Short Paper: Semantic Annotations for Sensor Open Data

Mikel Emaldi, Jon Lázaro, Unai Aguilera, Oscar Peña, and Diego López de Ipiña

Deusto Institute of Technology - DeustoTech, University of Deusto
Avda. Universidades 24, 48007, Bilbao, Spain
{m.emaldi, jlazaro, unai.aguilera, oscar.pena, dipina}@deusto.es

Abstract. Since the creation of the Open Data Euskadi (ODE) initiative in 2009, one of its challenges has been the publication of government Open Data following the Linked Data principles. On the other hand, one of the challenges for the Semantic Sensor Web is the integration and fusion of data from heterogeneous sensor networks. In this short paper we present the efforts made at the Bizkaisense[1] project on the alignment of different ontologies with the objective to fulfil these two challenges.

1 Introduction

Many governments across the world have realised the importance of opening data both as a service to promote transparency and as a way to enable businesses to make a better use of publicly available information. Open Data Euskadi[2] is a good example of it, in this case fostered by the Basque Country Government in Spain. Despite the efforts made by this initiative towards Linked Data, they have been mainly focused on the publication of raw data, directly taken from the computer systems from different administrations. The fact is that there is a limited number of datasets published as Linked Data or, at least, in any of the RDF serializations; but the good news is that external providers such as research centres or companies can treat raw data and publish them as Open Linked Data.

One of the datasets requiring this treatment is the one containing the data generated by the pollution sensors deployed throughout Basque Country[3]. According to [4] there are five challenges for the Semantic Sensor Web: 1) the first is about the abstraction level of the data extraction, process and management; 2) quality and Quality of Service of sensor data; 3) integration and fusion of sensor data; 4) identification and location of relevant sensor-based data sources; and 5) rapid development of applications. The Bizkaisense project is focused on the accomplishment of the third challenge. We think that the first step to integrate

[1] This research is founded by Bizkailab 2010 program, Bizkaiko Foru Aldundia - Diputación Foral de Bizkaia.

[2] http://opendata.euskadi.net/

[3] http://www.ingurumena.ejgv.euskadi.net/r49-n82/es/vima_ai_vigilancia/indice.apl?lenguaje=c

this Sensor Open Data with heterogeneous data sources is to publish them as Linked Data. In this short paper we present the efforts made at this first step, which are related to the mapping of the raw data to appropriate ontologies and the alignment between them.

The remainder of the paper is organized as follows. Section 2 discusses related work. Section 3 exposes the Open Data available about pollution sensors. Section 4 exposes the adopted solution to map the raw data to appropriate ontologies. Finally, Section 5 concludes and outlines the future work.

2 Related Work

Several efforts have been done in the field of semantic sensor networks; in this section we expose some examples. In [2], the SSN and SWEET ontologies are used to model sensor data and to allow a federated query system among them. However, these approaches do not use any ontology to represent the units of the measurements made by sensors. [3] presents a survey about the different ontologies to model different aspects of sensor networks. Linked Stream Middleware (LSM) provides wrappers for real time data collecting and publishing, a web interface to visualize and publish data and a SPARQL endpoint for querying data from heterogeneous sensor networks in an unified way [7]. In LSM, the user can import different ontologies to represent her/his sensors. This approach allows to annotate a wide variety of sensors, however, it can be an obstacle to manage the interoperability among different sensor networks. Related with the usage of custom ontologies, [1] uses its own ontology to represent the location of the sensors with a high granularity (floor, room, etc.). In [5], they map the sensors of Android powered smartphones to SSN and DUL ontologies, extending SSN with the proper instances to represent these sensors. The approach presented by [10] extends SSN with a collection of observations that their sensor network observes.

The AEMET Linked Data[4] project has a strong relationship with our Bizkaisense project. In this project the weather stations of AEMET (Agencia Estatal de Meteorología) have been annotated semantically. The data related to these stations have been extracted from CSV files provided by AEMET[5]. To annotate these weather stations, they have combined SSN ontology with *aemetonto*, a custom ontology made for the project which is used to represent the different meteorological phenomena that the stations can measure.

Summarizing, we can see that although there are many projects related to semantic sensor networks, usually, the way to fulfil the limitations of the ontologies is to introduce new custom ontologies, instead of reingenieering existing ontological resources. Similar reingeneering work can be seen at [5] or [10]. We think that this second approach, which is adopted by Bizkaisense, is more suitable to achieve the interoperability among heterogeneous sensor networks.

[4] `http://aemet.linkeddata.es/`

[5] `ftp://ftpdatos.aemet.es/datos_observacion/observaciones_diezminutales/`

3 Pollution Sensoring in Basque Country

There are 72 pollution sensors set up along the Basque Country, managed by the Basque Government. These sensors measure the air quality based on: chemical substances (xylene, sulfur dioxide, toluene, carbon monoxide, ozone, particulate matter (10 and 2,5 $\mu g/m^3$), nitrogen dioxide, hydrocarbons, hydrogen sulfide, ammonia, ethylbenzene, volatile organic compounds, benzene and smoke) and solar and ultraviolet radiation. They also measure different atmospheric and meteorological phenomena like wind speed and direction, temperature, barometric pressure and humidity. The data generated by these sensors is very useful, for example, to track the evolution of the air quality over time. However, the provided raw data is very difficult to analyse without the possibility of making complex queries over it. To solve this issue, all of these features have been semantically annotated as can be seen at Section 4.

The data gathered by these sensors can be accessed in two different ways: 1) through the historical records stored into CSV files since 2000 and 2) through real-time data published at each sensor's web page[6] extracted via web-scraping techniques. Two simple Python scripts have been built to parse both data sources. The generated RDF data is stored into an OpenLink Virtuoso semantic store and served through Pubby Linked Data interface. We do not go into this process of data transformation and publication in any depth because this paper focuses into the work done with ontologies used along the project.

4 Semantic Annotation of Pollution Sensors

Different ontologies have been used to semantically annotate these sensors' data, as depicted in Figure 1. The main ontology of the model is SSN (Semantic Sensor Network Ontology) [8]. This ontology, developed by the W3C Semantic Sensor Incubator Group, is used in Bizkaisense to annotate different aspects of sensors and their measurements. The location of a sensor is represented by an instance of the *Point* class from WGS84 Vocabulary[7] through `dul:hasLocation` property, and it is linked with the nearest *Feature* instance of GeoNames[8] through `dul:nearTo` property.

More reingenieering work has been done extending SWEET and MUO ontologies. SWEET 2.3 (Semantic Web for Earth and Environmental Terminology) is a collection of ontologies that describes both orthogonal concepts (space, time, physical quantities, etc.) and integrative science knowledge concepts (phenomena, events, etc.) [9]. Although the ontology is very complex and describes a wide variety of chemical substances, there are some concepts that it does not include. Concretely, we have extended the ontology[9] with classes representing *Ethylbenzene* ($C_6H_5C_2H_5$) and *HydrogenSulfide* (H_2S), as can be seen in Figure 2.

[6] `http://www.ingurumena.ejgv.euskadi.net`
[7] `http://www.w3.org/2003/01/geo/`
[8] `http://www.geonames.org/`
[9] `http://helheim.deusto.es/bizkaisense/sweetAll-extended.owl`

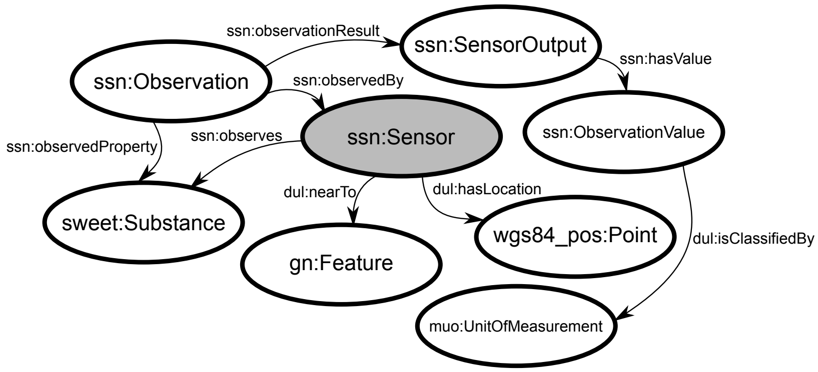

Fig. 1. Aligning the SSN ontology with SWEET, GeoNames, MUO, DUL and WGS84 ontologies.

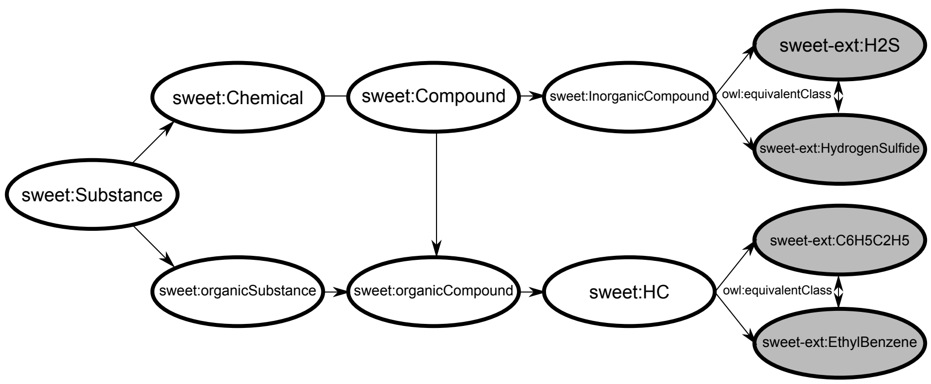

Fig. 2. Classes added to SWEET (colored) and their partial class hierarchy.

Regarding to MUO ontology (Measurement Units Ontology)[10], we have extended its instances[11] from the data extracted from UCUM (Unified Code for Units of Measure)[12]. These new instances are *cubic-meter, cubic-squared, microgram, millibar, milligram, milliwatt, meter-per-second, micro-gram-per-cubic-meter, milligram-per-cubic-meter, milliwatt-per-squared-meter, watt-per-squared-meter and watt-per-squared-meter*, as can be seen in Figure 3.

Finally, the Dublin Core [6] vocabulary is used to annotate common attributes like dates, titles, descriptions and so on. Code 1 shows an example of an observation made by a pollution sensor. More examples can be found at the project web page[13] and its SPARQL endpoint[14].

[10] `http://idi.fundacionctic.org/muo/muo-vocab.html`
[11] `http://helheim.deusto.es/bizkaisense/ucum-ext.owl`
[12] `http://idi.fundacionctic.org/muo/ucum-instances.html`
[13] `http://helheim.deusto.es/bizkaisense/`
[14] `http://helheim.deusto.es/bizkaisense/sparql`

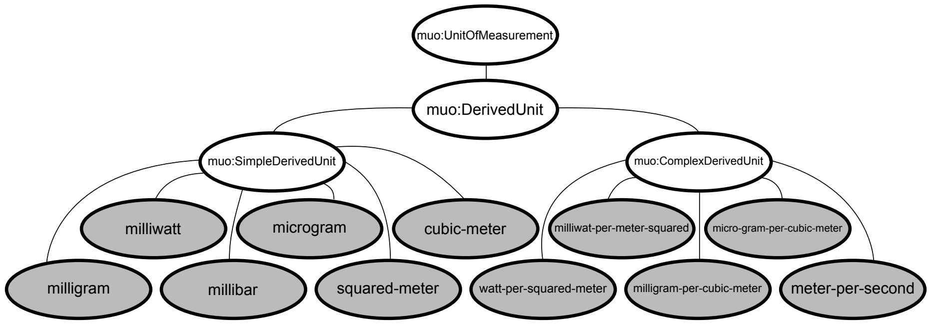

Fig. 3. Instances added to UCUM. The prefix `ucum-ext:` has been omitted from instances to ease the comprehension of the figure.

Code 1 Example of an observation made by a pollution sensor.

```
@prefix bizkaisense:   <http://helheim.deusto.es/bizkaisense/resource/station/> .
@prefix observation:   <.../bizkaisense/resource/station/ELCIEG/NO/01012011/00#> .
@prefix ssn:           <http://purl.oclc.org/NET/ssnx/ssn#> .
@prefix sweet:         <http://http://sweet.jpl.nasa.gov/2.3/propSpeed.owl#> .
@prefix dul:           <http://www.loa-cnr.it/ontologies/DUL.owl#> .
@prefix ucum-extended: <http://helheim.deusto.es/bizkaisense/ucum-extended.owl#> .

observation:             rdf:type               ssn:Observation ;
                         dc:date                "2011-01-01T00:00:00" ;
                         ssn:observedProperty   sweet:NO ;
                         ssn:observationResult  observation:sensoroutput ;
                         ssn:observedBy         bizkaisense:ELCIEG .
observation:sensoroutput rdf:type               ssn:SensorOutput ;
                         ssn:hasValue           observation:outputvalue .
observation:outputvalue  rdf:type               ssn:ObservationValue ;
                         dul:hasDataValue       3 ;
                         dul:isClassifiedBy     ucum-extended:microgram-per-cubic-meter .
```

5 Conclusion and Future Work

On this paper we have described the efforts made to semantically annotate the Sensor Open Data provided by Open Data Euskadi. These efforts include the analysis of different ontologies from the domains of sensor networks, chemistry and meteorology; and the extension of these ontologies to fulfil all the requirements of these pollution sensors. Even though the SWEET ontology belongs to a concrete domain, the extension of UCUM instances of MUO ontology can be reused in a wide variety of cross-domain projects. In addition, this semantic representation of pollution sensors allows us to make complex queries over their data, e. g. the queries used in Bizkaisense for calculating averages of certain substances in a region, as the one we can see in Code 2. Furthermore, this semantic model can be adopted by other sensor networks of the same domain. The approach of extending existing ontologies in contrast of creating new ad-hoc ontologies allows the interoperability among different sensor networks. On the other hand, this paper demonstrates the usefulness of Open Data platforms like ODE.

Code 2 Example of a complex query used in Bizkaisense.

```
SELECT (AVG(?value) as ?avg) WHERE {
    ?medition ssn:observedBy       ?station .
    ?station  dul:nearTo           <http://sws.geonames.org/3104499> .
    ?medition dc:date              ?date .
    ?medition ssn:observationResult ?res .
    ?medition ssn:observedProperty  sweet:NO .
    ?res      ssn:hasValue         ?val .
    ?val      dul:hasDataValue     ?value .
    ?val      dul:isClassifiedBy   ?obsunit .
    FILTER (xsd:dateTime(?date) >= xsd:dateTime("2011-02-17T00:00:00")) .
    FILTER (xsd:dateTime(?date) <= xsd:dateTime(("2011-02-21T00:00:00"))) . }
```

The next goal in Bizkaisense is the integration of data about pollution sensors with other data sources related to environmental domain, like solid and liquid wastes production of Basque Country. With the integration of more data sources we expect to appeal the experts of the domain to increase the features of the system and to demonstrate the real value of the Sensor Open Data vision.

References

1. Barnaghi, P., Presser, M., Moessner, K.: Publishing linked sensor data. In: Proceedings of the 3rd International Workshop on Semantic Sensor Networks (2010)
2. Calbimonte, J.P., Jeung, H., Corcho, O., Aberer, K.: Semantic sensor data search in a large-scale federated sensor network. 4th International Workshop on Semantic Sensor Networks (2011)
3. Compton, M., Henson, C., Lefort, L., Neuhaus, H., Sheth, A.: A survey of the semantic specification of sensors. Proc. Semantic Sensor Networks 17 (2009)
4. Corcho, O., García-Castro, R.: Five challenges for the semantic sensor web. Semantic Web 1(1), 121–125 (2010)
5. d'Aquin, M., Nikolov, A., Motta, E.: Enabling lightweight semantic sensor networks on android devices. 4th International Workshop on Semantic Sensor Networks (2011)
6. Initiative, D.C.M.: Dublin core metadata element set, version 1.1: Reference description (1999), `http://dublincore.org/documents/1999/07/02/dces/`
7. Le-Phuoc, D., Quoc, H.N.M., Parreira, J.X., Hauswirth, M.: The linked sensor Middleware–Connecting the real world and the semantic web. Proceedings of the Semantic Web Challenge (2011)
8. Lefort, L., Henson, C., Taylor, K., Barnaghi, P., Compton, M., Corcho, O., Garcia-Castro, R., Graybeal, J., Herzog, A., Janowicz, K.: Semantic sensor network XG final report. W3C Incubator Group Report (2011)
9. Raskin, R.G., Pan, M.J.: Knowledge representation in the semantic web for earth and environmental terminology (SWEET). Computers & Geosciences 31(9), 1119–1125 (2005)
10. Stasch, C., Schade, S., Llaves, A., Janowicz, K., Bröring, A.: Aggregating linked sensor data. 4th International Workshop on Semantic Sensor Networks p. 46 (2011)