and ii) the number of axioms contained in an ontology. Our evaluation has shown that the former factor may be crucial for the performance of reasoning, while the latter one has only a moderate influence. For instance, an attempt to integrate QUDT ontology modules specifying units of measure vaulted the time of reasoning carried out on a machine equipped with an Intel i7 2600 CPU and 8GB RAM to three hours. We believe the explanation for this was the highly expressive OWL-profile used for the QUDT specification. The reasoning time for RÉPENER global ontology using seven selected modules of SUMO upper-level ontology only (in this case, QUDT part was not imported) of a total size of 5.3 MB and containing 100 axioms as those described in IV.B achieved 1 minute 20 seconds on the same machine. When the number of axioms increased to 1,000 the reasoning time rose to 5 minutes 32 seconds (these measures are valid for the HermiT reasoner version 1.3.5). It should be mentioned that originally SUMO is specified using the KIF Knowledge Interchange Format (KIF) language, which has a high level of expressivity. When translated to OWL, however, SUMO modules lose many axioms which cannot be expressed in OWL one to one. Hence the translated version is on the $\mathcal{EL}$ level [33].

## VI. CONCLUSION AND FUTURE WORK

This paper has presented a case study on knowledge discovery, as was carried out in the context of a particular domain (Building Energy Performance) and aiming at the fulfilment of a particular task (development of an information system for the decision-making process support of stakeholders participating at different stages of a building's life cycle). Within the case study, we have shown several stages of the process of knowledge discovery and engineering:

1) *Vocabulary acquisition* from different realms related to the domain and to the task of interest: services exposed by the information system to be developed; structure of data sources to be integrated and canonical domain knowledge in form of standardization documents;
2) *Mapping of terms* onto each other for the purpose of defining a common vocabulary for all of the realms;
3) *Specification of relationships* between terms, which is an important step from the definition of a vocabulary towards ontological design: in the course of a relationship definition, a term became a concept.
4) *Building taxonomy* of concepts by integration the SUMO ontology as the taxonomy's upper level. At this step, the abstract knowledge based on philosophic, linguistic and engineering postulates, as discovered and constructed by a third party, became part of the ontology being constructed;
5) *Formal specification* of the discovered knowledge, i.e., the elaboration of this knowledge towards a formal ontology. This operation makes the new knowledge available for exploitation, particularly in the context of data management and decision-making support on which the RÉPENER project has its focus;
6) *Knowledge Evaluation*, using distinct criterion and methods.

This paper addressed issues related to knowledge discovery and ontological design, as were carried out in the context of the RÉPENER project aiming at development of an information system supporting the stakeholder in all phases of a given building's life cycle. Nevertheless, the paper did not address the implementation aspect of the information system. Neither can we argue if the evaluation of the ontology hereby presented can replace the evaluation of the information services, to be developed. This is specified in a separate paper [2], which presents further motivation and the context for the RÉPENER global ontology. However, we assume the existence of a strong correlation between the quality of the ontology and the quality of information services. The demonstration of such correlation will be the goal of further work. One of the most important tasks in this regard will be computability evaluation of the resulting information system using benchmarks for conjunctive queries addressing distributed data.

## ACKNOWLEDGEMENT

RÉPENER is being developed with the support of the research program BIA 2009-13365, co-funded by the Spanish National R+D+i Plan 2009-2012.

## REFERENCES

[1] T. Heath, and C. Bizer, Linked Data: Evolving the Web into a Global Data Space (1st ed.), Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 2011, pp. 1-136. Morgan & Claypool.

[2] L. Madrazo, A. Sicilia, M. Massetti, and F. Galan, "Semantic modeling of energy-related information throughout the whole building lifecycle," In Proceedings of the 9th European Conference on Product and Process Modelling ECPPM. Iceland, 2012.

[3] T. Gruber, “Towards principles for the design of ontologies used for knowledge sharing,” International Journal of Human Computer Studies, Vol. 43 (5-6), 1995, pp. 907-928.

[4] A. Kiryakov, B. Popov, I. Terziev, D. Manov, and D. Ognyanoff: “Semantic Annotation, Indexing, and Retrieval,” Journal of Web Semantics, 2(1) 2004, pp. 49-79.

[5] T. Berners-Lee, J. Hendler, and O. Lassila, “The Semantic Web. Scientific American”, 284, 2001, pp. 34-43.

[6] E. Oren, “Sindice.com: A document-oriented lookup index for open linked data,” International Journal of Metadata, Semantics and Ontologies, 3(1), 2008, pp.37-52.

[7] D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, and R. Rosati, “Description Logic Framework for Information Integration,” 6th International Conference on the Principles of Knowledge Representation and Reasoning (KR-98). Italy, 1998.

[8] M. Doerr, J. Hunter, and C. Lagoze, “Towards a Core Ontology for Informa-tion Integration,” Journal of Digital information, 4(1) 2003.

[9] M. Uschold, “Creating, Integrating and Maintaining Local and Global Ontologies,” Proc. of 14th European Conf. on Artificial Intelligence (ECAI'00), Germany, 2000.

[10] D. Calvanese, G. De Giacomo, D. Lembo, M .Lenzerini, R. Rosati, and M. Ruzzi, Using OWL in Data Integration, R. De Virgilio, F. Giunchiglia, and L. Tanca, (Eds.). Semantic Web Information Management, 2010 (Chapter 14, pp. 397-424), Springe-Verlag, Germany.

[11] J. Keistead, and K.H. Van Dam, “A Survey on the Application of Conceptualisations in Energy Systems Modelling,” In Formal Ontologies Meet Industry: Proceedings of the 5th International Workshop. Ed. P.E. Vermaas, V. Dignum, IOS Press, 2011, (pp. 50-62)

[12] N. Shah and K. Chao, “Energy Conservation Recommending Semantic Services,” In Alain Zarli (Ed.) conference CIB –W78, France, 2011.

[13] M.J. Kofler, Ch. Reinisch, and W. Kastner, A semantic representation of energy-related information in future smart homes. In Energy and Buildings, 47(0), pp.169 - 179, 2012.

[14] A. Noguero, N. Arana, and J. Martinez, “Enabling energy efficiency through device awareness using ontologies,” In Alain Zarli (Ed.) conference CIB –W78, France, 2011.

[15] K. Kabitzsch and J. Ploennings, “Ontology models and design patterns for building automation, ” In Alain Zarli (Ed.) conference CIB –W78, France, 2011.

[16] J. Han, Y. Jeong, and I. Lee, “Efficient Building Energy Management System Based on Ontology, Inference Rules, and Simulation,” International Conference on Intelligent Building and Management. Proc .of CSIT vol.5 IACSIT Press, Singapore, 2011.

[17] A. Wagner, S. Speiser, A. Harth, “Semantic Web Technologies for a Smart Energy Grid: Requirements and Challenges,” In Proceedings of the 9th International Semantic Web Conference (ISWC2010), 2010.

[18] H.M. Böhms, W. Plokker, B. Charvier, L. Madrazo, and A. Sicilia, “IntUBE energy information platform,” in 8TH European Conference on Prodcut and Process Modelling ECPPM, Ireland: CRC Press, 2010, pp. 339-344.

[19] M. Uschold, and M. King, “Towards methodology for building ontologies,” Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95, Canada, 1995.

[20] J. Contreras, and J. Martínez-Comeche, “Ontologías: ontologías y recuperación de información,” 2008. Retrieved March 9, 2012, from http://www.sedic.es/gt_normalizacion_tutorial_ontologias.pdf

[21] B. Swartout, R. Patil, K. Knight, and T. Russ, “Toward distributed use of large-scale ontologies,” In Proceedings of the 10th Knowledge AcquisitionWorkshop (KAW’96) Canada, 1996.

[22] Y. Sure, M. Erdmann, J. Angele, S. Staab, R. Studer, and D. Wenke, “OntoEdit: Collaborative Ontology Development for the SemanticWeb,” Horrocks and J. Hendler (Eds.) In Fisrt International Semantic Web Conference (ISWC 2002) (Vol. 2342 of LNCS, pp. 221–235). Springer-Verlag Berlin, 2002.

[23] T. Tudorache, N.F. Noy, and M.A. Musen, “Supporting collaborative ontology development in Protege,” In 7th Intl. Semantic Web Conference, ISWC 2008, Germany, 2008.

[24] V. Presutti, A. Gangemi, S. David, G. Aguado de Cea, M.C. Suarez Figueroa, E. Montiel-Ponsoda, and M. Poveda, “Library of design patterns for collaborative development of networked ontologies,” Deliverable D1.1.3, NeOn project, 2008.

[25] A. Gangemi, and V. Presutti, Handbook on Ontologies, chapter Ontology Design Patterns,2009, pp. 221. Springer.

[26] B. Cuenca Grau, I. Horrocks, Y. Kazakov, and U. Sattler, “Modular reuse of ontologies: Theory and practice,” Journal of Articial Intelligence Research, 31, 2008, pp. 273-318.

[27] K. Lomas, T. Oreszcyn, D. Shipworth, A. Wright, and A. Summerfield, “Carbon Reduction in Buildings (CaRB) - Understanding the Social and Technical Factors that Influence Energy Use in UK Homes,” In Proceedings of the RICS Annual Conference Cobra, London, UK, 2006.

[28] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, and R. Rosati, “Tractable reasoning and efficient query answering in description logics: The DL-Lite family,” Journal of Automated Reasoning, 39(3), 2007, pp. 385–429.

[29] A. Poggi, D. Lembo, D. Calvanese, G. De Giacomo, M. Lenzerini, and R. Rosati, “Linking data to ontologies,” Jounal on Data Semantics, 2008, pp.133–173.

[30] A. Gómez-Pérez, Ontology evaluation, In S. Staab, and R. Studer, (Eds.), Handbook on Ontologies (1st ed.), 2004, (Chapter 13, pp 251-274). Springer.

[31] L. Obrst, W. Ceusters, I. Mani, S. Ray, and B. Smith, The evaluationof ontologies, In C.J.O. Baker, and K.-H. Cheung, (Eds.), Revolutionizing Knowledge Discovery in the Life Sciences, 2007, (Chapter 7, pp. 139- 158). Springer.

[32] A. Gangemi, C. Catenacci, M. Ciaramita, and J. Lehmann, “Ontology evaluation and validation: an integrated formal model for the quality diagnostic task,” Technical report, Laboratory of Applied Ontologies-CNR, Italy, 2005.

[33] J. Brank, M. Grobelnik, and D. Mladenic, “A survey of ontology evaluation techniques,” 8th International Multi-Conference of the Information Society, 2005, (pp. 166-169).

[34] V. Corrado, S.P. Corgnati, M. Garbino, “Energy Consumption Data Collection with DATAMINE,” Energy, Climate and Indoor Comfort in Mediterranean Countries, AICARR (ITA), Climamed 2007, 5-7 Settembre 2007, 2007, (pp. 803-816).