Live Linked Open Sensor Database

Danh Le-Phuoc
Digital Enterprise Research Institute,
National University of Ireland, Galway
Galway, Ireland
danh.lephuoc@deri.org

Josiane Xavier Parreira
Digital Enterprise Research Institute,
National University of Ireland, Galway
Galway, Ireland
josiane.parreira@deri.org

Michael Hausenblas
Digital Enterprise Research Institute,
National University of Ireland, Galway
Galway, Ireland
michael.hausenblas@deri.org

Yuanbo Han
Digital Enterprise Research Institute,
National University of Ireland, Galway
Galway, Ireland
yuanbo.han@deri.org

Manfred Hauswirth
Digital Enterprise Research Institute,
National University of Ireland, Galway
Galway, Ireland
manfred.hauswirth@deri.org

ABSTRACT

There are millions of sensors being deployed all over the world. Data generated by these sensors is provided in different formats and interfaces and is rarely associated with semantics that describe its meaning. The heterogeneity and lack of semantic descriptions pose a big barrier for accessing sensor data and combining it with other data sources.

The *Live Linked Open Sensor Database* project is the first project to provide a live database of semantically enriched sensor data, where each sensor reading is extended by adding proper metadata and by linking it to resources in the Linked Open Data Cloud. Currently, the database provides information of approximately 200,000 sensors and we are currently working on expanding it to incorporate even more data sources.

Categories and Subject Descriptors

H.3.3 [**Information storage and retrieval**]: Information search and retrieval—*information filtering, query formulation*; H.3.5 [**Information storage and retrieval**]: On-line Information Services—*data sharing, web-based service*

Keywords

Linked sensor, sensor mashup, sensor stream

1. INTRODUCTION

Sensors have quickly become very popular and currently there are millions of sensors being deployed all over the world. However, data provided by these sensors usually comes in different formats without any semantics that describes its meaning, which limits the use of sensor data to a few applications domains. Making sensor-generated information usable as a key source of knowledge will require its integration into the existing information space of the Web. Gartner [14] predicts that...

> ...by 2015, wirelessly networked sensors in everything we own will form a new Web. But it will only be of value if the 'terabyte torrent' of data it generates can be collected, analyzed and interpreted [14].

Adding enough metadata to describe the meaning of sensor data is essential to make them more accessible. Meaning of a sensor data includes the feature of interest, the specification of measuring devices, accuracy, measuring condition, scenario of measurements, location, etc. Such metadata becomes even more important in scenarios involving a large numbers of sensors and gigabytes of sensor data. In particular, in cases where the user does not have an a priori knowledge of the nature of the sensors, the semantic description provides a starting point for a better understanding of what is available, what are the constrains, etc. For instance, a city planner may want to assess and monitor quality of life in certain area. To do so, he can start to navigate from his own domain of knowledge, then finds out that quality of life can be measured from noise, sunlight, humidity, air pollution, traffic condition, etc. Then he can filter out which sensor sources can provide such data in his area of interest.

As we mentioned before, data from different sources is heterogeneous. Data providers like SensorMap [17], EarthScope [7] and Earthcam [6] usually publish data in different interfaces, formats and standards. Such data sources are difficult to be integrated in order to make them accessible to other applications.

Our *Live Linked Open Sensor Database* project is the first project to provide a live database of semantically enriched sensor data, where each sensor reading is extended by adding proper metadata and by linking it to resources in the Linked Open Data Cloud. To enable easy access and integration of heterogenous sensor data sources, we collect sensor data published in a variety of formats such as XML,RSS,JSON and plain text, and enrich their meaning and context via ontologies and links to the Linked Data Cloud. This is done by using the data transformation and consolidation modules of DERI Pipes [11]. Eventually, the integrated sensor data is published under a *unifying linked data model* [12] at `http://sensormasher.deri.org/`. The metadata is accessible via both a SPARQL endpoint and a visual explorer. The live linked stream data from original sources are accessible in RDF format, which is mediated by our sensor middleware [1, 10].

Currently, the database provides information of approximately 200,000 sensors and we are currently working on expanding it to incorporate even more data sources.

The rest of this paper is organized as following. We introduce motivating use cases in Section 2. The sensor data sources currently supported by the project are listed in Section 3. Section 4 presents how we expose the sensor data sources as linked data sources, as well as how links to the Linked Data Cloud are created. The *SensorMasher*, a demo application for this Linked Sensor Data is described in Section 5. Finally, Section 6 presents conclusion and future work.

2. MOTIVATING USE CASES

Due to its very nature, the Linked Data Web is a global, decentralised information space. In order to utilise the data in an application, one depends on central-point-of-access services, allowing for the look-up of data sources based on their characteristics. In the following we discuss several cases where such a database for sensors can be of use.

A typical service provided by national weather agencies are storm and hail warning Web pages. Now, imagine a farmer who possesses an acre with, say, maize crops. Every now and then, he checks such a page to assess the potential danger for his seed; a repetitive and time-consuming task, one would rather have automatised and in-time. Having the Linked Sensor Database in place, it would be straightforward to implement a notification service, which, for example, takes into account the relevant sensors for a certain area (weather stations, social media streams, etc.) and—given that certain conditions are met, e.g., wind force and temperature—sends a text message to our farmer, warning him about a potential danger in advance.

Further, in disaster cases, such as disease outbreaks, people often report about the situation via media such as Twitter, and equally the officials want to get a quick overview in order to take immediate measures, like evacuating people. Combining the user-generated content from social platforms via the Linked Open Data cloud with relevant sensor data can help prevent the spread of a disease. One major challenge here is to determine what the relevant sensors are, that is, the selection of sensors based on their characteristics, for instance, location, type of sensor, and so forth, which can be provided by the Linked Sensor Database.

Eventually, mobile browser can augment the reality with information about local points of interest, such as historical sights, nearby bus stops and cafes. However, current implementations of mobile augmented reality have severe limitations regarding data source selection and integration of data sources, as we have recently noted[1]. With the Live Open Linked Sensor Database, the data source selection can be performed dynamically and based on the user's preferences and demands.

[1] `http://www.w3.org/2010/06/w3car/exploiting_lod_for_ar.pdf`

3. SENSOR DATA SOURCES

Currently, our Live Linked Open Sensor Database contains data coming from surface weather observations [19]. Surface weather observations are the most popular data sources for weather forecast and climatological warnings. Sensor data is taken by automated airport weather stations [2] or personal weather stations [16]. Sensors from weather stations report data about temperature, visibility, wind direction and speed, air pressure, dew point, etc., around the world. This data is provided as weather services such as Weather Underground [23], WeatherBug [22], Yahoo Weather [26] and NOOA [15], which together covers over 50,000 places in the globe. Among them, NOOA not only provides data from weather sensors but also from radar, satellite, river, sea and snow observations, etc. In addition to the weather service, Yahoo also provides traffic data [25] from traffic sensors. Moreover, Weather Bug allows access to cameras attached to weather stations. These cameras are cheap and easy to deploy, therefore, there are millions of them having been deployed to record real time images. Usually data from such cameras are made accessible via a simple web page, which makes it easy to link them to other data sources. For example, the New York city Department of Transportation[2] provides real time traffic cameras around the city, e.g., the Brooklyn Bridge at Centre Street[3]. EarthCam[4] is one of the biggest portals collecting live data from cameras.

4. SENSORS AS LINKED DATA SOURCES

The process of transforming raw, heterogeneous sensor data into linked and semantically enriched information that is available in real time involve many steps, which are described below.

We first access the data sources to identify how the data is provided and which format the data is in. The sensor data sources mentioned earlier are usually accessible via web services, HTTP or open protocols like openDAP[5]. Data formats can be XML, RSS, HTML, JSON or even raw data formats. For each of the different formats we build a wrapper that converts the data to a common representation used in the project. This common representation contains, among other things, a semantic description of the data source. The semantic description is given by the sensor ontology that we are currently developing in the W3C Semantic Sensor Network working group [18], which was extended to provide an ontology that describes the taxonomy of sensor readings like temperature, humidity, visibility and sea level. To improve usability, concepts of our ontology are linked to concepts in the SWEET ontology [20] via *owl:equivalentClass* link.

SWEET is a largely used ontology that provides vocabularies to describe basic science processes, substances, phenomena and realms. For instance, user is likely to be interested in natural phenomenas such as blizzards, snow fall, wind speed and satellite graphic instead of sensor hardware

[2] `http://nyctmc.org/`

[3] `http://nyctmc.org/google_popup.php?cid=175`

[4] `http://www.earthcam.com/`

[5] `http://opendap.org/`

measuring them. These vocabularies enable user to discover sensor data by following semantics of sensor readings other than technical specification of sensors. In addition, we use the FOAF vocabulary for representing organizational data (nature of data source, access policies, etc).

With the data source descriptions we are able to generate semantic data from the sensor readings and link them to resources in the Linked Open Data Cloud. For each type of sensor data source, we configure our data transformation operators in DERI Pipes [11] to represent the metadata in RDF triples and to generate live RDF data from the sensor readings, where each reading has a URI associated with it.

The generated triples are then linked to Geonames resources by using Geonames ontology [9]. To build the links to the geo-elements of Geonames we use lookup services such as Yahoo! GeoPlanet Data [24], Weather Underground and Weather Bug station lookup to find the correlations among the resources. For the data sources that do not provide any lookup service, we developed text analysis modules that extract geo-data from text patterns. For example, for sources providing images from live cameras, we use the Google search API to search for web pages that contain common URL patterns and HTML patterns of the cameras firmware's output, which are then parsed to get the URL of live images and the geo-text associated with it. With the geo-text we can link such readings to the Geonames database.

To publish the sensor meta data and the summaries of live sensor data readings we use D2R [5]. The SPARQL endpoint of D2R provides the interface for exploring the sensor data sources. Live sensor data is recorded by our linked stream processing engine [10], where the size of the time window for which the data must be archived can be configured (e.g., three months for weather data, one day for camera images and satellite graphics, etc). Live Sensor readings are available through URL in RDF format, where these RDF data fragments have links to sensor metadata.

Currently we have a database of approximate 200,000 sensors. The database allows access to approximately 20 million triples about sensor descriptions such as sensor specification, geo-information and summary of readings. The summary of the readings provides links to live readings as well as archived readings of sensor.

5. SENSORMASHER

The SensorMasher application, available at `http://sensormasher.deri.org/`, demonstrates some possible ways of using our Live Linked Sensor Database. SensorMasher provides, among other things, a GUI for non-technical users to explore sensors on the map as well as to filter sensor data sources with browsing facets (see Figure 1). With browsing facets, users can drill down their search results by using free text search as well as taxonomy and spatial hierarchy based filters.

Via this GUI, users can preview and compare real time and historical sensor data with the supported visualizations. Moreover, technical users can use the D2R server's interfaces to navigate through the triplified sensor data sources. Application developers can use SPARQL Endpoint to query meta data and retrieve live data to their applications. We also integrate the DERI Pipes mashup modules to enable users to visually and combine sensor data sources to create new virtual data sources, which can also be published (see Figure 2).

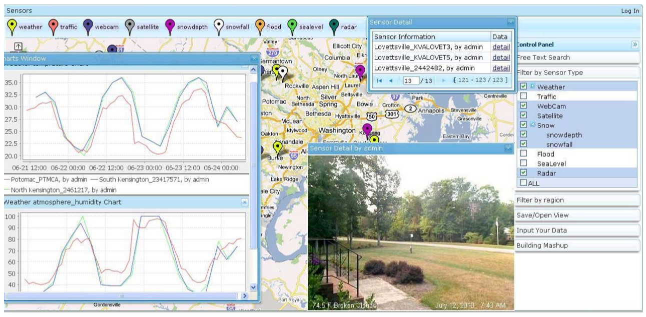

Figure 1: Screenshot of the SensorMasher application

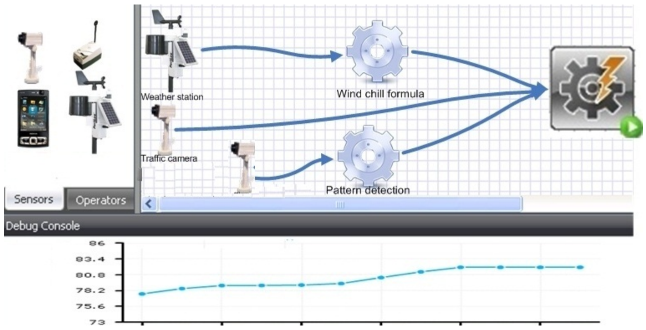

Figure 2: A screenshoot of Sensor Composer

In addition, with the aid of our CQELS (*Continuous Query Evaluation over Linked Streams*) [12] system, users are able to specify notification rules using an extended version of SPARQL that supports stream processing.

6. CONCLUSION

The Live Linked Open Sensor Database project is the first project to provide a live database of semantically enriched sensor data, where each sensor reading is extended by adding proper metadata and by linking it to resources in the Linked Open Data Cloud. With approximate 200,000 sensors, it not only streams live data from all over the world, but it also provides useful interfaces for application developers. This is a on-going project; currently we retrieve new sensor data sources and update the database every two weeks. In next steps, we plan to add more sensor types as well as more links to other data sets within the Linked Open Data Cloud such as LinkeGeoData [13], DBpedia [4], Eurostats [8], US Census [21] and Data-Gov [3].

Along with physical sensors, we also plan to consider dynamic web sources, e.g., blog posts, news feeds, social network applications (Facebook, Twitter, etc.) and search engine updates as virtual sensor data sources. Our goal is to seamlessly integrate physical and virtual sensor data sources into a unifying linked data set.

7. ACKNOWLEDGMENTS

The work presented in this paper is funded by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

8. REFERENCES

[1] Aberer, K., Hauswirth, M., and Salehi, A. 2007. Infrastructure for Data Processing in Large-Scale Interconnected Sensor Networks. In Proceedings of the 2007 international Conference on Mobile Data Management (May 01 - 01, 2007). International Conference On Mobile Data Management. IEEE Computer Society, Washington, DC, 198-205.

[2] Automated airport weather station, http://en.wikipedia.org/wiki/Automated_airport_weather_station

[3] Data-gov, http://data-gov.tw.rpi.edu/wiki

[4] DBpedia, http://dbpedia.org/

[5] D2R,Publishing Relational Databases on the Semantic Web, http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/

[6] EarthCam - Webcam Network, http://www.earthcam.com/

[7] EarthScope - An Earth Science Program, http://www.earthscope.org/

[8] EuroStat, http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData/EuroStat

[9] Geonames ontology, http://www.geonames.org/ontology/

[10] Le-Phuoc, D.,Hauswirth, M. Linked open data in sensor data mashups. http://ceur-ws.org/Vol-522/p3.pdf

[11] Le-Phuoc, D., Polleres, A., Hauswirth, M., Tummarello, G., and Morbidoni, C. Rapid prototyping of semantic mash-ups through semantic web pipes. In Proceedings of the 18th international Conference on World Wide Web (Madrid, Spain, April 20 - 24, 2009). WWW '09. ACM, New York, NY, 581-590.

[12] Le-Phuoc, D., Parreira, J. X.,Hausenblas, M. and Hauswirth, M. Unifying Stream Data and Linked Open Data. Technical Report, DERI,Galway, July 2010.

[13] LinkedGeoData, http://linkedgeodata.org/

[14] Mark Raskino, Jackie Fenn, and Alexander Linden. Extracting Value From the Massively Connected World of 2015. Gartner Research, 1 April 2005.http://www.gartner.com/resources/125900/125949/extracting_valu.pdf

[15] NOAA's National Weather Service, http://www.nws.noaa.gov/observations.php

[16] Personal Weather Station, http://en.wikipedia.org/wiki/Personal_weather_station

[17] Microsoft SensorMap, http://www.sensormap.org/sensewebv3/sensormap/

[18] Semantic Sensor Network Incubator group, http://www.w3.org/2005/Incubator/ssn/

[19] Surface weather observation, http://en.wikipedia.org/wiki/Surface_weather_observation

[20] SWEET ontolgies, http://sweet.jpl.nasa.gov/ontology/

[21] US Census, http://www.rdfabout.com/demo/census/

[22] Weather Bug, http://en.wikipedia.org/wiki/Surface_weather_observation

[23] Weather Underground, http://www.wunderground.com/

[24] Yahoo! GeoPlanet Data, http://developer.yahoo.com/geo/geoplanet/data/

[25] Yahoo Traffic, http://developer.yahoo.com/traffic/

[26] Yahoo Weather, http://weather.yahoo.com/