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

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science in Geomatics

by

Ivo de Liefde

June 2016

Ivo de Liefde: *Discovery, Integration and Aggregation of Sensor Data Using the Semantic Web* (2016)
This work is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit
http://creativecommons.org/licenses/by/4.0/.

The work in this thesis was made in the:

OTB - Research for the Built Environment
Faculty of Architecture & the Built Environment
Delft University of Technology

Supervisors: Drs. M. de Vries
dr.ir. B.M. Meijers

ABSTRACT

[Should fit on one page.]

Bacon ipsum dolour sit amet porchetta beef turkey, bacon turducken boudin hamburger venison ball tip. Brisket pork loin bresaola short loin ground round leberkas pastrami tongue jerky cow turducken beef ribs. Pork ribeye landjaeger prosciutto pig venison tenderloin. Swine beef ribs kielbasa, porchetta tenderloin salami venison pork belly tail. Bacon ipsum dolour sit amet porchetta beef turkey, bacon turducken boudin hamburger venison ball tip. Brisket pork loin bresaola short loin ground round leberkas pastrami tongue jerky cow turducken beef ribs. Pork ribeye landjaeger prosciutto pig venison tenderloin. Swine beef ribs kielbasa, porchetta tenderloin salami venison pork belly tail. Bacon ipsum dolour sit amet porchetta beef turkey, bacon turducken boudin hamburger venison ball tip. Brisket pork loin bresaola short loin ground round leberkas pastrami tongue jerky cow turducken beef ribs. Pork ribeye landjaeger prosciutto pig venison tenderloin. Swine beef ribs kielbasa, porchetta tenderloin salami venison pork belly tail.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

API Application Programming Interface . . . 2
CORINE Coordination of Information on the Environment . . . 17
DE-9IM Dimensionally Extended Nine-Intersection Model . . . 13
EEA European Environment Agency . . . vii
EU European Union . . . 1
GIS Geographical Information System . . . 17
HTTP HyperText Transfer Protocol . . . 10
INSPIRE Infrastructure for Spatial Information in Europe . . . 1
IoT Internet of Things . . . 1
IRCEL-CELINE Belgian interregional environment agency . . . vii
IRI International Resource Identifier . . . 12
ISO International Organisation for Standardisation . . . 1
OGC Open Geospatial Consortium . . . 1
O&M Observations and Measurements . . . 1
OWL Web Ontology Language . . . 1
PURL Persistent Uniform Resource Locator . . . 13
RDF Resource Description Framework . . . 1
REST Representational State Transfer . . . 5
RIVM Dutch national institute for public health and the environment . . . vii
SEL Semantic Enablement Layer . . . 5
Sem-SOS Semantically Enabled SOS . . . 5
SensorML Sensor Modelling Language . . . 1
SIR Sensor Instance Registry . . . 2
SOR Sensor Observable Registry . . . 2
SOS Sensor Observation Service . . . vii
SPARQL SPARQL Protocol and RDF Query Language . . . 1
SSNO Semantic Sensor Network Ontology . . . 6
SSW Semantic Sensor Web . . . 2
SWE Sensor Web Enablement . . . 1
UML Unified Modeling Language . . . 15
URI Uniform Resource Identifier . . . 3
URL Uniform Resource Locator . . . 9
W3C World Wide Web Consortium . . . 2
WCS Web Coverage Service . . . 2
WKT Well-Known Text . . . 12
WFS Web Feature Service . . . 2
WMS Web Map Service . . . 2

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 questions on how to structure it. This thesis aims to design a method to publish and link sensor metadata on the semantic web to improve the discovery, integration and aggregation of sensor data using SWE standards.

1.1 BACKGROUND

In 2008 the Open Geospatial Consortium (OGC) introduced a new set of standards called Sensor Web Enablement (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 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. The Web Ontology Language (OWL) is an extension of

RDF and was designed "to represent rich and complex knowledge about things, groups of things, and relations between things" [OWL working group, 2012]. Originally, the semantic web intended to add metadata to 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 Semantic Sensor Web (SSW) builds on standards by OGC and the World Wide Web Consortium (W_3C) "to provide enhanced descriptions and meaning to sensor data" [Sheth et al., 2008, p. 78]. W_3C 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 open standards 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 RIVM is one of the first ones to be developed. It has only recently been launched and contains data on air quality. A number of other organisations still use a custom Application Programming Interface (API) to retrieve data from sensors connected to the internet. The problem of these custom APIs is that it is very hard to create an application that automatically retrieves data from them, because they have not implemented standards regarding the content of their service, the metadata models behind it or the kind of requests that can be made. It forces the application to have knowledge built in on the specifics of the individual APIs that are being used.

It has been researched to what extent a catalogue service could be useful for discovering sensor data from a SOS using the web service interfaces Sensor Instance Registry (SIR) [Jirka and Nüst, 2010] and 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 the sensor data sources used in this paper no register or catalogue service has been implemented. Atkinson et al. [2015] also argues that catalogue services have a number of major disadvantages. It places a very high burden on the client to not only know where to find the catalogue service, but also to have knowledge on all kinds of other aspects (e.g. its organisation, access protocol, response format and response content) [Atkinson et al., 2015, p. 128]. Atkinson et al. suggest that linked data is therefore a much better solution for discovering sensor data.

However, for sensor data to be discovered on the semantic web there have to be inward links, from other sources linking towards the sensor (meta)data. Current research on the SSW has focused on publishing sensor data on the semantic web with links that point outwards [Atkinson et al., 2015; Janowicz et al., 2013; Pschorr, 2013]. This gives meaning to the data and is useful in order to work with the data, but it has a very limited effect on the discovery of the sensor data by others.

One of the challenges of using sensor data is the difficulty of integrating it 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 granularity, both in time and space. Spatio-temporal irregularities are a fundamental property of sensor data [Ganesan et al., 2004].

The question arises to what extent the semantic web could be a better solution for publishing sensor data than the current geoweb solutions like SOS. The geoweb has some very good qualities, such as very structured approaches through which (sensor) data can be retrieved using well defined services. These standardised services have been accepted by large organisations as OGC and ISO. Furthermore, they are often based on years of discussion. 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. Since data on the web has a distributed nature it can be questioned whether centralised catalogue services are feasible to create. It places a burden on the owner of the SOS to register with a catalogue service. Also, there could be multiple of these services on the web creating issue regarding the discovery of relevant catalogues. The semantic web could solve this issue by getting rid of the information silos and storing data directly on the web instead. This allows the interlinking and reuse of data on the web, which makes it easier to find related data. For automatic integration and aggregation it could be useful that the semantic web is machine understandable.

In conclusion, the problem to be addressed is the lack of knowledge on how to exploit the full potential of the sensor web using the semantic web. Creating the right links could greatly enhance the discovery, integration and aggregation of sensor data. However, there is no method yet to establish this linked metadata for sensors, while the standardised nature of a SOS should allow for generating it in an automated process. This thesis will create a design for such an automated process, research how to establish inward links and explore the advantages and disadvantages of publishing sensor metadata on the semantic web with a proof-of-concept implementation.

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 [Korteweg et al., 2007; Xiang et al., 2013]. 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, studies have 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 [Price Waterhouse Coopers, 2014]. Both experts and non-experts will be involved in this development. Experts will produce more data because of European legis-

lation (INSPIRE). Non-experts will be involved more often via smart cities and IoT developments where users or consumer electronics produce sensor data as well. 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. Having a automated process for this and being able to seamlessly integrate and aggregate data from different sources could be of great use for research such as van der Hoeven et al. [2014], Van der Hoeven and Wandl [2015] and Theunisse [2015]. They are examples of studies that try to understand phenomenon in the built environment using sensor data. Currently data collection and processing takes up a large part of the research, while with the implementation of SWE standards and the use of the semantic web this might be significantly reduced.

1.4 RESEARCH QUESTION

This thesis aims to design 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: *To what extent 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 geographical features 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.

Pschorr et al. [2010] has created a prototype that is able to find sensors from a SOS using linked data. The user can input a location and find sensors that are located nearby. They acknowledge the above mentioned advantages of linked data over a catalogue service. However, the method presented by Pschorr et al. [2010] is still limited to retrieving sensors from a single source in a buffer around a point location.

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. 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 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

Ontologies are necessary to provide meaning to data on the semantic web and to create semantic interoperability. Three recent efforts for developing a standard ontology for sensor data based on SWE standards will be discussed here.

2.3.1 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 they discuss together with additional relevant concepts.

2.3.2 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 were 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, and have a certain sample interval. In sensor optimisation the selected sen-

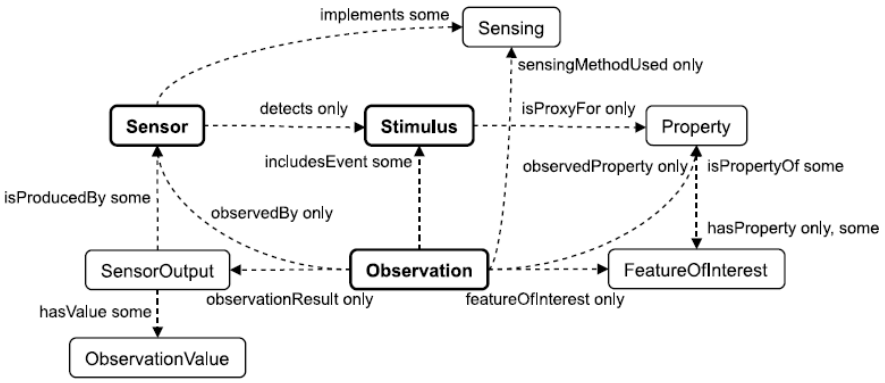

Figure 2.1: The stimulus–sensor–observation pattern [Compton et al., 2012, p. 28]

sors 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 and transmitted. In every evaluated model the same sensors can be described 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 formalised following the SensorML schema which can be queried by users via a 'Unified Sensor Capability Description Model-based Engine'.

2.3.3 Om-lite & sam-lite ontologies

Cox [2015b] has been working on new semantic ontologies 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. Therefore, Cox [2015b] proposes two new ontologies: OWL for observations or om-lite [Cox, 2015a], which defines the concepts from O&M regarding observations and OWL for sampling features or sam-lite, which defines the sampling feature concepts [Cox, 2015d]. A mapping of the SSNO to om-lite is also provided.

Cox [2015b] 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. Many other ontologies could be implemented in combination with om-lite and sam-lite, depending on the kind of observations that are being modelled and the data publisher's preference.

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 based on features-of-interest. The approach by Stasch et al. [2011b] is similar, but 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 feature-of-interest that is being sampled densely in some parts and more sparsely in other parts could lead to inaccurate results. To counter this 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. Not all kinds of aggregation are meaningful (e.g. taking the sum of temperature values). This requires a formalisation of expert knowledge which they call semantic reference systems.

3 METHODS

A number of studies related to this thesis have been reviewed in Chapter 2. This chapter discusses why the semantic web will be used for linking sensor metadata and which methods will be used to achieve this. The SWE standards, the om-lite and sam-lite ontologies, and RDF will be described.

3.1 SENSOR METADATA ON THE SEMANTIC WEB

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 can be collected for another area of research. However, Atkinson et al. [2015] do not mention the integration of complementary 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 only a 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 requested at a certain Uniform Resource Locator (URL) which has to be known to potential users.

Since data on the web has a distributed nature it can be questioned whether centralised catalogue services are feasible to create. It places a burden on the owner of the SOS to register with a catalogue service. Also, there could be multiple of these services on the web creating issue regarding the discovery of relevant catalogues. The semantic web could solve this issue by getting rid of the 'dataset-centric' approach and adding metadata directly to the web instead.

3.2 SENSOR OBSERVATION SERVICE

There are three core requests that can be made to retrieve sensor (meta)data from a SOS: `GetCapabilities`, `DescribeSensor` and `GetObservation`. `GetCapabilities` returns a complete overview of what the SOS has to offer. The `DescribeSensor` request returns detailed information about individual sensors. These three core requests are mandatory in a SOS under the 2.0 specifications [Bröring et al., 2012]. There are also a number of optional extensions to a SOS. Requests can be made as a HyperText Transfer Protocol (HTTP) GET request or a HTTP POST request. There can be different response formats. Usually there is at least the option to retrieve the response as an XML document. Based on the specification by Bröring et al. [2012] this paragraph describes the core and optional requests of a SOS, as well as the structure of their responses.

3.2.1 Get capabilities

The `GetCapabilities` request is the first step in communicating with a SOS. The request is made by taking the HTTP address of the SOS and adding `service=SOS&request=GetCapabilities`. It returns a document including information on what the service has to offer. The document contains a number of sections: service identification, service provider, operations metadata, filter capabilities and contents.

In the service identification section there is general information about the service, such as the title and supported SOS versions, but also whether there are fees or access constraints. The service provider section contains details on which organisation provides the SOS and lists their contact information. The operations metadata section lists the supported request types. It also contains an overview of all features-of-interest, observed properties, procedures and offerings. Offerings are similar to layers in a WMS, grouping together observations collected by one procedure.

The contents section describes the data that can be retrieved, grouped in offerings. Each offering has an identifier, together with information on the procedure, observable properties and the feature-of-interest type. Which filters can be applied in a request is described in the filter capabilities section. The supported parameters for both spatial and temporal filters are listed here.

3.2.2 Describe sensor

The `DescribeSensor` request gives detailed information on a specific sensor. The request is built by taking the HTTP address of the SOS and adding `service=SOS&version=2.0.0&request=DescribeSensor&procedure=aprocedure&proceduredescriptionformat=aformat` where the procedure and procedure description format have to contain values defined in the capabilities document.

3.2.3 Get observation

Using `GetObservation` actual measurements can be retrieved. The request is made by taking the HTTP address of the SOS and adding `service=SOS&version=2.0.0&request=GetObservation`. This returns a response with the

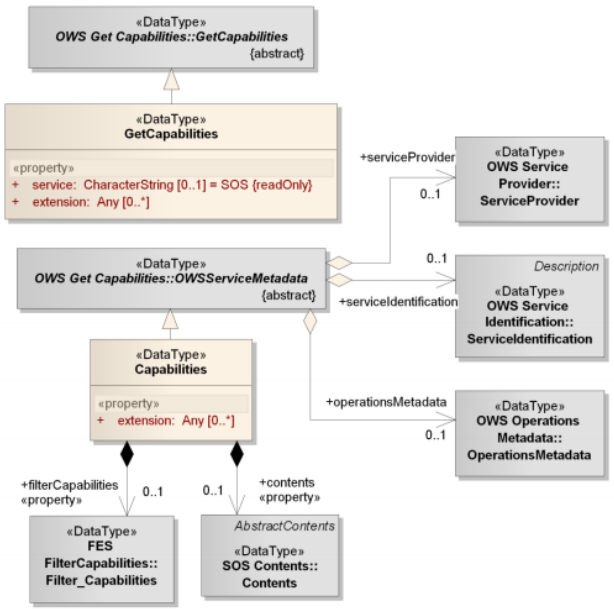

Figure 3.1: Data model behind the capabilities document of a SOS [Bröring et al., 2012]

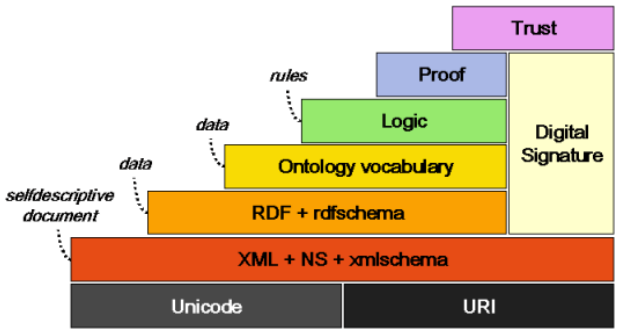

Figure 3.2: Hierarchy of the semantic web [Koivunen and Miller, 2002]

default parameters, which can differ from one SOS to another. To further specify the request, optional paramaters can be added such as: observed property, procedure, feature-of-interest, offering and outputformat. Spatial and temporal filters can be added if these are supported by the service.

3.3 SEMANTIC WEB

3.3.1 Resource Description Framework

For publishing geographic data on the semantic web a conversion of Shapefiles to RDF is required. For this the method by Missier [2015] will be used. 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 notation. 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].

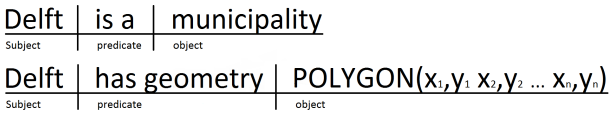

Figure 3.3: Triples of object, predicate and subject define Delft as a municipality with a geometry

In RDF data is stored as 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. For example, to define a geographic feature such as the municipality of Delft on the semantic web a number of triples can be made. Figure 3.3 shows how Delft can be defined as a municipality with a certain geometry using triples of subject, predicate and object.

Three types of data can make up these triples [Manola et al., 2014]. 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 a location or how it can be accessed. An IRI is a generalisation of an URI, and also allows non-ASCII characters. In the example of the municipality of Delft, IRIs can be used to define 'Delft' and 'Municipality', but also for the predicates 'is a' and 'has geometry'. 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. In the example of Delft, a literal could be used to store the actual geometry of the boundary: POLYGON((x_1, y_1 x_2, y_2 ... x_n, y_n, x_1, y_1)). A literal value can have a datatype specification [Cyganiak et al., 2014]. This is added to the literal with the ^^ symbols, followed by the IRI of the datatype specification. In Figure 3.4 the datatype is 'geo:wktLiteral'.

Sometimes it is 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].

3.3.2 Notation

There are a number of different notations 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]. Turtle will be used in this thesis, because it is commonly used notation which is also relatively easy to read for humans. The DBPedia IRI is used for the object 'Municipality'. The 'is a' predicate is represented by a built-in RDF predicate which can be written simple as 'a'. The second predicate is 'hasGeometry' for which the GeoSPARQL IRI is used. The geometry is a literal in the Well-Known Text (WKT) format. Note that the subject is only written once when there are multiple triples with the same subject. Triples that shares the same subject are divided by semicolons. A point marks the end of the last triple with a specific subject.

```
PREFIX geo: <http://www.opengis.net/ont/geosparql#>

<http://example.com/Delft> a <http://dbpedia/resource/Municipality> ;
    geo:hasGeometry "<http://www.opengis.net/def/crs/EPSG/0/4258> POLYGON(( x1 y1, x2 y2, ... xn yn, x1 y1 ))"^^geo:wktLiteral
```

Figure 3.4: Triples of Figure 3.3 in the Turtle notation

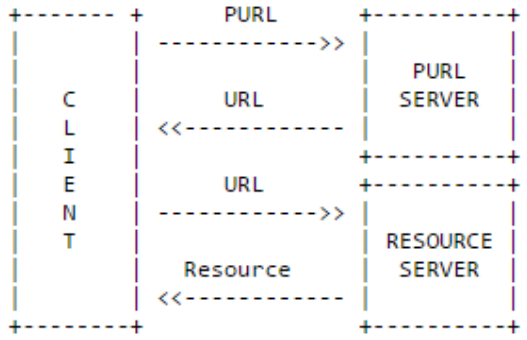

Figure 3.5: Persistent Uniform Resource Locator (PURL) resolves to the current resource location [Shafer et al., 2016]

3.3.3 Persistent Uniform Resource Locators

URLs are an essential part of the web. However, if an URL changes the existing links towards this URL are broken. To prevent this Persistent Uniform Resource Locators (PURLs) are being used. A Persistent Uniform Resource Locator is a "naming and resolution service for general Internet resources" [Shafer et al., 2016]. This allows organisations to change the location of their data without changing the URL to which can be linked. A PURL server receives the URL and redirects the client to the current location of the resource. If the location of the resource changes, the server can be informed. It will then redirect clients to the new location (Figure 3.5).

3.4 GEOSPARQL & STSPARQL

GeoSPARQL "defines a vocabulary for representing geospatial data in RDF, and it defines an extension to the SPARQL query language for processing geospatial data" [Perry and Herring, 2012, p. xvi]. It allows for defining geometric data in RDF and performing spatial queries. The Dimensionally Extended Nine-Intersection Model (DE-9IM) [Strobl, 2008] has been implemented to find topological relations between two geometries. GeoSPARQL has been implemented in the 'Parliament' SPARQL endpoint [Battle and Kolas, 2012]. The Strabon endpoint uses stRDF, which is "a constraint data model that extends RDF with the ability to represent spatial and temporal data" [Koubarakis and Kyzirakos, 2010, p. 425]. The stRDF model can be queried using stSPARQL, which syntax is similar to GeoSPARQL (listing 3.1 & 3.2). Both extensions of SPARQL use filter expressions to perform spatial operations. The definition of geometries and the syntax of the filter expression differ slightly.

Listing 3.1: A GeoSPARQL query to find the names of features that contain a point geometry

```
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
```

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT
?name
WHERE {
?feature geo:hasGeometry ?geom .
?feature foaf:name ?name.
FILTER
    (geof:sfContains(?geom,"<http://www.opengis.net/def/crs/EPSG/0/4258>
    POINT(4.289244 52.027337)"^^geo:wktLiteral))
}
```

Listing 3.2: A stSPARQL query to find the names of features that contain a point geometry

```
PREFIX strdf: <ttp://strdf.di.uoa.gr/ontology#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT
?name
WHERE {
?feature strdf:hasGeometry ?geom .
?feature foaf:name ?name.
FILTER (?geom contains "POINT(4.289244
    52.027337);<http://www.opengis.net/def/crs/EPSG/0/4258>"^^strdf:WKT)
}
```

3.5 ONTOLOGIES

When publishing data on the semantic web, ontologies are required to specify what things are and how they relate to other things. The evaluation of observation metadata ontologies by Hu et al. [2014] is interesting, since it exposes what the relevant aspects are in the process of observation discovery. However, their proposed model focusses mainly on including remote sensing and imagery data in metadata models that were not originally created for this kind of data. The SSNO is an ontology that clearly describes the process between sensor, stimulus and observation. However, Cox [2015b] points out that an important aspect of describing a sensor network is missing in this ontology: the sampling. Also, the om-lite and sam-lite ontologies by Cox [2015b] are lightweight ontologies that can be complemented by already existing linked data ontologies. They do not rely on the (heavy) ISO specifications that date from before the semantic web, unlike the SSNO. The om-lite and sam-lite ontologies will therefore be used in this thesis.

The SOS has a number of metadata attributes such as the service provider's details (including contact information), its spatial and temporal extent (spatialFiler & 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 feature which intents 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 ouput. 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 Unified Modeling Language (UML) diagram in Figure **??** the classes Observation, Process, ObservedProperty and FeatureOfInterest can be mapped to classes belonging to 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, 2012] and the PROV ontology for the sensor and sensor observation service classes [W3C Semantic Sensor Network Incubator Group, 2011].

3.6 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. Also, spatial aggregation techniques (based on neighbourhood analysis) can be considered to adjust for spatio-temporal irregularities as mentioned by Ganesan et al. [2004]. 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.

4 DATA

4.1 VECTOR DATA

4.1.1 Topography

The datasets of Dutch provinces (provincies, Figure 4.2) and municipalities (gemeenten, Figure 4.1) have been downloaded from https://www.pdok.nl/nl/producten/pdok-downloads/basis-registratie-kadaster/bestuurlijke-grenzen-actueel. For the Netherlands there are 12 features in the provinces and 393 in the municipalities dataset.

It has been challenging to obtain data of administrative boundaries of Belgium (even from the INSPIRE data portal). Therefore, all data for Belgium was retrieved from http://www.gadm.org/. There are also 12 features in the provinces (including the capitol region of Brussels) and there are 589 features in the municipalities dataset.

The country datasets have also been downloaded from http://www.gadm.org/ (Figure 4.3). The administrative unit data contains the names of the administrative units and their (polygon) geometry.

4.1.2 Land cover

Data on land cover will be used to complement the data of administrative units. A section of the 2012 dataset from the Coordination of Information on the Environment (CORINE) programme will has been selected for this (Figure 4.4). The entire CORINE dataset was retrieved from http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012. The features overlapping the Netherlands and Belgium have been retrieved from this dataset using the open source QGIS software and stored in a separate database in Postgres.

The database contains polygon geometries (Figure 4.5) with a unique identifier and a code that refers to the type of landcover. These codes can be looked up in the accompanied spreadsheet file containing the legend table of CORINE 2012.

4.2 RASTER DATA

Data is often used in a raster representation for computations in a Geographical Information System (GIS). For natural phenomenon a raster representation is especially well suited. The EEA reference grid is a standard grid which covers Europe. It is available with a resolution of 100km², 10km² and 1km². In this thesis the EEA grid cells with a resolution of 100km² (Figure 4.6) and 10km² (Figure 4.7) have been used that overlap the Netherlands and Belgium. 15 grid cells of 100km² and 843 grid cells of 10km² have been selected from the original dataset.

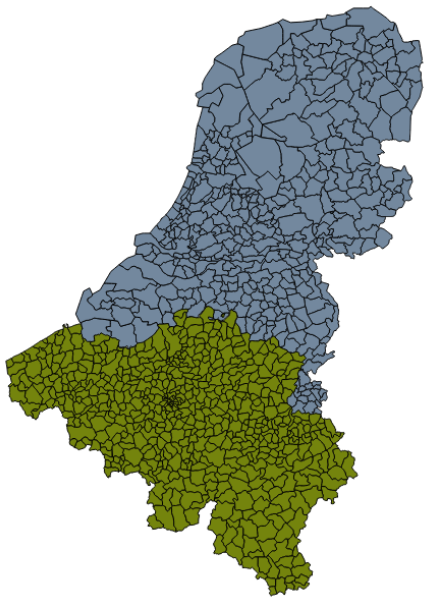

Figure 4.1: Dataset of municipalities in the Netherlands and Belgium in 2015 (from Dutch cadaster and GADM.org)

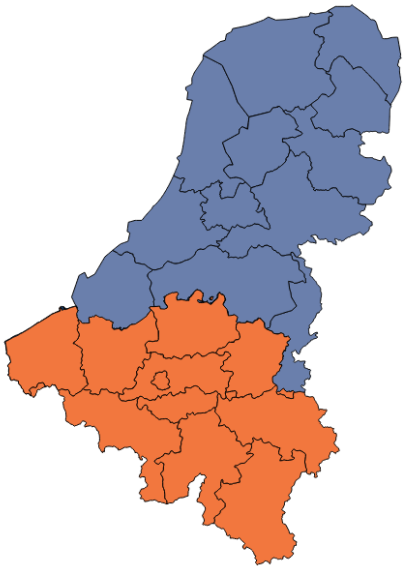

Figure 4.2: Dataset of provinces in the Netherlands and Belgium in 2015 (from Dutch cadaster and GADM.org)

Figure 4.3: Dataset of the Netherlands and Belgium in 2015 (from GADM.org)

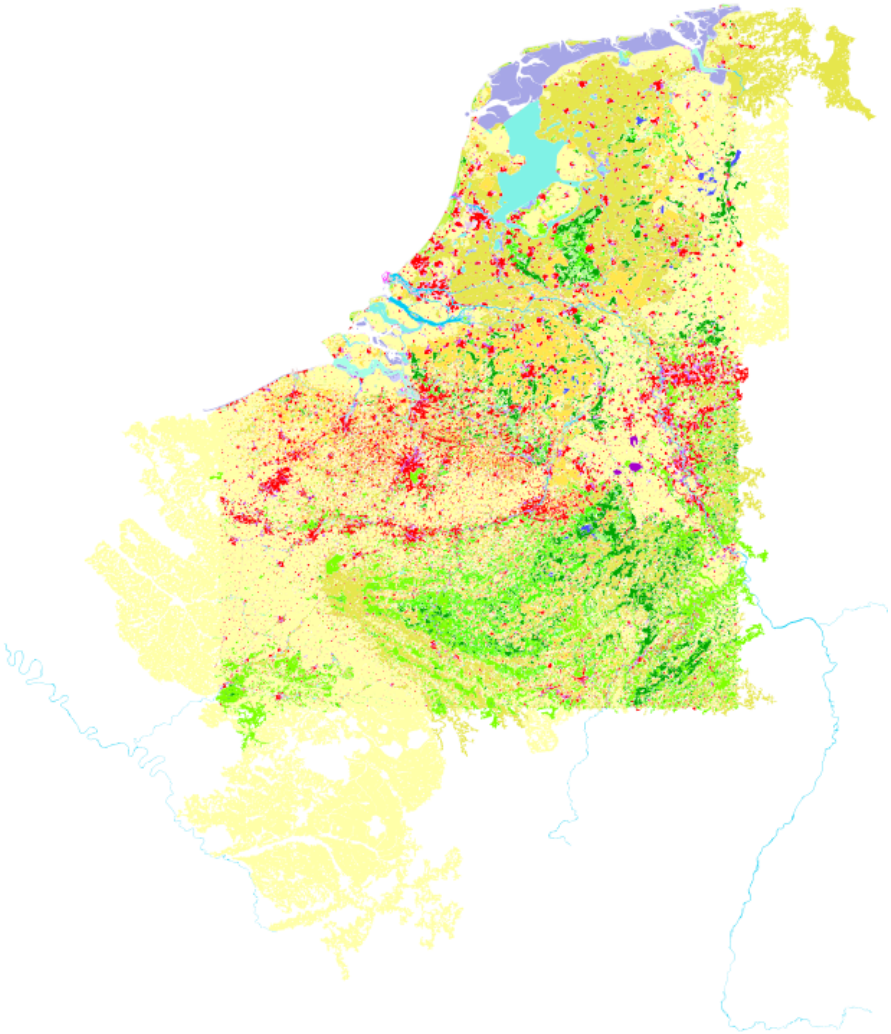

Figure 4.4: Dataset of landcover in the Netherlands and Belgium in 2012 (from Copernicus The European Earth Observation Programme)

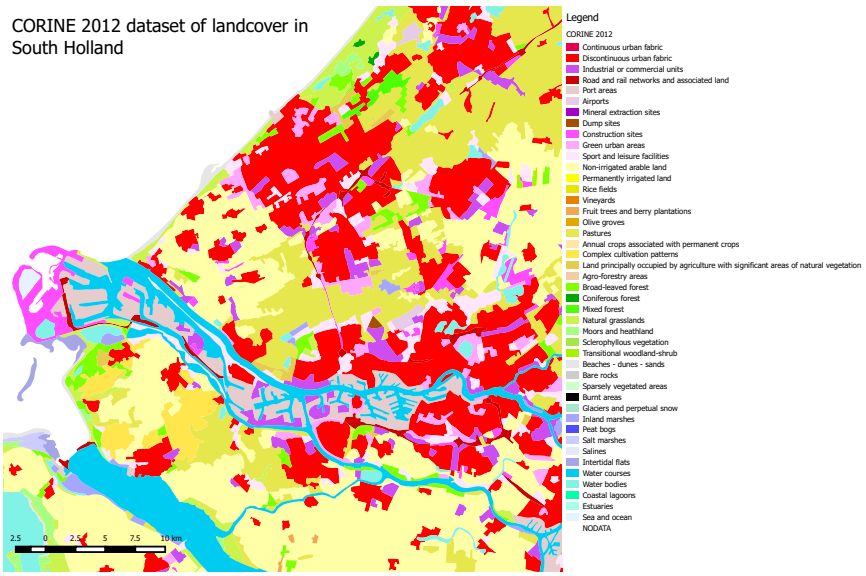

Figure 4.5: Landcover of the province of South Holland (subsection of the dataset from Figure 4.4)

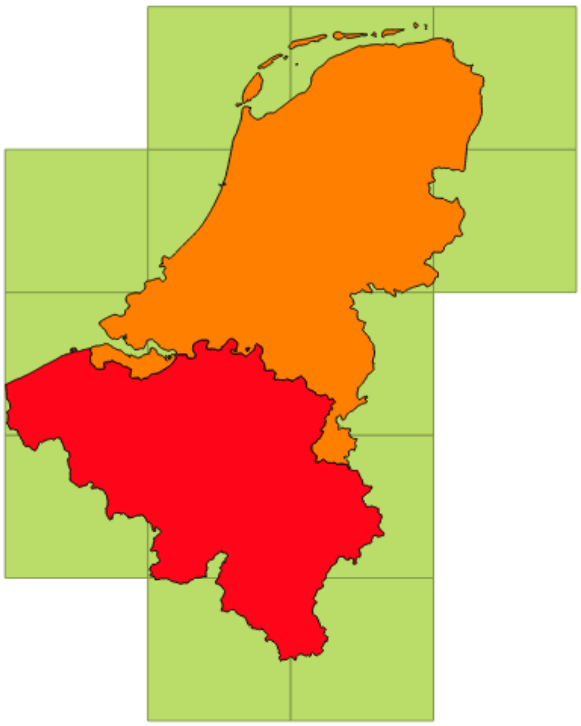

Figure 4.6: EEA reference grid cells with a resolution of 100km² overlapping the Netherlands and Belgium

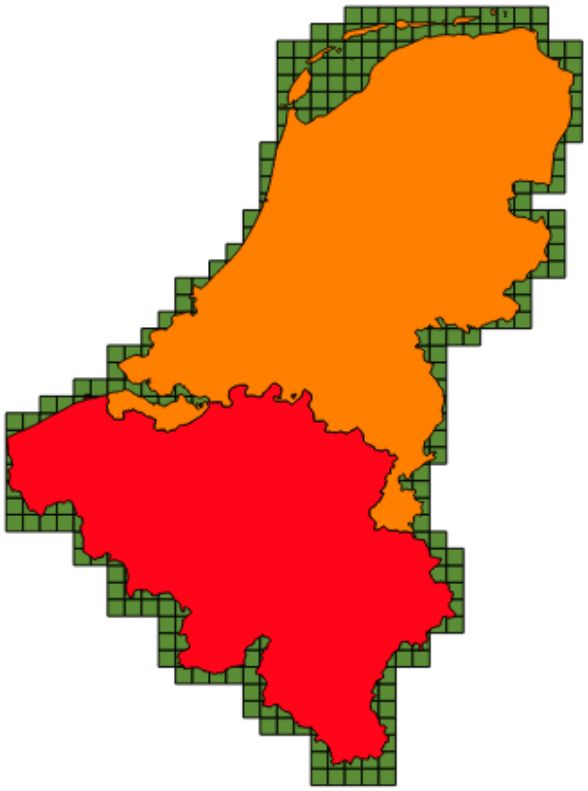

Figure 4.7: EEA reference grid cells with a resolution of 10km² overlapping the Netherlands and Belgium

Figure 4.8: Webmap by the RIVM showing their air quality sensor network (http://www.lml.rivm.nl/meetnet)

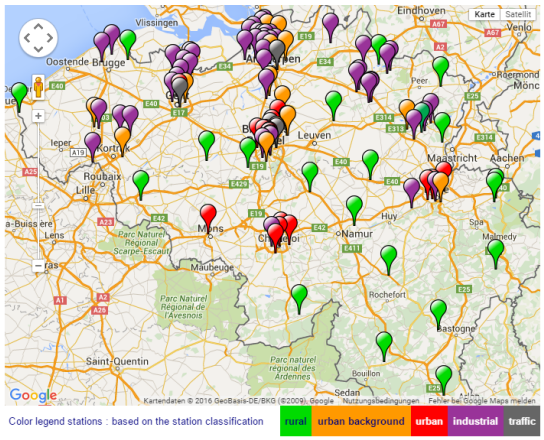

Figure 4.9: Webmap by IRCEL-CELINE showing their air quality sensor network (http://www.irceline.be/en/air-quality/measurements/monitoring-stations/)

4.3 SENSOR DATA

Air quality sensor data will be used from the RIVM (http://inspire.rivm.nl/sos/) and from the IRCEL-CELINE (http://sos.irceline.be/). Both of these organisations have a SOS where data can be retrieved according to the SWE standards. The one of the RIVM has been online since the 21st of August, 2015. IRCEL-CELINE already made the SOS available on the first of January, 2011. Figure 4.8 and Figure 4.9 show the sensor networks of both organisations. They provide different kinds of sensor data, such as particulate matter (PM_{10}), nitrogen dioxide (NO^2) and ozone (O^3). Figure 4.10 shows one of the sensor locations in the city center of Amsterdam.

Figure 4.10: Google Streetview image of RIVM sensor location in Amsterdam in 2015

5 IMPLEMENTATION

Implementation of the methods are described in this chapter.

5.1 PREPARING LINKED DATA

Linked data has been prepared that is used to retrieve and process sensor data on the semantic web (Figure 5.1). This is done for vector data sets of administrative units and land cover features, and for raster data sets of EEA grids with a resolution of 10km² and 100km².

Three types of administrative units have been converted to linked data: countries, provinces and municipalities. Every administrative unit has a name, 'type' and (multi)polygon geometry assigned to it (Figure 5.1). The administrative unit type is defined by DBPedia URIs of country, province and municipality.

The CORINE 2012 land cover dataset contains features with an identifier, a land cover type and a (multi)polygon geometry (Figure 5.1). The identifier has the form of: 'EU-' plus a unique seven digit number. The land cover type is defined by a three digit number, which can looked up in the provided spreadsheet containing the legend.

The EEA reference grid with resolutions of 10km² and 100km². Every feature is defined by an identifier, a resolution and a point geometry of the origin (Figure 5.1). The identifier is a code given to a feature by the EEA and has the form of: resolution + 'E' + x coordinate + 'N' + y coordinate.

5.2 PUBLISHING LINKED DATA

Setting up the Strabon (Figure 5.2), Apache Tomcat and Pubby software.

The Strabon and Parliament endpoint have been tested since they both handle GeoSPARQL queries. Strabon has been used in the final implementation, because the Parliament endpoint rejected certain longer queries (see 6.3).

Pubby software in combination with Apache Tomcat allows for a user interface that is easier to navigate through for humans. The links stored in RDF triples are represented as hyperlinks which can be used to navigate between pages about different concepts.

For creating Persistent Uniform Resource Locators the Purlz software (http://www.purlz.org/) has been used. All URIs that are created get a PURL assigned to it. The PURL resolves the URI to a DESCRIBE query at the endpoint. This query is structured as a get request: `http://localhost/strabon-endpoint-3.3.2-SNAPSHOT/Describe?submit=describe&view=HTML&handle=download&format=turtle&query=DESCRIBE<an_URI>`. The re-

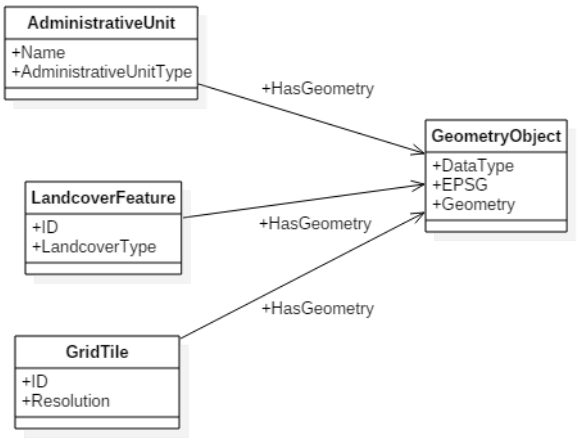

Figure 5.1: Model of vector and raster features

Table 5.1: Types of PURLs [Shafer et al., 2016]

PURL Type	Meaning	HTTP Shorthand
301	Moved permanently to a target URL	Moved Permanently
302	Simple redirection to a target URL	Found
303	See other URLs (use for Semantic Web resources)	See Other
307	Temporary redirect to a target URL	Temporary Redirect
404	Temporarily gone	Not Found
410	Permanently gone	Gone

quest has '`/Describe?submit=describe`' to call the script that deals with describe queries and to tell it that the request is also submitting this kind of query. The parameters '`view=HTML&handle=download`' indicate that the endpoint's website is requested, but the returned data should be a download file instead of an HTML page. The parameter '`&format=turtle`' sets the RDF notation of the download file to Turtle and '`&query=DESCRIBE <an_URI>`' is the SPARQL query that contains the URI between brackets.

Every URI is written to an XML file with the parameters: ID, PURL type, and target address. Optionally, information about the person or organisation maintaining the PURL can be added. The ID is the original URI that is being resolved to the target address. The PURL type is set to 303, which means that it refers the client to the target address. Alternative types can be found in Table 5.1. After all URIs have been added to the so called XML 'batch' file [PURL, 2016], the file can be posted to the Purlz server.

Setting it up on the university server

5.3 RETRIEVING METADATA FROM THE SENSOR OBSERVATION SERVICE

The metadata is automatically retrieved from the SOS according to Figure 5.3. A SOS is maintained by an organisation, of which the name is retrieved, as well as whether they charge fees or have implemented access constraints for using the SOS. In most cases the use is free of charge and without access

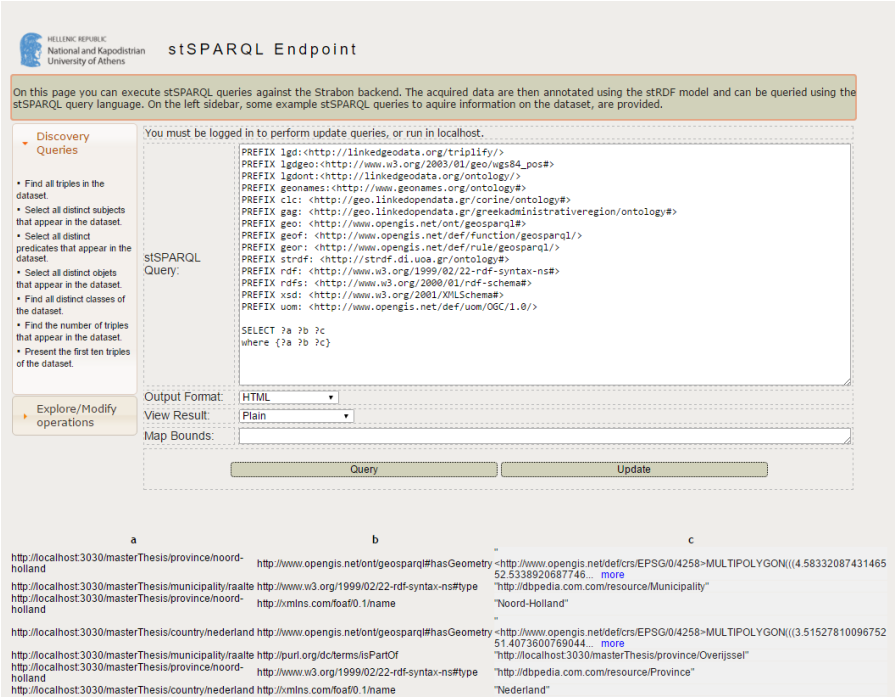

Figure 5.2: Strabon endpoint

constraints. However, it is possible for an organisation to restrict the use of the SOS in these ways.

In the SWE standards a sensor is modelled using two entities: a procedure and a feature of interest. The procedure is the method of sensing and the feature of interest is the feature of which the sensor is sensing a certain property. Therefore, the observable property ties together the procedure and feature of interest. It should be noted that the geometry of a feature of interest is not necessarily always a point geometry. It can also be generalized into larger features (e.g. multiple sensors observing different parts of one lake).

An offering is a grouping of features of interest, which have a common procedure. The purpose of offerings is to allow users to query the observation data more efficiently. Features of interest that are often queried together are grouped into the same offering.

A Python class object is created for the SOS based on Figure 5.3. This class contains the different variables and has built in functions to automatically retrieve the metadata. To collect all the required metadata a number of request have to be made, which is done using the `SOSclass.request()` method which requires only the HTTP address of the SOS as input. First, the capabilities document is retrieved to collect information about the organisation, supported SOS versions and response formats. It also contains lists with identifiers for all features of interest, offerings, observable properties and procedures. In this document the offerings are linked to a certain procedure and to a certain observable property.

Unfortunately, the capabilities document is not able to provide information about which procedure is being applied for which feature of interest. Also, the features' geometries cannot be retrieved from it. Therefore, a `GetFeatureOfInterest` request is made to retrieve the location of each features of interest. However, the `GetFeatureOfInterest` document does

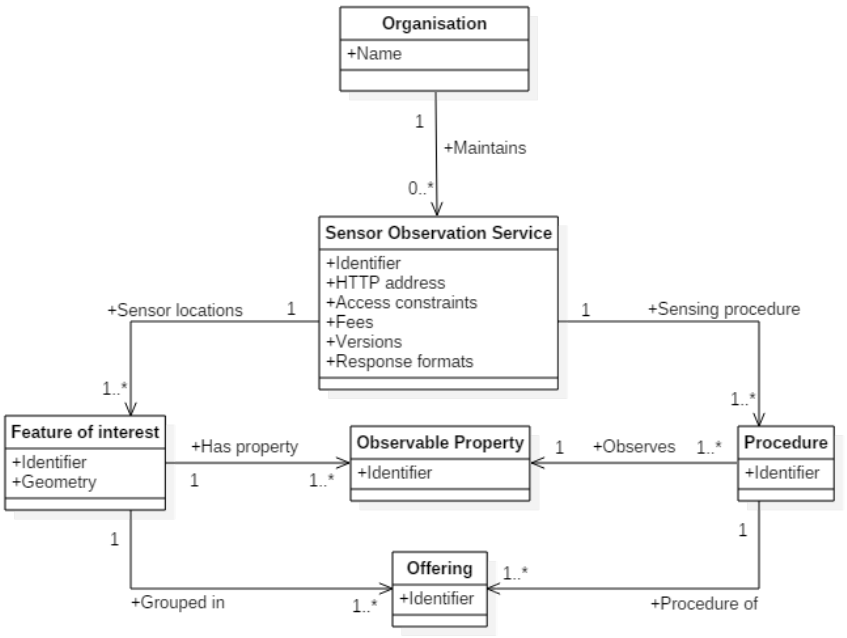

Figure 5.3: Metadata automatically retrieved from a SOS

not necessarily provide information about the procedures that are related to a certain feature of interest. When a `GetObservation` request is made for an offering, the returned observation data is grouped per feature of interest. Therefore, small amounts of data are retrieved from each offering using `GetObservation` requests, when possible with a temporal filter to limit the data traffic. Every procedure and offering can now be related to a set of features of interest with point geometries.

5.4 MODELLING WITH THE OM-LITE AND SAM-LITE ONTOLOGIES

After the metadata has been retrieved from the SOS (Figure 5.3) it has to be converted to linked data. For this the om-lite and sam-lite ontologies are being used in combination with the PROV and GeoSPARQL ontologies. Figure 5.4 shows the semantic relations between the metadata that are stored in RDF triples. Every (instance of a) class in this figure is represented by a URI.

A SOS is modelled as an agent with a specific name, that acts on behalf of a certain organisation. The organisation, access constraints, fees, versions and response formats are properties of the SOS. Every sensor is described by a procedure and a certain feature of interest. The sensor class was not present in the model of Figure 5.3, because the SOS does not define sensors. However, the sensor class has been added to the semantic model to make the relation between procedure and sampling point explicit.

In Figure 5.4 the collection of sampling features only contains sampling points. This is because the feature of interest of an air quality sensor is equal to the bubble of air directly around the sampling point. Other sampling features can be added when the application requires this. Sampling features

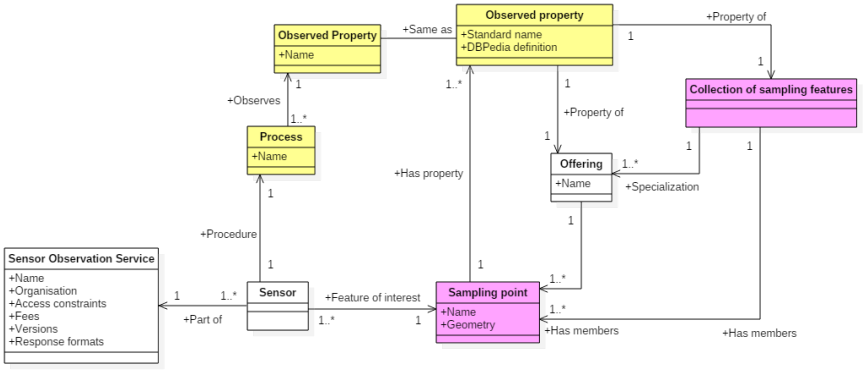

Figure 5.4: Sensor metadata as modelled in RDF (om-lite classes in yellow and sam-lite classes in purple)

are grouped into collections of features of which the same observed property is measured. These collections can contain sampling points from multiple Sensor Observation Services. The offering class is a specialization of the collection of sampling features. It contains a subset of the sampling points that are all part of the same offering at a particular SOS.

Every observed property that is defined in a SOS relates to a certain observed property as defined by DBPedia. Since SOS requests require their own identifiers as input the observed property class exists twice in the model: one as defined by the SOS and one as defined by DBPedia. For the same reason all sampling points, processes and offerings have a 'name' attribute in addition to their URI. These store the original identifier that they were given by the SOS.

5.5 ESTABLISHING INWARD LINKS FROM DBPEDIA

Sending triples to DBPedia the project to be uploaded.

Contains links from DBPedia page of 'Sensor Observation Service' to the semantic definition of the input SOS, from the DBPedia page of 'ozone', 'particulate matter' and 'nitrogen dioxide' to the corresponding collections of sampling features.

Also, the data of administrative units, land cover features and EEA reference grid cells will be linked to from DBPedia.

5.6 PROTOTYPE IMPLEMENTATION

The prototype implementation serves as a proof of concept. It looks on the semantic web for sensors that observe a certain property in a specific area. It collects the data for these sensors at their corresponding Sensor Observation Service. When multiple data sources are found the data is integrated. The sensor data is aggregated before it is returned to the user.

5.6.1 Input parameters

The prototype takes a number of input parameters. First of all, a list with observed properties which will be the 'layers' that the process returns. The second parameter is the category of input features. This can be set to administrative units (country, province or municipality), land cover or raster. The third parameter is a list of input feature. This is a list of names or identifiers that correspond to the category. The next parameter is the temporal range. This has to be a list of two ISO datetime strings representing the start and end time. The fifth parameter is the temporal granularity, represented by an ISO delta datetime string. The sixth input parameter is the method of spatial aggregation. This method will be applied to aggregate the data based on the input features. The last input parameter is the temporal aggregation method to aggregate data between start and end time to the required granularity.

5.6.2 Retrieving geometries

The input category is a starting point for the process to find the geometries of the input features. It creates a SPARQL query to retrieve the geometries of these features.

5.6.3 Spatial queries

With the found geometries a SPARQL query is made to find a sensor collection that has a certain observed property. From this collection sensors can be selected that overlap the previously found geometries. Unfortunately SPARQL queries are not allowed to exceed a certain number of lines. This creates problems when querying larger vector geometries (provinces and countries). For these queries two alternatives have been implemented: using the EEA reference grid as a spatial index for vector geometries and using the bounding box vector geometry.

5.6.4 Retrieve sensor data

From these sensors data is collected.

5.6.5 Integrate data sources

Data is integrated from multiple sources.

5.6.6 Data aggregation

Data is aggregated.

5.7 SETTING UP THE WEB PROCESSING SERVICES

Creating two Web Processing Services using PyWPS.

5.8 CREATING AN ONLINE DASHBOARD

6 RESULTS

6.1 IMPLEMENTATION DIFFERENCES BETWEEN SENSOR OBSERVATION SERVICES

Even though both Sensor Observation Services use the SWE standards they are not exactly the same. They have different approaches for making identifiers and have slightly different content in their capabilities and describe sensor documents. This has caused the implemented to be more complex than expected. Especially providing a describe sensor document with as much information as possible (containing related features of interest, observed property and offerings) is important for making sense of the metadata. This should also be done using the appropriate SensorML tags, instead of the more general 'swe:keywords' for example.

When retrieving sensor data from the SOS using `GetObservation` requests one implementation provides this data as an array of comma separated values, using the SWE Array Observation class. The other implementation provides the same kind of data embedded in XML tags using the O&M Measurement class.

One of the SOS implementations allowed multiple observed properties to be related to an observation procedure. This was an unexpected finding, because a procedure is defined as a "method, algorithm or instrument, or system of these which may be used in making an observation" [ISO, 2011, p. 4]. When a procedure has different observed properties it is rather a combination of procedures. It is unlikely that the same procedure can observe one property at a certain time and a completely other property at another time. The original implementation of the WPS had to be adjusted to deal with this.

6.2 SEMANTICS IN SENSOR OBSERVATION SERVICES

6.2.1 mapping of observable properties

The metadata in Sensor Observation Services does not necessarily contain semantics. This is an issue when automatically retrieving metadata from different services. In the implementation this has caused a problem with identifying which observable properties in a SOS are the same as observable properties found in another SOS.

6.2.2 Automatically creating an URI scheme

If an URI is automatically created for data from an unknown source there is an amount of uncertainty to what the URI will look like. Either a random identifier can be assigned to it, or the identifier that is already provided by

the data source. This identifier most likely contains some reference to the nature of the real world thing the URI represents and is therefore preferable to a a completely random identifier. Adding the given identifier to the URI can create very long and strange URIs because different data sources have a different way of creating URIs since the O&M schema allows simply 'any URI'.

6.3 SPATIAL QUERIES WITH STSPARQL

For retrieving data about a vector feature three methods for spatial querying have been implemented. First of all, spatial queries in which the server side receives the complete geometry of a feature and checks which point geometries it contains. Second of all, spatial queries in which the server side receives the bounding box of the complete geometry and check which points are within it. Third of all, spatial queries in which the server side receives the EEA reference grid cell that overlap the geometry and checks which points each cell contains. With the first method the client does not have to perform any spatial queries anymore. The second and third method will also return points that are not inside the geometry of the vector feature. These points could be filtered out by the client.

6.3.1 Vector queries

GeoSPARQL allows spatial queries in which a geometry can be inserted in the query. The implementation uses this functionality to test for spatial relations between geometries. For example between the point geometry of a sensor and the polygon geometry of an administrative unit. However, when geometries become more complicated their WKT definition becomes more verbose. This leads to the query being rejected based on its number of characters by the endpoint. This indicates that vector queries with complex geometries are not very efficient to include in a SPARQL query.

6.3.2 Raster queries

All raster cells are retrieved that overlap with the vector geometry. For these raster cell sensors are retrieved and later the excess ones are filtered out.

6.3.3 Bounding box queries

Creating a bounding box around a vector feature and make `GetObservation` to the SOS.

6.3.4 Latitude and longitude order

During the implementation a problem has come up regarding the order in which latitude and longitude are being presented in the SOS. The SOS of the RIVM provides point geometries in WGS84 as longitude, latitude and height. However, the Strabon endpoint expects the order to be latitude, longitude and height. This results false outcomes of spatial queries.

6.4 OUTPUT DATA

Data is now outputted in XML according to the O&M schema. However, can this schema really handle aggregated data?

7 DISCUSSION

7.1 METADATA DUPLICATION

The method presented in this thesis takes metadata from a SOS, converts it to linked data and publishes it on the semantic web. Although the metadata has taken another form, it is now stored twice in two different locations. This may not be desirable, for instance when data is updated in the original source and its linked data equivalent is not. Also more storage space is required for the same amount of data. However, extra functionality is achieved in return.

7.2 METADATA QUALITY

The quality of the metadata in the SOS influences the quality of the metadata SPARQL endpoint.

7.3 AUTOMATED PROCESS

If there is no meaning added to definitions like observed property, the metadata is not machine understandable. In this case, manual work has to be done to make it machine understandable. Only after this manual process it can be published on the semantic web.

7.4 EXPLICIT TOPOLOGICAL RELATIONS

Spatial features have topological relations with other spatial features. These relations can be made explicit on the semantic web. However, in this thesis they have not been made explicit and are calculated on-the-fly with spatial queries using GeoSPARQL. Making topological relations explicit in a subject-predicate-object structure could improve query speed, as they are likely less expensive than spatial queries. However, this is a trade-off with the required storage space. Furthermore, the chances of incorrect or broken links increase as both features and topological relations can change over time.

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.

Battle, R. and Kolas, D. (2012). Enabling the geospatial semantic web with parliament and geosparql. *Semantic Web*, 3(4):355–370.

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 9th, 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 9th, 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.

Bröring, A., Stasch, C., and Echterhoff, J. (2012). OGC Sensor observation service interface standard.

Cambridge Semantics (2015). Introduction to the Semantic Web. [online] https://www.cambridgesemantics.com/semantic-university/introduction-semantic-web [accessed on December 8th, 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 1st, 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 24th, 2015].

Cox, S. J. D. (2015d). OWL for Sampling Features. [online] http://def.seegrid.csiro.au/ontology/om/sam-lite [accessed on November 24th, 2015].

Cyganiak, R., Wood, D., and Lanthaler, M. (2014). RDF 1.1 Concepts and Abstract Syntax. [online] https://www.w3.org/TR/rdf11-concepts [accessed on February 2nd, 2016].

Gandon, F. and Schreiber, G. (2014). W3C RDF 1.1 XML Syntax. [online] http://www.w3.org/TR/rdf-syntax-grammar/ [accessed on December 9th, 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 2nd, 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 [accessed on December 2 nd, 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.

Koivunen, M.-R. and Miller, E. (2002). W3C Semantic Web Activity. In Hyvönen, E., editor, *Semantic Web Kick-off Seminar in Finland*.

Korteweg, P., Marchetti-Spaccamela, A., Stougie, L., and Vitaletti, A. (2007). *Data aggregation in sensor networks: Balancing communication and delay costs*. Springer.

Koubarakis, M. and Kyzirakos, K. (2010). Modeling and querying metadata in the semantic sensor web: The model stRDF and the query language stSPARQL. In *The semantic web: research and applications*, pages 425–439. 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 8th, 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 11th, 2015].

Manola, F., Miller, E., and McBride, B. (2014). W3C RDF Primer. [online] http://www.w3.org/TR/rdf11-primer/ [accessed on December 9th, 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.

OWL working group (2012). Web Ontology Language (OWL). [online] http://www.w3.org/2001/sw/wiki/OWL [accessed on December 18th, 2015].

Percivall, G. (2015). OGC Smart Cities Spatial Information Framework. OGC Internal reference number: 14-115.

Perry, M. and Herring, J. (2012). GeoSPARQL - A Geographic Query Language for RDF Data.

Price Waterhouse Coopers (2014). Sensing the future of the Internet of Things. [online] https://www.pwc.com/us/en/increasing-it-effectiveness/assets/future-of-the-internet-of-things.pdf [accessed on December 18th, 2015].

Pschorr, J., Henson, C. A., Patni, H. K., and Sheth, A. P. (2010). Sensor discovery on linked data.

Pschorr, J. K. (2013). SemSOS: an Architecture for Query, Insertion, and Discovery for Semantic Sensor Networks. Master's thesis, Wright State University.

PURL (2016). Batch Uploading to a PURL Server v1.0-1.6.x. [online] https://code.google.com/archive/p/persistenturls/wikis/PURLBatchUploadingVersionOne.wiki [accessed on February 19th, 2016].

Shafer, K., Weibel, S., Jul, E., and Fausey, J. (2016). Introduction to Persistent Uniform Resource Locators. [online] https://purl.oclc.org/docs/long_intro.html [accessed on February 18th, 2016].

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.

Strobl, C. (2008). *Dimensionally Extended Nine-Intersection Model (DE-9IM)*. Springer.

Theunisse, I. A. H. (2015). The Visualization of Urban Heat Island Indoor Temperatures. Master's thesis, TU Delft, Delft University of Technology.

van der Hoeven, F., Wandl, A., Demir, B., Dikmans, S., Hagoort, J., Moretto, M., Sefkatli, P., Snijder, F., Songsri, S., Stijger, P., et al. (2014). Sensing Hotterdam: Crowd sensing the Rotterdam urban heat island. *SPOOL*, 1(2):43–58.

Van der Hoeven, F. D. and Wandl, A. (2015). Hotterdam: How space is making Rotterdam warmer, how this affects the health of its inhabitants, and what can be done about it. Technical report, TU Delft, Faculty of Architecture and the Built Environment.

W3C Semantic Sensor Network Incubator Group (2011). Semantic Sensor Network Ontology. [online] http://www.w3.org/2005/Incubator/ssn/ssnx/ssn [accessed on December 9th, 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.

Xiang, L., Luo, J., and Rosenberg, C. (2013). Compressed data aggregation: Energy-efficient and high-fidelity data collection. *Networking, IEEE/ACM Transactions on*, 21(6):1722–1735.

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.