AGGREGATION OF SEMANTIC SENSOR DATA

Graduation proposal

by

Ivo de Liefde

December 4, 2015

CONTENTS

ACRONYMS

EU european union . . . 1

GIS geographical information system . . . 4

HTTP hypertext transfer protocol . . . 7

INSPIRE infrastructure for spatial information in Europe . . . 1

IoT internet of things . . . 1

IRCEL-CELINE Belgian interregional environment agency . . . 11

ISO international organisation for standardisation . . . 1

OGC open geospatial consortium . . . 1

O&M observations and measurements . . . 1

OWL web ontology language . . . 7

RDF resource description framework . . . 1

REST representational state transfer . . . 3

RIVM Dutch national institute for public health and the environment . . . 11

SensorML sensor modelling language . . . 1

SEL semantic enablement layer . . . 3

Sem-SOS semantically enabled sos . . . 3

SOS sensor observation service . . . 1

SSNO semantic sensor network ontology . . . 3

SSW semantic sensor web . . . 1

SWE sensor web enablement . . . 1

URI uniform resource identifier . . . 3

W3C world wide web consortium . . . 1

WPS web processing service . . . 4

XML extensible markup language . . . 3

1 INTRODUCTION

From 2020 onwards all member states of the european union (EU) should provide sensor data to the infrastructure for spatial information in Europe (INSPIRE) in order to comply with annex II and III of the INSPIRE directive (INSPIRE, 2015). For this a number of sensor web enablement (SWE) standards are required to be used (INSPIRE, 2014). The sensor web is a relatively new development and there are still many question on how to structure it. This thesis aims to develop a method to publish and link sensor metadata on the semantic web for discovering, integrating and aggregating sensor data.

1.1 BACKGROUND

In 2008 the open geospatial consortium (OGC) introduced a new set of standards called SWE. These standards make it possible to connect sensors to the internet and retrieve data in a uniform way. This allows users or applications to retrieve all kinds of sensor data, regardless of the type of observations or the sensor's manufacturer (Botts et al., 2008). Among other standards SWE includes the observations and measurements (O&M) which is a model for encoding sensor data, the sensor modelling language (SensorML) which is a model for describing sensor metadata and the sensor observation service (SOS) which is a service for retrieving sensor data (Botts et al., 2007). O&M has also been adopted by the international organisation for standardisation (ISO) under ISO 19156:2011 (ISO, 2011).

Recently OGC has defined the role which their standards could play in smart city developments (Percivall, 2015). Smart cities can be defined as "enhanced city systems which use data and technology to achieve integrated management and interoperability" (Moir et al., 2014, p. 18). Research on smart cities has shown a great potential for using sensor data in urban areas. Often this is presented in the context of the internet of things (IoT) (Zanella et al., 2014; Wang et al., 2015a). The IoT can be described as "the pervasive presence around us of a variety of *things* or *objects* ... [which] are able to interact with each other and cooperate with their neighbors to reach common goals" (Atzori et al., 2010, p. 2787).

Parallel to the development of the sensor web other research has focused on the semantic web, as proposed by Berners-Lee et al. (2001). This is a response to the traditional way of using the web, where information is only available for humans to read. The semantic web is an extension of the internet which contains meaningful data that machines can interpret as well. Rather than publishing documents on the internet the semantic web contains linked data using the resource description framework (RDF) (Bizer et al., 2009).

Sheth et al. (2008) proposes to use semantic web technologies in the sensor web. This so-called semantic sensor web (SSW) builds on standards by OGC and the world wide web consortium (W3C) "to provide enhanced descriptions and meaning to sensor data" (Sheth et al., 2008, p.78). W3C responded to this development by developing a standard ontology for sensor data on the semantic web (Compton et al., 2012).

1.2 PROBLEM STATEMENT

The implementation of the sensor web is still in an early stage.

A number of companies and organisations still use their own custom APIs to connect sensors to the internet.

It is hard to find SOS services on the internet.

It has been argued that it is difficult to integrate sensor data from different sources to perform data fusion (Corcho and Garcia-Castro, 2010; Ji et al., 2014; Wang et al., 2015b). Data fusion is "a data processing technique that associates, combines, aggregates, and integrates data from different sources" (Wang et al., 2015a, p. 2).

1.3 RESEARCH QUESTION

This thesis aims develop a method that uses the semantic web to improve sensor data discovery as well as the integration and aggregation of sensor data from heterogeneous sources. The following question will be answered in this research: *How can the semantic web improve the discovery, integration and aggregation of distributed sensor data?*

2 RELATED WORK

a related work section in which the relevant literature is presented and linked to the project;

2.1 SEMANTIC SENSOR DATA MIDDLEWARE

Henson et al. (2009) and Pschorr (2013) suggest adding semantic annotations to a SOS which they call semantically enabled sos (Sem-SOS). In Sem-SOS the raw sensor data goes through a process of semantic annotating before it can be requested with a SOS service. The retrieved data is still an extensible markup language (XML) document, but with embedded semantic terminology as defined in an ontology model. The data retrieved from Sem-SOS is therefore semantically enriched.

Janowicz et al. (2013) has specified a method that uses a representational state transfer (REST)ful proxy as a façade for SOS. When a specific uniform resource identifier (URI) is requested the so-called semantic enablement layer (SEL) translates this to a SOS request, fetches the data and translates the results back to RDF. In this method the sensor data is converted to RDF on-the-fly. This allows the data to be interpreted by both humans and machines.

Atkinson et al. (2015) have identified that ”distributed heterogeneous data sources are a necessary reality in the case of widespread phenomena with multiple stakeholder perspectives” (Atkinson et al., 2015, p.129). Therefore, they propose that methods should be developed to move away from the traditional dataset centric approaches and towards using linked data for cataloguing. This has the potential to bring together data and knowledge from different areas of research about the same (or similar) features-of-interest. It is also argued that using both linked data services and data-specific services could ease the transition into the linked data world.

2.2 SENSOR DATA ONTOLOGIES

W3C has developed an ontology based on the stimulus-sensor-observation pattern (Compton et al., 2012).

Hu et al. (2014) has reviewed a number of metadata models (including SensorML and semantic sensor network ontology (SSNO)). They argue that all of the current metadata models are not sufficient for sensor data discovery. Therefore, they propose a metadata model that ”reuses and extends the existing sensor observation-related metadata standards” (Hu et al., 2014, p. 10546).

Cox (2015a) has been working on an improved semantic ontology based on O&M.

2.3 SENSOR DATA AGGREGATION

Stasch et al. (2011b) propose to aggregate sensor data based on the geometry of sampling features. Stasch et al. (2011a) proposes a web processing service (WPS) that takes sensor data right from a SOS service in order to aggregate it. The approach by Stasch et al. (2011b) takes sensor data as input that is already published on the semantic web.

Stasch et al. (2014) argue that in order for automatic aggregation to work there needs to be semantics on which kind of aggregation methods are appropriate for a specific kind of sensor data. This requires a formalisation of expert knowledge which they call semantic reference systems.

2.4 CONCLUSION

Sem-SOS (Henson et al., 2009; Pschorr, 2013) as well as SEL (Janowicz et al., 2013) focus on combining the sensor web with the semantic web, but do not address the integration and aggregation of sensor data. Similarly, Atkinson et al. (2015) proposes to expose sensor data to the semantic web in order to find other kinds of related data about the same feature-of-interest. Data that is collected from another area of research for example. Also Atkinson et al. (2015) does not mention the integration of sensor data from heterogeneous sources. Stasch et al. (2011b) and Stasch et al. (2011a) suggest interesting methods for aggregating sensor data based on features-of-interest. However, also these studies use sensor data from a only single source into account. Moreover, Corcho and Garcia-Castro (2010) and Ji et al. (2014) argue that methods for integration and fusion of sensor data on the semantic web is still an area for future research. Data fusion is "a data processing technique that associates, combines, aggregates, and integrates data from different sources" (Wang et al., 2015a, p. 2). This thesis therefore focuses on the discovery, integration and aggregation of sensor data, building on some of the principles proposed by related research discussed in this chapter.

The idea by Jones et al. (2014) of delivering data to users through a service with which they are already familiar is very appealing, because it would enable sensor data to be immediately used in any existing geographical information system (GIS). This is also suggested by Atkinson et al. (2015) to ease the transition to the linked data world. However, current research has mainly been concerned with static geographic data, not with (aggregated) sensor data. Therefore, this thesis aims to provide the service for integrating and aggregating sensor data as a WPS.

3 RESEARCH OBJECTIVES

the research objectives and/or research questions are clearly defined, along with the scope (ie what you will not be doing);

3.1 RESEARCH QUESTION

The main question this thesis will try to answer is:
How can the semantic sensor web improve the discovery, integration and aggregation of distributed sensor data?

To answer the main question a number of sub-questions need to be answered:

- How can sensor metadata be retrieved from a SOS and published on the semantic web with links to features-of-interest in an automated process?
- How can aggregation methods be represented on the semantic web to formalise expert knowledge and prevent meaningless aggregation?
- What are the advantages and disadvantages of integrating sensor data from different sources?
- To what extent can already existing standards for retrieving geographic data be used for a service that supplies integrated and aggregated sensor data?

3.2 OBJECTIVES

This thesis explores a method to store metadata of sensors on the semantic web, and to link it to real world features-of-interest and appropriate methods for aggregation. This should improve the discovery of sensor data through links to other related data on the internet.

To improve the integration of sensor data a middleware architecture will be developed that can return sensor data for features-of-interest from different sources. The returned sensor data will be aggregated. Only appropriate methods of aggregation are offered for each kind of observations, based on a formalisation of expert knowledge on the semantic web.

3.3 SCOPE

4 METHODS

overview of the methodology to be used;

4.1 SENSOR OBSERVATION SERVICE

Retrieve sensor metadata from the sensor observation service. There are a number of different requests that can be made: GetCapabilities, DescribeSensor and GetObservation. These requests can be made as a hypertext transfer protocol (HTTP) GET request or a HTTP POST request. The response is an XML document using the O&M (for GetObservation) or SensorML (for DescribeSensor).

4.2 RESOURCE DESCRIPTION FRAMEWORK

Publishing static geographic data on the semantic web requires a conversion of Shapefile to RDF. First the Shapefile is loaded into a Postgis database. After that a Python script retrieves the records from the database, maps it to an ontology and writes it to an RDF file. The final step is to publish the RDF online (Missier, 2015).

The sensor metadata is also being published on the semantic web. To do this an XML document is automatically retrieved from a SOS by a Python script. This script then extracts the relevant data from the XML and maps it to an ontology. It outputs an RDF file that will be published online. When new sources of sensor data are added the RDF document will be updated.

4.3 ONTOLOGY MAPPING

To publish data on the semantic web ontologies are required to specify the different classes and their relations. An ontology for static geographic data has to be connected to an ontology for sensor metadata. For this the ontologies by Cox (2015b) will be used: web ontology language (OWL) for observations and OWL for sampling features.

4.4 SENSOR DATA AGGREGATION

There are many different ways to aggregate sensor data, for example by taking the minimum value, the maximum value, the average value, the sum, etc. In order to determine which method of aggregation is applicable for a specific kind of sensor data the sensor metadata will contain links to appropriate aggregation methods. However, which methods are appropriate should be based on expert knowledge. Therefore, this requires a literature analysis.

5 PLANNING

time planning—having a Gantt chart is probably a better idea then just a list;

5.1 GANTT CHART

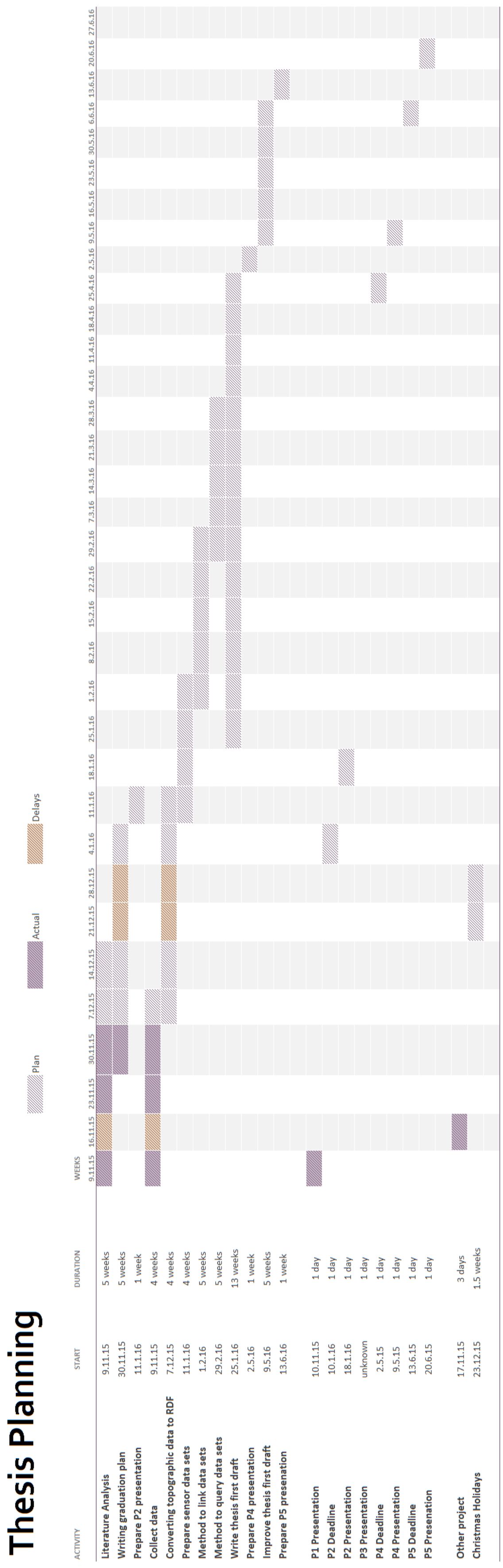

Figure 5.1: GANTT chart showing the planning of the thesis

6 TOOLS AND DATA

since specific data and tools have to be used, it's good to present these concretely, so that the mentors know that you have a grasp of all aspects of the project;

6.1 DATA

Topographic data of neighbourhoods, city districts, municipalities and provinces Air quality sensor data from the Dutch national institute for public health and the environment (RIVM) (http://inspire.rivm.nl/sos/) and from the Belgian interregional environment agency (IRCEL-CELINE) (http://sos.irceline.be/).

6.2 DATABASE

A Postgres database will be used with the Postgis extension.

6.3 SERVER

Prototyping will be done using a localhost at first, but in the end it could be hosted on the university server.

6.4 PROTOTYPE

- The Python programming language will be used for scripting a prototype.
- Psycopg2 will be used to connect a Python script to a Postgres database.
- Python's Request library will be used for making HTTP POST and GET requests.
- For working with XML Python's xml package will be used.
- To create RDF documents the Python library RDFLib will be used.

BIBLIOGRAPHY

Atkinson, R. A., Taylor, P., Squire, G., Car, N. J., Smith, D., and Menzel, M. (2015). Joining the dots: Using linked data to navigate between features and observational data. In *Environmental Software Systems. Infrastructures, Services and Applications*, pages 121–130. Springer.

Atzori, L., Iera, A., and Morabito, G. (2010). The internet of things: A survey. *Computer networks*, 54(15):2787–2805.

Berners-Lee, T., Hendler, J., Lassila, O., et al. (2001). The semantic web. *Scientific american*, 284(5):28–37.

Bizer, C., Heath, T., and Berners-Lee, T. (2009). Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, pages 205–227.

Botts, M., Percivall, G., Reed, C., and Davidson, J. (2007). Ogc sensor web enablement: Overview and high level architecture. OGC document 06-021r1.

Botts, M., Percivall, G., Reed, C., and Davidson, J. (2008). Ogc sensor web enablement: Overview and high level architecture. In *GeoSensor networks*, pages 175–190. Springer.

Compton, M., Barnaghi, P., Bermudez, L., GarcíA-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., et al. (2012). The ssn ontology of the w3c semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17:25–32.

Corcho, O. and Garcia-Castro, R. (2010). Five challenges for the semantic sensor web. *Semantic Web-Interoperability, Usability, Applicability*, 1.1(2):121–125.

Cox, S. J. D. (2015a). Observations and sampling. [online] https://www.seegrid.csiro.au/wiki/AppSchemas/ObservationsAndSampling [accessed on December 1st, 2015].

Cox, S. J. D. (2015b). Ontology for observations and sampling features, with alignments to existing models.

Henson, C., Pschorr, J. K., Sheth, A. P., Thirunarayan, K., et al. (2009). Semsos: Semantic sensor observation service. In *Collaborative Technologies and Systems, 2009. CTS'09. International Symposium on*, pages 44–53. IEEE.

Hu, C., Guan, Q., Chen, N., Li, J., Zhong, X., and Han, Y. (2014). An observation capability metadata model for eo sensor discovery in sensor web enablement environments. *Remote Sensing*, 6(11):10546–10570.

INSPIRE (2014). Guidelines for the use of observations & measurements and sensor web enablement-related standards in inspire annex ii and iii data specification development.

INSPIRE (2015). Inspire roadmap.

ISO (2011). Iso 19156:2011; geographic information – observations and measurements.

Janowicz, K., Broring, A., Stasch, C., Schad, S., Everding, T., and Llaves, A. (2013). A restful proxy and data model for linked sensor data. *International Journal of Digital Earth*, 6(3):233–254.

Ji, C., Liu, J., and Wang, X. (2014). A review for semantic sensor web research and applications. *Advanced Science and Technology Letters*, 48:31–36.

Jones, J., Kuhn, W., Keßler, C., and Scheider, S. (2014). Making the web of data available via web feature services. In *Connecting a Digital Europe Through Location and Place*, pages 341–361. Springer.

Missier, G. A. (2015). Towards a web application for viewing spatial linked open data of rotterdam. Master's thesis, Delft University of Technology.

Moir, E., Moonen, T., and Clark, G. (2014). What are future cities: Origins, meanings and uses.

Percivall, G. (2015). Ogc smart cities spatial information framework. OGC Internal reference number: 14-115.

Pschorr, J. K. (2013). Semsos: an architecture for query, insertion, and discovery for semantic sensor networks. Master's thesis, Wright State University.

Sheth, A., Henson, C., and Sahoo, S. S. (2008). Semantic sensor web. *IEEE Internet Computing*, 12(4):78–83.

Stasch, C., Autermann, C., Foerster, T., and Pebesma, E. (2011a). Towards a spatiotemporal aggregation service in the sensor web. poster presentation. In *The 14th AGILE International Conference on Geographic Information Science*.

Stasch, C., Schade, S., Llaves, A., Janowicz, K., and Bröring, A. (2011b). Aggregating linked sensor data. In Taylor, K., Ayyagari, A., and de Roure, D., editors, *Proceedings of the 4th International Workshop on Semantic Sensor Networks*, page 46.

Stasch, C., Scheider, S., Pebesma, E., and Kuhn, W. (2014). Meaningful spatial prediction and aggregation. *Environmental Modelling & Software*, 51:149–165.

Wang, M., Perera, C., Jayaraman, P. P., Zhang, M., Strazdins, P., and Ranjan, R. (2015a). City data fusion: Sensor data fusion in the internet of things.

Wang, X., Zhang, X., and Li, M. (2015b). A review of studies on semantic sensor web. *Advanced Science and Technology Letters*, 83:94–97.

Zanella, A., Bui, N., Castellani, A., Vangelista, L., and Zorzi, M. (2014). Internet of things for smart cities. *Internet of Things Journal, IEEE*, 1(1):22–32.

COLOPHON

This document was typeset using LaTeX. The document layout was generated using the `arsclassica` package by Lorenzo Pantieri, which is an adaption of the original `classicthesis` package from André Miede.