AGGREGATION OF SEMANTIC SENSOR DATA

Graduation proposal

by

Ivo de Liefde

December 3, 2015

CONTENTS

ACRONYMS

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

Sem-SOS (Henson et al., 2009; Pschorr, 2013) as well as semantic enablement layer (SEL) (Janowicz et al., 2013) focus on publishing meaningful sensor data, but do not address the integration and aggregation of sensor data. Stasch et al. (2011b) suggests an interesting method for aggregating sensor data and publishing it on the semantic web based on spatial 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 on the semantic web as linked data. This leads to a number of issues regarding the validity of the links after the aggregation process. 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). However, Jones et al. (2014) has been mainly concerned with static geographic data, not with (aggregated) sensor data.

This thesis aims to build on the recent literature by creating 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 or 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.

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 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.

2.2 SENSOR DATA ONTOLOGIES

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

Cox (2015) 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. (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.

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 will explore a method that stores metadata of sensors on the semantic web, and links it to real world features of interest and to 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. The returned sensor data will be aggregated. Only appropriate methods of aggregation are offered for each kind of observations, based on formalised expert knowledge on the semantic web.

3.3 SCOPE

4 METHODS

overview of the methodology to be used;

5 PLANNING

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

5.1 GANTT CHART

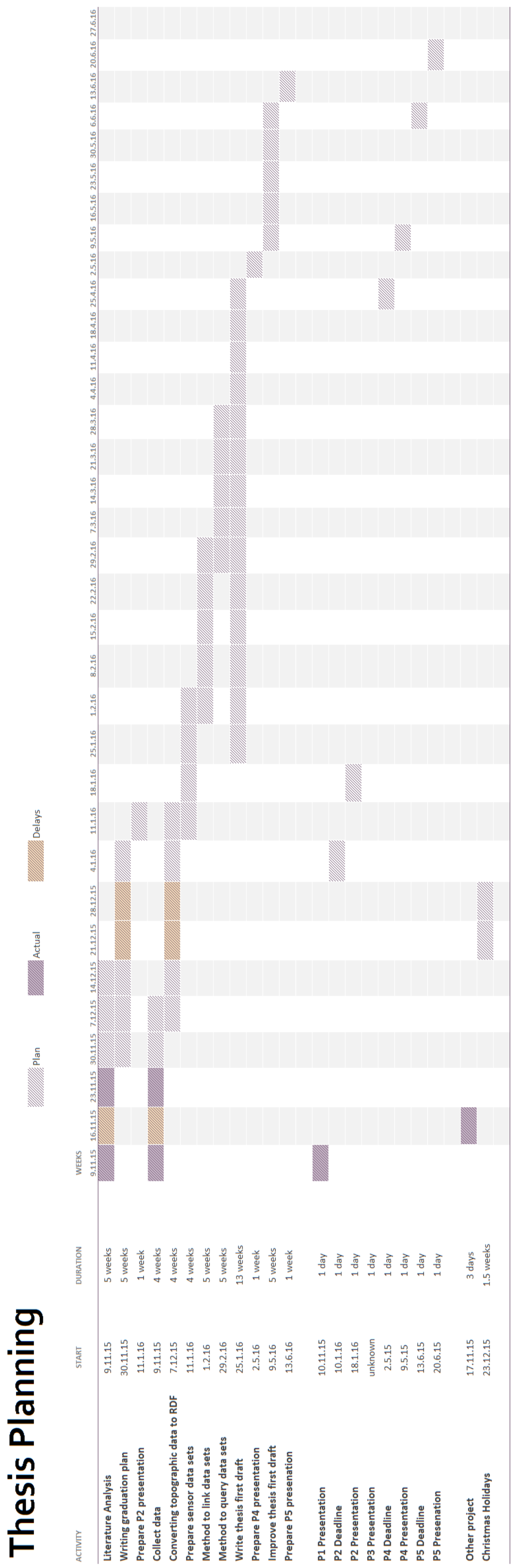

Figure 5.1: GANTT chart showing the planning of the thesis

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;

BIBLIOGRAPHY

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

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.

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.

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.