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

Deliverable D2.1: Strategy for Energy Management System Interoperability

Document Details

Delivery date:	*M3*
Lead Beneficiary:	*INRIA*
Dissemination Level (*):	*PU*
Version:	*2.4*
Preparation Date:	*2014-01-08*
Prepared by:	*Luz Maria Priego (INRIA), Jérôme Euzenat (INRIA), Raúl García-Castro (UPM), María Poveda-Villalón (UPM), Filip Radulovic (UPM), Matthias Weise (AEC3)*
Reviewed by:	*Jan Peters-Anders (AIT), Thanasis Tryferidis (CERTH)*
Approved by:	*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:	Rogelio Segovia

Revision History

Date	Author	Partner	Content	Ver.
2013-10-01	Jérôme Euzenat	INRIA	Filled in structure	0.1
2013-11-06	Jérôme Euzenat	INRIA	Change template, added material in Chapter 1 and 2	0.2
2013-11-15	Raúl García-Castro	UPM	Deliverable structure	0.3
2013-11-27	María Poveda-Villalón	UPM	Structure draft and first contributions to sections 2.1 and 3	0.4
2013-12-03	María Poveda-Villalón and Raúl García-Castro	UPM	Update of sections 2, 3 and 4	0.5
2013-12-08	Jérôme Euzenat, Luz Maria Priego	INRIA	Improvement of two first chapters, addition of Alignment section of the others	0.6
2013-12-09	Jérôme Euzenat, Matthias Weise	AEC3	Contribution to Chapter 2 and 3	0.7
2013-12-10	Luz Maria Priego	INRIA	Updated document from the content in the wiki First version of the document	1.0
2013-12-19	Luz Maria Priego	INRIA	Updated document from reviewers comments Second version of the document	2.0
2013-12-23	Luz Maria Priego	INRIA	Updated document from project members' comments Alignment with D3.1	2.1
2014-01-06	Luz Maria Priego	INRIA	Synchronized with D3.1 common sections	2.2
2014-01-07	Luz Maria Priego	INRIA	Dates standardization Document details changes	2.3
2014-01-08	Luz Maria Priego	INRIA	Chapter 3 updated based on UPM comments	2.4

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

The goal of the Ready4SmartCities project is to support energy data interoperability in the context of SmartCities. It keeps a precise focus on building and urban data.

Work package 2 is more specifically concerned with identifying the knowledge and data resources available or needed, that support energy management system interoperability. This deliverable defines the strategy to be used in WP2 for achieving its goal.

It is made of two parts: identifying domains and stakeholders specific to the WP2 activity and the methodology used in WP2 and WP3.

Relevant domains (Section 2) and stakeholders (Section 3) are separated in two parts: level 1 and level 2 depending on their direct involvement in energy management. Level 1 is typically concerned with urban energy consumers and producers as well as resources necessary for expressing their energy data. Level 2 is concerned with peripheral domains or stakeholders such as weather information and tourism industry.

Concerning the methodology, we consider as resources the semantic resources that may be used and shared by stakeholders. These are ontologies, data sets and alignments.

The collection of resources (Section4) is a critical part of the project and will take advantage of as many sources as possible: partners knowledge, literature review, norm analysis, resource directories and data set investigations. It will also have to identify missing resources.

The identification of resources (Section 5) will be carried out with standard metadata vocabularies covering content, provenance, rights and maintenance of resources. It will mostly reuse state-of-the-art vocabularies.

The assessment of the identified resources (Section 6) will be carried out according to two set of standard indicators (5 star rating and open data index indicators) allowing for better qualifying available resource usability.

Finally the information about such resources (Section 7) will be published, in addition to the formal deliverables, in RDF on the project web site. Some of the resources, such as alignments, may be directly available from a server as well.

Glossary

Alignment	The result of analysing 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 colours.
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.

Table of Contents

1 Introduction

1.1 Purpose of this Document

The purpose of this document is to define the strategy to be followed by Work Package 2 for carrying out its task of supporting energy efficiency data interoperability. Such strategy will influence the results of this work package which will be further used in other parts of the project, as Figure 1 shows.

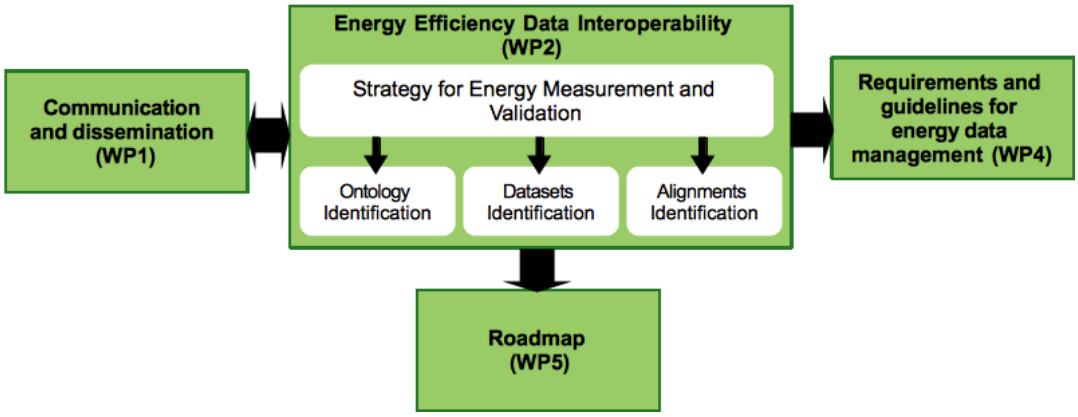

Figure 1. Relationships between Work Package 2 and other Work Packages

More precisely, it identifies which domains fall in the perimeter of the work package and which stakeholders are active in this domain. In addition, it provides a strategy for identifying resources for helping interoperability (ontologies, datasets and alignments) and characterising their use. It also further refines the task of assessing the quality of such resources and the way they could be published or rather publicised.

This work package, and more prominently its strategy, shares a lot with work package 3 dedicated to energy measurement and validation. Hence, both deliverables overlap: they have the same structure and some of their sections are identical. The identical sections are those related to the strategy to be followed for collecting, identifying, assessing and publishing the relevant ontologies, datasets and alignments; this way, the strategy followed in both work packages is homogeneous.

SmartCities are cities which make the most of the Information and CommunicationTechnology (ICT) infrastructure to provide adapted services to the citizens. According to the FP7 SmartCity program, a smart city is more precisely considered as an urban area with efficiently managed energy systems involving a collaborative network of cross-domain organizations. This requires to share relevant knowledge and data, support reasoning and offer solutions for energy systems. Although the Ready4SmartCity project focusses on exchange of energy and energy-related data, the SmartCity context is broader and concerns any type of city related data.

There are two approaches for identifying domains and stakeholders:

1. To define use cases for Smart Cities as far as possible and derive required domains, stakeholders and data requirements;
2. To make a survey of data, domains and stakeholders related to energy management systems.

Given the open-ended aspect of SmartCities, we will adopt the second attitude rather than the first one which would restrict too much the scope of the project. However, during the project, we will have to follow the use case and scenarios developed for the roadmap in work package 5 to ensure that the activities of work packages 2 and

3 actually cover those scenarios. In order to support new energy trends (e.g. marine energy) we will identify the missing resources stated in section 4.7

1.2 Document Structure

This document's structure is shared with D3.1. After this introduction, we will investigate the domains (Section 2) and stakeholder categories (Section 3) involved in the exchange of energy data within SmartCities. This is the part specific to D2.1.

In the remainder, we will provide the strategy for collecting (Section 4), identifying (Section 5), assessing (Section 6) and publishing resources (Section 7) for the purpose of exchanging energy data, i.e., ontologies, datasets and alignments. These sections are identical in both deliverables.

1.3 Contribution of Partners

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

- Introduction and conclusions. UPM, INRIA, AEC3
- Relevant domains and stakeholders. UPM, INRIA, AEC3
- Collection, identification, assessment, and publication methods. UPM, INRIA
- Comments before quality control stages. AIT, CERTH, VTT

2 Relevant domains

The main domain of the Ready4SmartCities project and WP2 in particular is the exchange of energy data in SmartCities. We define a first level in which we identify the principal domains and stakeholders dealing with energy data interchange. However, SmartCities is a never ending field and interoperability should spread to other components. Hence, we define a second level involving contexts with data of interest for exploiting and complementing the primary data.

The strategy to be deployed at the early stage of the project will be to collect all stakeholders and domains in SmartCities and to try to show their interdependencies on a high level. But going into details (ontologies, data and alignment) we will reduce the scope of work. We may certainly focus on buildings and urban planning in the centre of our investigation and to arrange all other domains around.

2.1 Level 1 domains

At this level, domains will concern data about **energy systems in SmartCities**. This involves defining types of energy as well as type of production, distribution and consumption.

Within cities, energy production and distribution should cover:

- Energy sources and forms:
  - a primary energy source is a non converted or transformed energy like fossil (coal, natural gas, oil), solar, wind, etc.
  - an energy form such as electricity, heating, cooling, transport fuels are transformed from a primary energy source;
- Distribution network and storage:
  - energy distribution network (power grids, heat supply, gas supply, district heating and cooling, gas stations/e-charge, fuel supply, etc.),
  - energy storage systems for electricity and heat energy (hourly, daily and seasonal storages, such as pumping plant and reservoir, night-storage heater, storage battery, etc.).

Energy consumption should cover:

- Buildings:
  - building envelope (material and construction types, e.g., thermal insulation, roof type, orientation),
  - building automation, HVAC (Heating, Ventilating, Air Conditioning) and electrical systems (all kinds of technical equipment of the building for heating, cooling, ventilation, lightning and hot domestic water supply),
  - facilities management (data for operation and maintenance);
- Transportation:
  - transport management including equipment (public transportation, time-tables, rental-cars, car-pools, private hybrid or electrical car/bicycle and their charging infrastructure),
- Outdoor and street lighting,
- Waste and water management.

This will also cover data about energy quantity and its characterization (i. e., to precise in which context such data applies):

- Temporal framework (When is energy consumed? e.g., date/ time);
- Spatial framework (Where is energy consumed? e.g., place/ physical address):
  - Geographical framework,

    - Urban framework: urban planning (residential, commercial, industry areas), urban road system (private and public transport) including traffic engineering;
- Organisational framework (for aggregating Who determines energy measurements? e.g., functional units / normative organizations / standards).

2.2 Level 2 domains

Beyond the sheer energy data, exchanges may also cover factors influencing this data. Such information is typically what can be traded in the context of SmartCities and which is precious to analyse the level 1 data.

These extend the concerned domain data to the wide range of dynamic information:

- Activity data about company and individual behaviour and expected/actual consumption:
    - occupancy and user behaviour (user data to energy profiles for office work, commerce, residential use, etc.),
    - consumer device (use, consumption);
- Weather data which may affect both the consumption and the production of energy (in case of solar or wind energy);
- Climate zones establishing seasons, weather patterns influencing heat/cold/humid days;
- Energy market data which may affect the prices of energy as well as the demand for it.

3 Relevant stakeholders

Similarly as in the case of the relevant domains, we define two levels of stakeholders related to the domain levels.

3.1 Level 1 stakeholders

The stakeholders of level 1 are necessarily **providers and consumers** of energy.

Providers may be divided according to the source used to produce energy (gas, wind, coal) or the form of energy provided (heating, cooling, transport fuel).

Consumers may be considered according to their activities:

- Citizens,
- Building managers,
- Tenants and owners,
- Municipality and governance bodies:
  - local and city councils (see D3.1),
  - municipal services,
- Companies.

This classification may be further refined.

We may add to the providers and consumers any intermediary bodies such as energy distributors (producers or cities may be distributors as well).

Finally, in the framework of local provision of energy, consumers themselves may become "prosumers", i.e., both producers and consumers of energy.

3.2 Level 2 stakeholders

From the concerned domains, stakeholders are naturally:

- Energy markets,
- The construction industry (architects, energy experts, HVAC engineer, etc.),
- The manufacturing industry (heater, solar panels, etc.),
- Citizens and city bodies (as providing these level 2 data),
- Normative bodies,
- All sorts of companies having their activities in the city:
  - the tourist industry,
  - energy providers (see D3.1),
  - the car industry,
  - the waste industry.

To this we may add stakeholders who produce data that may be used by citizens for controlling their energy consumption:

- weather forecast / observation,
- traffic management.

4 Collection methods

The methods to be used in READY4SmartCities for collecting ontologies, datasets and alignments may vary depending on the type of resource, however they can be organised according to the type of collection method. Next, we describe all the methods that will be used; these methods will be applied concurrently in the work package.

4.1 Involve project partners

Clearly, project partners will be a primary source for collecting information about ontologies, datasets and alignments. Therefore, their active involvement will be required during the project lifetime.

Most project partners belong to the Energy Efficient Building Association (E2BA) or have a deep expertise in ICT in the building sector. To mitigate this limited technical scope we will involve experts from the partners' strong network of contacts, with ties to the ICT community, energy sector, and public authorities.

4.2 Contact stakeholders

In order to collect ontology, dataset and alignment information from stakeholders, a survey will be prepared, focused on stakeholders participating in European projects. Furthermore, the organisation of VoCamps will also be an ideal way to contact stakeholders and collecting information about existing and needed resources.

4.3 Review literature

Two particular types of literature will be considered for identifying ontologies, datasets and alignments:

- Research literature, and
- European project production (see list of relevant projects as well as newly accepted projects).

4.4 Analyse standardization and institutional bodies

One important source of ontological knowledge is normative standards and legislation. Standardization bodies, such as buildingSmart, ETSI (European Telecommunications Standards Institute), CEN (European Committee for Standardization), CENELEC (European Committee for Electrotechnical Standardization), W3C (World Wide Web Consortium), OASIS (Advancing Open Standards for the Information Society), OMG (Object Management Group), ISO (International Organization for Standardization), and OGC (Open Geospatial Consortium) usually provide standards that consist of data models with precise and useful descriptions of concepts (or even ontologies). Besides, regarding datasets, institutional bodies (e.g., governments, statistical institutes) may be a source of official data sets.

4.5 Lookup resource catalogues

There are several catalogues that may be looked up for identifying ontologies, datasets and alignments relevant to READY4SmartCities:

- Ontology search engines:
  - Watson http://watson.kmi.open.ac.uk/

- Swoogle http://swoogle.umbc.edu
- Linked Open Vocabularies (LOV) http://lov.okfn.org/dataset/lov/

- Linked Data dataset catalogues:
  - Datahub (Data Management Systems) http://datahub.io
  - Reegle http://data.reegle.info/
  - Open Energy Information (OpenEI) http://en.openei.org/datasets/
- Open data catalogues:
  - Open Government Data http://opengovernmentdata.org/data/
  - Open Data Index https://index.okfn.org/
  - European Union Open Data Portal http://open-data.europa.eu/en/data/
  - Linking Open Government Data http://logd.tw.rpi.edu/
  - The European Open Government Data Initiative http://www.govdata.eu/en/europeanopen.aspx
  - Geonames http://www.geonames.org
  - Open Data initiatives from EU countries
- Web portals, which contain data sets from a concrete organization or a domain
  - The Buildings Performance Institute Europe (BPIE) http://www.buildingsdata.eu/data-search
  - University of Missouri https://library.missouri.edu/guides/data/inter-data/
  - The World Bank http://datacatalog.worldbank.org
  - Clean Web Initiative http://cleanweb.co
  - Engage project http://www.engagedata.eu/

4.6 Dataset investigation

The investigation of available data sets will allow to identify the ontologies and vocabularies that they use. They may also provide links to other data sets as well as information about connections between these ontologies. Alignment may be built from existing links.

4.7 Identification of missing resources

The output of the collection task will be a list of relevant ontologies, datasets and alignments (classified into domains). For each of these resources, proper information about them will be provided in terms of metadata; these metadata are discussed in the next section.

5 Identification methods

The identification of ontologies, datasets and alignments will be carried out by means of forms to be filled by READY4SmartCities partners. For designing and sharing the forms the Google Drive technology will be used as it offers an easy interface for creating the forms and gathering the responses which are stored in the form of a spreadsheet. In addition, answers might be modified or completed afterwards, if needed. Once the identification process is finished, the contents of the spreadsheet will be transformed to RDF and published as explain in Section 7.

In READY4SmartCities we will use a common set of metadata to describe the ontologies, datasets and alignments to be identified in the different work packages of the project (WP2 and WP3). Furthermore, we will identify the vocabulary terms that can be used to represent those metadata in RDF, in order to allow the automated processing of such metadata. The following sections describe those metadata.

5.1 Metadata for Ontologies

Once the number of ontologies available in the Web (coming not only from research but also from industry) was significant, it was clear the need for adding a set of metadata to ontologies in order to facilitate their reuse. The OMV[1] (Ontology Metadata Vocabulary) [Hartmann et al., 2005] was the first attempt to define an ontology that could be used to represent ontology metadata. However, one limitation of OMV is that it does not reuse terms already defined in other well-known ontologies.

Years later, Vocabulary of a Friend[2] (VOAF) was defined following a similar motivation: plenty of ontologies are being used in the Linked Data cloud and knowing the relationships between those ontologies could be used to enhance their reusability by defining networks of ontologies. VOAF reuses as many properties as possible from other ontologies (such as Dublin Core or VANN), which is one of the best practices when developing ontologies. This way, VOAF relies on existing well-known ontologies for generic ontology metadata (e.g., the name or the version of the ontology) and defines new properties for expressing the different ways ontologies can rely on, extend, specify, annotate, or link to each other.

Other approaches for describing ontology metadata exist; however, they are specific to the concrete ontology repositories that use them (e.g., BioPortal[3] [Whetzel et al., 2011] in the biomedical domain or OntoSelect[4] [Buitelaar and Eigner, 2008]).

In order to describe the ontologies that are identified in the project we will use a common set of metadata. These metadata have been selected after analysing 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). Our decision has been to choose VOAF as a seed ontology, we have removed those properties that are not needed in this setting right now, and we have added those from OMV in those cases not covered by VOAF. In our case, we are not using any of the properties defined under the VOAF namespace because they mainly deal with describing relationships between ontologies (e.g., *voaf:reliesOn* or *voaf:usedBy*), which are out of the scope of our catalogue at this moment.

Table 1 summarizes the metadata to be collected from ontologies in the energy efficiency domain, along with the corresponding properties to be used in the RDF representation and their range.

1 http://omv2.sourceforge.net/

2 http://lov.okfn.org/vocab/voaf

3 http://bioportal.bioontology.org/

4 http://olp.dfki.de/OntoSelect/

Identifier	Definition	Range	RDF property (and range)
Name	The name given to the ontology	String	dc:title (xsd:string)
Description	A free-text account of the ontology	String	dc:description (xsd:string)
Domains	The different domains covered by the ontology	String	omv:hasDomain (skos:Concept)
Version	The version of the ontology	String	owl:versionInfo (xsd:string)
Creation date	The date of formal issuance of the ontology	Date	dc:issued (xsd:dateTime)
Last update	Most recent date on which the ontology was changed, updated or modified	Date	dc:modified (xsd:dateTime)
Contact person	The person(s) primarily responsible for making the ontology	String	dc:creator (foaf:Person)
Publisher	The organization that published the ontology	String	dc:publisher (foaf:Organization)
License	The license of the ontology	String	cc:license (cc:License)
URI	The URI of the ontology	URI	omv:URI (xsd:anyURI)
Namespace	The preferred namespace URI to use when using terms from this vocabulary	URI	vann:preferredNamespaceUri (xsd:anyURI)
Format	The format of the ontology	String (e.g., RDF/XML, Turtle, N3)	omv:hasOntologySyntax (omv:OntologySyntax)
Ontology language	The language in which the ontology is implemented	String (e.g., OWL, RDF-S)	omv:hasOntologyLanguage (omv:OntologyLanguage)
Language	The language of the resource	String (using RFC 4646)	dc:language (dc:LinguisticSystem)
Comments	Further information about the ontology in the context of our catalogue	String	rdfs:comment (xsd:string)
References	Resources that might provide additional information (documents, deliverables, papers, etc.)	URI	rdfs:seeAlso (xsd:anyURI)

Table 1. Ontology metadata to be collected in Ready4SmartCities

Since the set of metadata that we are interested in gathering is not fully covered by either ontology we cannot directly reuse these approaches. Therefore, we have reused the properties used in VOAF and added those from OMV in the cases not covered by VOAF. In our case, we are not using any of the properties defined under the VOAF namespace because they mainly deal with describing relationships between ontologies (e.g., *voaf:reliesOn* or *voaf:usedBy*), which are out of the scope of our catalogue. However, we follow their guidelines on reusing properties from other ontologies; VOAF reuses as many properties as possible from other ontologies instead of reinventing the wheel, which is one of the best practices when developing ontologies and vocabularies.

As can be seen in Table 1, we reuse vocabulary terms from *cc* (Creative Commons Rights Expression Language), *dc* (DCMI Metadata Terms), *foaf* (Friend of a Friend), *omv* (Ontology Metadata Vocabulary), and *vann* (VANN: A vocabulary for annotating vocabulary descriptions). Table 2 lists the vocabularies reused and their URIs.

Vocabulary	Prefix	URI
Creative Commons Rights Expression Language	cc	http://creativecommons.org/ns
DCMI Metadata Terms	dc	http://purl.org/dc/terms/
Friend of a Friend	foaf	http://xmlns.com/foaf/0.1/
Ontology metadata vocabulary	omv	http://omv.ontoware.org/2005/05/ontology#
VANN: A vocabulary for annotating vocabulary descriptions	vann	http://purl.org/vocab/vann/

Table 2. Vocabularies, prefixes and URIs relation

Notwithstanding, we also include in Table 3 the mappings between the metadata to be used in READY4SmartCities and their equivalent terms in VOAF and OMV. As can be seen, the mapping between our proposal and VOAF is almost direct.

Identifier	READY4SmartCities	Mapping to VOAF	Mapping to OMV
Name	**dc:title**	dc:title	omv:name
Description	**dc:description**	dc:description	omv:description
Domains	**omv:hasDomain**	dc:isPartOf	omv:hasDomain
Version	**owl:versionInfo**	owl:versionInfo	omv:version
Creation Date	**dc:issued**	dc:issued	omv:creationDate
Last update	**dc:modified**	dc:modified	omv:modificationDate
Contact person	**dc:creator**	dc:creator	omv:hasCreator
Publisher	**dc:pubisher**	dc:publisher	--
Licence	**cc:license**	cc:license	omv:hasLicense
URI	**omv:URI**	-- (uses the ontology URI)	omv:URI
Namespace	**vann:preferredNamespaceUri**	vann:preferredNamespace	--
Format	**omv:hasOntologySyntax**	--	omv:hasOntologySyntax
Ontology language	**omv:hasOntologyLanguage**	--	omv:hasOntologyLanguage
Language	**dc:language**	dc:language	omv:naturalLanguage
Comments	**rdfs:comment**	rdfs:comment	--

References	rdfs:seeAlso	rdfs:seeAlso	omv:documentation

Table 3. Mapping of vocabulary terms to VOAF and OMV

5.2 Metadata for Datasets

In order to describe the datasets that are identified in the project we will use a common set of metadata. In this case, these metadata have been selected from the Data Catalog Vocabulary[5] (DCAT)), which is the ontology that is in the process of being standardised by the W3C for describing dataset metadata and that extensively reuses terms from other well-known vocabularies, such as Dublin Core, FOAF and SKOS. One advantage of DCAT is that it does not make any assumption about the format of the datasets used in a catalogue, while the other possible approach for describing dataset metadata, VoiD[6] [Alexander et al, 2011], is only valid for RDF datasets.

The list of metadata to be collected from datasets is defined in Table 4, along with the corresponding properties to be used in the RDF representation and their range.

Identifier	Definition	Range	RDF property (and range)
Name	The name given to the dataset	String	dc:title (xsd:sting)
Description	A free-text account of the dataset	String	dc:description (xsd:string)
Domains	The different domains covered by the dataset	String	dcat:theme (skos:Concept)
Version	The version of the dataset	String	dcat:distribution (dcat:Distribution)
Publication date	Publication date of the dataset	Date	dc:issued (xsd:date)
Last update	Most recent date on which the dataset was changed, updated or modified	Date	dc:modified (xsd:date)
Update frequency	The frequency at which the dataset is updated	String	dc:accuralPeriodicity (dc:Frequency)
Contact person	Relevant contact information	String	dcat:contactPoint (vcard:Kind)
Publisher	The organization that published the dataset	String	dc:publisher (foaf:Agent)
License	The license of the dataset	String	dc:license (dc:LicenseDocument)
URI	The URI of the dataset	URI	dcat:accessURL (rdfs:Resource)
Format	The format of the dataset	String	dc:format (dc:MediaTypeOrExtent)
Language	The language of the dataset	String (using RFC 4646)	dc:language (dc:LinguisticSystem)

5 http://www.w3.org/TR/vocab-dcat/

6 http://vocab.deri.ie/void/

Comments	Further information about the dataset in the context of our catalogue	String	rdfs:comment (xsd:string)
References	Resources that might provide additional information (documents, deliverables, papers, etc.)	URI	rdfs:seeAlso (xsd:anyURI)

Table 4. Dataset metadata to be collected in Ready4SmartCities

As can be seen in the Table 4, DCAT reuses vocabulary terms from *dc* (DCMI Metadata Terms), *foaf* (Friend of a Friend) and *vcard* (W3C vCard ontology). Table 5 lists the vocabularies reused and their URIs.

Vocabulary	Prefix	URI
Vocabulary of a Friend	dcat	http://purl.org/vocommons/voaf
DCMI Metadata Terms	dc	http://purl.org/dc/terms/
Friend of a Friend	foaf	http://xmlns.com/foaf/0.1/
W3C vCard ontology	vcard	http://www.w3.org/2006/vcard/ns#

Table 5. Vocabularies, prefixes and URIs relation

5.3 Metadata for Alignments

An ontology alignment is a set of correspondences between semantically related entities from two ontologies [Euzenat and Shvaiko, 2013]; these correspondences can be used for various tasks, such as ontology merging, query answering, or data translation. In order to describe alignments, we will use the Alignment format set of tags[7] complemented by the W3C Provenance ontology[8]. The Provenance ontology will replace the previous use of OMV in the Alignment format.

The list of metadata to be collected for alignments is defined in Table 6, along with the corresponding properties to be used in the RDF representation and their range.

Identifier	Definition	Range	RDF property (and range)
URI	The URI of the alignment	URI	align:id (xsd:anyURI)
Name	The name given to the alignment	String	align:pretty (xsd:string)
Comments	Further information about the alignment in the context of our catalogue	String	rdfs:comment (xsd:string)
Aligned ontology	The first aligned ontology	URI	align:onto1(xsd:anyURI)
Aligned ontology	The second aligned ontology	URI	align:onto2(xsd:anyURI)

7 http://alignapi.gforge.inria.fr/labels.html

8 http://www.w3.org/TR/prov-o/

Level	The type of language used for expressing correspondences	String	align:level (xsd:string)
Type	The parity of the alignment (1:1, *:*, etc.)	String	align:type (xsd:string)
Purpose	The purpose of the alignment	String	omwg:purpose (xsd:string)
License	The license of the alignment	String	cc:license (cc:License)
Generation process	The process that generated the alignment	Activity	provo:wasGeneratedBy (Activity)
Creation date	The date of formal issuance of the alignment	Date	provo:endedAtTime (xsd:dateTime)
Generator	The person or organisation having performed the activity that has produced the alignment	Person	provo:agent (foaf:Person)
Method	The method used in the process	Classname	align:method (xsd:string)
Method version	The version of the method used	String	align:methodVersion (xsd:string)
Derived form	An alignment used in the process	Alignment	provo:used (xsd:anyURI)

Table 6. Alignment metadata to be collected in Ready4SmartCities

These metadata deprecate several of the properties used with the Alignment API [David et al, 2011].

- `align:method` is now a property of `Matching`;
- `align:methodVersion` may also be added to the `Matching` properties;
- `align:derivedFrom` could be replaced by `provo:used`;
- `align:time` should now be applied to the `Activity`;
- `align:provenance` was not really well defined and is abandoned;
- `dc:creator` is replaced or can still be used as a subproperty of `provo:agent`;
- `dc:date` should be replaced by `provo:endedAtTime`.

As can be seen in the Table 6, apart from the Alignment format and the Provenance ontology we reuse vocabulary terms from *cc* (Creative Commons Rights Expression Language), *dc* (DCMI Metadata Terms), *foaf* (Friend of a friend), *omv* (Ontology metadata vocabulary) and *omwg* (Ontology Mapping Working Group). Table 7 lists the vocabularies reused and their URIs.

Vocabulary	Prefix	URI
Alignment format	align	http://knowledgeweb.semanticweb.org/heterogeneity/alignment#
Creative Commons Rights Expression Language	cc	http://creativecommons.org/ns
DCMI Metadata Terms	dc	http://purl.org/dc/terms/
Friend of a Friend	foaf	http://xmlns.com/foaf/0.1/

Ontology metadata vocabulary	omv	http://omv.ontoware.org/2005/05/ontology#
Ontology Mapping Working Group	omwg	http://www.omwg.org/TR/d7/d7.2/
Provenance ontology	provo	http://www.w3.org/ns/prov#

Table 7 Vocabularies, prefixes and URIs relation used for annotating alignments

Mappings between *prov-o* and Dublin core have already been documented[9]; those between this setting and the OMV mapping will be documented on the Alignment format page.

This section has described the common set of metadata that will be used to describe the ontologies, datasets and alignments to be identified in the different work packages of the project (WP2 and WP3). After starting building the catalogues it may happen that new metadata would be required. In that case, we will extend these metadata as needed.

9 http://www.w3.org/TR/prov-dc/

6 Assessment

The assessment of ontologies, datasets and alignments will be carried out manually by READY4SmartCities partners. In READY4SmartCities we will use a common set of methods to assess the ontologies, datasets and alignments identified in the different work packages of the project (WP2 and WP3).

The main guidelines for publishing data over the web are the extremely well-known Linked Data principles and the Linked Open Data 5 Star rating system defined by Tim Berners-Lee[10]. More precisely, the rating system defines the following levels (taken literally from the source):

LOD1. *Available on the web (whatever format) but with an open licence, to be Open Data*

LOD2. *Available as machine-readable structured data (e.g., excel instead of image scan of a table)*

LOD3. *As (2.) plus non-proprietary format (e.g., CSV instead of excel)*

LOD4. *All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff*

LOD5. *All the above plus Link your data to other people's data to provide context*

The Open Data Index[11] also provides nine indicators used to measure how open is the data from a given country in a given domain. These indicators are:

ODI1. *Does the data exist?*

ODI2. *Is it in digital form?*

ODI3. *Is it publicly available?*

ODI4. *Is it free of charge?*

ODI5. *Is it online?*

ODI6. *Is it machine-readable?*

ODI7. *Is it available in bulk?*

ODI8. *Is it openly licensed?*

ODI9. *Is it up to date?*

The following sections present the assessment methods to be used for ontologies, datasets and alignments. In them, we will refer to the indicators stated in these rating systems as LOD or ODI plus its ordinal numeration according to the lists above. Furthermore, for each indicator we will collect an answer or "Yes", "No", or "Unknown". We must note that we have not taken into account every indicator for every type of resource; for instance, while it makes sense to assess whether some government data is public or not (i.e., ODI3), such assessment makes no sense when dealing with ontologies or alignments.

6.1 Ontology Assessment

In first instance, the indicators for assessing ontologies taken into account are:

10 http://www.w3.org/DesignIssues/LinkedData.html

11 https://index.okfn.org/

- **Whether the ontology is available on the Web (Whatever format).** This indicator is related to LOD1 and ODI5.
- **Whether the ontology is available following the W3C standards (SKOS, RDF-S or OWL).** This indicator is related to LOD4 and ODI6.
- **Whether the ontology is available under an open license.** This indicator is related to LOD1 and ODI8.

In addition, in order to provide a more detailed assessment (e.g. related to good modelling practices), the OWL ontologies available on the web could be evaluated by means of external evaluation services such as OOPS![12] (OntOlogy Pitfall Scanner!) [Poveda-Villalón et al., 2012] which is an on-line application to identify pitfalls in ontologies. Other services for detecting good practices might also be used [Poveda-Villalón et al., 2013].

6.2 Dataset Assessment

For assessing dataset the following indicators, mainly based on the Open Data Index criteria, will be used:

- **Whether the dataset is in digital form.** This indicator is related to ODI2.
- **Whether the dataset is publicly available.** This indicator is related to ODI3
- **Whether the dataset is free of charge.** This indicator is related to ODI4
- **Whether the dataset is online.** This indicator is related to ODI5 and LOD1.
- **Whether the dataset is machine-readable.** This indicator is related to ODI6 and LOD2.
- **Whether the dataset is available in bulk.** This indicator is related to ODI7.
- **Whether the dataset is openly licensed.** This indicator is related to ODI8 and LOD1.
- **Whether the dataset is up to date.** This indicator is related to ODI9.

In a first approximation no existing services for dataset assessment or validation can be used due to the diversity of types in which the dataset could be distributed and to the fact that existing services are too specific (e.g., Databugger[13])

6.3 Alignment assessment

In first instance, the indicators for assessing alignments taken into account are:

- **Whether an alignment is identifiable for a pair of ontologies.**
- **Whether the alignment is in digital form.** This indicator is related to ODI2.
- **Whether the alignment is available on the Web (whatever format).** This indicator is related to LOD1 and ODI5.
- **Whether the alignment is available following the standard Alignment format.** This indicator is related to LOD4 and ODI6.

12 http://www.oeg-upm.net/oops/

13 http://databugger.aksw.org/

- **Whether the alignment is available under an open license.** This indicator is related to LOD1 and ODI8.
- **Whether the alignment maintained is up to date (in particular with respect to ontologies)**. This indicator is related to ODI9.

In addition, if necessary, we may start using tools for assessing the consistency of alignments (e.g., LogMap[14] or Alcomo[15]).

This section has described the common set of methods that will be used to assess the ontologies, datasets and alignments identified in the different work packages of the project (WP2 and WP3). After starting building the catalogues it may happen that we see the need for using new assessment methods (e.g., that support an automated assessment or are more informative); in that case, we will use such methods as needed.

14 http://www.cs.ox.ac.uk/isg/tools/LogMap/

15 http://web.informatik.uni-mannheim.de/alcomo/

7 Publication methods

In READY4SmartCities we will publish catalogues including the collected information about ontology, datasets and alignments both by means of a web page where the catalogues will be displayed and by means of machine-processable data using Web standards (i.e., RDF); together with the public deliverables of the project.

In the catalogues we will only publish the metadata described in the previous section. Our goal is not to publish the resources ourselves, but to render them accessible. However, the case of alignments is different because we expect many alignments to be missing or not to be in relevant formats; hence we plan to set up alignment servers supporting them.

The catalogues in the RDF format will be published in RDF files on a web server. In addition, a web page will show the catalogues in a human-readable way. A preliminary prototype of the web pages of the ontology, dataset and alignment catalogues are shown in Figure 2, 3 and 4, respectively

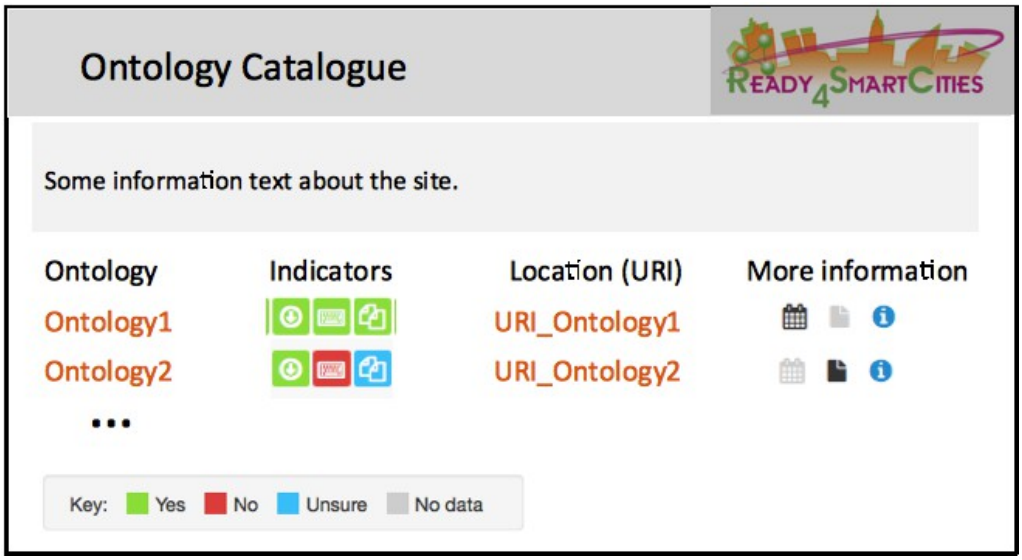

Figure 2. Ontology catalogue web interface prototype

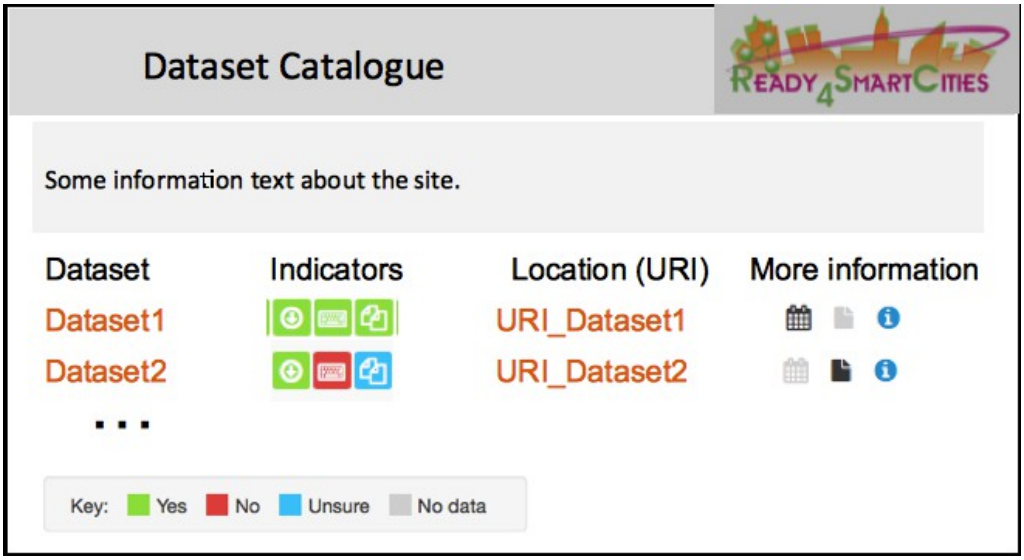

Figure 3. Dataset catalogue web interface prototype

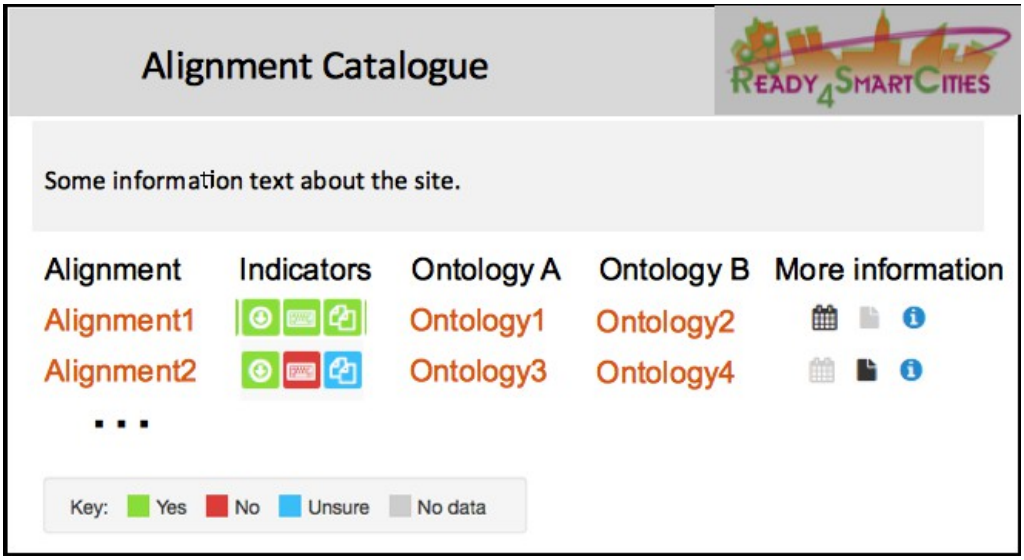

Figure 4. Alignment catalogue web interface prototype

As can be seen, the result representation provided at Open Data Index[16] will be taken as reference. That is, representing the different values for each indicator observed by means of an icon and colours. In this way, a green icon will mean "Yes", a red icon will mean "No", and a blue icon will mean "Unknown" [17].

In addition, it may be possible that we provide advanced programmatic access to the catalogues under other ways such as SPARQL endpoints or Alignment servers.

16 https://index.okfn.org/

17 An example could be found at https://index.okfn.org/country/overview/Spain/ (last access 27th November, 2013)

8 Conclusions

This deliverable established the Work package 2 strategy for fostering energy efficiency data interoperability. It circumscribed the domains and category of stakeholders to be involved and considered for the interoperability of energy data to be exchanged in SmartCities, and especially in relation with Energy management systems. It then described the same strategy as workpackage 3 for collecting, identifying, assessing and publishing resources, i.e., ontologies, data and alignments.

We plan to follow a common strategy in work packages 2 and 3, even if the relevant domains and stakeholders differ in both work packages. One advantage of this is that the way of documenting and presenting the catalogues to people outside the project will be homogeneous. Furthermore, the use of information combined from these catalogues will be easier and, to this end, we will also provide the catalogues in a machine-processable format. We also expect that good ideas and practices can so easily move from one work package to the other.

The identification of concrete stakeholders and domains will also help in other tasks in the project, such as in defining concrete dissemination or community engagement activities (work package 1); adapting the guidelines for energy data generation, publication and exploitation (work package 4); or defining the roadmap (work package 5). This stakeholder and domains may evolve as the R4SC project

Furthermore, the ontology, dataset and alignment catalogues will help not only identifying the resources that support interoperability but also highlighting the current gaps where effort should be put and the existing trends in the different domains (across e.g., harmonisation of ontologies, availability and privacy of energy data).

9 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.

[Buitelaar and Eigner, 2008] P. Buitelaar and T. Eigner. Ontology Search with the OntoSelect Ontology Library. 6th International Conference on Language Resources and Evaluation (LREC'08). Marrakech, Morocco. 28-30 May 2008.

[David et al, 2011] J. David, J. Euzenat, F. Scharffe, and C.T. dos Santos, "The Alignment Api 4.0," Semantic Web J., vol. 2, no. 1, pp. 3-10, 2011.

[Euzenat and Shvaiko, 2013] J. Euzenat, P. Shvaiko, Ontology matching, Springer-Verlag, Heidelberg (DE), 520p., 2013

[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.

[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, Irlanda, del 8 al 12 de octubre de 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 Pattern (WOP2013) Workshop at the 12th International Semantic Web Conference (ISWC2013). Sydney, Australia. 21 Octubre 2013.

[Whetzel et al., 2011] P.L. Whetzel, N.F. Noy, N.H. Shah, P.R. Alexander, C. Nyulas, T. Tudorache, M.A. Musen. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res. 14 June 2011.