

DISCOVERY, INTEGRATION AND  
AGGREGATION OF SENSOR DATA USING THE  
SEMANTIC WEB

Graduation proposal

by

Ivo de Liefde

December 15, 2015

# CONTENTS

1	INTRODUCTION	1
1.1	Background . . . . .	1
1.2	Problem statement . . . . .	2
1.3	Scientific relevance . . . . .	3
1.4	Research question . . . . .	3
2	RELATED WORK	4
2.1	Sensor data catalogue service . . . . .	4
2.2	Semantic sensor data middleware . . . . .	4
2.3	Sensor data ontologies . . . . .	5
2.4	Sensor data aggregation . . . . .	7
2.5	Discussion . . . . .	7
3	RESEARCH OBJECTIVES	9
3.1	Research question . . . . .	9
3.2	Objectives . . . . .	9
3.3	Scope . . . . .	9
4	METHODS	11
4.1	Sensor observation service . . . . .	11
4.2	Resource description framework . . . . .	11
4.3	Ontology mapping . . . . .	12
4.4	Sensor data aggregation . . . . .	12
5	PLANNING	14
5.1	Deadlines . . . . .	14
5.2	GANTT Chart . . . . .	14
6	TOOLS AND DATA	16
6.1	Data . . . . .	16
6.2	Database . . . . .	16
6.3	Server . . . . .	16
6.4	Prototype . . . . .	16

## ACRONYMS

API	application program interface	2
EU	european union	1
GIS	geographical information system	8
HTTP	hypertext transfer protocol	11
INSPIRE	infrastructure for spatial information in Europe	1
IoT	internet of things	1
IRCEL-CELINE	Belgian interregional environment agency	16
IRI	international resource identifier	11
ISO	international organisation for standardisation	1
OGC	open geospatial consortium	1
O&M	observations and measurements	1
OWL	web ontology language	12
RDF	resource description framework	1
REST	representational state transfer	4
RIVM	Dutch national institute for public health and the environment	2
SEL	semantic enablement layer	4
Sem-SOS	semantically enabled sos	4
SensorML	sensor modelling language	1
SIR	sensor instance registry	2
SOR	sensor observable registry	2
SOS	sensor observation service	1
SPARQL	sparql protocol and rdf query language	1
SSNO	semantic sensor network ontology	5
SSW	semantic sensor web	2
SWE	sensor web enablement	1
UML	unified modeling language	12
URI	uniform resource identifier	2
URL	uniform resource locator	8
W3C	world wide web consortium	2
WCS	web coverage service	2
WFS	web feature service	2
WMS	web map service	2
WPS	web processing service	7
XML	extensible markup language	2

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 questions 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 sensor data through standard protocols, 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 understand as well. Rather than publishing documents on the internet the semantic web contains linked data using the resource description framework (RDF), also known as the 'web of data' (Bizer et al., 2009). Data in RDF can be queried using the sparql protocol and rdf query language (SPARQL) at so called SPARQL endpoints. Originally, the semantic web intended to add metadata on the internet (Lassila and Swick, 1999). However, today it is being used for linking

any kind of data from one source to another in a meaningful way ([Cambridge Semantics, 2015](#)).

[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 creating a standard ontology for sensor data on the semantic web ([Compton et al., 2012](#)).

## 1.2 PROBLEM STATEMENT

Finding sensor data that can be retrieved using a [SOS](#) is not easy. The implementation of the sensor web is still in an early stage. At the moment there are only a limited number of [SOS](#) implementations available on the web and they contain a limited amount of data. In the Netherlands the [SOS](#) by the Dutch national institute for public health and the environment ([RIVM](#)) is one of the first ones to be developed. It has only recently been created and contains data on air quality. A number of other organisations still use a custom application program interface ([API](#)) to retrieve data from sensors connected to the internet. It has been researched to what extent a catalogue services could be useful for discovering sensor data: the sensor instance registry ([SIR](#)) ([Jirka and Nüst, 2010](#)) and the sensor observable registry ([SOR](#)) ([Jirka and Bröring, 2009](#)). Catalogue services have already been available for example for the web map service ([WMS](#)), web feature service ([WFS](#)) or web coverage service ([WCS](#)) ([Nebert et al., 2007](#)). However, for discovering sensor data from the [SOS](#) services used in this paper no register or catalogue service has been implemented.

It has been argued that one of the challenges of using sensor data is the difficulty of integrating 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](#)). Even if the sources comply with the [SWE](#) standards it is challenging, since the data can be of a different scale, both in time and space.

A question that comes to mind is to what extent the semantic web could be a better solution for publishing sensor data than using a [SOS](#). The geoweb has some very good qualities, such as very structured approaches in which (sensor) data can be retrieved using well defined services. This is different from for example web pages where content can be completely unstructured. The response of a [SOS](#) also contains some semantics about sensor data. There can be x-links inside the extensible markup language ([XML](#)) with uniform resource identifier ([URI](#))s that point to semantic definitions of objects. Still, the semantic web could be beneficial for the geoweb as it is machine understandable which could be useful for automatic integration and aggregation. It also contains links to other relevant data which could make discovering sensor data more easy.

### 1.3 SCIENTIFIC RELEVANCE

Sensor data ties together many different fields of research. On the one hand there is research on how to create the most efficient sensor networks that uses the least amount of power to transfer the observed data over long distances. This involves academic fields such as mathematics, physics and electrical engineering. On the other hand there is research that uses sensor data to gain insights into real world phenomenon. This involves academic fields such as geography, environmental studies and urbanism. In order to connect these scientific fields, research has been focused on the use of computer science and standardisation for transferring sensor data over the internet.

In the future more sensor data is expected to be produced, on the one hand by experts because of European legislation ([INSPIRE](#)). However, on the other hand also non-experts will be involved more often via smart cities and [IoT](#) developments where users or consumer electronics produce sensor data. This vast amount of data could be very useful for academic research, provided researchers are able to find the data they need online and are able to integrate and aggregate data from heterogeneous sources. Publishing sensor metadata on the semantic web could make it easier to find what you need through related data on the internet. A [SOR](#) or [SIR](#) is only useful for users if it is already known in advance where to find a specific catalogue service, as content inside the service cannot be linked to from other parts of the web. It is also dependent on whether people producing data have invested time and effort to register their sensors to such a service.

### 1.4 RESEARCH QUESTION

This thesis aims to 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 number of research topics are relevant for this thesis: how to use existing standards for publishing sensor data to the semantic web, developing ontologies that are suitable for many different kinds of sensor data and how to aggregate sensor data based on features-of-interest and time. This chapter discusses the recent relevant literature on these topics.

### 2.1 SENSOR DATA CATALOGUE SERVICE

The [SOR](#) is "a web service interface for managing the definitions of phenomena measured by sensors as well as exploring semantic relationships between these phenomena" ([Jirka and Bröring, 2009](#), p. vi). This is a web service developed by [OGC](#) to enable semantic reasoning on sensor networks, especially concerning phenomenon definitions. This should make it easier to discover sensors that observe a certain phenomenon and to interpret sensor data.

Another web service interface specification by [OGC](#) is [SIR](#). [SIR](#) is aimed at "managing the metadata and status information of sensors" ([Jirka and Nüst, 2010](#), p. xii). The goal of this web service is to close the gap between metadata models based on [SensorML](#), which is used in [SWE](#), and the metadata model used in [OGC](#) catalogue services. Furthermore, it provides functionalities to discover sensors, to harvest sensor metadata from a [SOS](#), to handle status information about sensors and to link [SIR](#) instances to [OGC](#) catalogue services.

### 2.2 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 [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 [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

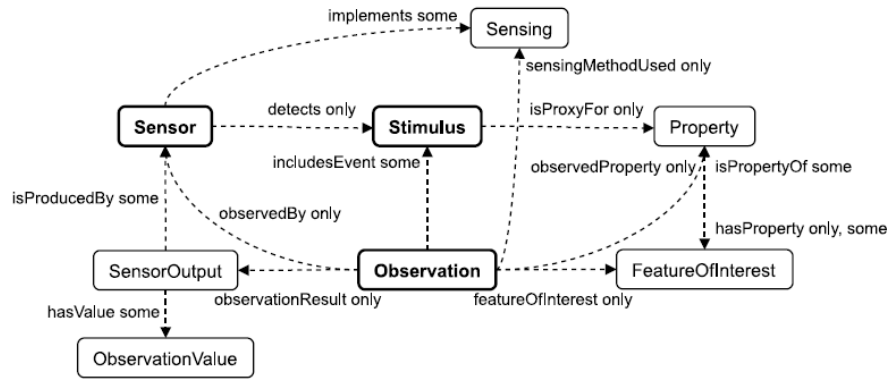


Figure 2.1: The stimulus-sensor-observation pattern (Compton et al., 2012, p. 28)

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.

## 2.3 SENSOR DATA ONTOLOGIES

### SEMANTIC SENSOR NETWORK ONTOLOGY

W3C has developed an ontology for sensors and observations called the semantic sensor network ontology (SSNO). This ontology aims to address semantic interoperability on top of the syntactic operability that the SWE standards provide. To accommodate different definitions of the same concepts the broadest definitions have been used. Depending on the interpretation these can be further defined with subconcepts. The SSNO is based on the stimulus-sensor-observation pattern, describing the relations between a sensor, a stimulus and observations (figure 2.1). Sensors are defined as “physical objects ... that observe, transforming incoming stimuli ... into another, often digital, representation”, stimuli are defined as “changes or states ... in an environment that a sensor can detect and use to measure a property” and observations are defined as “contexts for interpreting incoming stimuli and fixing parameters such as time and location” (Compton et al., 2012, p 28). The ontology can be used to model sensor networks from four different perspectives (sensor, observation, system, and feature & property), which have been discussed together with additional relevant concepts.

### OBSERVATION CAPABILITY METADATA MODEL

Hu et al. (2014) have reviewed a number of metadata models (including SensorML and SSNO) for the use of earth observation (including remote sensing). They argue that all of the current metadata models are not sufficient for sensor data discovery. This conclusion is based on an evaluation of six criteria. Three steps have been identified in the process of obtaining relevant sensor data for earth observation, which have been used to derive criteria for their evaluation framework. These steps are sensor filtration, sensor optimisation and sensor dispatch. The filtration of sensors should result in a set of sensors that meets the requirements of the application: it should measure the right phenomenon, be active, be inside the spatial and temporal range,



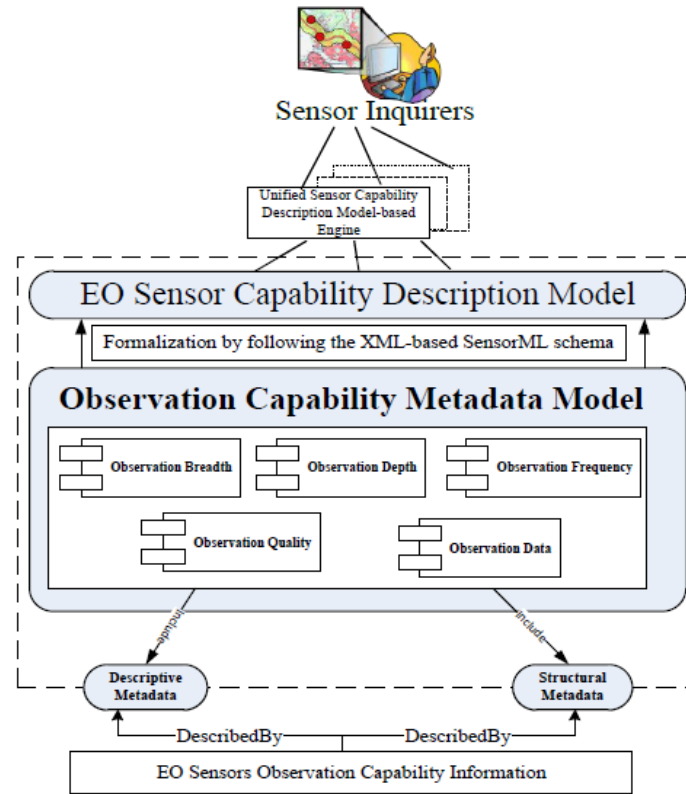


Figure 2.2: Architecture of the sensor discovery system (Hu et al., 2014, p. 10553)

and have a certain sample interval. In sensor optimisation the selected sensors should be combined to complement or enhance each other. To do this, the observation quality, coverage and application is relevant. In the last step - sensor dispatch - the data should be retrieved, stored transmitted. In every evaluated model sensors can describe in different ways or only partially, which affects the outcome of the sensor dispatch.

Therefore, a metadata model is proposed that “reuses and extends the existing sensor observation-related metadata standards” (Hu et al., 2014, p. 10546). It is composed of five modules: observation breadth, observation depth, observation frequency, observation quality and observation data. They should be derived from metadata elements described using the Dublin Core metadata element set. These five modules can then be formalized following the [SensorML](#) schema which can be queried by users via the ‘Unified Sensor Capability Description Model-based Engine’ (figure 2.2).

#### OM-LITE & SAM-LITE ONTOLOGIES

Cox (2015b) has been working on a new semantic ontology based on O&M. Previous efforts, such as the [SSNO](#) have been using pre-existing ontologies and frameworks. However, there are already many linked data ontologies that could be useful for describing observation metadata, such as space and time concepts. Also, the [SSNO](#) does not take sampling features into account. The proposed om-lite ontology defines the concepts from O&M regarding observations, while the sam-lite defines the sampling feature concepts. The author also provides a mapping of the [SSNO](#) to om-lite.

Cox (2015b) also describes how the PROV ontology (Lebo et al., 2013) can be directly used inside om-lite. The PROV ontology is “concerned with the

production and transformation of Entities through time-bounded Activities, under the influence or control of Agents” (Cox, 2015b, p. 12). This is a very convenient ontology for modelling real world entities, such as sensors, observation processes and sampling processes.

## 2.4 SENSOR DATA AGGREGATION

Sensor data aggregation can be performed for two purposes: to reduce the energy constraint of sensor networks (Korteweg et al., 2007) or to sample a feature-of-interest in space and/or time (INSPIRE, 2014). Sampling is performed when a feature-of-interest is not accessible, in which case “observations are made on a subset of the complete feature, with the intention that the sample represents the whole” (Cox, 2015a). Stasch et al. (2011a) proposes a web processing service (WPS) that retrieves sensor data 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.

Ganesan et al. (2004) stresses that spatio-temporal irregularities are fundamental to sensor networks. Irregular sampling can have a potentially large influence on the accuracy of the aggregated outcome. For example, averaging sensor data from a sampling feature that is being sampled densely in some parts and more sparsely in other parts could lead to inaccurate results. In this case the values of the densely sampled area should have a lower weight than the values from the sparsely sampled area. The same holds true for temporal irregularities (Ganesan et al., 2004). Also, 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.

## 2.5 DISCUSSION

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. Atkinson et al. (2015) do not mention the integration of sensor data from heterogeneous sources either. 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).

Jirka and Nüst (2010) and Jirka and Bröring (2009) present methods for including SOS services in an OGC catalogue service using SOR and SIR. Making sensor metadata available in a catalogue service will improve the discovery. However, discovery through the semantic web is likely to be more

effective, since links can be created towards the sensor data from many different sources of related information. Another advantage is that links can be created by everybody that publishes linked data on the web, allowing sensor data to be used for implementations that were not identified beforehand by the publisher. Also, the semantic web will be easier to access, while the catalogue service can only be accessed on at a certain uniform resource locator ([URL](#)) which has to be known to potential users. This thesis therefore focuses on the discovery, integration and aggregation of sensor data using the semantic web.

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

## 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 automatically published on the semantic web?
- How can metadata on the semantic web be linked to relevant features-of-interest using existing vocabularies?
- How can incoming links be automatically generated from DBPedia?
- How can aggregation methods be represented on the semantic web to formalise expert knowledge and prevent meaningless aggregation?
- 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. The service can be accessed as a [WPS](#).

## 3.3 SCOPE

The focus of this thesis is to explore the role of the semantic web in discovering, integrating and aggregating sensor data from heterogeneous sources. Therefore, accepted [OGC](#) and [ISO](#) standards will be, such as [SOS](#), [SensorML](#) and [O&M](#). The om-lite and sam-lite ontologies will be used in combination

with geoSPARQL. An in-depth analysis of different upcoming standards lies outside the scope of this research.

# 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 Postgres database with the Postgis extension. After that a Python script retrieves the records from the database. Attributes of the records will be mapped to classes from predefined ontologies. Then the script creates an [RDF](#) graph and serialises it to a certain [RDF](#) language. This is written to a file. The final step is to publish the [RDF](#) on the web and create a [SPARQL](#) endpoint to query the data ([Missier, 2015](#)).

In [RDF](#) data is stored in so-called 'triples'. These triples are structured as: subject, predicate and object [Berners-Lee et al. \(2001\)](#). The subject and the object are things and the predicate is the relation between these two things. Three types of data can make up these triples. The first type is an international resource identifier ([IRI](#)). This is a reference to a resource and can be used for all positions of the triple. A [URL](#) is an example of an [IRI](#), but [IRIs](#) can also refer to resources without stating where a location or how it can be accessed. It is a generalisation of an [URI](#), also allowing non-ASCII characters. The second type of data is a literal. A literal is a value which is not an [IRI](#), such as strings, numbers or dates. These values can only be used as object in a triple. Sometimes it's useful to refer to things without assigning them with a global identifier. The third type is the blank node and can be used as an subject or object without using an [IRI](#) or literal ([Manola et al., 2014](#)).

There are a number of different languages for writing down these triples (serialisation), such as [XML](#) ([Gandon and Schreiber, 2014](#)), [N3](#) ([Berners-Lee and Connolly, 2011](#)) and [Turtle](#) ([Beckett et al., 2014](#)).

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](#) documents will be updated.

### 4.3 ONTOLOGY MAPPING

The unified modeling language ([UML](#)) diagram (figure [4.1](#)) describes different components of a [SOS](#). The [SOS](#) has a number of metadata attributes such as the service provider's details (including contact information), its spatial and temporal extent (spatialFilter & temporalFilter) and the capabilities to query a subset of this extent. It receives data from a sensor which makes observations. An observation can be defined as "an action whose result is an estimate of the value of some property of the feature-of-interest, obtained using a specified procedure" ([Cox, 2015a](#)). The sensor is placed at a sampling point. The sampling point is part of a sampling features which intends to resemble the feature-of-interest. In the case of air quality the feature-of-interest is the bubble of air surrounding the sensor, therefore the sampling point equals the feature-of-interest ([INSPIRE, 2014](#)). The design is that an observation of the sampling feature describes the feature-of-interest through measuring one of its properties. The measurement procedure is described by a short string of text, input and output parameters and the units of measurement of the output. The relation between feature-of-interest and administrative units is added to improve the discovery of sensor data on the semantic web.

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. From the [UML](#) diagram in figure [4.1](#) the classes Observation, Process, ObservedProperty and FeatureOfInterest can be mapped to classes belonging to web ontology language ([OWL](#)) for observations ([Cox, 2015c](#)). SamplingFeature and Sampling point can be mapped to classes from [OWL](#) for sampling features [Cox \(2015d\)](#). GeoSPARQL can be used for the administrativeUnit class ([Perry and Herring, 2011](#)) and the [SSNO](#) for the sensor and sensor observation service classes ([W3C Semantic Sensor Network Incubator Group, 2011](#)).

### 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.

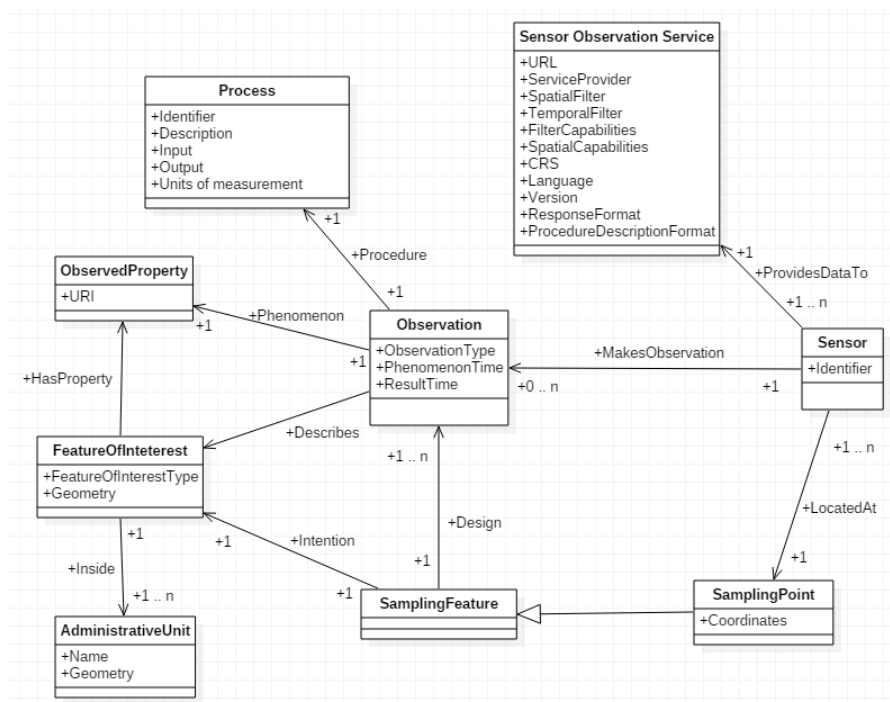


Figure 4.1: UML diagram of sensor observations service



# 5 | PLANNING

This chapter provides an overview of the deadlines for the thesis, ranging from the start of the thesis on November tenth, 2015 to the final deadline on June twentieth, 2016. Based on these deadlines a planning has been made, which breaks down the total workload into twelve parts. A GANTT chart is presented as a graphical presentation of the thesis planning.

## 5.1 DEADLINES

For the planning of the thesis a number of deadlines are import. At five moments in time the status of the thesis has to be presented and at three of these moments it is required to hand in a report. The five deadlines are referred to as P1 to P5 in the graduation manual. P1 took place on November tenth, 2015. The general idea for the thesis was presented here and a number of students and staff was present to provide feedback. This document is part of the deliverables for P2. At P2 the research proposal for the thesis needs to be handed in. This contains the research question, its relevance, a literature analysis of related work, a description of the methods, a planning and an overview of the tools and data that will be used. The research proposal is due January eleventh, 2016. The P2 presentation is scheduled on January eighteenth, 2016. Preliminary results are presented at P3, for which the date is still unknown. The P4 deadline is on May second, 2016. This is the deadline for handing in the first draft of the thesis report. It should also be presented on May seventh, 2016. The final deadline (P5) is on the thirteenth of June, 2016. The thesis outcomes will be presented on June twentieth, 2016.

## 5.2 GANTT CHART

# Thesis Planning

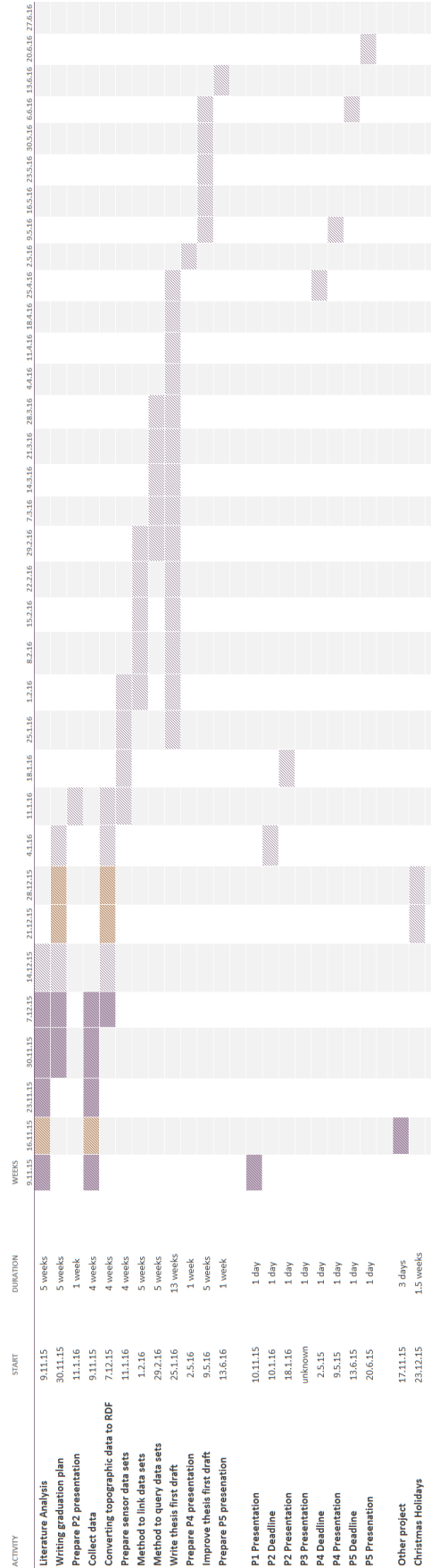


Figure 5.1: GANTT chart showing the planning of the thesis

# 6 | TOOLS AND DATA

## 6.1 DATA

Topographic data of neighbourhoods, city districts, municipalities and provinces

Air quality sensor data from the [RIVM](http://inspire.rivm.nl/sos/) (<http://inspire.rivm.nl/sos/>) and from the Belgian interregional environment agency ([IRCEL-CELINE](http://sos.irceline.be/)) (<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.
- The scripts will be part of a [WPS](#) using [PyWPS](#)

## 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.
- Beckett, D., Berners-Lee, T., Prud’hommeaux, E., and Carothers, G. (2014). W3c rdf 1.1 turtle. [online] <http://www.w3.org/TR/turtle/> [accessed on December 9<sup>th</sup>, 2015].
- Berners-Lee, T. and Connolly, D. (2011). W3c notation3 (n3): A readable rdf syntax. [online] <http://www.w3.org/TeamSubmission/n3/> [accessed on December 9<sup>th</sup>, 2015].
- 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.
- Cambridge Semantics (2015). Introduction to the semantic web. [online] <https://www.cambridgesemantics.com/semantic-university/introduction-semantic-web> [accessed on December 8<sup>th</sup>, 2015].
- 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 1<sup>st</sup>, 2015].
- Cox, S. J. D. (2015b). Ontology for observations and sampling features, with alignments to existing models.

- Cox, S. J. D. (2015c). Owl for observations. [online] <http://def.seegrid.csiro.au/ontology/om/om-lite> [accessed on November 24<sup>th</sup>, 2015].
- Cox, S. J. D. (2015d). Owl for sampling features. [online] <http://def.seegrid.csiro.au/ontology/om/sam-lite> [accessed on November 24<sup>th</sup>, 2015].
- Gandon, F. and Schreiber, G. (2014). W3c rdf 1.1 xml syntax. [online] <http://www.w3.org/TR/rdf-syntax-grammar/> [accessed on December 9<sup>th</sup>, 2015].
- Ganesan, D., Ratnasamy, S., Wang, H., and Estrin, D. (2004). Coping with irregular spatio-temporal sampling in sensor networks. *ACM SIGCOMM Computer Communication Review*, 34(1):125–130.
- 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. [online] <http://inspire.ec.europa.eu/index.cfm/pageid/44> [accessed on December 2<sup>nd</sup>, 2015].
- ISO (2011). Iso 19156:2011; geographic information – observations and measurements. [online] [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=32574](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32574) [accessed on December 2<sup>nd</sup>, 2015].
- 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.
- Jirka, S. and Bröring, A. (2009). Ogc sensor observable registry discussion paper. Reference number: OGC 09-112.
- Jirka, S. and Nüst, D. (2010). Ogc sensor instance registry discussion paper. Reference number: OGC 10-171.
- 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.
- Korteweg, P., Marchetti-Spaccamela, A., Stougie, L., and Vitaletti, A. (2007). *Data aggregation in sensor networks: Balancing communication and delay costs*. Springer.
- Lassila, O. and Swick, R. R. (1999). Resource description framework (rdf) model and syntax specification. [online] <http://www.w3.org/TR/PR-rdf-syntax/> [accessed on December 8<sup>th</sup>, 2015].

- Lebo, T., Sahoo, S., and McGuinness, D. (2013). Prov-o: The prov ontology. [online] <http://www.w3.org/TR/prov-o/> [accessed on December 11<sup>th</sup>, 2015].
- Manola, F., Miller, E., and McBride, B. (2014). W3c rdf primer. [online] <http://www.w3.org/TR/rdf11-primer/> [accessed on December 9<sup>th</sup>, 2015].
- 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.
- Nebert, D., Whiteside, A., and Vretanos, P. (2007). Opengis catalogue services specification.
- Percivall, G. (2015). Ogc smart cities spatial information framework. OGC Internal reference number: 14-115.
- Perry, M. and Herring, J. (2011). Geosparql - a geographic query language for rdf data. [online] <http://www.opengeospatial.org/standards/geosparql> [accessed on December 9<sup>th</sup>, 2015].
- 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.
- W3C Semantic Sensor Network Incubator Group (2011). Semantic sensor network ontology. [online] <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn> [accessed on December 9<sup>th</sup>, 2015].
- 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 L<sup>A</sup>T<sub>E</sub>X. The document layout was generated using the `arsclassica` package by Lorenzo Pantieri, which is an adaption of the original `classicthesis` package from André Miede.