A System for Publishing Sensor Data on the Semantic Web

Alexandra Moraru[1], Carolina Fortuna[2] and Dunja Mladenic[1]

[1] Artificial Intelligence Laboratory, J. Stefan Institute, Ljubljana, Slovenia
[2] Department of Communication Systems, J. Stefan Institute, Ljubljana, Slovenia

The development of sensor technologies in the last years offers support for new applications for the Internet of Things. The availability of the data resulting from sensor networks deployed in different environments is important for a number of advanced services, such as traffic flow prediction and power consumption monitoring.

We propose a system for publishing sensor data following the linked data principles and providing hereby integration with the Semantic Web. The main components are the Semantic Enrichment component and the Data Publishing component, while for storing sensor data we use a relational database.

Keywords: semantic web, linked data, sensor web

1. Introduction

The Internet of Things is gaining popularity with the development of the Internet towards a network of interconnected objects. From mobile phones to transportation cargos, to electrical appliances and to any type of sensing devices new applications are finding their way to daily life. This development will provide new services and will enable new directions for communication. (i.e. "things-to-persons" or "thing-to-thing") [1].

A major role in the development of Internet of Things is played by "web accessible sensor networks and archived sensor data that can be discovered and accessed using standard protocols and application program interfaces" [2], defined as Sensor Web by the Open Geospatial Consortium (OGC). However, for obtaining the maximum impact that Sensor Webs can have on the development of Internet of Thing, challenges have appeared not only for the physical infrastructures, but also for data management and processing, as the amount of data that sensors can produce is significant [3].

Traditional approach applied for handling existing sensor data (i.e., measurements collected by a sensor network) is to store it in a database and process it at a later time. Recent research in the context of relational databases is tackling problems such as prioritization in case of limited bandwidth, modeling and cleaning of sensor readings, uncertain data management [4]. In data mining and machine learning communities, technology has been developed to process and build models based on streams of data [5], therefore being able to build real-time applications based on such streams (e.g., intrusion detection systems using stream mining).

Up to this point, the systems are built in such a way that the integration of sensor data from two different systems is non-trivial to achieve. By describing the meanings of data streams and the context under which these data were collected in a computer understandable manner, truly large scale distributed sensing systems can be built. This is where the semantic web protocol stack and semantic technologies can play a key role. Extending Sensor Web towards applying semantic technologies leads to the development of Semantic Sensor Web (SSW) [6] enabling the integration of different systems. Some of the directions adopted for achieving such integration are related to Linked Open Data (LOD) (including Linked sensor data [7][8][9]), or to semantic annotation and composition of web services [10].

In this paper, we propose and implement a

system for publishing real-world sensor data collected from a sensor network that includes sensing devices for monitoring environmental conditions such as temperature, humidity, luminance and pressure. A semantic layer is added for enriching the sensor data (measurements obtained from sensors) and metadata (properties of sensors including location and observed phenomena). The semantic enrichment is based on mapping rules that are connecting the database that stores the sensor data and metadata to a standardized ontology in the sensor network domain. The output of the system is a dataset of semantic sensor descriptions and real-time measurements, published following the principles of linked open data and therefore web accessible.

The rest of this paper is structured as follows. Section 2 describes related work and puts the proposed system in the perspective of the existing research. Section 3 discusses requirements for publishing real world data on the semantic web. In Section 4 we describe the proposed system, while in Section 5 we present an implementation for publishing enriched sensor data. Section 6 gives an example of possible interaction with the system and we conclude the paper in Section 7.

2. Related Work

One way of making sensor descriptions and measurements available on the Web is to publish them on LOD cloud. The advantages and challenges of Linked Sensor Data are discussed by Keßler and Janowicz [9]. Integrating the descriptions of sensors into LOD can provide better sensor data accessibility without introducing very high complexity. The paper stresses the importance of finding the appropriate links between different datasets from LOD and proposes a semiautomatic way for generating them. Patni et al. [7] were the first to add to the LOD cloud a large dataset of sensor descriptions and measurement, by first representing it in Observation and Measurements (O&M) standard and then converting it to a standard representation in Resource Description Framework (RDF). Barnaghi and Presser [8] propose a platform for publishing linked sensor data defined by spatial, temporal and thematic attributes. The platform offers an interface for publishing linked sensor

data without requiring from the users a semantic technological background. However, the user is requested to manually enter relevant keywords that describe the sensors for obtaining a list of suggested concepts from on-line repositories.

In contrast, in our work we propose manually mapping structured sensor data and metadata from a database to semantic concepts. In addition, we chose a W3C standardized ontology that contains the semantic concepts needed. The advantage of using a standardized ontology is that it can offer mediation between different representations of sensor descriptions. While we do not cover the problem of database population, we assume it can be automatized through different protocols [11].

Similar to our work, Le-Phouc et al. [12] propose a system for publishing sensor data based on D2R Server and linking it to existing ontologies. In addition, the authors approach the problem of having multiple sensor sources and suggest using specialized wrappers for every different format of sensor data. In our work, we focus on a single sensor source and optimize the semantic enrichment using a standardized ontology.

3. Integration of Sensor Data in the Semantic Web

The Semantic Web [13], uses representation languages such as RDF and OWL to provide a formal description of data and knowledge. The purpose of the Sematic Web is to provide a machine understandable representation of data, while creating links in these data makes it easier to discover and allows reasoning.

One of the requirements for building the SSW refers to knowledge representation. The advantages that semantic technologies bring in knowledge representation are better scalability and interoperability, since adding or changing new information to a set of programs that use the same model resumes to changing the external model, while the design of those programs can remain the same, without the need of human involvement [14]. The model used for knowledge representation, from the point of view of SSW, must be suitable for capturing all the semantics required for enrichment of sensor data and must also provide a shared vocabulary used for

publishing such data. One of the categories of knowledge representation appropriated for the model required is by means of ontologies.

With the development of semantic sensor networks a number of ontologies describing the sensor networks domain have been brought forth in the last years. A detailed survey is provided in [15], where eleven sensor network ontologies are analyzed. Therefore, considering the need of standardization regarding sensor networks ontologies, an incubator group from W3C has been formed with the purpose of developing ontologies for sensor networks – referred to as Semantic Sensor Network (SSN) ontology[1] – and search for appropriate methods for enhancing available standards with semantic technologies.

4. System Description

The process of publishing sensor data on the web involves several steps and poses different challenges. Starting with storing the data from the sensing devices and resulting with public datasets containing semantically enriched measurements and the related metadata, various methods can be applied in the process. We present a system that includes all the steps of the publishing process. Although this is not the

only way to treat the problem of storing and publishing linked sensor data, the proposed system offers an efficient and user friendly way of solving it, being optimized for a real deployment of a sensor network.

Figure 1 illustrates the proposed system showing the main components of the system and the relation between them. The sensor data results from a sensor network deployed in the environmental domain which sends streams of raw measurements and metadata to an intermediate server. Next, a relational database is updated when the server receives new messages.

The main components are the Semantic Enrichment component and Data Publishing component. The first component contains modules which allow semantic enrichment of metadata and measurements. This component uses vocabularies (i.e. ontologies) and mapping rules for defining the relationships between and within the data and metadata from the database.

The Data Publishing component contains modules which permit exposing the sensor information on the Web. We have used a publicly available D2R Server[2] as an implementation of this component, as it offers an interface to the database and does not require replicating the database content into a separate RDF triplestore. In order to provide semantic meaning
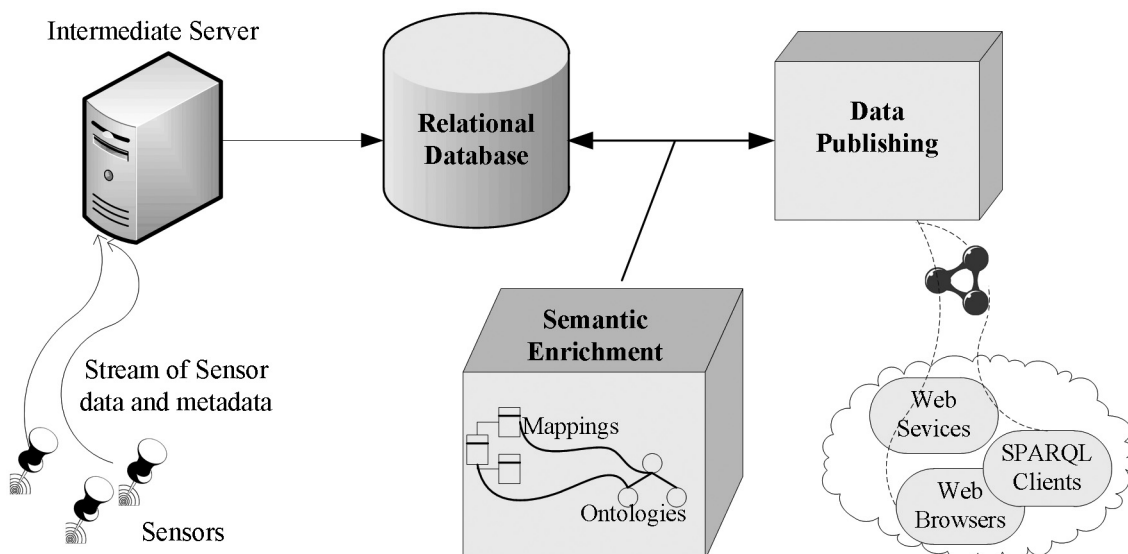

Figure 1. Architecture of the proposed system

---

for the sensor data and metadata contained in the database, we have used existing ontologies for editing the mapping file used by the D2R Server. Finally, the database content can be accessed by different web services, SPARQL clients or browsed using a HTML interface.

5. Publishing Semantically Enriched Sensor Data

Making sensor data available for a large number of users enables the development of new and useful applications. The methods for publishing sensor data can vary from standardized web services, such as OGS's Sensor Observation Service (SOS), to applications specific methods, as the ones used by Pachube[3] or Sensorpedia[4]. However, such methods require prior knowledge of the infrastructures used, while Linked Sensor Data would enable better accessibility. Moreover, the support for integration with existing knowledge would increase also the usability of data.

Properly exposing and linking data is essential for the success of the Semantic Web. According to the principles of Linked Data, each entity should have at least three corresponding dereferenceable URIs [17]: one URI for the real-world object itself and two more URIs for related information resources that describes the real-world object and have a HTML and a RDF/XML representation.

5.1. Semantic Vocabulary

In order to provide meaning for the data published, we have chosen representative concepts from related ontologies. The majority of the concepts used are described in the SSN ontology, which is a sensor network specific ontology. Its alignment with the DOLCE Ultra Lite upper ontology helps in describing linked data resources.

As our experimentation work required just a subset of the SSN ontology concepts and relationships, we will provide details just on those concepts, as they are illustrated in Figure 2. The class *System* represents a sensing infrastructure and its sub-systems can be represented using the predicate *hasSubSystem*. The physical location of a system is represented by the class *Platform*, which can have multiple systems attached.

The class *Sensor* is used to represent concrete sensing objects and the predicate observes and indicates the *Property* observed by a sensor (e.g. temperature, humidity). In order to represent sensor devices, the class *SensingDevice* can be used, which inherits all the properties of the classes Sensor and Device. For representing sensor measurements we have used three classes: *Observation*, *SensorOutput* and *ObservationValue*. An observation represents a situation in which a value of a property is estimated by a sensor. The result of an observation is a sensor output which has a specific value.
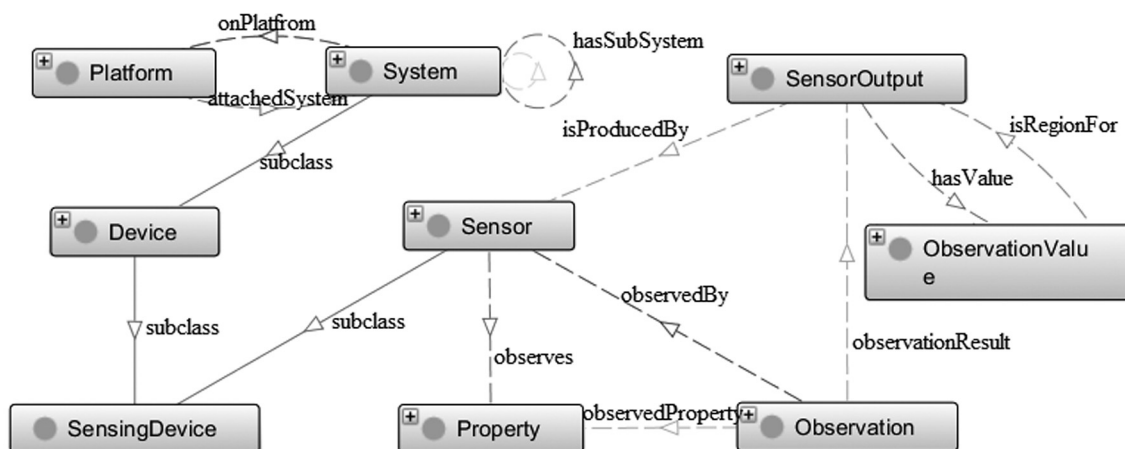

Figure 2. Subset of concepts and relationships from SSN ontology

---

[3] http://www.pachube.com/
[4] http://www.sensorpedia.com/

For the geographical location of the platforms on which the sensors are attached we have used concepts from other ontologies that are also part of linked data cloud. For latitude and longitude coordinates we have used the Basic GeoWGS84 Vocabulary[5], which provides the namespace for representing the coordinates. In addition, for geographical names we have used GeoNames[6] database, in RDF format, and links to platforms were represented using the *based_near* predicate from FOAF ontology[7]. Each platform was linked to the nearby location, based on geographical coordinates and determined using the *findNearbyPlaceName* web service provided by GeoNames.

5.2. Mapping Rules for Semantic Enrichment

For the mapping between the database content and the corresponding semantic concepts we have used the D2RQ language [16]. The mapping was done respecting the SSN ontology and not the database structure, therefore some of the instances in the database were mapped to more than one concept, as it will be explained further.

The information about the systems deployed in the network and the sensing devices attached, were represented in separate tables. Each row of a table was describing an instance, which was uniquely identified by the primary key of the table. Therefore, these instances were mapped directly to *System*, respectively to the subclasses of *SensingDevice* classes. However for the concrete instances of the *Platform* class, representing the physical objects to which the systems are attached, there is no unique identifier in the database. Therefore, we used the geographical coordinates to identify the platforms, with the assumption that at one exact point in space there cannot be more than one platform. The D2RQ mapping corresponding is represented in Figure 3. It can be observed that the platforms are defined as individuals of the *LightPole* class, which is then defined as a subclass of Platform and *SpatialThing* (the latter class is from the GeoWGS84 Vocabulary) classes. The creation of the *LightPole* class is related to the real world scenario of the deployment used.

```
map:Platform a d2rq:ClassMap;
d2rq:dataStorage map:database;
d2rq:uriPattern
"platforms/LightPole@@sensor_node_table.gps_latitude@@-
@@sensor_node_table.gps_longitude@@";
d2rq:class vocab:LightPole;
d2rq:classDefinitionLabel "Light Pole";
d2rq:additionalClassDefinitionProperty map:
PlatformSubclass;
d2rq:additionalClassDefinitionProperty map:
SpatialThingSubclass.

map:LightPolePlatformSubclass a d2rq:AdditionalProperty;
    d2rq:propertyName rdfs:subClassOf;
    d2rq:propertyValue ssn:Platform.

map:LightPoleSpatialThingSubclass a
d2rq:AdditionalProperty;
    d2rq:propertyName rdfs:subClassOf;
    d2rq:propertyValue geo:SpatialThing.
```

Figure 3. Platfrom mapping using D2RQ

```
map:DeviceType a d2rq:ClassMap;
d2rq:dataStorage map:database;
d2rq:uriPattern
"http://sensorlab.ijs.si:2020/vocab/resource/sensor/SensorType
@@sensor_type_table.st_uid@@".

map:sensor_type a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:DeviceType;
d2rq:property rdf:type;
d2rq:constantValue rdfs:Class.

map:sensor_type_subclass a d2rq:PropertyBridge;
d2rq:belongsToClassMap map:DeviceType;
d2rq:property rdfs:subClassOf;
d2rq:constantValue ssn:SensingDevice.

#sensor type mapping
map:sensor_class a d2rq:PropertyBridge;
    d2rq:property rdf:type;
    d2rq:belongsToClassMap map:sensor;
    d2rq:join "sensor_table.st_uid =
sensor_type_table.st_uid";
    d2rq:uriPattern
"http://localhost:2020/vocab/resource/sensor/SensorType@@s
ensor_type_table.st_uid@@";.
```

Figure 4. Sensing Device subclasses using D2RQ

Another piece of information stored in the database is the type of the sensing device placed on a sensor node. We have extracted types of sensing devices from a table of the database, which were mapped as subclasses of the *SensingDevice* class of SSN ontology and next used for specifying the type of individuals from sensing devices, using the "join" command from D2RQ language (Figure 4).

For the content related to sensor observations the unique ids from the database were used to identify all the individuals corresponding to *Observation*, *SensorOutput* and *ObservationValue*. The sensor measurements registered over the last day are mapped one by one, while for the past measurements average of the last hour is calculated in a separate view of the database and then mapped to RDF representation.

Figure 5. Web Interface of the D2R Server presenting all the classes used in mapping

The D2R Server manages the URIs created for the RDF representation of the database content and allows data to be browsed and searched using web browsers or SPARQL endpoints. All the classes can be browsed for listing their instances, as it can be observed in Figure 5, and also they can be checked at `http://sensorlab.ijs.si:2020/`.

6. Conclusions

Following the LOD principles is a promising approach for making sensor data available on the Web, as it addresses users from diverse domains of interest – through the openness of the format – which can access and process it. Also, by facilitating the integration with other existing knowledge already represented in LOD, it provides the means for a more complete and detailed description of the data published.

The proposed system facilitates publishing sensor data on the Semantic Web, contributing also to the development of the Internet of Things. It follows the principles of LOD and uses standardized vocabularies for describing the data. The sensor data is stored in a relational database, while access to the RDF representation of database content is provided using a publicly available tool. The semantic enrichment of sensor data is done through mapping rules and the meaning of data is mainly represented using the SSN ontology. The system presents also the advantage of publishing real time data for the last day, while for the archived data averages for each hour are made available.

References

[1] Commission of the European Communities. Internet of Things — An action plan for Europe. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions; 2009 June 18; Brussels.

[2] M. BOTTS ET AL., OGC White Paper OGC® Sensor Web Enablement: Overview and High Level Architecture. White Paper. OpenGIS. 2007.

[3] A sea of sensors. The Economist. The Economist Newspaper Limited 2010 Nov 4; London.

[4] S. MADDEN, Database Abstractions for Managing Sensor Network Data. *Proceedings of the IEEE 2010*; 98(11): 1879–1886.

[5] M. G. MOHAMED, A. ZASLAVSKY, S. KRISHNASWAMY, Mining data streams: a review. *SIGMOD Rec. 2005*; 34(2): 18–26.

[6] A. SHETH, C. HENSON, S. S. SAHOO, Semantic Sensor Web. *IEEE Internet Computing 2008*; 12(4): 78–83.

[7] H. PATNI, C. HENSON, A. SHETH, Linked sensor data. In *2010 International Symposium on Collaborative Technologies and Systems*; 2010 May 17–21; Chicago, IL,USA. p. 362–370.

[8] P. BARNAGHI, M. PRESSER, Publishing Linked Sensor Data. In *Proceedings of the 3rd International Workshop on Semantic Sensor Networks*; 2010 November.

[9] C. KESSLER, K. JANOWICZ, Linking Sensor data – Why, to What and How? In *Proceedings of the 3rd International Workshop on Semantic Sensor Networks*; 2010 November.

[10] Envision EU Project, `http://www.envision-project.eu/` [02/15/2011]

[11] A. MORARU ET AL., Exposing Real World Information for the Web of Things. In *Proceedings of the 8th International Workshop on Information Integration on the Web*, 2011 March 28, Hyderabad, India.

[12] D. LE-PHUOC ET AL., Live linked open sensor database. In *Proceedings of the 6th International Conference on Semantic Systems*. New York, NY, USA: ACM, Article 46, 4 pages.

[13] G. ANTONIU, F. VAN HARMELEN, *A Semantic Web Primer*. The MIT Press; March 31 2008.

[14] T. BERNERS-LEE, J. HENDLER, O. LASSILA, The semantic web a new form of web content that is meaningful to computers will unleash a revolution of new possibilities, *Scientific American*; 2001 May 17.

[15] M. COMPTON ET AL., A Survey of the Semantic Specification of Sensors. In *Proceedings of the 2nd International Workshop on Semantic Sensor Networks*; 2009 October.

[16] C. BIZER ET AL., The D2RQ Platform v 0.7 – Treating Non-RDF Relational Databases as Virtual Graphs. `http://www4.wiwiss.fu-berlin.de/bizer/d2rq/spec/` [02/01/2011]

[17] T. HEATH, C. BIZER, *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool 2011. ISBN 978-1608454303.

Contact addresses:
Alexandra Moraru
Artificial Intelligence Laboratory
J. Stefan Institute
Jamova 39
Ljubljana
Slovenia
e-mail: `alexandra.moraru@ijs.si`

Carolina Fortuna
Department of Communication Systems
J. Stefan Institute
Jamova 39
Ljubljana
Slovenia
e-mail: `carolina.fortuna@ijs.si`

Dunja Mladenić
Artificial Intelligence Laboratory
J. Stefan Institute
Jamova 39
Ljubljana
Slovenia
e-mail: `dunja.mladenic@ijs.si`

CAROLINA FORTUNA is a senior research assistant and a PhD student working at the Department of Communication Systems, Jozef Stefan Institute. She received her BSc in electrical engineering from the Technical University of Cluj-Napoca, Romania. Her research is interdisciplinary focusing on semantic technologies with applications in modelling of communication and sensor systems and on combining semantic technologies, statistical learning and networks for analyzing large datasets. She has published papers in refereed conferences and journals, served in the program committee of conferences such as ICC, Globecom and WCNC. She has also been to industry internships at Bloomberg LP and Siemens PSE.

DUNJA MLADENIĆ is an expert on study and development of machine learning, data/text mining, semantic technology techniques and their application in real-world problems. She has been associated with the J. Stefan Institute since 1987, first as a student and since 1992 employed as a researcher. She is leading Artificial Intelligence Laboratory of the J. Stefan Institute since 2011. She received her MSc and PhD degrees in computer science from the University of Ljubljana in 1995 and 1998 respectively. She was a visiting researcher at the School of Computer Science, Carnegie Mellon University, USA in 1996-1997 and in 2000-2001. She has published papers in refereed conferences and journals, served in the program committee of international conferences and organized international events in the area of text mining, link analysis and data mining. She is co-editor of several books including "Data Mining and Decision Support: Integration and Collaboration", Kluwer Academic Publishers 2003, "Semantic Knowledge Management: Integrating Ontology Management, Knowledge Discovery, and Human Language Technologies" Springer 2008, "Web Mining: from Web to Semantic Web", Springer 2004, "Semantics, Web and Mining" Springer 2006, "From Web to Social Web: discovering and deploying user and content profiles", Springer 2007, "Knowledge Discovery Enhanced with Semantic and Social Information" , Springer 2009.

ALEXANDRA MORARU is a PhD student at the J. Stefan International Postgraduate School in the Information and Communication Technologies program. She received her BSc in computer science from the Technical University of Cluj-Napoca in 2009 and her MSc in information and communication technology from the J. Stefan International Postgraduate School in 2011. She started her collaboration with J. Stefan Institute in 2008, with a 2 months internship program, and since 2009 she has been a student there. Her general research interests are in the area of Semantic Web and semantic technologies, more specifically, the applicability of semantics in sensor networks.