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

Report from the VoCamp - Energy measurement data in municipalities

Vienna, Austria (April 22nd-23rd, 2015)

Organization: Universidad Politécnica de Madrid
Local and support organization: Austrian Institute of Technology and D'Appolonia

1 Introduction

This report summarizes the work carried out, presentations and discussions held during the “VoCamp in energy measurement data in municipalities”[1] which took place at the Austrian Institute of Technology - AIT (Energy Department) in Vienna, from 22nd to 23rd of April, 2015. This event was organized and supported by the READY4SmartCities project[2], as part of the planned training route[3].

A VoCamp is an informal event where participants 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 them for publishing data on the Web.

This VoCamp was focused on how municipalities can represent their data about energy measurement in order to publish it online (e.g., as open data). The interest to this question arises from the expected benefits such as the ability to easily reuse these data by third parties or to link them to other relevant data for further processing (e.g., building information models, climate, occupancy). The VoCamp aimed at obtaining a common ontology that can be used by municipalities to represent their energy measurement data in order to publish such data online as Linked Data. In order to obtain such vocabulary the VoCamp was organized in a series of practical session as shown in Figure 1, where the output from some session represents the input for the following activity. In summary, starting from several datasets brought by the participants, the ontological requirements were extracted; then, the vocabularies were modelled according to such requirements; after that, the different models were put in common in order to agree on one common vocabulary which was translated into nine languages during the last session.

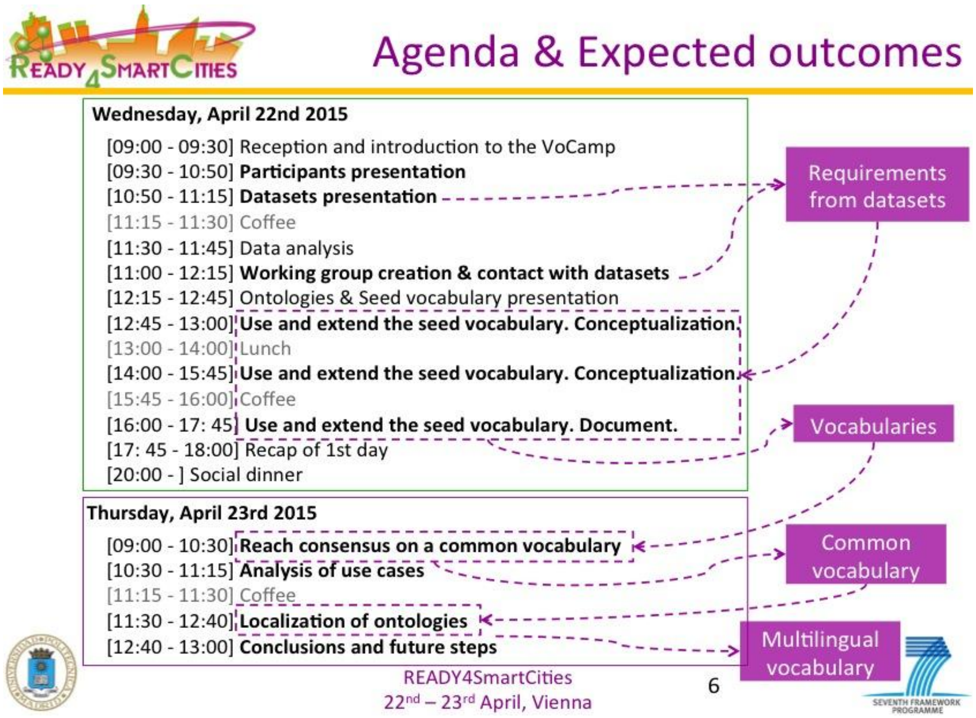


Figure 1. Agenda and expected outcomes of the VoCamp

[1] http://smartcity.linkeddata.es/LD4SC/VoCamp/
[2] http://www.ready4smartcities.eu/
[3] http://www.ready4smartcities.eu/training

The event was organized in a sequential number of activities following the basic steps to build a common multilingual vocabulary for measurement data in municipalities. That is, from requirement extraction to ontology localization, including ontology modelling and conceptualization. There were 20 attendees coming from the following institutions and countries as shown in Table 1.

Table 1 Participants list

Nr.	Name	Surname	Organization	Country
1	Marko	Batic	Institute Mihajlo Pupin	Serbia
2	Andrea	Cavallaro	D'Appolonia	Italy
3	João	Encarnação	Energaia - Energy Agency	Portugal
4	Raúl	García-Castro	UPM - Ontology Engineering Group	Spain
5	Asunción	Gómez Pérez	UPM - Ontology Engineering Group	Spain
6	Johannes	Gruber	Ingenieurbüro conGRUent	Austria
7	Dimosthenis	Ioannidis	CERTH/ITI	Greece
8	Florian	Nadler	AIT	Austria
9	Anna	Osello	Politecnico di Torino	Italy
10	Pieter	Pauwels	Ghent University	Belgium
11	Jan	Peters-Anders	AIT	Austria
12	Stenzel	Pit	Fraunhofer IIS/EAS	Germany
13	María	Poveda-Villalón	UPM - Ontology Engineering Group	Spain
14	Filip	Radulovic	UPM - Ontology Engineering Group	Spain
15	Simon	Robinson	empirica	Germany
16	Mari	Sepponen	VTT	Finland
17	Graeme	Stuart	Institute of Energy and Sustainable Development, De Montfort University	UK
18	Nikola	Tomasevic	Institute Mihajlo Pupin	Serbia
19	Mikko	Tuomisto	VTT	Finland
20	Stefan	Vielguth	AIT	Austria

The following sections summarize and describe the sessions on dataset presentation and analysis, ontology conceptualization, consensus on a common model, use cases analysis and ontology localization. Finally, some concluding remarks and future steps are shown.

2 Dataset presentation

Participants to the VoCamp brought and presented the following 6 datasets:

- **Dataset 1:** The smartspaces (http://smartspaces.dmu.ac.uk/) dataset contains electricity, gas, and water consumption measurements about public buildings in Leicester (See Figure 2).

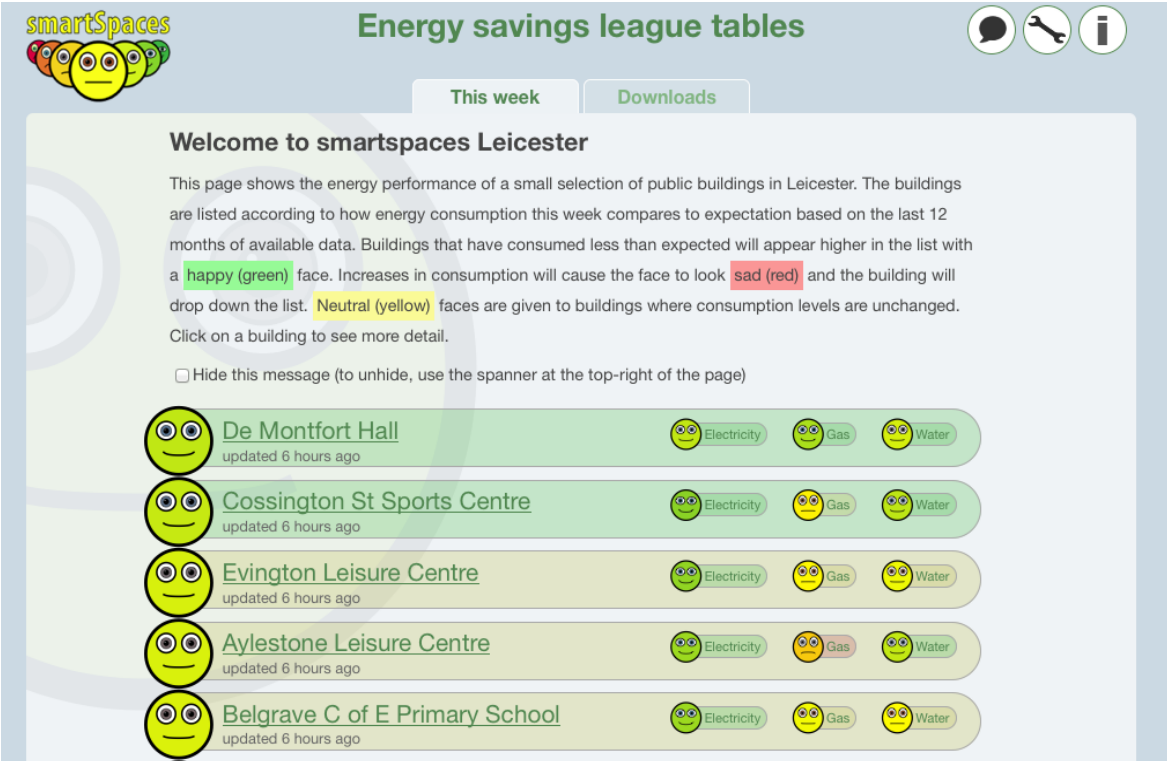


Figure 2 smartspaces dataset screenshot

- **Dataset 2:** The energaia observatory (http://observatorio.energaia.pt) provides data visualizations about consumption, emission and population from the metropolitan area of Porto (See Figure 3).

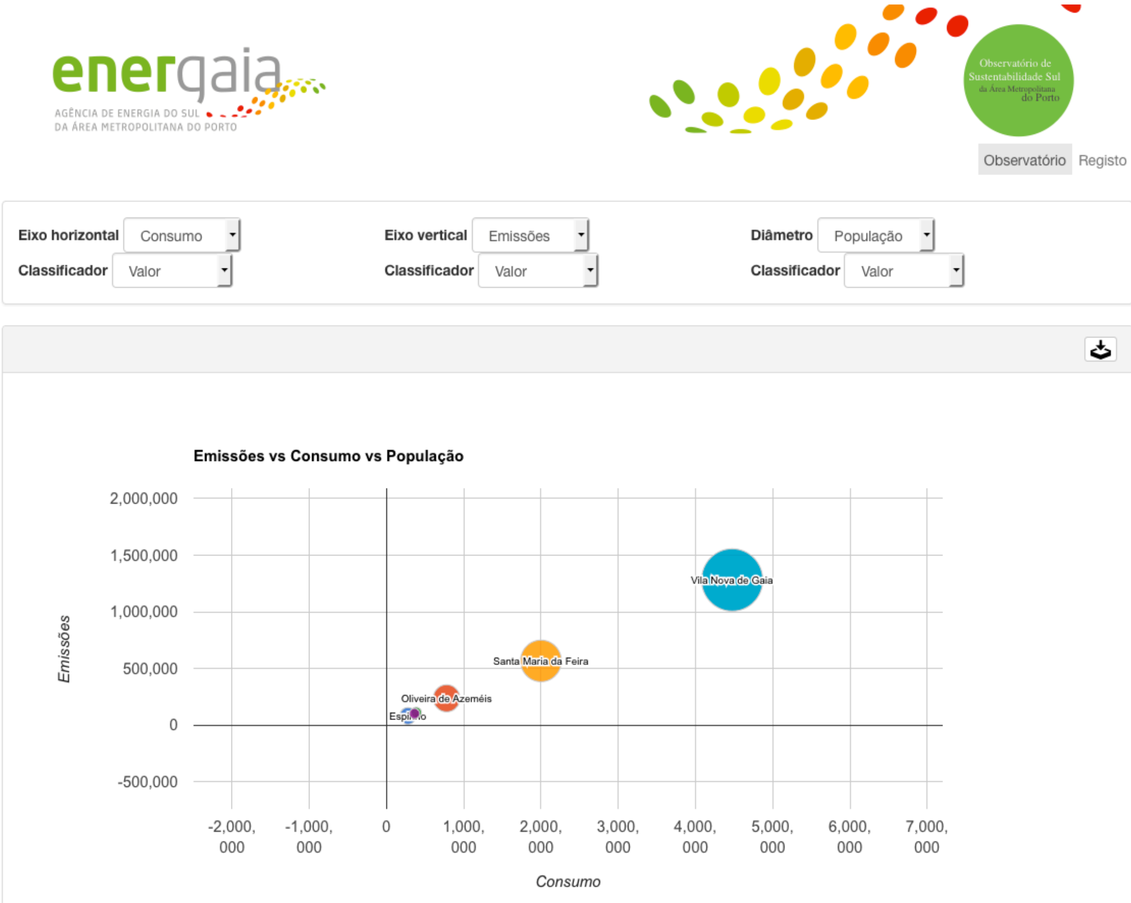


Figure 3 energaia dataset screenshot

- **Dataset 3: D**ataset about Sim models and building performance carried out before the building is built. This data is the basis for simulation before construction and to compare how a building is performing (after construction) versus how it is supposed to be performing. Figure 4 and Figure 5 show the data presented and a slide from the presentation respectively.

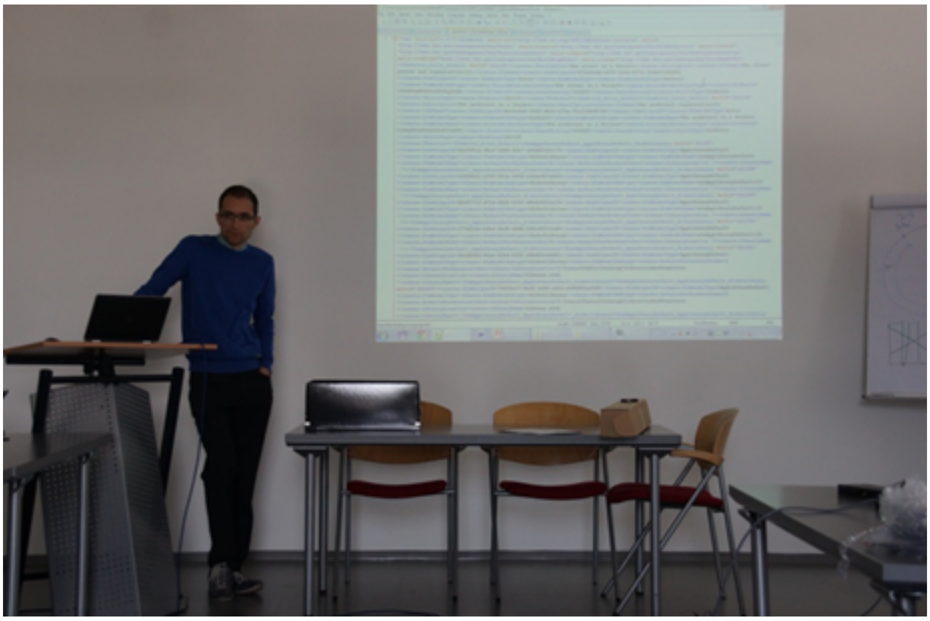


Figure 4 Pieter Pauwels presenting the SimModel dataset

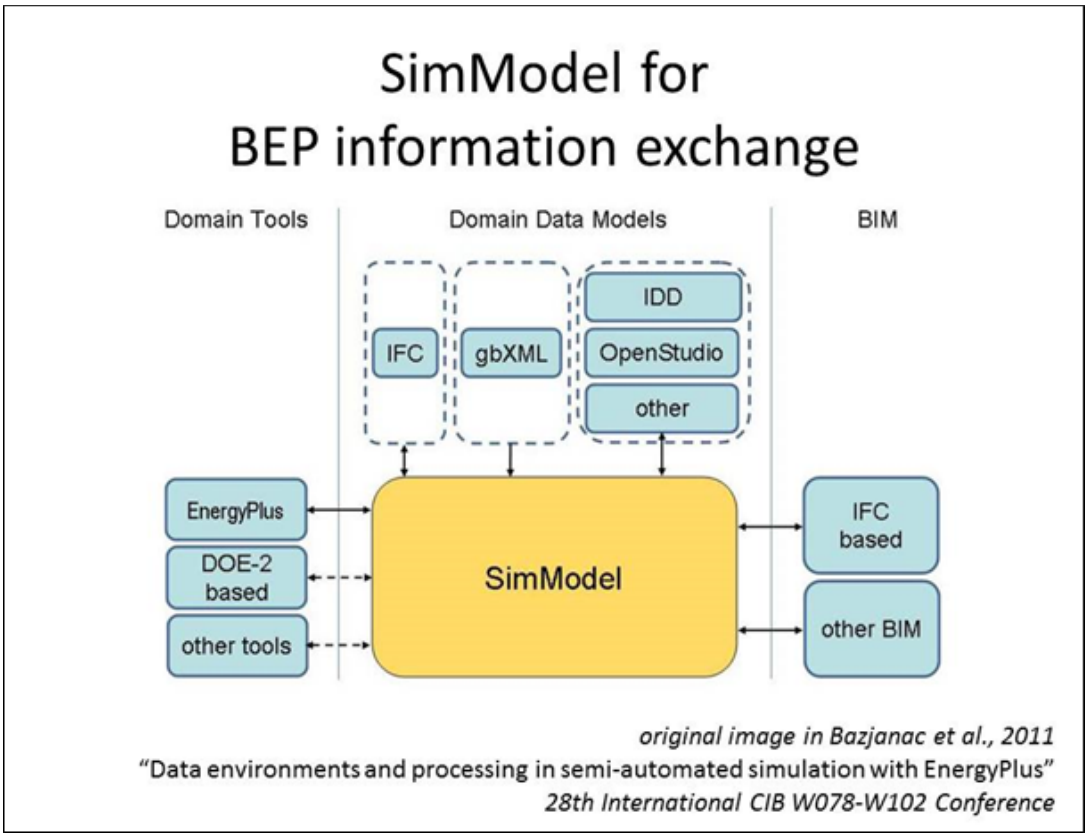


Figure 5 SimModel for BEP information exchange

- **Dataset 4:** Dataset about Institute Mihajlo Pupin campus in Belgrade, Serbia. This dataset contains measurements from energy production, storage and use for electricity and heat in intervals of 1 minute and 1 second, depending on the measurement. This dataset is used for energy optimization and energy flow control (See Figure 6).

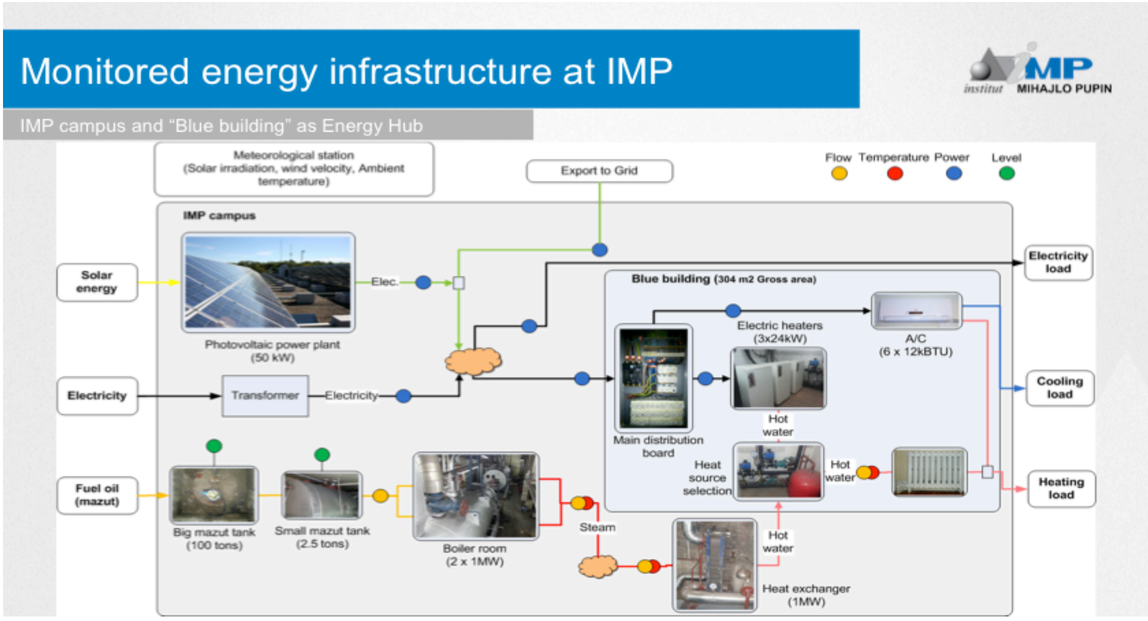


Figure 6 IMP campus data flow

- **Dataset 5:** Dataset about the Distribution of photovoltaic collectors in Vienna gathering measurements about energy production entities and photovoltaic collectors in a particular geolocation (See Figure 7).

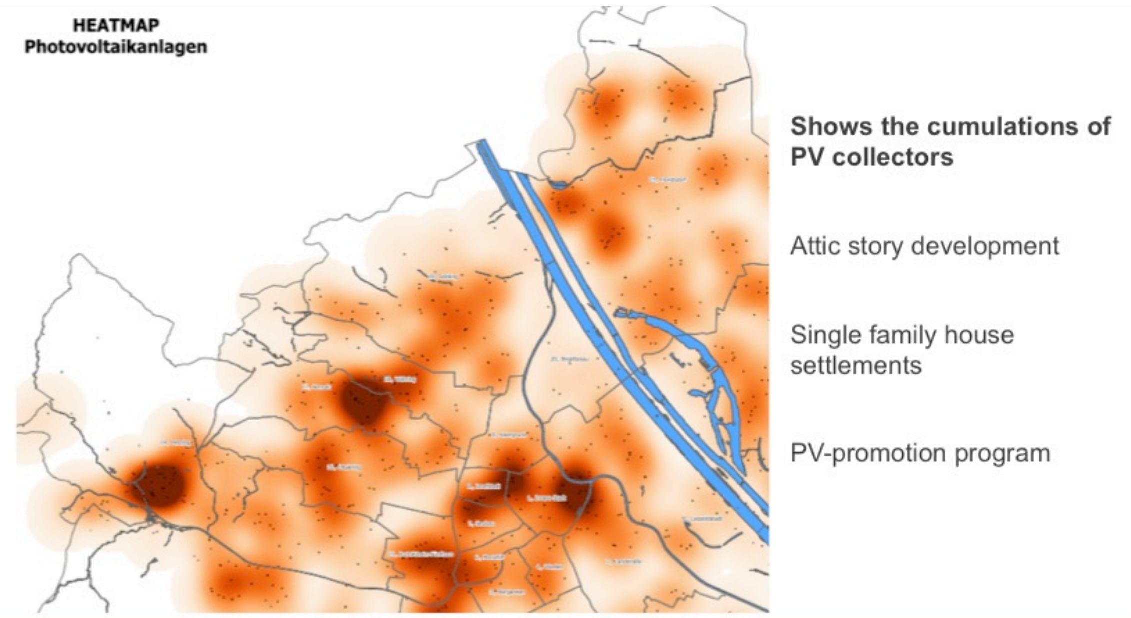
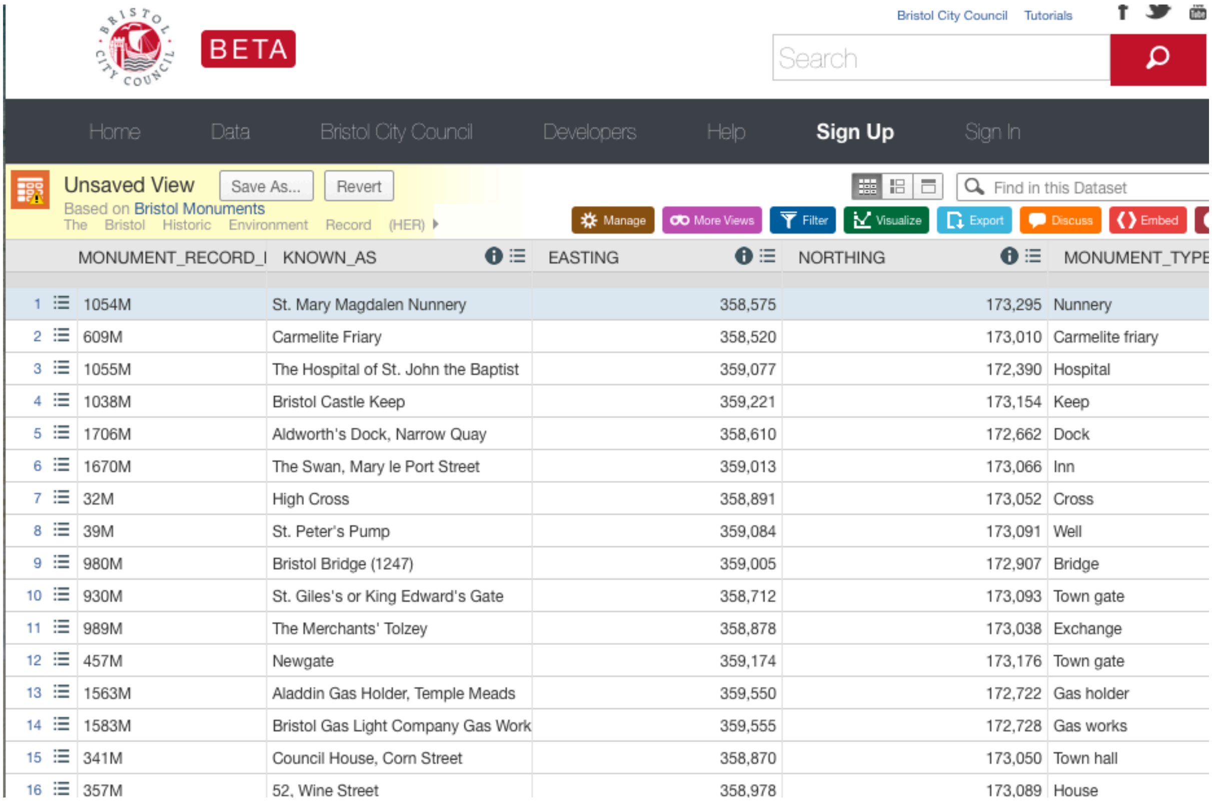


Figure 7 Distribution of photovoltaic collectors in Vienna

- **Dataset 6:** Dataset about Bristol public data (https://opendata.bristol.gov.uk/category/Energy). The focus was in the solar panels data about the installed solar photovoltaic capacity in Bristol each month and in energy consumption (See Figure 8).



	MONUMENT_RECORD_I	KNOWN_AS	EASTING	NORTHING	MONUMENT_TYPE
1	1054M	St. Mary Magdalen Nunnery	358,575	173,295	Nunnery
2	609M	Carmelite Friary	358,520	173,010	Carmelite friary
3	1055M	The Hospital of St. John the Baptist	359,077	172,390	Hospital
4	1038M	Bristol Castle Keep	359,221	173,154	Keep
5	1706M	Aldworth's Dock, Narrow Quay	358,610	172,662	Dock
6	1670M	The Swan, Mary le Port Street	359,013	173,066	Inn
7	32M	High Cross	358,891	173,052	Cross
8	39M	St. Peter's Pump	359,084	173,091	Well
9	980M	Bristol Bridge (1247)	359,005	172,907	Bridge
10	930M	St. Giles's or King Edward's Gate	358,712	173,093	Town gate
11	989M	The Merchants' Tolzey	358,878	173,038	Exchange
12	457M	Newgate	359,174	173,176	Town gate
13	1563M	Aladdin Gas Holder, Temple Meads	359,550	172,722	Gas holder
14	1583M	Bristol Gas Light Company Gas Work	359,555	172,728	Gas works
15	341M	Council House, Corn Street	358,870	173,050	Town hall
16	357M	52, Wine Street	358,978	173,089	House

Figure 8 Bristol energy consumption in buildings

3 Dataset analysis

During this session the VoCamp organizers provided some notions about how to analyze the data in order to extract main terms and relations from the datasets (See Figure 9). We also discussed the need for a common wording in the measurement datasets, as well as the need for multilingualism since datasets will be also used by local technicians, e.g., which might need the data in their local language.

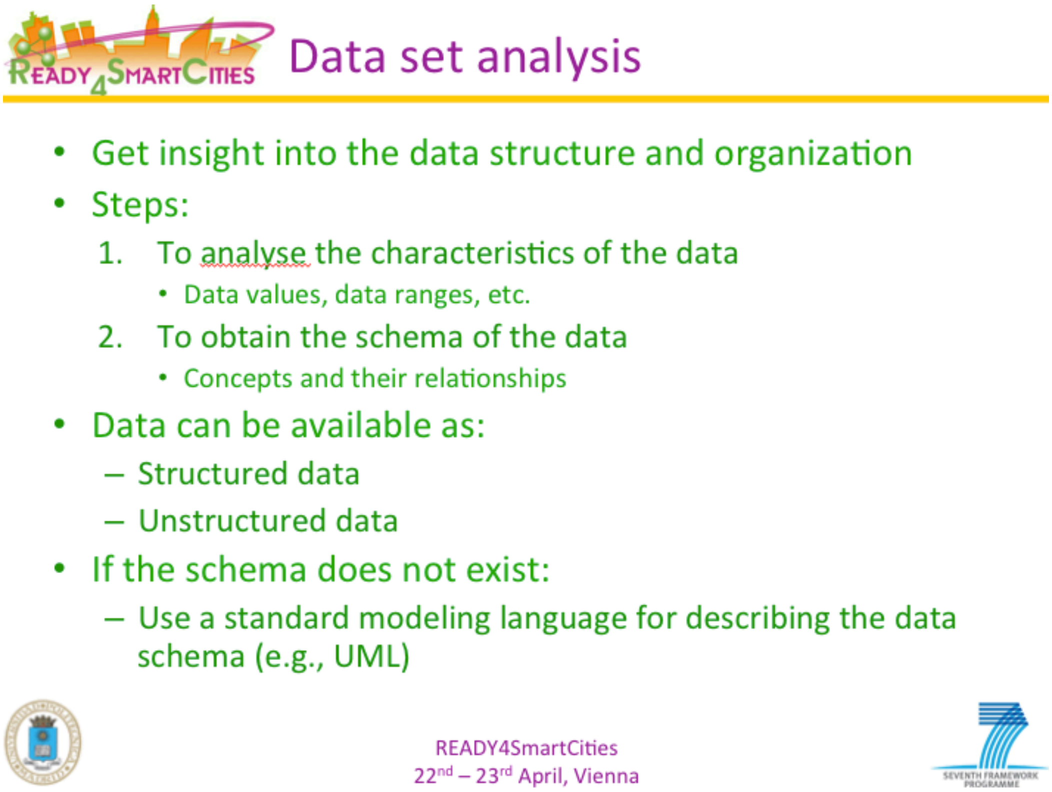


Figure 9 Brief introduction to data set analysis

In order to move to the practical exercises, participants were divided into three groups, as shown in Figure 10, where each group had the following two datasets assigned:

- Group 1: dataset 4 (Institute Mihajlo Pupin campus) and dataset 6 (Bristol public data)
- Group 2: dataset 2 (energaia) and dataset 5 (photovoltaic collectors)
- Group 3: dataset 1 (smartspaces) and dataset 3 (Sim models)

Figure 10 Discussion groups

The goal of this activity was to become familiar with the datasets and to start eliciting requirements of the vocabulary that the participants would need to develop.

This grouping was maintained during the dataset analysis and conceptualization activities so that each group carried out the exercises over the assigned datasets.

4 Conceptualization

Along this section attendees were provided with some notions and tips about (a) extracting ontology requirements, (b) addressing the conceptualization and (c) reusing of ontological resources.

First, Asunción Gómez Pérez explained fundamental concepts of semantic vocabulary creation via an example of the electricity consumption at the Prado museum (See Figure 11).

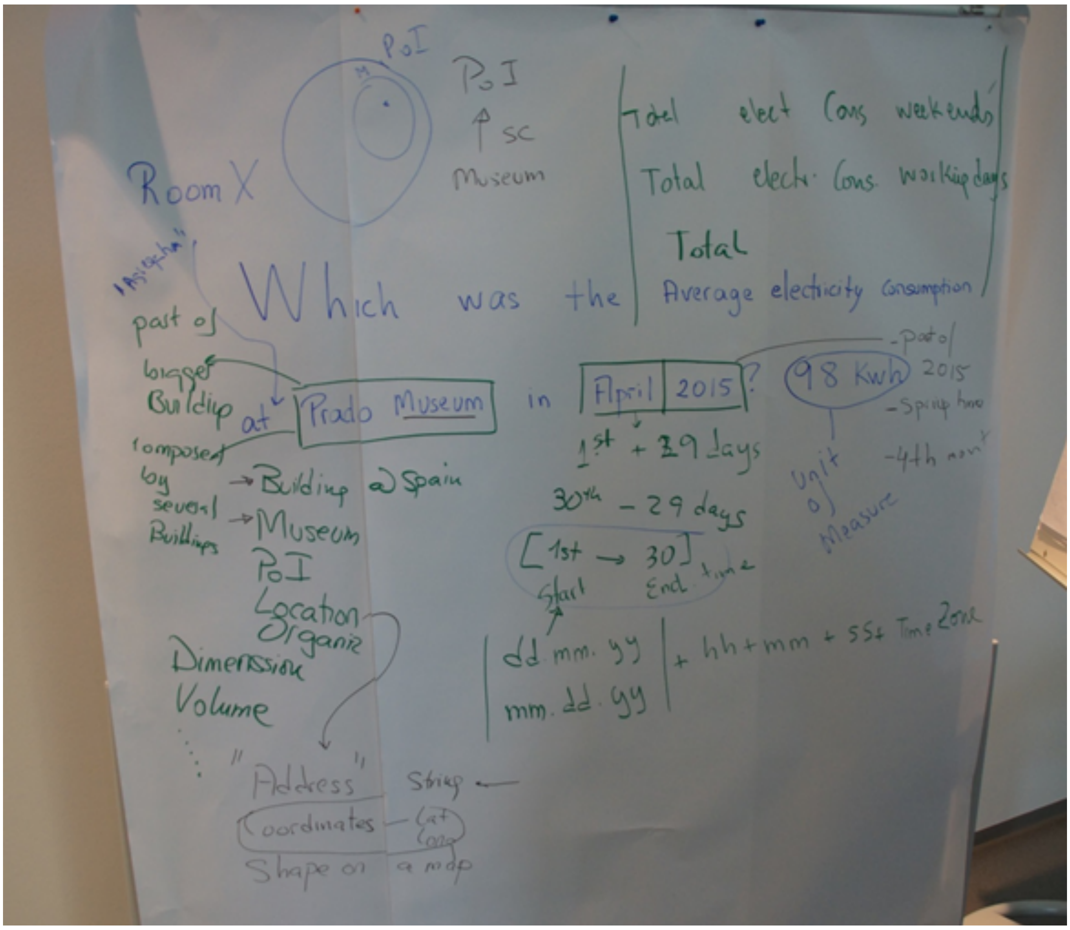


Figure 11 Example of the Prado museum

Next, the participants were ready to prepare their ontological requirements for their datasets in the form of Competency Questions. The 18 competency questions shown in Table 2 were developed by the participants.

Then the participants were taught how to use a Seed vocabulary (See Figure 12) in order to extend it to create the vocabulary to cover the needs of each group, according to the competency questions defined by them. This vocabulary, formed mainly by reused terms from existing vocabularies, was proposed by organizers in order to ease the conceptualization activity. The reuse of existing conceptualizations and ontologies is a good practice in knowledge representation.

Table 2. Competency Questions list

	Competency Question	Group
CQ1	Which building in Bristol consumed the most electricity in October 2014?	1
CQ2	Which is the building with the highest electrical consumption in last year?	1
CQ3	Which building in Bristol used much more electricity than expected in simulation model?	1
CQ4	What is the total area of solar panels that should be installed on all public buildings in Bristol?	1
CQ5	The 10% of buildings that have used more electricity than expected in April 2013?	1
CQ6	What is the consumption/cost (total or average) of a specific client (or clients in a district)?	2
CQ7	What is most productive PV plant in the City	2
CQ8	What is the best district in PV power production?	2
CQ9	What was the most energy consuming building in the district	2
CQ10	Which buildings are producing more than spending in the district?	2
CQ11	How does the promotion program for PV plants work out?	2
CQ12	What is the half hourly consumption from building x between date1 and date2 (is it to be cleaned? is it to be interpolated? using what method?)	3
CQ13	From which sources does the building produce/consume its energy and how much?	3
CQ14	How much energy WILL BE consumed/produced next week at building X - based on historic data/simulation	3
CQ15	Which consumption profiles are off/on the expected observation values	3
CQ16	How is my building doing? - compared to ...	3
CQ17	In my set of buildings (e.g. in Vienna) which ones should I look at?	3
CQ18	How much money has been saved due to my energy efficiency investment?	3

Figure 12 Filip Radulovic presenting the Seed vocabulary

The afternoon was filled with creating a vocabulary for the datasets that the participants brought with them. They had a look at the meaning of their attributes and the interrelationships of their entities. It became obvious that this exercise was made more complicated by the domains that were involved in the production and use of the data (and it would have been easier to analyse the data and to produce a vocabulary without knowing the purpose of the data).

Finally, each group developed a starting model to describe the data contained in the assigned data sets (See Figure 13, Figure 14 and Figure 15).

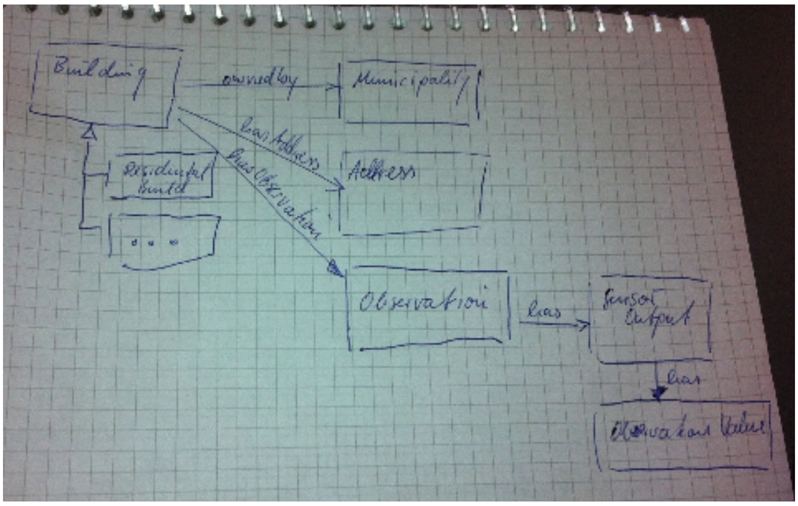


Figure 13 Partial models from group 1

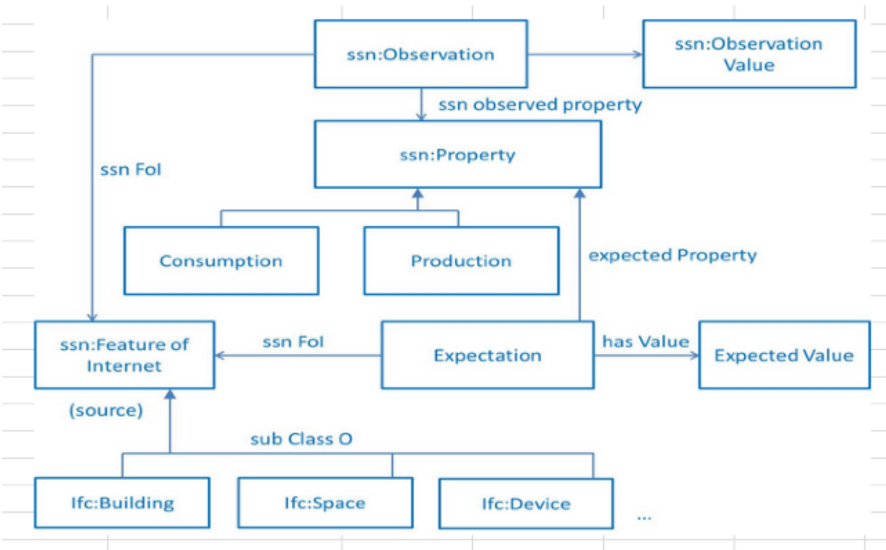


Figure 14 Partial models from group 2

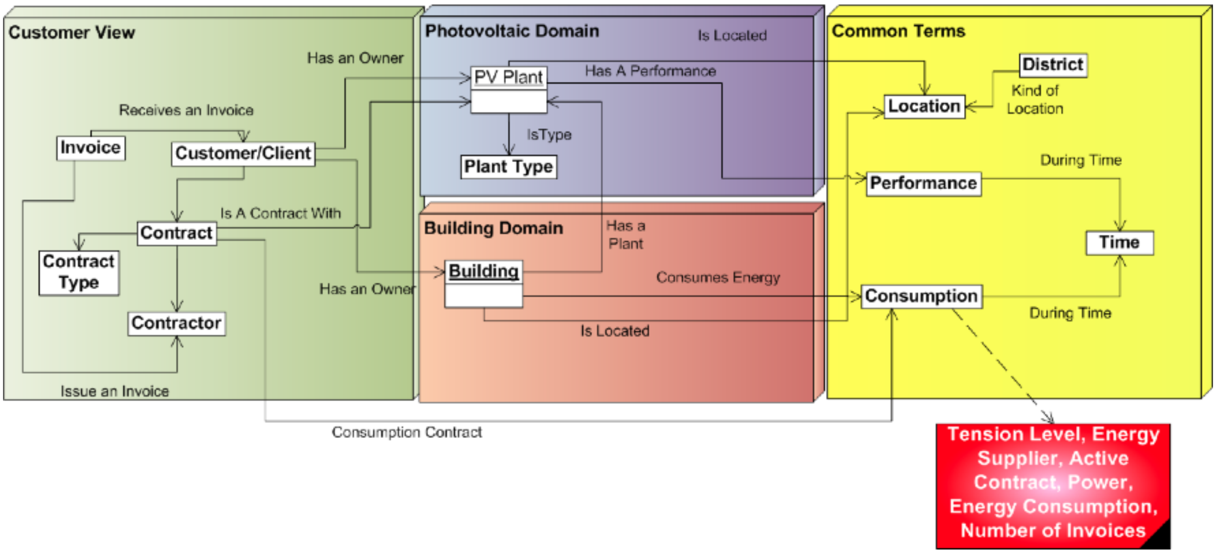


Figure 15 Partial models from group 3

5 Reaching consensus

In the morning of the second day, Raúl García-Castro started developing a more in depth vocabulary with the help of the participants' work of day 1 and through questions to the auditorium. The aim of this session was to generate a common vocabulary to describe the data assigned to every group (See Figure 13, Figure 14 and Figure 15).

Figure 16 Raúl García-Castro developing a common vocabulary with the participants

He also discussed the Semantic Sensor Network Ontology[4] (ssn), which describes sensors, observations, and related concepts (See Figure 16 and Figure 17), that appear in the datasets analyzed by participants in previous sessions.

[4] http://www.w3.org/2005/Incubator/ssn/ssnx/ssn

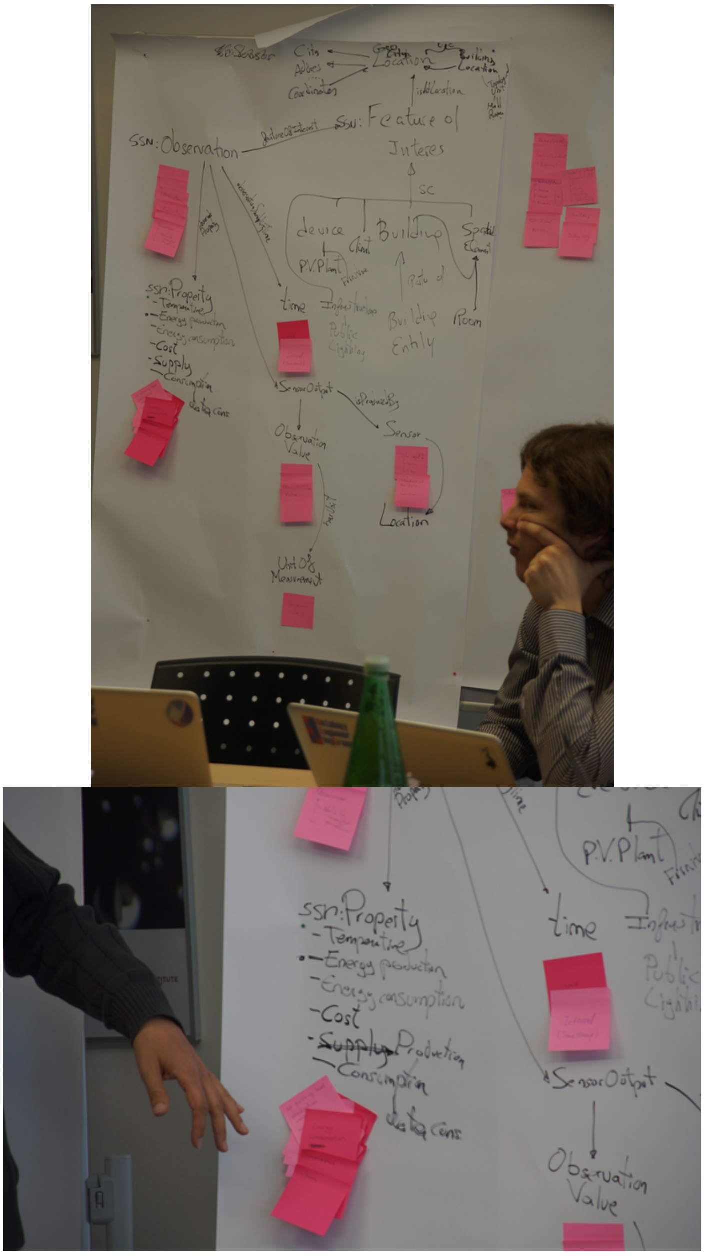


Figure 17 Description of the common vocabulary

As output of this session a common model were proposed listing in a spreadsheet the concepts and needed relationships between them. This terminology represented the basis for the ontology implementation and the ontology localization activities(See Section 7).

6 Use case analysis

During the use cases session the focus was in the uses cases related to the datasets that participants had presented. However, other uses cases arose. As a summary the following use cases were mentioned:

- Energy data publication and visualization
- Trust of data and benefits from provenance information
- Building performance simulation
- Performance comparison between buildings
- Performance optimization
- Real time energy monitoring
- Energy flow control
- Assessment of energy saving/performance innovation and proof of the energy saving
- Energy contracting process

Some scenarios were also drawn where several use cases could take place sequentially. These scenarios are:

- Energy consumption monitoring → data publishing → performance calculation or simulation → performance comparison → performance optimization
- Building (or other city energy consumption system as street lighting, water, gas, etc.) model → simulation → performance estimation → performance optimization

7 Ontology localization

The ontology localization activity refers to the adaptation of an ontology to a particular language and culture. During the localization session main concepts, terms and properties appearing in the model were translated from English into nine different languages, namely: German (@de), Greek (@el), Spanish (@es), Finnish (@fi), French (@fr), Italian (@it), Portuguese (@pt), Dutch (@nl), and Serbian (@sr). These languages represent all the native and secondary languages spoken by attendees. Figure 18 presents an excerpt of the table used to collect labels in the different languages.

Ontology term @en	@de	@el	@es	@fi	@fr	@it	@pt	@nl	@sr
address	Adresse	Διεύθυνση	dirección	osoite	adresse	Indirizzo	morada	adres	adresa
building	Gebäude	Κτίριο	edificio	rakennus	bâtiment	Edificio	Edifício	gebouw	zgrada
building entity	Gebäudeeinheit	Οντότητα Κτίριο	elemento del edificio	rakennuskokonais	entité de bâtiment	elementi propri dell'edificio	Entidade do edifício	gebouwentiteit	segment zgrade
building location	Gebäudeposition		localización del edificio	rakennuksen sijainti	localisation du bâtiment	posizione dell'edificio	Localização do edifício	gebouwlocatie	lokacija zgrade
calibration	Kalibrierung	Βαθμονόμη	calibración	kalibrointi	calibration	Calibrazione	calibração	kalibratie	kalibracija
city	Stadt	Πόλη	ciudad	kaupunki	ville	Città	cidade	stad	grad
client	Kunde	Πελάτης	cliente	asiakas	client	Cliente	cliente	klant	klijent
client, customer	Abnehmer	Τελικός Χρήστης/Πελάτης	cliente	asiakas, kuluttaja	client	Clinte, consumatore	cliente	klant	korisnik
consumer	Verbraucher		consumidor	kuluttaja	consommateur	consumatore	consumidor	verbruiker	
consumption	Verbrauch	Κατανάλωσ	consumo	kulutus	consommation	consumo	consumo	verbruik	potrosnja
coordinates	Koordinaten		coordenadas	koordinaatti	coordonnée	Coordinate	coordenadas	coördinaten	koordinate
cost	Kosten	Κόστος	coste	kustannus	coût	costo	custo	kostprijs	trosak
demand	Bedarf	Ζήτηση	demanda	tarve	demande	domanda	procura	vraag	potraznja
device	Gerät	Συσκευή	dispositivo	laite	dispositif	dispositivo	dispositivo	toestel	uredjaj
energy consumption	Energieverbrauc	Ενεργειακή Κατανάλωση	consumo enegético	energiankulutus	consommation d'énergie	energia consumata	consumo de energia	energieverbruil	energetska potrosnja

Figure 18 Excerpt of the ontology terms in 10 different languages

During the VoCamp only the labels or tags were attached to the identified terms of the ontology, that is, no descriptions in natural language were assigned. Therefore, the localization activity covered the labels identified during the conceptualization activity. As future work it would be advisable doing the same for the natural language description defining the meaning of such terms.

8 Conclusions and further steps

The main product obtained from this VoCamp is the multilingual ontology available at http://smartcity.linkeddata.es/LD4SC/def/ontology#.

Finally, after a round of feedback from the attendees the following open issues and suggestion for further events were proposed:

- Further development of use cases and scenarios
- Focus on cities in a broader view:
  - having a deeper look at the energy demand
  - not focusing on buildings only
  - district level
  - traffic data
  - territorial aspects
- Need to add saving results
- Need for modelling experimentation
- Need for modelling series of observations
- Simulations requires a complete different set of data
- More definition of terms, what does every concept means and how to use it
- Another VoCamp including city partners
- Define specific use cases, extend the use cases with other concepts defined by the industry as flexibility, aggregators, etc.
- Interest in public building comparison
- Combine observation with simulation data for the planning
- Start from the model already created and use it to transform data, get linked data from the process
- Mapping among ontologies
- Develop applications from the actual data and the ontology
- Focus more in the ontology reuse activity, look for existing ontologies, see what could be reused, what is needed, etc.

Figure 19 UPM VoCamp team
(From left to right: Filip Radulovic, María Poveda-Villalón, Asunción Gómez-Pérez, Raúl García-Castro)