READY4SmartCities - ICT Roadmap and Data Interoperability for Energy Systems in Smart Cities

Deliverable D2.2: Ontologies and datasets for Energy Management System interoperability v1

Document Details

Delivery date:	*M12*
Lead Beneficiary:	*AEC3 Ltd.*
Dissemination Level (*):	*PU*
Version:	*1.0*
Preparation Date:	*06/10/2014*
Prepared by:	*Matthias Weise (AEC3), María Poveda-Villalón (UPM), Mari Carmen Suárez-Figueroa (UPM), Raúl García-Castro (UPM), Jérôme Euzenat (INRIA), Luz-Maria Priego-Roche (INRIA), Bruno Fies (CSTB), Andrea Cavallaro (DAPP), Jan Peters-Anders (AIT), Kleopatra Zoi (CERT/IT)*
Reviewed by:	*Anna Osello (POLITIO), Bruno Fies (CSTB)*
Approved by:	*Coordinator (DAPP), Technical Coordinator (UPM)*

(*) Only one choice between:

- PU = Public
- PP = Restricted to other programme participants (including the Commission Services)
- RE = Restricted to a group specified by the consortium (including the Commission Services)
- CO = Confidential, only for members of the consortium (including the Commission Services)

Project Contractual Details

Project Title:	ICT Roadmap and Data Interoperability for Energy Systems in Smart Cities
Project Acronym:	READY4SmartCities
Grant Agreement No.:	608711
Project Start Date:	2013-10-01
Project End Date:	2015-09-30
Duration:	24 months
Project Officer:	Svetoslav Mihaylov

SEVENTH FRAMEWORK PROGRAMME

Revision History

Date	Author	Partner	Content	Ver.
April 2014	Raúl García-Castro Matthias Weise	UPM AEC3	Deliverable structure	0.1
30/05/2013	Raúl García-Castro	UPM	Structure draft and first contributions to sections relevant to UPM	0.2
03/06/2014	Jerome Euzenat	INRIA	Structure draft and first contributions to sections relevant to INRIA	0.3
16/06/2014	Strahil Birov	EMP	Draft of chapters Aim and Collection methods	0.4
09/07/2014	María Poveda-Villalón, Mari Carmen Suárez-Figueroa, Raúl García Castro	UPM	Contributions to sections 1, 2, 4	0.5
03/09/2014	Strahil Birov	EMP	Division of content into D2.2 and D3.2	0.6
19/09/2014	Matthias Weise	AEC3, UPM, INRIA	Finalisation of D2.2 for internal review	0.7
22/09/2014 30/09/2014	Anna Osello Bruno Fies	Polito CSTB	Internal review	0.8
02/10/2014	Asunción Gómez Pérez	UPM	Final internal review	0.9
06/10/2014	Mari Figueroa, Matthias Weise, Andrea Cavallaro	UPM, AEC3, DAPP	Submission	1.0

Statement of originality:
This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Statement of financial support:
The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. FP7-SMARTCITIES 2013-608711

Executive Summary

This document presents the status of work with regards to work package 2 of the READY4SmartCities project, whose goal it is to identify the knowledge and data resources that support interoperability for Energy Management Systems (EMS). The work has been divided into two parts: (I) work that is intedepend from the EMS domain and thus was carried out together with work package 3, and (II) identification of ontologies, datasets and alignments specific for EMS interoperability. This decision was taken to avoid overlap between both work packages and to harmonize used data collection methods and publication of results.

For the **collection of ontologies and datasets**, a special online catalogue has been developed to ensure that resources are collected and recorded in a standardised way. The catalogue also allows for ease of understanding and use in terms of submission of new content, visualisation of existing resources and handling of recorded items. For the **collection of alignments**, an alignment server has been set up in order to identify and document links and alignments among the identified ontologies and datasets. The server will be made available as a web service so that the ontology and dataset catalogue can refer to it.

Various **collection methods** are continuously being used in order to identify and collect relevant ontologies, datasets and explore possible alignments. The methods include the set-up and administration of an online survey addressed to relevant experts, stakeholders in the domains identified in the previous deliverable, literature review by the study team, analysis of standardisation and institutional bodies, and screening of resource catalogues. Stakeholder engagement is crucial for the collection process, not only to be informed about yet available resources but also to raise awareness to publish data based on well-defined ontologies. WP2 and WP3 both support creation of communities to implement the vision of the Ready4SmartCities project.

II: Collected resources of EMS interoperability

Within the first project phase WP2 collected 25 ontologies. As there is no clear borderline between WP2 and WP3 there are even more ontology resources in the R4SC ontology catalogue that are of interest for EMS interoperability. From a technical point of view the majority of ontologies is based on OWL, RDF/XML and is written in English. While there is a common basis for ontology development (and more important ontology alignment), actual use of those resources is more difficult due to the lack of license information, missing online-availability or improper content negotiation mechanisms. Further assessment of those ontologies is more difficult. Some of them seem to be quite mature and widely accepted, while others have been developed for a very specific purpose and may disappear due to lack of further support. While our catalogue is developed to be a neutral source of information, possible issues are shown through colour codes in order to be aware of potential difficulties.

The availability of open linked data(sets) related to energy in general is scarce. In sum, only nine datasets could be collected, from which only three have been assigned to Energy Management System interoperability. Interesting datasets were either not available in an open data format or people hesitate to publish their data because of privacy, security, technical or other issues. Thus, beside extending the domains for further search and revising our methodology we plan to put more efforts in investigating reasons for this. The study team is committed to actively engaging the relevant stakeholders and pursuing the aspect of datasets in the domains of energy management systems and energy measurement and validation.

While intermediate results about ontologies are acceptable the results about datasets are far behind our expectations. Based on these results we are sceptical about availability of a substantial number of datasets for EMS interoperability, even if it is planned to extend the scope of domains. Therefore, for the second project phase we suggest to slightly change the strategy by putting more efforts in (a) working on examples for transferring available energy data to open linked data and (b) providing answers to most urgent question from potential data owners.

Glossary

Alignment	The result of analyzing multiple vocabularies to determine terms that are common across them.
Dataset	A collection of RDF data, comprising one or more RDF graphs that is published, maintained, or aggregated by a single provider. In SPARQL, an RDF Dataset represents a collection of RDF graphs over which a query may be performed.
Linked Data	A pattern for hyperlinking machine-readable data sets to each other using Semantic Web techniques, especially via the use of RDF and URIs. Enables distributed SPARQL queries of the data sets and a browsing or discovery approach to finding information (as compared to a search strategy). Linked Data is intended for access by both humans and machines. Linked Data uses the RDF family of standards for data interchange (e.g., RDF/XML, RDFa, Turtle) and query (SPARQL).
Ontology	A formal model that allows knowledge to be represented for a specific domain. An ontology describes the types of things that exist (classes), the relationships between them (properties) and the logical ways those classes and properties can be used together (axioms).
Open Data	Refers to content that is published on the public Web in a variety of non-proprietary formats.
OWL	Web Ontology Language (OWL) is a family of knowledge representation and vocabulary description languages for authoring ontologies, based on RDF and standardized by the W3C.
RDF	Resource Description Framework (RDF) is a family of international standards for data interchange on the Web produced by W3C. RDF is based on the idea of identifying things using Web identifiers or HTTP URIs, and describing resources in terms of simple properties and property values.
SKOS	Simple Knowledge Organisation System (SKOS) is a vocabulary description language for RDF designed for representing traditional knowledge organization systems such as enterprise taxonomies in RDF.
SPARQL	SPARQL Protocol and RDF Query Language (SPARQL) defines a query language for RDF data, analogous to the Structured Query Language (SQL) for relational databases. It is a family of standards of the World Wide Web Consortium.
URI	A global identifier standardized by joint action of the World Wide Web Consortium and Internet Engineering Task Force. A Uniform Resource Identifier (URI) may or may not be resolvable on the Web. URIs can be used to uniquely identify virtually anything including a physical building or more abstract concepts such as colors.
VoCamp	A VoCamp is an informal event where people can spend some dedicated time creating lightweight vocabularies/ontologies for the Semantic Web/Web of Data. The emphasis of the events is not on creating the perfect ontology in a particular domain, but on creating vocabularies that are good enough for people to start using for publishing data on the Web.
BIM	A new method to support collaborative work in the AEC and FM industry. BIM is an abbreviation for Building Information Model(ling) and defined to be *“an improved planning, design, construction, operation, and maintenance process using a standardized machine-readable information model for each facility, new or old, which contains all appropriate information created or gathered about that facility in a format useable by all throughout its lifecycle.”* (definition from the NBIMS committee).

Table of Contents

Document Details ... 1
Project Contractual Details ... 1
Revision History ... 2
Executive Summary ... 3
Glossary ... 4
Table of Contents ... 5
How to read this document ... 7
Part I: Approach and methodology ... 8
1 Collection of ontologies and datasets ... 9
1.1 Project Partner involvement ... 9
1.2 Stakeholder involvement ... 9
1.3 Review literature for ontology seeking ... 11
1.4 Review literature for datasets ... 12
1.5 Analysis of Standardization and Institutional Bodies ... 12
1.6 Lookup Resource Catalogues ... 14
2 Recording of ontologies and datasets ... 15
2.1 Ontology catalogue ... 15
2.1.1 Overview of the ontology catalogue ... 15
2.1.2 Catalogue generation ... 16
2.1.3 Web application ... 19
2.2 Dataset catalogue ... 20
2.2.1 Overview of the dataset catalogue ... 20
2.2.2 Catalogue generation ... 20
2.2.3 Web application ... 21
2.3 Alignments catalogue ... 22
2.3.1 Overview of the Alignment server ... 22
2.3.2 Example methodology of alignment generation ... 24
2.3.3 Description of the Alignment server as web service and link to the ontology catalogue ... 24
2.4 Overview of ontologies and datasets gathered during the first project year ... 26
2.4.1 Ontologies, vocabularies and standards ... 26
2.4.2 Datasets ... 36

2.4.3 Ontology alignments and data links ... 36
2.4.3.1 Content analysis ... 37
2.4.3.2 Reference analysis ... 37
2.4.3.3 Distance analysis ... 39
2.4.3.4 Correspondence analysis ... 42
2.4.3.5 Alignment server ... 47
Part II: Ontologies and Datasets for Energy Management Systems ... 48
3 Differentiation between Energy Management System and Energy Measurement and Validation ... 48
4 Collected ontologies relating to Energy Management Systems ... 52
4.1 Gap analysis ... 52
4.2 List of ontologies ... 53
5 Collected datasets relating to Energy Management Systems ... 65
5.1 Gap analysis ... 65
5.2 List of datasets ... 65
Part III: Conclusions and next steps ... 68
References ... 70
Appendix: prefix list ... 71

How to read this document

Both work package 2 and 3 of the READY4SmartCities project set out to identify and collect ontologies and datasets that support interoperability for energy management systems and energy measurement and validation respectively. While they address different domains, the processes for identification and collection as well as analysis and presentation are identical. The partners involved in the project decided to use a uniform approach in order to share the initial effort of developing the necessary infrastructure (online catalogue, alignments server, online survey) and focus on the process of identification and collection.

The deliverable is therefore divided into three parts that document the activities of the consortium with regards to ontology and dataset collection.

Part I describes in detail the collection process and methods used, the catalogue developed to support resource collection and storage, and the alignments server used to identify links and connections between the collected resources. An overview of the collected ontologies (more statistical) and datasets (more descriptive due to the low number of datasets) is also present in this part. Part I is identical in D2.2 and D3.2, while the other parts of these deliverables differ in content. Therefore the reader should refer to part I only once.

Part II presents how the two domains differ from one another and provides an overview of the collected resources that support interoperability for energy management systems. A gap analysis is also an element of this section, used as an indicator as to what resources need to be collected and thus gives insights into the work to be done during the second project year. The resources are documented using standardised tables that cover general, information about the ontology/dataset, its licence and format, scope as well as possible use cases and links to other resources. This information partially overlaps with information that is provided in the online catalogue. While the catalogue is limited to objective data about identified ontologies and datasets this part provides a first rough assessment.

Part III contains conclusions and draws next steps to be taken in order to increase the effectiveness of the process and collect more ontologies and datasets during the second project year.

Part I: Approach and methodology

A new energy data ecosystem that will enable interoperability and exploitation of data in the context of Smart Cities by adopting Semantic Web standards and technologies is one of the main tasks in the READY4SmartCities project. Such data ecosystem will accommodate cross-domain data and will allow the exploitation of such data by identifying the set of relevant ontologies and datasets that are collected in a catalogue that can be used by interested parties as a starting base for exploitation of the resources.

Work packages 2 and 3 of the READY4SmartCities project promise to identify the knowledge and data resources that support interoperability for energy management systems as well as energy measurement and validation. These resources include relevant ontologies, vocabularies and standards, datasets, and ontology alignments. Energy data along different topics is covered, e.g. energy use and production, behaviour of construction elements, equipment, and users (WP2) as well as measurement methods, baselines and measured data, key performance indicators, etc. (WP3).

The purpose of this document is to present the approach chosen to identify and collect the resources as well as the results achieved so far. This includes elaborating the developed online catalogue for recording ontologies and datasets and the collection methods used to engage relevant stakeholders as source of information about these resources. In addition, the ontologies and datasets are introduced, including a domain coverage analysis, a gap analysis of missing resources, and statistics on the collected metadata and the quality indicators.

These results will be updated in a second version of this document due at the end of the READY4SmartCities project in 2015.

Contribution of Partners.

The following list states which partners have contributed to the different sections of this part of the deliverable.

- Introduction and conclusions. EMP, AEC3, UPM, INRIA
- Catalogue development: UPM
- Alignments catalogue: INRIA
- Collection of ontologies and datasets: AEC3, EMP, UPM, INRIA, CSTB, DAPP, AIT, CERT/IT, POLITO
- Relevant ontologies and datasets: UPM, EMP, AEC3
- Ontology alignments and data links: INRIA

1 Collection of ontologies and datasets

1.1 Project Partner involvement

The involvement of project partners started early on in the project by discussing and planning tasks and objectives in work packages 2 and 3 during the monthly consortium telcos. It was then decided to create a more focused group comprising partners involved in WP2 and 3, which would discuss the progress of both work packages on a weekly basis (starting from 12.3.2014). The weekly telcos broadly cover the following topics:

- **Organisation** of work and distribution of tasks: continuous status reporting and assignment of new tasks in accordance with the plan chosen and the available documents, i.e. the DoW, as well as discussion of open issues.
- **Communication** of early results and invitations of stakeholders to partake in related activities and events. This includes setting up online surveys, validation of identified ontologies and datasets during VoCamps, as well as other events. Online media (twitter, linked-in) is also part of the strategy to reach a wider range of stakeholders. Invitations and news on relevant websites such as ValMet, eeSemantics, and the project website have also been continuously announced.
- **Research**: one of the main sources of finding ontologies and datasets in the relevant domains comes from the expertise of the involved partners in R4SC. In addition, continuous research by the focus group is performed using the sources described in D2.1/D3.1.

1.2 Stakeholder involvement

An **online survey** was set up and launched in March 2014 to enable capturing contributions by the stakeholder community. The idea of the survey is to provide an easy way for stakeholders to take part in the project activities, while also offering the possibility for more experienced stakeholders to provide detailed information. This has been realised by creating two versions of the survey. The first asks stakeholders to only provide the location (URL) of the resource they are aware of, and the follow up research of the resource is done by the project partners. A second survey provides an interface with all information necessary to record an ontology or dataset. If filled by a stakeholder, this information is saved in the database and only needs to be checked by the curator of this database (for the ontology catalogue, this is UPM, empirica is the curator for the gathered datasets). The survey links will remain active throughout the project lifetime in order to provide a way for new ontologies and datasets to be included. The following links are used for this purpose:

- http://survey.ready4smartcities.eu/index.php/638667/ - short ontology survey
- https://docs.google.com/forms/d/1kTrNUKRnAlN5bBnOwTzQjWwQLinKFQcW4EqXDOYbFsQ/viewform - long ontology survey
- http://survey.ready4smartcities.eu/index.php/162877/ - short dataset survey
- https://docs.google.com/forms/d/1EUISLPLpVHmBaUy2gI76LjE_UPkgPaSW9J1nDruKS0U/viewform - long dataset survey

The target audience for the online survey consisted primarily of stakeholders having access or connected somehow to energy-related data. Such stakeholders were reached through various channels as listed below:

- Mailing list of relevant partners/projects – each partner from the READY4SmartCities consortium shared a number of their partners from other projects based on their background and their relevance to the survey. The mailing list created counted more than 1000 people and was used to introduce the R4SC project and to invite interested people to fill in the survey.
- eeSemantics wiki – CERTH partner is responsible for the maintenance of the eeSemantics wiki, forum and document library on Semantic Interoperability of Energy Efficiency ICT Tools for eeBuildings and beyond and therefore has access to the whole member list of relevant stakeholders (counting more than 500 members).

An introduction to the R4SC project and concept was sent, followed by an invitation to participate in the survey, by both a post in the Forum and an email sent to the mailing list.

- READY4SmartCities Portal – the survey was made available and promoted on the R4SC website http://www.ready4smartcities.eu/ and was posted on the website's newsletter.
- Social Networks – the questionnaire invitation was published through the R4SC project's social networks, namely LinkedIn and Twitter, early established in the project.
- VoCamp Participants – during the VoCamps in Germany and Finland, participants with high relevance to energy-related data were approached and were requested to dedicate some time to answer the survey.

Up to July 2014 (five months running time) there have been:

- 5 legitimate entries for ontologies, all of them have been covered by the catalogue
- Just 1 entry for datasets

resulting directly from the survey, a rather disappointing number considering that the survey page has been visited in the same time period by more than 20 times more users than the submitted entries, which shows that either the users were not aware of any ontologies/datasets from the relevant domains and therefore could not contribute (the more probable explanation), or that they did not want to share results.

The ontology catalogue has been presented during the following events:

- 4th VoCamp on "Integrating multiple domains and scales" (Barcelona, Spain, 13-14 of February 2014): During this event a preliminary version of the catalogue was presented. Main feedback was about providing the catalogue metadata in RDF (which is currently implemented in the catalogue).
- 5th VoCamp on "Device & Sensor Ontologies" (Bonn, Germany, 20-21 May 2014): During this VoCamp the ontology catalogue was presented obtaining the following comments:
  - When clicking in a domain, it would be nice to see all the ontologies about that domain: this feature is planned to be implemented in future versions.
  - When clicking in a concrete syntax, the application should return the appropriate format: this point involves some technical restrictions so far, therefore it is not planned to be implemented in immediate versions but consider as future work.
  - To split somehow what labels give only information and which ones retrieve information: it is considered to be implemented in the catalogue.
  - To include the ontologies reused by BETaas ontology and HYDRA ontology, EBBITS ontology, SCO: Smart Campus Ontology and DER modeling.
    - At the moment of writing this deliverable, SCO and DER was not available and UPM is still waiting response about HYDRA and EBBITS.
    - Regarding the ontologies reused by BETaas ontology, CF[1] and Phenonet[2] will be included in the next version of the catalogue. The rest of reused ontologies currently appear in the catalogue.
- Joint workshop on Linked Data in Architecture and Construction (2nd LDAC Workshop & 6th eeSemantics VoCamp) (Espoo, Finland, 26-27 May 2014): During this VoCamp the ontology catalogue was presented in a brief slot instead of a full presentation; hence, there was no time for giving details. However, there was interest in reusing the RDF serialization of the metadata gathered in the catalogue. Other ontologies to be included in the catalogue were also proposed, for example, the "CB-NL: a common ontology" and the SEMANCO ontology, which was already consider by other approaches.

[1] http://www.w3.org/2005/Incubator/ssn/ssnx/cf/cf-feature

[2] http://www.w3.org/2005/Incubator/ssn/ssnx/meteo/phenonet

1.3 Review literature for ontology seeking

Some of the ontologies included in the READY4SmartCities catalogue[3] have been gathered through the revision of related literature. It is important to mention that the search has been focused on ontologies or vocabularies already implemented in an ontology language, such as RDF and OWL. Thus, when the ontology was only a non-implemented model, such ontology was not taken into account.
The general ontology collection process was:

- UPM read each corresponding document and search for references to ontologies
- When a reference to a relevant ontology is found in the text, two different situations can occur:
  - Such a reference directly leads to a link in which the ontology (implemented in an ontology language) is available. In this case, UPM downloaded the ontology and reviewed the ontology code. After that, UPM acted as catalogue populator by means of providing ontology metadata through the online form already mentioned in Section 1.2.
  - Such a reference is just a textual reference (normally the ontology name). In this case, UPM performed a broad search in the Internet looking for documents about such ontology. When documents were found, UPM started again the general process. On the contrary, UPM had to contact people involved in the ontology development and/or related with such an ontology. UPM directly contacted paper authors, deliverable contributors and/or project coordinators in order to ask for (a) other relevant papers and/or documents in which the ontology is described, (b) information about the ontology files (e.g., if exists, the site in which the ontology is available for downloading), and (c) any other relevant data. However, UPM discovered cases in which it were not possible to contact people (document authors, project coordinators, etc.) involved in the ontology development or related to the ontology building.

    As a result of the contacts conducted, the possible responses obtained were:

    - Confirmation that the ontology is not available on-line, but the ontology file was sent via email
    - Confirmation that there is no ontology implemented
    - Confirmation that the ontology is not public
    - Information about the current status of the ontology development (e.g., the ontology implementation is in progress, our plans includes the development of an ontology).
    - No reply was obtained at the moment of writing this document

The revision of related literature included the following sources:

- *eeSemantics wiki*[4]. UPM has reviewed pages in the wiki looking for ontologies related to the energy efficiency domain. In particular, pages on the 'Examples and Implementations' and 'eeBuilding Data Models' sections were inspected. In some cases, it was also needed to search for related papers and/or documents. As a result of reviewing this source, five ontologies were included in the catalogue.
- *eeBuilding Data Models workshop proceedings*. Proceedings of 2012 and 2013 editions of this series of workshops were reviewed in order to find related ontologies. The ontologies found in such proceedings were already included in the catalogue while checking other sources.
- *ETSI Smart Appliances workshop report*. The document, D-S1 Interim Study Report, presents a list of existing semantic assets and use case assets, describes their semantic coverage, and proposes an initial semantic mapping. In some cases, it was also needed to search for related papers and/or documents. As a result of the revision of this report one ontology has been included in the catalogue.

---

[3] http://smartcity.linkeddata.es/

[4] https://webgate.ec.europa.eu/fpfis/wikis/display/eeSemantics/Home

- *European project production*. Documents produced within 70 energy-related projects (such as STREAMER, SESAME-S, S4EEB, HYDRA, and SEEMPUBS) have been reviewed. As an outcome of this literature checking, five ontologies were included in the catalogue by UPM acting as a catalogue populator. It is worth mentioning that nine projects are currently developing ontologies (such as ee-DIM ontology) and/or have in their plans the ontology building. In addition, 18 out of the 70 contacted projects do not develop ontologies and UPM is still waiting response for 38 projects.
- *Other related research literature*. Papers in the area of energy efficiency have been reviewed. UPM included in the catalogue eight ontologies (e.g., DogOnt, ontologies developed in the context of ThinkHome project) found during the inspection of this source.

Finally, it is also important to mention that UPM has checked READY4SmartCitites Deliverable D4.1 in order to include in the catalogue those ontologies mentioned in the described guidelines. In addition, UPM considered useful to have ontologies in the geographical domain, thus literature in such an area was reviewed. The effect of this revision was the inclusion of two ontologies (OGC GeoSPARQL and WGS84 Geo Positioning).

1.4 Review literature for datasets

The datasets included in the READY4SmartCities catalogue have been gathered mainly through desk research, which, however, relates also to surveying related literature sources. It is important to mention that the search has been focused on datasets that are linked and open, i.e. the data should be in RDF. This meant that other datasets which weren't linked or open were not added to the catalogue, they were, however, taken into account specifically for the gap analysis (see chapter 8).

Relevant sources for the datasets came from the expertise of the involved project partners, the survey entries, and suggestions from experts and stakeholders contacted by the consortium as part of WP1 activities. Some of the portals that were pointed as possible sources of information include:

- **Reegle**[5]: the gateway has already established itself as a popular information portal in the fields of renewable energy and energy efficiency. It offers all of its data under W3C standards, i.e. it is open and Linked Data in a non-proprietary format (RDF).
- **OpenEI**: a collaborative knowledge-sharing platform with free and open access to energy- related data, models, tools, and information. OpenEI features over 55,000 content pages, more than 600 downloadable data sets, regional gateways on a variety of energy-related topics, and numerous online tools.
- **Datahub**: this powerful data management platform covering a wide range of topics. It offers data collections, some of which are linked and open.

The dataset collection process is similar to the one used to collect ontologies. An identified dataset that meets the requirements of Linked Open Data is added to the catalogue by the dataset curator EMP (only metadata) through the corresponding online form.

1.5 Analysis of Standardization and Institutional Bodies

In general, standardization and institutional bodies are a valuable source of information when it comes to identify agreements for information exchange and reuse of data. Seamless exchange of digital data has been an issue from the very beginning of computer based work and a lot of efforts have already been made to reach consensus between different parties about how to organize and structure shared data. The Open Linked Data Approach based on general webstandards like URI, XML, RDF, OWL and SPARQL is a relatively new approach compared to other technologies like SQL, IDEF or STEP-EXPRESS. The main use case of (Open) Linked Data is to publish and interlink pieces of information and thus differs from current exchange and integration approaches.

[5] http://www.reegle.info/

Meanwhile, after several years of research, standardization bodies took notice of this new technology and its potential benefits. While there are still ongoing discussions about use cases and how to position OLD to existing developments, it became clear that both approaches can benefit from each other. On one side there are rich vocabularies, model schemata and business logic developed in many years of standardization work and on the other side there is a new technology to support the web of data with all promised advantages. While our search for ontologies and open datasets published by standardization bodies was not really successful we realized that there are ongoing discussions and preparation work for further standardisation. A short summary of the current situation as well as activities of R4SC towards support actions is given below.

W3C

W3C is seen as the most relevant standardization body for OWL-based ontologies. The partner UPM is active in working groups related to the standardization of different technologies in the W3C. Different ontologies and vocabularies developed in the W3C and widely used were included in the catalogue for representing generic concepts (e.g., time, organizations) and some specific ones (e.g., sensor networks, statistical data). More domain specific W3C standards are currently developed or discussed for instance with support from OGC (Spatial Data on the Web Working Group)[6] or AEC researches (Linked Building Data Community Group)[7].

ETSI

From summer 2013, the European Commission has the intention to launch a standardization exercise at ETSI to propose a high-level model (an ontology) for smart appliances, as an ETSI standard. The first step consists in a pre-normative study that will be done by the Dutch TNO. This project is called "Study on Semantic Assets for Smart Appliances Interoperability" and consists in defining/ identifying a common vocabulary for appliances product information, commands, signals and in a second step agrees on an abstract architecture compatible with the current machine-to-machine (M2M) standards. The outcomes of this study is highly relevant for our project and already ontologies coming from 17 relevant initaitives or project have been translated into Turtle language and are available for download (https://sites.google.com/site/smartappliancesproject/ontologies).

UPM and other project partners participated in the DG CONNECT & ETSI Workshop on Smart M2M Appliances, held in Brussels on 27-28 May 2014. In that workshop, a study on available semantics assets for the interoperability of smart appliances was presented. The document, D-S1 Interim Study Report, presents a list of existing semantic assets and use case assets, describes their semantic coverage, and proposes an initial semantic mapping. We took into account the ontologies described in that document and, in some cases, we also needed to search for related papers and/or documents

AENOR

UPM is member of the AENOR (the Spanish standardization body) Technical Committee for Smart Cities (CTN 178). For this version of the catalogue a current working draft of a standard on open data for smart cities was analysed in order to search for relevant ontologies.

buildingSMART

buildingSMART is an international non-profit organization that develops open standards for the AEC and FM industry. Since nearly 20 years buildingSMART is pushing the BIM technology. Meanwhile its open IFC standard is supported by all major CAD software tools. AEC3 is very active in this organization and started to facilitate discussions about an ifcOWL standard as a baseline for further developments. The Joint workshop on Linked Data in Architecture and Construction (2nd LDAC Workshop & 6th eeSemantics VoCamp, Espoo/Finland, 26-27

[6] http://www.w3.org/2014/05/geo-charter

[7] http://www.w3.org/community/lbd/

May 2014), co-organised and supported by the Ready4SmartCities project, brought together ontology and AEC experts and was used to discuss two main topics: (1) use case scenarios for linked building data and (2) requirements for a unified ifcOWL[8] representation. Also, it was decided to give feedback to the buildingSMART organization and to facilitate a buildingSMART working group that puts this topic on its agenda.

ISO

ISO is a well known international standardization body for a broad spectrum of engineering applications. The partner AEC3 is involved in standardization work in the building and construction sector, in particular in publishing the IFC model as an ISO standard (ISO 16739). OWL ontologies are not yet a topic, but there are similarities to XML schema-based definitions. Within the STEP familiy of standards (ISO 10303) the EXPRESS language as used for the IFC specification is defined. For support of XML schema a mapping approach is used that includes a standard mapping configuration that can also be adapted to specific purposes. This approach fits to proposals that have been made by several researchers to transfer the EXPRESS-based IFC model to an OWL representation. These proposals could be a baseline for a general mapping approach that then would allow to map other EXPRESS-based standards to a W3C conform representation.

Other Standardisation and Institutional Bodies

There are a couple of efforts towards the aim of Ready 4 Smart Cities, e.g. the Energy Performance Buildings Directive from CEN or the draft about a Facility Smart Grid Information Model from ASHRAE. Also, there are a couple of data exchange standards that are relevant in context of smart cities use cases. However, they typically do not make use of the Open Linked Data approach or underlying technologies so that we decided to ignore such efforts for our catalogue or further discussions.

1.6 Lookup Resource Catalogues

There are several ontology search engines that UPM has analysed for identifying ontologies that are relevant to READY4SmartCities: Watson[9], Swoogle[10], and Linked Open Vocabularies (LOV)[11].

The main resource used during the ontology catalogue has been LOV as it includes information about creators, maintainers and publishers that are not always included in the ontology encoding nor the documentation associated, if any. As LOV does not cover all the ontologies gathered during this collection process this approach does not ensure to find such metadata for all possible cases.

Another catalogue that UPM analysed was the Collaborative platform Joinup[12]. This platform offers several services that aim to help e-Government professionals share their experience regarding interoperability solutions with each other. Although the vocabularies are not directly related to the energy efficiency or the smart cities domain, UPM considered useful to review ontologies and vocabularies recommended in such a platform. The effect of this inspection was the inclusion of the Registered Organization Vocabulary in the ontology catalogue.

---

[8] As buildingSMART already publishes a mature, object-oriented data model the strategy from researchers has been to work on mapping proposals from the EXPRESS language to a proper OWL representation of IFC. Depending on use case scenarios and used ontology toolsets there are different flavours for such mapping definitions. Thus, while all available ifcOWL representations are derived from the original IFC specification there is not yet a common agreement within this community which of those should be preferred or the "standard" representation.

[9] http://watson.kmi.open.ac.uk/

[10] http://swoogle.umbc.edu/

[11] http://lov.okfn.org/dataset/lov/

[12] https://joinup.ec.europa.eu/

2 Recording of ontologies and datasets

2.1 Ontology catalogue

2.1.1 Overview of the ontology catalogue

In order to collect ontologies we follow a semi-automatic process that involves different people with different roles: a) **contributors**, who suggest ontologies to be included in the catalogue or even provide their descriptions (i.e., metadata) through an on-line form; b) **populators**, who include new ontologies into the catalogue by describing them through the on-line form; and c) metadata **curators**, who review, improve, and complete the metadata of the ontologies inserted by contributors and populators.

These roles and their interaction with the ontology collection process are illustrated in Figure 1. As shown in such figure, the process consists of the following steps:

1. Contributors and populators provide ontology metadata through an on-line form[13]. There is also an option for contributors to provide minimal information for ontologies by means of filling in a short on-line form[14]; in this case, the metadata curators will be in charge of completing the ontology metadata.
2. The metadata is received by the curators, that is, the catalogue maintainers, who review, improve, and complete such data if needed. This step implies some manual evaluation of the collected metadata.
3. Once the metadata is curated, both an RDF [Brickley, 2004] and an HTML representation of the catalogue information are generated. During this process some evaluation tasks are carried out over the ontologies. It should be noted that since the process contains a manual component (i.e., metadata curation) the catalogue is not immediately updated when a new ontology is introduced through the on-line form.

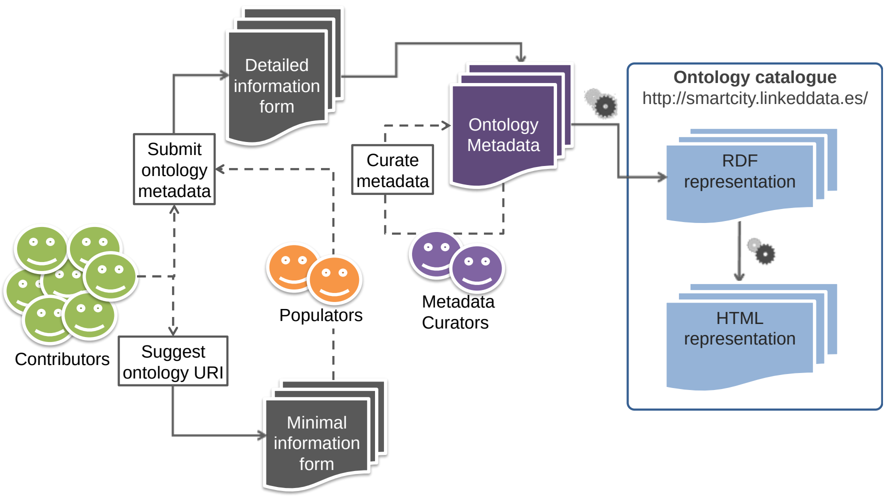

Figure 1. Proposed process to collect ontology metadata and generate the ontology catalogue

[13] http://goo.gl/SG0pMA

[14] http://survey.ready4smartcities.eu/index.php/638667/

2.1.2 Catalogue generation

As already mentioned, the main advantages of reusing existing ontologies for describing the data of the ontology catalogue are that the catalogue data will be more interoperable with existing data and that the time of developing the ontology for the catalogue decreases. For these reasons, a common set of metadata vocabularies has been reused to describe the ontologies that are included in the catalogue.

These metadata have been selected after analyzing two well-known ontologies that can be used to describe ontology metadata, namely, OMV (Ontology Metadata Vocabulary) [Hartmann et al. 2005] and VOAF (Vocabulary of a Friend[15]) as explained in [García-Castro et al, 2014].

One limitation of OMV is that it does not reuse terms already defined in other well-known ontologies. For this reason we follow the VOAF approach that consists on reusing terms already defined in other vocabularies and only add those that are strictly necessary. As a result, five vocabularies have been reused for describing the ontologies of the catalogue; their titles, prefixes and URIs are listed in Table 1.

Table 1. Vocabularies reused for describing the ontologies of the catalogue

Vocabulary	Prefix	URI
Creative Commons Rights Expression Language	cc	http://creativecommons.org/ns
Dublin Core Metadata Initiative Metadata Terms	dc	http://purl.org/dc/terms/
Vocabulary of a Friend	voaf	http://purl.org/vocommons/voaf#
Ontology Metadata Vocabulary	omv	http://omv.ontoware.org/2005/05/ontology#
VANN: A vocabulary for annotating vocabulary descriptions	vann	http://purl.org/vocab/vann/

Figure 2 shows the ontology used to describe the ontologies included in the catalogue. As represented in such figure, the central class of the model is *voaf:Vocabulary*, that is used to represent ontologies. This class contains some attributes (or datatype properties) to represent the ontology title (*dc:title*), its description in natural language (*dc:description*), its creation date (*dc:issued*), its last modification date (*dc:modified*), its prefix (*vann:preferredNamespacePrefix*), and its namespace (*vann:preferredNamespaceUri*).

[15] http://purl.org/vocommons/voa

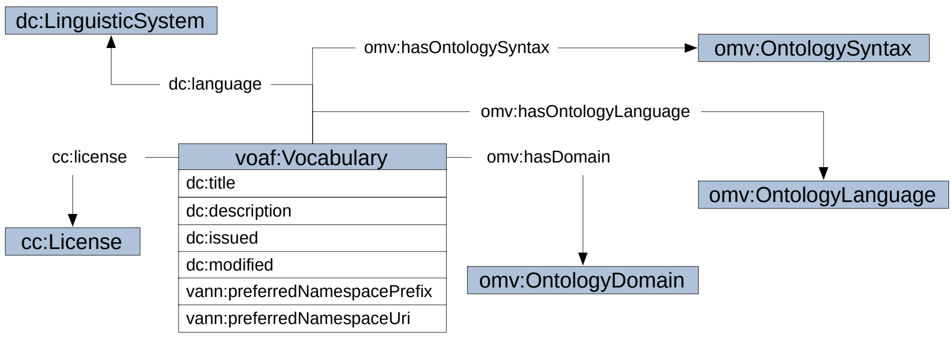

Figure 2. Ontology to represent ontology metadata

Individuals belonging to the class *voaf:Vocabulary* could be related to individuals belonging to other classes. This way, the ontology language in which an ontology is developed is stated through the relationship *omv:hasOntologyLanguage* between the classes voaf:Vocabulary and *omv:OntologyLanguage*; the syntax in which an ontology is available is represented by the relationship *omv:hasOntologySyntax* between the classes *voaf:Vocabulary* and *omv:OntologySyntax*; the domains covered by the ontology are indicated by means of the relationship *omv:hasDomain* between the classes *voaf:Vocabulary* and *omv:OntologyDomain*; the language in which the ontology is expressed is stated by the relationship *dc:language* between the classes *voaf:Vocabulary* and *dc:LinguisticSystem*; and the license of the ontology is indicated through the property *dc:license* between the classes *voaf:Vocabulary* and *cc:License*.

Once the model for describing ontology metadata is defined, the collected data from the on-line form is transformed into RDF according to such model. The data gathered from the form is stored in a Comma-Separated-Value (CSV) file where each row contains the data related to a given ontology. For each ontology, an individual of *voaf:Vocabulary* is created, its attributes are filled in with the values introduced by the contributors or curators and, finally, the individual is linked to other individuals that represent ontology syntaxes, ontology implementation languages, languages, licenses, and domains. An example of an ontology annotated following the ontology presented above is shown in Figure 3.

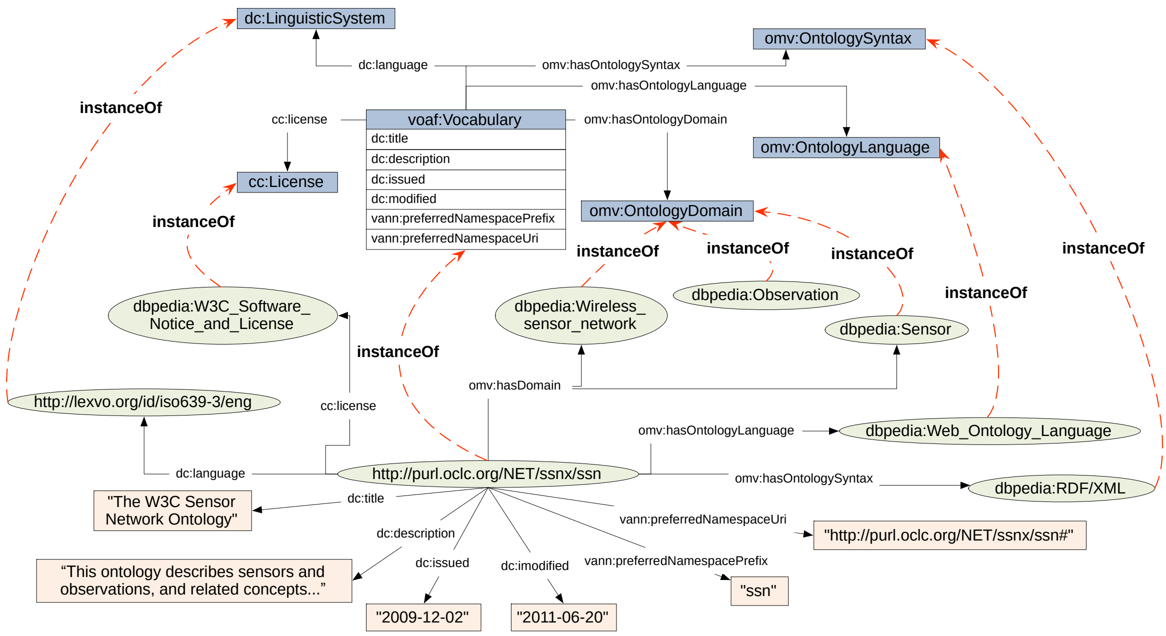

Figure 3. Example of ontology metadata representation in RDF

It is worth noting that the generated RDF data is linked to existing datasets, following widely-used recommendations for publishing Linked Data [Bizer, Heath, & Berners-Lee, 2009]. In this case we use the DBpedia[16] and Lexvo[17] datasets in order to link to general domain and to linguistic entities, respectively. DBpedia may be considered the nucleus for the Web of Data since it contains information about a number of domains (e.g., geographic information, people, companies, online communities, films, music, books, among others) describing around 4.0 million entities for the English version. Lexvo has been selected to represent languages as it brings information over 7,000 language identifiers and ensures that these identifiers are dereferenceable and highly interconnected as well as externally linked to a variety of resources on the Web.

During this process there can appear different scenarios for attaching property values to a given individual of the class *voaf:Vocabulary*. The easiest case is when filling in the values for attributes because the value gathered from the CSV file is used as a Literal and directly linked to the vocabulary through the corresponding datatype property (for example *dc:title* in Figure 2).

When the link between the vocabulary individual and the values to be attached is an object property (for example *omv:OntologyLanguage* in Figure 2), it means that the target value takes the form of another individual, instead of a Literal. For these cases there are two possible ways of linking a given individual to other individuals. For the object properties *omv:hasOntologyLanguage*, *omv:hasOntologySyntax*, and *cc:License* there are sets of individuals pre-defined in the model because the possible values are an enumerated set. It should be noted that the current set of individuals might not cover all the cases; for example, for licenses, when a new license is included into the system a new individual for representing such a license is created. For the case of the object property *dc:language*, the Lexvo dataset is used in order to represent individuals of the class *dc:LinguisticSystem*.

In order to represent individuals form the classes *omv:OntologyLanguage*, *omv:OntologySyntax* and *cc:License* we give priority to URIs defined in official namespaces, that is, in namespaces controlled by the organism that created or maintains such concept or term. In this regard, we use URIs defined in the cc namespace to identify cc licenses (e.g., http://creativecommons.org/licenses/by/3.0/). If there is no official URI defining a given individual, we link to the corresponding DBpedia entity; for example, for representing the OWL ontology language we use "http://dbpedia.org/resource/Web Ontology Language".

Finally, for representing the domains that a given ontology might cover there is no fixed set of possible individuals, that is, this field is a free text box in the on-line form where the contributor or curator could include any value or set of values. In order lo link the ontologies to the domains they are related to, we first try to find existing entities representing such domains. For doing so, ontology grounding techniques are used in order to determine links between the unrestricted terminology of users and resources of the Web of Data (particularly DBpedia), making easier the interoperability and later alignment among models [Lozano et al. 2012]. If no entity from DBpedia is found, a new individual is created in a namespace under our control, as recommended in Linked Data development guidelines [Bizer, Heath, & Berners-Lee, 2009]. Among the advantages of linking ontology domains to DBpedia entities it should be noted: (a) the connection of the dataset being built with the Web of Data through a well-connected dataset, DBpedia, and (b) the avoidance of duplicates due to different lexicalizations of the same concept.

As previously explained the collected data is generated both in machine-processable format (i.e., RDF data following the Linked Data principles) and in a human-readable documentation (i.e., an HTML website that is automatically generated from the RDF).

[16] http://dbpedia.org

[17] http://www.lexvo.org/)

2.1.3 Web application

The catalogue of ontologies about smart cities, energy and other related fields can be accessed through a web application available at http://smartcity.linkeddata.es/ontologies/.

As shown in Figure 4, the catalogue allows visualizing metadata about the listed ontologies. For each ontology, the metadata are shown in the columns: “Open License”, “Ontology Language”, “Syntax”, “Domain”, and “Natural Language”. The values shown in each cell of the table contain different information both represented by text and by color; for ontology metadata, colors have the following meaning: “plain information” for blue and “unknown” for grey. Furthermore, in addition to the color, each cell contains detailed information when available.

Apart from ontology metadata, the catalogue presents in the first three columns the quality indicators for the ontologies defined in [Garcia-Castro et al, 2014]: “Online Availability’, “Open License”, and “Ontology Language”. For the quality indicators, colors have the following meaning: “success” for green, “warning” for orange, “danger” for red, and “unknown” for grey.

The values of the “Open License” and “Ontology Language” indicators are taken from the ontology metadata and the evaluation results are stated using color. For example, in the column “Open License” we can see that the ontologies “Units of Measure (OM)” and “The W3C Organization Ontology” are both published under an open license as the color of the cell is green, while detailed information about the licenses is also provided. More precisely, these licenses are “CC-BY 3.0” (Creative Commons Attribution 3.0 Unported) and “W3C” (W3C Software Notice and License) respectively, as shown in Figure 4.

The “Online Availability” indicator represents whether the ontology is available in the Web in RDF and in HTML format. The evaluation of this indicator is performed on execution time when the catalogue is generated, that is, it is updated every time the catalogue is rebuilt.

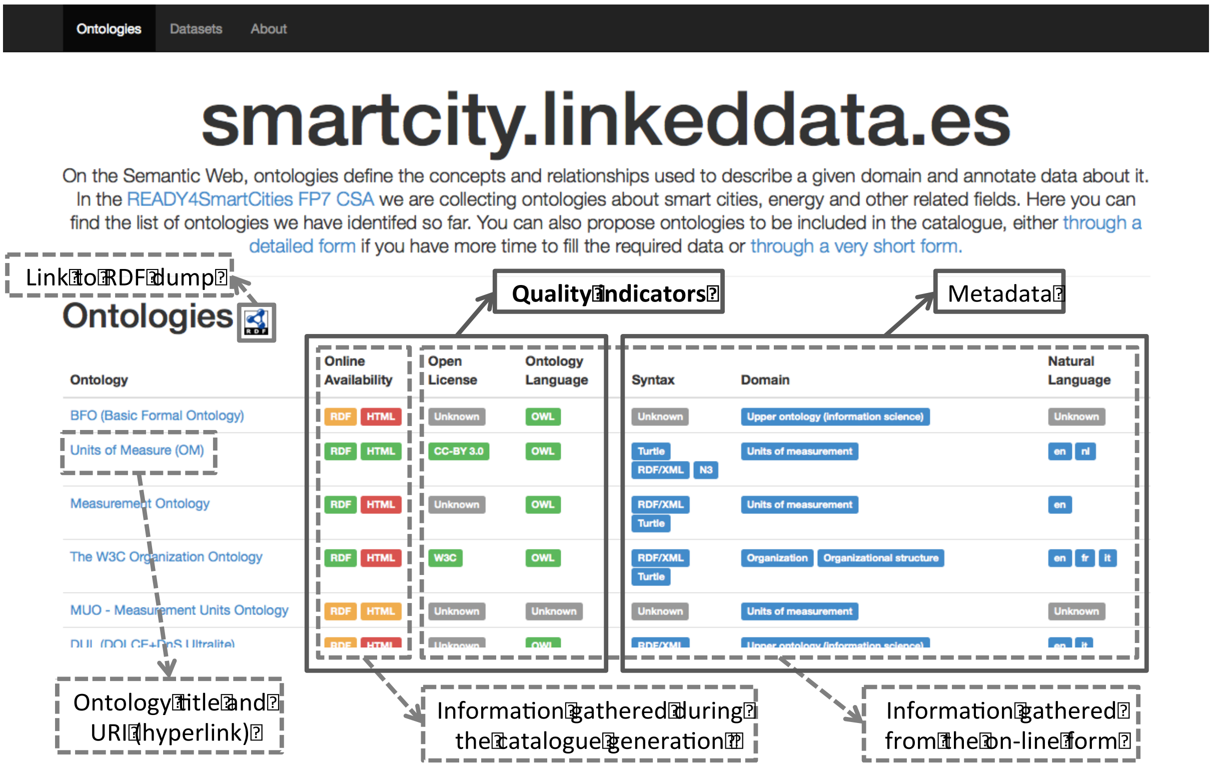

Figure 4. Screenshot of the ontology catalogue home page

Regarding this indicator, it is a good practice to provide information in different formats for a given resource in the Web [Bizer, Heath, & Berners-Lee, 2009]. In our case, the resource is the ontology itself and the different formats are its human-readable documentation (for example, an HTML page) and machine-readable information (for example, an RDF serialization). In addition, the mechanisms to provide both the HTML and RDF versions of the ontology should be compliant with the content negotiation recommendations (http://www.w3.org/TR/swbp-vocab-pub/). In order to automatically evaluate the indicator we first use Vapour (http://validator.linkeddata.org/vapour) to check whether the ontology URI provides RDF and HTML in a way compliant to content negotiation recommendations; if so, the color associated to the format is green. If for any format there is no correct content negotiation mechanism implemented, we next check whether any RDF and/or HTML resource is available even though the technical implementation is not compliant with content negotiation best practices; in that case, the format (RDF or HTML) is shown in orange. Finally, if one or both formats are not available through the URI, the format is represented in red.

It should be noted that one ontology could provide one format according to content negotiation mechanisms and the other in a non-compliant way or not even provide it. Different combinations of these cases are shown in Figure 4.

2.2 Dataset catalogue

2.2.1 Overview of the dataset catalogue

The approach followed for gathering datasets is equivalent to the one for collecting ontologies already described in Section 2.1.1. In this case, contributors and populators provide dataset metadata through an on-line form available at http://goo.gl/0ENc5h. The option for contributors to provide minimal information for datasets by means of filling in a short on-line form is available at http://goo.gl/Hvo5yX.

2.2.2 Catalogue generation

For the persistence of the dataset metadata we have followed the same approach as presented in Section 2.1.2. For this case we have reused the ontologies listed in Table 2 for describing the datasets of the catalogue.

Table 2. Vocabularies reused for describing the ontologies of the catalogue

Vocabulary	Prefix	URI
Creative Commons Rights Expression Language	cc	http://creativecommons.org/ns
DCMI Metadata Terms	dc	http://purl.org/dc/terms/
Data Catalog Vocabulary	dcat	http://www.w3.org/ns/dcat#
Ontology Metadata Vocabulary	omv	http://omv.ontoware.org/2005/05/ontology#

Figure 5 shows the ontology used to describe the datasets included in the catalogue. As represented in such figure, the central class of the model is *dcat:Dataset*, that is used to represent datasets. This class contains some attributes (or datatype properties) to represent the dataset title (*dc:title*), its description in natural language (*dc:description*), its creation date (*dc:issued*), its last modification date (*dc:modified*), and its URL (*dcat:landingPage*) that are reused from other vocabularies. For other indicators some attributes have been created in the namespace http://smartcity.linkeddata.es/def# (marked in grey in Figure 5), this attributes represent by means of Boolean values whether the dataset is available online (*availableOnline*), whether a bulk can be downloaded (*bulkAvailable*), whether the dataset can be found in digital form (*digitalForm*), whether it is free of charge (*freeOfCharge*), whether it is available in a machine readable format (*machineReadableFormat*), whether it is publicly available (*publiclyAvailable*) and whether it is up to date (*updated*).

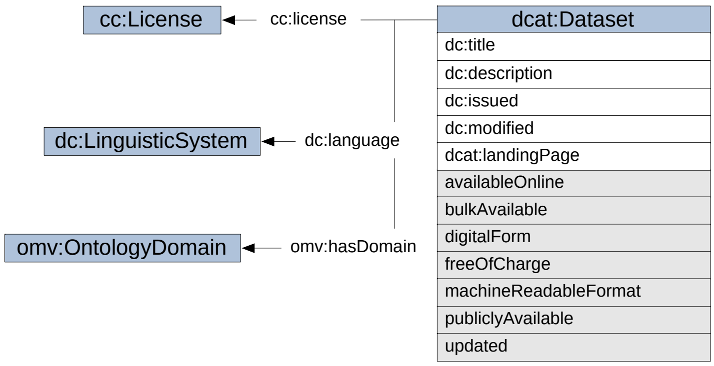

Figure 5. Ontology to represent dataset metadata

Individuals belonging to the class *dcat:Dataset* could be related to individuals belonging to other classes. This way, the domains covered by the dataset are indicated by means of the relationship *omv:hasDomain* between the classes *dcat:Dataset* and *omv:OntologyDomain*; the language in which the dataset is expressed is stated by the relationship *dc:language* between the classes *dcat:Dataset* and *dc:LinguisticSystem*; and the license of the dataset is indicated through the property *dc:license* between the classes *dcat:Dataset* and *cc:License*. An example of a dataset annotated following the ontology presented in Figure 5 is shown in Figure 6.

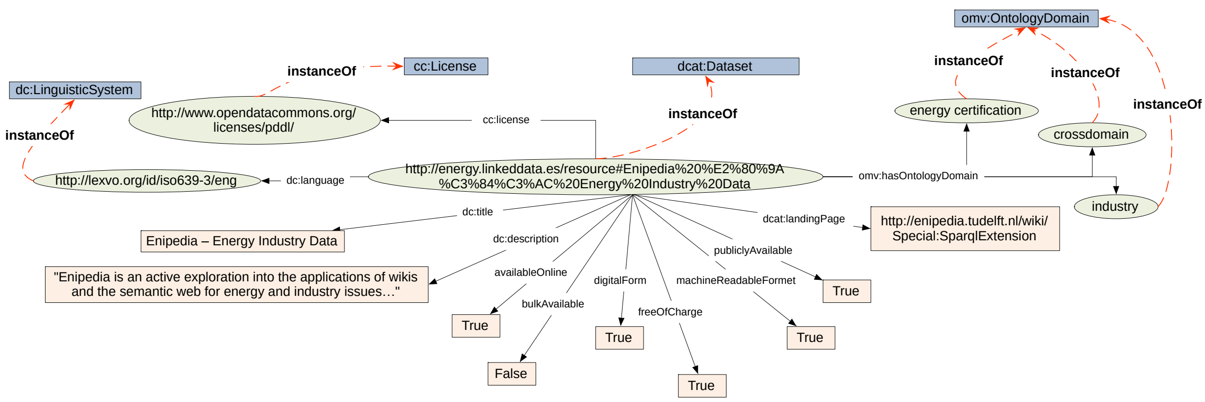

Figure 6. Example of dataset metadata representation in RDF

2.2.3 Web application

The catalogue of datasets about smart cities, energy and other related fields can be accessed through a web application available at http://smartcity.linkeddata.es/datasets/.

As shown in Figure 7 the catalogue allows visualizing metadata about the listed datasets. For each dataset, the metadata are shown in the columns. More precisely the columns “Digial form”, “Publicly available”, “Free of charge”, “Available online”, “Machine readable”, “Available in bulk”, “Open License” and “Up to date”, represent the considered quality indicators as defined in [Garcia-Castro et al, 2014] while the columns “Domain” and “Natural language” provide general information about the dataset. The values shown in each cell of the table contain different information both represented by text and by color; for ontology metadata, colors have the following meaning: "plain information" for blue and "unknown" for grey. Furthermore, in addition to the color, each cell contains detailed information when available.

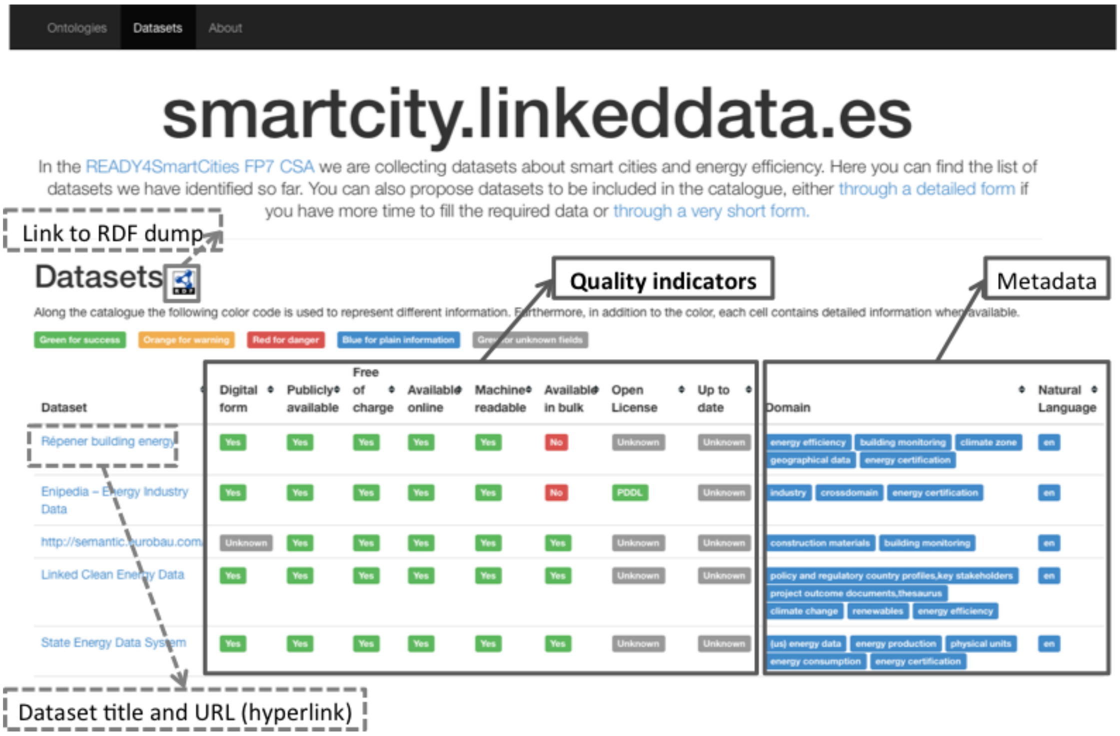

Figure 7. Screenshot of the dataset catalogue home page

2.3 Alignments catalogue

The alignment catalogue is implemented as an alignment server sharing alignments on the web. This server should be directly connected to the ontology catalogue and be able to update itself upon changes in this catalogue.

Below, we describe briefly the architecture of the alignment server.

2.3.1 Overview of the Alignment server

The Alignment server can supply alignments for people to inspect and for systems to reuse. More than a simple catalogue, it offers the opportunity to generate, organise and manipulate alignments online.

The goal of the Alignment server is that different actors can share available alignments and methods for finding alignments. Such a server enables to match ontologies, store the resulting alignment, store manually provided alignments, extract merger, transformer, mediators from those alignments.

The Alignment server is built around the Alignment API. It thus provides access to all the features of this API. The server architecture is made of three layers:

- **A storage system** providing persistent storage and retrieval of alignments. It implements only basic storage and runtime memory caching functions. The storage is made through a DBMS interface and can be replaced by any database management system as soon as it is supported by jdbc.
- **A protocol manager** which handles the server protocol. It accepts the queries from plug-in interfaces and uses the server resources for answering them. It uses the storage system for caching results.

- **Protocol plugs-in** which accept incoming queries in a particular communication system and invoke the protocol manager in order to satisfy them. These plugs-in are ideally stateless and only translator for the external queries.

This infrastructure is able to store and retrieve alignments as well as providing them on the fly. We call it an infrastructure because it will be shared by the applications using ontologies on the semantic web. However, it may be seen as a directory or a service by web services, as an agent by agents, as a library in ambient computing applications, etc.

Services that are provided by the Alignment server are:

- storing alignments, whether they are provided by automatic means or by hand;
- storing annotations in order for the clients to evaluate alignments and to decide to use one of them or to start from it (this starts with the information about the matching algorithms, the justifications for correspondences that can be used in agent argumentation, as well as properties of the alignment);
- producing alignments on the fly through various algorithms that can be extended and parametrised;
- manipulating alignments by inverting them, applying thresholds;
- generating knowledge processors such as mediators, transformations, translators, rules as well as to process these processors if necessary;
- finding similar ontologies and contacting other such services in order to ask them for operations that the current service cannot provide by itself.

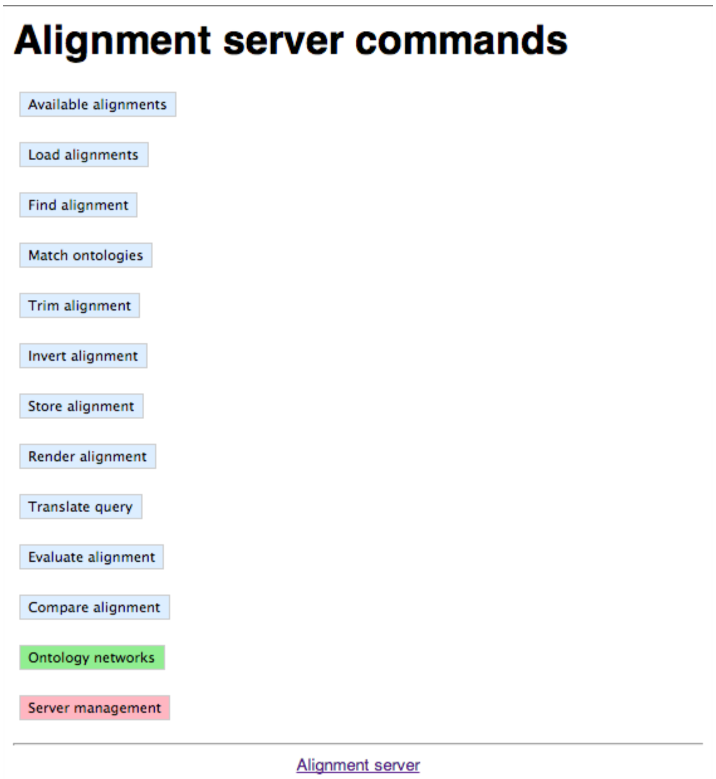

Figure 8. Menu of the services provided through the Alignment server

The menu of these services through the HTML plug-in is seen on Figure 8. For Ready4SmartCities, we introduced in the server the notion of ontology network which group together a set of ontologies and a set of alignments for better visibility.

2.3.2 Example methodology of alignment generation

In order to illustrate the Alignment API in the R4SC project we have proceeded as shown in Figure 9 and explained below.

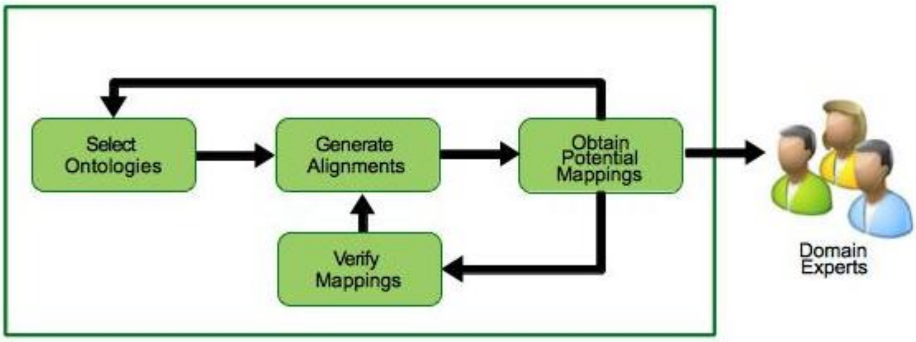

Figure 9. Obtaining potential mappings with the Alignment API

(1) Select ontologies.

(2) Generate alignments.

(3) Obtain potential alignments

(4) Verify results

This methodology has been followed to generate the alignments described in section 2.4.3.

2.3.3 Description of the Alignment server as web service and link to the ontology catalogue

This section serves as a documentation for the connection between the ontology catalogue and the Alignment server. The main point would be that it is possible to link these. This has to be performed through web services call invocation. We describe here the REST interface, however a SOAP interface is also available.

There are two main way which can be used to connect the Ontology catalogue to the Alignment server.

The first option is that the ontology catalogue redirect from one ontology to the Alignment server. This is achieved by generating the following URL in the ontology catalogue.

http://al4sc.inrialpes.fr/html/listalignments?uri1=http://www.geonames.org/ontology&uri2=all

This would redirect to the list of all alignments involving the geoname ontology as shown in the following figure:

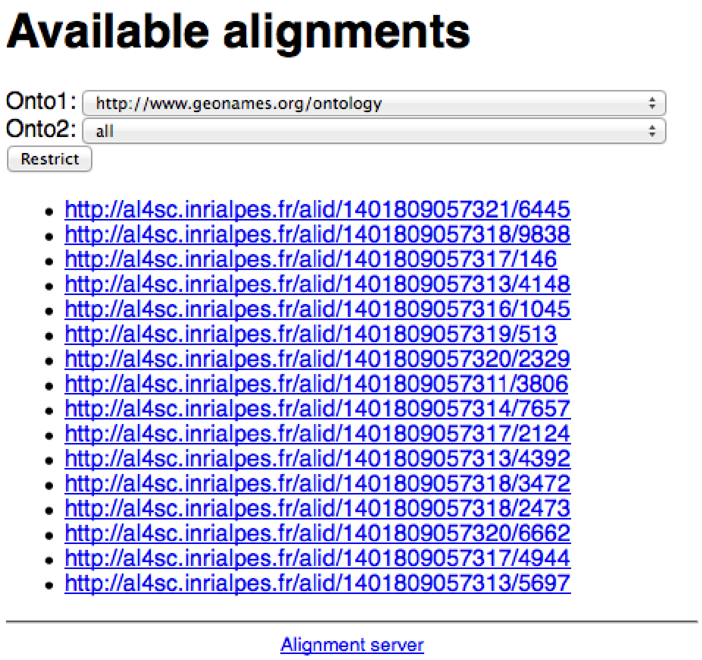

Figure 10. List of alignments involving the geoname ontology

The second option is that the Ontology catalogue uses the REST interface in order to obtain the list of available alignments. This can be achieved by using the following URI:

http://al4sc.inrialpes.fr/rest/find?onto2=http://www.geonames.org/ontology

which, in this case, may return:


```
    <findResponse
       xml:base='http://exmo.inrialpes.fr/align/service#'
       xmlns='http://exmo.inrialpes.fr/align/service#'>
      <id>29</id>
      <sender>http://al4sc.inrialpes.fr</sender>
<alignmentList>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057313/5697</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057317/4944</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057318/3472</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057313/4392</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057314/7657</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057311/3806</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057320/2329</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057319/513</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057316/1045</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057313/4148</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057317/146</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057318/9838</alid>
        <alid>http://al4sc.inrialpes.fr/alid/1401809057321/6445</alid>
      </alignmentList>
    </findResponse>
```

The obtained alignments URI may be used redirecting to the Alignment server or for further exploiting alignments through the REST interface. The REST interface is further documented at:
http://alignapi.gforge.inria.fr/rest.html

2.4 Overview of ontologies and datasets gathered during the first project year

2.4.1 Ontologies, vocabularies and standards

General overview of the Ontology Catalogue

- At the moment of writing this deliverable, the Ready4SmartCities Ontology Catalogue contained **42 ontologies**.
- UPM analysed these ontologies in order to provide a general overview of the ontology languages and format used, the natural languages in which ontologies are expressed, and the licenses attached to these ontologies.
- INRIA performed a content analysis covering other relevant aspects

The corpus ranges from fairly tiny ontologies (reorganization vocabulary: 7 entities) to huge ones (sumo: 90971 entities). It is not always easy to determine which of these entities are local and which belong to other ontologies because standard namespace is not always set-up appropriately.

The most common ontology language in the Ready4SmartCities Catalogue is **OWL**, followed by RDF-S. 40 ontologies are implemented in OWL, while only 3 ontologies are coded in RDF-S. The distribution of ontology languages in the catalogue is shown in Figure 11. It is worth mentioning that five ontologies are in more than one ontology language. These ontologies are Timeline Ontology, Data Cube, DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), SUMO (Suggested Upper Merged Ontology), and BFO (Basic Formal Ontology).

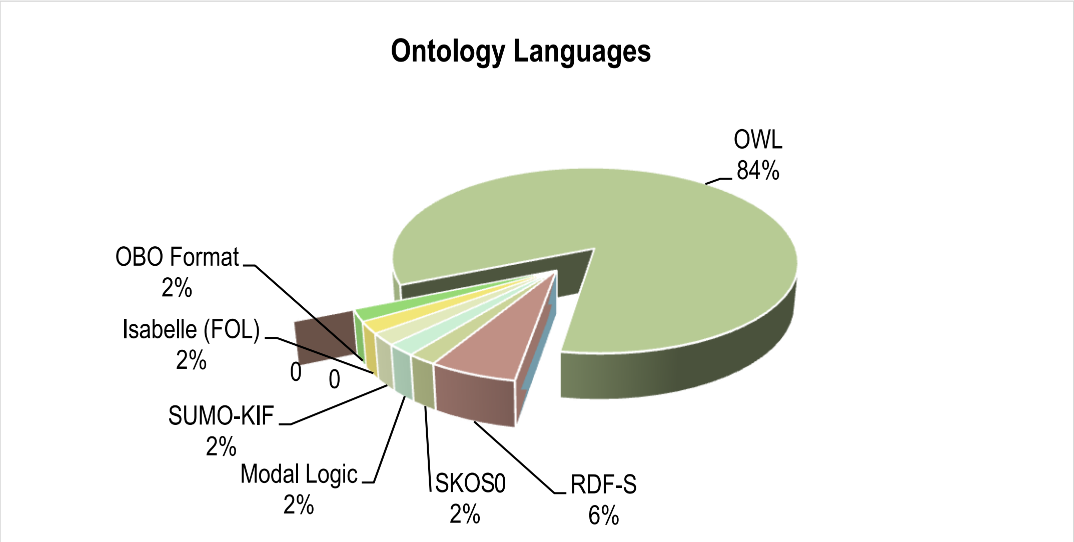

Figure 11. Ontology languages distribution

Regarding the ontology syntaxes, **RDF/XML** is the most usual one followed by Turtle. 37 ontologies are written using the RDF/XML syntax, while 8 are using the Turtle syntax. As in the case of ontology languages, there are six ontologies in the catalogue provided with more than one format. These ontologies are Units of Measure (OM), Measurement Ontology, The W3C Organization Ontology, IFC2X3 - University of Ghent, Places Ontology, and Registered Organization Vocabulary. It is important to mention that for two ontologies the ontology syntax is not known. The distribution of ontology formats in the catalogue is shown in Figure 12.

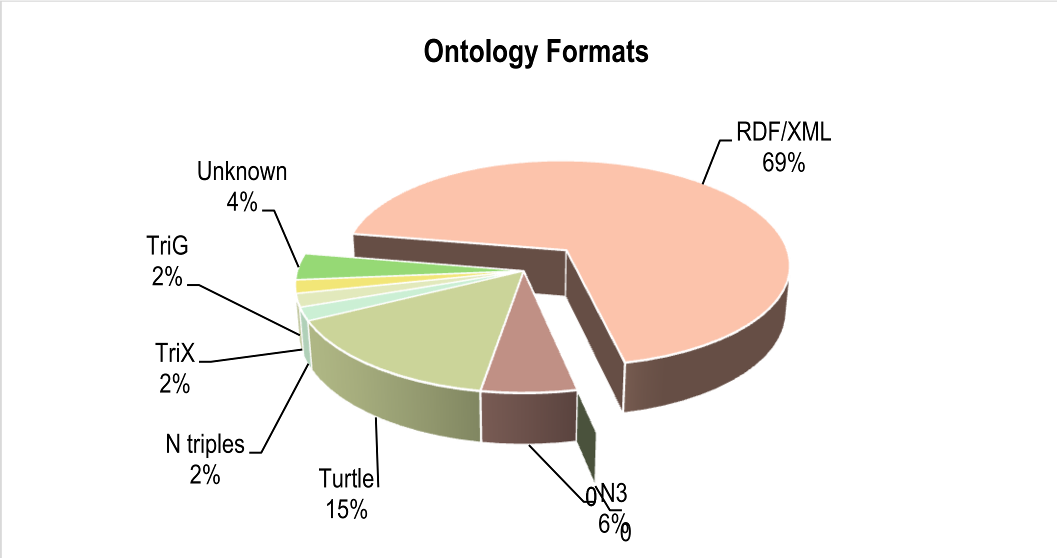

Figure 12. Ontology formats distribution

With respect to the natural language used for naming ontology elements, the most common one is **English** (40 ontologies are written in such a language). Surprisingly, the second position in the natural language ranking is for 'Unknwon' (two ontologies are annotated with unknown language[18]) and for Italian (2 ontologies in the catalogue are in Italian). There are four ontologies in the catalogue that are written in more than one natural language. These ontologies are Geonames, Units of Measure (OM), The W3C Organization Ontology, and DUL (DOLCE+DnS Ultralite). The distribution of natural languages used in the catalogue is shown in Figure 13.

[18] This situation occurs because the ontology documention does not provide information about the natural language used. In addition, the code for those ontologies was not available, so it was not possible to discovery the language used for naming ontology elements.

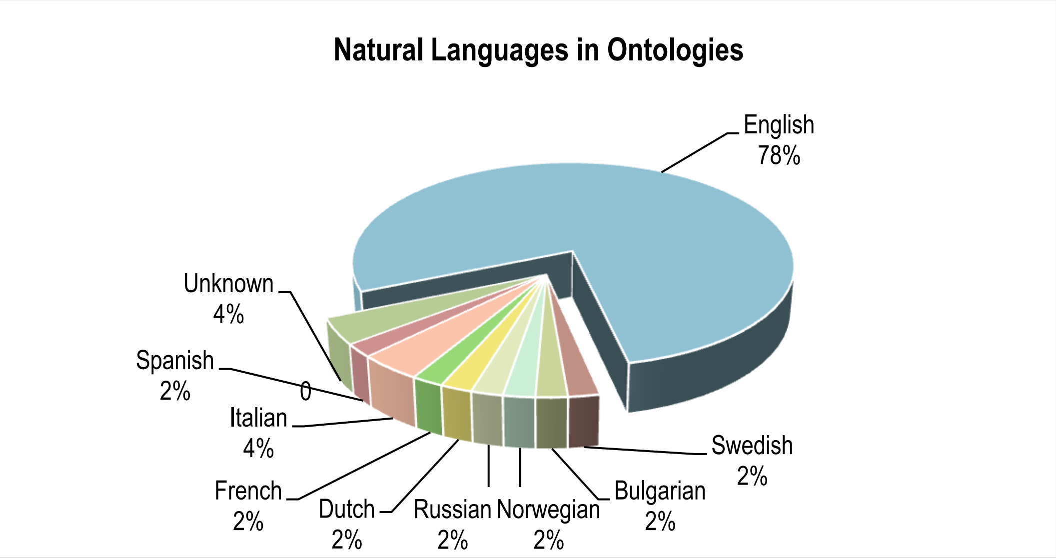

Figure 13. Distribution of natural languages in ontologies

Most of the ontologies (28 out of 42) in the catalogue have no information about licenses (ontology license is **Unknown**). In those cases in which authors provide license information, the most usual licenses are the W3C software license (4 ontologies have this type of license) and the CC-BY Creative Commons Attribution Unported (Open) (another 4 ontologies have this kind of license). The distribution of ontology licenses in the catalogue is shown in Figure 14.

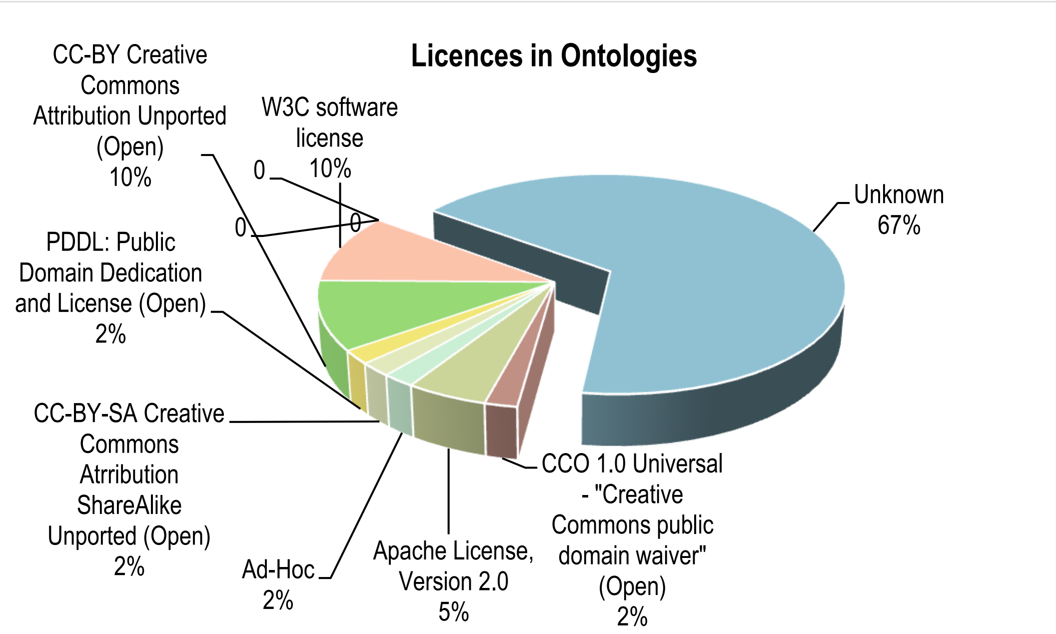

Figure 14. Ontology licenses distribution

UPM also analyzed the 42 ontologies in the catalogue with respect to the following quality indicators: online availability of ontologies and open license attached to the ontologies.

Regarding the online availability of ontologies, UPM performed two analyses: the first one refers to the availability of ontology code (RDF) and the second one refers to the availability of ontology documentation (HTML). In both cases[19] the study refers to:

- whether the corresponding content (RDF or HTML) can be retrieved in the given format according to content negotiation best practices for publishing RDF vocabularies ("Content Negotiation")
- whether the content can be retrieved even though no content negotiation mechanisms are properly set up ("No Content Negotiation")
- whether the content can not be retrieved ("Not Available")
- other situations[20] ("Unknown")

In the first case, **32 out of 42 ontologies can be retrieved in RDF**. However, 22 out of these 32 are retrieved although content negotiation mechanisms have not been properly set up. In addition, 4 ontologies cannot be retrieved in RDF and 6 probably are not available or are published in a wrong way. The distribution of RDF availability in the catalogue is shown in Figure 15.

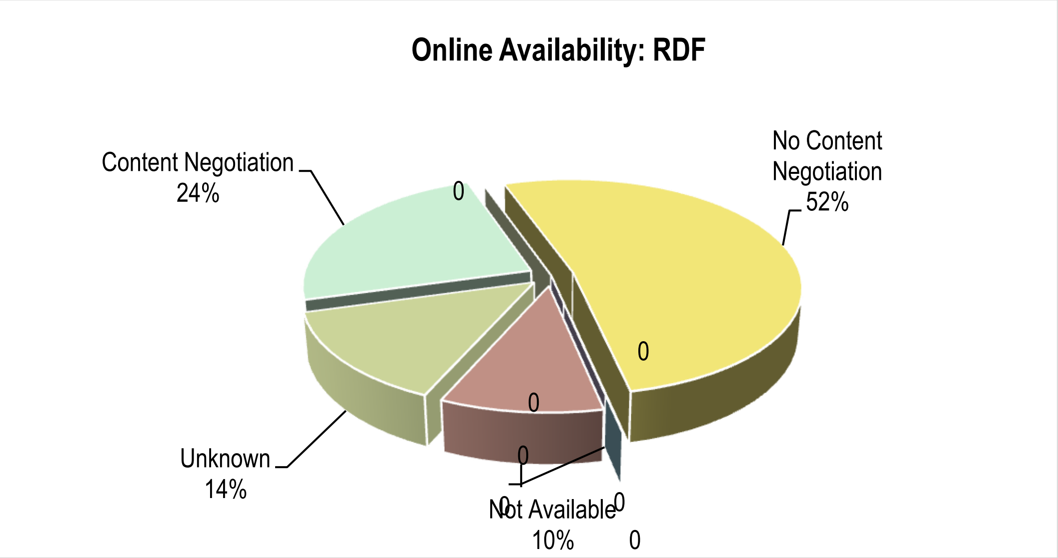

Figure 15. Distribution of RDF availability

In the second case, **13 out of 42 ontologies can be retrieved in HTML**; 1 out of these 13 is retrieved though content negotiation mechanisms have not been properly set up. In addition, 18 ontologies cannot be retrieved in HTML and 11 probably are not available or are published in a wrong way. The distribution of HTML availability in the catalogue is shown in Figure 16.

[19] In order to check content negotiation mechanisms for RDF and HTML formats, the linked data validator Vapour (http://validator.linkeddata.org/vapour) is used while the RDF content of the available ontologies are loaded in a JENA (http://jena.apache.org/) model.

[20] This means that Vapour provides an exception.

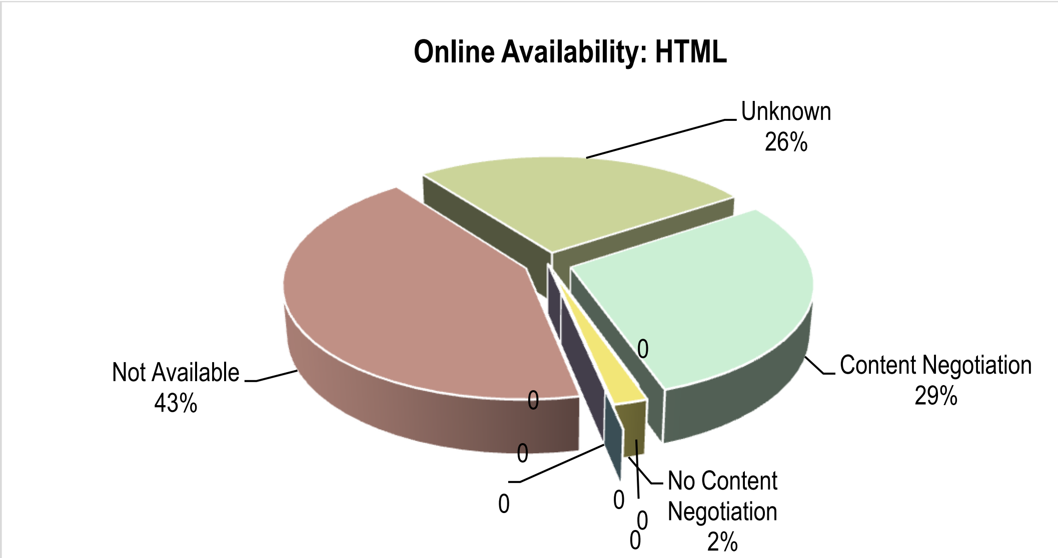

Figure 16. Distribution of HTML availability

With respect to the licenses used for the ontologies, **14 out of 42 ontologies have an open license**. However, there are 28 ontologies in the catalogue that have no information about license. The distribution of licenses types is shown in Figure 17.

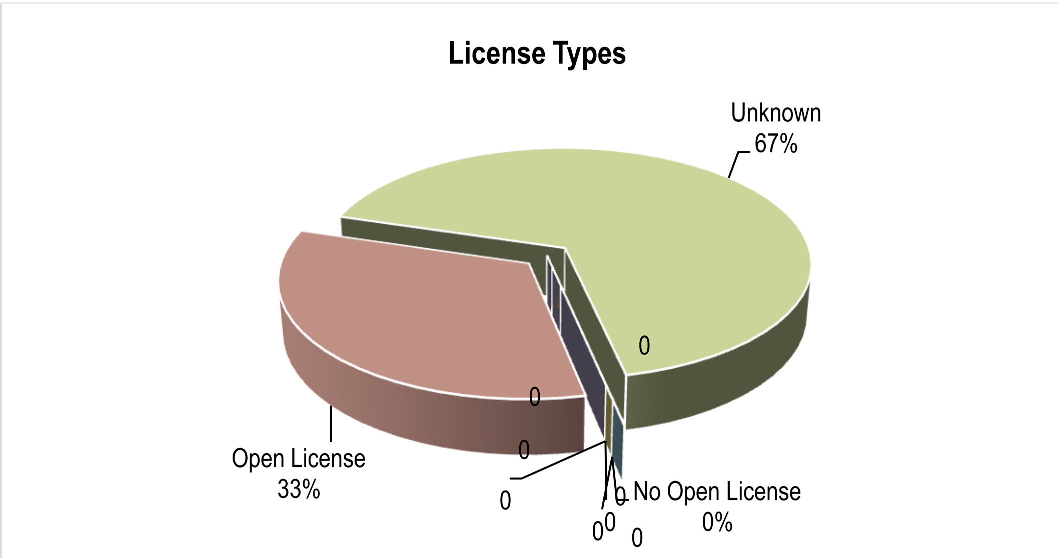

Figure 17. Distribution of licenses types

Domain coverage analysis

Regarding the specific domains identified in Deliverable D3.1, at first the set of ontologies in the catalogue covers

- the ***five domains identified for Level 1***, that is, Temporal, Organisational, Statistical, Spatial/Geographical, and Measurement
- ***3 out 7 domains identified for Level 2***. These domains are Energy, Weather, and Building. Thus, Climate Zone, Environmental, Occupancy, and User Behaviour do not seem to be covered.

Total figures of ontologies related with Level 1 domains and with Level 2 domains are shown respectively in Figure 18 and Figure 19.

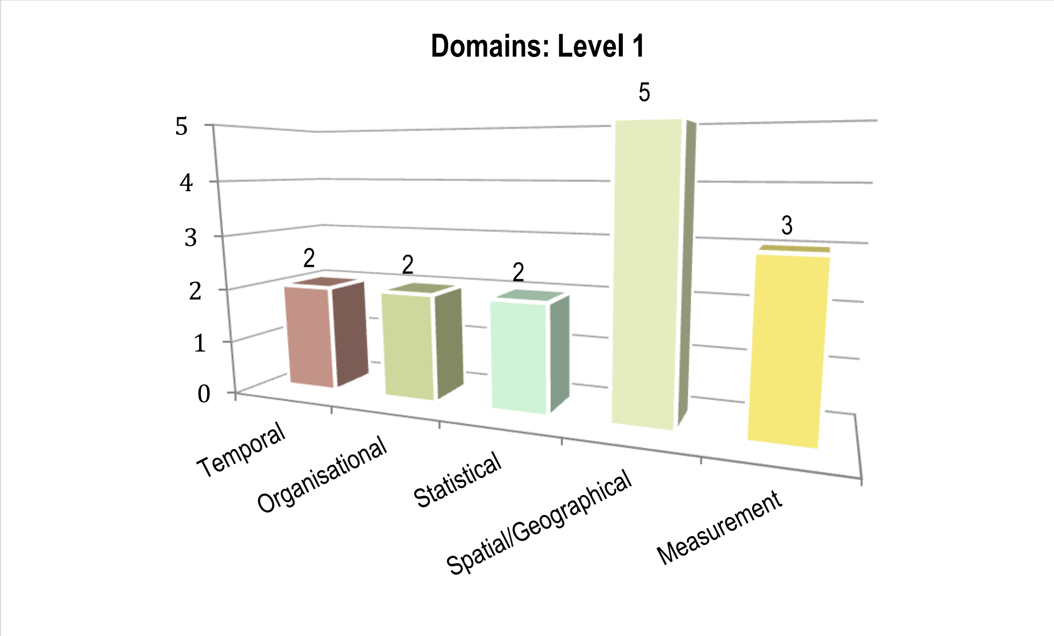

Figure 18. Number of ontologies in Level 1 domains

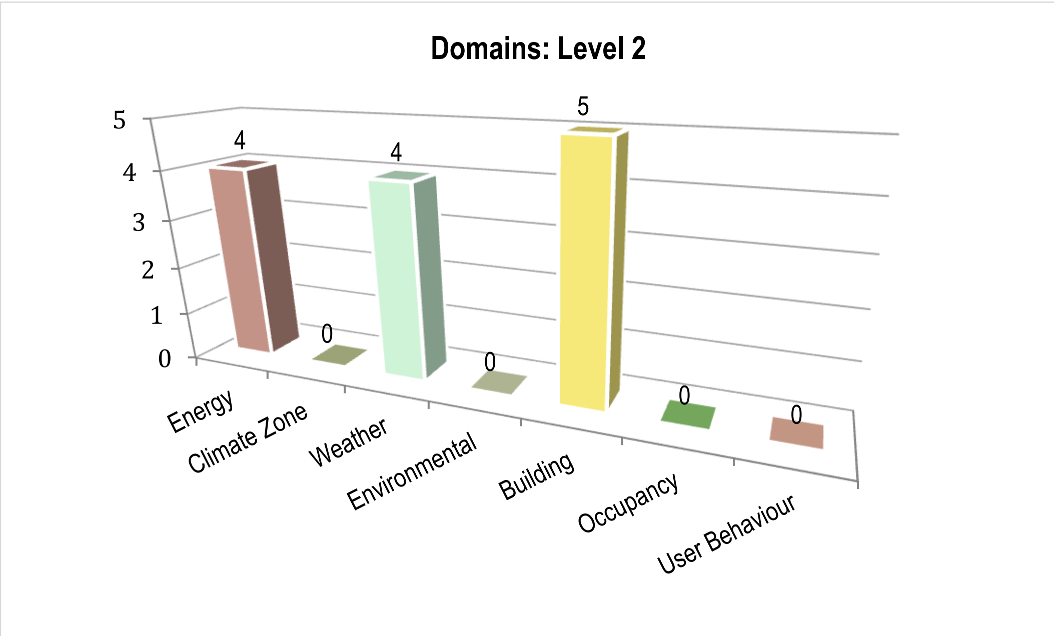

Figure 19. Number of ontologies in Level 2 domains

UPM also analyzed the list of domains attached to the ontologies by catalogue populators. As a result of this analysis, 16 new domains were identified. They are shown in Table 4. Such domains were studied with the aim of finding some relations with the domains established in Deliverable D3.1. UPM found the following outcomes

- the Indicator domain can be considered to be a subdomain of Measurement (a Level 1 domain)

- the Airport, School Building and Building Performance domains are related to Building domain (a Level 2 domain)
- the Building Usage and Preferences domains can be considered as subdomains of User Behaviour (a Level 2 domain).

These findings imply new figures for ontologies related to Level 1 domains and to Level 2 domains. In the latter case, a new domain is covered. Thus, ***4 out 7 domains identified for Level 2*** are covered. Comparisons between number of ontologies related to domains strictly identified in Deliverable D3.1 and number of ontologies related to those domains identified in UPM analysis are shown in Figure 20.

Table 3. New domains identified

Indicator	Airport	School Building
Building Performance	Building Usage	Preferences
Provenance	Top Level	Generic
Device	Sensor	IFC
Smart Environment	Home Automation	Process
Urban Planning		

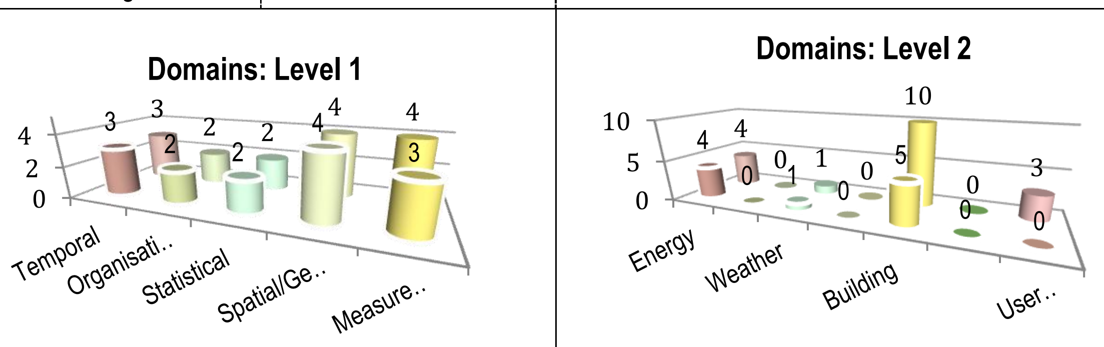

Figure 20. Number of ontologies per domain

In addition, the following ***ten new domains*** are also covered: Provenance, Top Level, Generic, Device, Sensor, IFC, Smart Environment, Home Automation, Process and Urban Planning. The numbers of ontologies per new domain are shown in Figure 21.

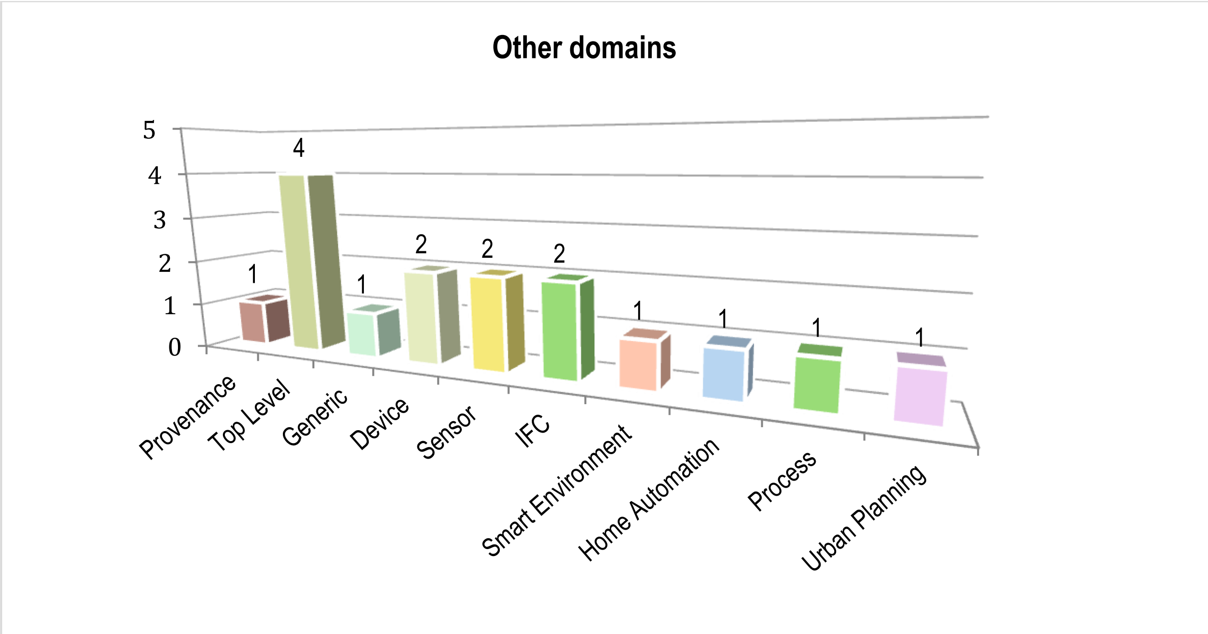

Figure 21. Number of ontologies in new domains

Thus, the map of domains (level 1, level 2 and others) and ontologies in the current version of the Ready4SmartCities catalogue can be represented as shown in the following figures[21].

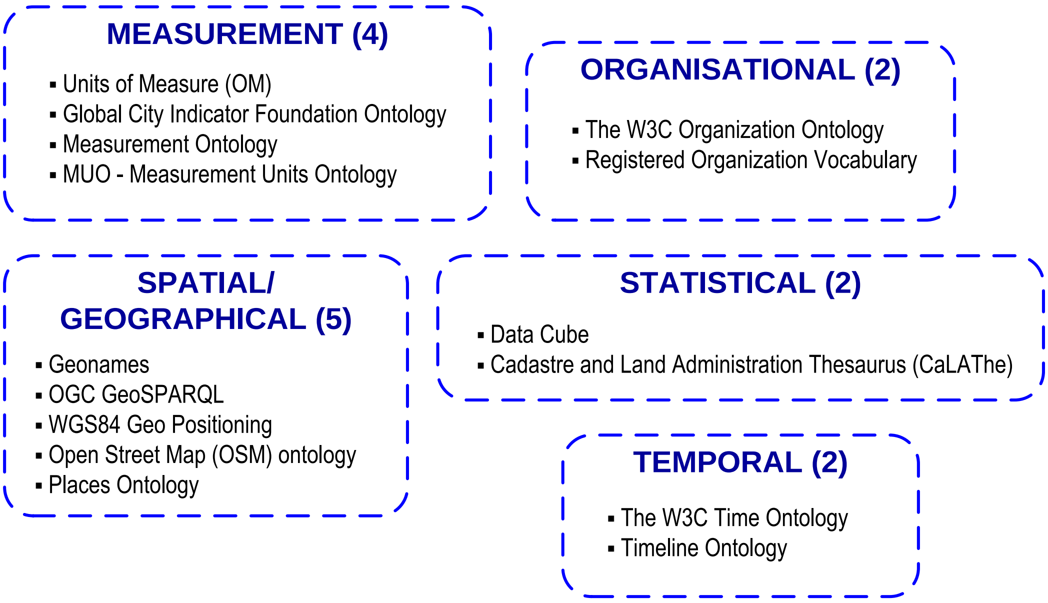

Figure 22 Ontologies in Level 1 Domains

[21] Domains with no ontologies are represented using striped squares.

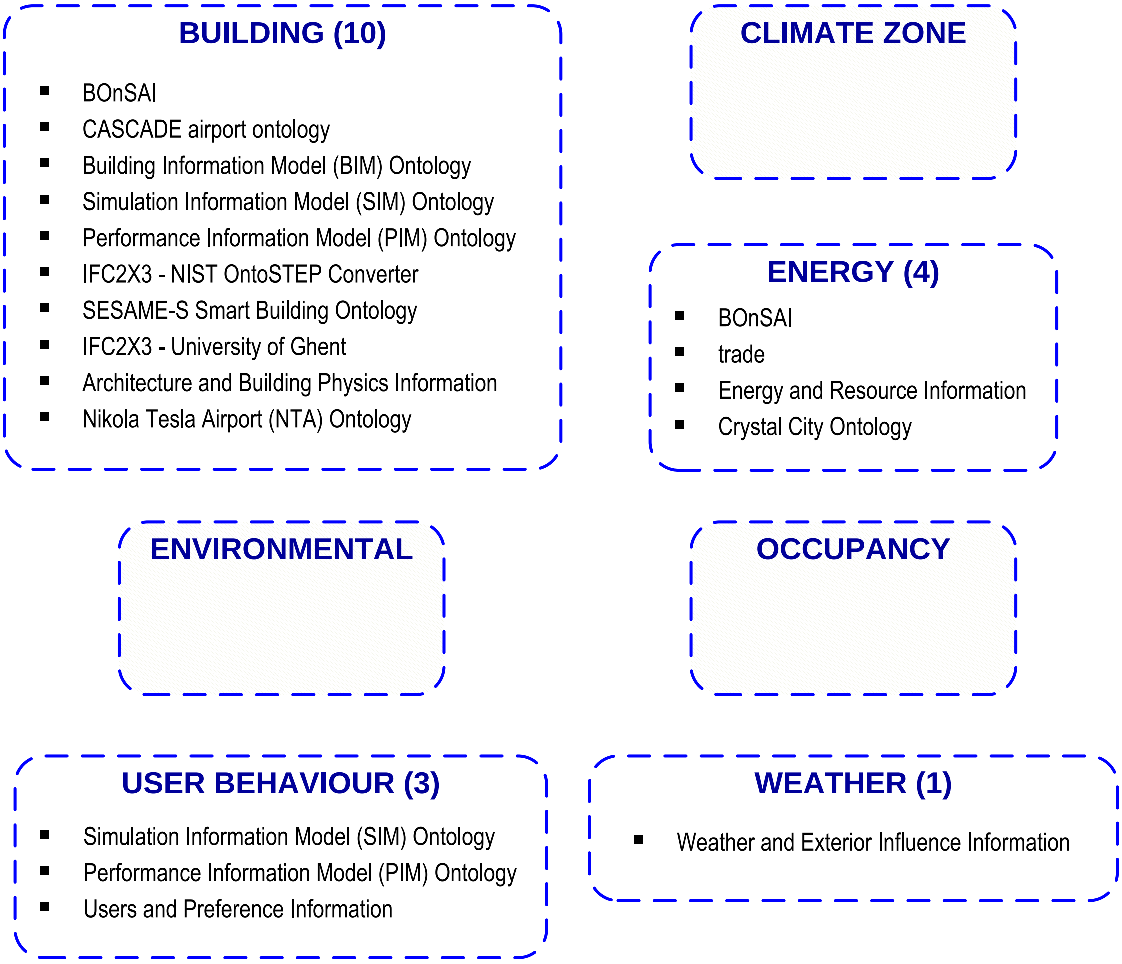

Figure 23 Ontologies in Level 2 Domains

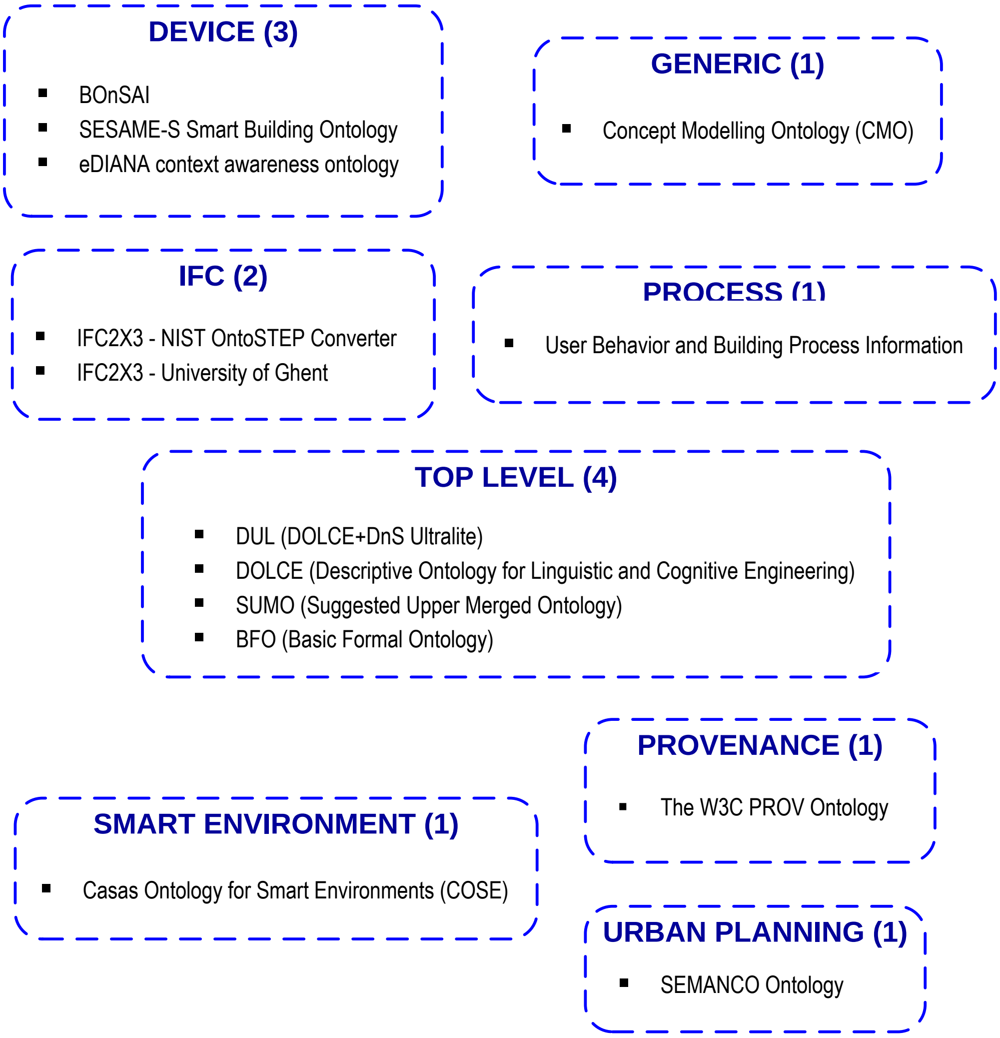

Figure 24 Ontologies in other Domains

Appreciation of some of the ontologies

It can be observed that some of the presented ontologies have played the game of extending and refining existing ontologies. Others have taken into account other ontologies.

On the other side, some other ontologies started from scratch ignoring previous related effort. This can be for good reasons (no compatible ontologies available). However, when these ontologies do not even use other ontologies for documentation purposes, this is a clear sign that their designers did not made effort to include them in a wider landscape. This definitely concerns those ontologies which only use xsd, rdf, rdfs and owl as namespace. The result is a scattered set of ontologies which indeed would benefit from alignment (see chapter2.3). This is what we consider below.

While investigating these ontologies we remarked that one of the ontology had many terms starting with “Ifc”. This is not anymore good practice on the worldwide web because, this Ifc string does not scale. URI are used for qualify term and should be used for that purpose. Since this prefix can be an obstacle to matching, and computing distances, we duplicated this ontology creating an ontology ifc2 which is used below.

2.4.2 Datasets

At the moment of writing this deliverable, the Ready4SmartCities Dataset Catalogue contained nine datasets. Due to the small sample, a statistical analysis does not make sense in this case; therefore a summary of the main characteristics of the datasets is presented here.

The datasets cover the domains *building design and measurement*, *building operation*, *outcome metrics*, and *weather and climate data*. The availability of creation dates and update frequencies for the identified datasets suggest that they have all been created in the last 2 to 5 years. For some datasets no license has been given (unknown); the datasets with a license include *CC-BY-SA Creative Commons Attribution-ShareAlike Unported (Open)*, *OGL* and *PDDL*.

The format of the datasets is usually N triples and RDF. Out of the nine datasets in the catalogue, just two have been recorded as originating from a European project. Two of the datasets are not available in bulk.

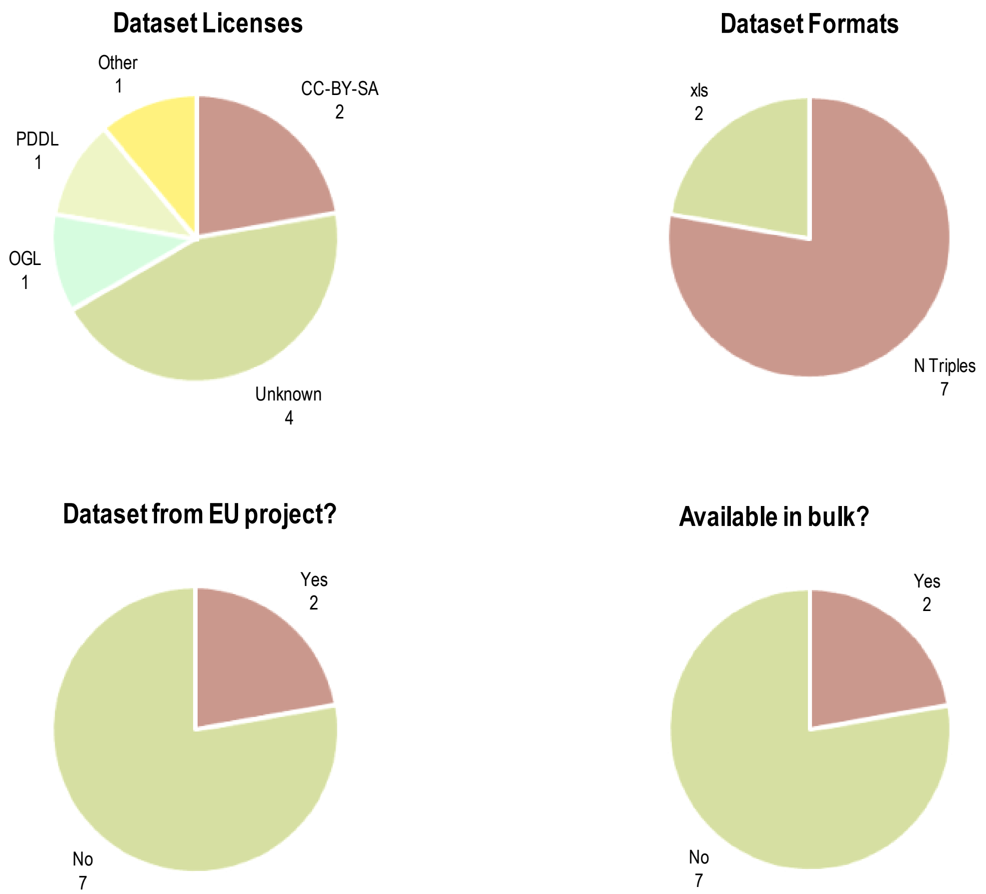

Figure 25. Overview of some dataset attributes

2.4.3 Ontology alignments and data links

This chapter deals with finding ontology alignments, and later, links from the ontologies and data source of the corresponding libraries. Indeed, as expected, there are not many alignments available. Some stakeholders told us that such alignments were part of their proprietary ontologies. Isolating and sharing alignments, however, has the benefit that it can be adopted and improved by others.

So, we take the active step of trying to obtain alignments from the ontologies themselves. For this, we first compute various distance measures between the available ontologies (see 2.4.3.3) in order to gain an insight of ontologies that would be more easily matched. Then, we perform ontology matching and observe what are the actual correspondences between the existing ontologies (see 2.4.3.4).

2.4.3.1 Content analysis

Content analysis inspects the ontologies and provides some statistics about their content.

- 12 ontologies were not available for download.
- 30 ontologies were downloaded. Among them, 28 were in XML and 2 in Turtle. 27 contained an OWL ontology, 1 contained an RDF ontology and 2 RDF Descriptions.
- The corpus ranges from fairly tiny ontologies (reorganisation vocabulary: 7 entities) to huge ones (sumo: 90971 entities). It is not always easy to determine which of these entities are local and which belong to other ontologies because standard namespace is not always set-up appropriately.

2.4.3.2 Reference analysis

Reference analysis considers how these ontologies are connected to each other by their designers. Such references may indicate alignments which have been embedded within an ontology and which would gain being made explicitly as alignments.

This analysis has been processed manually; it could be possible to do it automatically.

We identified various external URIs found in these ontologies. There can roughly be three types of references to another ontology within an ontology:

- declaration as a prefix;
- explicitly imported through owl:import
- directly used by referencing the full URI of an entity of another ontology.

We have identified the two first categories and sampled for the others. The non-systematic use of prefixes in ontologies renders the task difficult.

We considered all the 38 other vocabularies as interesting to study. However, quite some of them are also technical vocabularies which may not be interesting. In fact, the declared namespaces are more often the mark of the tool which created the ontology rather than that of the ontology itself.

Most ontologies use a common core of prefixes for XML Schema, RDF, RDFS, OWL. These are the vocabularies in which ontologies are expressed and we will not consider them further. There also are references to other vocabularies for expressing ontologies: OWL2XML, SKOS and DAML (an ancestor of OWL), SWRL for rules, or protege and owlapi corresponding to the tools used for representing ontologies.

Some other ontologies are mostly used for documenting the ontologies: DC (Dublin core vocabulary), dcterms, CC (Creative commons), vann (Vocabulary annotation), adms (asset description metadata schema), void (linked data set description), voaf (vocabulary of a friend).

In the end, the relevant vocabularies are the following:

- Minimal or base vocabularies: foaf (people and organisations), time (temporal location), timezone (temporal location), temporal (temporal relations), vcard (people), bibo (bibliography), wgs84 (geographical location), qudt (units), gml (geographical markup), scovo (statistical core vocabulary).
- Technical vocabularies: owls (services)

- General purpose vocabularies: CoDaMoS (Pervasive services), opencyc, goodrelations, schema, sumo, dul

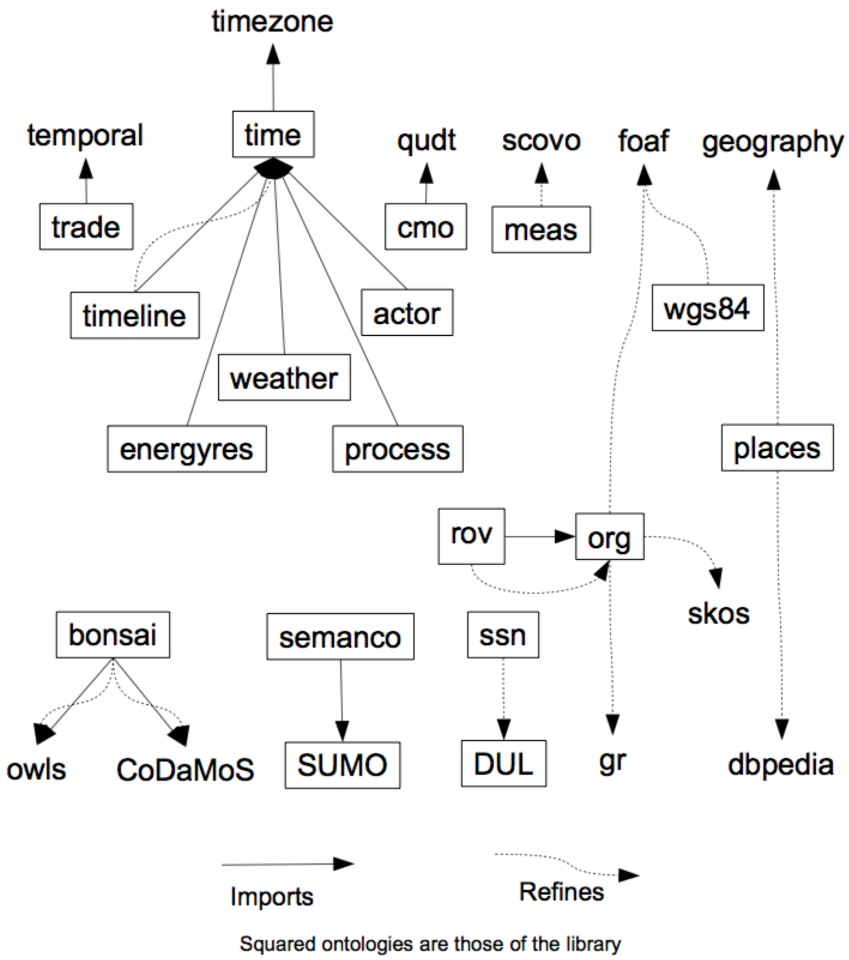

Figure 26. Relations between the 16 ontologies which import or refine others. At the bottom of the picture are general ontologies and at the top specialized ontologies.

Very few ontologies extend the others, in the sense that they refine some of their instances. The table below presents examples of refinement within the 30 ontologies. These 9 ontologies are likely to be the only ontologies that refine others.

Ontology source	Ontology « refined »	Entity refined
Bonsai	CoDaMoS	Resource
	owl-s	Service
Org	foaf	Organization
	foaf	Agent
	skos	Concept

	skos	notation
	gr	BusinessEntity
	prov	wasDerivedFrom
Places	dbpedia	City
	dbpedia	Continent
	geograophy	Continent
Rov	org	FormalOrganization
	org	classification
	dcterms	Agent
Semanco	sumo	Identifier
	sumo	TimeDuration
	sumo	TemperatureMeasure
	sumo	StationaryArtifact
	sumo	Building
Ssn	dul	DesignedArtefact
Timeline	time	Instant
	time	hasEnd
Um	om	Unit_of_measure
	om	Ratio_scale
wgs84	foaf	based_near

The W3C time ontology is the one which is the most reused (5 times), foaf is refined twice. All other ontologies are imported or refined only once. Some ontologies of the panel are reused or imported by other ontologies from the panel (time, org, SUMO and DUL).

2.4.3.3 Distance analysis

Distance analysis takes the ontologies and computes distances between them. We have applied it to the 30 ontologies above but for sumo because it is too large (we could include it later). To these we have added the created ifc2 ontology. We have used for that purpose the OntoSim library (http://ontosim.gforge.inria.fr) and we have used two simple distances:

- TF-IDF simple gathers all strings associated to an ontology (i.e., labels plus comments) and creates a bag of words for each ontology. These are compared with the TF-IDF metrics;
- lexical+hungarian method compares in each ontology the names of concepts and computes a similarity with Jaro-Winckler measure, then matches them one-to-one with the Hungarian method. The average number of matched concepts similarity gives the similarity between ontologies.

Two ontologies provided no results when computing distances: wgs84 and prov. The first one is in RDFS and the second one is a syntactically incorrect OWL ontology.

	actor	bfo	bonsai	building	cmo	cose	cube	dog	dolce	dul	energyresource	geonames	ifc2	ifc2x3	meas	muo	ogc	org	places	process	prov	rov	semanco	ssn	time	timeline	trade	um	weather	wgs84
actor	1.00	.05	.04	.16	.09	.02	.04	.06	.07	.07	.08	.03	.01	.01	.03	.10	.02	.02	.03	.26	.00	.07	.08	.08	.01	.04	.09	.04	.15	.00
bfo		1.00	.06	.06	.09	.02	.05	.08	.28	.14	.08	.06	.01	.01	.04	.11	.07	.04	.06	.04	.00	.11	.13	.19	.00	.08	.10	.07	.03	.00
bonsai			1.00	.05	.06	.04	.03	.11	.09	.07	.13	.03	.00	.00	.01	.09	.02	.02	.02	.08	.00	.06	.09	.10	.01	.06	.07	.07	.09	.00
building				1.00	.07	.02	.05	.08	.07	.09	.10	.04	.01	.01	.02	.06	.03	.01	.02	.19	.00	.03	.08	.09	.00	.05	.06	.04	.12	.00
cmo					1.00	.03	.11	.11	.14	.14	.14	.06	.01	.01	.16	.25	.08	.05	.06	.08	.00	.08	.16	.17	.02	.10	.12	.33	.10	.00
cose						1.00	.01	.04	.03	.02	.04	.01	.00	.00	.01	.03	.00	.01	.00	.02	.00	.01	.03	.03	.01	.02	.01	.02	.01	.00
cube							.92	.07	.08	.10	.06	.03	.01	.01	.11	.10	.08	.04	.03	.06	.00	.09	.08	.13	.00	.05	.07	.07	.05	.00
dog								1.00	.08	.08	.56	.07	.01	.01	.08	.11	.03	.02	.05	.09	.00	.04	.13	.12	.01	.08	.08	.10	.08	.00
dolce									1.00	.31	.09	.08	.01	.01	.08	.24	.08	.06	.07	.10	.00	.11	.21	.37	.00	.13	.15	.11	.10	.00
dul										1.00	.07	.06	.01	.01	.06	.14	.05	.08	.05	.10	.00	.13	.14	.55	.00	.09	.11	.08	.07	.00
energyresource											1.00	.07	.01	.00	.04	.09	.03	.02	.04	.13	.00	.04	.15	.10	.01	.07	.10	.09	.08	.00
geonames												1.00	.01	.00	.03	.04	.03	.02	.13	.04	.00	.04	.08	.08	.00	.03	.07	.04	.04	.00
ifc2													1.00	.53	.00	.00	.00	.00	.01	.01	.00	.00	.01	.01	.00	.00	.01	.00	.00	.00
ifc2x3														1.00	.00	.00	.00	.00	.00	.01	.00	.00	.01	.01	.00	.00	.01	.00	.00	.00
meas															1.00	.23	.03	.02	.04	.03	.00	.04	.10	.08	.00	.04	.08	.15	.03	.00
muo																1.00	.03	.03	.03	.10	.00	.08	.15	.19	.01	.09	.10	.21	.10	.00
ogc																	1.00	.02	.03	.02	.00	.04	.05	.06	.00	.03	.03	.03	.01	.00
org																		1.00	.04	.02	.00	.08	.06	.11	.00	.02	.12	.03	.02	.00
places																			1.00	.02	.00	.03	.07	.08	.00	.05	.05	.03	.03	.00
process																				1.00	.00	.06	.08	.13	.01	.07	.08	.06	.14	.00
prov																					.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
rov																						1.00	.08	.15	.00	.04	.14	.03	.03	.00
semanco																							1.00	.22	.01	.12	.17	.14	.09	.00
ssn																								1.00	.00	.12	.15	.13	.10	.00
time																									1.00	.03	.00	.02	.01	.00
timeline																										1.00	.10	.09	.07	.00
trade																											1.00	.07	.05	.00
um																												1.00	.09	.00
weather																													1.00	.00
wgs84																														.00

Figure 27. TF-IDF distance analysis

	actor	bfo	bonsai	building	cmo	cose	cube	dog	dolce	dul	energyresource	geonames	ifc2	ifc2x3	meas	muo	ogc	org	places	process	prov	rov	semanco	ssn	time	timeline	trade	um	weather	wgs84
actor	1.00	.66	.72	.67	.75	.71	.66	.75	.66	.72	.75	.72	.80	.79	.68	.68	.67	.69	.66	.77	NaN	.70	.79	.72	.64	.66	.71	.76	.73	NaN
bfo		1.00	.68	.63	.76	.78	.66	.70	.68	.65	.71	.61	.87	.84	.62	.66	.51	.63	.50	.73	NaN	.64	.73	.69	.67	.64	.68	.71	.68	NaN
bonsai			1.00	.71	.77	.75	.73	.78	.66	.72	.78	.72	.82	.80	.69	.70	.71	.74	.72	.72	NaN	.72	.78	.72	.71	.71	.70	.78	.70	NaN
building				1.00	.70	.75	.66	.72	.64	.68	.75	.67	.87	.82	.63	.63	.61	.64	.61	.76	NaN	.60	.78	.70	.66	.67	.69	.71	.71	NaN
cmo					1.00	.80	.82	.71	.69	.74	.72	.60	.79	.77	.84	.81	.70	.75	.74	.73	NaN	.75	.69	.74	.83	.77	.68	.74	.74	NaN
cose						1.00	.74	.80	.71	.72	.79	.76	.81	.78	.85	.78	.75	.77	.77	.70	NaN	.74	.78	.73	.74	.75	.68	.81	.70	NaN
cube							1.00	.75	.64	.71	.74	.72	.83	.81	.65	.68	.65	.63	.62	.71	NaN	.69	.75	.72	.65	.68	.70	.74	.69	NaN
dog								1.00	.68	.69	.83	.57	.78	.76	.73	.73	.70	.74	.71	.75	NaN	.76	.72	.69	.79	.75	.67	.76	.75	NaN
dolce									1.00	.70	.68	.62	.80	.77	.65	.72	.59	.68	.60	.69	NaN	.64	.72	.70	.63	.65	.66	.70	.65	NaN
dul										1.00	.70	.61	.83	.82	.68	.69	.59	.65	.57	.69	NaN	.68	.74	.84	.70	.68	.66	.69	.66	NaN
energyresource											1.00	.57	.80	.77	.71	.70	.70	.72	.71	.75	NaN	.74	.75	.70	.79	.73	.68	.75	.75	NaN
geonames												1.00	.79	.77	.68	.66	.60	.63	.62	.70	NaN	.64	.69	.64	.77	.69	.62	.61	.71	NaN
ifc2													1.00	.94	.89	.90	.83	.88	.85	.79	NaN	.83	.72	.85	.81	.85	.76	.77	.77	NaN
ifc2x3														1.00	.88	.86	.81	.86	.83	.78	NaN	.83	.70	.83	.81	.84	.75	.75	.76	NaN
meas															1.00	.75	.57	.66	.61	.70	NaN	.59	.74	.69	.82	.68	.64	.81	.68	NaN
muo																1.00	.56	.64	.58	.71	NaN	.64	.76	.71	.72	.67	.66	.80	.69	NaN
ogc																	1.00	.50	.47	.68	NaN	.58	.73	.61	.69	.66	.63	.67	.69	NaN
org																		1.00	.51	.72	NaN	.68	.77	.69	.69	.65	.68	.70	.72	NaN
places																			1.00	.69	NaN	.61	.74	.62	.69	.63	.62	.67	.68	NaN
process																				1.00	NaN	.71	.77	.69	.74	.71	.70	.75	.71	NaN
prov																					NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
rov																						1.00	.77	.70	.66	.66	.65	.71	.71	NaN
semanco																							1.00	.75	.81	.77	.74	.69	.76	NaN
ssn																								1.00	.71	.70	.65	.71	.66	NaN
time																									1.00	.75	.69	.86	.69	NaN
timeline																										1.00	.66	.73	.68	NaN
trade																											1.00	.71	.68	NaN
um																												1.00	.76	NaN
weather																													1.00	NaN
wgs84																														NaN

Figure 28. Lexical+hungarian distance analysis

We have presented the results through filtering the TF-IDF over .25 (green) and .15 (yellow) and lexical+hungarian over .85 (green) and .75 (yellow). We observe that ifc2x3 and ifc2 (that we generated) behave very closely, but that ifc2 is always closer to other ontologies than ifc2x3, so we will only consider the latter.

The two measures find different patterns of similarity between ontologies. The lexical measure finds that all ontologies are well matched to IFC, it also finds "um" close to "time"; the TF*IDF measure finds some clusters such as dolce-DUL-ssn (this is quite relevant since dolce is a source of dul). It also finds that these are quite close to bfo, process is closed to actor, and dog to energyresource.

This is not very conclusive. They allow for clustering ontologies hierarchically.

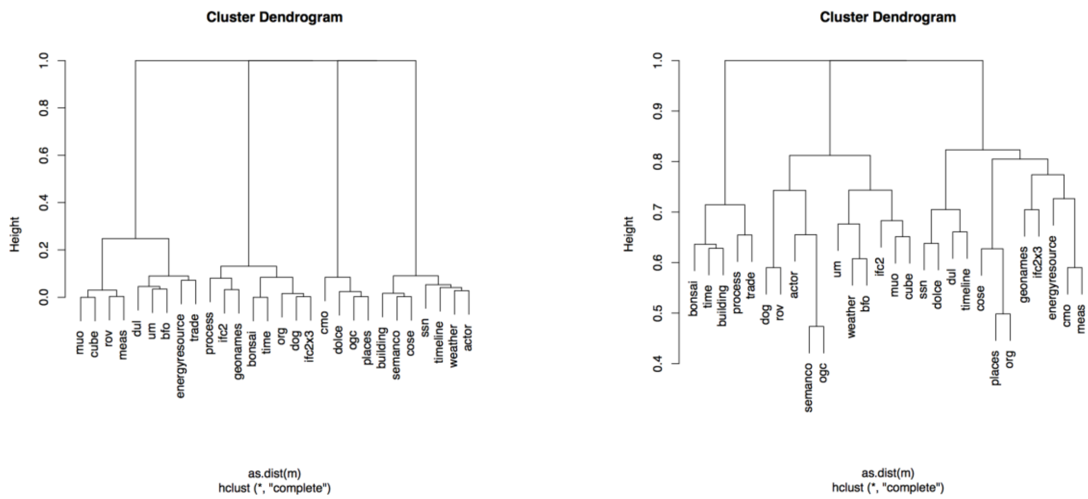

Figure 29. Cluster Dendrograms for TF-IDF (left) and lexical-hungarian (right)

2.4.3.4 Correspondence analysis

Correspondence analysis runs ontology matchers against the ontologies. We have exploited very simple measures comparing of the Alignment API (http://alignapi.gforge.inria.fr):

- *exact* finds only concepts which have the exact same name;
- *edna* computes edit distance between concept names;
- *smoa* computes a distance which takes into accounts habits of computer scientists for defining terms (use of _ or Camel convention).

The advantage of such methods is that they can be easily run. Similarly to the case with the similarity computation, it has not be possible to obtain alignments for prov and wgs84_pos.

The first simple method tells which the common names are; we have used it for counting the number of exact common names.

	actor	bfo	bonsai	building	cmo	cose	cube	dog	dolce	dul	energyresource	geonames	ifc2	ifc2x3	measurement	muo	ogc	org	places	process	prov	rov	semanco	ssn	time	timeline	trade	um	weather	wgs84
actor			1								1		1					1		2			5	1					2	
bfo									2	6			4	1				2		1				1						
bonsai				3		4		9		7	7	1	4	2									1	4			1	9	3	
building						2		1		1	1		3																	
cmo																								1						
cose								14	1	6	10		9	4									1					6		
cube																								1				1		
dog									1		398		6						1				6	1		1	1			
dolce										7	1	1	1						3	1				1		1	1			
dul													9	1				3	2	2			2	1		1	1	1		
energyresource													6	3					1	1			4			1	1	1		
geonames													1	1																
ifc2														###	3	2		5	2	6			3	6	1	1	1	54	1	
ifc2x3															3		3			4			1	2	1	1		53	1	
measurement																												2		
muo																												2		
ogc																														
org																				1			1						1	
places																							2			1	4			
process																							1	2						
prov																														
rov																														
semanco																											4	6	4	
ssn																												1		
time																										4				
timeline																														
trade																												1		
um																													2	
weather																														
wgs84																														

Figure 30. Number of exact common names between entities of two ontologies

We draw the graph of relations with strongly connected ontologies (at least 6 common names). This in fact favours larger ontologies. Such ontologies are all connected together which a core of 7 ontologies (bonsai, cose, dog, dul, energyresource, ifc2, um). The very strong connection between dog and energy resources and um and ufc2 suggest that some of these ontologies are likely from the same origin.

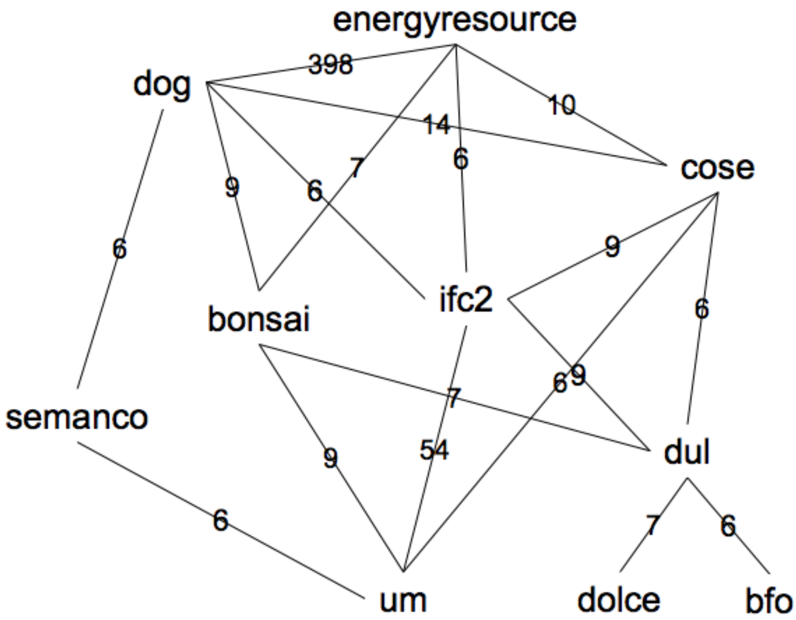

Figure 31. Graph of ontologies sharing at least 6 labels

It is likely that some ontologies represent by instances what other express as classes. This may not have been caught by these methods.

These first methods suggested place where there should be interesting matches:

We ran more elaborated methods and we selected thresholds so that they add around 2000 more matches to the exact matches. This is done according to the following table:

Table 4. Number of maches by methods according to thresholds

Threshold	SMOA	Edit distance
0.	132933	171750
.5	54649	20597
.6	33323	10597
.7	19400	**7820**
.8	11976	6978
.9	**8159**	6161
Exact match	6036	6036

	actor	bfo	bonsai	building	cmo	cose	cube	dog	dolce	dul	energyresource	geonames	ifc2	ifc2x3	measurement	muo	ogc	org	places	process	prov	rov	semanco	ssn	time	timeline	trade	um	weather	wgs84
actor			1					1		1	2		4	1				1		2			9	2	1			2	2	
bfo								1	4	10	1		6	3	1	1		2		1				1				3		
bonsai				4		5		23		8	22	1	19	10				2		6			14	5	2	3	2	12	3	
building						3		3		3	5		19	5									5			1		1		
cmo								1			1													1						
cose								26	1	7	19		29	11		1			2	1			16	3			1	23		
cube										2		1	2	2		2		1						4			1	2		
dog									2	2	486	1	30	8					3	9			22	7	1	1	2	12	7	
dolce										18	1		6	2	1	2		2	3	1			1	1	1	1	2	1		
dul													29	13	1	3		6	2	6			9	2	1	3	2	7	1	
energyresource													29	11					3	11			48	6		2	4	15	13	
geonames													14	13	1			1					2					4	1	
ifc2														###	14	13		12	5	9			48	14	4	4	6	199	11	
ifc2x3															5	5		8	1	8			24	10	4	4	4	147	9	
measurement																3		1						1	1			6		
muo																							1					4		
ogc																														
org																				1			2				3	1	1	
places																							7			1	5	1		
process																							9	4		1		7	7	
prov																														
rov																														
semanco																								10	6	4	7	54	31	
ssn																										1		1	3	
time																										5		2	1	
timeline																												2	1	
trade																												2		
um																													13	
weather																														
wgs84																														

Figure 32. SMOA with threshold .9

	actor	bfo	bonsai	building	cmo	cose	cube	dog	dolce	dul	energyresource	geonames	ifc2	ifc2x3	measurement	muo	ogc	org	places	process	prov	rov	semanco	ssn	time	timeline	trade	um	weather	wgs84
actor			3					2		1	4		2	1				1		3			8	1				1	4	
bfo									5	9			4	2	1			2	1	1				1				3		
bonsai				3		6	1	15	2	10	12	1	13	8				3					8	7	1	1	2	12	5	
building						2		1		2	1		7	3						4			4							
cmo								1			1													1						
cose							1	23	1	8	16		21	10	1	2			2	1			12	2				12	1	
cube										1			8	4		1								5				2		
dog									2	6	539		22	9				4	2	6			19	3	1	2	3	6	3	
dolce										20	3	1	2	1	1	1		2	4	1			2	1		1	1	2		
dul											4		29	12	1	1		6	4	5			8	3	1	1	2	5		
energyresource													24	10					2	9			25	1		2	5	8	5	
geonames													7	7				1					3					1		
ifc2														###	5	8		12	4	9			38	12	3	3	1	112	13	
ifc2x3															4	4		10		8			25	10	3	3	1	104	13	
measurement																1												3		
muo																							1	1				2		
ogc																														
org																				1			2					1	2	
places																							5			1	5	1		
process																							6	4	1	1		4	5	
prov																														
rov																														
semanco																								3	1	3	7	37	19	
ssn																												5	3	
time																										9		1	1	
timeline																														
trade																												3		
um																													7	
weather																														
wgs84																														

Figure 33. EDIT DISTANCE with threshold .7

We considered SMOA with the .9 thresholds and we drew the graphs of connection between ontologies. With a threshold of 10 correspondences, we obtain an extension of the previous exact match graph with added new ontologies (weather, ssn, building, process, org, muo, measurements and geonames) and a reinforcement of the connections across the graph (especially for semanco and bonsai). The new ontologies are, with the exception of weather (4 connections) and ssn (2 connections) poorly connected. The connection between bonsai and dul cannot remain because it has only 8 correspondences.

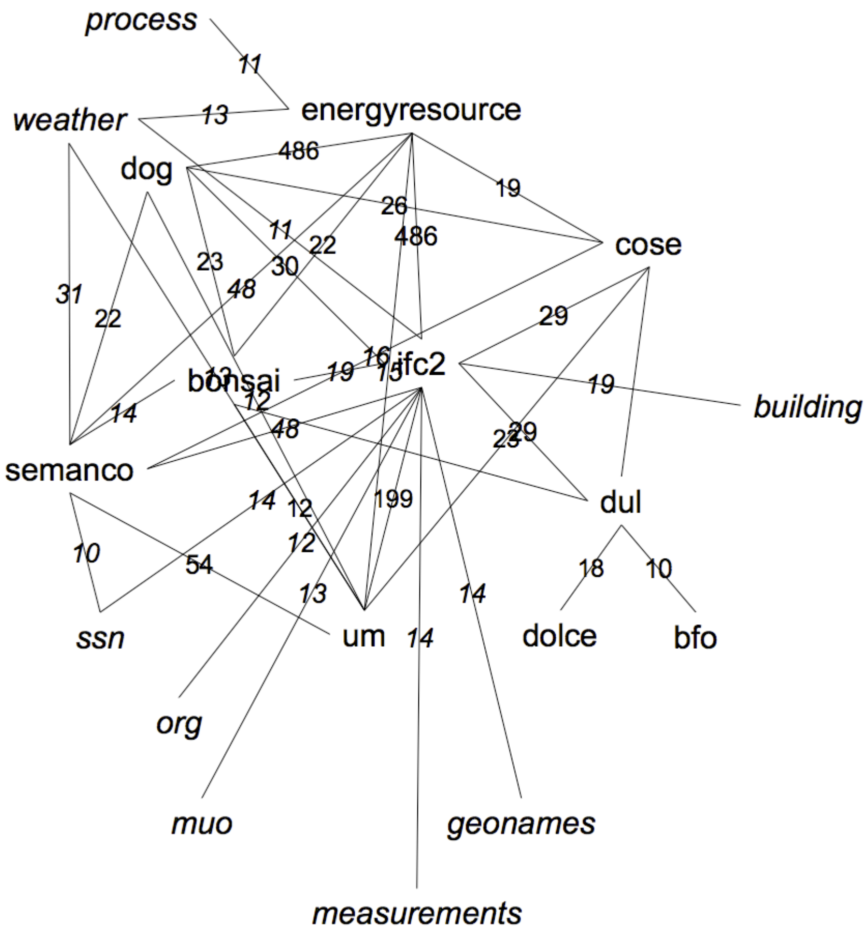

Figure 34. SMOA threshold .9 connections with at least 10 correspondences; new ontologies and connections are in italics, old discarded ontologies and connections are dotted

With a threshold of 15 correspondences, we roughly come back to the initial graph. Only weather and building have strong enough connections to remain in the graph. The graph is still more strongly connected than the initial graph. Only two connections are lost with respect to the initial graph: bfo had only one connection and disappears; um is not sufficiently connected to bonsai.

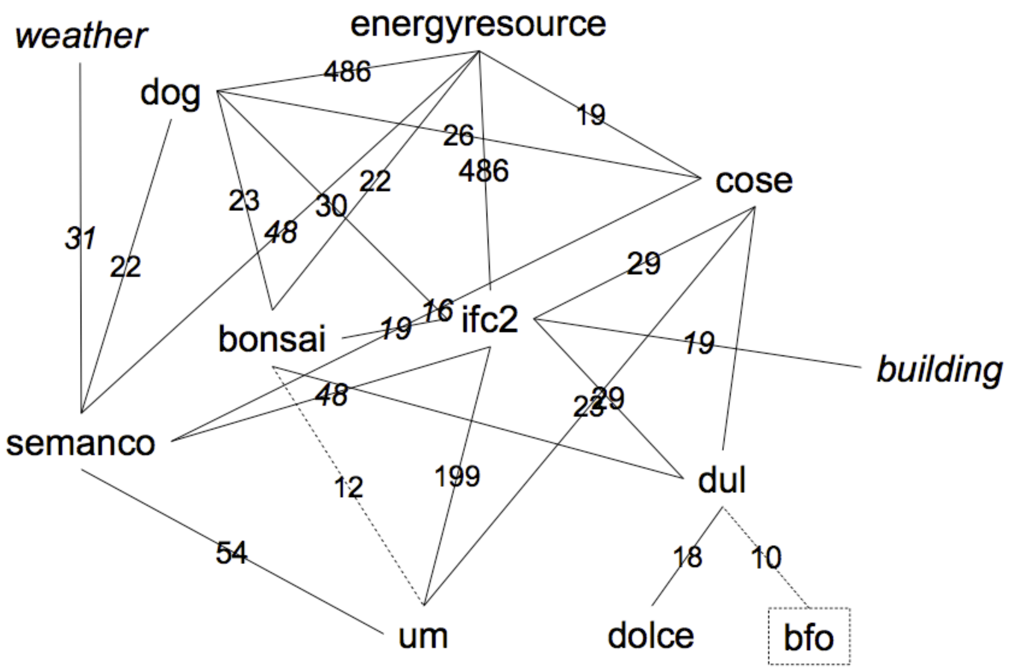

Figure 35. SMOA threshold .9 connections with at least 15 correspondences; new ontologies and connections are in italics, old discarded ontologies and connections are dotted

If we look a bit more qualitatively in one alignment, here ifc2-bonsai, we obtain the following correspondences:

IFC	**Bonsai**	**SMOA .9**	**EDNA .7**	**Observation**
parameter	Parameter	1.0	1.0	
Building	Building	1.0	1.0	
frequency	frequency	1.0	1.0	
Point	Point	1.0	1.0	
values	Value	.97	.83	?
mode	Model	.97	.8	#
ActuatorType	Actuator	.94		hasType
inputPhase	HasInput	.94		?
Condition	AirCondition	.93	.75	>
ParameterList0	Parameter	.93		?
ParameterValue	Parameter	.93		hasValue
ServiceLife	Service	.93		?
SensorType	Sensor	.92		hasType
BuildingStorey	Building	.92		< ?

PressureMeasure	Pressure	.91		
pointParameter	parameter	.91		< ?
BuildingElement	Building	.91		isPartOf
rateDateTime	dateTime	.9		< ?
ActuatorTypeEnum	Actuator	.9		

Clearly the four first correspondences seem to be correct, then half of the supplementary correspondences. EDNA thresholded at .7 finds fewer correspondences (13) which are, in general, less meaningful.

2.4.3.5 Alignment server

The results of some of these computation are provided in the public alignment server http://al4sc.inrialpes.fr. It can be used for creating new alignments or network of ontologies or for browsing the existing ones.

We provide three such networks here:

- Refine: a network containing relations between the 10 core ontologies curated by hands..
- ExactMatch: a network containing correspondences between entities which have exact matching terms between their elements
- Extended: a further elaborated network in which more matches are provided.

Part II: Ontologies and Datasets for Energy Management Systems

3 Differentiation between Energy Management System and Energy Measurement and Validation

The domains covered in work packages two and three come from two main application areas which have common aspects that not only allow to follow the same methodology within both work packages but also to share a lot of resources in terms of ontologies, datasets and alignments. There is no clear borderline as one may expect, which finally led to the decision to have a single point of information for the catalogues. Nevertheless, there are important differences between the two application areas that are described below. However, using linked data we expect that both application areas will more and more converge in the future, which will lead to more robust and flexible solutions for both application areas. In order to pinpoint the common areas, the two tables below should provide the scope of work in both packages. The first table tries to characterize and compare both application areas, whereas the subsequent tables show typical domains covered by the work packages. To avoid overlap, common ontologies and datasets are introduced either in D2.2 or D3.2.

WP2 is reviewing the linked data situation for Energy Management Systems (EMS). In general, EMS has a very broad scope and includes a lot of domains and stakeholders that depend on each other and must interact in order to be able to control and monitor energy production and consumption of electro-mechanical facilities. For several reasons it was decided in WP2 to first focus on the construction sector, which not only is a major energy consumer with high potentials for energy savings and peak energy balancing but is also an energy producer and even a way for energy storage. There are a lot of use cases for smart cities that directly or indirectly relate to buildings, e.g. prediction of energy demands (based on the heating, cooling and lighting demands of buildings that is also linked to user behaviours) or traffic management (for e.g. travelling between office and residential areas). Also, the construction industry is an interesting environment for testing and promoting the linked data approach as there are many different stakeholders that must collaborate and share information.

WP3 addresses the need to validate the results of energy-efficiency actions by analysing their measured impact. Measuring consumption in smart cities provides the source of data to be validated (including measurement methods, predictive models and algorithms), but other factors also play a role in the analysis, such as weather and climate data, building characteristics, user behaviour, etc.

Measurement and validation requires complete terminology for experimentation and piloting including experimental group, control group, statistical significance, outcome metrics (key performance indicators, KPIs), modelling parameters (e.g. occupancy, comfort levels, meteorology, etc.).

The ontologies and datasets described in the next sections therefore have been selected because they address one or more of the topics WPs two and three are focussed on.

Concerning alignments, their generation in a nearly blind way already allows for clustering ontologies and identifying clusters of ontologies related to these topics.

Table 5: Relevance criteria for comparing ontologies and datasets used in WP2 and WP3 scenarios

	Energy Management System (WP2)	**Energy Measurement and Validation (WP3)**
Main application area	Controlling a "single" electro-mechanical system either for energy production or energy consumption, automation of systems (machine-to-machine communication)	Measure and validate energy consumption and/or production to provide key figures for strategic and operative decisions, decision support and awareness services
Characteristics of used data		
degree of standardization	Medium	Low
degree of structured data	Very high	Medium
degree of complexity	High	Medium
degree of openness	Very low (outside of the "system" environment) Medium (within the "system", if different players must work together)	Medium to High
fault tolerance	Low to very low	Medium
security requirements	Very high	Low to medium
amount of data	Medium to high	Very high
real-time requirements	Medium to very high	Low to medium

Table 6: Assignment of identified ontologies and domains to WP2 and WP3

		Metrics and indicators (e.g. temporal, organisational, statistical, spatial)	Methods of measurement (incl. Scales, units, classifications)	Predictive models / Energy analysis	User behaviour	Building design and refurbishment	Monitoring	Controlling	Optimizing performance	Building operation	GIS	Systems: BACS, BEMS	Groups (experimental, control)	Statistics	Outcome metrics (KPIs)	Modelling parameters (e.g. occupancy, comfort levels, meteorology, climate)	Piloting	Organisation	Energy data	Weather and climate data	Environmental data (e.g. pollution)	Upper Ontologies	Measurement	Time	Devices/Sensors	Provenance
		2&3	WP3	WP3	2&3	WP2	WP2	WP2	2&3	WP2	WP3	WP2	WP3	2&3	WP3	2&3	WP3	WP3	2&3	WP3	WP3	2&3	WP3	2&3	2&3	2&3
Architecture and Building Physics Information	WP2					2																				
The W3C Organization Ontology	WP3																3	3								
IFC2X3 - University of Ghent	WP2					2																				
IFC2X3 - NIST OntoSTEP Converter	WP2					2																				
The W3C Time Ontology	WP3	3	3																					3		
BFO (Basic Formal Ontology)	WP3																					UPM?				
Weather and Exterior Influence Information	WP3							3								3				3	3			3		
Units of Measure (OM)	WP3	3																					3	3		
Measurement Ontology	WP3		3																				3			
Users and Preference Information	WP2				2																					
Energy and Resource Information	WP2					2													2							
MUO - Measurement Units Ontology	WP3																						3			
Casas Ontology for Smart Environments (COSE)	WP2	2																								
DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering)	WP2																					2				
DUL (DOLCE+DnS Ultralite)	WP3																					3				
Timeline Ontology	WP3																							3		

		Metrics and indicators (e.g. temporal, organisational, statistical, spatial)	Methods of measurement (incl. Scales, units, classifications)	Predictive models / Energy analysis	User behaviour	Building design and refurbishment	Monitoring	Controlling	Optimizing performance	Building operation	GIS	Systems: BACS, BEMS	Groups (experimental, control)	Statistics	Outcome metrics (KPIs)	Modelling parameters (e.g. occupancy, comfort levels, meteorology, climate)	Piloting	Organisation	Energy data	Weather and climate data	Environmental data (e.g. pollution)	Upper Ontologies	Measurement	Time	Devices/Sensors	Provenance
		2&3	WP3	WP3	2&3	WP2	WP2	WP2	2&3	WP2	WP3	WP2	WP3	2&3	WP3	2&3	WP3	WP3	2&3	WP3	WP3	2&3	WP3	2&3	2&3	2&3
SESAME-S Smart Building Ontology	WP2																									
Simulation Information Model (SIM) Ontology	WP2																									
Performance Information Model (PIM) Ontology	WP2																									
The W3C SemanticThe W3C Sensor Network Ontology	WP2						x	x																		
Building Information Model (BIM) Ontology	WP2					2																				
Global City Indicator Foundation Ontology	WP3													3									3			
User Behavior and Building Process Information	WP2				2																					
Cadastre and Land Administration Thesaurus (CaLAThe)	WP2					2																				
CASCADE airport ontology	WP2					2	2																			
Nikola Tesla Airport (NTA) Ontology	WP2					2	2																			
trade	WP3								3																	
Geonames	WP2					2																				
Data Cube	WP3													3												
The W3C PROV Ontology	WP3																									UPM
DogOnt	WP3																								3	
SUMO (Suggested Upper Merged Ontology)	WP3																					UPM				
BOnSAI	WP3																								3	

4 Collected ontologies relating to Energy Management Systems

4.1 Gap analysis

The current version of the Ready4SmartCities Ontology Catalogue contained **42 ontologies**; out of them **25 ontologies are particularly related to WP2** (that is, 60% of the ontologies in the catalogue).

In this section, we are going to provide the analysis of such 25 ontologies regarding the ontology metadata gathered in the catalogue, namely ontology language, ontology syntax, natural language, license, and availability.

Regarding the ontology language, 23 ontologies are implemented in OWL, one of the two most common languages for developing ontologies, one ontology (WGS84 Geo Positioning) is implemented in RDF-S, and one ontology (Cadastre and Land Administration Thesaurus (CaLAThe)) is written in SKOS. Only one ontology (DOLCE) is implemented using more than one ontology language; in this case OWL and Modal Logic. In order to benefit the interoperability and the usability of ontologies in different contexts, it could be beneficial to have the whole set of ontologies both in OWL and in RDF-S. In addition, for the ontology implemented in SKOS, it could be appropriate to translate the ontology into such languages.

With respect to the syntaxes or formats for WP2 ontologies, there are 23 ontologies in RDF/XML, three ontologies in Turtle, and one ontology (Cadastre and Land Administration Thesaurus (CaLAThe)) whose format is not known because the ontology code is not available. Only two ontologies in the catalogue are provided with more than one syntaxes. Thus, it could be also useful to have more ontologies with different formats.

The most common natural language in the catalogue is English, which is also the most common natural language in research tasks. Indeed, 24 ontologies out of the whole set of WP2 ontologies in the catalogue are written in English. In addition, there is one ontology (Cadastre and Land Administration Thesaurus (CaLAThe)) whose natural language is unknown because the ontology code is not available. Currently, there is one ontology (Geonames) written in more than one natural language (these are English, Swedish, Bulgarian, Norwegian, and Russian). Since multilingualism is a key issue, the catalogue should include more ontologies written in different natural languages.

A good point in the catalogue is that only open licenses are attached to those ontologies with license information. Talking about ontologies specially related to WP2, only five ontologies provide information about license issues; the rest of the ontologies (that is, 20 ontologies) has no license information.

With respect to the online availability of the WP2 ontologies in the catalogue, 64% of these ontologies can be retrieved in RDF. However, 52% of the ontologies do not have content negotiation mechanisms properly set up for this format and 36% cannot be retrieved in RDF. This situation should be corrected. Regarding HTML availability, 32% of the ontologies can be retrieved in such a format. However, 68% of the ontologies cannot be retrieved in HTML, which normally provides ontology documentation. Thus, in order to benefit the understanding and reuse of the ontologies, this situation should be also improved.

In addition, it is worth mentioning that in some cases the negotiation mechanisms seem to be good established, however the retrieved content does not correspond with the expected ones. This occurs when the ontology URI follows the pattern "www.owl-ontologies.com/xxx" or contains only names (e.g., "CityEnergyInvestmentStudy"). This situation should also be corrected.

As a summary, it is crucial to resolve the following issues

- to provide useful information about those metadata whose current value is 'Unknown'. This is the case of
  - ontology syntax and natural language for Cadastre and Land Administration Thesaurus (CaLAThe)[22]
  - ontology license for most of the ontologies (20 out of 25)
- to properly set up content negotiation mechanisms for ontology code and ontology documentation
- to obtain correct information for those ontologies that cannot be retrieved in RDF and/or in HTML. This is the case of
  - 3 ontologies cannot be retrieved in RDF and 6 probably are not available or are published in a wrong way
  - 7 ontologies cannot be retrieved in HTML and 10 probably are not available or are published in a wrong way

4.2 List of ontologies

IFC Ontology

Name	IFC2X3 - University of Ghent
Author and License	Davy Van Deursen, Pieter Pauwels (mapping configuration from IFC2x3 Express specification from buildingSMART), unknown license
URL	http://multimedialab.elis.ugent.be/organon/ontologies/IFC2X3#
Description	OWL representation of the buildingSMART data model. The IFC data model is written in an EXPRESS schema (IFC2x3). This ontology is the result of an automated transformation of this EXPRESS schema into an OWL ontology. In this transformation process, every EXPRESS element that has a direct equivalent in OWL is mapped onto this equivalent. More specifically, for each ENTITY element in EXPRESS a corresponding OWL class is generated, EXPRESS attributes are converted into the appropriate OWL properties, etc.
Scope (Domain)	Buildings, AEC industry, BIM
Use cases (Motivation, Relevance)	The IFC data model supports data sharing of BIM data. It supports coordination of design activities and hand-over of design and maintenance data. There are many use cases for smart cities where building data is of relevance, either to be referenced (in particular for EMV use cases) or actively used by building simulation and maintenance (EMS use cases).
Data sets	IFC datasets can be generated by all major CAD tools. However, these tools export IFC data in the original SPF format only and thus has to be mapped to an RDF representation according to this ontology. Public IFC files are available from pilot and research projects mainly. However, building data is typically not published as it is mainly shared within the design team only or are handed over to contractual partners.
Open issues/	There is an agreed standard developed by the buildingSMART organisation. This is the

[22] This information could not be gathered because the ontology code was not available.

Challenges	baseline definition from which an ifcOWL representation can be derived and enriched. So far, there are several mapping approaches, all of them dealing with slightly different requirements and boundaries. All mapping approaches will lose some sort of information as OWL is not able to deal with all constraints specified in the original IFC EXPRESS definition. Also, none of available mapping approaches is enriching the original definition.
Tool support	

NIST OntoSTEP Converter plugin for Protégé

Name	IFC2x3 NIST OntoSTEP Converter
Author and License	Rachuri Sudarsan, Raphael Barbau, Sylvere Krima; (developer of this tool) (mapping configuration from IFC2x3 Express specification from buildingSMART -> OWL-DL representation), unknown license
URL	http://www.nist.gov/OntoSTEP/ifc2x3 (download of the tool)
Description	See IFC Ontology
Scope (Domain)	Buildings, AEC industry, BIM
Use cases (Motivation, Relevance)	See IFC Ontology
Data sets	See IFC Ontology
Open issues/ Challenges	See IFC Ontology
Tool support	Plugin for Protégé that enables to convert EXPRESS schemata and SPF datasets.

Architecture and Building Physics Information

Name	Architecture and Building Physics Information
Author and License	Institute of Computer Aided Automation, Vienna University, Austria unknown license
URL	https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/BuildingOntologySharedVocabulary.owl
Description	An ontology representing building information (e.g. structure, material, architecture) for Smart Home Systems. Classes, axioms and customized datatypes have been retrieved from gbXML (www.gbxml.org). (for further information see: https://www.auto.tuwien.ac.at/projectsites/thinkhome/building-information.html). The mapping from gbXML is done through an XSLT script, which is also available on the website.

Scope (Domain)	Buildings, Energy Analysis
Use cases (Motivation, Relevance)	There are many use cases for smart cities where building data is of relevance. gbXML data is typically used for energy analysis, which is done in the design phase of a building.
Data sets	gbXML datasets can be generated and imported by many CAD and energy analysis tools. However, these tools export a XML file according to the gbXML schema definition and thus has to be mapped to an RDF representation according to this ontology. Sample gbXML files are available at www.gbxml.org. However, building data is typically not published as it is mainly shared within the design team only or are handed over to contractual partners.
Open issues/ Challenges	There is an agreed schema developed by the gbXML initiative. This is the baseline definition from which this ontology was derived based on an XSLT script.
Tool support	Population of the ontology through mapping approaches from traditional CAD tools.

SESAME-S Smart Building Ontology

Name	SESAME-S Smart Building Ontology
Author and License	Research Centre for Telecommunication (FTW, http://www.ftw.at/), Austria unknown license
URL	http://datahub.io/dataset/smartbuilding-sesames https://commondatastorage.googleapis.com/ckannet-storage/2012-08-20T165445/SmartBuildingv3.owl
Description	SESAME-S = Semantic Smart Metering Services for Energy Efficient Houses This ontology is a typical example of a purpose-built ontology. It was developed within the SESAME project, which is already finished. The ontology is not maintained anymore and no further documentation is available. It contains about 20 class and 30 property definitions, thus being a rather small ontology in terms of size and scope. It is focused on the data that has been managed in the two real-world examples, e.g. measurements of temperature, humidity, light and presence of persons.
Scope (Domain)	Smart Sensors, Devices
Use cases (Motivation, Relevance)	The ontology was developed to show the “next generation of energy efficient buildings”. It is part of a prototype development to proof the concept of semantic smart metering and providing services for energy efficiency. One of the goals was to raise awareness for taking care of reducing energy consumption within a building by providing measured data to people who are using the facilities. More information about the project is available on their website (http://sesame-s.ftw.at/) and in a number of research publiciations, e.g. in: Girtelschmid, S., Steinbauer, M., Kumar, V., Fensel, A., Kotsis, G. "On the Application of Big Data in Future Large Scale Intelligent Smart City Installations", International Journal of Pervasive Computing and Communications, Emerald Group Publishing, Vol. 10 Iss: 2 (2014).

Data sets	Data from two real-world examples have been managed with this ontology. One example from Austria, the Kirchdorf school example, and another one from Russia, the Chernogolovka factory example. The datasets are not public available as there are strong concern regarding security and privacy issues (actual energy consumption, usage patterns of the building).
Open issues/ Challenges	There is no maintenance and no further documentation of this ontology. It is used by the authors as a baseline for follow-up projects.
Tool support	Prototypes/tools developed in SESAME-S project.

Simulation Information Model (SIM) Ontology

Name	Simulation Information Model (SIM) Ontology
Author and License	unknown license
URL	http://www.modelservers.org/public/ontologies
Description	Developed and used in the IntUBE project (Intelligent Use of Building's Energy Information), which was carried out from 2007 to 2010.
Scope (Domain)	Building usage, Building performance
Use cases (Motivation, Relevance)	Simulated data generated by energy simulation tools (including their input parameters)
Data sets	Examples from the IntUBE project available (see URL).
Open issues/ Challenges	The status of the ontology is unclear. Website (and domain) is not available
Tool support	Data managed in the "Energy-information integration platform"

Building Information Model (BIM) Ontology

Name	Building Information Model (BIM) Ontology
Author and License	unknown license
URL	http://www.modelservers.org/public/ontologies
Description	IntUBE project (Intelligent Use of Building's Energy Information) – finished 2011 (project website no more available)
Scope (Domain)	Building
Use cases	Static data about the building in general, such as building location, process stage, spaces,

(Motivation, Relevance)	envelopes and building services
Data sets	
Open issues/ Challenges	see Simulation Information Model (SIM) Ontology
Tool support	

Performance Information Model (PIM) Ontology

Name	Performance Information Model (PIM) Ontology
Author and License	unknown license
URL	http://www.modelservers.org/public/ontologies
Description	IntUBE project (Intelligent Use of Building's Energy Information) – finished 2011 (project website no more available)
Scope (Domain)	Building usage, Building performance
Use cases (Motivation, Relevance)	dynamic data obtained from monitoring systems, including climate, building use and energy performance
Data sets	
Open issues/ Challenges	see Simulation Information Model (SIM) Ontology
Tool support	

Energy and Resource Information

Name	Energy and Resource Information
Author and License	TU Vienna
URL	https://www.auto.tuwien.ac.at/projectsites/thinkhome/facilities-and-energy-information.html https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/energy/changelog/EnergyResource_Revision_1.03.txt
Description	An ontology representing energy information for Smart Home Systems.
Scope (Domain)	Home Automation
Use cases	

(Motivation, Relevance)	
Data sets	https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/EnergyResourceOntology.owl https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/EnergyResourceOntologyExample.owl
Open issues/ Challenges	
Tool support	

Cadastre and Land Administration Thesaurus

Name	Cadastre and Land Administration Thesaurus (CaLAThe)
Author and License	Professor Erik Stubkjær, Department of Planning, Aalborg University, Denmark, Dr. Volkan Cagdas, Department of Surveying Engineering, Yildiz Technical University, Turkey. licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License
URL	http://www.cadastralvocabulary.org (available on request)
Description	This ontology provides a controlled vocabulary, which is derived mainly from the ISO/DIS 19152 Land Administration Domain Model and related to existing thesauri, primarily the GEMET thesaurus, the AGROVOC thesaurus, and the STW Thesaurus for Economics.
Scope (Domain)	Buildings, cadastre, geography
Use cases (Motivation, Relevance)	In smart cities application, it could be useful where certain buildings data are needed; e.g. geographical positioning, internal divisions (apartments), spatial representation.
Data sets	
Open issues/ Challenges	CaLAThe is encoded as a Simple Knowledge Organization System (SKOS), according to specifications developed by the World Wide Web Consortium (W3C).
Tool support	

CASCADE airport ontology

Name	CASCADE airport ontology
Author and License	Institute Mihajlo Pupin, Sanja Vranes, Nikola Tomasevic, Marko Batic, CASCADE ICT for Energy Efficient Airports Unknown license
URL	https://webgate.ec.europa.eu/fpfis/wikis/display/eeSemantics/CASCADE+Modelling+Ontology https://webgate.ec.europa.eu/fpfis/wikis/download/attachments/44483343/CASCADE%20Core%20Airport%20Ontology%20%28class%29.owl?version=1&modificationDate=1399554858401&api=v2

Description	The CASCADE Core airport ontology provides a generic model of the airport facility as a set of concepts and corresponding relationships among them. The purpose of the Core airport ontology is to provide the modelling guidelines and to describe the technical characteristics/relations of related systems installed at the site, their topological profile, as well as to facilitate the interpretation of signals.
Scope (Domain)	Airports, automated buildings
Use cases (Motivation, Relevance)	Even if this ontology is oriented to create a model of airport facility, it can be used also in generic buildings modelling, particularly public buildings or complexes, due to the commonality with airport sub-functions.
Data sets	
Open issues/ Challenges	The CASCADE deontology is characterized by a partial superposition with other ontologies taken into account (regarding geography or buildings). It would be expectable to reach a
Tool support	

GeoNames Ontology

Name	GeoNames Ontology
Author and License	Bernard Vatant, GeoNames. Creative Commons CC BY 3.0
URL	http://www.geonames.org/ontology/ontology_v3.1.rdf
Description	The GeoNames Ontology makes it possible to add geospatial semantic information to the World Wide Web. All over 8.3 million geonames toponyms now have a unique URL with a corresponding RDF web service. Other services describe the relation between toponyms.
Scope (Domain)	Geography
Use cases (Motivation, Relevance)	Relevant to guarantee unique reference to toponyms and easy information access through the GeoNames database (http://sws.geonames.org), especially geographic position.
Data sets	At http://www.geonames.org/advanced-search.html all of the rdf produced by GeoNames are available.
Open issues/ Challenges	
Tool support	

DOLCE

Name	DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering)

Author and License	Claudio Masolo License unknown.
URL	http://www.loa-cnr.it/ontologies/DOLCE-Lite.owl#
Description	DOLCE is the first module of the WonderWeb Foundational Ontologies Library (WFOL). DOLCE has a clear cognitive bias, in the sense that it aims at capturing the ontological categories underlying natural language and human common-sense. its authors believe that such bias is very important for the Semantic Web. DOLCE is an ontology of particulars, in the sense that its domain of discourse is restricted to them. A basic choice we make in DOLCE is the so-called multiplicative approach: different entities can be co-located in the same space-time (e.g. the vase and the amount of clay).
Scope (Domain)	Top level ontology
Use cases (Motivation, Relevance)	Upper level ontologies could be used for data integration across datasets
Data sets	Upper level ontologies could be used in a high number of datasets as they represent top concepts
Open issues/ Challenges	
Tool support	

Users and Preference Information

Name	Users and Preference Information
Author and License	Institute of Computer Aided Automation, Vienna University, Austria unknown license
URL	https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/ActorOntology.owl
Description	An ontology describing user information and preferences for Smart Home Systems. User profiling knowledge includes information about human characteristics (e.g age and gender) and preferences (e.g. visual and thermal habits) allowing the formulation of different habit patterns. This ontology came as a result of ThinkHome project, which utilizes artificial intelligence to improve control of home automation functions provided by dedicated automation systems. (for further information see https://www.auto.tuwien.ac.at/projectsites/thinkhome/user-information.html)
Scope (Domain)	User Preferences, User Profiling, User Scheduling, Energy Management
Use cases (Motivation,	There are many use cases for smart cities where smart home occupancy data is of relevance.

Relevance)	In particular, these data offer valuable information about : • thermal and visual preferences • configured schedules for energy profiling Advanced control automations related to this data can significantly improve energy-efficiency and energy-saving, yet preserving used comfort and preferences.
Data sets	As reported in ThinkHome project, all data collected will be publicly available through a dedicated web-site. There is no other evidence that this ontology has already been used by other projects/applications, in order to seek for more available data-sets.
Open issues/ Challenges	
Tool support	

User Behaviour and Building Process Information

Name	User Behaviour and Building Process Information
Author and License	Institute of Computer Aided Automation, Vienna University, Austria unknown license
URL	https://www.auto.tuwien.ac.at/downloads/thinkhome/ontology/ProcessOntology.owl
Description	This is an ontology representing application, processes, profiles and patterns for smart home operation. It contains elementary operations that could describe basic human actions in the building during specified time periods. This ontology came as a result of ThinkHome project, which utilizes artificial intelligence to improve control of home automation functions provided by dedicated automation systems. (For further information see: https://www.auto.tuwien.ac.at/projectsites/thinkhome/process-information.html)
Scope (Domain)	User Actions, Building Processes, Energy Management and Automated Control
Use cases (Motivation, Relevance)	There are many use cases for smart cities where processing data is of relevance. User actions and elementary operations provide valuable information about: • Interrelationship between occupants, devices and building spaces • Real time consumption actions • Optimization of control strategies based real time events and data.
Data sets	As reported in ThinkHome project, all data collected will be publicly available through a dedicated web-site. There is no other evidence that this ontology has already been used by other projects/applications, in order to seek for more available data-sets.
Open issues/ Challenges	

Tool support	

The W3C Semantic The W3C Sensor Network Ontology

Name	The W3C SemanticThe W3C Sensor Network Ontology
Author and License	W3C Semantic Sensor Network Incubator Group W3C Software Notice and License
URL	http://purl.oclc.org/NET/ssnx/ssn
Description	This ontology describes sensors and observations, and related concepts. It does not describe domain concepts, time, locations, etc. these are intended to be included from other ontologies via OWL imports. (For further information see : http://www.w3.org/2005/Incubator/ssn/wiki/Report_Work_on_the_SSN_ontology)
Scope (Domain)	Sensors, Sensors Measuring, Monitoring, Devices
Use cases (Motivation, Relevance)	Measuring and Monitoring support the basis of the intelligent operation. Valuable information about • Sensors as a device () • Measuring operations and measuring capability. • Device http://purl.oclc.org/NET/ssnx/ssn#Device Five working examples are already included in the reference wiki page, illustrating the application of different parts of this ontology, such as: University deployment, Smart product, Wind sensor, Agriculture Meteorology and Linked Sensor Data.
Data sets	The W3C Semantic Sensor Network Incubator Group maintains hosts a wiki reference page since 2005, providing the respective ontologies for public uses and allowing interaction with public via open data and communication methods via a W3C list (public-xg-ssn@w3.org) It is expected that Data-sets based on this ontology may already exist from other projects.
Open issues/ Challenges	
Tool support	

Casas Ontology for Smart Environments (COSE)

Name	Casas Ontology for Smart Environments (COSE)
Author and License	School of Electrical Engineering and Computer Science, Washington State University, Box 642752, Pullman, WA, 99164-275

URL	-
Description	The number of smart appliances and devices in the home and office has grown dramatically in recent years. Unfortunately, these devices rarely interact with each other or the environment. In order to move from environments filled with smart devices to smart environments, there must be a framework for devices to communicate with each other and with the environment. This enables reasoners and automated decision makers to understand the environment and the data collected from it. Semantic web technologies provide this framework in a well-documented and flexible package. In this paper we present the Casas Ontology for Smart Environments (COSE) and accompanying data from a test smart environment and discuss the current and future challenges associated with a Smart Environment on the Semantic Web.
Scope (Domain)	Smart Environments, Ambient Assisted Living
Use cases (Motivation, Relevance)	
Data sets	
Open issues/ Challenges	
Tool support	

Nikola Tesla Airport (NTA) Ontology

Name	Nikola Tesla Airport (NTA) Ontology
Author and License	Possibly: University of Belgrade, Institute Mihajlo Pupin
URL	
Description	The ontology facilitates the interpretation and semantic enrichment of SCADA signals using the underlying spatial and topological model of the airport infrastructure as well as vendor data regarding the equipment characteristics, protocols and standards used. http://www.e-drustvo.org/icist/2012/html/pdf/495.pdf
Scope (Domain)	airport managament, emergency management, facility management
Use cases (Motivation, Relevance)	"Nikola Tesla" airport Belgrade "For improving and providing more intelligent , holistic, airport facility management systems that rely on contemporary management platforms such as Supervisory Control and Data Acquisition (SCADA) systems, classification and description of various information/data within

	the airport infrastructure"[23]
Data sets	
Open issues/ Challenges	
Tool support	

[23] http://www.e-drustvo.org/icist/2012/html/pdf/495.pdf

5 Collected datasets relating to Energy Management Systems

5.1 Gap analysis

As shown in the comparison

Table 5 Energy Management System interoperability is typically characterized by high demands regarding security and privacy issues. Also, there are rather complex datastructures and a huge amount of data so that it seems that there is a natural barrier for publishing data on the web. In that respect, there are still a lot of open questions to be discussed and solved. Additionally, there is still a lack of clear business cases for data owners to open their data and to justify additional efforts to transfer and host the data in the web. All these circumstances might explain why there is only very few open linked data available. In general, found datasets are either results of research projects or somehow driven by public authorities. From industry a natural interest is driven by marketing use cases, i.e. provision of open data to advertise their products. Accordingly, they typically focus on unique selling features instead of providing neutral and comparable product descriptions. Our preliminary conclusion about availability of open datasets in the area of Energy Management Systems is quite disappointing. The following section summarizes the result of our research and, not claiming to give a complete picture of the current situation, it shows the challenges of providing a critical mass of data to be a sound basis to build new applications or point of information.

5.2 List of datasets

The European Building and Construction Materials Database for the Semantic Web

Name	The European Building and Construction Materials Database for the Semantic Web
Author and License	Andreas Radinger, Martin Hepp, Otto Handle unknown license (data mapped from the Eurobau database available at http://eurobau.com/)
URL	http://semantic.eurobau.com/sitemap.xml (for fetching all data) http://semantic.eurobau.com/eurobau-utility.owl (ontology) http://linkeddata.uriburner.com/sparql (public SPARQL endpoint) http://eurobau.com/ (source)
Description	Major dataset of the European building and construction materials market for the Semantic Web on the basis of the GoodRelations Web Vocabulary for E-Commerce. (see http://semantic.eurobau.com/)
Scope (Domain)	Construction Materials
Use cases (Motivation, Relevance)	Comparison of products? Search for products
Statistics	81 Manufacturers / Brands 19 Resellers 183 Warehouse locations 56.360 Product types (including variants) 1.783.798 Offerings 95 % of the product models include rich FreeClassOWL descriptions
Questions	

State Energy Data System (SEDS)

Name	State Energy Data System (SEDS)
Author and License	U.S. Energy Information Administration (EIA) unknown license The data collected by EIA surveys forms (http://www.eia.gov/survey/) are for the most part not proprietary and available. For users eager to dive deeper there are assembled tools to access searchable databases.
URL	Assembled tools are available to customize searches, view specific data sets, study detailed documentation, and access time-series data. • http://api.eia.gov/ Application Programming Interface (API) is a machine readable format which can serve all customers for free, though a registration key is needed for access. (For further information see: http://www.eia.gov/developer/) • http://www.eia.gov/beta/api/bulkfiles.cfm The bulk download facility provides the entire contents of each major API data set in a single ZIP file. • http://www.eia.gov/tools/models/datatools.cfm Additional set of data tools for exploiting data from different domains.
Description	The State Energy Data System (SEDS) is the source of the U.S. Energy Information Administration's (EIA) comprehensive state energy statistics. SEDS is aimed to create historical time series of energy production, consumption, prices, and expenditures by state for analysis and forecasting purposes. (For further information see: http://www.eia.gov/state/seds/)
Scope (Domain)	Consumption, Prices and Expenditures, Production
Use cases (Motivation, Relevance)	There are many use cases for smart cities where energy data system is of relevance: • Historical time series of energy production / consumption, prices and expenditures • Energy Analysis • Exploitation of data for prediction purposes
Statistics	408,000 electricity series organized into 29,000 categories 30,000 State Energy Data System series organized into 600 categories 115,052 petroleum series and associated categories 11,989 natural gas series and associated categories 132,331 coal series and associated categories (released Feb 25, 2014) 3,872 Short-Term Energy Outlook series and associated categories (released May 27, 2014) 368,466 Annual Energy Outlook series and associated categories (released May 27, 2014)
Questions	
Name	State Energy Data System (SEDS)

Repener building energy

Name	Repener building energy
Author and License	Álvaro Sicilia et.al. Creative Commons Attribution
URL	http://arcdev.housing.salle.url.edu/repener/sparql
Description	Integrated information of the Spanish territory, regarding energy certification, building monitoring, and geographical data
Scope (Domain)	energy efficiency, energy certification
Use cases (Motivation, Relevance)	
Data sets	
Open issues/ Challenges	
Tool support	

Part III: Conclusions and next steps

In general, the decision to **join efforts** between workpackage 2 and 3 and to work on common solutions has been proven to be the right decision. Several advantages can be mentioned: (1) more resources could be spent on working on the methodology and provision of general tool support. The developed online catalogue and the alignment tool very much fits to the proposed idea of providing data (in that case our findings about ontologies, datasets and alignments) in an open, interlinkable and thus easily reusable way. This is seen as an important outcome of WP2 and WP3. (2) throughout our regular telecons there have been lively discussions and information exchange that not only helped to get insights into other domains but also to avoid duplication of work. Although we finally decided to assign ontologies either to D2.2 or D3.2, many ontologies are relevant for both application areas. The decisions which ontologies (or parts of an ontology) are needed depend to a high degree on the use case scenario. But this was not part of this work package. Therefore, with the aim of having a neutral catalogue it is reasonable to offer a single point of information covering both application areas. (3) external experts are seen as a main source of information. To get access to that information we developed a survey that was sent out to those experts using different communication channels like e-mail, twitter etc. Rather than sending specific surveys from WP2 and WP3 a more generic survey was prepared and communicated only once to the community.

The **process** of gathering ontologies will continue to be applied in order to increase the set of ontologies included in the catalogue. As already mentioned in Section 2, there are various projects (a) in which the ontology development is in progress in the moment of writing or (b) in which their plan includes the building of ontologies in the near future. In addition, there are still in-situ correspondences where information about ontologies developed is expected. With regards to dataset collection, the period up to writing this deliverable has showed that datasets are far harder to find, mainly due to the facts that (a) most dataset are not the result of EU projects and therefore are not subject to specific requirements (e.g. openness, recommended formats), (b) the availability of linked open data related to energy in general is scarce, or (c) when available, the data is not linked and/or open. Stronger activities are needed in order to record more datasets in the next period, including more active research, committed stakeholder engagement, as well as putting stronger focus on datasets at workshops and other events.

Regarding the **catalogue**, as immediate future lines of work UPM plans to include search features and to provide a SPARQL endpoint so that users can query the RDF version of the catalogue. In addition, in order to provide a more detailed assessment (e.g., related to good modeling practices), the OWL ontologies available on the Web will be evaluated by means of external evaluation services such as OOPS! (OntOlogy Pitfall Scanner![24]), an on-line application used to identify pitfalls in ontologies.

The **alignment server** will be further improved in order to serve better the project:

- Connection to the ontology catalogue: we documented the interface of the Alignment server in order for the ontology catalogue to link to the alignments from the ontologies, this would ease the navigation between the two tools.
- Automatic update: in the other direction, we plan to have automatic recomputation of alignments when the ontology from the catalog changes.
- Alignments from linked data: we plan to develop alignment inference from the links available between resources of linked data.
- Online network of ontology edition: we would like to offer users to create their own ontology networks from available alignments. At the moment, this is only possible by creating offline an ontology network and loading it to the server.

---

[24] http://www.oeg-upm.net/oops

The **domains** covered by the catalogue are currently 5 identified for Level 1, (Temporal, Organisational, Statistical, Spatial/Geographical, and Measurement) and 4 out 7 domains identified for Level 2 (Energy, Weather, Building, and User Behaviour). Thus, effort should be put in trying to cover in the next version of the catalogue the following three domains: Climate Zone, Environmental, and Occupancy. In addition, there are 10 new domains covered by ontologies in the catalogue. These domains are Provenance, Top Level, Generic, Device, Sensor, IFC, Smart Environment, Home Automation, Process and Urban Planning.

A specific conclusion for Energy Management System interoperability is to put more efforts in reusing already existing data models or database schemas. In many areas of WP2 there already exist agreements that have much in common with ontologies and thus have a good chance to be (fully or semi-automatically) transferred into open linked data by some data publishing process. Thus, the guideline developed in WP4 and described in deliverable 4.1 could be used to help people to transfer selected datasets into open linked data. Such practical help could be of interest in two cases (1) a rich dataset from a specific stakeholder shall be transferred (in that case data is stored in a database or a proprietary dataformat) and (2) a generic mapping approach for a widely-used data format (preferrable an open standard like IFC) can be developed that can be applied to datasets of different stakeholders. In case data is already public available, but not as Open Linked Data, the stakeholder needs to be convinced to move to W3C standards. Different scenarios could be developed for this, in particular showing how to establish links to other datasets and thus to enrich own data. However, in many cases data is still kept in closed enironments due to a number of reasons, such as lack of knowledge and tools, security, privacy and licensing issues or lack of incentives to do so. In those cases, additional solutions might be necessary to deal with those concerns. For instance, it would be possible to specify a data subset that shall be published or to protect parts of the dataset by access control mechanisms. Accordingly, based

We should encourage ontology developers to provide their ontologies in more than one ontology language as well as localized in different natural languages. Regarding ontology syntaxes, it would be useful to have ontologies in more than one ontology format. Furthermore, ontology developers should be animated to improve those cases in which the content negotiation mechanisms have not been set up properly (both for code and for documentation).

In addition we should resolve those metadata that currently have as values 'Unknown'. For future ontology metadata, we should animate catalogue populators to provide complete information. In particular it is key to have information about the license for the ontologies.

References

[Alexander et al, 2011] K. Alexander; R. Cyganiak; M. Hausenblas; J. Zhao. *Describing Linked Datasets with the VoID Vocabulary*. 3 March 2011. W3C Note.

[Bizer, Heath, & Berners-Lee 2009] Bizer, C., T. Heath, & T. Berners-Lee (2009). *Linked Data – the story so far*. International Journal on Semantic Web and Information Systems 5, 1–22.

[Brickley, 2004] Brickley, D. (2004). *RDF vocabulary description language 1.0: RDF schema*. http://www. w3. org/tr/rdf-schema/.

[Garcia-Castro et al, 2014] García-Castro, R., Poveda-Villalón, M., Radulovic, F., Gómez-Pérez, A., Euzenat, J., Priego-Roche L.M., Vogt, G., Robinson, S., Birov, S., Fies, B. (2013). *Deliverable 3.1: Strategy for Energy Management System Interoperability*. READY4SmartCities project.

[Hartmann et al., 2005] Jens Hartmann, Raúl Palma, York Sure, Mari del Carmen Suárez-Figueroa, Peter Haase, Asunción Gómez-Pérez, Rudi Studer. *Ontology Metadata Vocabulary and Applications*. Workshop on Web Semantics (SWWS2005). Agia Napa, Cyprus. 1-2 November 2005.

[Lozano et al. 2012] Lozano, E., J. Gracia, O. Corcho, R. A. Noble, & A. Gómez-Pérez (2012, November). *Problem-based learning supported by semantic techniques. Interactive Learning Environments*.

[Poveda-Villalón et al., 2012] M. Poveda-Villalón, M.C. Suárez-Figueroa, A. Gómez-Pérez. *Validating ontologies with OOPS!*. 18th International Conference on Knowledge Engineering and Knowledge Management. Galway, Ireland. 8-12 October 2012.

[Poveda-Villalón et al., 2013] M. Poveda-Villalón, B. Vatant, M.C. Suárez-Figueroa, A. Gómez-Pérez. *Detecting Good Practices and Pitfalls when Publishing Vocabularies on the Web*. 4th Workshop on Ontology Patterns (WOP2013). Sydney, Australia. 21 October 2013.

Appendix: prefix list

prefix	URI	#referenc	#nsdecl	#imports	#use
xsd	http://www.w3.org/2001/XMLSchema#	28	19		9
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	29	26		30
rdfs	http://www.w3.org/2000/01/rdf-schema#"	29	26		30
owl	http://www.w3.org/2002/07/owl#	28	25		28
owl2xml	http://www.w3.org/2006/12/owl2-xml#	2	2		0
skos	http://www.w3.org/2004/02/skos/core#	7	6	1	6
daml	http://www.daml.org/2001/03/daml+oil#	2	0		0
dc	http://purl.org/dc/elements/1.1/	12	11	1	9
dcterms	http://purl.org/dc/terms/	8	7		4
cc	http://creativecommons.org/ns	3	3		3
time	http://www.w3.org/2006/time	6	6	5	2
timeent	http://www.w3.org/2006/time-entry	1	0	1	
foaf	http://xmlns.com/foaf/0.1/	9	7		0
protege	http://protege.stanford.edu/plugins/owl/protege#	3	3		1
xsp	http://www.owl-ontologies.com/2005/08/07/xsp.owl#	3	3		0
owl-s	http://www.daml.org/services/owl-s/1.2/Service.owl#	1	1	1	0
swrlb	http://www.w3.org/2003/11/swrlb#	4	4		0
codamos	http://lpis.csd.auth.gr/ontologies/CoDAMoS/CoDAMoS.owl#	1	1	1	0
swrlq	http://sqwrl.stanford.edu/ontologies/built-ins/3.4/sqwrl.owl#	2	2		0
swrla	http://swrl.stanford.edu/ontologies/3.3/swrla.owl#	2	2		0
opencyc	http://sw.opencyc.org/concept/	1	1		0
vann	http://purl.org/vocab/vann/	4	4		4
swrl	http://www.w3.org/2003/11/swrl#	4	4		0
timezone	http://www.w3.org/2006/timezone#	4	3		0
owlapi	http://www.semanticweb.org/owlapi#	1	1		0
scovo, scv	http://purl.org/NET/scovo#	2	1		1
gr	http://purl.org/goodrelations/v1#	1	1		1
prov	http://www.w3.org/ns/prov#	2	2		2
vcard	http://www.w3.org/2006/vcard/ns#	1	1		0
status	http://www.w3.org/2003/06/sw-vocab-status/ns#	2	2		1
schema	http://schema.org/	2	1		2
wdrs	http://www.w3.org/2007/05/powder-s#	1	1		1
adms	http://www.w3.org/ns/adms#	2	2		2
voaf	http://purl.org/vocommons/voaf#	1	1		1
sumo	http://www.ontologyportal.org/SUMO.owl#	2	1	1	0
dul	http://www.loa-cnr.it/ontologies/DUL.owl#	1	1		1
p1	http://www.owl-ontologies.com/assert.owl#	1	1		0
p2	http://www.owl-ontologies.com/Ontology1148042246.owl#	1	1		0
meta	http://www.co-ode.org/ontologies/meta/2005/06/15/meta.owl#	1	1		0
abox	http://swrl.stanford.edu/ontologies/built-ins/3.3/abox.owl#"	1	1		0
tbox	http://swrl.stanford.edu/ontologies/built-ins/3.3/tbox.owl#	1	1		0
swrlx	http://swrl.stanford.edu/ontologies/built-ins/3.3/swrlx.owl#	1	1		0
swrlm	http://swrl.stanford.edu/ontologies/built-ins/3.4/swrlm.owl#	1	1		0
sqwrl	http://sqwrl.stanford.edu/ontologies/built-ins/3.4/sqwrl.owl#	2	2		0
temporal	http://swrl.stanford.edu/ontologies/built-ins/3.3/temporal.owl#	1	1	1	0
bibo	http://purl.org/ontology/bibo/	1	1		1
ombibo	http://www.wurvoc.org/bibliography/om-1.8/	1	1		1
wgs84_pos	http://www.w3.org/2003/01/geo/wgs84_pos#	3	1		0
unit	http://qudt.org/vocab/unit	1	0	1	0
void	http://rdfs.org/ns/void#	1	0		1