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

Deliverable D3.1: Strategy for Energy Measurement and Interoperability

Document Details

Delivery date:	*M3*
Lead Beneficiary:	*Universidad Politécnica de Madrid*
Dissemination Level (*):	*PU*
Version:	*2.0*
Preparation Date:	*15/01/2014*
Prepared by:	*Raúl García-Castro (UPM), María Poveda-Villalón (UPM), Filip Radulovic (UPM), Asunción Gómez-Pérez (UPM), Jérôme Euzenat (INRIA), Luz-Maria Priego-Roche (INRIA), Georg Vogt (EMP), Simon Robinson (EMP), Strahil Birov (EMP), Bruno Fies (CSTB)*
Reviewed by:	*Jan Peters-Anders (AIT), Luz-Maria Priego-Roche (INRIA)*
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:	Rogelio Segovia

SEVENTH FRAMEWORK PROGRAMME

Revision History

Date	Author	Partner	Content	Ver.
15/11/2013	Raúl García-Castro	UPM	Deliverable structure	0.1
27/11/2013	María Poveda-Villalón	UPM	Structure draft and first contributions to sections 2.1 and 3	0.2
28/11/2013	Filip Radulovic	UPM	First contributions to section 2.2	0.3
3/12/2013	María Poveda-Villalón and Raúl García-Castro	UPM	Update of sections 2, 3 and 4	0.4
9/12/2013	Raúl García-Castro, Jérôme Euzenat, Georg Vogt, Simon Robinson, Strahil Birov	UPM, INRIA, EMP	Updated document from the content in the wiki First version of the document	1.0
17/12/2013	Raúl García-Castro, Strahil Birov	UPM, EMP	Implemented comments from reviewers: Mari Sepponen (VTT), Jan Peters-Anders and Luz-Maria Priego-Roche (INRIA)	1.1
27/12/2013	Raúl García-Castro, Luz-Maria Priego-Roche	UPM, INRIA, CSTB	Implemented comments from reviewers	1.2
15/01/2014	Raúl García-Castro, Bruno Fies, Simon Robinson	UPM, CSTB, EMP	Implemented last comments from reviewers Second version of the document	2.0

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This deliverable contains original unpublished work except where clearly indicated otherwise.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

The goal of the READY4SmartCities project is to support energy data interoperability in the context of Smart Cities. Work package 3 is more specifically concerned with identifying the knowledge and data resources, available or needed, that support interoperability in energy measurement and validation activities. This deliverable defines the strategy to be used in this work package for achieving its goal.

More precisely, this document identifies which domains fall in the perimeter of the work package and which stakeholders are active in this domain. In addition, it provides a concrete set of methods for collecting, identifying, assessing, and publishing the different resources that enable such interoperability (ontologies, datasets and alignments); these methods are common for work packages 2 and 3.

Relevant domains (Section 2) and stakeholders (Section 3) are separated into two levels: Level 1 contains the principal domains and stakeholders for energy measurement and validation and Level 2 involves those contexts that could enrich those identified in Level 1.

Concerning the methodology, we take into account those semantic resources that may be used and shared by stakeholders. These are ontologies, datasets and alignments.

The collection of resources (Section 4) is a critical part of the project and will take advantage of as many sources as possible: partners knowledge, literature reviews, norm analyses, resource directories, and data set investigations.

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 sets of standard indicators (5 star rating and open data index indicators) allowing for better qualifying available resource usability.

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

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 (related to, e.g., harmonisation of ontologies, availability and privacy of energy data, etc.).

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.

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 3 for carrying out its task of supporting interoperability in energy measurement and validation activities. 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.

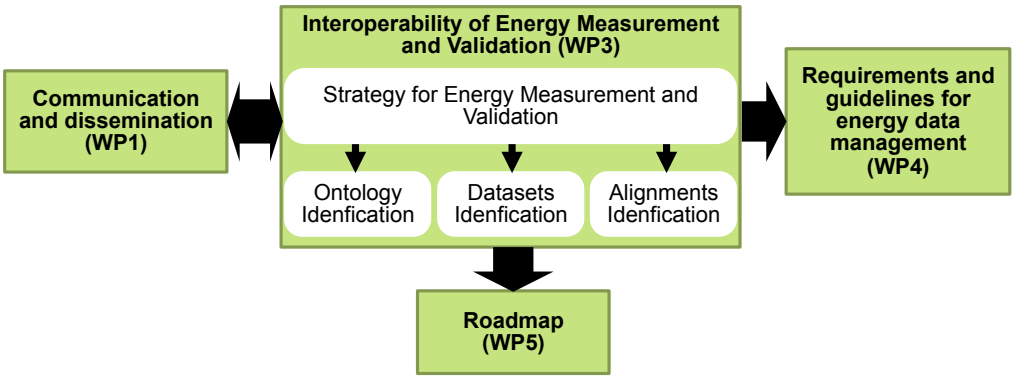

Figure 1. Relationships between work package 3 and other work packages

More precisely, this document identifies which domains fall in the perimeter of the work package and which stakeholders are active in this domain. In addition, it will provide a strategy for identifying resources for helping interoperability (ontologies, datasets and alignments) and characterising their use. It will also further refine 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 2 dedicated to energy management systems interoperability. 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.

Smart Cities are cities that make the most of Information and Communication Technology (ICT) infrastructures to provide adapted services to the citizens. According to the FP7 Smart City program, a Smart City more precisely considered as an urban area with efficiently managed energy resources 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 READY4SmartCities project focuses on exchange of energy-related data, the Smart City context is broader and concerns any type of city-related data.

At the Smart City level, it is quite complicated to draw a unique scenario illustrating the need for exchanging data at the different levels of the city and among the different stakeholders concerned. Nevertheless at such high level the main drivers are already identified.

One of the main challenges for the coming years is to develop strategies and corresponding services in order to reduce of course the overall energy consumption but also the consumption peaks. The risk of network collapse is particularly high during peak periods and for identified areas. In order to avoid a global collapse, the current strategy called “load shedding” consists in stopping energy supply to identified parts of the network in order to keep the other parts safe. With the emergence of so called “smart grid” the objective is to setup a new strategy, called either “load shifting” or “load shaping”, which consists in smoothing the peak loads by steering the demand. This load shaping can be seen as a cost-effective alternative to the installation of new production facilities (to supply the peaks) and a response to the difficulty of storing energy for later use. This approach can also be used to offset the intermittent production from renewable energy sources.

The load shaping is already in use in the electricity supply industry on a voluntary basis. When needed, an industrial sector and/or company agree to erase or decrease its electricity consumption during a certain period, for a fee. It is for the industrial concern to pause some of the electrical equipment, according to the precise terms of its contract. At the city level, the challenge is now to involve citizens in the load shaping.

This is where ICT has a crucial role to play by providing means for a seamless adaptation of buildings' energy behaviour following external directives and taking into account local constraints. The energy system will have to provide comfort to citizens by consuming locally and anticipate external demands based on a computation of various information (local measures, forecast, learned behaviours, external data and constraints, etc.). This situation is often called "Demand Response 2.0". Its objective is mainly based on the ad-hoc integration of renewables sources of energy (taking into account their intermittent production capacity) in the existing grid composed at its lower level by different energy devices. This is where the mechanisms for "Energy Trading" take place involving different stakeholders at different levels or domains.

There are two approaches for identifying domains and stakeholders:

1. To define use cases for Smart Cities as far as possible and derive the required domains and stakeholders
2. To make a survey of domains and stakeholders related to energy measurement and validation

Given the open-ended aspect of Smart Cities, we will adopt the second approach (to survey existing domains and stakeholders) rather than the first one (to define use cases) that 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 and to be aware of new trends in the state of the art to ensure that the activities of work packages 2 and 3 actually cover those scenarios and are up to date.

1.2 Document Structure

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

In the remainder, we will provide the strategy for collecting (Section 4), identifying (Section 5), assessing (Section 6) and publishing (Section 7) resources 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
- Relevant domains and stakeholders. EMP, CSTB
- Collection, identification, assessment, and publication methods. UPM, INRIA

2 Relevant Domains

The relevant domains in work package 3 are those that allow assessing the success of innovation projects that deal with the interchange of energy data in Smart Cities. In particular, we differentiate between two levels of domains and stakeholders. Level 1 includes the domains needed to automate and support workflows and create tools to provide to stakeholders and their associated business processes in daily energy consumption and energy innovation (primarily the latter). Due to the magnitude of the Smart Cities domain a second level is defined; it involves contextual domains that complement the primary data identified at level 1 and provide further relevant insight. These lists of domains are not exhaustive; they will be enriched empirically using the ontologies we find.

2.1 Energy Using or Producing Products

The eeSemantics initiative[1] has defined the notion of EupP for representing any **E**nergy **u**sing or **p**roducing **P**roduct. From our perspective, this abstract notion can be seen as a black box from which energy-related information can be measured or extracted. From the "Domain" point of view an "EupP" can be considered either as:

- Any "Energy element" which represents any energy-related equipment.
- Any "Energy Zone" which is supposed to represent for instance zones of a building (simple rooms, areas in a building) but also houses, flats as a composition of EupPs or Electric vehicles. It could also represent a whole building as a composition of sub zones, rooms, corridors, flats, etc.
- Any "Energy Neighbourhood" which corresponds to the aggregation of "Energy Zones" of the energy network (e.g., buildings, streets lights, electrical vehicles)
- Any "Energy District" which corresponds to an aggregation of "Energy Neighbourhoods". The border between the notion of neighbourhood and the notion of district is given by considering that a district corresponds to an administrative denomination of an area (which comprises several buildings and neighbourhoods) and a neighbourhood is the aggregation of buildings that are in the same geographic area (for instance, in the same block or in the same residence).

The upper level can be the Smart City level composed of one EupP resulting from the aggregation of sub-EupPs.

In addition to this urban-based approach, Level 1 and Level 2 domains also have to be taken into account and related to the adapted level of EupP defined above.

2.2 Level 1 Domains

The main domains of WP3 deal with **assessing the demand for energy in Smart Cities.** Energy consumption is typically measured at one or more meters located between the wider grid and the city. These basic measurements of consumption cover all appliances and activities (including changes) in the city while allowing for assessment of energy efficiency and financial cost. They involve data about measuring energy consumption in different frameworks:

- Temporal (When / How often is energy usage measured? e.g., date, time, interval)
- Organisational (Who participates in measuring energy consumption? e.g., entity, body; which are its characteristics? e.g., legal identity, contracts, financial standing, stakeholdings, etc.)
- Statistical (Why is statistical analysis beneficial for assessing energy consumption? How and where is historical data stored? e.g., algorithms, statistical methods, baselines, control groups)

[1] See: **eeSemantics wiki.** Semantic Interoperability of Energy Efficiency ICT Tools for eeBuildings and beyond. https://webgate.ec.europa.eu/fpfis/wikis/display/eeSemantics/Home

- Spatial (Where in the supply chain does energy measurement take place?)
- Measurement (How is measurement performed? e.g., scales, metrics, units, classifications)

A particular attention should be paid to the notion of individual privacy; collecting energy measurements must be done in accordance with the corresponding local regulation (if any).

2.3 Level 2 Domains

There are some factors that influence measurement and need to be taken into account. These factors affect the consumption and therefore measurement, which reflects changes in consumption. Analysing and assessing such changed data affects the measurement results. E.g., there may be a huge spike in the measured data, this can be attributed to sudden changes in the weather conditions such as heating degree days, which depend on the climate of the observed site.

Therefore, the level 2 domains will cover information such as:

- Energy data (e.g., energy type, energy demand, energy offer)
- Climate zone (e.g., rainfall, sunshine hours)
- Weather data (e.g., outside temperature, wind speed)
- Environmental data (e.g., pollution)
- Building characteristics (e.g., insulation, spatial location, postal address, owner, manager)
- Occupancy (e.g., based on user's schedule, etc.)
- User behaviour and characteristics (e.g., practices for using devices)

3 Relevant Stakeholders

In this deliverable we have categorized the different abstract roles (i.e., stakeholdings) that a stakeholder plays in a concrete business case related to the provision of energy measurement and validation services. Stakeholders can be either legal entities (LE – i.e., natural persons or organisations), collections of LEs (consortiums), or quasi-legal entities (QLE).

A stakeholder has at least one stakeholding and is the social/legal location of the benefits and costs of a stakeholding. It will be also the case that the QLE and/or its agents *manage* the stakeholding, that is, act to avoid loss or to maximise net benefit. Furthermore, a stakeholder can have more than one stakeholding.

Examples of stakeholders and possible stakeholdings are the following:

- City or local **councils** often have multiple stakeholdings, e.g., as Public Energy User, Public Building Operator. Each of such stakeholdings may be managed by a department, subsidiary or contracted out to an enterprise.
- **Citizens** can be both Energy Consumers and Private Energy Providers (i.e., "prosumers").
- **Enterprises** include publicly-owned organisations and can also take on multiple stakeholdings, which may start with a policy or mission, move to public offers to contract and then be managed under contracts.
- A **social housing** cooperative in Germany is a LE with membership and personnel, usually with the stakeholdings of Private Energy User and Public Building Operator.
- European research **projects** are consortia (i.e., QLE) and have stakeholdings such as Energy Saving Intervention Trial Evaluator or Energy Saving Intervention Developer.

Similarly as in the case of the relevant domains, we define the following two levels of stakeholdings:

- **Level 1 stakeholdings** relate to everyday energy use; those stakeholdings in daily energy consumption where reverse revenue streams follow energy flows from sources to consumption in heating, electric devices, etc.
- **Level 2 stakeholdings** relate to energy innovation activities; those stakeholdings in energy measurement and validation related to innovation for energy saving (invention, research, development, trials) or to exploitation of new ideas, in decisions on deploying new services or approaches.

3.1 Level 1 Stakeholdings

- **Public Housing Provider**: provides residential rented, esp. social housing services with heating, cooling and access to electricity and water; invests in and maintains the residential energy and resource infrastructure (pipes, plant, networks).
- **Public Building Owner**: owns non-residential public buildings, buildings used by the public and buildings of public interest; ensures provision of heating, cooling and access to electricity and water. Invests in and maintains the residential energy and resource infrastructure (pipes, plant, networks).
- **Public Space Provider**: liable by contract or statute to provide safe and lighted public spaces such as roads and public squares.
- **Private energy user** (household, citizen, tenant, enterprise): pays for heating and electricity; is "end user" of electricity, gas, water or heat; and may wish to save energy and bills.
- **Tax payer**: pays taxes which cover local public building energy costs, public space energy costs, public subsidies, etc.
- **Energy Regulator and Policy-Maker**: provides subsidies to citizens and businesses; sets taxes, charges, and tariffs; plans energy measures and interventions. Typical LE types are cities, regional or central governments but also development agencies, NGOs, independent agencies, or specialised government agencies.

- **Private Energy Provider** (prosumer): is a private individual, investor in and owner of photo voltaic or other generator equipment, is not a Public Energy Provider.
- **Public Energy Provider**: generates electricity, contracts for gas supply, sells to Public Energy Provider or Energy Retailer.
- **Grid Provider**: maintains an electricity grid and balancing supply and demand, sells to Public Energy Provider or Energy Retailer.
- **Public Building Operator**, manages public energy consumption in single or multiple buildings, in the framework of energy provision contracts, including contract-based demand response management, e.g., by guiding and steering when electricity is being used or sold back to the smart grid.
- **Energy Service Company** (ESCO): contracts to deliver energy against an energy use and energy saving plan; contract execution requires an Energy Reference Use Estimator.
- **Energy Retailer**: contracts with consumers for electricity or heating, setting and publishing tariffs; contract execution requires an Energy Billing Service Provider. Also contracts with a Public Energy Provider, Private Energy Provider, Grid Provider etc. to meet demand. Is the owner of revenue flows in both directions.
- **Energy Monitoring Service Provider**: provides information about energy use behaviour and equipment performance to building tenants, building managers, ESCOs, etc. to enable actions to be taken to optimise energy use, to motivate and focus behavioural response by building users or managers. Contract execution requires an Energy Reference Use Estimator and Energy Measurement Service Provider.
- **Energy Billing Service Provider**: calculates retail bills and presents the invoices to consumers, building owners etc.; receives payments and manages accounts receivable. Revenue from contract with Energy Retailer. Contract execution requires a Measurement Service Provider. This stakeholding is often integrated with Energy Retailer.
- **Energy Measurement Service Provider**: installs measurement equipment and operates this to provide raw measures of consumption of energy and other resources to a contracting party, e.g., Energy Billing Service Provider, ESCO, Energy Monitoring Service Provider. The stakeholding may be combined with others including Energy Retailer, Public Housing Provider, etc.
- **IT Service Provider**: provides network, storage and computing services to Measurement Service Provider and Energy Billing Service Provider.

3.2 Level 2 Stakeholdings

- **Energy Saving Intervention Trial Evaluator**: designs trials, carries out pilot testing and delivers evidence relating to the effectiveness of interventions – new services, products, procedures – in reducing energy consumption vis a vis the counterfactual "without intervention" amount in buildings included in the trial (not directly applicable to other buildings). Requires: Energy Reference Use Estimator, Energy Measurement Service Provider.
- **Energy Saving Intervention Developer**: researches into and develops new services, products, procedures with the aim of reducing energy consumption in spaces or buildings or transport.
- **Energy Saving Intervention Decision-maker**: takes evidence from trials and decides whether an intervention – new services, products, procedures – will reduce energy consumption in specific buildings in the future (applies also to buildings in the trial, as it relates to future consumption). Contract execution requires: Energy Saving Intervention Trial Evaluator and Energy Savings Estimator.
- **Energy Savings Estimator**: takes evidence of savings from buildings in a trial and uses this in an energy use estimation model to estimate future energy savings in specified buildings over a future period. Delivered to Energy Saving Intervention Decision-makers.
- **Energy Reference Use Estimator**: takes relevant data on energy use, e.g., baseline same building, control group, and uses this in an energy use estimation model to estimate counterfactual energy

consumption in a specified building over a specified period, as fixed value or weighted probability distribution. Delivered to ESCOs, Energy Saving Intervention Trial Evaluators and Energy Monitoring Service Providers.

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 all along the project life.

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 (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, national or international 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 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 identifying 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. Alignments 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 into RDF and published as explained 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 Ontology Metadata Vocabulary[2] (OMV) [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[3] (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[4] [Whetzel et al., 2011] in the biomedical domain or OntoSelect[5] [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.

---

[2] http://omv2.sourceforge.net/

[3] http://lov.okfn.org/vocab/voaf

[44] http://bioportal.bioontology.org/

[5] 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 ontology	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

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	--
License	cc:license	cc:license	omv:hasLicense
URI	omv:URI	-- (uses the ontology URI)	omv:URI
Namespace	vann:preferredNamespaceUri	vann:preferredNamespaceUri	--
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[6] (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[7] [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)
Comments	Further information about the dataset in the context of our catalogue	String	rdfs:comment (xsd:string)

[6] http://www.w3.org/TR/vocab-dcat/

[7] http://vocab.deri.ie/void/

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 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[8] complemented by the W3C Provenance ontology[9]. The Provenance ontology will replace the previous use of OMV with 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)
Level	The type of language used for expressing correspondences	String	align:level (xsd:string)

[8] http://alignapi.gforge.inria.fr/labels.html

[9] http://www.w3.org/TR/prov-o/

Type	The arity 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 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[10]; 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.

[10] 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[11]. 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[12] 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 of "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:

[11] http://www.w3.org/DesignIssues/LinkedData.html

[12] 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![13] (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 this 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[14])

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.
- **Whether the alignment is available under an open license.** This indicator is related to LOD1 and ODI8.

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

[14] http://databugger.aksw.org/

- **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[15] or Alcomo[16]).

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.

[15] http://www.cs.ox.ac.uk/isg/tools/LogMap/

[16] http://web.informatik.uni-mannheim.de/alcomo/

7 Publication Methods

In READY4SmartCities we will publish catalogues including the collected information about ontologies, datasets and alignments by means of a web page where the catalogues will be displayed and by means of machine-processable data using the 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. Preliminary prototypes of the web pages of the ontology, dataset and alignment catalogues are shown in Figure 2, Figure 3 and Figure 4, respectively.

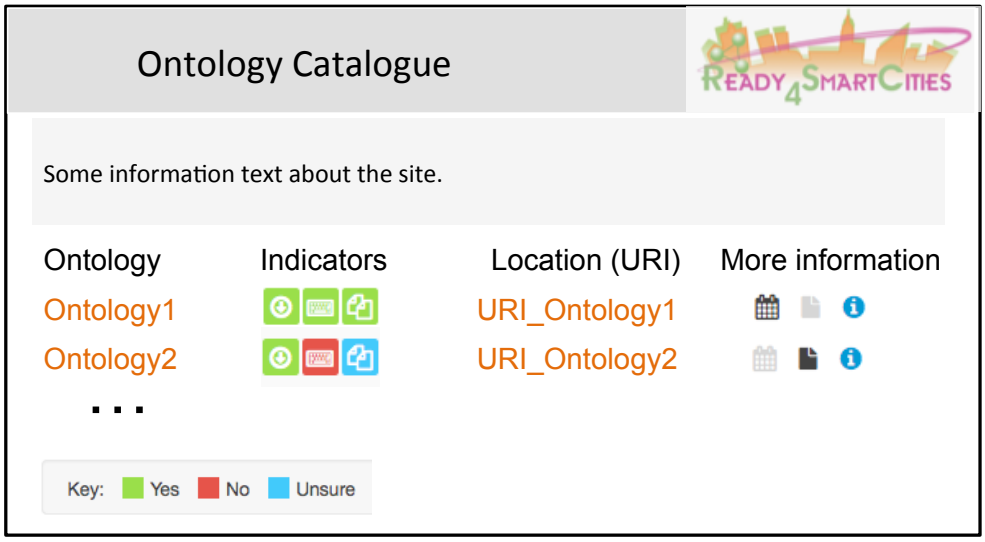

Figure 2. Ontology catalogue web interface prototype

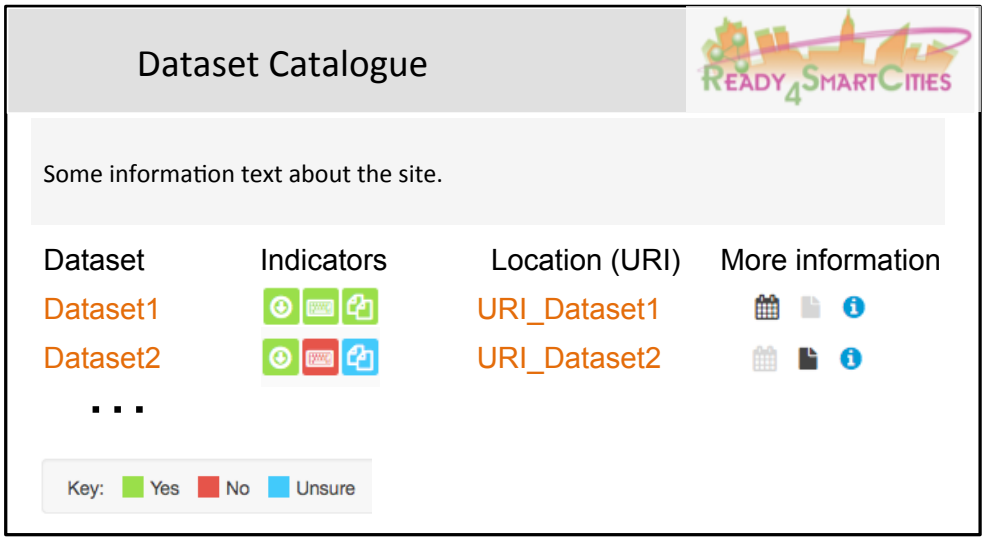

Figure 3. Dataset catalogue web interface prototype

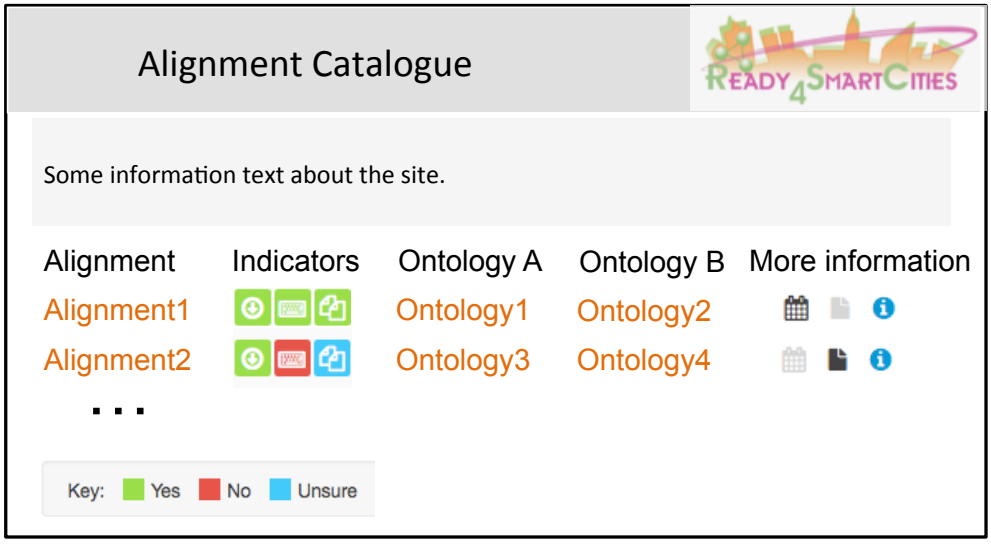

Figure 4. Alignment catalogue web interface prototype

As can be seen, the result representation provided at Open Data Index[17] 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"[18].

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.

Furthermore, even if we plan to maintain the catalogues alive after the submission of the corresponding project deliverable, a snapshot of them will be published in the ValMet wiki.

[17] https://index.okfn.org/

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

8 Conclusions

This document sets the ground for the work to be performed in the next tasks of work package 3. On the one hand, the concrete scope of the work package has been defined by identifying the relevant domains and stakeholders related to energy measurement and validation. On the other hand, a concrete set of methods for collecting, identifying, assessing and publishing the catalogues of ontologies, datasets and alignments has been defined.

We plan to follow a common strategy in work packages 2 and 3, even if the relevant domains and stakeholders differ between 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 combined information 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 will 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). Due to feedback from other work packages, these lists of stakeholders and domains are expected to evolve during the 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 (related to, e.g., harmonisation of ontologies, availability and privacy of energy data, etc.).

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. Trojan dos Santos. *The Alignment Api 4.0.* Semantic Web Journal, vol. 2, no. 1, pp. 3-10, 2011.

[Euzenat and Shvaiko, 2013] J. Euzenat and 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, 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.

[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.