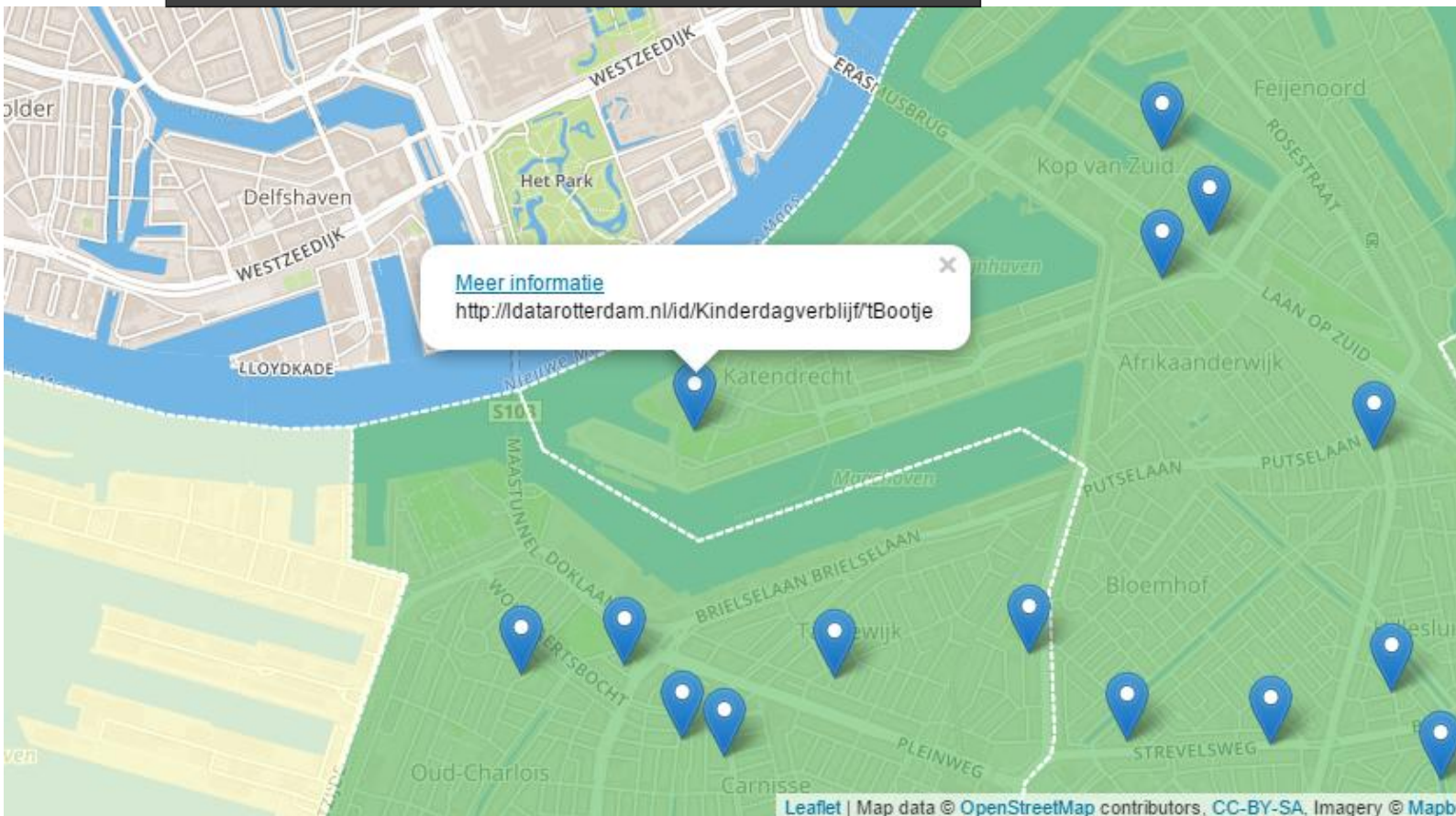


Towards a Web application for viewing Spatial Linked Open Data of Rotterdam

Godelief Abhilakh Missier



On the cover: The Web map showing the locations of the kindergartens from a dataset from the Rotterdam Open Data Store overlaid on OpenStreetMap. The colors indicate the percentage of households with children per area, in where darker green means more households with children in that area).

Towards a Web application for viewing Spatial Linked Open Data of Rotterdam

By

Godelief Abhilakh Missier

in partial fulfilment of the requirements for the degree of

Master of Science

in Geomatics for the Built Environment
Faculty of Architecture and the Built Environment



Graduation Professor	:	Prof. dr. ir. P.J.M. van Oosterom
Supervisor	:	drs. M.E. de Vries
Co-reader	:	dr.ir. B. van Loenen
Delegate of Board of Examiners	:	Not yet defined



Supervisor	:	ir. J. Hermans- van Ree
Supervisor	:	ing. R. Poll-van Dasselaar

ACKNOWLEDGEMENTS

This Master Thesis forms the research performed during the past 8 months regarding the topic spatial linked open data. This thesis is performed at the municipality of Rotterdam. In this section I would like to thank everybody who has helped and supported me during this time.

First my supervisors, Marian de Vries for her support, guidance and always making time for meetings. Peter van Oosterom for his time and valuable comments to keep improve my work. My co-reader Bastiaan van Loenen for his time to comment on my thesis.

From the municipality I want to thank my supervisors Jane Hermans- van Ree and Rob Poll- van Dasselaar for their guidance, interest, valuable input and providing me with the opportunity to do an internship.

From the municipality, I would also like to thank Michiel Boelhouwer, Rianne Miedema, Louis Smit, Joris Goos, Christian Veldhuis, Ron Keijser, Rob Weber, Mark Verschuur and Arjen Hazelebach and everybody else who commented on my thesis, inspired me and made me feel welcome at the municipality.

Also I would like to thank Frans Knibbe from Geodan for taking the time to introduce me to the various topics of Linked Data when I just started this thesis, back in November and his help. I would also like to thank Linda van den Brink from Geonovum for her time and valuable comments.

And also Marcel van Mackelenbergh for introducing me to linked data during his guest lecture in 2014 and his help with my thesis.

And last but not least of course my family for their love, support, understanding and for always believing in me!

TABLE OF CONTENT

Acknowledgements	5
Table of Content	7
List of Figures	10
List of Tables	12
Acronyms	13
Abstract.....	15
1. Introduction	1
1.1. Motivation	3
1.2. Research objective.....	4
1.3. Thesis outline	5
2. Case study	7
2.1. Focus of the case study	7
2.2. Research area.....	8
2.3. Choice of datasets	8
2.4. Functionality Web Application	13
2.5. End Users.....	13
3. Background.....	17
3.1. The 5 star scheme.....	17
3.2. Linked Data Methodology	18
3.3. Spatial Linked Data projects abroad.....	19
3.4. Linked Data in the Netherlands	21
3.5. BAG as Linked Data.....	22
3.6. Summary.....	25
4. Methodology.....	27
4.1. Outline of the methodology	27

4.2. Phase 1. Preparation	28
4.3. Phase 2. Modeling.....	29
4.4. Phase 3. Conversion.....	30
4.5. Output: Web application.....	32
4.6. Linked Data Management	33
5. Linked Data Technology	35
5.1. RDF	35
5.2. RDF Syntax	36
5.3. Ontologies and vocabularies	38
5.4. SPARQL.....	40
5.5. Modeling spatial data in RDF	42
5.6. Triple store functionality	46
5.7. On-the-fly conversion.....	49
5.8. Selected software and tools	49
6. Conversion to Spatial Linked Open data	53
6.1. Conversion of the municipality Open Data	53
6.2. Conversion of the statistical neighborhood and area data.....	61
6.3. Conversion of the Base Register for Addresses and Buildings (BAG)	63
6.4. Conclusion & discussion.....	73
7. Possibilities of using Spatial Linked Open Data	75
7.1. Spatial analysis of the created Spatial Linked Data	75
7.2. Linking to other datasets	79
7.3. Link the data with the metadata	82
7.4. Conclusion & discussion.....	84

8. Visualizing Spatial Linked Open Data.....	87
8.1. Current situation of publishing open data in Rotterdam	87
8.2. RDF on the Web	88
8.3. RDF on a web map	89
8.4. Integration into the existing system.....	92
8.5. Conclusion & discussion	93
9. Conclusions and recommendations	95
9.1. Conclusion.....	95
9.2. Reflection case study	97
9.3. Future work.....	99
9.4. Policy	100
9.5. Open research questions.....	102
10. Reflection.....	103
References	107
Appendix 1: Used prefixes.....	113

LIST OF FIGURES

Fig. 1. Example of a triple	2
Fig. 4. Municipality of Rotterdam.....	7
Fig. 5. Rotterdam Zuid in detail.....	8
Fig.6. Example of the BAG e.....	10
Fig.7.GisWeb.....	11
Fig.8. CBS area and neighbourhood map of 2013	12
Fig. 9. Layers in the Web Application.....	13
Fig. 10. Screenshot of the Web viewer of Ordnance Survey viewing London.....	20
Fig. 11. URI-pattern from the URI-strategy.....	22
Fig. 12. Flowchart of creating a Linked Data Web application	28
Fig. 13. Relation between the real world and data.....	29
Fig.14. RDF storage approach	30
Fig.15. On- the- fly to RDF approach	30
Fig. 16. Schematic way of linking datasets.....	32
Fig. 17. System Architecture of a Linked Data Web Application.....	33
Fig.18. Lifecycle model of publishing linked open data.....	34
Fig.19. Example of a directed graph.....	36
Fig. 20.Vocabularies in the Linked Open Vocabulary cloud	40
Fig.21. GeoSPARQL ontology	44
Fig. 22. RCC8.....	45
Fig. 23. Egenhofer's 9-intersection model.....	45
Fig. 24. OGC Simple Features Topological relationships.....	45
Fig.25. System Architecture of Strabon.....	48
Fig.26. the Kindergarten dataset from the Rotterdam Open Data Store.....	54
Fig.27. Attribute specification of a kindergarten in QGIS.....	54
Fig.28. Example of a kindergarten.....	56
Fig.29. Printscreen of the resources	59
Fig. 30. Printscreen of one resource.....	59
Fig. 31. Graph view of a selection of the household data.....	61
Fig. 32. CBS data for Rotterdam Zuid	62
Fig.33. Area dataset.....	62
Fig.34. Neighbourhood dataset.....	62
Fig. 35. UML- model with the classes in the BAG.....	65
Fig. 36. Incorrect premises and residential units in red.....	66
Fig. 37. Ontology of the BAG	68
Fig. 38. Concept of a premise with its versions	69
Fig. 37. a premise with its different versions	70
Fig.40. a premise and residential unit	72

Fig.41. Printscreen of the result from a federated query in Strabon.....	77
Fig.42. Web page of the Waalhaven on DBpedia.....	78
Fig. 43. Printscreen of the results from a federated query in Strabon.....	78
Fig. 44. Printscreen of the metadata query in Strabon.....	83
Fig. 45. Example of a premise and its definitions	84
Fig.46. the result of a query asking for all the kindergartens.....	88
Fig.47. Concept of visualizing the geometry	89
Fig.48. Part of the Web map showing the locations of the kindergartens	90
Fig.49. Information about one kindergarten object	91
Fig.50. Printscreen of the Web page about kindergarten 't Bootje	91
Fig. 51. Implementation of linked open data option 1	92
Fig. 52. Implementation of linked open data option 2	93

LIST OF TABLES

Table 1. Chosen open datasets	9
Table 2. 5 star scheme and data format.....	17
Table 3. Versions of the BAG.....	23
Table 4. Topological relations in GeoSPARQL	45
Adaptation from (Battle & Kolas, 2011a).....	45
Table 5. Non-Topological relations in GeoSPARQL	46
Table 6. Comparison of a selection of triple stores.....	47
Table 7. Tools per component used for the research	50
Table 8. Chosen vocabularies for the Kindergarten dataset	55
Table 9. Chosen vocabularies	61
Table 10. Chosen vocabularies for the CBS data	63
Table 11. Existing tables in the BAG- data	64
Table 12. Chosen vocabularies for the BAG.....	67
Table 13. Concept of the historical data in the BAG translated to URIs	69
Table 14. Comparison of the two methods.....	74
Table 15. Result spatial query.....	76
Table 16. Results of the performed query.....	79
Table 17. Result of the explicit linking query.....	80
Table 18. Result of the dynamic linking query	81
Table 19. Result of the metadata query	83
Table 20 . Possibility of spatial queries	85
Table 21. Used Prefixes and namespaces of existing vocabularies	113
Table 22. Used Prefixes and defined namespaces for converted data.....	113

ACRONYMS

<i>BAG</i>	Basisregistratie Adressen en Gebouwen, base registration of addresses and buildings
<i>CBS</i>	Centraal Bureau voor de Statistiek, Statistics Netherlands
<i>CRS</i>	Coordinate Reference System
<i>CSV</i>	Comma separated value file, a non-proprietary file format
<i>GBK</i>	Grootschalige Basis Kaart – Large scale map of Rotterdam, property of the Municipality of Rotterdam
<i>GIS</i>	Geographical Information Systems
<i>Graph</i>	A collection of triples. The subject and object are represented as nodes. The predicate is represented as an edge.
<i>INPSIRE</i>	Infrastructure for Spatial Information in the European Community
<i>JSON</i>	JavaScript Object Notation
<i>KML</i>	Keyhole Markup Language
<i>Linked Data</i>	Interrelated data on the Web, written in RDF.
<i>OGC</i>	Open Geospatial Consortium
<i>OWL</i>	Web Ontology Language
<i>RDF</i>	Resource Descriptive Framework in where data is written in the syntax of a triple. Standard for writing Linked Data.
<i>RDFS</i>	Resource Descriptive Framework Schema
<i>SKOS</i>	Simple Knowledge Organization System
<i>SPARQL</i>	A Standard, the query language for triples.
<i>Triple store</i>	RDF based storage solution for triples.
<i>Triple</i>	(s,p,o), where s is subject, p is predicate and o is object. In a triple a relation between the subject and object is given by the predicate.
<i>URI</i>	Unique Resource Identifier which identifies a resource. Mostly written and preferred as a HTTP URL.
<i>VoID</i>	Vocabulary of Interlinked Datasets
<i>W3C</i>	World Wide Web Consortium
<i>WFS</i>	Web Feature Service, an OGC standard for publishing spatial data as vector data
<i>WMS</i>	Web Map Service, an OGC standard for publishing maps on the Web

ABSTRACT

Linked Data is a method to publish data on the Web and originates from the Informatics sector. The data model exists of subject and object which are connected through the predicate. These three elements can be depicted with web pages. Governments and municipalities open up their data and publish spatial data on the Web, to be transparent to their citizens and to stimulate innovation. This spatial open data can be confusing for users since data is published as either typical GIS-formats or metadata is missing or sparsely available. Linked Data could benefit to that regard. The 5 star scheme from Tim Berners-Lee defines a model to publish open data as linked data on the Web so that additional information can be looked up. Data as linked data is considered more user friendly, understandable, re-useable and accessible for all users and converting spatial data to spatial linked data aids in integrating spatial data on the Web. Ordnance Survey has already published spatial datasets as spatial linked data and lately in the Netherlands organizations such as Geonovum and the Kadaster and companies are exploring this subject.

The research done for this master thesis focuses on how to apply the linked data method for spatial open data in order to profit from the benefits of linked data. This research provides a methodology to create spatial linked open data. In the case study for the Municipality of Rotterdam, four spatial open datasets concerning the composition of areas and neighborhoods in Rotterdam Zuid are converted to spatial linked open data to exemplify the possibilities and limitations for spatial linked open data.

The case study showed that there are multiple users who can benefit from spatial linked open data, divided in two user types do layman users benefit in a better understanding of the data better and finding related information. The expert users benefit in the asking of spatial queries over the data. This can be used to create an inventory or simple analysis over the data. For more complicated queries, for transformations and data processing a GIS is needed.

The results showed that modeling spatial data according to the OGC standard for spatial linked data, the GeoSPARQL ontology, provides the possibility to ask spatial queries over the data to do a simple analysis or inventory over multiple heterogeneous and distributed datasets, by dynamically and implicitly linking datasets together on their location. As of now, not all tools for creating linked data support spatial data yet, therefore it can be time consuming to convert the data to spatial linked data. The on-the-fly method, where the actual data is converted on the fly to linked data would be ideal, in terms of data management, since the data is more up-to-date and not copied. Spatial linked open data has benefit in cross knowledge domain issues where multiple heterogeneous datasets are required, such as in emergency management and urban planning.

1. INTRODUCTION

Most of the data that exists, has a spatial context. This spatial data used to be only for GIS experts, but since the EU INSPIRE Directive mandates governments to be transparent, open data is more and more accessible on the Web for everyone (data.gov.uk, n.d.). Recently the W3C and OGC have announced to work together to improve the integration of spatial information on the Web. This as result of the value and potential spatial information on the Web has to a broad range of parties (Geonovum, 2015). Search engine providers collect spatial data to aid location based searching (Goodwin, Hart, & Dolbear, 2009) and many online Web/ and smartphone applications provide location based content, such as Twitter and Facebook. Users on the Web ask often spatially related questions (“I am here, where is the nearest supermarket?”). Such a question can be easily answered with a GIS. But the users also would like to know more, for example the opening hours and also the public transport timetables. This linking or combining of information from a GIS with the Web automatically proves currently to be difficult (Geonovum, 2015). And users that do not have experience with a GIS have to spend some time on finding answers on the Web.

Ongoing activities, such as the development of the Semantic Web help to break spatial data out of its isolation (Knibbe, 2014b). The Semantic Web, also known as the Web of Data, is a place where all the data is connected to each other so that users can discover and browse endlessly (W3C, 2013b). But the Semantic Web is not only about relating data from other sources, but also about data sharing and data interoperability. To achieve all this, Linked Data is used (W3C, 2013a).

Linked Data is a way of representing data with its context to other data. It is written in the standard RDF where data is represented in the format of subject - predicate – object, which is called a triple. The predicate is the link or relation between the subject and object. The subject, predicate and object are all identified by an URI, which is preferred to be a HTTP-link (Hartig & Langegger, 2010). This is explained in figure 1. The building ‘De Rotterdam’ is located in the neighbourhood Kop van Zuid. The HTTP-links refer to information on the Web, which ensures users to find additional information (Hartig & Langegger, 2010). For example, on the website of ‘De Rotterdam’, information about the architect and the opening hours of the building can be found.

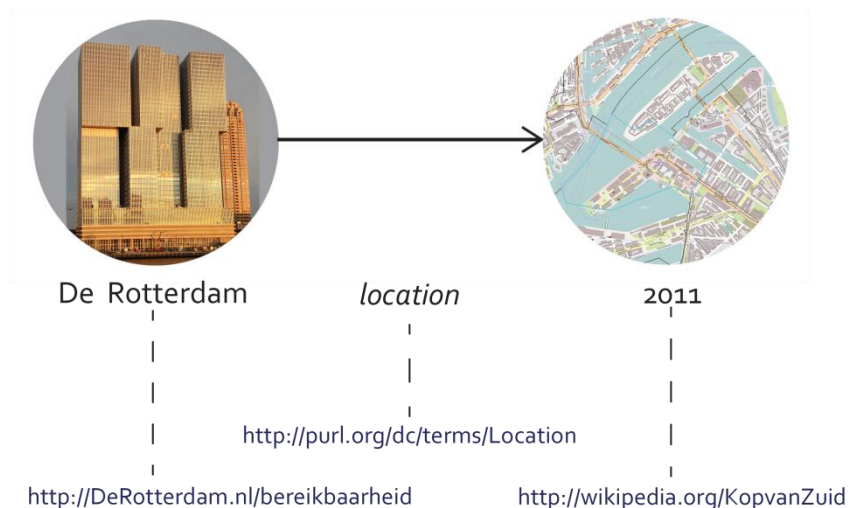


Fig. 1. Example of a triple

One of the advantages of Linked Data is that by modeling data in the same standard, datasets can be combined even though their data schemas differ. And with Linked Data comes the possibility to do SQL-like queries on the Web, with the therefore designed standard SPARQL (see section 5.4 for an example).

According to literature, the Linked Data method can aid in combining spatial data with other datasets, to provide more content to users, helping solve cross knowledge domain issues and helps in aiding location based services (Battle & Kolas, 2011a). Linked Data is an important incentive for open data (Kuhn, Kauppinen, & Janowicz, 2014). And Kuhn, Kauppinen & Janowicz even state : “As for Open Data and Linked Data in general, the GIScience community has a great opportunity to help exploit location as an integrator across platforms, domains and disciplines,” (Kuhn et al., 2014).

More and more organizations and companies in the Netherlands are interested in (spatial) linked (open) data and are researching its use. These researches are exploratory. For example the Tax Administration, the Ministry of Infrastructure and Environment, the Kadaster, TNO and Geonovum. The PiLOD (Platform Implementatie Linked Open Data) is founded to manage and research the use of Linked Data and by the growing and active organization concerning Linked Data (Platform Linked Data Nederland, nd.).

Despite the many interest for spatial linked open data, there are still a lot of challenges regarding the publishing of spatial Linked Data on the Web. Spatial data contains often geometries. These challenges include specifying of the topology of the data (how the spatial objects are related to each other), the conversion of geometry to RDF, the way to link metadata with spatial data (Knibbe, 2014b) and what the best method is to link spatial data with other types of data (W3C, 2014).

This thesis will therefore provide insight in how to apply the linked data method for spatial open data in order to profit from the benefits of linked data. This

research provides a methodology to create spatial linked open data, making use of existing standards and tools. The municipality of Rotterdam has stated its interest in the possibilities of spatial linked open data, which creates the opportunity for a case study to research what spatial linked open data can mean for the municipality. In this case study, the municipality will be introduced to the possibilities of spatial linked open data and insight will be provided in if and how their data can be converted to spatial linked open data. The outcome of this research will guide municipalities and others in creating spatial linked data as well as provide insight in its possibilities and limitations. The research will provide the incentive for municipalities to open up their data in a way that it is facilitated for society, so that not only experts users can make use of spatial data, but also layman users.

1.1. MOTIVATION

The Municipality of Rotterdam, along with other municipalities, provides open data to its citizens to stimulate innovation and to create a transparent government. This data can be useful for education, business and research (Gemeente Rotterdam, n.d.-b). The main website of Rotterdam (<http://www.rotterdam.nl/>) and the website Rotterdam Open Data Store (<http://www.rotterdamopendata.nl/dataset>) provide open (spatial) datasets, with a maximum of 3 stars on the 5 star data model (see section 3.1). Also on GisWeb, an online GIS with an authorization option for employees of the Municipality are open spatial datasets of Rotterdam visualized on a map.

However, for non-professional users the available open data can be difficult accessible and therefore an answer to their spatially related question is hard to find. For example, users need to browse through all the file folders first or view them on a map, to discover what each dataset is about, to find the right dataset they are searching for (see figure 2). They rely heavily on the metadata, in particular the semantic description of terms and the provenance of the data, while this is mostly sparse or not available. Also, some datasets are only available in specific formats such as Shapefiles, WMS/WFS, KML and more (see figure 3). Users who are not familiar with these formats can have trouble opening these files, since they need additional software for this.

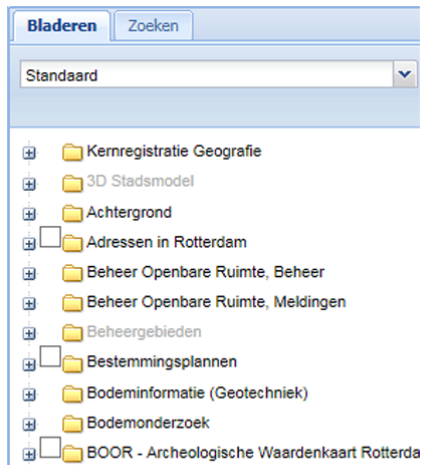


Fig 2. File folder structure containing different datasets in GisWeb (GIS-diensten, n.d.)



Fig.3. On the right Examples of 'specific' GIS related data formats in the Rotterdam Open Data Store (RotterdamOpenDataStore, n.d.)

Spatial Open data is thus available and can be downloaded, but a more accessible way of obtaining and viewing the data would make these websites more user-friendly and helps the Municipality with their task to be transparent and to provide data to its citizens. By providing data as linked data, the notion Anybody can say Anything about Any topic (AAA). Which means that anyone can access, process and make conclusions about the data. This results in different opinions and perspectives (Platform Linked Data Nederland, 2014). Which can also stimulate innovation. With spatial Linked Open Data more value to the existing websites can be added in terms of information services, the accessibility of the data, quality of the data and in terms of user friendliness.

1.2. RESEARCH OBJECTIVE

Therefore the following research question for this graduation thesis is formed:

WHAT ARE POSSIBILITIES AND LIMITATIONS OF THE USE OF SPATIAL LINKED OPEN DATA FOR THE MUNICIPALITY OF ROTTERDAM?

To answer the research question, a literature study and a case study will be done. The literature study provides the needed knowledge about Linked Data for the case study. The case study will exist of creating and visualizing the spatial Linked Data on a Web map. To create this, various aspects of linked data are touched upon, such as the conversion to linked data, the publishing of the linked

data on the Web and the querying of the data. With the Web map the Municipality can be introduced to Linked Data and its possibilities in a visual stimulating way. The end result of the case study will be a conclusion if Linked Data is of importance to the Municipality. Assessing the research will be done by providing an answer on the question if spatial Linked Open Data provides new services and or provides a more effective and efficient data management for the Municipality of Rotterdam.

The case study will entail the interest of the Municipality, namely a graphic visualization of statistical data, especially the composition of the neighborhood (demography, amount of facilities). The goal of the application will be to show the composition of neighborhoods and areas.

The sub questions that define the research are:

1. How to publish spatial data as Linked Data on the Web?

Two approaches, namely the conversion and explicit storage of RDF data in a triple store and the live conversion of spatial data in the database to RDF on the Web are tested and compared to find the most effective method.

Also the OGC standard, the GeoSPARQL ontology (Perry & Herring, 2012) for querying and modeling spatial data, will be implemented

2. What are options for linking datasets?

To link datasets with each other, different possibilities arise, depending on the nature of the datasets and are described.

3. How to query distributed and heterogeneous spatial datasets?

The querying of Linked Data is another form of bringing multiple datasets together. With help of the query language for RDF: SPARQL (see appendix 1 for an example) an answer to this question can be found.

4. How to create a Web application with linked data?

With a Web application, an interactive map displayed on the Web is meant.

The sub questions cover the main themes of linked data, namely the modeling of the data in RDF, the storage of triples, publishing data to the Web, the linking with other distributed linked datasets and the querying and visualizing of linked datasets. The answer regarding the main objective will therefore also cover these themes.

1.3. THESIS OUTLINE

The thesis will first focus on the case study of the web application for viewing spatial linked open data of Rotterdam in chapter two. It will outline the data that will be used, the users and the functionality of the research. The third chapter will summarize some related projects that deal with (spatial) Linked (open) Data. The fourth chapter will outline the methodology and clarify shortly the steps that will be

taken. Chapter five will represent the literature study done on all aspects as told in the methodology. From then on the literature part is finished and the research part will start with chapter six that focuses on creating the spatial linked open data. It describes the conversion considering the steps from the methodology. Chapter seven outlines the possibilities with the obtained spatial linked open data from the previous chapter. Chapter eight will focus on the last step of the research, the visualizing of spatial Linked Data. The thesis will end with the conclusions and recommendations for future work in chapter nine and the reflection is stated in chapter ten. In appendix 1 are the used prefixes and namespaces mentioned which will be used throughout several chapters.

2. CASE STUDY

In order to answer the research question, it is important to state beforehand who are the users of the spatial Linked Open Data and what kind of questions the application will be answering. This is important to specify, to guide the conversion to Linked Data in a certain direction and show the possibilities of it more clearly. The following sections explain therefore the goal of the case study: the spatial Linked Data Web application in more detail.

2.1. FOCUS OF THE CASE STUDY

The Municipality of Rotterdam (encircled in black in fig. 4.) is located in the province Zuid-Holland. Rotterdam Zuid, which will be the research area, is the encircled green part in the figure. Rotterdam Zuid consists of seven areas, which are then divided into several neighborhoods.

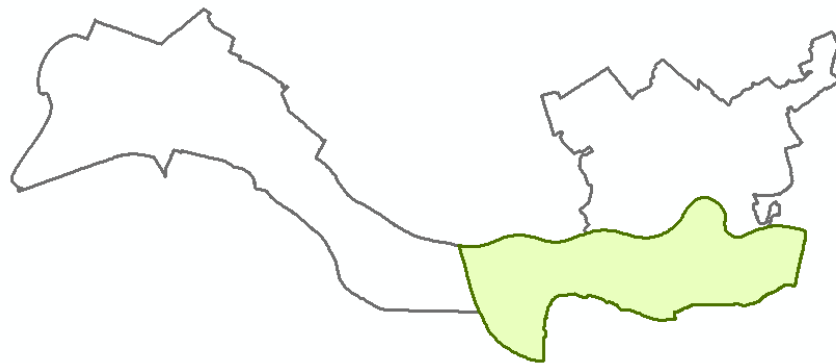


Fig. 4. Municipality of Rotterdam where Rotterdam Zuid is highlighted in green

The focus will lie on these seven areas, since the Web application can be of importance for The national program Rotterdam Zuid (NPRZ). This project focuses on improving the quality of living in Rotterdam Zuid for the citizens of Rotterdam Zuid (Gemeente Rotterdam, n.d.-a). The project focuses on three themes, namely living, education and work.

By visualizing statistics about the composition of the neighborhoods in Rotterdam Zuid on a map, new insights can arise of how to improve the quality. By visualizing data in a different manner, people can look at things in a new way. By showing links to other data, a more complete picture arises which can lead to new conclusions being formed. That is also one of the advantages of Linked Open Data,

that with Linked Data a different way of looking at data is assumed: the Open World Assumption. This means that to have the full picture, data from various sources is needed and that people realize that digital systems do not hold all the information, but a combination of data sets is needed (Platform Linked Data Nederland, 2014).

2.2. RESEARCH AREA

Rotterdam Zuid consists of the areas Feijenoord, Charlois, IJsselmonde, Hoogvliet, Pernis, Waalhaven and Vondelingenplaat (see figure 5). The Waalhaven is now mostly an inactive harbor area, Vondelingenplaat is also a harbor area and is still an active industrial area.

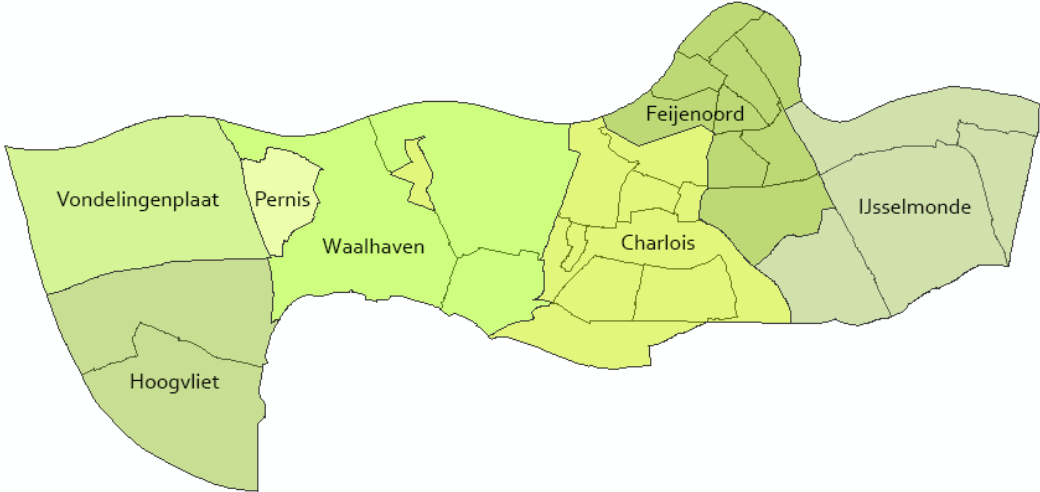


Fig.5. Rotterdam Zuid in detail

The areas Hoogvliet, Charlois, IJsselmonde and Feijenoord are residential areas. The neighborhood Kop van Zuid is a famous area in Feijenoord, since it houses many high-rise buildings famous for its architecture, such as De Rotterdam, where (part of) the Municipality is located.

2.3. CHOICE OF DATASETS

Several datasets/data sources that provide an overview of the areas and neighborhoods in Rotterdam Zuid are listed in the table below. Since the focus of this thesis is Open Data, only open datasets are chosen (see table 1).

Type of data	Open Dataset
Geometry of the boundaries of the area and neighborhood	CBS statistical neighborhood and area data (CBS wijk- en buurtkaart)
Statistical data of Rotterdam	Data from Rotterdam in Cijfers Data from Rotterdam Open Data Store
Base map	GBK or OpenStreetMap
Enrichment	BAG DBpedia

Table 1. Chosen open datasets

In the following sections, these datasets will be explained.

2.3.1 OPEN DATA OF THE CENTRAL GOVERNMENT

The Dutch central government established twelve base registrations (the system of base registrations) to provide efficiency while performing public tasks. Municipalities and other authorities that perform public tasks are mandated to use the system of base registrations. The data covers topics such as topographic data, data about vehicles and personal data (Digitale overheid, n.d.). A few of these base registrations are available as open data.

One of these base registration that is available as open data is the BAG. The Building register for Addresses and Buildings (BAG) exists out of two base registers: the register for addresses and the register for buildings. This dataset describes the addresses and locations of, among others, premises, mooring locations and pitch locations in the Netherlands, along with administrative and informative attributes (see figure 6). Each Municipality is responsible for their part of the dataset and the Kadaster manages the whole (Kadaster, n.d.).

The BAG is chosen because it has some advantages that make it interesting to publish as Linked Data: it contains addresses and buildings with attributes which provides an overview of the areas and neighborhoods. Therefore, statistics can be linked to the addresses and/or buildings. Also, because the BAG is a spatial dataset that contains point and polygon geometry, research can be done on how to convert those geometries to RDF. Furthermore is the BAG an high quality dataset which is publicly available and is relatively complicated, since multiple files/tables need to be combined to obtain relevant data, which can be used as an example to find out how difficult it is to convert the BAG to Linked Data.

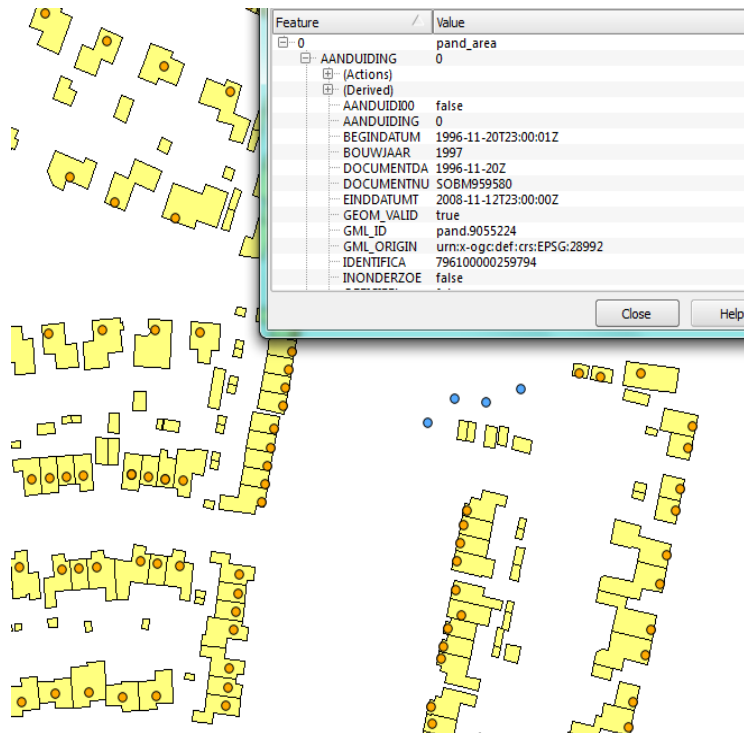


Fig. 6. Example of the BAG with the premises in yellow along with the attributes, the residential units in orange and the address points in blue

2.3.2 OPEN DATA OF ROTTERDAM

Also the Municipalities provides open data at different websites. An overview of the websites where spatial open data (in different formats, but not in RDF) is provided, is given below.

MAIN WEBSITE OF ROTTERDAM

www.rotterdam.nl

This website is the main website of Rotterdam for the citizens of Rotterdam. It consists of all the information about the Municipality of Rotterdam. Also, open data is available for download, such as for example the 3D City model of Rotterdam.

GISWEB 2.1

<http://www.gis.rotterdam.nl/gisweb2/default.aspx>

The GisWeb website is an online GIS that shows open data, created by the municipality itself. There is an authorisation function for municipality employees to provide access to not open data. Users of GisWeb are thus the Municipality to assist them in their task of managing the public space and also other (unknown) users. The open data that is available on GisWeb is mostly also available on the RotterdamOpenDatastore and on the main website of Rotterdam and comes from the Municipality itself, corporate and private parties (GIS-diensten, n.d.).

In the online GIS, first a base map (aerial foto/black and white map/GBK map) can be chosen. Different layers can then be overlaid on the chosen base map.

Functionality in the form of filtering on column name or spatially in the form of features inside a buffer around a point, line and polygon is possible. Also, a WMS with GetFeature functionality is provided to add additional information to the features (see figure 7). The data can be downloaded as bitmap images. The metadata is given according to the ISO19115 standard (Metadata standard for spatial data), added with some additional fields created by the Municipality.

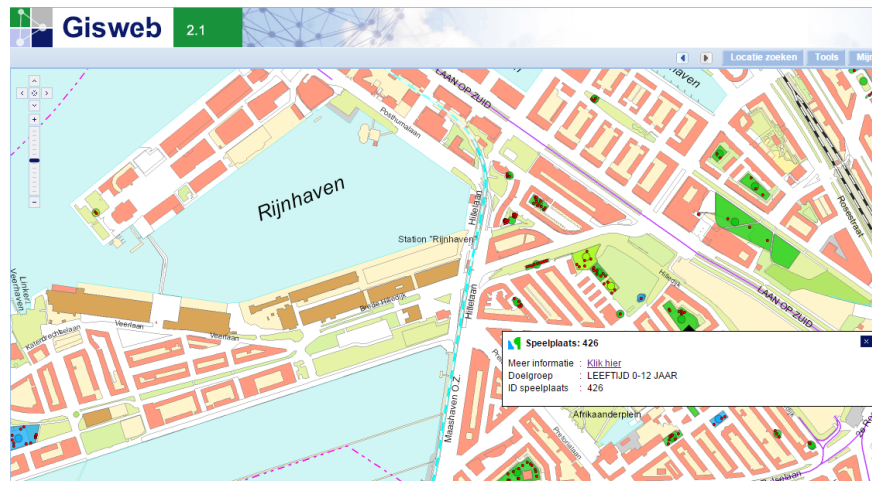


Fig.7. GisWeb viewer showing layers with playground equipment and playgrounds.

THE ROTTERDAM OPEN DATA STORE

<http://www.rotterdamopendata.nl/dataset>

The Rotterdam Open Data Store is founded in 2012 and is an initiative from the Hogeschool Rotterdam, the municipality and various other companies/organizations (Rotterdam Open Data Store, n.d.). The Rotterdam Open Data Store provides a download service of data from different sources, though the Municipality is by far the biggest supplier of open data. Currently (may 2015), there are 181 datasets uploaded in the Data Store, consisting of different data formats. The most used way to provide data (from the Municipality) is by a WMS/WFS service. The topics concerning the data consist of, among others, administrative data, height data, aerial imagery and safety indexes of neighbourhoods. Metadata is very briefly provided and the data store seems to focus more on expert users of GIS/ geo formats.

ROTTERDAM IN CIJFERS

www.rotterdamincijfers.nl

This website is maintained by the Research and Business Intelligence (OBI) department of the Municipality. They provide statistical data of the Municipality from their own registrations and external sources such as the business and employment register Zuid- Holland and crime and notifications data from the Police Region Rotterdam-Rijnmond). The data covers subjects as demography, education, liveability and safety, data about dwellings and economy. This data is

updated yearly and can be downloaded. Most of the consumers are employees of the municipality, employees of other municipalities, HBO and WO students, researchers from universities, housing associations and others (not known) (R.Weber, personal communication, 2015) .

Datasets from the Rotterdam Open Data Store and Rotterdam in Cijfers will be used to convert to Linked Data, since these datasets are downloadable for use.

2.3.3. CBS STATISTICAL NEIGHBORHOOD AND AREA DATA

Statistics Netherlands (CBS) and the Cadastre (Kadaster) provide the statistical neighborhood and area data (CBS wijk- en buurtkaart). This dataset consist of a set of three corresponding shapefiles with the boundaries of municipalities, areas and neighborhoods of the whole of the Netherlands.

BU_CODE	BU_NAAM	WK_CODE	GM_CODE	GM_NAAM	WATER	POSTCODE	DEK PERC	AANT INW	AANT MAN	AANT VROUW	P 00 14 JR	P 15 24 JR	P 25 44 JR
BU05990110	Stadsdriehoek	WK059901	GM0599	Rotterdam	NEE	3011	1	13110	6855	6255	7	15	48
BU05990111	Oude Westen	WK059901	GM0599	Rotterdam	NEE	3014	2	9495	4860	4635	15	16	30
BU05990112	Cool	WK059901	GM0599	Rotterdam	NEE	3012	1	4320	2155	2160	8	13	42

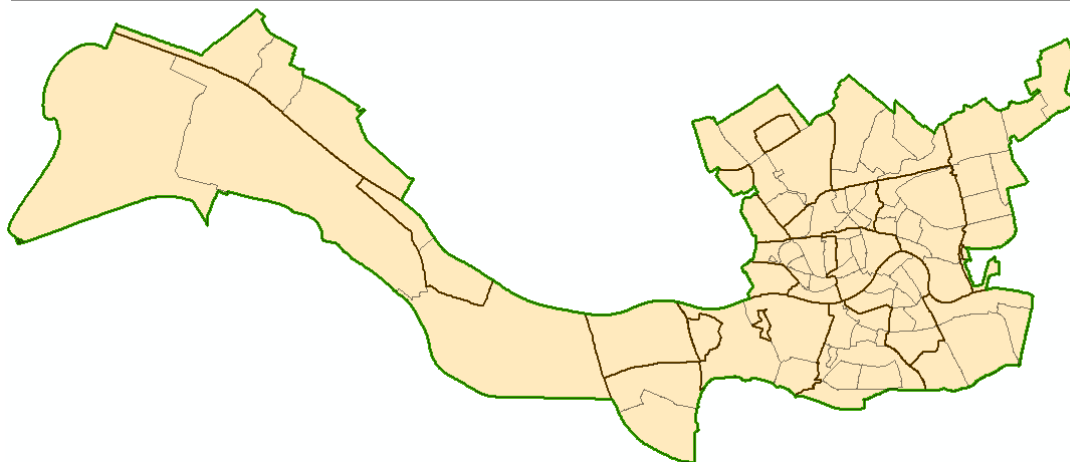


Fig.8. CBS area and neighbour map of 2013 and part of the neighbour attribute table

In figure 8 these corresponding shapefiles are visualized for the Municipality of Rotterdam. The green encircled area is the municipality boundary, the brown encircled areas the areas of Rotterdam and the light grey lines visualize the neighborhoods in Rotterdam. Each of these three datasets contain demographical statistical data such as the number of households, number of cars and the age distribution. Every year an updated version of the dataset is published. The boundary geometry comes from among other the base registration Kadaster and the statistical data from the CBS itself (CBS, 2015).

The latest update of the dataset is from 2014, which will also be converted to Linked Data, since this dataset contains boundary geometry and also statistical data. By creating a mix between (statistical) data from different sources, a more reliable application can be made. One source can have more/additional information about the same theme than another source, which results in enrichment of information.

2.4. FUNCTIONALITY WEB APPLICATION

A concept of the web application and the used datasets can be seen in the image below (see figure 9).

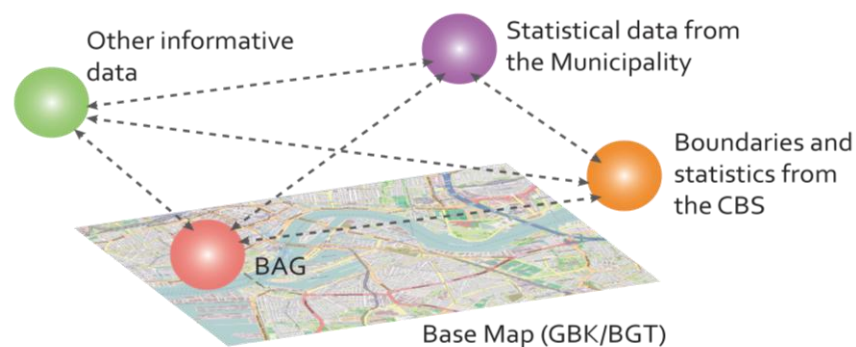


Fig. 9. Layers in the Web Application

Basically the web app will consist of a base map with displayed on that the result of a spatial question (query) over one or more linked datasets. The base map is preferred to be the large scale map of the municipality itself, since the Web app shows data from the municipality. Otherwise, OpenStreetMap can be used. Since the application is considered to be used also by non-professional GIS users, there should be an user friendly option for doing the queries, such as predefined buttons that perform a simple query. A SPARQL interface for professional users is on the other hand also preferred, to be able to do more complicated queries.

The users will, apart from the Linked Data functionality such as discovering additional information, not notice that the application is working with Linked Data. Users who are familiar with Linked Data will notice this however, by the possibility to ask spatially related questions.

2.5. END USERS

There are different end users who can make use of the web application, though they all want to have an answer to the general question: *“What does the neighborhood look*

like"? Since the data is open data, it is considered that everyone who is interested in his/her surroundings can make use of the application. Basically there are two types of users who benefit from the web application, namely the layman and the expert users.

2.5.1. USER TYPE 1: LAYMAN USERS

The layman users are the users who do not have experience with Linked Data Technologies, such as RDF and SPARQL. They will not notice that the application works on linked data, but they will however notice the advantages of Linked Data by finding additional information.

A few examples of end users and questions they might want to have answered are given below:

- Citizens; who want to move to a new house and want to investigate the neighborhood:

Which utilities are in a 10 km radius around my new house?

Dependent on the chosen datasets: a list/visualization on a map of schools, supermarkets, parks, stores and sport clubs will come up, along with the opening hours and advertisements.

- Community organizations; which represents the interests of the neighborhood and want to fit their entertainment program to the average age:

How is the age distribution in my neighborhood?

Dependent on the chosen datasets: a list/visualization on a map of the age distribution will be shown, along with their social-economic classes.

- Shops/companies; that want to settle themselves in a neighborhood:

What kind of stores/companies are already in this neighborhood?

Dependent on the chosen datasets: a list/visualization on a map of the stores and companies, their functions, opening hours and other branches comes up.

FUNCTIONALITY WEB APPLICATION

For the layman users the web application screen provides the following functions. Showing the data on a web map with clickable objects and predefined buttons or menu options that send a query to the SPARQL endpoint, without the user having to write the query itself.

2.5.2. USER TYPE 2: EXPERT USERS

The expert users have knowledge of Linked Data Technologies and the Semantic Web. They would notice the usage of Linked Data from the use of the URIs as identifiers and the links that are provided to other datasets.

These users could be:

- PHD students/researchers in the Semantic Web
- Information software companies focusing on Semantic Technologies
- Individuals who are already familiar with the Semantic Web

These users can make use of the RDF data in their own web application environment.

FUNCTIONALITY WEB APPLICATION

For the expert users the web application should have a SPARQL- endpoint. Also, the data for download in RDF is preferred.

The cases study provides insight into the use of spatial linked open data. It shows how the spatial linked open data looks, how to do spatial queries over the data and how to link and combine the data. Therefore the case study shows the possibilities and limitations. The expected result is that the both types of users benefit from spatial linked open data.

3. BACKGROUND

This chapter describes the background information and related work concerning spatial Linked Open Data.

3.1. THE 5 STAR SCHEME

Tim Berners-Lee, the creator of the Web and the pioneer of Linked Data, designed in 2006 the 5 star scheme to reach the full potential of open data, namely providing additional information to users so they can keep browsing. The scheme can be explained as follows.

1. Data is available on the Web with an open license
2. The data is structured and available on the Web with an open license
3. The data is available as a non-proprietary format
4. URIs are used as identifiers, so that there can be pointed to individual things, and using existing standards such as RDF and SPARQL
5. Links between the dataset and other datasets are made
(Berners-Lee, 2006)

The first three stars indicate Open Data, while the fourth and fifth stars indicate Linked Open Data. The 5 star scheme has a lot to do with what data format the data is published to the Web. In table 2 the number of stars and the accompanying data format is depicted. According to (Gillies, 2013), the 5 star scheme can as follows be implemented for spatial data.

Stars	Rule	Data format Source: (5stardata, 2012)	Spatial data format Source: (Gillies, 2013)
1 star	Open license	PDF,png,jpeg	PDF,png,jpeg
2 stars	Structured data	Excel	Shp, geo- database
3 stars	Non-proprietary format	CSV	KML, GeoJSON, RDF
4 stars	URIs as IDs	RDF	RDF
5 stars	Links to other data	RDF	RDF

Table2. 5 star scheme and data format

3.2. LINKED DATA METHODOLOGY

To create linked data, a guideline can be used. One of them is the W3C Linked Data Cookbook of 2011, as developed by the W3C Government Linked Data Working Group, which aims to provide guidance to (governments) for publishing open data, in particular 4 and 5 star linked data. The cookbook is intended for everyone that is working on the process of opening up governmental data (Hyland & Villazón Terrazas, 2011). In the Cookbook, seven Best Practices are defined which will be shortly described here:

1. Data modeling; This consist of identifying, modeling the data to existing vocabularies, the naming, in where you use URIs as identifiers and test the modeled data if its logical. This can be a time consuming event and difficult to do, but this effort is only done one time and then the many benefits of linked data can be reaped.

2. Use URIs as identifiers; in where it is a best practice to use HTTP URIs. Also take in advance an URI strategy.

3. Use existing vocabularies if possible to model the data.

4. Provide metadata with descriptions about the data

5. Conversion of the data to RDF; This can be done in three ways, namely automatically (good for large datasets, but further not so much recommended since links to another dataset are not made and thus misses a potential of linked data), partially scripted and the last one modeled first and then scripted. This second to last and last method are recommended, since it ensures a higher quality of data. Also, a recommendation is to not model the data for one application or for a specific use, since that would diminish the use case of the linked data.

6. Choose an appropriate use license for the linked dataset

7. Publish the linked data to the Web and advertise the data; Publish the data on the Web means that the data is available and accessible for everyone on a Web page. (Hyland & Villazón Terrazas, 2011).

Another guideline is from TNO, who created a step-by-step guide for publishing Linked Open Data. To exemplify this manual, they converted open energy data from Liander, an energy company, to linked data in their project Linking Open Energy Data. The guide consists of nine steps to publish the data, namely:

1. Select the data you want to convert to linked data.

2. Prepare the data; by which is meant that the data is filtered, anonymized and the quality is improved if needed.

3. Model the data; if possible, according to existing vocabularies or ontologies.

4. Define a naming scheme; recommended is the implementation of the Dutch URI- strategy.

5. Convert the data; to convert the data, three strategies are defined, namely writing a script to go from data in a database data to RDF. Or using a tool such as OpenRefine (from Google) to do the mapping. Or use another program to make a direct mapping between the database and RDF data.

6. Organize governance; by using the BOMOD model, which is a model for management and development for Open Data.

7. Add metadata; Add triples concerning the metadata, according to the Dublin Core and VoID vocabularies (see section 5.3.).

8. Publish the data; which can be done in four ways, namely, making the data publicly available on a Web server, or storing the data in a triple store (storage solution for triples) and making it accessible with a SPARQL endpoint, or provide mappings on-the fly from a relational database and the last approach is to create a web application on top of the triple store.

9. Link the data; in this last step, the Linked Dataset is linked to ontologies (existing data models) and to other datasets. These steps are exemplified and applied to the energy data (Steen, Eckartz, & Folmer, 2014).

Lately, multiple governments/organizations have worked on creating and publishing Linked Data as well as spatial Linked Data. The next section describes a few related projects.

3.3. SPATIAL LINKED DATA PROJECTS ABROAD

One of the first examples of publishing spatial Linked Data on the Web is from the Ordnance Survey in the UK. For the project “Making Public Data Public”, three open data sets in the domain of administrative geography as Linked Data were published. The datasets are a list of administrative regions (polygon data), the Code-Point Open Linked Data (point data) and the Boundary line Linked Data (line data). The data descriptions contain the name and spatial relationships. These datasets are shown on a map in a web viewer and can be queried (Ordnance Survey, n.d.), as can be seen in figure 10.

London



London is a European Region.	
Objects related to "London"	
Extent	41428-1
Westminster Constituency	Uxbridge and South Ruislip West Ham Feltham and Heston Romford Streatham Westminster North Chipping Barnet Ealing North Lewisham East Chingford and Woodford Green
Within	England
Touches	Eastern South East

Core facts about "London"	
Type	European Region
Label	London
Pref Label	London
Northing	176392.3
Easting	529613.8
Lat	51.489632
Long	-0.134505
Area Code	EUR
Gss Code	E15000007
Point	51.489632 -0.134505
UnitID	41428

Fig. 10. Screenshot of the Web viewer of Ordnance Survey viewing London (Ordnance Survey, 2015)

The used vocabularies that describe the spatial relations between administrative units are defined by their own Administrative Geography vocabulary. These spatial relations are explicitly stated in the predicates (Goodwin et al., 2009). The map viewer works well, however the possibility to do queries in the SPARQL endpoint seems only to work with the example queries. If another query is done, no answer is given back.

A second noticeable example is from Germany where the Institute of Computer Science of the University of Leipzig has published OpenStreetMap (OSM) data as RDF on the web and made available on a map in a web viewer. The name of this project is LinkedGeoData. To convert the OSM data, stored in a relational database, they used Triplify to convert the results from queries on the relational database, to RDF. The generation of the triples with Triplify is done 'on the fly'. In Triplify a mapping file needs to be constructed by the user to generate the triples (Auer, Lehmann, & Hellmann, 2009). Currently the web application does only show an OpenStreetMap map with no Linked Data visualized on it anymore.

Also in other countries are developments for publishing spatial data as Linked Data, namely in Ecuador, where they is an initiative to create the Ecuadorian Geospatial Linked Data repository (Saquicela, Espinoza, Piedra, & Villazón-Terrazas, 2014). Also in Spain for the GeoLinked Data Case, where they created a

Web browser to visualize point location data (Vilches-Blázquez, Terrazas, Leon, Priyatna, & Corcho, 2010).

3.4. LINKED DATA IN THE NETHERLANDS

A spatial linked data project from the Netherlands is from the DEN, the Digital Heritage Netherlands which is an organization that encourages heritage institutions with their ICT. DEN is sponsored by the Ministry of Education, Cultural Affairs and Science (DEN, 2014). One of their projects is Heritage & Location, which focuses on the sharing of historical information and innovative services to attract more public (Veer van 't, 2014a). For this project, a web application is developed where archeological monuments are visualized on a map. They implemented the GeoSPARQL ontology (OGC standard to model and define spatial relations and spatial querying, see section 5.5.2) and used a number of programs. The data in the PostgreSQL database was converted on the fly to RDF with the D2RQ platform (see section 5.8). As triple store USeekM was used to perform the spatial queries and for the web application OpenLayers is used (Veer van 't, 2014b). The demo is still online and shows the result of a spatial intersect query.

3.4.1. PILOD USE CASES

Platform Linked Data Netherlands has also several working groups which focus on (spatial) data and/or base registrations as linked data. One of them is the recent finished project the BGT as Linked Data. This project is performed by Geonovum and a team of other experts, on behalf of the Ministry of Infrastructure and Environment. In this proof of concept, a conversion of the Base registration large scale topography (Basisregistratie Grootchalige Topografie- BGT), valuation data of properties (Waardering Onroerende Zaken-WOZ), trade register (Handelsregister-NHR) and management of public space data (Beheer Openbare Ruimte-BOR) for the test area Leiden are converted to linked data. Also, the BAG as linked data is used. All these linked datasets and made available through a Web application.

One of the challenges they faced during the project was that it required a long time for collecting all the data and creating the linked data. Another challenge was generating links between the datasets. Generating links in corresponding datasets was easy to facilitate, but for datasets that did not comply with each other was more difficult and needed to be done manually. The geometry (points) were generated by a WFS and the possibility to do spatial queries with GeoSPARQL did not yet succeed in the available time frame (Brink, 2014).

Another use case is the URI-strategy. The URI-strategy is developed by one of the working groups from the PiLOD community, because they felt the need to identify

governmental data. They proposed a standard way of defining an URI to identify this type of data and to develop a national URI-strategy for the government in the Netherlands. Also the UK has such a strategy for their government. The URI-strategy is mainly for the base registrations and other Dutch standards, because they contain data that is relevant for all sorts of applications. The strategy is not intended for data that is not from the base registrations and Dutch standards, since they mean that this kind of data is less linked to. The URI- strategy is currently in development to become a standard (Overbeek & Brink van den, 2013).

The URI-strategy is composed of the following pattern (see figure 11):

`http://{domain}/{type}/{concept}/{reference}`

Fig. 11. URI-pattern from the URI-strategy (Overbeek & Brink van den, 2013)

The domain basically represents the reliability and recognizability of the data publisher. The type can be one of three things:

- id: which identifies the data
- doc: which defines the metadata about an object
- def: which describes a term in an ontology

The concept defines a group or name what the data is about. The concept helps to create unique ids and a comprehensible URI. The reference identifies an individual element of the data. (Overbeek & Brink van den, 2013). Examples of Linked Data that already implement the URI-strategy are the BAG as Linked Data and the BGT as Linked Data.

3.5. BAG AS LINKED DATA

There are a few organizations/companies that have created or are working on the BAG as Linked Data. The reason that the BAG is popular is because the BAG is an interesting dataset where many datasets can be combined with, since it contains addresses and premises and is rather complicated. The table below shows the versions of the BAG as linked data, which are currently published on the Web (see table 3). In the table the versions of the BAG are ranked. The criteria selected are based on an evaluation how good the original data is modeled as linked data. That includes if queries can be asked over the data, how would the dataset rate on the 5 star data scheme, is the data complete, thus does it contain the versions of objects in history and metadata, is the geometry modeled and if yes, is it done according to the GeoSPARQL ontology and is the data published to the Web so as to discover

additional information and is metadata published. The last row shows how the different versions score on these seven criteria.

	Kadaster	Geodan	Netage	Waag Society
Queryable	Yes	Yes	Yes	No
Number of stars	4	5	5	4
Including historical data	Yes	Yes	No	No
Including geometry	Yes	Yes	Yes	Yes
Modeled according to GeoSPARQL	Partially	Partially	No	No
Published to the Web	Yes	Yes	Yes	No
Contains metadata	Yes	Yes	Yes	Yes
Score	5,5/7	6,5/7	5/7	2/7

Table 3. Versions of the BAG

KADASTER

SPARQL endpoint: <http://almere.pilod.nl/sparql>

Vocabulary: <http://bag.kadaster.nl/>

A group of parties, in particular the Kadaster, have converted the BAG to linked data. The BAG as linked data is available on the server of the Platform Linked Data Community. This project focuses on the further development of the base registration where the focus now mostly lies on the semantics. Their project is ongoing (Brattinga, Overbeek, & Santema, 2014).

From querying the data in the SPARQL endpoint, can be seen that the URI strategy is used to define the http- links of the data. The total BAG is modeled, including the older versions of objects. The document number and data + time are combined together as 'gegevenscombinatie', which identifies an element. The data objects in the BAG contain geometry modeled according to the GeoSPARQL standard, but since the data is stored in the Virtuoso triple store, the data type for geometry according for the Virtuoso triple store is used.

GEODAN

SPARQL endpoint: <http://lod.geodan.nl/sparql/>

Vocabulary: <http://lod.geodan.nl/vocab/bag/index.html>

The geo company Geodan recently did a project on publishing the BAG as Linked Data. They linked the BAG to DBpedia (Knibbe, 2014a) to make it 5 star data. From looking up the vocabulary and the data in the SPARQL endpoint, can be seen that the linked data is stored in the Virtuoso triple store. The geometry is modeled according to the GeoSPARQL standard, but since also the Virtuoso software is used, the data is not modeled fully to GeoSPARQL.

NETAGE

SPARQL endpoint: <http://data.resc.info/bag/sparql>

Vocabulary: <http://data.resc.info/bag/>

Netage, a software company, is one of the owners of a version of the BAG. The BAG from Netage is converted with the D2RQ platform to Linked Data (on-the-fly conversion), which can be seen when looking up the data vocabulary. The data is accessible by a SPARQL endpoint (named Snorql). All the columns of the BAG are converted, from only the valid premises. Therefore does this version of the BAG not contain any historical data. The Basic Geo vocabulary for point objects in the WGS84 CRS (see section 5.5.1) is used to define the lat/long positions of the point geometry for the addresses and residential units. The polygons of the premises are defined by a predicate from their own defined BAG vocabulary, also in WGS84. Important to mention is that this version of the BAG has a specific use case for the fire brigade (Platform Linked Data Nederland, 2015).

WAAG SOCIETY

Web application: <http://citysdk.waag.org/map/#layers/bag.pand/objects>

The CitySDK Linked Data API of Amsterdam is developed by Waag Society and shows open data from different sources as linked data on an OpenStreetMap of the Netherlands, but mostly for Amsterdam. The project is designed that users can instead of only viewing data, also add data themselves. The API is a web service that shows the data and where applicable links it to other datasets (Waag Society, n.d.-b) The web service is a standard used in several other cities in Europe, namely Manchester, Helsinki, Lisbon, Rome, Lamia (Greece) and Istanbul. The project is part of the European Union project, CitySDK project which focuses on improving information services and therefore provides a set of tools for cities (Waag Society, n.d.-a).

From looking up the BAG data in the CitySDK Linked Data API can be seen that this version of the BAG is differently made available to the users than the other versions. The BAG is as JSON-LD format (see section 8.3) layered over OpenStreetMap. The predicates are visible in the JSON-LD format on the right. Not all the data is converted for this version. The premises along with their construction year, state and identifier are visible on a map. The mooring locations of Amsterdam are viewable as a separate layer; they contain the address and identifier as attributes. If there was made a link to another dataset and if the data was published to the Web, thus if the data could be looked up on the Web, the data would be 5 stars.

Considering these observations, the BAG as linked data from the company Geodan, seems to be the most elaborated version that is available on the internet. Though, since the data is stored in the Virtuoso triple store and therefore contains geometry data types of Virtuoso, spatial queries according to the Virtuoso software can be asked, but GeoSPARQL queries cannot be performed on this data. This makes the linked data software dependent for spatial queries, which is not desirable. In this thesis I will therefore try to model the data according to the standard for spatial data.

3.6. SUMMARY

The discussed example projects show that there are, also in the Netherlands, a lot of parties researching spatial linked open data. Though every party focuses on different parts, from the conversion to the structure of the URI, to the semantics. The demo from Erfgoed and Locatie seems to be the most advanced Web application. The demo works well, however the data model is complicated and this makes the use of the app difficult. It seems that even though GeoSPARQL is implemented, other spatial queries cannot be asked.

This research aims to add to the existing projects a detailed description of a use case for the municipality, in where the full process of creating spatial linked open data, querying and the visualizing of spatial Linked Open Data will be addressed.

The URI strategy will also be implemented. Even though the URI- strategy is not intended for other types of data than the base registrations, it seems that this strategy would also benefit for data of municipalities, since they have a public task and assumed that their published open data will be much used, the URIs need to be defined well for citizens to find and understand the municipality data.

Also, a few parties have already converted the BAG to linked data. The idea of using the BAG for this thesis is to model the BAG according to the GeoSPARQL ontology. This requires to model the BAG again.

The guidelines for creating linked open data as described here, differ a bit from each other. The Cookbook focuses on creating the highest quality of linked data (5 stars), while this step comes last in the methodology of TNO (see step 9). The TNO guideline emphasizes the preparing, filtering, anonymizing of the data, while this step is not so much mentioned in the Cookbook. This has probably to do with the fact that TNO developed the guideline during a project and thus fit their steps to their project.

The guidelines for publishing linked open data could act as guides for my research. In the next chapter, the steps to perform the research will be described.

4. METHODOLOGY

This chapter describes and explains the steps that are to be taken to create spatial Linked Open Data and to visualize this data on a map on the Web.

4.1. OUTLINE OF THE METHODOLOGY

The methodology that will be used is based on the Cookbook and guide from TNO, with some changes. For creating linked data it is important to understand, analyze and filter the data. This, as a preparation step for the actual data modeling, to know which vocabularies to use. Also since not all data needs to be modeled to linked open data, depending on the use case. For example, the BAG contains administrative data such as document number and document data, this is more useful for expert users of the BAG and not so much for other users of open data. They are probably more interested in the address data or the area of a premise. For spatial data in particular it is important to prepare the data also, for example change the CRS to WGS84 to make more links to other existing datasets in the Semantic Web. Or for example filter out invalid polygons from the data to, so spatial queries will not go wrong. Also with spatial data, you might want to show the data on a map, since then the data is better readable in its context.

The Cookbook is against modeling data for a specific purpose, to make the data applicable for as many use cases, but I think that depending on the dataset and the use case for spatial data it is better to know beforehand the use case and the users, thus show the data on a web map.

Another observation is that it seems that the order of these steps of the guidelines seem to entail what the definitions is of Linked Data. In my opinion, the linking to other ontologies/vocabularies/datasets should happen before the publishing of the data and not as a last step as is done at the TNO step-by-step guide. This to ensure to create 5 star linked data.

To carry out the proposed research and create the spatial Linked Data Web application, several steps are taken. These steps can be seen in the flowchart below (see figure 12).

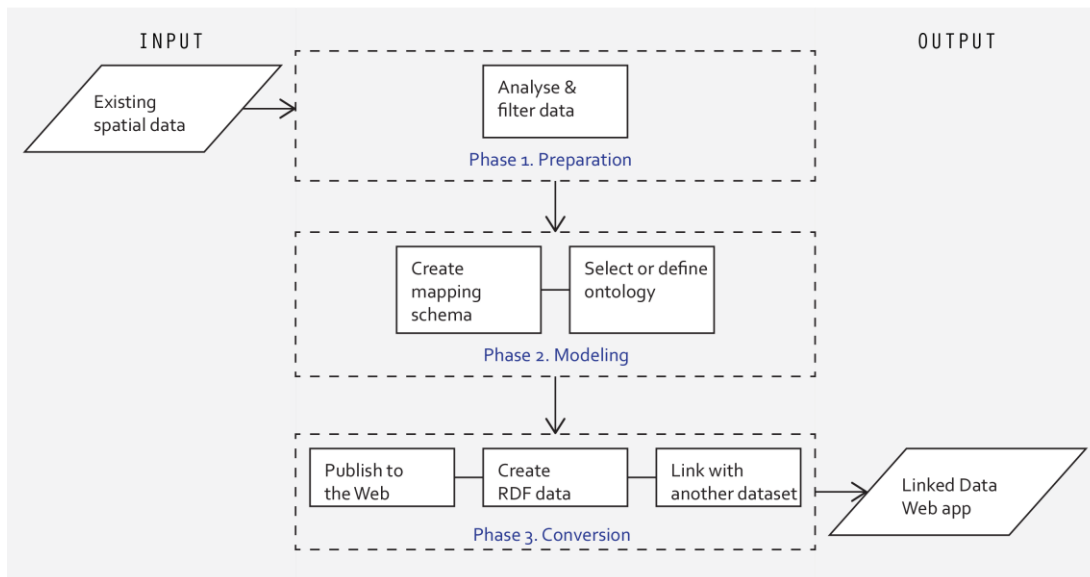


Fig. 12. Flowchart of creating a Linked Data Web application

Tim Berners-Lee has designed four rules to create Linked Data. The first refers to using URIs for the data. The second rule addresses the use of HTTP URIs, so users of the data can look the data up on the Web. The third rule states that information about a resource should be given, by using standards such as vocabularies and SPARQL. The last rule refers to providing links to other datasets, to provide more additional information (Berners-Lee, 2006). These rules will be also taken into account in the methodology. In the following sections, a few steps of the flowchart are described into more detail.

4.2. PHASE 1. PREPARATION

The first steps consist of acquiring the data, analyzing and filtering the data. The first step is to know in which data format the data is, in order to choose a suitable tool/software or approach to convert the data. Analyzing of the data is important to understand the data and for modeling it correctly to linked data. Filtering the data is important to select which part of the data needs to be mapped to linked data and to, if needed, improve the data quality. As mentioned at the start of this chapter, also for spatial data these steps are important.

4.3. PHASE 2. MODELING

4.3.1. DEFINING AN ONTOLOGY

After analyzing and filtering the data, the first step is to define an ontology. This ontology is part of the data model. This data model, which is required to re-use data, because if data is not understandable, it cannot be re-used, contains the semantics of how objects in the real world are mapped into a dataset (see figure 13). An ontology defines a knowledge domain and includes vocabularies. (LinkedDataTools.com, n.d.-a). Vocabularies are the semantics that describe the dataset. If an ontology for the theme of the spatial dataset does not exist yet, a new one needs to be created. Though it is preferred to use existing ontologies/vocabularies, to be able to link to datasets on the Web (Bizer, Cyganiak, & Heath, 2007).

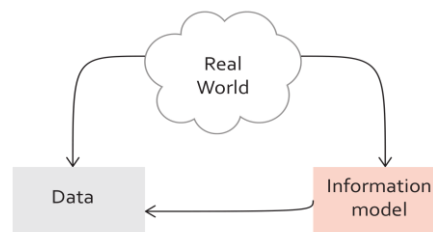


Fig. 13. Relation between the real world and data

First needs to be determined what type of data will be converted to RDF and determine which vocabularies fit that type of data, to define an ontology. A vocabulary exists out of several terms with the same meaning, regardless the context of the data (LinkedDataTools.com, n.d.-a). These terms form the predicates of the RDF data. Since vocabularies are used for the predicates and are described as semantics, the semantics are basically in the data. This makes the data more understandable. And by re-using predicates or vocabularies of others, data is even more understandable, because the meaning of the re-used predicates is more common and therefore known.

Another advantage of re-using predicates, is that data can be more easily merged, since data that uses the same predicates is in a way similar (Wood, 2011). Since there are datasets about different themes, different vocabularies that describe relations between different types of data exist (Linked Open Vocabulary, n.d.).

To be able to use existing terms, knowledge of the different vocabularies is needed.

4.3.2. MAPPING SCHEMA

Based on the analysis of the data and the chosen vocabularies as described above, the data model or graph model can be defined. This graph model is used as a mapping schema or mapping file for converting the data to RDF. Basically, certain tables, classes or groups of same objects in the data, become nodes in the graph (Bizer et al., 2007). The nodes are linked according to the relations these classes, tables or groups have between them.

4.4. PHASE 3. CONVERSION

4.4.1. CONVERSION

The next step will consist of converting the spatial data to RDF. Two approaches will be tested in this thesis on the most effective method. The most effective method is the one that performs best on data management and speed of conversion.

The two approaches make a distinction between actually storing the RDF triples in a triple store and converting the spatial data to RDF 'on the fly'. These approaches are visualized in a schematic way in the figures below.

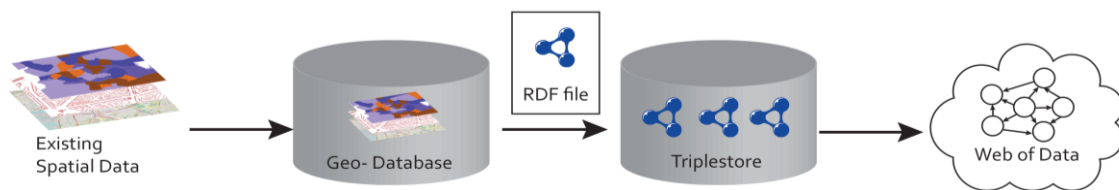


Fig.14. RDF storage approach

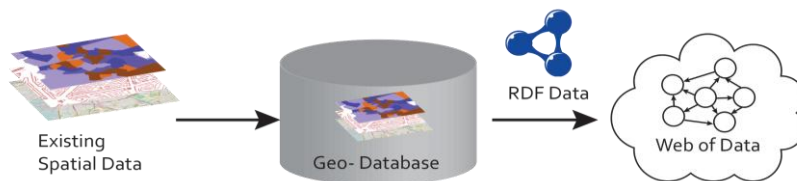


Fig.15. On- the- fly to RDF approach

The first approach (in figure 14) is the RDF storage approach which consist of converting the spatial dataset to RDF, then store it in a RDF triple store (storage system for triples) and publishing the RDF data to the Web. The second approach (as seen in figure 15) is the on- the- fly to RDF approach which consist of converting the spatial data in a geo-database to RDF 'on the fly', while publishing the data to the web. The second approach would seem to be better, since there is no duplication of data.

4.4.2. LINKING

The next step thereafter is to link the spatial Linked dataset to other, preferably distributed, datasets, to make actually Linked Data of the RDF data and to exploit the full potential of Linked Data.

To link to other existing Linked Data sets that describe similar objects owl:sameAs, rdfs:seeAlso foaf:page predicates can be used. Datasets can be linked on common keys in the data, label matching and ontology matching (Hassanzadeh, 2011). The linking can be done manually, but when the dataset is large, an automatic method is preferred. Examples of this can be a pattern-based or property-based algorithm (Bizer et al., 2007).

For this case study, there are several aspects to link datasets on. These are schematically visualized in figure 16. With matching based on geometry is meant performing a spatial query regarding the geometry. The BAG contains geometry of the premises and address points, the CBS neighborhood data also contains geometries of neighborhoods and areas. Therefore, a link between the BAG and the CBS data can be made by doing a spatial contain/intersect function for all the point or polygon objects of the BAG that are in the geometry of the CBS data. This is a dynamic link, since it is made through a query.

With semantic linking is meant that attributes that have the same meaning can be linked. In the figure can be seen that the zip code from the BAG is linked to an area or neighborhood from the CBS dataset to where that zip code belongs to.

With string matching is meant that the same names are linked, in the figure this is the same neighborhood name. Essential for this is that the names are spelled accordingly. These three types of linking can be done through either a query (the matching based on geometry), and the semantic and string matching can also be made by manually creating triples.

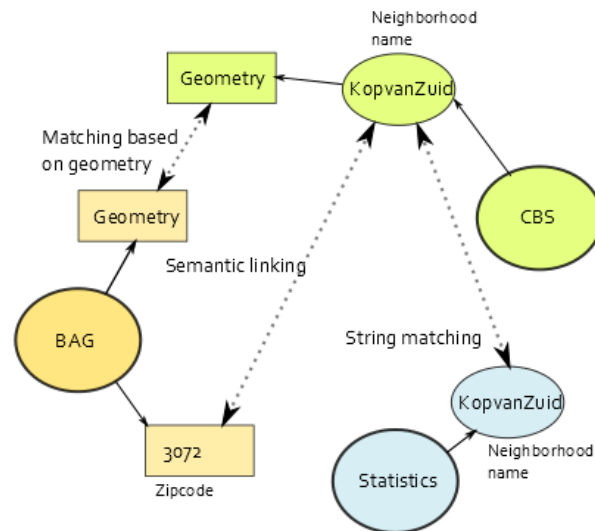


Fig. 16. Schematic way of linking datasets

To link to other distributed datasets such as DBpedia, the neighborhood name and `rdfs:seeAlso` can be used.

4.5. OUTPUT: WEB APPLICATION

The last step is visualizing and querying: it exists of creating the web application: making the spatial Linked Open Data available on a map in a Web viewer. And making the data query able by using a SPARQL interface connected to a SPARQL endpoint or hard coding an user-friendly option to ask questions to the application. As described in the linked data methodology in section 3.2., are there different options to publish the linked open data to the Web. Since this research is about spatial linked open data, showing the data on the Web in a map seems the best option to show the data in its context.

In the figure on the next page (figure 17) the system architecture of a Linked Data Web application can be seen. This is based on the conversion of RDF 'on the fly'. For publishing the Linked Data on the web, a Linked Data API is needed. This API makes a connection with the database to answer specific queries. The HTTP endpoint is a webpage that forms the interface. To make a Web application, a REST API is used. This REST API gives the possibility to give the Linked Data in various formats back to the user, depending on what the user wants. The Linked Data can be made visible in HTML on a website as HTTP links, or as RDF data in a browser or the most user friendly: a graphic visualization in a Web application (Brattinga et al., 2014).

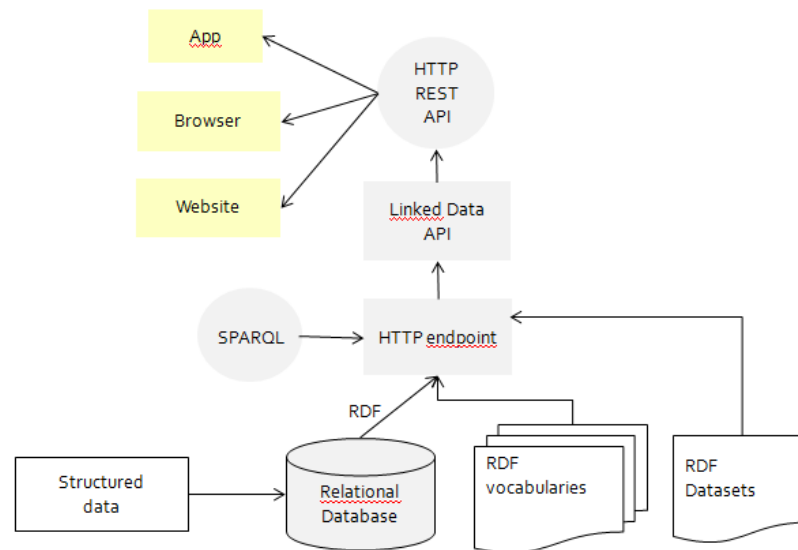


Fig. 17. System Architecture of a Linked Data Web Application-adapted from (Brattinga et al., 2014)

4.6. LINKED DATA MANAGEMENT

This last section focuses on data management. It explains how the linked open data could be maintained and what the costs are to upgrade open data to linked open data.

4.6.1. MAINTAINING LINKED OPEN DATA

An important part of publishing open data is data governance, quality control and maintenance of the data. To provide qualitative good datasets, the data needs to be updated regularly. Also, someone should be in charge of the data to prevent from errors (Folmer, Reuvers, Quak, & Brandt, 2014)

4.6.2. LIFECYCLE MODEL FOR PUBLISHING LINKED OPEN DATA

According to the W3C Linked Data Cookbook, should every project that is about opening up governmental data have a life cycle, to guide the process (Hyland & Villazón Terrazas, 2011). A lifecycle model is a model that describes all phases of a process. By assessing the development of a process, a plan of action can be made (Broek, Veenstra, & Folmer, 2013). Several life cycles models are explained in the Cookbook, but the lifecycle model as described next, is developed during a use case in the Netherlands, namely for the opening up of datasets at TNO, which could be a bit similar to the research done in this thesis and is therefore elaborated on.

In this model, five phases that are common to publish open data as Linked data can be identified (see figure 18).

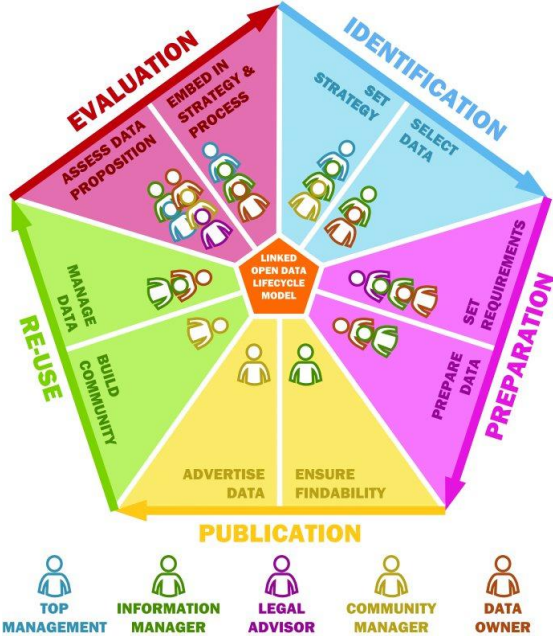


Fig.18. Lifecycle model of publishing linked open data (Broek et al., 2013)

The phases identification, preparation, publication, re-use and evaluation all exist of two steps and played by different roles. They are shortly described here. The first phase is identification, in where a vision or strategy is formed in how opening up of data contributes to the mission of the organization. Relevant stakeholders, such as businesses and the community can be involved to know which datasets are of interest to them. Secondly, there should be thought of which datasets are suitable to open up. The second phase consists of preparing the data. First the requirements are set. These include the technical, economic and legal requirements, and in the next step the data is converted to Linked Data. The third phase is publication. In the first step it is important to make sure the published data can be found by others, for example by opening up a data portal and then in the next step the data is advertised. The fourth phase focuses on re-use. To promote the re-use of the data, a network around the data can be built with partners that can actively use the data. This can also ensure feedback to improve the quality of the data. The managing of the data exists of checking if all the requirements are implemented. And the next step consists of managing the data. The last phase is the evaluation phase in where all stakeholders reflect on the linked open data implementation and see how the process can be better adapted to the strategy (if needed). If the strategy needs to be adapted, the process becomes an iterative cycle (Broek et al., 2013).

5. LINKED DATA TECHNOLOGY

This chapter describes the literature study performed for this thesis. The literature includes elements of Linked Data and steps from the methodology. Section 5.1. up to and including 5.4. explains more about RDF, the syntax of RDF, vocabularies and SPARQL. The sections thereafter focus on the conversion of spatial data and this chapter ends with linked data management.

5.1. RDF

In the introduction is shortly written that Linked Data is written in the W3C standard RDF. With data in RDF, information on the Web can be described. Information on the Web is also called a resource or an entity. The data model of RDF is a directed graph, which exists of a set of triples. A triple consists of a subject, predicate and object and represents a RDF statement. The elements of a triple are all written with a HTTP URI. Though the object does not have to be an URI, it can also be a literal value, such as a number, name (string) or a date. With a literal value, the data type has to be specified with data types from the XML schema. These could be string, integer, float, date, boolean and more (Cyganiak, Wood, & Lanthaler, 2014). In the example below a triple about the construction year of the building De Rotterdam is displayed. The construction year is a literal and is specified as an integer.

Subject	Predicate	Object
<code>http://www.DeRotterdam.nl</code>	<code>www.exampledatarotterdam/constructionYear</code>	<code>"2013"</code> <code>^^xsd:integer</code>

A collection of RDF statements or a set of triples is seen in the image below (see figure 19).

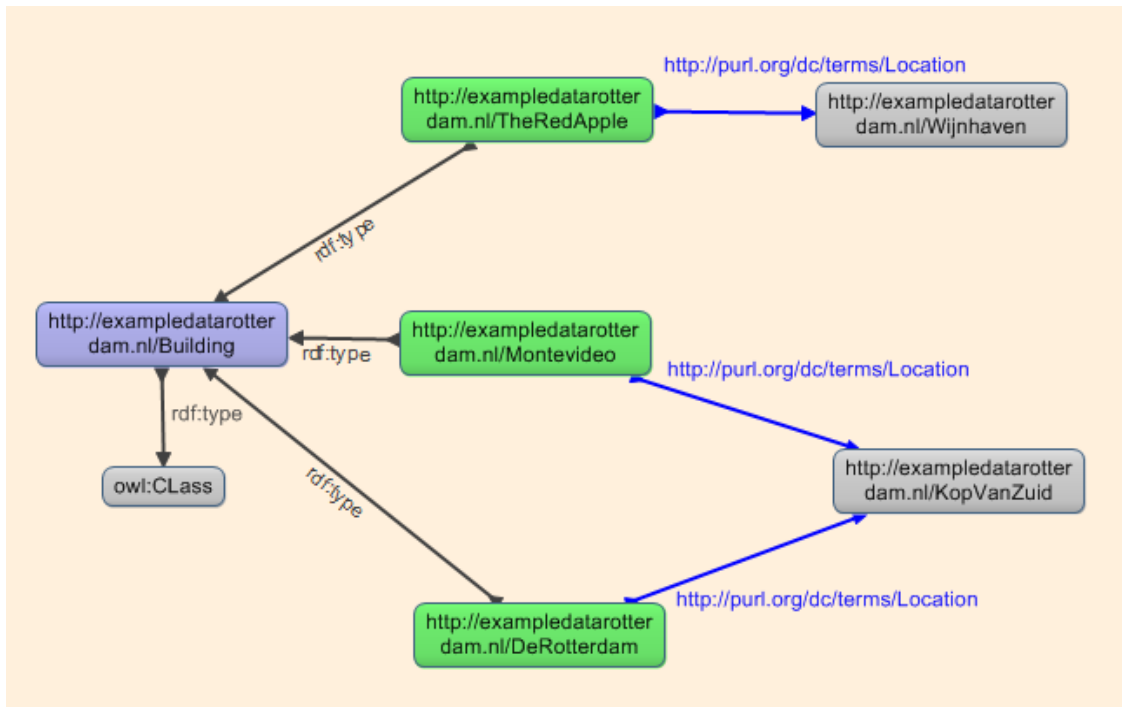


Fig.19. Example of a directed graph, visualized in Gruff

In the example above, a simple directed graph of three high-rise buildings and their location in Rotterdam is displayed. A dataset can exist of multiple graphs, for example the BAG data can exist of a premise graph, a mooring location graph and a pitch location graph (Cyganiak et al., 2014).

In this example the common part of the URI is <http://exampledatarotterdam.nl>, this is called the namespace of the dataset. This can be shortened to a namespace prefix such as *example*. The resource De Rotterdam (<http://exampledatarotterdam.nl/DeRotterdam>) can then be written as *example:DeRotterdam*, which makes a URI more readable.

Common prefixes are RDF and OWL (see figure 19) and XSD, for the XML schema (see example triple above).

5.2. RDF SYNTAX

RDF can be written in different formats. They differ from each other in human readability and use case. The most common formats are N-triples, Turtle and RDF/XML.

The N-triple syntax exists of the subject, predicate and object written with URIs/literals and ends with a space and then a dot (Beckett, 2014). In the example below is the same data as in figure 19 shown, in 7 triples. In the N-triples syntax, the data model of RDF is recognizable. Though this syntax is harder to read when

having a large file, it is much used when generating large RDF documents, since it is easy to parse.

```
<http://exampledatarotterdam.nl/Building> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/2002/07/owl#Class> .
<http://exampledatarotterdam.nl/TheRedApple> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://exampledatarotterdam.nl/Building> .
<http://exampledatarotterdam.nl/TheRedApple> <http://purl.org/dc/terms/Location>
<http://exampledatarotterdam.nl/Wijnhaven> .
<http://exampledatarotterdam.nl/DeRotterdam> <http://purl.org/dc/terms/Location>
<http://exampledatarotterdam.nl/KopVanZuid> .
<http://exampledatarotterdam.nl/Montevideo> <http://purl.org/dc/terms/Location>
<http://exampledatarotterdam.nl/KopVanZuid> .
<http://exampledatarotterdam.nl/Montevideo> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://exampledatarotterdam.nl/Building> .
<http://exampledatarotterdam.nl/DeRotterdam> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://exampledatarotterdam.nl/Building> .
```

Example of N-triples syntax

The Turtle syntax (the Terse RDF Triple Language) is a language that describes a RDF graph in text (Beckett, Berners-Lee, Prud'hommeaux, & Carothers, 2014). First the prefixes are listed, and then a subject with all its predicates and objects are listed. This way of writing RDF is more humanly readable, but is much less used to generate large RDF files because it can be harder to parse. In the image below, some example triples can be seen.

```
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix example: <http://exampledatarotterdam.nl/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

example:DeRotterdam a example:Building ;
  dcterms:Location example:KopVanZuid .

example:Montevideo a example:Building ;
  dcterms:Location example:KopVanZuid .

example:TheRedApple a example:Building ;
  dcterms:Location example:Wijnhaven .

example:Building a <http://www.w3.org/2002/07/owl#Class> .
```

Example of Turtle syntax

An example of the RDF/XML format can be seen in the figure below.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
>
  <rdf:Description rdf:about="http://exampledatarotterdam.nl/DeRotterdam">
    <rdf:type rdf:resource="http://exampledatarotterdam.nl/Building"/>
    <dcterms:Location rdf:resource="http://exampledatarotterdam.nl/KopVanZuid"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://exampledatarotterdam.nl/Building">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://exampledatarotterdam.nl/Montevideo">
    <dcterms:Location rdf:resource="http://exampledatarotterdam.nl/KopVanZuid"/>
    <rdf:type rdf:resource="http://exampledatarotterdam.nl/Building"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://exampledatarotterdam.nl/TheRedApple">
    <rdf:type rdf:resource="http://exampledatarotterdam.nl/Building"/>
    <dcterms:Location rdf:resource="http://exampledatarotterdam.nl/Wijnhaven"/>
  </rdf:Description>
</rdf:RDF>
```

Example of RDF/XML syntax

Basically it consists of node elements (subject/object) and property elements (predicates). Each subject is identified with the `rdf:about` tag, and the corresponding object is identified with the `rdf:resource` tag. Each new subject is filed under a `rdf:Description` tag. There are more syntaxes of RDF, such as JSON-LD and GeoJSON (Gandon & Schreiber, 2014), which will be described in section 8.3. All the different syntaxes have pros and cons and are better to use for a certain purpose, but in the end they can all be used to write Linked Data.

5.3. ONTOLOGIES AND VOCABULARIES

As described in section 4.2 do ontology form part of the data model and do they exists of vocabularies which define the predicates in the RDF data. By re-using an ontology or vocabulary, data becomes more understandable and that promotes re-use of the data. These vocabularies are also written in RDF, which means that the semantics of the data are in the same format as the data itself.

Two much used vocabularies (meta models) which add semantics to the data (LinkedDataTools.com, n.d.-b) are RDFS and OWL. Other vocabularies are SKOS,

Dublin Core and VoID, which are described below. All these vocabularies are data independent, which means they can be used for most types of data.

RDFS

RDFS, or RDF Schema, is an extension to RDF and exists of several classes and properties. RDFS is used to model groups of corresponding resources and to define the relations between them. With RDFS a group of related resources are named classes. An individual of this class is an instance. To state that a resource is a member of a class, the predicate `rdf:type` can be used (Brickley & Guha, 2014), (see fig. 20 where De Rotterdam is an instance of the class Building).

In a graph, there is basically no hierarchy. Though with the use of RDFS, a hierarchical data structure can be made by for example using the predicate `rdfs:subClassOf` to state that a class is a subclass of another class. This hierarchical structure is useful for understanding the data and to query the data.

OWL

OWL, or shortly Web Ontology Language, is an extension to RDFS (LinkedDataTools.com, n.d.-a) and provides a much richer data- modeling vocabulary for RDF. With OWL, the RDF data becomes even more understandable than with RDFS alone. OWL adds to RDFS the `owl:Class` which can be used to define a class as an OWL class and specifies relations between classes (such as cardinality, equality and so on) (McGuinness & Harmelen van, 2004).

RDFS and OWL are much used in Linked Data to add the semantics of the data model to the RDF data.

SKOS

SKOS is mainly used for classifications. It describes relations between elements of knowledge organization systems (W3C Semantic Web, 2012).

DUBLIN CORE

For the metadata the Dublin Core vocabulary is much used (see the dark blue circle in figure 5). This standard is defined by the Dublin Core Metadata Initiative which is an organization for promoting the re-using of metadata terms (Metadata Innovation Dublin Core Metadata Initiative, n.d.).

VoID

VoID stands for Vocabulary of Interlinked Datasets and is used to document the metadata of a RDF dataset. It divides metadata in four types, namely general metadata where the Dublin Core vocabulary is used for, access metadata, which describes how the RDF data can be accessed, structural metadata, which describes the data model and structure of the datasets and the last type is the description of

links between datasets which describes how multiple datasets are related to each other (Alexander, Cyganiak, Hausenblas, & Zhao, 2011).

Knowledge of vocabularies that exists is needed, to be able to re-use the vocabularies. On the website of the Linked Open Vocabularies (<http://lov.okfn.org/dataset/lov/>) are a lot of vocabularies that exists listed (see figure 20). The size of the circles means the links that are made from that particular vocabulary to other vocabularies and the colors define their knowledge domains. For example, Foaf (friend of a friend) is a much used vocabulary to describe a page about people on the Web. The website of the Linked Open Vocabularies provides users to search for vocabularies in a knowledge domain.

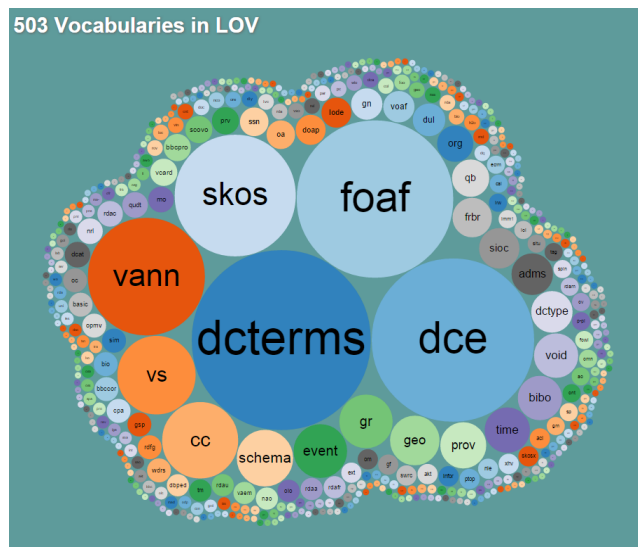


Fig. 20. Vocabularies in the Linked Open Vocabulary cloud as of 24-04-2015 (Linked Open Vocabulary, n.d.)

5.4. SPARQL

The standard SPARQL is a standard for querying RDF data. Basically a SPARQL query asks for a part of a graph (sub graph) of the RDF data that matches a certain pattern. This pattern is called a basic graph pattern and exists of the three elements of a triple (subject, predicate and object), where at least one of the three elements is known (Prud'hommeaux & Seaborne, 2008). In the SPARQL query example below is first asked for all the subjects that are of type housing and then for all the subjects that match that pattern and also are located in Kop van Zuid.


```
PREFIX example: <http://exampledatarotterdam.nl>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

```
SELECT *
WHERE {
?subject rdf:type example:Building .
?subject dcterms:Location example:KopVanZuid .}
```

Example of a SPARQL query

In the example above can be seen that a query is constructed of listing the used prefixes first and then do a select and where and optional a filter. The elements that are not known, are recognizable at the question marks in the query entry.

SPARQL is similar to SQL in that it also uses the same pattern of select and where and other constructors. In the figure below is for the same query as above (select the buildings in Kop van Zuid that have are of type Building), the SQL query given. In this query are from the fictional table arearotterdam all the elements selected that are of the type “building” and that also have as location “Kop van Zuid”.

```
SELECT *
FROM arearotterdam
WHERE type = "Building"
And location = "Kop van Zuid";
```

Example of a SQL query

In the SPARQL query is first asked for all the buildings that are of type “Building” and from those results only the buildings that are located in “Kop van Zuid” are selected. The main difference between SPARQL and SQL is that SQL does not work with data on the Web.

With SPARQL different types of queries can be asked. There are three types, namely tuple, graph and boolean queries and differ from each other in how they are constructed and the result that gets back. Tuple queries are queries that produce a specific element from the stored RDF data. These queries are written with Select, as seen in the example above.

Graph queries are queries that return a graph. This can be used to return a subset of the data, if the data exists of multiple graphs. These are written by replacing the Select for Construct or Describe.

Boolean queries are queries that return a binary value, such as true or false. These queries are useful for analyzing if a dataset contains specific information. These queries are written with Ask instead of the Select (Sesame, 2014)

With a SPARQL endpoint (website that provides a SPARQL interface and query engine) RDF data can be queried. Depending on the support of the query engine, more or less complicated queries can be asked. The tuple queries are the most common and are mostly provided by any SPARQL endpoint that is available. Currently a new SPARQL recommendation is available: SPARQL 1.1 which adds more functionality, such as inserting and deleting triples (Prud'hommeaux & Seaborne, 2008).

5.5. MODELING SPATIAL DATA IN RDF

Spatial data differs from other types of data, because spatial data is bound to a location or a position. That location can be mentioned by a place name and the position by coordinates. Spatial data will often have a geometry, such as a point, line or a polygon. Spatial data contains topology, to define how the spatial objects are related to each other. Metadata is especially important for spatial data, to prevent from modeling the data incorrectly and to be able to do a good analysis. The metadata for spatial data can contain among others the accuracy, the timestamp, the location the data applies to and the CRS when modeling locations with coordinates.

5.5.1. VOCABULARIES FOR SPATIAL DATA

The need to convert spatial data as Linked Data, has lead to different parties developing different vocabularies, such as the ISA Programme Location Core Vocabulary, NeoGeo Vocabulary, schema.org, the OGC standard GeoSPARQL, the W3C Geospatial Vocabulary and stRDF.

The ISA Programme Location Core Vocabulary is developed to aid in publishing spatial data in the context of the EU INSPIRE Directive. The vocabulary consists of three classes, namely the Location, Address and Geometry and specific properties relating to the classes. The status of the Geometry class stated to be unstable (EU ISA Programme Core Vocabularies Working Group (Location Task Force), 2013).

The NeoGeo Vocabulary is developed through many vocabulary community meetings. In NeoGeo a distinction is made between spatial features which can have geometries and is the Region Connection Calculus (RCC) used to define spatial relations. In RCC, binary relations between two spatial objects are used. Functions for spatial reasoning are not defined (Norton et al., 2012).

Schema.org is a vocabulary defined for webmasters to markup their HTML page in such a way that search engines can access the structured data in the database

where the web page is generated from. This vocabulary is thus defined for developers of web pages (schema.org, n.d.-b). In the vocabulary, an entity can have a place which can have a geo-aspect namely the GeoCoordinates to define the position in WGS84 or a GeoShape (box, circle, line and polygon) which defines the geometry (schema.org, n.d.-a).

The W3C Geospatial Vocabulary is an update to the Basic Geo Vocabulary. The Basic Geo Vocabulary was one of the first vocabularies for spatial data and it defines points with its longitude, latitude and altitude in the WGS84 reference system (Brickley, 2006). It is still a much used vocabulary. The Geospatial Vocabulary defines geometry types according to the GeoRSS Feature model, where a feature can be represented as a point, line, box and polygon geometry. Spatial relations were not yet defined in the document (Lieberman, Singh, & Goad, 2007). The GeoRSS model is a model that is used to add a location to RSS feeds. It is based on XML and supports the WGS84 reference system (Koubarakis, Karpathiotakis, Kyzirakos, Nikolaou, & Sioutis, 2012).

GeoSPARQL is an OGC standard and ontology for representing spatial features and geometries and provides functions for spatial reasoning, all based on OGC standards (Battle & Kolas, 2011a).

Another way to represent spatial data as Linked Data is through stRDF and stSPARQL. It is developed in the context of the European project TELEIOS which focuses on Linked Earth Observation data. stRDF is an extension to RDF wherein a triple is extended with a fourth element: time, which is not yet exploited. The query language for stRDF is stSPARQL. To represent geometry, the OGC standards WKT and GML are used and different CRS can be added as a literal. stRDF /stSPARQL is similar to GeoSPARQL (Koubarakis, Karpathiotakis, Kyzirakos, Nikolaou, Vassos, et al., 2012).

The vocabularies listed here differ from each other in the degree to which they have been developed. Some of the vocabularies only deal with point data and some even provide predicates for defining spatial relations. The vocabularies do not talk about 3D data and only the stRDF and stSPARQL languages seem to talk about time. Time is very important for spatial data, since objects in the real world mostly exist during a time period.

For this research the GeoSPARQL ontology is used, since this is the OGC standard for spatial data in RDF. Also, this standard defines functions for spatial querying and defines a vocabulary to define geometry in different CRS. In the next section, a more detailed description of GeoSPARQL is given.

5.5.2. THE GEOSPARQL SPECIFICATION

The GeoSPARQL specification consist of a few components, including the ontology for representing features, geometries and their relations, a set of spatial functions and a set of transformation rules for easier querying.

The Simple Feature Specifications from the OGC is used as input for the GeoSPARQL ontology. In this ontology, a distinction is made between features and geometries, in where features which are objects in the real world, can have a geometry, which is a representation of the real world location. Both are subclasses of a spatial object. These three classes can be used to connect to another ontology (see figure 21) (Battle & Kolas, 2011a).

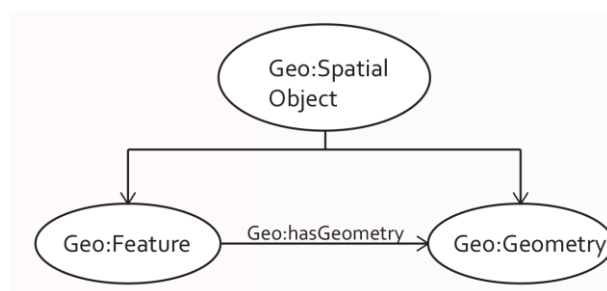


Fig.21. GeoSPARQL ontology

Features can have multiple geometries, dependent on how detailed the representation is modelled. These geometries are represented in either GML or as a WKT and contain a literal which specifies the CRS (Battle & Kolas, 2011a). Preferred for this research is WKT, since this way of representing geometries is more easily imported in the web application later.

5.5.3. SPATIAL OPERATIONS

The other main component of the GeoSPARQL specification is spatial operations. Spatial reasoning can be used to do filter functions on the dataset to return a selection of the dataset, based on the spatial component of the objects. For these topological filter functions, the relations according to three spatial models can be applied to the spatial Linked Data, which are implemented in GeoSPARQL.

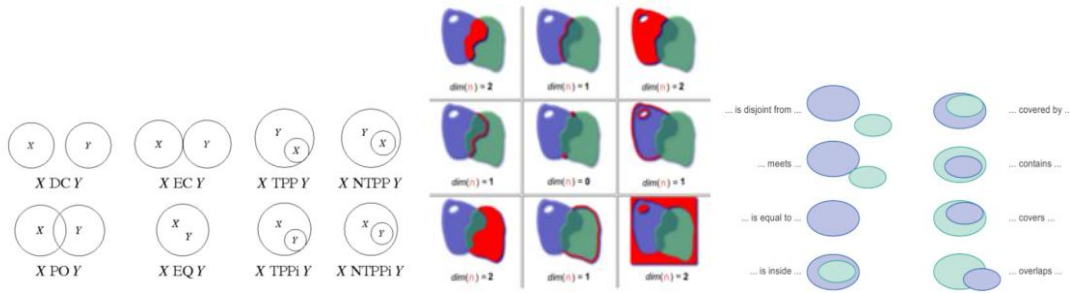


Fig.22. (left) RCC8 (Wikipedia, 2013)

Fig. 23. (middle) Egenhofer's 9-intersection model (Wikipedia, 2015)

Fig.24. (right). OGC Simple Features Topological relationships (Arroyo Ohori K. & Ledoux, 2013)

The Region Connection Calculus (RCC) is a model that defines functions and relations for expressing and reasoning over spatial regions. RCC8 is a subset of the RCC, which defines 8 binary relations (see figure 22). The same relations as described in the RCC8 are also described with different names by Egenhofer in his 9- Intersection model. This model adds the functionality to describe relations between different dimensions of geometries. A distinction is made between interior, exterior and boundary of the two objects (see figure 23). The OGC Simple Features Specification uses the 9- intersection model to describe spatial relations in GIS (see figure 24). (Battle & Kolas, 2011a). Most of the functions are equivalent to each other, as can be seen in the table below (see table 4).

RCC8	Egenhofer	Simple Features
rcc8eq	ehEquals	sfEquals
rcc8dc	ehDisjoint	sfDisjoint
-	-	sfIntersects
rcc8ec	ehMeet	sfTouches
rcc8ntpp + rcc8tpp	ehInside + ehCoveredBy	sfWithin
rcc8ntppi + rcc8tppi	ehContains + ehCovers	sfContains
rcc8po	ehOverlap	sfOverlaps
-	-	sfCrosses

Table 4. Topological relations in GeoSPARQL
Adaptation from (Battle & Kolas, 2011a)

The geometry class defines also functions for standard vector operations in a GIS, which are defined as non-topological functions, according to the GeoSPARQL standard (see table 5).

Non-topological functions	
intersection	envelope
boundary	getSRID
buffer	relate
convexHull	symDifference
difference	union
distance	

Table 5. Non-Topological relations in GeoSPARQL from the namespace for the spatial functions for the GeoSPARQL ontology: <http://www.opengis.net/def/function/geosparql/>

These functions return a geometry and can be used to do a spatial analysis on data. For example the buffer functions can be used to create a buffer around a point object of a certain distance and then filter all the objects in this buffer to search in the neighborhood. The distance function can be used to calculate the distance between two objects. Noticeable is that there are no functions for transformations of the CRS.

5.6. TRIPLE STORE FUNCTIONALITY

To store the spatial linked data in a triple store, the triple store should meet some requirements. One of the requirements as defined by literature, is that the storage layer should support the storing of spatial objects. Also, (spatial) indexing of objects is required (Battle & Kolas, 2011b). Since the development of the GeoSPARQL language is fairly new, not all triple stores support the standard yet. Some of the available triple stores have their own vocabulary for modeling spatial linked data and performing spatial queries. In the table below a comparison of a selection of triple stores is given (see table 6). The selection of triple stores is based on the stores as came across in the literature study.

	Support for spatial objects	Implementation of GeoSPARQL	Spatial indexing
AllegroGraph	Yes	No	Yes
Openlink Virtuoso Universal Server	Yes	Work in progress	Yes
Oracle Spatial & Graph	Yes	Yes	Yes
Parliament	Yes	Yes	Yes
Strabon	Yes	Yes	Yes

Table 6. Comparison of a selection of triple stores

This table is shortly explained in the following sections.

ALLEGROGRAPH

AllegroGraph is developed by Franz Inc. and is a graph database and provides a SPARQL endpoint (Franz Inc., 2015a). AllegroGraph has its own vocabulary for spatial data types and since version 5.0.1. it is named N-dimensional geospatial. Multi-dimensional spatial objects can be stored in the store (Franz Inc., 2015b). Spatial indexing is done by dividing the space in strips of different widths (Battle & Kolas, 2011b). To query the spatial objects, AllegroGraph uses a different way of reasoning in SPARQL, named SPARQL Magic Properties (Franz Inc., 2015d). Also from Franz Inc. is the RDF browser Gruff that can be used to show visual graphs of the RDF data in AllegroGraph (Franz Inc., 2015c).

OPENLINK VIRTUOSO

Virtuoso (from Openlink software inc.) is a software package that provides amongst others a RDBMS, a triple store, a Linked Data Server and a Web Application Server in one program (Openlink Software, 2015). Virtuoso could be used to go from data in a RDBMS to RDF data and publish it to the Web in one program. Also Virtuoso (from version 6.01.3126) provides support for spatial objects in the database and the triple store. To index the spatial objects, an R-tree is used. The Basic Geo Vocabulary is used as CRS, though with SQL-functions ST_Transform and ST_SetSRID, it can be changed to another SRS. (Openlink Software, n.d.-a). The functions ST_intersects, ST_contains and ST_within can be used to query the geometry objects and refer to the SQL functions and predicates (Openlink Software, n.d.-b). From version 7.1. Virtuoso supports the geometry types and WKT representations of OGC data types and currently the functions are mapped to GeoSPARQL predicates (Williams, 2014).

ORACLE SPATIAL & GRAPH

Oracle Database 12c is a RDBMS which provides also a triple store. The triple store RDF Semantic Graph is part of Oracle Spatial and Graph (Oracle, n.d.) and is able to

convert the data in the database on the fly to RDF. Since version 12c, the full implementation of GeoSPARQL is supported.

PARLIAMENT

Parliament is a triple store developed by BBN Technologies. It uses an additional query processor (Jena or Sesame) to provide SPARQL querying over RDF data (Emmons, 2014). Parliament supports the GeoSPARQL standard and uses a temporal and spatial index (Parliament, n.d.). The spatial index is an R-tree which splits a spatial SPARQL query into separate parts, to provide an optimized query plan between the parts of the query that ask for spatial information and the parts that ask for non-spatial information (Battle & Kolas, 2011b).

STRABON

Strabon is originally developed for the European project SemSorGrid4Env, but its functionality is extended for the European project Teleios. Strabon is a RDF store where spatial objects that change over time can be stored (see figure 25) (Strabon, n.d.). As DBMS either MonetDB and PostgreSQL (temporal) + PostGIS can be used. The RDF store Sesame is used and extended for the development of Strabon, to support the spatial and temporal functions (Teleios project, n.d.). Strabon is designed to work with stRDF and stSPARQL, but can also answer GeoSPARQL queries, though not all are implemented (Strabon, n.d.).

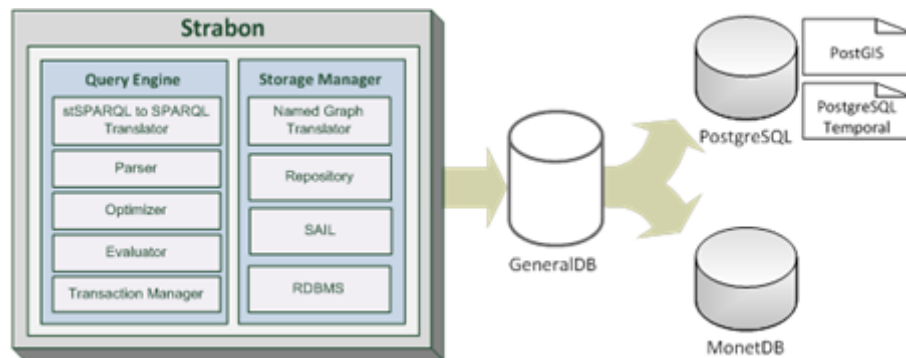


Fig.25. System Architecture of Strabon (Teleios project, n.d.)

The research done for this thesis requires the implementation of GeoSPARQL. Since Oracle Spatial & Graph was not available in Rotterdam for this research, Strabon seemed to be the best choice, since it could be easily installed and has a nice interface.

5.7. ON-THE-FLY CONVERSION

This section describes a bit of theory about the on-the-fly to RDF approach which will be tested in this research.

5.7.1. R2RML

Since data often resides in a relational database there are a few languages that define mappings between data in a relational database to RDF. One of them is the language R2RML, in which a (conceptual) mapping is made. The language can be differently implemented in software, namely by providing a browser to access and view the RDF (virtual mapping), or a SPARQL endpoint queries the database and provides the result in RDF (virtual) (see example on page 60). or the software can generate a RDF file with the triples resulting from the data in the database.

There is also another language for creating RDF from data in a relational database, namely DM (direct mapping). This language maps the table as is and provides thus not much customization of the data (Das, Sundara, & Cyganiak, 2012).

Customization is in any way needed to map the data in the table to existing vocabularies and ontologies.

The following section describes the used software and tools for creating and visualizing the spatial Linked Open Data, which will be described in the next chapter.

5.8. SELECTED SOFTWARE AND TOOLS

The table below (see table 7) shows the tools for each of the components of the Linked Data Web application that are used. The selection of tools/software is based on experience with the software or the availability of the software.

Components	Purpose	Tools
Data Analysis	UML-editor	Microsoft Visual Studio
	GIS	ArcGIS and QGIS
	Geo-database	PostgreSQL/PostGIS
Defining an ontology	Graph visualizing program	Gruff/AllegroGraph 3.3
	Creating data model	Python with libraries psycopg2 and RDFlib
Creating RDF data	RDF storage approach	
	Geo- database	PostgreSQL/PostGIS
	Creating RDF file	Python with libraries psycopg2 and RDFlib
	Triple Store	Strabon on Apache Tomcat
	Publish data to the Web	Pubby
	On-the-fly to RDF approach	
	Geo- database	PostgreSQL/PostGIS
	On-the-fly conversion	D2RQ Platform
Querying and Visualizing RDF Data	Graph visualizing program	Gruff /AllegroGraph 3.3
	Triple store and	Strabon on Apache
	SPARQL endpoint	Tomcat
Visualizing spatial Linked Open Data	Map control	Leaflet
	Base map	WMTS OpenStreetMap from Mapbox

Table 7. Tools per component used for the research

A few of the tools are explained in this section.

PSYCOPG2

<https://pypi.python.org/pypi/psycopg2>

Psycopg2 is a Python library that provides access to the PostgreSQL database through Python.

RDFLIB

<https://pypi.python.org/pypi/rdfliB>

RDFlib is a python library that provides a lot of functions for working with RDF data, such as storage of triples, implementation of SPARQL and converting RDF in different syntaxes.

PUBBY

<http://wifo5-03.informatik.uni-mannheim.de/pubby/>

Pubby is used to publish the data on the web. A SPARQL endpoint is needed and a mapping file needs to be made to show the RDF data on a Web page.

D2RQ PLATFORM

<http://d2rq.org/>

The D2RQ Platform is a program that converts data from a relational database to RDF data on the Web. It does not use the R2RML language as explained in section 5.8.1., but uses its own mapping language D2RQ. The program provides a generate mapping option which automatically provides a mapping file and publishes the data on the Web. It also provides a SPARQL endpoint, named SNORQL. Another possibility is to generate a RDF file with the D2RQ program. The mapping file is written in turtle syntax (Cyganiak, Bizer, Garbers, Maresch, & Becker, 2012).

LEAFLET

<http://leafletjs.com/>

To create the Web API, Leaflet is used. Leaflet is a JavaScript library to create interactive maps.

6. CONVERSION TO SPATIAL LINKED OPEN DATA

This chapter describes the conversion of open spatial data to linked open spatial data. First the conversion of two open datasets from the municipality are described, then conversion of the CBS data and finally the conversion of the BAG of Rotterdam Zuid. In total, data from four datasets are transformed. The datasets have different data formats, namely shapefiles, a CSV and tables in a relational database. This chapter provides insight in the sub question of how to publish spatial open data as spatial linked open data on the Web. The datasets from the Municipality are chosen to be relevant for the NPRZ project, in particular for the education theme.

6.1. CONVERSION OF THE MUNICIPALITY OPEN DATA

6.1.1. DATA FROM THE ROTTERDAM OPEN DATA STORE

DATA ACQUISITION

The dataset selected from the Rotterdam Open Data Store is a dataset containing kindergartens in Rotterdam. The file contains geometry since the kindergartens are indicated with a point. Attributes are the address and name of the kindergarten. The dataset is accessible by a WMS or WFS. To acquire the data, the dataset was loaded as a WFS into QGIS and saved as a Shapefile.

DATA ANALYSIS AND PREPARATION

In QGIS only the kindergartens in Rotterdam Zuid are selected. According to the data, there are 34 kindergartens in Rotterdam Zuid (see figure 26). From every kindergarten: the name, zip code and house number are documented (see figure 27). The house number column is apparently duplicated in the dataset.

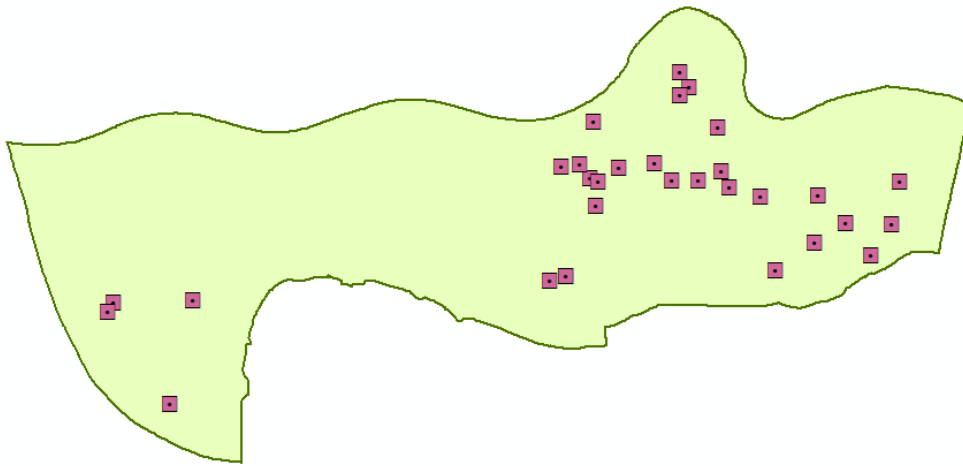


Fig.26. the Kindergarten dataset from the Rotterdam Open Data Store. The purple squares represent the kindergartens in Rotterdam Zuid

Feature	Value
0	kinderdagverblijven
ID	1018053
(Actions)	
(Derived)	
HUISNR	25
HUISNUMMER	25
ID	1018053
OMSCHRIJV	Kinderdagverblijf 't Bootje
POSTCODE	3072AS
TYPE	5474
TYPE_OMSCH	Kinderdagverblijven

Fig.27. Attribute specification of a kindergarten in QGIS

The Shapefile is loaded into PostgreSQL/PostGIS with the shp2pgsql and the psql tool. This results in a table with the 34 kindergartens, their attributes and the point geometry in the Dutch spatial reference system EPSG:28992.

DEFINING AN ONTOLOGY

The small kindergarten dataset is fairly simple to understand. Also the attributes in the dataset are self-explanatory. Based on the data that is in the kindergarten dataset and the literature study done on vocabularies, the following vocabularies are chosen (see table 8).

Type of data	Description	Vocabulary
General	General relations between resources	RDF Schema (RDFS) Web Ontology Language (OWL)
Spatial features + geometry	Point objects	GeoSPARQL
Metadata	Descriptive data	Dublin Core Vocabulary of Interlinked Datasets (VOID)

Table 8. Chosen vocabularies for the Kindergarten dataset

Basically all the 34 kindergartens are a member of the class Kindergarten, modelled in RDFS and OWL. There does exist a vocabulary to model addresses specifically, namely the Location Core Vocabulary (as described in section 5.6.1). However, the vocabulary is in English, while all the data is in Dutch. A triple using the Location Core Vocabulary would look as follows (using prefixes):

Kinderdagverblijf:tBootje – locn:postCode – "3072AS"^^xsd:string

In the context of re-use, it would be good to model the data also in English. But since the end users of the Web app will be people living in Rotterdam and probably will speak Dutch, there is chosen to not use this vocabulary.

IMPLEMENTING THE URI-STRATEGY

The URI-strategy (see figure 12), as described in section 3.3.1 is implemented, since this is in accordance with the current developments in the Netherlands.

The domain name is <http://ldatarotterdam.nl>. Since this is the actual data, the type is 'id' and the concept is a kindergarten (kinderdagverblijf in Dutch). The reference is the name of the kindergarten. Since a URI cannot contain spaces, the words are concatenated and punctuation marks are removed. The full URI of one resource is then:

<http://ldatarotterdam.nl/id/Kinderdagverblijf/tBootje>. The metadata is implemented in the 'doc' type of the URI strategy: <http://ldatarotterdam.nl/doc/Kindergarten>. The def type is not used, since the dataset is self explanatory.

MAPPING- SPATIAL DATA

The point geometry of the locations of the kindergartens is mapped according to the GeoSPARQL ontology. First the class Kindergarten is said to be a member of the

Spatial Feature class. Then a resource (tBootje) is made a member of the class Kindergarten. This resource has a geometry of the Simple Features Type point. The geometry of the resource has a WKT representation. Important is to model the CRS as a string and to specify the data type, namely a WKT literal of the geometry. The figure below shows the defined triples (see figure 28).

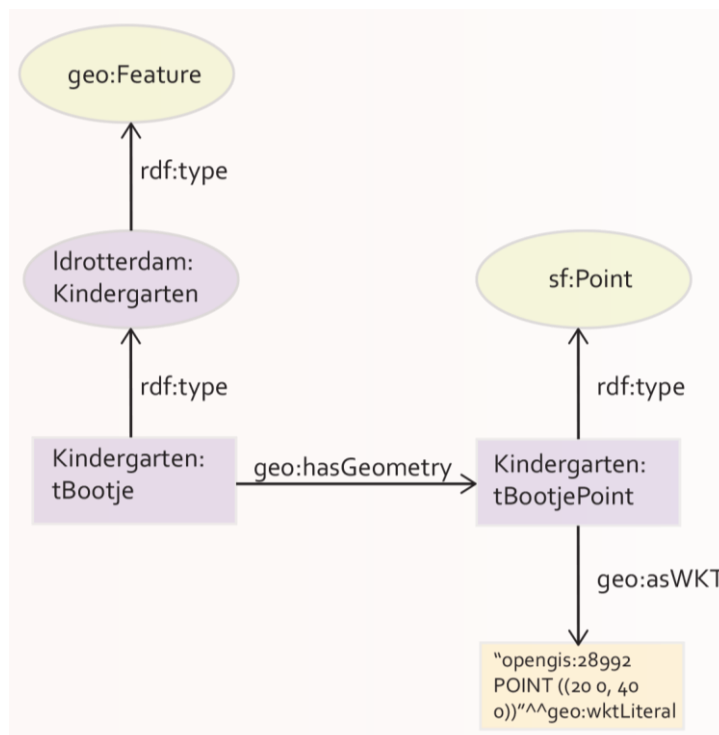


Fig.28. Example of a kindergarten modelled according to the GeoSPARQL vocabulary

MAPPING- METADATA

For the metadata the VOID vocabulary is used. The VOID vocabulary references the Dublin Core vocabulary and the Friend of a Friend vocabulary (FOAF). For every dataset, the following predicates are specified:

- rdf:tpe void:Dataset
- foaf:homepage Homepage for the dataset
- dcterms:title Title of the dataset
- dcterms:description Description of the dataset
- dcterms:license The license with the use restrictions
- dcterms:creator The organization who created the dataset
- dcterms:publisher The publisher who made the data available
- dcterms:issued Date when the data was published
- void:inDataset To link a resource to the void:Dataset
- dcterms:created The date when the dataset was created
- (dcterms:requires If necessary, a file that is needed by the dataset)

These predicates ensure the title, a short summary, the source of the dataset, the publisher and a few statements about the data. The `dcterms:modified` can be used to let users know when changes were made to the dataset. The `dcterms:license` specification is very important to know the user restrictions to use the data.

Though the `VOID` vocabulary can be used for every dataset, spatial datasets are preferred to have some additional metadata, such as the accuracy, the dimension or a bounding box. The Dublin Core vocabulary has a few predicates that can be used for spatial data.

The following predicates are therefore also used for every dataset. A further research is not done on adding spatial metadata to linked data.

- `dcterms:spatial` Spatial coverage
- `dcterms:temporal` Temporal coverage

CONVERSION: RDF STORAGE APPROACH

The RDF storage approach is implemented by writing a mapping file in Python. For this, the library `RDFlib` is used. In the Python script a connection is made with the database (through the library `Psycopg2`) and then the table is converted to RDF. For the predicate names, the column names are used. The output of this script is a RDF file with N-triples. This file is uploaded to the triple store `Strabon`.

PUBLISHING THE DATA

Publishing the data is done with `Pubby`. `Pubby` works as a web application and is installed in `Apache Tomcat`. `Pubby` publishes the data on the Web by performing a describe query on the dataset through a SPARQL endpoint, or by reading a static RDF file. It proved to be difficult to get `Pubby` to work in combination with `Strabon`, which possible had to do because they are installed in the same `Apache Tomcat`. A static RDF file containing the kindergarten RDF data was therefore uploaded to `Pubby`. In figure 49, can be seen how the Web page `Pubby` creates of the linked data looks.

CONVERSION: ON-THE-FLY TO RDF APPROACH

The on-the-fly approach is performed with the `D2RQ` platform. To use the `D2RQ` platform, a primary key is generated on the table. The automatic generate mapping tool creates a basic mapping, where all the kindergartens are specified with a number. Therefore the mapping file is manually changed to change the URI and the name of the resources. Part of the mapping file can be seen below. First there is a section that defines some basic properties for the class `Kindergarten`. The second section defines what the name is of each resource (the name of the kindergarten).

The third section describes how the column zip code should be mapped. The green font shows the adjustments I made to the data, which consist of changing the URI-pattern, the vocabulary for the data, the labels and the data type.

```
# Table kvnew
map:kvnew a d2rq:ClassMap;
  d2rq:dataStorage map:database;
  d2rq:uriPattern "@@kvnew.omschrijvi|urlify@@";
  d2rq:class vocab:Kinderdagverblijf;
  d2rq:classDefinitionLabel "Kinderdagverblijf in Rotterdam Zuid";
.
map:kvnew__label a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:kvnew;
  d2rq:property rdfs:label;
  d2rq:pattern "@@kvnew.omschrijvi@@";
.
map:kvnew_postcode a d2rq:PropertyBridge;
  d2rq:belongsToClassMap map:kvnew;
  d2rq:property vocab:Kinderdagverblijf_postcode;
  d2rq:propertyDefinitionLabel "Postcode";
  d2rq:column "kvnew.postcode";
  d2rq:datatype xsd:string;
```

The mapping file for the kindergarten dataset written in the D2RQ language

The URI however, could not be changed to the one defined in the URI strategy, since the URI should start with <http://localhost:2020/resource>, since the data is locally stored and not made publicly available on a Web server. The printscreen below show all the resources that belong to the class Kindergarten (see figure 29).



Fig. 29. Printscreen of the resources belonging to class "Kinderdagverblijf", made with the D2RQ platform

And the printscreen below shows one resource (see figure 30). The D2RQ program automatically provides metadata according to the VoID vocabulary.



Fig. 30. Printscreen of one resource, made with the D2RQ platform

The D2RQ platform also provides a nice SPARQL endpoint (named Snorql). But literature ensures that the program does not provide GeoSPARQL functionality (Veer van 't, 2014b).

Though a mapping where the data is modeled according to the GeoSPARQL ontology can be made, there is no functionality to actually perform spatial queries on the data. Therefore, this method it is not suitable yet to use on spatial data that contains geometry.

6.1.2. DATA FROM ROTTERDAM IN CIJFERS

DATA ACQUISITION, ANALYSIS & PREPARATION

Data from the Research and Business Intelligence of the department of the municipality (OBI) is accessible through a Web application. The dataset that was chosen is a dataset about the number of children per household. In the web application, a selection of the areas in Rotterdam Zuid can already be made. The data is then downloadable in different formats such as a PDF, a Microsoft Office Document, a Microsoft excel spreadsheet and a CSV. The CSV format is chosen since it can be easily read by a script. Part of the CSV-file is shown below.

```

"", "0 kinderen", "1 kind", "2 kinderen", "3 kinderen", "4 kinderen", "5 of meer kinderen"
"Feijenoord", "23462", "5699", "3847", "1574", "488", "162"
"IJsselmonde", "19631", "4688", "3208", "1102", "289", "113"
"Charlois", "23969", "4999", "3260", "1135", "323", "126"
"Pernis", "1426", "391", "304", "72", "15", "1"
"Hoogvliet", "10642", "2719", "2105", "560", "130", "29"
"Haven- en industriegebieden", "97", "11", "4", "2", "-", "-"

```

Part of the CSV-file containing the number of households in areas of Rotterdam Zuid in 2014

DEFINING AN ONTOLOGY

Just like the kindergarden dataset, the data is immediately clear and therefore the mapping can be fairly simple. Based on the data that is in the dataset and the literature study done on vocabularies, the following vocabularies are chosen (see table 6).

Type of data	Description	Vocabulary
General	General relations	RDFS
	between resources, area names	OWL
Metadata	Descriptive data	Dublin Core
		VOID

Table 9. Chosen vocabularies

CONVERSION: RDF STORAGE APPROACH

The CSV-file can be converted to RDF with a Python script and the RDFlib library. Also in this dataset were strings that contained spaces which were thus removed. The URI for the dataset is [http://ldatarotterdam.nl/id/Gebied/\(Gebiedsnaam\)](http://ldatarotterdam.nl/id/Gebied/(Gebiedsnaam)) for the areas and [http://ldatarotterdam.nl/id/Huishoudens/\(number of children\)](http://ldatarotterdam.nl/id/Huishoudens/(number of children)) to define the number of children. As output, a RDF-file in N-triples syntax is produced and uploaded to Strabon. The same metadata as specified for the kindergarten dataset is also specified for this dataset. The image below shows some of the RDF data in graph view (see figure 31).

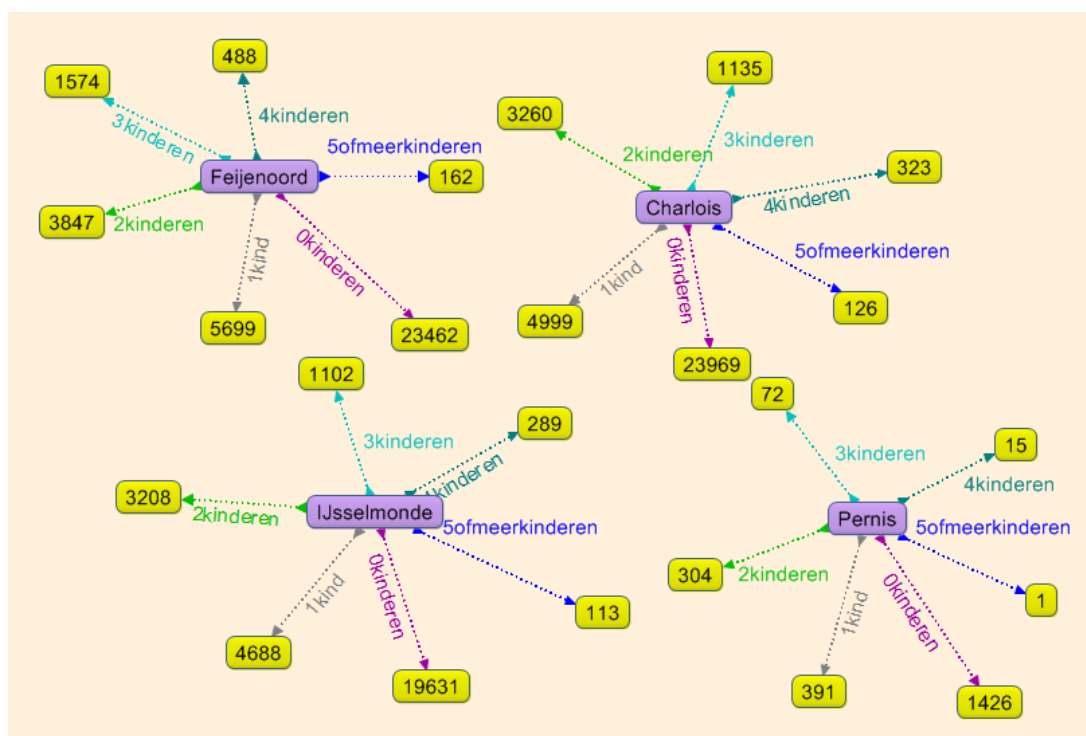


Fig.31. Graph view of a selection of the household data of OBI visualized in Gruff

6.2. CONVERSION OF THE STATISTICAL NEIGHBORHOOD AND AREA DATA

DATA ACQUISITION

The area and neighborhood map of 2014 from the CBS can be downloaded from the CBS website. It comes with a PDF file containing the use restrictions and explanation of terms. When downloading the file folder, it contains three Shapefiles: one for the municipalities, one for the areas and one for the neighborhoods (see figure 32).

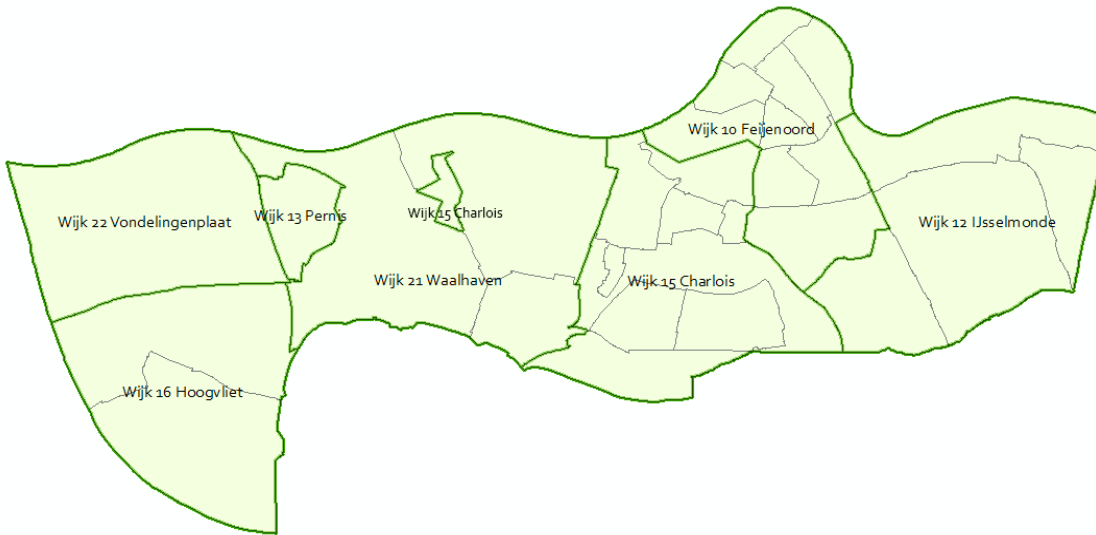


Fig. 32. CBS data for Rotterdam Zuid

DATA ANALYSIS AND PREPARATION

First, only the data from the municipality of Rotterdam is selected in QGIS (see figure 31). Since the Shapefile with the boundaries of the municipality is not needed, this Shapefile is not taken into account. The neighborhood and area Shapefiles contain the names, geometry with the boundaries and a lot of statistics concerning the demography of the neighborhoods and areas. These Shapefiles have corresponding columns with codes, to link them together (see in figures 33 and 34 the corresponding column `wk_code`). These corresponding relations are explicitly mapped (see section 7.2.1). Just like the data from the Rotterdam Open Data Store, this Shapefile is loaded to a table in the geo-database, to make the data accessible for conversion.

	gid [PK] serial	bu_code character varying(8)	bu_naam character varying(60)	wk_code character varying(8)
1	1	BU05991017	Kop van Zuid	WK059910
2	2	BU05991079	Kop van Zuid-Entrepot	WK059910
3	3	BU05991080	Vreewijk	WK059910
4	4	BU05991081	Bloemhof	WK059910
5	5	BU05991082	Hillesluis	WK059910
6	6	BU05991085	Katendrecht	WK059910

Fig.33. Area dataset

	gid [PK] serial	wk_code character varying(8)	wk_naam character varying(60)
1	1	WK059910	Wijk 10 Feijenoord
2	2	WK059912	Wijk 12 IJsselmonde
3	3	WK059913	Wijk 13 Pernis
4	4	WK059915	Wijk 15 Charlois
5	5	WK059916	Wijk 16 Hoogvliet
6	6	WK059921	Wijk 21 Waalhaven

Fig.34. Neighbourhood dataset

DEFINING AN ONTOLOGY

The data is a bit more complicated than the municipality data, since there are multiple columns and two tables that correspond to each other. In the next chapter this will be described in more detail. For the CBS data, the following vocabularies are used (see table 10).

Type of data	Description	Vocabulary
General	General relations between resources	RDFS OWL
Spatial features + geometry	Boundaries (polygons)	GeoSPARQL
Metadata	Descriptive data	Dublin Core VOID

Table.10. Chosen vocabularies for the CBS data

CONVERSION: RDF STORAGE APPROACH

To convert this dataset, the RDF storage approach is again performed, since this dataset contains geometry that will be modeled according to the GeoSPARQL ontology. The geometry consists of polygons which are a member of the Simple Features class Polygon.

The URI for the dataset is [http://ldatarotterdam.nl/id/Wijk/\(Wijknaam\)](http://ldatarotterdam.nl/id/Wijk/(Wijknaam)) for the areas and [http://ldatarotterdam.nl/id/Buurt/\(Buurtnaam\)](http://ldatarotterdam.nl/id/Buurt/(Buurtnaam)) to define the neighbourhoods.

6.3. CONVERSION OF THE BASE REGISTER FOR ADDRESSES AND BUILDINGS (BAG)

6.3.1. DATA ACQUISITION

The BAG is open data which can be downloaded from the Nationaal Georegister. This website is the national website for describing and downloading spatial data of the Netherlands (Nationaal Georegister, n.d.). The BAG, named *INSPIRE downloadservice voor PDOK* on the Nationaal Georegister website exists of multiple files in XML/GML-format. The tool NLEextract is therefore developed to convert and visualize open governmental data, such as the base registers and TOP10NL. With this tool it is possible to load data in a geo-database (NLEextract, n.d.). A dump of the BAG from 8 November 2014 was loaded into the geo-database

PostgreSQL/PostGIS, which resulted in 17 tables (see table 11). The highlighted rows list the tables that contain geometry.

1.	Addressable object Secondary Address	10.	Public Space
2.	Municipality	11.	Premise
3.	Municipality_Province	12.	Province
4.	Municipality_Place of Residence	13.	Mooring Location
5.	Gt_pk_Metadata	14.	Residential Unit
6.	Pitch Location	15.	Residential Unit Purpose of Use
7.	nlx_bag_Info	16.	Residential Unit Premise
8.	nlx_bag_Log	17.	Place of Residence
9.	Number Indication	18.	Addresses (separate CSV file)

Table 11. Existing tables in the BAG- data

The downloaded BAG dump did not contain an address table. The addresses of the residential units were later added to an empty table in the database from a CSV- file that was found as a separate file in the NLExtract tool.

6.3.2. DATA ANALYSIS

To understand the data model and to know which data is relevant for end users, the data needs to be analyzed. The UML- model (see figure 35) describes the relations between the different tables. Each table from the BAG- data is defined as a class in the UML-model along with the columns as properties.

The extra class addressable object (adresseerbaar object) is added to form the link between the other classes. An addressable object is related to the number indication (nummeraanduiding). The table number indication holds all the addresses and is related to a place of residence (woonplaats) and public space (openbare ruimte). An addressable object can have multiple addresses which are stored in the table Addressable Object Secondary Address (adresseerbaarobjectnevenadres) (E-overheid, n.d.-a). A mooring location (ligplaats), pitch location (standplaats) and a residential unit (verblijfsobject) are all addressable objects and contain an address. A premise (pand) can have one or more residential units (verblijfsobject), but not every premise has to be a residential unit.

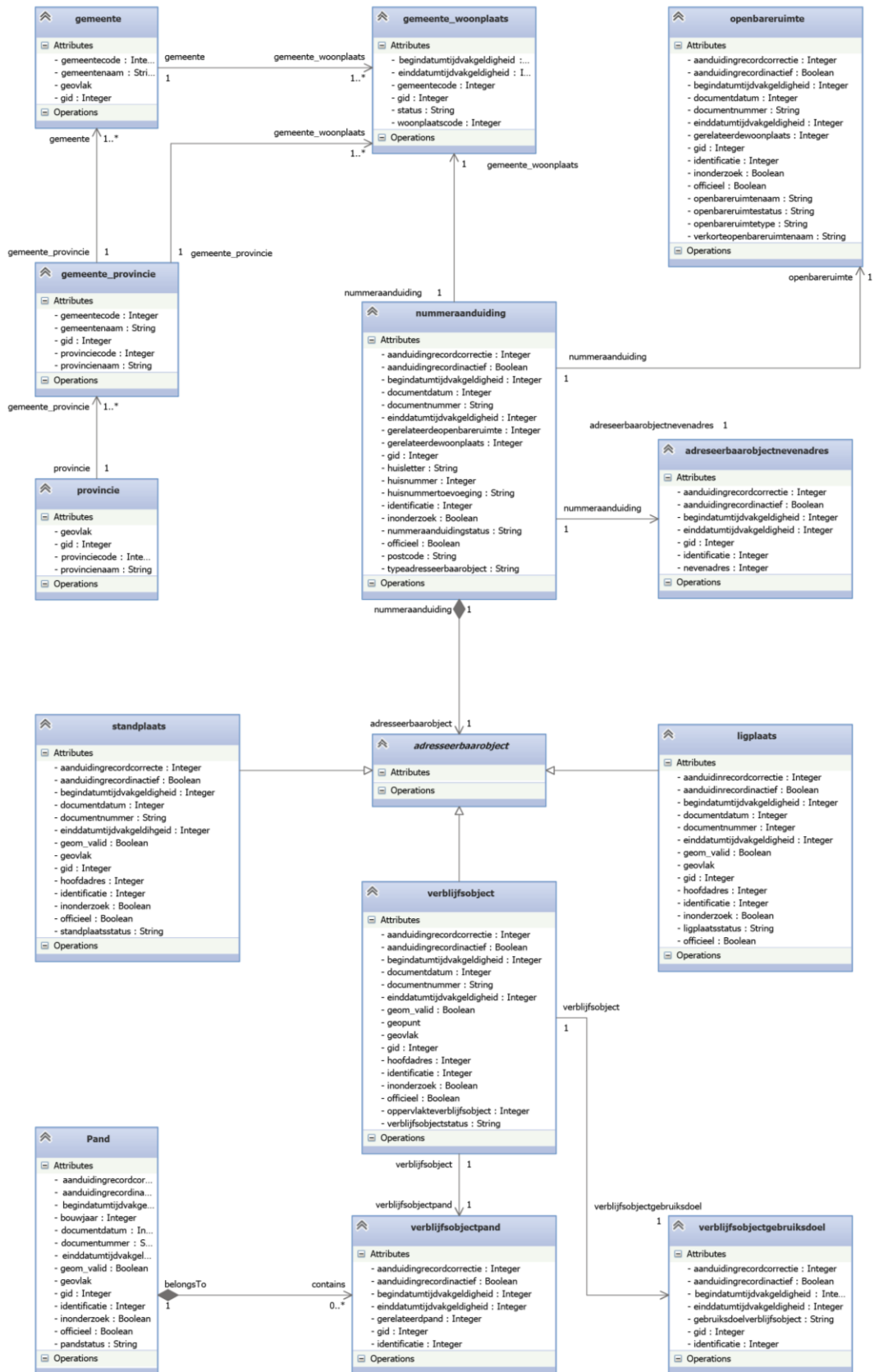


Fig. 35. UML- model with the classes in the BAG

6.3.3. DATA PREPARATION

After analyzing the data, filtering of the data is needed because not every class, column or row needs to be mapped for this case study. Filtering is done on:

- select all the data inside Rotterdam Zuid
- select only the 'interesting' columns
- select all the data with measured geometries
- `geom_valid = TRUE`

Since only the data from Rotterdam Zuid is relevant for this research, these rows are filtered from the data of the whole of the Netherlands. By performing SQL-queries were all the objects inside the borders of Rotterdam Zuid were selected, tables with only entries for Rotterdam Zuid were created.

Also, not all the tables seemed relevant for the case study. Only data that seemed of interest for potential end users will be converted to RDF. One of the reasons to use the BAG is to link informative spatial datasets to addresses and geometry of the premises in the BAG, to provide more information. Interesting data from the BAG could be information that explains details about the building itself, such as the construction year or area size. Administrative information however, such as document number or document date, seemed to be only of interest for the (expert) users of the BAG and where therefore not mapped to RDF.

When showing the layers in QGIS, it can be seen that some premises contain geometry that is not represented correctly. These geometries show up as large triangles (see figure 36). To prevent errors when performing queries based on geometry in the application later on, these incorrect geometries were removed from the table in the database. Also, geometry that was not valid, was deleted from the tables.



Fig.36. Incorrect premises and residential units in red

To conclude, the data that was converted to RDF consists of:

- The premises located in Rotterdam Zuid;

Containing the start- and enddate, the state of the object, the year of construction and the geometry. Different versions of the premises exist.

- The residential units in Rotterdam Zuid.

Containing the area and purpose of use of the residential units. Not all the residential units are labelled with a function though. And just like the premise table, the table with the residential units contains different versions.

- The addresses corresponding to the residential units in Rotterdam Zuid.

The table addresses contains addresses of residential units, geometry and historical data. The addresses are related to a residential unit, but not to an actual version of the residential unit. Thus one residential unit, can contain multiple addresses located on the same place.

By not converting all the classes and attributes from the BAG, the creation of unnecessary triples not relevant for end users is diminished, which makes the dataset more clear and transparent for users that are not familiar with the BAG.

6.3.4. DEFINING AN ONTOLOGY

The data model of the BAG is more complex than the previously discussed datasets, since the BAG exists of multiple corresponding tables. The table below (table 12) describes the used vocabularies to describe the BAG dataset.

Type of data	Description	Vocabulary
General	General relations between resources	RDFS OWL Simple Knowledge Organization System (SKOS)
Spatial features + geometry	Premises and Addresses (Point and Polygon-objects)	GeoSPARQL
Metadata	Descriptive data	Dublin Core VOID

Table.12. Chosen vocabularies for the BAG

Since the data model of the BAG is more complex, an ontology is defined, based on the UML-model and the vocabularies described in the table above, to explain the RDF mapping. This model (see figure 37) shows the relations in the RDF dataset.

Basically, the classes to map in the UML-model, become nodes in the graph. These nodes, addressable object (addresses), residential unit and premise are all OWL-classes and are hierarchically ordered by being subclasses of another. Since they are connected to each other, only the addressable object has to be mapped to a spatial feature.

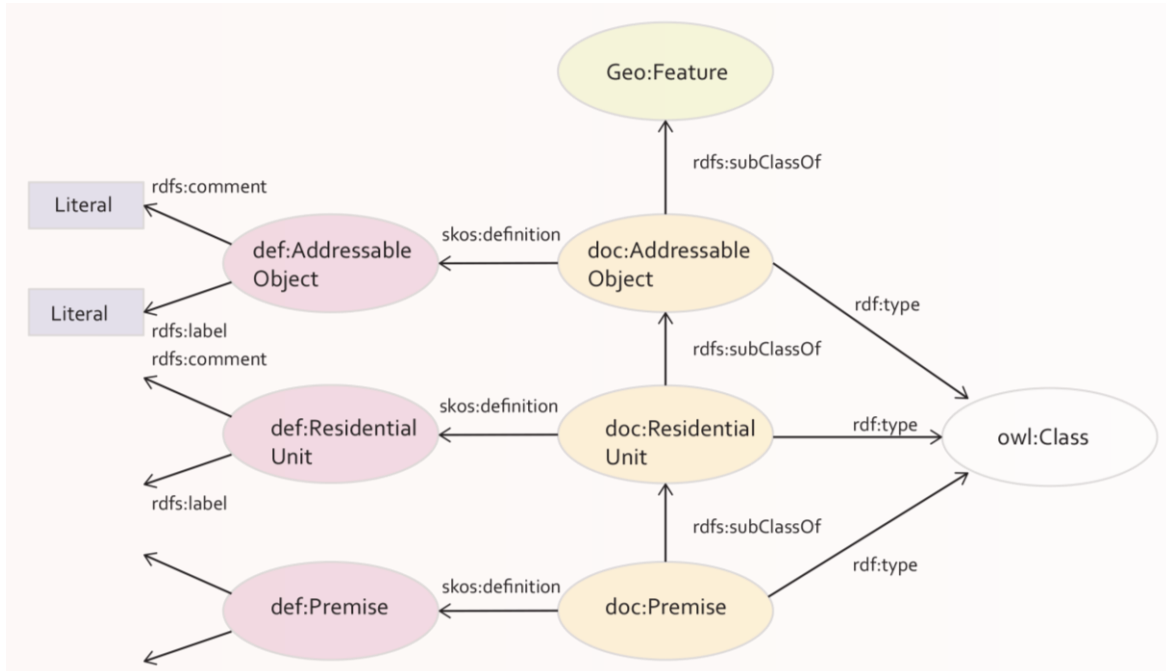


Fig.37. Ontology of the BAG

SKOS is added to relate the classes with the definitions used in the BAG data. Since the definitions of terms in the BAG warrant some description.

6.3.5. IMPLEMENTING THE URI STRATEGY

The BAG ontology is defined in the namespace `doc` from the URI- strategy: <http://ldatarotterdam.nl/bag/doc>. Also the metadata is defined under the same `doc` identifier. The definitions of the classes (as seen as the purple nodes in figure 37) are defined in the `def` namespace, where the link then becomes <http://ldatarotterdam.nl/bag/def>. The actual data, according to the URI-strategy is under the `id` namespace: <http://ldatarotterdam.nl/bag/id>.

MAPPING - METADATA

The 'def' type contains all the definitions for each class. Each definition has a relation with a class from the data model (see figure 37). The definitions are taken from the Stelselcatalogus, which is a document that ensures non-ambiguous definitions for the base registrations (E-overheid, n.d.-b). This is also referenced in the metadata for the linked dataset. Each class in the BAG is given a label and a comment with these definitions. This is exemplified in section 7.3.

MAPPING – HISTORICAL DATA

The BAG is a historical dataset, which means there are multiple versions of the same spatial objects through time in the data. A premise or residential unit is stored with its identification number (for a premise in Rotterdam this starts with 5991, for a residential unit in Rotterdam it starts with 59901). Whenever there is an update of a premise or residential unit, a new entry is made in the BAG table, in where the object is stored with the same identification number, but with different ids (named gid). The more recent the update is, the higher the id is.

For example, a premise can be stored once at the start of the construction and again when the premise is in use. The image below shows how the different versions of an object are mapped (see figure 38). First one triple is made of the object with the version, and then other triples are defined with the version and the actual related data to that premise.

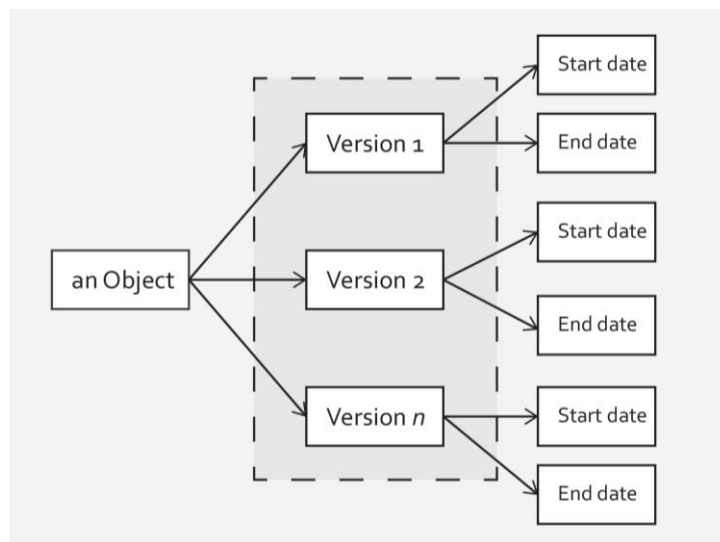


Fig.38. Concept of a premise with its versions

The table below (see table 13) shows how this concept of figure 36 is mapped to URIs.

	Subject (object in fig.38)	Predicate	Object (version in fig. 38)
Node	'599100100002429'		'7931344'
URI	http://ldatarotterdam.nl/bag/id/Pand/599100100002429	gid	http://ldatarotterdam.nl/bag/id/Pand/599100100002429/7931344

Table 13. Concept of the historical data in the BAG translated to URIs

6.3.6. CONVERSION: RDF STORAGE APPROACH

To convert this dataset, the RDF storage approach is performed, since this dataset contains geometry that will be modeled according to the GeoSPARQL ontology.

The RDF script was constructed the same way as for the other datasets. When converted to RDF an example of the data can be seen in the image below (see figure 39, in where a premise is modelled with its versions. The geometry is omitted in this figure to clarify the structure of the data).

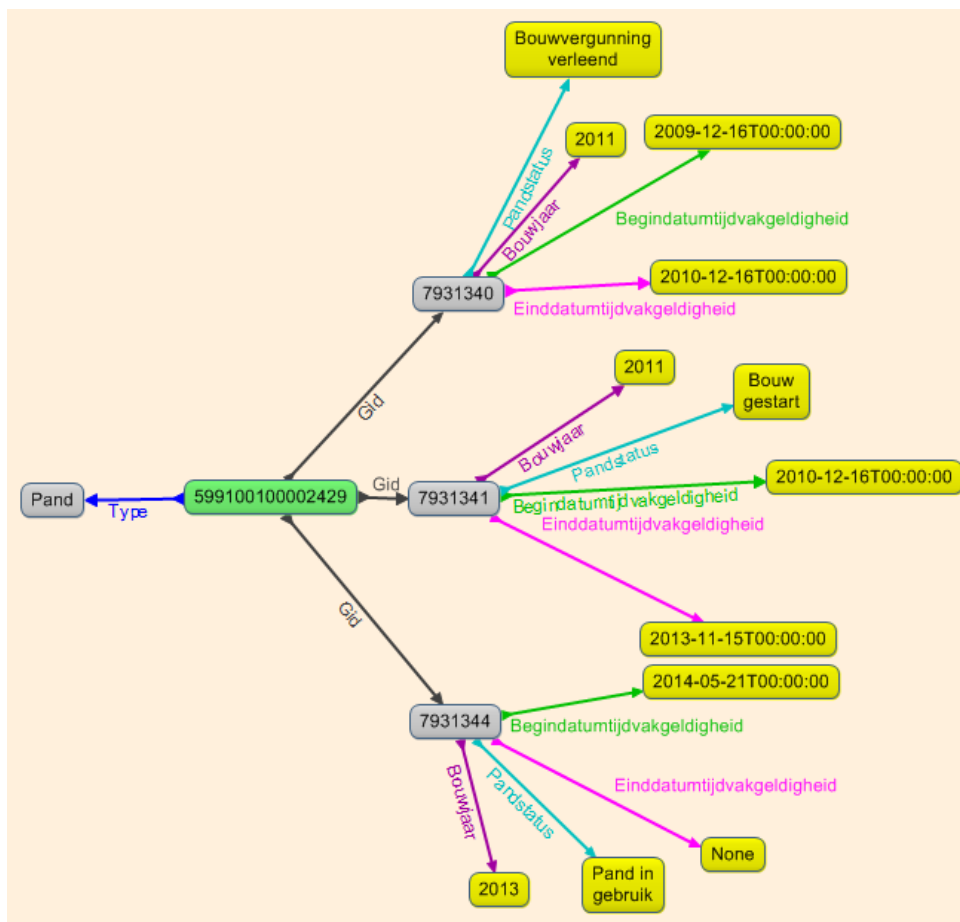


Fig.37. a premise with its different versions- modelled in Gruff

Now it is possible to ask for a certain version in time. In the query which can be seen below, is asked for a version of the premise form figure 39 between 2010 and 2014, which will result in version 7931341.

```

SELECT ?version
WHERE {
  < http://ldatarotterdam.nl/bag/id/Pand/599100100002429> Pand:gid ?version .
  ?version Pand:begindatumtijdvakgeldigheid ?bdate .
  ?version Pand:einddatumtijdvakgeldigheid ?edate .
  FILTER(xsd:dateTime(?bdate)>
xsd:dateTime("2010-01-01T00:00:00Z"))
  FILTER(xsd:dateTime(?edate)<
xsd:dateTime("2014-01-01T00:00:00Z"))}

```

To get the last version of the premise, the query as seen below can be performed. Basically there is asked for the highest value of the version, since the highest version of a premise is also the latest version of this premise.

```

SELECT (MAX(?version) AS ?version)
WHERE {
  < http://ldatarotterdam.nl/bag/id/Pand/599100100002429> Pand:gid ?version .}

```

Finally, the figure below shows a subset of triples of the BAG in RDF containing a premise with a few versions with the attributes and the geometry. The premise is related to (one of the) residential units along with the attributes and the corresponding address with the point geometry (see figure 40 on next page).

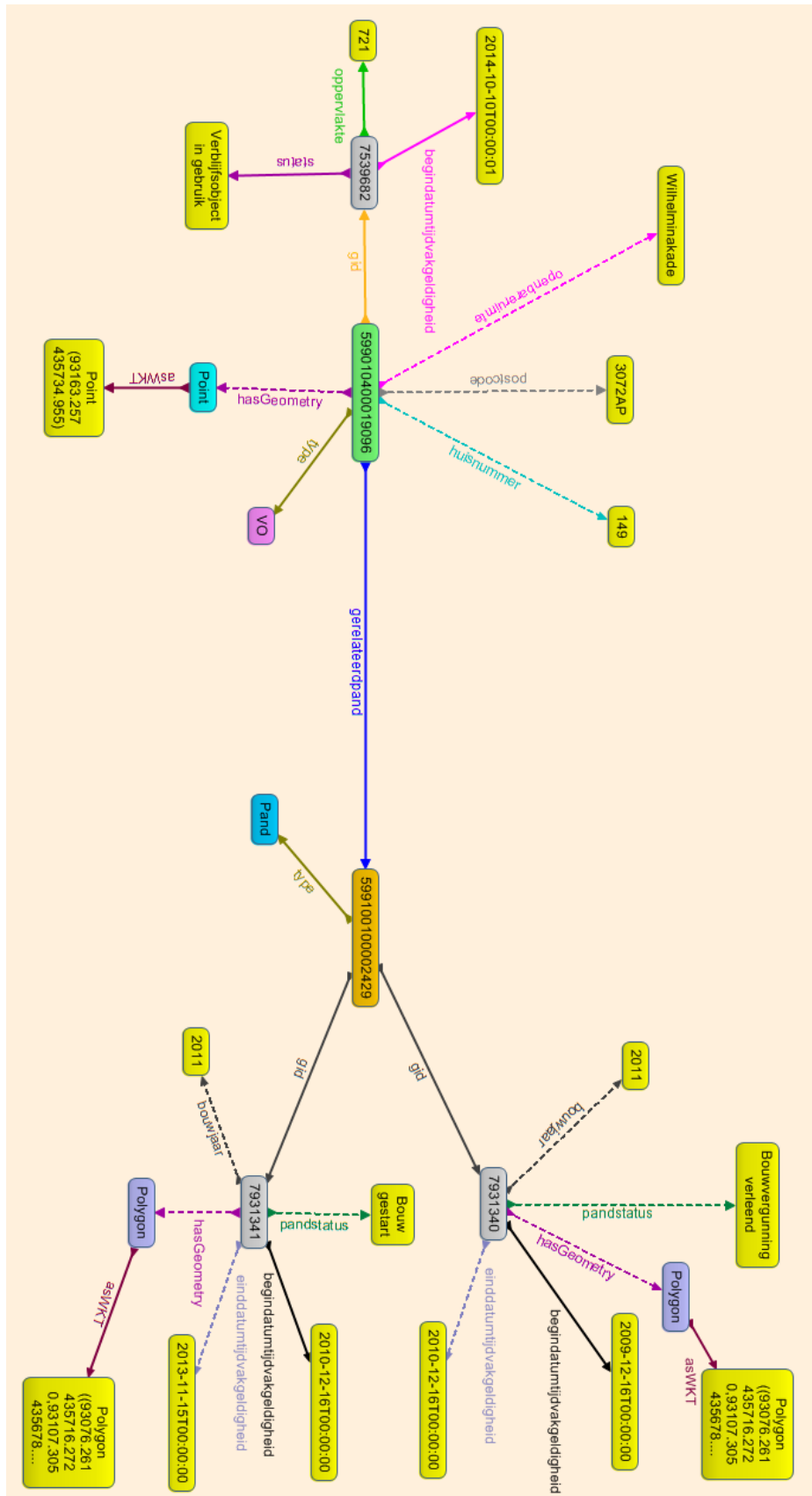


Fig.40. a premise and residential unit along with their attributes from the BAG visualized in Gruff

DIFFERENCES WITH THE BAG FROM OTHER SOURCES

Since as described in section 3.4, currently multiple versions of the BAG exists. The difference with this version of the BAG is that this version contains geometry modeled according to the GeoSPARQL ontology. Also, the data model is very much simplified by not converting all the BAG data to RDF. The data quality is updated, since there are made links to the Stelselcatalogus to explain definitions in the BAG.

6.4. CONCLUSION & DISCUSSION

This chapter provided insight into the conversion of different data formats to RDF. A general remark is that for working with spatial data it is currently a task to find the right software and tools to support the GeoSPARQL ontology. A consideration needs to be done, which conversion tool, triple store and query engine will be used that all support the GeoSPARQL ontology.

An observation is that datasets with a simple data model, such as the OBI and kindergarten data, were obviously more easily and faster converted than datasets with a difficult data model, such as the BAG.

Even though the URI-strategy is not meant to be used for data that does not belong to base registration or standards, it provided clear and recognizable URIs for all datasets. The Municipality would benefit from having a recognizable URI with the same domain name, so that layman users are able to find data from the Municipality.

6.4.1. COMPARISON OF THE APPROACHES

Looking at the two approaches, the D2RQ approach performs a fast mapping, even when the table in the database contains many rows (such as the BAG). Another advantage is that the data is already published to the Web, thus no additional tool is needed. Furthermore, some metadata is added automatically. The generate mapping tool provides a start with the mapping, but manual changing is needed to change the URIs and if the dataset has a complex data model, the mapping needs to be altered as well. Also the data types and when reusing existing vocabularies/ontologies, these need to be added manually. Lastly, it takes some time to understand the D2RQ language. For spatial data though, the D2RQ platform can be used to create the mapping according to the GeoSPARQL ontology, but a SPARQL endpoint that can perform GeoSPARQL queries is needed. I found it faster to do the mapping via the RDF storage approach, because the mapping file generated by D2RQ in the end needed a lot of customization.

The RDF storage approach scored high in customization of the RDF data. Also uploading the RDF files to the triple store Strabon went very fast and storing the triples in an actual triple store, provided an overview of the data. Dependent on the

triple store, the stores provide more functionality than a virtual mapping does. The script for creating the RDF file however, worked fine for small datasets, but took very long to generate the BAG RDF file for example. This approach took more manual labor, but this is a onetime effort. Also, to diminish the manual labor, this method could be automated. In table 14, these approaches are compared. The RDF storage approached performs well on customization and is suitable for spatial data, but is more manual labor and is slower. The on-the-fly approach performs well in speed, but provides fewer possibilities for customization without manual labor.

	RDF storage (Python + Strabon)	On-the-fly to RDF (D2RQ)
Customization	++	--
Speed	--	++
For spatial data	++	--
Manual labor	--	-+
Data up to date	--	++

Table 14. Comparison of the two methods

In the end the on-the-fly approach is preferred, since this would ensure that the data is up to data which is important for the data quality. One change made in the database would ensure that also in the on-the-fly version of the RDF data the change is visible. With the RDF storage approach, there are two versions of the data and the data publisher needs to implement changes in both versions of the data. This is a lot more work, but could be an automated process.

6.4.2. EVALUATION METHODOLOGY

In this chapter four datasets with different datasets were converted to linked data. The methodology as described in chapter 4 is used for these datasets. As a general guideline, the methodology proved to be working. In the end though, the steps were a bit extended. In phase 2, the modeling is added the implementation of the URI strategy. In phase 3, the conversion is also taken into account the conversion of the metadata. Other datasets, such as for example energy data, can be transformed to linked data with the same methodology, though the output can be different, depending on the type of data and who the users will be of the data.

In the W3C Cookbook and step-by-step guide from TNO, adding governance is added to the steps. The methodology provided in this thesis does not include this. This methodology only focuses on creating spatial linked open data. The methodology can be used in addition with a lifecycle model as seen in section 4.7.2.

7. POSSIBILITIES OF USING SPATIAL LINKED OPEN DATA

This chapter describes the possibilities and limitations of the generated spatial linked open data from the previous chapter. Descriptions of how to query the data, the linking of multiple datasets and how to integrate the metadata with the dataset are provided. An answer to the sub questions: ‘How to query distributed and heterogeneous spatial datasets?’ and ‘What are options for linking datasets?’ are given.

7.1. SPATIAL ANALYSIS OF THE CREATED SPATIAL LINKED DATA

One of the advantages of Linked Data is the possibility to do SQL-like queries on the Web, with the standard SPARQL. With SPARQL, you can ask a question over the data and get a specific bit of data back. In order to do that, knowledge of the SPARQL standard, as well as (some of) the URIs that are used in the datasets and (some of the) relations that exist in the data is needed. Knowing some of the URIs and/or relations in the data is needed as a starting point to discover the rest of the data. With the standard GeoSPARQL and a spatial dataset containing geometry, spatial queries can be done. In the section below, an example of this is given, based on the data that is converted to spatial linked data. This analysis illustrates as an example of some questions that can be asked over the data.

7.1.1. SPATIAL QUERY

Suppose someone wants to open up a kindergarten somewhere in Rotterdam Zuid. A question would be: How many kindergartens are in which areas in Rotterdam Zuid? This to know what would be the best place to open up this new service. First a simple query can be done on the CBS area dataset to know which areas there actually are. The original question can be answered with a spatial query, done on the CBS area and neighborhood map for the boundaries of the areas and the dataset about the Kindergartens from the Rotterdam Open Data Store.

```

SELECT ?wijk (COUNT(?kinderdagverblijf) AS ?amount)
WHERE{
?kinderdagverblijf a <http://ldatarotterdam.nl/id/Kinderdagverblijf/> .
?kinderdagverblijf geo:hasGeometry ?geo .
?geo geo:asWKT ?wkt .
?wijk a <http://ldatarotterdam.nl/id/Wijk/> .
?wijk geo:hasGeometry ?wgeo .
?wgeo geo:asWKT ?wwkt .
FILTER(geof:sfContains(?wwkt, ?wkt) .}
GROUP BY ?wijk

```

The query can be seen in the selection above. First there is asked for every Kindergarten in the dataset and their geometry. Then the geometry from the CBS data is asked. A filter with the Simple Features Contains function gives the following result as seen in table 15 back. Important to notice is that the spatial functions are performed on the representations of the geometries, thus the WKT objects.

wijk (area)	amount
http://ldatarotterdam.nl/id/Wijk/Charlois	"9"^^<http://www.w3.org/2001/XMLSchema#integer>
http://ldatarotterdam.nl/id/Wijk/Feijenoord	"10"^^<http://www.w3.org/2001/XMLSchema#integer>
http://ldatarotterdam.nl/id/Wijk/IJsselmonde	"7"^^<http://www.w3.org/2001/XMLSchema#integer>
http://ldatarotterdam.nl/id/Wijk/Hoogvliet	"4"^^<http://www.w3.org/2001/XMLSchema#integer>

Table 15. Result spatial query

This results show that there are only kindergartens in 4 areas. Considering this, a new question may arise, why are there no kindergartens in the areas Vondelingenplaat, Pernis and Waalhaven? General information about the areas in Rotterdam Zuid can be useful, which can be found at DBpedia, the Linked Data variant of Wikipedia. To access the information in Strabon, a federated query will be performed.

7.1.2. FEDERATED QUERY

A federated query allows users to combine distributed data across the Web. Basically, a SPARQL endpoint sends part of the query to another SPARQL endpoint, which sends the result back and present to the user the combined result. With a federated query it is thus possible to query locally stored RDF data in combination with RDF data that is published on the Web by others. This type of query is only possible if the query processor supports the SPARQL 1.1 Query Language (Prud'hommeaux & Buil-Aranda, 2013).

Assuming that the URI of the page of Wikipedia is not known, firstly a search is done to see if there are actually information resources about these three areas. First, the area Waalhaven is researched. The Service statement in the query contains the SPARQL endpoint for the DBpedia dataset.

```
SELECT *
WHERE {
SERVICE <http://nl.dbpedia.org/sparql> {
dbpedia:Waalhaven ?p ?o . }}

```

Apparently, there are more resources that contain the name Waalhaven (see figure 41).

P	O
http://www.w3.org/2000/01/rdf-schema#label	"Waalhaven"@nl
http://dbpedia.org/ontology/wikiPageDisambiguates	http://nl.dbpedia.org/resource/Vliegveld_Waalhaven
http://dbpedia.org/ontology/wikiPageDisambiguates	http://nl.dbpedia.org/resource/Waalhaven_(Rotterdam)
http://dbpedia.org/ontology/wikiPageDisambiguates	http://nl.dbpedia.org/resource/Waalhaven_(wijk)
http://dbpedia.org/ontology/wikiPageDisambiguates	http://nl.dbpedia.org/resource/Waalhaven_(Nijmegen)

Fig.41. Printscreen of the result from a federated query in Strabon

By clicking on the link in the object section for the Waalhaven in Rotterdam we can access the DBpedia page about the Waalhaven were a lot of information about this area can be found (see figure 42). Apparently, the Waalhaven is a harbor area.



Fig.42. Web page of the Waalhaven on DBpedia (DBpedia, 2015)

The queries for the other areas are visible below. The Information about Pernis is asked by performing the same query as the one for Vondelingenplaat, but then by changing the name to Pernis. Since there is only one resource for each area, there is directly asked for a comment, which is mostly a short description of the information resource. The comment is directly visible in Strabon, which means there is no redirecting to another page.



Fig.43. Printscreen of the results from a federated query in Strabon

The results show that the area Vondelingenplaat and Waalhaven are industrial areas which makes it quite likely that there is a very small population and thus no need for a kindergarten. The area Pernis consist of the village Pernis and is surrounded by harbours (see figure 43).

To be sure about the size of the population in the three areas, a follow-up question could be: how many children are actually in the areas in Rotterdam Zuid? By performing a query on the dataset from the CBS, which contains demographic statistics about the areas in Rotterdam Zuid, this question can be answered.

```

SELECT?wijk ?numberofhouseholds ?percentagewithchildren
WHERE{
?wijk wijk:aantal_hh ?numberofhouseholds .
?wijk wijk:p_hh_m_k ?percentagewithchildren .}

```

wijk	number of households	percentage with children
http://datarotterdam.nl/id/Wijk/Charlois	"32580"	"30"
http://datarotterdam.nl/id/Wijk/Pernis	"2140"	"37"
http://datarotterdam.nl/id/Wijk/Vondelingenplaat	"0"	"-99999997"
http://datarotterdam.nl/id/Wijk/Hoogvliet	"15575"	"36"
http://datarotterdam.nl/id/Wijk/Feijenoord	"34970"	"34"
http://datarotterdam.nl/id/Wijk/IJsselmonde	"27720"	"34"
http://datarotterdam.nl/id/Wijk/Waalhaven	"25"	"13"

Table 16. Results of the performed query

This result in table 16 shows that the presumptions are true, there are actually not much households in the harbor areas (zero in Vondelingenplaat). Pernis has actually the highest percentage of children, though one of the lowest amounts of households. We can specify this amount of children further by linking the data from the CBS dataset with the dataset from Rotterdam in Cijfers, which is described in the following section.

7.2. LINKING TO OTHER DATASETS

By linking to other datasets, more additional information is given to the user. There are different ways to make links between the datasets.

7.2.1. BY CREATING EXPLICIT LINKS IN THE RDF DATA

The CBS dataset contains area names and the Rotterdam in Cijfers datasets also. One way of linking these two datasets, is to make explicit links between the datasets at the level of conversion to RDF. Thus this should be considered during the mapping. New triples are created to define that there is a relation between the two datasets, namely the owl:sameAs predicate, which is used to define that two URIs define the same resource. Part of the mapping can be seen below.

gebied:Feijenoord,	owl:sameAs,	wijk:Feijenoord
gebied:Charlois,	owl:sameAs,	wijk:Charlois

These triples are also added to the triple store Strabon. Then a query as seen below, can be performed. To know more information about the number of children per household in the area Pernis, the corresponding subject for Pernis in the Rotterdam in Cijfers dataset can be found. The result can be seen in table 17.

```
SELECT DISTINCT ?p ?o
WHERE{
?gebied owl:sameAs wijk:Pernis .
?gebied ?p ?o
FILTER(datatype(?o) != xsd:integer) .}
GROUP BY ?p ?o
ORDER BY ASC(?p)
```

p	o
http://ldatarotterdam.nl/id/Huishoudens/okinderen	"1426"
http://ldatarotterdam.nl/id/Huishoudens/1kind	"391"
http://ldatarotterdam.nl/id/Huishoudens/2kinderen	"304"
http://ldatarotterdam.nl/id/Huishoudens/3kinderen	"72"
http://ldatarotterdam.nl/id/Huishoudens/4kinderen	"15"
http://ldatarotterdam.nl/id/Huishoudens/5ofmeerkinderen	"1"

Table 17. Result of the explicit linking query

7.2.2. LINKING ON TOPOLOGY USING GEOSPARQL

The next query shows an example where the BAG is combined with the CBS data in where dynamic linking is used.

In the query there is asked for which addresses lie in the area Feijenoord. This is done by performing a Filter with the Simple Features function `sfWithin`.

```
SELECT ?street ?number ?zipcode
WHERE {
  ?vo rdf:type <http://bag.testp2.nl/doc/VO>.
  ?vo geo:hasGeometry ?geo .
  ?geo geo:asWKT ?wkt .
  ?vo vo:openbareruimte ?street.
  ?vo vo:huisnummer ?number.
  ?vo vo:postcode ?zipcode.
  <http://ldatarotterdam.nl/Wijk/Feijenoord> geo:hasGeometry ?fgeo .
  ?fgeo geo:asWKT ?fwkt .
  FILTER(geof:sfWithin(?wkt, ?fwkt)) .}
```

street	number	zip code
"Anjerstraat"	"1"	"3071VA"
"Wilhelminakade"	"455"	"3072AP"
"Tolhuislaan"	"198"	"3072LL"
"Koninginnenhoofd"	"1"	"3072AD"
"Hortensiastraat"	"3"	"3073TB"

Table 18. Result of the dynamic linking query

The result can be seen in table 18. In this query the BAG is linked by topology with the CBS data. No mapping has to be made during the level of converting the data to RDF.

For this query, a selection of addresses from the BAG of Rotterdam Zuid is stored in the triple store, thus remarks about the scalability/speed of querying cannot be made.

7.2.3. BY DEFINING A SPARQL QUERY ON A FOREIGN KEY IN THE DATA

A third way is to create a SPARQL query based on a shared predicate the datasets have. To show this option, the CBS area and neighbourhood datasets are used. In the tables belonging to the CBS data, can be seen that both the Area dataset and the Neighbourhood dataset have the corresponding column *wk_code* to link the neighbourhoods to the corresponding area (see figure 33 and 34).

To answer the question: “Which neighbourhoods are in the area Feijenoord?”, a SPARQL query as seen below can be performed.

```
SELECT ?neighborhoods
WHERE {
  <http://ldatarotterdam.nl/id/Wijk/Feijenoord> wijk:wk_code ?o .
  ?neighborhoods buurt:wk_code ?o .}
```

This results in a list with the neighborhoods in area Feijenoord. Besides shared predicates, also shared objects such as strings can be used.

7.3. LINK THE DATA WITH THE METADATA

As mentioned in section 7.1.1., every dataset has metadata, which is according to the URI-strategy implemented under the link <http://ldatarotterdam.nl/doc/>.

For the kindergarten dataset, this can be retrieved from the store, by posing the following query.

```
SELECT *
WHERE{
  <http://ldatarotterdam.nl/doc/Kinderdagverblijf/> ?property ?metadata}
```

property	metadata
http://purl.org/dc/terms/description	"Adressen, namen en locaties van kinderdagverblijven in Rotterdam Zuid"
http://purl.org/dc/terms/source	http://www.rotterdamopendata.nl/dataset/kinderdagverblijven
http://xmlns.com/foaf/spec/homepage	http://www.rotterdamopendata.nl/dataset/kinderdagverblijven
http://purl.org/dc/terms/created	"31-07-2012"
http://purl.org/dc/terms/date	"28-04-2015"
http://purl.org/dc/terms/publisher	"Me"
http://purl.org/dc/terms/temporal	"2015"
http://purl.org/dc/terms/license	"Geen licentie verstrekt"
http://purl.org/dc/terms/modified	"04-04-2014"
http://purl.org/dc/terms/creator	"GIS-Diensten"
http://purl.org/dc/terms/title	"Kinderdagverblijven in Rotterdam Zuid"
http://purl.org/dc/terms/spatial	"Rotterdam Zuid"
http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://rdfs.org/ns/void#Dataset

Fig.44. Printscreen of the metadata query in Strabon

Now by posing a simple query on an object of the dataset, the metadata can be directly retrieved. For this, some knowledge of the URI- strategy is needed to know the doc type in the URI. And also, the terms are in English which is maybe a bit strange, considering the dataset is in Dutch. By looking up the whole document, all the metadata can be retrieved (see figure 44).

When doing a query, the metadata can be directly requested, since the metadata is in the same data format as the actual data (see example below and the result in table 19).

```
SELECT ?dataset ?creator ?title
WHERE {
?s ?p "Feijenoord" .
?s void:inDataset ?dataset .
?dataset dcterms:creator ?creator .
?dataset dcterms:title ?title .}
```

dataset	creator	title
http://ldatarotterdam.nl/doc/Wijk-enBuurtkaart	"Kadaster/Centraal Bureau voor de Statistiek"	"Wijk- en Buurtkaart 2014"

Table 19. Result of the metadata query

As explained in section 7.3.5, is the BAG linked to definitions from the Stelselcatalogus. Now when exploring the dataset, the definitions, residing in the namespace <http://ldatarotterdam.nl/def/pand> can be requested (see fig.45).

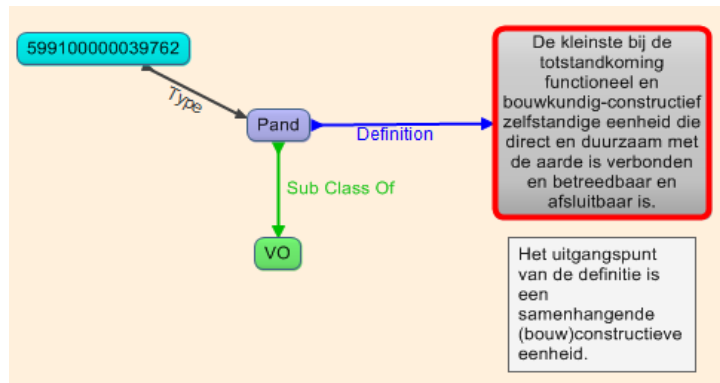


Fig. 45. Example of a premise and its definitions taken from the Stelselcatalogus-visualized in Gruff

7.4. CONCLUSION & DISCUSSION

7.4.1. CONCLUSIONS REGARDING SPATIAL QUERYING

The queries shown in this chapter exemplify the potential of spatial Linked Data. It showed that a simple analysis of multiple datasets can be done with the help of functions based on the GeoSPARQL ontology. Though not all functions are yet supported in Strabon.

According to literature, six types of spatial queries are distinguished (Veer van 't, 2014a). Namely, reverse geocoding, which means finding an element based on a named location. Range queries, for finding spatial entities in a bounding box. Spatial join, in where one type of geometries intersect other type of geometries. Nearest neighbour, finding the closest element on a specified location. Spatial aggregation, in where spatial elements are combined or buffered. Transformations, such as spatial reprojection, scaling or rotation (Veer van 't, 2014a). In table 20 the performance of the different type of queries that are implemented yet in Strabon can be seen. Spatial aggregation functions, such as a buffer or union do sometimes work. But sometimes an error occurs.

Type of spatial query	Possible
Reverse geocoding	Yes
Range queries	Yes
Spatial join	Yes
Nearest neighbor	Yes
Spatial aggregation	Partially
Transformations	No

Table 20 . Possibility of spatial queries

Considering both type of users and the possible queries they may want to ask for the goal of the Web app, the following functions may want to be used:

- geof:sfContains/ehContains/ehCovers: “Which schools are inside this area?”
- geof:sfWithin/ehInside/ehCoveredBy: “Which schools are inside this area?”
- geof:buffer: “Which facilities are in a range of 5 km around my house?”
- geof:distance : “What is the shortest distance between my house and a school?”

Thus only the buffer function provides sometimes a problem.

Linked data seems to have an explore and analysis function. If someone needs to do an analysis, multiple datasets can be researched by using spatial linked data. If a more complex operation or data processing needs to be done, a GIS or geo-database should be used. More often, complex operations need to be done on a subset of the data and visualized in a way that is difficult for semantic web based systems (Veer van 't, 2014a).

The analysis also showed that distributed and heterogeneous datasets can be linked to the locally stored ‘own’ data by performing a federated query. Performing a federated query depends on the functionality of the query engine. Not all query engines support this type of query, possibly because it can take a heavy load on the server where the query engine is stored, to perform these types of queries. A federated query can be useful to add additional information to the data to provide more context and also to help solve cross knowledge domain issues, since for this type of problems different datasets from different sources is needed. An example of this is the risk map.

7.4.2. CONCLUSIONS REGARDING LINKING

The linked of datasets can be done in three ways. These three ways of linking, show that at different levels links between datasets can be made. At the level of mapping where explicit links between the data objects can be made. And at the level of querying the data, where linking on a foreign key and on topology is possible.

Making explicit links seems to be useful when designing an application with Linked Data. Since then in advance can be thought of which datasets to combine and how, this will result in easier and faster querying, but this takes more manual work.

Linking on foreign key can only be done between datasets that already correspond to each other, like in the example or between the BGT which has a corresponding column that points to a BAG premise. This way of linking seems useful for a set of related datasets, such as the base registrations. A requirement though, is that the shared predicate needs to be known in advance to pose the query.

Linking on topology seems to be the most effective method, since you do not have to know in advance the predicate or have to create links at the level of mapping. Linking on topology can be used to link datasets from the Rotterdam Open Data Store, since what connects these datasets together, is the location. The TIR is a dataset that consist of the boundaries of areas and neighborhoods in Rotterdam. The datasets from the open data portal can be linked to the TIR.

Though a limitation for linking on topology to other datasets that reside on the Web, such as DBpedia, the spatial data needs to be converted to WGS84, since many spatial datasets use that reference system.

This chapter also proved insight in how the metadata can be requested along with a query. Since the metadata is of the same data format as the data, they can be easily linked. Especially with Linked Open Data it is important to add the provenance of the data, to be able to correctly interpret the data. Also, adding the provenance of the data makes the data more reliable and improves the quality of the dataset. Different datasets concerning the same subject/theme can then be compared (Harth & Gil, 2014).

Since Strabon is originally designed to be used in comparison with the standards stRDF and stSPARQL, these were also tested briefly. It appeared to be used far more easy to map triples to the stRDF standard and to use stSPARQL, since stRDF does not define a spatial ontology, thus the classes in a dataset does not need to be mapped to the stRDF ontology. The only requirement is that the geometry has the correct data type as defined in stRDF. Also the functionality that stSPARQL provided seemed similar to GeoSPARQL.

8. VISUALIZING SPATIAL LINKED OPEN DATA

In the previous chapter various SPARQL queries were performed to show the potential of linked data. Though, assuming the end users of the Web app are not familiar with SPARQL, a Web application can give access to the data in a more user friendly manner in the form of a map with clickable objects. Also, spatial data is preferred to be visualized on a map.

This chapter provides insight in the Web app and the visualization of linked data. First the current situation of providing open data by the municipality is discussed, will be described in more detail, thereafter the Web map and a possible implementation of the linked data system architecture.

8.1. CURRENT SITUATION OF PUBLISHING OPEN DATA IN ROTTERDAM

Most of the data from the municipality is stored in a database management system (DBMS). From there, multiple applications and services are connected to the database and each other, to provide the functionality for managing objects in public space, an inventory of these objects and the registration of activities in public space. Consumers of this system are GisWeb for maps as well the Rotterdam Open Data Store. The data that is provided for GisWeb and the Rotterdam Open Data Store comes from a data warehouse. The data in the warehouse consist of administrative data of the public space (Beheer Openbare Ruimte). These thematic datasets in the data warehouse are exported every night from the central database and copied to GisWeb and to the Rotterdam Open Data Store. Mutations in the original data are thus visible one day later (R. Keijser, personal communication, 2015)

For GisWeb, the Open Source Tools Mapserver and OpenLayers are used to show data on a map. For the Open Data Store, Mapserver is used to support the WMS/WFS that is used for publishing the data.

8.2. RDF ON THE WEB

As mentioned in section 4.5, the RDF data can be published to the Web in several ways. For spatial data though, it is preferred to show the polygons or point locations on a map. To view the data in its context, to be able to make better conclusions and to analyze the data properly. Therefore, a Web application seems the best option to show spatial data, since through a Web API, the data can be shown on a map and made interactively, through the possibility of clicking on objects.

8.2.1. VISUALIZING IN STRABON

The triple store Strabon also provides the possibility to show data on a map. When posing a GeoSPARQL query, the RDF result is mapped to KML to show the output on Google Maps (see figure 46).

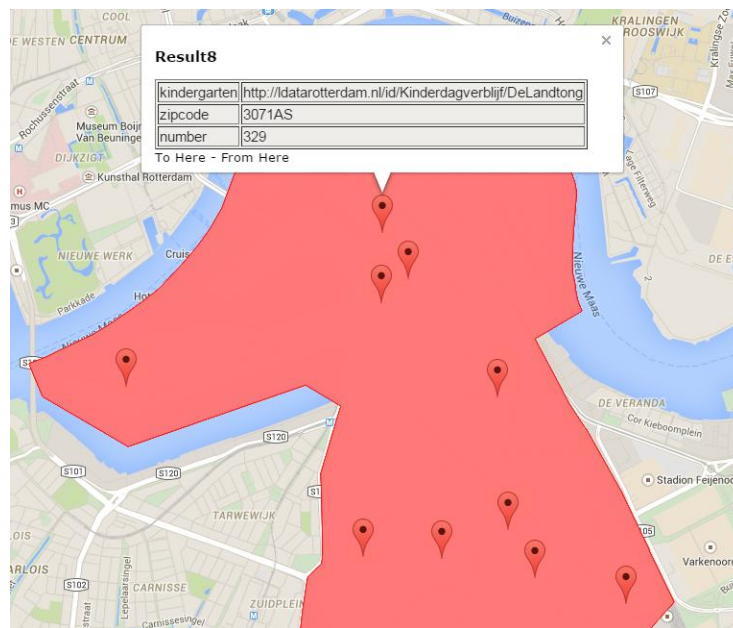


Fig.46. the result of a query asking for all the kindergartens including zip code and house number in the area Feijenoord visualized in Strabon

Though this is a very nice option in Strabon, it requires to first do a query, before a visualization is made. Also this requires users to have Strabon installed. Therefore, an additional Web map is required which shows the result of a query as spatial linked open data on the screen. Strabon can then be used as a backend for the Web application in where the application communicates with Strabon for the queries and to access the linked data in the triples tore.

8.3. RDF ON A WEB MAP

To show RDF on the Web in a web page (or map), the RDF needs to be embedded in HTML. Web pages are written in HTML in combination with JavaScript to provide more functionality such as clicking on objects.

Since RDF consist of triples and a triple consist of URIs, these URIs should be embedded in HTML. This can be done in several ways, namely by writing RDF as JSON-LD and GeoJSON.

JSON-LD

JSON-LD is based on JSON, which is a machine readable format, but also human readable. JSON is much used in web applications since web browsers can directly read it, because it uses an encoding similar to JavaScript. Since JSON cannot directly express URIS, JSON-LD is developed (Folmer et al., 2014)

GeoJSON

GeoJSON is also based on JSON, though GeoJSON is a data format especially for spatial data. Point, LineString, Polygon and more are geometry types that are supported. Every GeoJSON object can represent a geometry, or if the geometry has additional properties it is represented as a feature (Butler et al., 2008)

Since Strabon can also return the answer to a GeoSPARQL query in GeoJSON, this format is used to show the spatial data on the map. GeoJSON can be added to a map in combination with Leaflet (see figure 47).

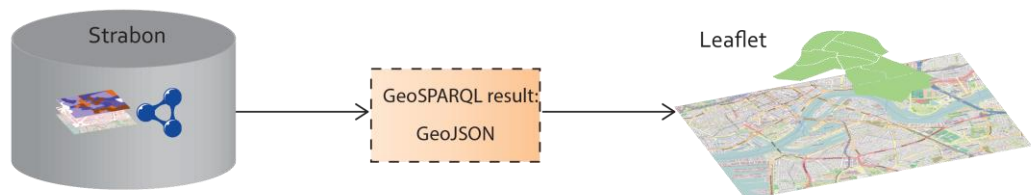


Fig.47. Concept of visualizing the geometry

An example of the GeoJSON format can be seen in the figure below. Here a part of the kindergarten dataset is depicted. The kindergarten dataset exists of several features, which have a CRS and a geometry of type point. This geometry has additional properties, such as the house number. All these features are grouped in a feature collection.

```

{
  "type": "FeatureCollection",
  "features": [
    { "type": "Feature",
      "crs": {
        "type": "name",
        "properties": {
          "name": "EPSG:28992"
        }
      },
      "geometry": {
        "type": "Point",
        "coordinates": [92312.63, 434898.75]
      },
      "properties": {
        "huisnummer": "\"25\"^<http://www.w3.org/2001/XMLSchema#integer\"",
        "type": "Feature",
        "crs": {

```

Example of the geometry of a kindergarten in GeoJSON

The Web Map is made with Leaflet for the map control and using Mapbox for the tile service. The base map is OpenStreetMap. The Web Map looks as follows. see figure 48, in where a part of the Web map can be seen with the spatial data result of one spatial query performed in Strabon.

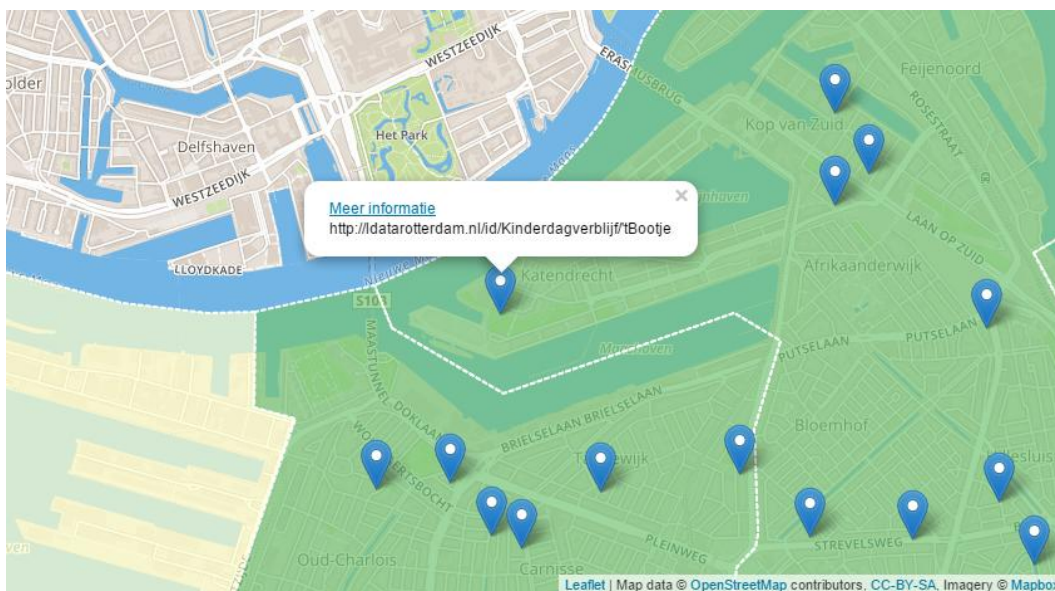


Fig.48. Part of the Web map showing the locations of the kindergartens, the colours indicate the percentage of households with children in where darker green means more households with children)

When clicking on 'Meer informatie' (more information) in the Web page, the following screen is shown (see figure 49). On this screen the properties /predicates

and objects/values are visible. This screen is automatically made with Pubby. If a new triple was added, this would be visible, since the web page shows all relations with this object. A user can now point to individual elements of this object, through their URI.



Fig.49. Information about one kindergarten object, showing the house number and zip code, made with Pubby

To make this kindergarten dataset 5 star data, the data is linked to the webpage from the kindergarten 't Bootje, with the rdfs:seeAlso predicate. This link is made manually. On this web page, additional information such as the telephone number, the facilities and so can be found (see figure 50).

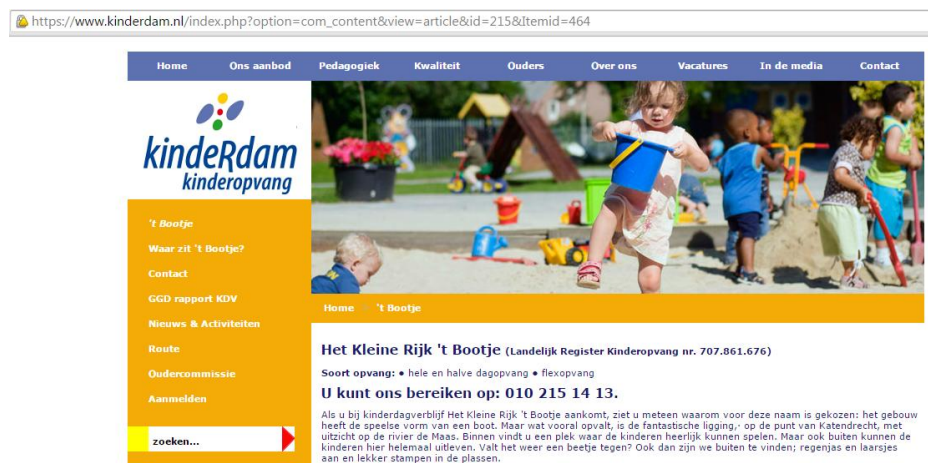


Fig.50. Printscreen of the Web page about kindergarten 't Bootje (kindeRdam kinderopvang, n.d.)

The web map proves also to be useful for layman users, since they can find additional information on the Web and accessing the spatial data as made available by the municipality. They can view the data on the Web, without downloading the data from the Open Data Store and they data seems more understandable with the predicates. If an attribute is not clear, it can be looked up, since it has an URI. A web

application that provided buttons for performing simple describe queries, would provide a flexible interface for layman users and a start for them to browse through.

8.4. INTEGRATION INTO THE EXISTING SYSTEM

To integrate a linked data web application into the existing systems of Rotterdam two simplified options are described in this section. The Web application uses as backend Strabon for the SPARQL endpoint and to access the triples. A conversion from the data from the Rotterdam Open Data Store to RDF data in Strabon should be made. The RDF storage approach, as described in chapter 6, should then be automated. Mapserver, which provides the base map for the GisWeb application, can also be used to provide the base map for the linked data web application (see figure 51).

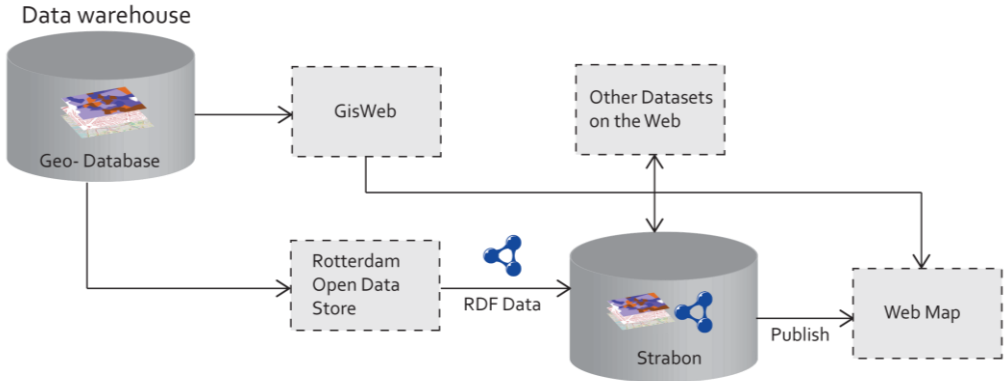


Fig. 51. Implementation of linked open data option 1

Another option is that the data from the data warehouse is directly converted to RDF and published on the Web (see figure 52). Strabon is then still needed for providing the GeoSPARQL endpoint or another endpoint that can do GeoSPARQL queries is needed. For this option, the On-the-fly to RDF option is used to convert the data from the data warehouse and publish it directly to the Web. The GeoSPARQL endpoint can access the published data through that URI and spatial queries can be performed and visualized on a map. This option would be more ideal, since this requires no additional copies of the data.

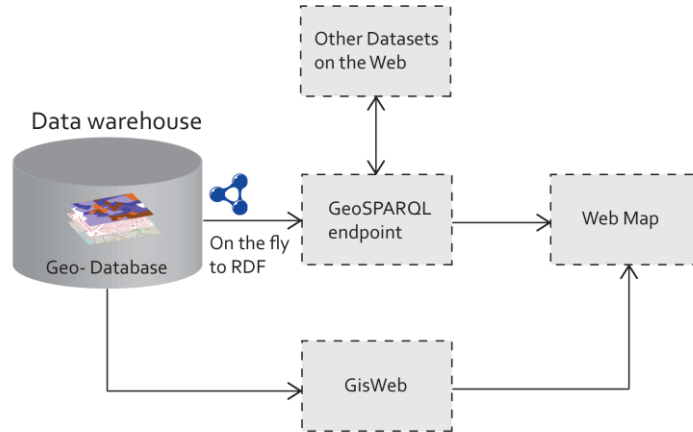


Fig. 52. Implementation of linked open data option 2

8.5. CONCLUSION & DISCUSSION

This chapter proved some insight in how to show spatial linked open data on a map. Two simplified implementations in how the web application can be implemented in the current situation of providing open data at the Municipality were a bit elaborated, though there is not that much research done for this.

The Case study focused on performing an analysis for the National Program Rotterdam Zuid. As demonstrated in the previous chapter, the combining of multiple spatial and non-spatial datasets is possible. Deducing questions from the results and asking follow-up questions on the Web, without making use of a GIS. Also, the spatial linked data can be made visible on a Web map in where the potential of finding additional information is exemplified. A Web application that is able to send GeoSPARQL queries to the triple store and show the result on the screen is not achieved. Though this would be of benefit for the expert users. Providing clickable example queries, with an explanation in what they mean, or buttons that provide the action for a query could help making this option available for layman users. The advantages of linked data that are more for expert users are the combining and querying of the data. Layman users benefit from the sharing of data, the finding of additional information and the better understanding of data.

9. CONCLUSIONS AND RECOMMENDATIONS

9.1. CONCLUSION

The research done for this thesis resulted in a guideline that proved insight in how to create spatial linked open data from existing spatial open data to make use of the advantages of linked data. These advantages benefits expert users and also layman users.

9.1.1. CONCLUSIONS RELATING TO THE SUB QUESTIONS

1. How to publish spatial data as Linked Data on the Web?

This thesis provided a methodology for converting existing spatial open data to spatial linked open data. The methodology consists of the following steps: analyzing and filtering the data. This step consists of understanding and preparing the data in order for it to be transformed. The second steps consist of select or define an ontology. Depending on the data, an ontology to be reused, may exist already. If not, an new ontology needs to be defined, consisting of existing vocabularies. The follow-up step is to create the mapping schema, this can be done by using an existing tool or to write it in a script. Also, a URI strategy is implemented. The next steps consist of converting the data to RDF, making links to another dataset (explicit links) to create 5-star data, implement metadata and publish the data to the Web. The last step is the output, which can exist of the raw data published to the web in HTML/RDF, or serve the data through a web app. In short, the methodology exists of three phases for converting. The methodology can be applied to all sorts of data.

The actual conversion phase is tested in this thesis in two approaches: namely the conversion and explicit storage of RDF data in a triple store and the live conversion of spatial open data in the database to RDF and published on the Web. The on-the-fly method is faster and in theory would be ideal since the data is not copied and this method ensures the up-to-dateness of the data. However, manual adaptation to customize the triples is still needed. The RDF storage approach proved to be slow, but reliable. This method can be automated, but still requires to make a copy of the data.

2. What are options for linking datasets?

This research defined three ways to link datasets, divided into explicit linking or the dynamic/implicit linking which can be done in two ways. The explicit linking requires manual creation of the triples. This is useful in a pre-defined web

application. The implicit linking exists of doing SPARQL queries based on a foreign key or performing a GeoSPARQL query. The SPARQL queries link datasets together on a common element, such as a name or address. The GeoSPARQL queries are spatial queries in where topological relations between two datasets that contain geometry can be asked.

Depending on the nature of the datasets, either method can be chosen, though a prerequisite is to know a bit of the data model of the datasets you want to link.

3. How to query distributed and heterogeneous spatial datasets?

Heterogeneous spatial datasets that contain geometry, can be queried with GeoSPARQL queries. These queries are based on three topological models, namely the RCC8, the 9-IM and the Simple Features Specification. Distributed datasets can be queried with a federated query, which sends a query to another SPARQL endpoint and uses that result to provide an answer. Both of these type of queries require a SPARQL endpoint that is familiar with federated queries and GeoSPARQL queries. For this research the triple store Strabon was used, which can perform federated queries as well as GeoSPARQL queries (not all are implemented yet).

4. How to create a Web application with Linked Data?

A web map was realized in this thesis, which visualized the result of a GeoSPARQL query on the screen. The triple store Strabon can provide an answer to a spatial query in GeoJSON. With Leaflet, the GeoJSON geometry was visualized on the OpenStreetMap by Mapbox. Clicking on an object, produces a pop-up and with Pubby, the additional RDF data was provided on a separate webpage.

In the introduction some challenges regarding literature were mentioned.

Some of these challenges as described in the introduction regarding spatial linked open data are covered in this thesis.

The specifying of the topology is done by modeling the data according to the GeoSPARQL ontology. In this ontology, a feature can have a geometry which is spatially indexed in the triple store Strabon and functions according to three topological models (RCC8, 9-IM and Simple Features Specification) can be asked over the data. Other spatial queries that return a geometry can also be asked over the data, such as a buffer. In Strabon, reverse geocoding, range queries, spatial join, nearest neighbor queries are all possible. Spatial aggregate functions such as buffer and union partially, since sometimes a bug occurs. Transformations are not in the GeoSPARQL ontology and thus cannot be performed.

Linked Data is a method for publishing data on the Web, so that it can be re-used, that is understandable and that it can be easily combined with other datasets as well as that additional information can be found. In my opinion should linked data not be a replacement of a GIS and are transformations therefore not needed.

The conversion of geometry to RDF is also done according to the GeoSPARQL ontology. In this ontology, geometries can be represented in WKT- format or GML.

Important for the storage of geometry in RDF according to GeoSPARQL is the definition of the geometry data type, as well as specifying the CRS. The storage of the numbers of coordinates depends on the backend used. Since for the triple store Strabon is used, with as backend PostgreSQL, the same rules apply.

Metadata is added to the datasets by using the Dublin Core and VOID vocabulary. The Dublin Core vocabulary provides some predicates for spatial data, but not all such as accuracy or specifying a bounding box. This is not further researched in this thesis.

A method to link spatial data with non-spatial data can be done by creating explicit triples or dynamic/implicitly linking on foreign key, as explained at sub question 2. A prerequisite is that a bit of the data model is known to be able to ask the queries.

9.2. REFLECTION CASE STUDY

The municipality of Rotterdam provides multiple spatial open datasets on the Rotterdam Open Data Store with a maximum level of 3 stars according to the 5 star scheme of Tim Berners- Lee. By converting their spatial open data to spatial linked open data (5 stars), they can reach the full potential of open data, namely the sharing, re-using and combining of data. Showing this linked open data on a map with the ability to ask spatial queries, provides the two type of users, namely laymen and expert users, the possibility to access, view and use the data according to their needs.

The expected result for the case study was that spatial linked open data would be of benefit to both expert and layman users. The case study showed that the expert user benefits from spatial linked open data to ask spatial related questions over the data. The layman users benefit from spatial linked open data in clicking on the map and browsing through the data to find additional information. Predefined buttons in a web application that perform spatial queries over the data that is stored in the triple store, would provide layman users to exploit the data even better.

The case study also showed that the Rotterdam Open Data Store sometimes lacks in providing qualitative good data. Datasets that are provided there are sometimes incomplete, the data format tag is most of the time incorrect, licenses are not specified and metadata is missing. This can lead to users losing their interest in the open data.

This research provided insight into the possibilities of spatial linked open data for the municipality of Rotterdam. And it showed that spatial open data from Rotterdam can be converted to spatial linked open data. The research question: What are possibilities and limitations of the use of spatial linked open data for the

municipality of Rotterdam is answered by providing a list of possibilities and limitations as can be seen below.

POSSIBILITIES

- Combining spatial/non-spatial heterogeneous and distributed datasets to discover additional information and help analyze cross knowledge domain issues.

By performing federated queries, different datasets, stored on different servers can be accessed. The possibility of doing federated queries is dependent on the SPARQL query engine used. Possible implementation of Linked Data could be the composition of neighborhoods and areas for the purpose of improving the quality of these neighborhoods and areas, thus for urban planning. The Energie Atlas, which is a project that focuses on showing the potentials of sustainable energy on a map. The risk map, which every municipality needs to make.

- Adding metadata to a spatial dataset

Metadata is written in the same standard as the data and can therefore easily be connected to the dataset and made accessible. There are vocabularies that can be used for adding metadata to a dataset. Provenance tracking improves the quality of the data.

- Querying spatial data on the Web

When spatial linked data that contains geometry is modelled according to the GeoSPARQL ontology, spatial queries on the Web can be performed. With spatial queries on the Web, linking of datasets can be done on topology. Many datasets are linked by location to each other. Linking on topology shows that easier querying of datasets can be done and costs less manual labour.

- Sharing of spatial datasets

When the data is published on the Web (on a map), it can be accessed and queried by anyone, if the URI of the Web map or SPARQL endpoint is known.

LIMITATIONS

- Not a complete support in tools for spatial data/GeoSPARQL

Though, since the development of the GeoSPARQL ontology and the conversion of spatial data to spatial linked data is fairly new, not all tools support this standard yet. An assessment of the tools that are needed and implement the GeoSPARQL ontology is needed, before doing a conversion for 'production' purposes.

- The linked data Web app is not a replacement of an online GIS

Spatial linked Data and GeoSPARQL does not replace an online GIS, since typical GIS functions are not implemented yet in the triple store researched for this thesis. At the moment, linked data can be used to provide an overview or inventory of the data. Spatial linked data can be used to do a simple analysis on multiple different datasets and provides a way to explore. However for more advanced spatial

operations on the data, a database or GIS would be preferred because of the far more extended functionality.

- It can be time consuming to prepare and model the data to linked data. Though this is a onetime activity, the municipality should consider this before starting a pilot.

9.2.1. OBSERVATIONS

One of the observations during this research was that the use of duplicate prefixes for different vocabularies is confusing. By browsing a dataset, the namespaces or used vocabularies should be explicitly stated on the webpage, otherwise it is not clear which vocabulary is used. For example the much used prefix geo is used for the Basic Geo vocabulary, but also for the GeoSPARQL ontology. It can be confusing to know which vocabulary is used for the dataset and to know if spatial queries can be performed on the data or not. A solution for this would be to set predefined names for the prefixes to use for a vocabulary.

Furthermore can the language of the vocabularies be confusing. Linked Data is about re-using existing vocabularies and ontologies. But for Dutch datasets such as the base registrations, it would be logical to have Dutch vocabularies or terms. For users who are not familiar with the data, it would provide a better understanding if the predicates were also in Dutch, since not everyone is familiar with the English language. A label specifying the translation of the much used vocabularies would be beneficial.

The GeoSPARQL vocabulary provides topological functions and functions that return a geometry. In the use case an area consisting of a multipolygon showed up. During the research the polygons were separately linked to the area, to be able to select one another, since it could be a case that an user only wants to select one of the polygons. There needs to be a general understanding of how to deal with multipolygons.

9.3. FUTURE WORK

The research done for this thesis can be extended by researching a few other topics, namely:

- Third approach:

In addition to the approaches in section 4.4.1, a third approach can be added, namely the direct conversion of the WFS from the Rotterdam Open Data Store to RDF and published to the Web. The WFS gives back GML data and since

standardized data proves to be more easily converted, a conversion from GML to RDF can be researched.

- The use of different software/tools

The use of Oracle Spatial & Graph, considering the administration data from the Municipality is also stored in an Oracle database. Oracle Spatial & Graph provides functionality to convert on the fly to RDF. This can be tested on the data from Rotterdam. This could be an effective method to create linked data from the data of Rotterdam.

For the on the fly approach, other tools that provide on the fly conversion and support the R2RML standard, such as Sparqlify could have been tested on use for spatial data. For this thesis, a selection of software-tools is used and conclusions are drawn from them. Preferred is to provide an overview of all the software-tools that can be used for spatial linked open data.

- Create a Web application

A web application that provides the functionality to perform a query over the data from the triple store and show the geometry on the web map would be an ideal situation. More research needs to be done on how to achieve such a dynamic web application.

- Scalability

Scalability is not researched in this thesis, but is an important aspect in linked data. The scalability of the triple store as well as the time it takes for certain queries to be performed on large dumps of RDF data, would provide more content to the research question.

- Specifying spatial metadata

As mentioned, is this not further researched in this thesis. For spatial data is the specifying of the metadata important, to analyze and understand the data correctly. This should be further researched.

9.4. POLICY

STEPS THAT COULD BE TAKEN BY THE MUNICIPALITY

The lifecycle model as described in section 4.6.2 as defined by (Broek et al., 2013) could act as a guideline for the Municipality of Rotterdam to reach the 5 star level to update the data from the Rotterdam Open Data Store, if they wish to do so.

IDENTIFICATION

The vision from the municipality regarding open data can be used here. However, the Municipality can decide if they want to promote the bottom- up approach or the top down approach. With the top down approach, the use of open data can be

structured in a certain direction. For example, certain datasets with the same theme can be made available on the Web. If only datasets of a certain theme are published, innovation is stimulated in that direction. The next step will be selecting which datasets to convert to RDF first. Since not all the data can be converted to RDF at once, it would be easier to select a few test datasets first, for example in a pilot project. Considering linked data principles, the bottom- up approach would be recommended.

PREPARATION

Preparing the data is a minor step, since the requirements set for the publishing of open data can be re-used.

PUBLICATION

The third phase however, the transformation of the data to RDF, is of course very important to generate linked data. According to the 5 star data website, it can become time consuming to convert the data to RDF (in case of 4 stars) for the data publisher (5stardata, 2012). Also, to reach 5 stars, the costs for a data publisher are that it can be time consuming to link your data to other datasets on the Web and that broken or incorrect links need to be taken care of (5stardata, 2012). Therefore, selecting a subset of the data to convert to RDF first can act as a test and is easier to maintain. The in this thesis described flowchart (see figure 13) can act as a starting point of how to convert the data. It would be good to implement the URI-strategy, to keep up with developments in the Netherlands.

When the data is converted to RDF, it can be uploaded to the Open Data portal of Rotterdam and then raise awareness to the linked open data by for example involving social media. Also, another Hackathon can be planned to promote the use of spatial linked open data.

RE-USE

In the re-use step, other parties can be approached to publish their linked open data of Rotterdam on the same portal, such as for example energy companies. Also the managing of the data is of course important. Under managing the linked open data falls to update the data and metadata frequently, improve the data quality by asking feedback from users, keep linking to other datasets and keep track of the downloads of the data to know which datasets are of interest (Broek et al., 2013)

EVALUATION

In this step could be looked at how much the linked open data is used (by looking at the number of downloads of the RDF data or queries) and what sort of users and the applications that are made (Broek et al., 2013). One adaptation can be that the data needs to be presented differently, with more or less functionality in a Web application. Or no Web application at all, but just a RDF browser. Another

adaptation could be that users can upload their own datasets and make links to the municipality data. Conclusions can be drawn if the strategy needs to be adapted.

The publication phase would seem to be the most difficult to achieve. However, this phase also ensures that faults in the dataset may be noticed, for example if addresses are not BAG- conform or that no metadata or use license is specified. This helps improve the quality of the data.

The municipality of Rotterdam plays an important role in innovation regarding (spatial) data. They are the first municipality in the Netherlands that provided a 3D city model as open data. Currently, they are also investigating the use of Big Data with several use cases and a Hackathon. Considering this, a pilot where some test spatial datasets from the open data portal are converted to spatial Linked Open Data can provide the municipality a new opportunity for the municipality of Rotterdam to show its innovative position.

9.5. OPEN RESEARCH QUESTIONS

By performing this research, several new research questions came up, regarding spatial linked open data.

- One of them is, can 3D data also be modelled as linked data and how can that be done? Making links from a 3D city model to other (2D) datasets, could be another interesting use case for spatial linked data.
- In this thesis only vector data is converted. Would raster data be possible to convert to linked data and in for what kind of use cases would this be useful?

10. REFLECTION

The performed research took place in the course of 8 months, starting at November 2014 till the end of June 2015 and was performed at the municipality of Rotterdam. The research is therefore also conducted for the municipality. The subject for this thesis is spatial linked open data. This research focused on how a new technology/method to publish data on the Web (linked data) could be applied to spatial data to benefit from the advantages. This thesis provided a guideline for the municipality of how to convert their open data to linked data and showed the possibilities and limitations of spatial linked open data on an interactive web map to prove insight for the municipality.

For Geomatics the steps to take in a project are generally data capture, data storage, analysis, communication/visualization and quality control regarding all these steps. The methodology used for this research also includes these steps, except for the first step data capture, since the research had as starting point using existing spatial open data and transforming that to spatial linked open data.

The first step to carry out the research consisted of a literature study about the topic, to understand better how linked data worked and what it exactly entails. The next steps focused on choosing the open datasets, analyzing, preparing and converting them to linked data done by two different approaches. The steps thereafter consisted of the querying and thus analyzing the acquired linked data and visualizing the data on a map. The chosen datasets and analysis is done in the context of a case study which entailed the interest of the municipality.

In this thesis the acquired knowledge during the courses of Geomatics is applied, especially knowledge from the courses Python Programming, Geoweb, Geo Database Management Systems, Geo datasets and Quality and GIS and Cartography. During the research new knowledge was learned, since linked data originates from the field of Informatics and thus required some new knowledge.

The planning proved to be feasible. Though a web application that can perform queries was not achieved in the available time, the interactive map which was sufficient enough in showing the acclaimed possibilities of spatial linked open data and exemplifies how the spatial linked open data can be shown on a map.

The process for this research proceeded without any problems, due to sticking to the planning. The small challenges that occurred during the research were mostly solved by using other tools/software.

This research strongly relates to the Geomatics program, since important for working with spatial (open) data, is to show the data on a map, preferable on the Web to be able to share it with others. linked data provides a method to do just that and provides additional advantages to existing methods to show data on the Web. One advantage is the possibility to ask spatially related questions over the spatial data on the Web, while combining the data with other (open) data from the Web. This yet proves to be difficult to be doing with a GIS. Spatial open data on the Web aids in location based services, which is one of the many applications of Geomatics. For location based services, a combination of different datasets is needed, depending on the application the services are used for. Also, in many more cross knowledge domains such as urban planning, crisis and emergency management and sustainability (energy) combining of different datasets is needed. Linked data provides to more easily combine different datasets on the Web. The performed research is therefore not limited to one application of Geomatics, but to all that deal with combining of spatial data on the Web. The research proved some of the acclaimed possibilities of spatial linked data, as well as some of its limitations as of now.

The research could especially be of importance for a municipality, which is a place that deals with the built environment and has as one of its tasks to manage the public space. It therefore deals with a lot of combining of different spatial datasets from different knowledge domains for different issues.

Many applications for spatial linked data arise, such as the risk map and Energie Atlas. Furthermore, Open Data ensures a transparent government and stimulates innovation. The more user friendly, accessible and understandable the open data is provided, thus as linked open data, the more it can be used by others such as companies, organizations and individuals. This can also ensure more participation by citizens, which is more and more sought after by municipalities.

Also by providing open data as spatial linked open data, users can benefit from it, whether they are more expert users which are able to ask queries over the data, or layman users who can discover new things through the combinations of data that are made and visualized on a map.

Performing this research at the municipality led to being the research not only focusing on theoretical applicability of spatial linked data, but provided a case study for spatial linked data. The research got therefore more context and exemplified the conclusions of this research better.

Scientifically this research is relevant, since linked data is a new topic and as literature explains has much potential for spatial data. Literature also indicates that much research is needed for spatial linked data. These challenges were touched upon in this research. In the Netherlands an active community, named Platform Linked Data Nederland, is engaged in dealing with Linked Data and certain

organizations such as Geonovum as well as the company Geodan are also researching spatial Linked Data. For this research, contact with them and others was sought, so as to learn from everyone and to be able to try to contribute new knowledge to this research subject.

REFERENCES

- 5stardata. (2012). 5 star Open Data Retrieved 05-05, 2015, from <http://5stardata.info/>
- Alexander, K., Cyganiak, R., Hausenblas, M., & Zhao, J. (2011). Describing Linked Datasets with the VoID Vocabulary Retrieved 28-04, 2015, from <http://www.w3.org/TR/void/>
- Auer, S., Lehmann, J., & Hellmann, S. (2009). LinkedGeoData - Adding a Spatial Dimension to the Web of Data. Retrieved from http://jens-lehmann.org/files/2009/linkedgeodata_iswc.pdf
- Battle, R., & Kolas, D. (2011a). GeoSPARQL: Enabling a Geospatial Semantic Web. Retrieved from http://www.semantic-web-journal.net/sites/default/files/swj176_0.pdf
- Battle, R., & Kolas, D. (2011b). Linking Geospatial Data with GeoSPARQL. Retrieved from <http://www.semantic-web-journal.net/sites/default/files/swj176.pdf>
- Beckett, D. (2014). RDF 1.1. N-Triples Retrieved 28-04, 2015, from <http://www.w3.org/TR/n-triples/>
- Beckett, D., Berners-Lee, T., Prud'hommeaux, E., & Carothers, G. (2014). RDF 1.1. Turtle Retrieved 28-04, 2015, from <http://www.w3.org/TR/turtle/>
- Berners-Lee, T. (2006). Linked Data Retrieved 28-04, 2015, from <http://www.w3.org/DesignIssues/LinkedData.html>
- Bizer, C., Cyganiak, R., & Heath, T. (2007). How to Publish Linked Data on the Web Retrieved 08-11, 2014, from <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/>
- Brattinga, M., Overbeek, H., & Santema, A. (2014). Best Practices for meaningful connected computing Retrieved 21-01, 2015, from <http://bp4mc2.org/>
- Brickley, D. (2006). W3C Semantic Web interest Group Retrieved 12-04, 2015, from <http://www.w3.org/2003/01/geo/>
- Brickley, D., & Guha, R. V. (2014). RDF Schema 1.1 Retrieved 15-12, 2014, from <http://www.w3.org/TR/rdf-schema/>
- Brink, L. v. d. (2014). Linked Data Pioniers. Retrieved from <http://www.geonovum.nl/documenten/presentaties-linked-data-bgt>
- Broek, T. v. d., Veenstra, A. F. v., & Folmer, E. (2013). Walking the extra byte: a lifecycle model for linked open data Pilot Linked Open Data Nederland Deel 2- De Verdieping (pp. 157-174). Retrieved from <http://www.pilod.nl/wiki/Boek/BroekEtAl.>
- Butler, H., Daly, M., Doyle, A., Gillies, S., Schaub, T., & Schmidt, C. (2008). The GeoJSON Format Specification Retrieved 04-05, 2015, from <http://geojson.org/geojson-spec.html>
- CBS. (2015). Wijk- en Buurtkaart 2014 Retrieved 28-04, 2015, from <http://www.cbs.nl/nl-NL/menu/themas/dossiers/nederland-regionaal/publicaties/geografische-data/archief/2015/wijk-en-buurkaart-2014-art.htm>
- Cyganiak, R., Bizer, C., Garbers, J., Maresch, O., & Becker, C. (2012). The D2RQ Mapping Language Retrieved 04-05, 2015, from <http://d2rq.org/d2rq-language>

- Cygniak, R., Wood, D., & Lanthaler, M. (2014). RDF 1.1. Concepts and Abstract Syntax Retrieved 05-05, 2015, from <http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/>
- Das, S., Sundara, S., & Cygniak, R. (2012). R2RML: RDB to RDF Mapping Language Retrieved 28-04, 2015, from <http://www.w3.org/TR/r2rml/>
- data.gov.uk. (n.d.). The INSPIRE Directive. Retrieved from http://data.gov.uk/sites/default/files/Factsheet%20-%20INSPIRE%20-%20Directivepdf_10.pdf
- DBpedia. (2015). Waalhaven (Rotterdam) Retrieved 14-06, 2015, from [http://nl.dbpedia.org/page/Waalhaven_\(Rotterdam\)](http://nl.dbpedia.org/page/Waalhaven_(Rotterdam))
- DEN. (2014). What is the DEN Foundation? Retrieved 04-05, 2015, from <http://www.den.nl/english>
- Digitale overheid. (n.d.). Stelsel van Basisregistraties Retrieved 28-4, 2015, from <http://www.digitaleoverheid.nl/onderwerpen/stelselinformatiepunt/stelsel-van-basisregistraties>
- E-overheid. (n.d.-a). Aanduiding nevenadressen Verblijfsobject Retrieved 14-01, 2015, from http://www.e-overheid.nl/onderwerpen/stelselinformatiepunt/stelsel-van-basisregistraties/stelselvoorzieningen/stelselcatalogus/begrippen/Adres/BAG/Verblijfsobject/Aanduiding_nevenadressen_Verblijfsobject-Verblijfsobject
- E-overheid. (n.d.-b). Stelselcatalogus Retrieved 14-01, 2015, from <http://www.e-overheid.nl/onderwerpen/stelselinformatiepunt/stelsel-van-basisregistraties/stelselvoorzieningen/stelselcatalogus>
- Emmons, I. (2014). Parliament User Guide Retrieved from <http://parliament.semwebcentral.org/ParliamentUserGuide.pdf>
- EU ISA Programme Core Vocabularies Working Group (Location Task Force). (2013). ISA Programme Location Core Vocabulary Retrieved 06-01, 2015, from <http://www.w3.org/ns/locn>
- Folmer, E., Reuvers, M., Quak, W., & Brandt, P. (Eds.). (2014). *Pilot Linked Open Data Nederland* Amersfoort: Klomp Bizzprint.
- Franz Inc. (2015a). AllegroGraph Retrieved 18-04, 2015, from <http://franz.com/agraph/allegrograph/>
- Franz Inc. (2015b). AllegroGraph 5.0.1 N-dimensional Geospatial Facility: Introduction and Overview Retrieved 18-04, 2015, from <http://franz.com/agraph/support/documentation/current/geospatial-nd.html>
- Franz Inc. (2015c). Gruff: A Grapher- Based Triple Store Browser for AllegroGraph Retrieved 18-04, 2015, from <http://franz.com/agraph/gruff/>
- Franz Inc. (2015d). SPARQL Magic Properties Retrieved 18-04, 2015, from <http://franz.com/agraph/support/documentation/current/magic-properties.html>
- Gandon, F., & Schreiber, G. (2014). RDF 1.1. XML Syntax Retrieved 28-04, 2015, from <http://www.w3.org/TR/rdf-syntax-grammar/>
- Gemeente Rotterdam. (n.d.-a). nationaal programma rotterdam zuid Retrieved 15-12, 2014, from <http://www.rotterdam.nl/nprz>
- Gemeente Rotterdam. (n.d.-b). open data Retrieved 01-02, 2015, from <http://www.rotterdam.nl/opendata>
- Geonovum. (2015). W3C en OGC werken aan integratie GIS en Web Retrieved 29-01, 2015, from <http://www.geonovum.nl/nieuws/w3c-en-ogc-werken-aan-integratie-gis-en-web>
- Gillies, S. (2013). Intro geospatial linked data Retrieved 05-06, 2015, from <http://sgillies.github.io/lawdi-2013-intro-geospatial-linked-data/#/2>

- GIS-diensten. (n.d.). Gisweb 2.1 Retrieved 05-02, 2015, from <http://www.gis.rotterdam.nl/gisweb2/default.aspx>
- Goodwin, J., Hart, G., & Dolbear, C. (2009). Geographical Linked Data: The Administrative Geography of Great Britain on the Semantic Web. *Transactions in GIS*, 12(Supplement s1), 19-30.
- Harth, A., & Gil, Y. (2014). Geospatial Data Integration with Linked Data and Provenance Tracking. Retrieved from http://www.w3.org/2014/03/igd/papers/igd14_submission_54
- Hartig, O., & Langegger, A. (2010). A Database Perspective on Consuming Linked Data on the Web. Retrieved from <http://link.springer.com/article/10.1007%2Fs13222-010-0021-7>
- Hassanzadeh, O. (2011). Publishing Relational Databases as Linked Data. Retrieved from <http://www.cs.toronto.edu/~oktie/slides/publishing-relational-databases-as-linked-data.pdf>
- Hyland, B., & Villazón Terrazas, B. (2011). Linked Data Cookbook Retrieved 14-06, 2015, from http://www.w3.org/2011/gld/wiki/Linked_Data_Cookbook
- Kadaster. (n.d.). Wat is de BAG? Retrieved 14-01, 2015, from <http://www.kadaster.nl/web/Themas/Registraties/BAG/BAGartikelen/Wat-is-de-BAG.htm>
- Keijser, R., personal communication, 2015
- kinderDdam kinderopvang. (n.d.). Het Kleine Rijk 't Bootje Retrieved 14-06, 2015, from https://www.kinderdam.nl/index.php?option=com_content&view=article&id=215&Itemid=464
- Knibbe, F. (2014a). Basisregistratie Adressen & Gebouwen (BAG) volgens principes van Linked Data | inleiding Retrieved 04-11, 2014, from <http://www.geodan.nl/basisregistratie-adressen-gebouwen-bag-vijfsterrendataset/>
- Knibbe, F. (2014b). Linked Data and geoinformatics - a gap analysis. Retrieved from http://www.w3.org/2014/03/igd/papers/igd14_submission_4
- Koubarakis, M., Karpathiotakis, M., Kyzirakos, K., Nikolaou, C., & Sioutis, M. (2012). Data Models and Query Languages for Linked Geospatial Data. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.352.2565&rep=rep1&type=pdf#page=298>
- Koubarakis, M., Karpathiotakis, M., Kyzirakos, K., Nikolaou, C., Vassos, S., Garbis, G., . . . Kersten, M. (2012). Building Virtual Earth Observatories Using Ontologies and Linked Geospatial Data. *Web Reasoning and Rule Systems*, 229-233.
- Kuhn, W., Kauppinen, T., & Janowicz, K. (2014). Linked Data- A Paradigm Shift for Geographic Information Science. *Geographic Information Science*, 173-186.
- Lieberman, J., Singh, R., & Goad, C. (2007). W3C Geospatial Vocabulary Retrieved 12-04, 2015, from <http://www.w3.org/2005/Incubator/geo/XGR-geo-20071023/>
- Linked Open Vocabulary. (n.d.). Linked Open Vocabularies Retrieved 08-12, 2014, from <http://lov.okfn.org/dataset/lov/>
- LinkedDataTools.com. (n.d.-a). Tutorial 3: Semantic Modeling Retrieved 11-12, 2014, from <http://www.linkeddatatools.com/semantic-modeling>

- LinkedDataTools.com. (n.d.-b). Tutorial 4: Introducing RDFS & OWL
Retrieved 11-12, 2014, from <http://www.linkeddatatools.com/introducing-rdfs-owl>
- McGuinness, D. L., & Harmelen van, F. (2004). OWL Web Ontology Language Overview Retrieved 22-12, 2014, from <http://www.w3.org/TR/owl-features/>
- Metadata Innovation Dublin Core Metadata Initiative. (n.d.). Background Retrieved 08-12, 2014, from <http://dublincore.org/metadata-basics/>
- Nationaal Georegister. (n.d.). Nationaal Georegister Retrieved 15-12, 2014, from <http://www.nationaalgeoregister.nl/geonetwork/srv/dut/search>
- NLExtract. (n.d.). Project Retrieved 15-12, 2014, from <http://www.nlextract.nl/project-definition>
- Norton, B., Vilches, L. M., León, A. D., Goodwin, J., Stadler, C., Anand, S., . . . Atemezing, G. A. (2012). NeoGeo Vocabulary Specification - Madrid Edition Retrieved 12-04, 2015, from <http://geovocab.org/doc/neogeo/>
- Openlink Software. (2015). Home Retrieved 18-04, 2015, from <http://virtuoso.openlinksw.com/>
- Openlink Software. (n.d.-a). 9.34. Geometry Data types and Spatial Index Support Retrieved 2015, 18-04, from <http://docs.openlinksw.com/virtuoso/sqlrefgeospatial.html>
- Openlink Software. (n.d.-b). 16.15. RDF and Geometry Retrieved 18-04, 2015, from <http://docs.openlinksw.com/virtuoso/rdfsparqlgeospat.html>
- Oracle. (n.d.). Oracle Spatial and Graph RDF Semantic Graph Retrieved 02-02-2015, 2015, from <http://www.oracle.com/technetwork/database/options/spatialandgraph/overview/rdfsemantic-graph-1902016.html>
- Ordnance Survey. (2015). SO16 0AS Retrieved 15-01, 2015, from <http://data.ordnancesurvey.co.uk/doc/postcodeunit/SO160AS>
- Ordnance Survey. (n.d.). The linked data web Retrieved 04-11, 2014, from <http://www.ordnancesurvey.co.uk/education-research/research/linked-data-web.html>
- Overbeek, H., & Brink van den, L. (2013). Aanzet tot een nationale URI-Strategie voor Linked Data van de Nederlandse overheid. Retrieved from <http://www.pilod.nl/wiki/Boek/URI-strategie>
- Parliament. (n.d.). Parliament A High- Performance Triple Store, SPARQL Endpoint, and Reasoner Retrieved 18-04, 2015, from <http://parliament.semwebcentral.org/>
- Perry, M., & Herring, J. (2012). OGC GeoSPARQL- A Geographic Query Language for RDF Data. Retrieved from <http://www.opengis.net/doc/IS/geosparql/1.0>
- Platform Linked Data Nederland. (2014). Linked Data in beeld. Retrieved from <http://www.geonovum.nl/sites/default/files/Linked%20Data%20in%20beeld-2014-Locked.pdf>
- Platform Linked Data Nederland. (2015). Aanpak Basisregistraties Retrieved 14-06, 2015, from http://www.pilod.nl/wiki/Aanpak_Basisregistraties
- Platform Linked Data Nederland. (nd.). Over Platform Linked Data Nederland Retrieved 26-01, 2015, from http://www.pilod.nl/wiki/Over_Platform_Linked_Data_Nederland
- Prud'hommeaux, E., & Buil-Aranda, C. (2013). SPARQL 1.1 Federated Query Retrieved 05-05, 2015, from <http://www.w3.org/TR/2013/REC-sparql11-federated-query-20130321/>

- Prud'hommeaux, E., & Seaborne, A. (2008). SPARQL Query Language for RDF Retrieved 28-04, 2015, from <http://www.w3.org/TR/rdf-sparql-query/>
- Rotterdam Open Data Store. (n.d.). Wie zijn wij? Retrieved 28-04, 2015, from <http://www.rotterdamopendata.org/faq>
- Saquicela, V., Espinoza, M., Piedra, N., & Villazón-Terrazas, B. (2014). Ecuadorian Geospatial Linked Data. Retrieved from schema.org. (n.d.-a). Thing>Place Retrieved 12-04, 2015, from <https://schema.org/>
- schema.org. (n.d.-b). What is Schema.org? Retrieved 12-04, 2015, from <https://schema.org/>
- Sesame. (2014). Querying a repository Retrieved 28-04, 2015, from <http://rdf4j.org/sesame/2.7/docs/articles/repository-api/querying.docbook?view> .
- Steen, M., Eckartz, S., & Folmer, E. (2014). Linking Open Energy Data. Retrieved from http://www.pilod.nl/w/images/4/4f/20141126_Ordina_Steen.pdf
- Strabon. (n.d.). Strabon Retrieved 05-05, 2015, from <http://www.strabon.di.uoa.gr/about>
- Teleios project. (n.d.). Strabon Retrieved 05-05, 2015, from <http://www.earthobservatory.eu/Strabon>
- Veer van 't, R. (2014a). Position Paper Heritage & Location Towards a geotemporal semantic infrastructure for Dutch cultural historical data. Retrieved from http://www.w3.org/2014/03/lgd/papers/lgd14_submission_24
- Veer van 't, R. (2014b). Technische toelichting GeoSPARQL- demo Retrieved 03-05, 2015, from <http://erfgoedenlocatie.nl/pilots/geosparql-demo/technische-toelichting-geosparql-demo/>
- Vilches-Blázquez, L. M., Terrazas, B. V.-., Leon, A. D., Priyatna, F., & Corcho, O. (2010). An Approach to Publish Spatial Data on the Web: The GeoLinked Data Case. Retrieved from <http://stko.geog.ucsb.edu/lstd2010/paper%207.pdf>
- W3C. (2013a). Linked Data Retrieved 07-11, 2014, from <http://www.w3.org/standards/semanticweb/data>
- W3C. (2013b). W3C Semantic Web Activity Retrieved 08-11, 2014, from <http://www.w3.org/2001/sw/>
- W3C. (2014). Call for Participation Retrieved 08-11, 2014, from <http://www.w3.org/2014/03/lgd/>
- W3C Semantic Web. (2012). Introduction to SKOS Retrieved 08-01, 2015, from <http://www.w3.org/2004/02/skos/intro>
- Waag Society. (n.d.-a). CitySDK Linked Data API Retrieved 13-06, 2015, from <http://citysdk.waag.org/>
- Waag Society. (n.d.-b). FAQ Retrieved 13-06, 2015, from <http://citysdk.waag.org/about/>
- Weber, R., personal communication, 2015
- Wikipedia. (2013). Region connection calculus Retrieved 05-05, 2015, from http://en.wikipedia.org/wiki/Region_connection_calculus
- Wikipedia. (2015). DE-9IM Retrieved 05-05, 2015, from <http://en.wikipedia.org/wiki/DE-9IM>
- Williams, H. (2014). OpenLink Virtuoso (Open- Source Edition) Retrieved 18-04, 2015, from <http://ehc.ac/p/virtuoso/mailman/virtuoso-users/thread/00525FFD-C4FC-46AD-AA9E-A031B18A950B@openlinksw.com/>

Wood, D. (2011). Intro to Linked Data: Data Modeling. Retrieved from <http://www.slideshare.net/prototipo/intro-to-linked-data-data-modeling>

APPENDIX 1: USED PREFIXES

The used prefixes and namespaces are listed in the tables 21 and 22.

Prefix	Namespace
dbpedia:	http://nl.dbpedia.org/resource/
dcterms	http://purl.org/dc/terms/
geo	http://www.opengis.net/ont/geosparql#
geof	http://www.opengis.net/def/function/geosparql/
owl	http://www.w3.org/2002/07/owl#
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs	http://www.w3.org/2000/01/rdf-schema#
skos	http://www.w3.org/2004/02/skos/core#
units	http://www.opengis.net/def/uom/OGC/1.0/
void	http://rdfs.org/ns/void#
xsd	http://www.w3.org/2001/XMLSchema#

Table. 21. Used Prefixes and namespaces of existing vocabularies

Prefix	Namespace
buurt	http://ldatarotterdam.nl/id/Buurt/
gebied	http://ldatarotterdam.nl/id/Gebied/
kv	http://ldatarotterdam.nl/id/Kinderdagverblijf/
pand	http://ldatarotterdam.nl/id/Pand/
vo	http://ldatarotterdam.nl/id/VO/
wijk	http://ldatarotterdam.nl/id/Wijk/

Table.22. Used Prefixes and defined namespaces for converted data