DEVELOPING A LINKED DATA FRAMEWORK FOR THE EXCHANGE OF AS-IS RAILWAY ASSET DESIGN AND MAINTENANCE INFORMATION

Summary MSc thesis

The current methods used for the management of railway asset design and maintenance data in the Dutch railway domain have led to difficulties in the retrieval of required information. Railway asset data is scattered amongst the many stakeholders involved in the asset management processes and the data is often difficult to retrieve due to a lack of data standardization. Each of these stakeholders within the domain, such as clients and maintenance contractors, create and maintain data that is relevant to their work processes. However, for a holistic approach to the management of assets, these stakeholders could benefit from the integration of the various datasets available.

These issues have become even more apparent with the recent start of the implementation of the European Railway Traffic Management System (ERTMS). This attempt at unifying the traffic management systems in use in the European Union also requires data restructuring. The design and implementation of this system depends on the current railway track layout and requires the data that is currently difficult to retrieve. Integration of the various datasets could help ease the retrieval of required information. However, data integration does not always offer a definitive solution, since the used schemas could change when no standards are established. Moreover, data integration can be difficult if there are no relationships known between the various datasets. Therefore, a better method for the sharing of information is required.

A substantial amount of research has been done on what the best practises are for asset information management throughout the lifecycle stages. Many standards have been created that could allow for improved interoperability between the many stakeholders involved within an established domain. However, standards are created with domain specific stakeholders in mind, limiting its usability when incorporating cross domain data. Currently, no standard exists that fully encompasses all of the railway design and maintenance data available. Moreover, the usage of standards does not necessarily provide a holistic overview of the available data when working with decentralised datasets.

An opportunity for large scale data integration can be found in the application of linked data. The usage of linked data could provide more than a one-off solution to data integration and could allow for improvements in data retrieval when working with decentralised data by making use of web-based applications. The usage of linked data requires a semantic web ontology that could be used to semantically relate the data available. However, currently no such ontology exists for railway asset design and maintenance data, impeding implementation. In this study a semantic web ontology was developed and tested in order to determine if this could facilitate large scale data integration of the non-standardized and decentralised data from the Dutch railway domain.

Together with the stakeholders involved in this research, an assessment was made of the needs that users have for data integration and the three main applications of a semantic web ontology were formulated (use cases): (1) identification of design and maintenance information for railway assets, (2) identification of the interface with subsurface infrastructure and railway assets and (3) assessment of the validity of the data for railway assets. From these three use cases the author derived several functional requirements in the form of competency questions, which were used to develop a conceptual ontology. In this ontology, functionalities from Web Ontology Language were incorporated, which could assist in data integration of the available datasets.

After the development of the conceptual ontology, the author implemented the ontology using three datasets from the Dutch railway domain. A dataset from a maintenance contractor containing design and maintenance data from a track segment was directly mapped out using the developed ontology towards Resource Description Framework (RDF). Moreover, railway assets instances from another RDF dataset were semantically related to the developed ontology, while making use of various object breakdown structures and properties. By using SPARQL queries, it was shown that it is possible to retrieve all relevant information for specific railway assets. The outcome of this study was a verified and tested semantic web ontology for railway asset design and maintenance data which was successfully used to facilitate large scale data integration.

DEVELOPING A LINKED DATA FRAMEWORK FOR THE EXCHANGE OF AS-IS RAILWAY ASSET DESIGN AND MAINTENANCE INFORMATION

Rick Akster

Construction Management & Engineering, Faculty of Engineering Technology

Submission date: May 29, 2023 (Version 1.0)

Within the Dutch railway sector, railway asset data is scattered and often difficult to retrieve. This is caused by a lack of standardization and the usage of decentralized storage of data. Integration of the various datasets could help ease the retrieval of relevant information for the construction and maintenance processes. However, data integration is often considered difficult and not a definitive solution. Therefore, a better method for the sharing of information is required. According to the literature, linked data could provide more than a one-off solution to data integration and could allow for improved methods in data retrieval when working with decentralized data. However, data integration using linked data in the railway design and maintenance domain is not well explored. The usage of linked data requires a semantic web ontology that could be used to semantically relate the data available. Currently no such ontology exists for railway asset design and maintenance data, impeding implementation. In this study an ontology using Web Ontology Language was developed for the railway asset design and maintenance domain and afterwards implemented for the Dutch railway domain. In this study it was shown that by using a semantic web ontology large scale data integration is possible.

Keywords: Linked data; Semantic Web Ontology; Information management; Railway Assets; Data Integration

---

1. Introduction

Many of the member states of the European Union developed their train networks individually, which has led to a heterogeneous network of railway infrastructure throughout Europe. Part of this heterogeneity originates from the various safety systems in use, which causes challenges for international railway travel to operate more safely and efficiently. As part of an integration process, the European Union has set the goal of standardizing the different safety systems in use by implementing a uniform system across all its member states [1].

This new system, the European Railway Traffic Management System (ERTMS), will be implemented along the highest intensity track segments in the Netherlands over the coming 10 years [2]. However, due to several reasons the end date for implementation has become uncertain [3]. One of the issues experienced is the retrieval of relevant asset information necessary for the design of the ERTMS track layout and for the execution of this design. Moreover, maintenance data is required to determine the actual state of the existing infrastructure that interfaces with the ERTMS design.

Issues arise when trying to retrieve information that is relevant to the ERTMS implementation process. Currently, the relevant data is managed by the client ProRail and various third parties such as maintenance contractors. Since each of these parties manages their own data, the datasets cannot be easily integrated or standardized by other parties. Asset data is often exchanged on a request basis and afterwards integrated with self-managed data. The data is stored using various schemas and data standards, and this data is often stored without any direct relationship to the datasets of other parties. Improvements in asset information management are required.

A substantial amount of research has been conducted on what the best practices are for the information management of assets throughout their lifecycles and how the exchange of this information can be standardized to allow for better interoperability between the information systems used [4], [5]. In the construction industry, data is often exchanged by making use of information systems such as BIM [6], which are also seen as a capable tool for sharing asset data [7]. To increase interoperability, data is often shared using open standards such as IFC [8]. However, BIM is often used in domains it was not originally meant for [9] and open standards are often created with domain-specific stakeholders in mind, limiting its usability in other situations.

Because of this limitation, this study explored alternative means for the management of asset information that would allow for increased interoperability and the efficient sharing of data amongst multiple parties. An opportunity for how the various silos of information can be connected has been identified in various literature [10]–[13], which is the usage of linked data principles. Linked data could assist in the communication of the relevant data without disruptions to the currently existing databases [13]. Moreover, linked data could provide a solution to the sharing of data when multiple schemas are used.

In order to implement linked data, a semantic web ontology is required that can contain the relevant railway asset data. Moreover, such an ontology should be used to map the currently existing data towards linked data standards. Currently, no semantic web ontology exists that is comprehensive enough to contain this data, impeding implementation. Moreover, the various datasets are currently managed by multiple parties without any coherency. It is currently unknown how these datasets could be semantically related to each other.

During this research, the author explored these issues by creating and testing an ontology that is based on semantic web standards. The main objective for this research was as follows:

"To create a linked data framework for the exchange of railway asset design and maintenance data that is capable of semantically relating data originating from various sources."

The academic contribution of this research is a semantic web ontology specific for the railway assets design and maintenance domain, which was shown to be capable of large-scale data integration using noncentralized data.

The ontology was developed in two parts by the author. Firstly, a conceptual ontology was developed for railway asset design and maintenance data that was created based on a number of competency questions. The intention of this conceptual ontology is for broader reuse and further development within the railway asset design and maintenance domain. Secondly, this conceptual ontology was implemented so that it could be directly used by the Dutch railway domain and it was tested if the ontology was capable of integrating a number of datasets. This implemented ontology was designed based on the existing brownfield environment with incomplete and error-filled data within the Dutch railway domain.

The implemented ontology was designed, modelled, implemented and tested using a synthetic and real dataset. This allowed for the distinguishing of shortcomings that originates from the ontology or from the real datasets. The author highlighted several issues that could occur when using a semantic web ontology for data integration and showed how this was solved in the ontology. The combination of the ontology, dataset and linked data triple store formed the linked data framework that was used to improve the management of information within the railway domain.

The created framework formed the connection between the various datasets and allowed for easier retrieval of all relevant information regarding ERTMS related assets. This was shown by making use of a case study, in which data could be efficiently retrieved using several SPARQL queries.

In section 2 of this research the theoretical background is given, where linked data and currently existing ontologies are further explored. In section 3, the methodology is presented that guided this research. Moreover, the approach for the development of the conceptual and implemented ontology are presented. In section 4, the use cases of the ontology and a number of derived competency questions. In section 5, the developed conceptual ontology is presented. In section 6, the developed conceptual ontology was implemented and tested with a case study. The presented ontology and the outcomes of this research were validated 7. In sections 8 and 9 the discussion and conclusion can be found.

2. Theoretical background

The theoretical background explores four topics that are central to this research: (1) the use case of linked data as a means of overcoming interoperability issues and increasing information sharing, (2) the developments of ontologies using linked data, (3) the application of linked data in the creation of knowledge graphs and (4) the modelling of geospatial semantics.

2.1. Large scale data integration

During the life cycle stages of an asset, information is created and afterwards utilized by the various stakeholders involved. As stated by Curry et al. [13], It is important to utilize all available information that is stored across various databases to manage assets holistically. Data that is created over the life cycle stages are often used and stored in a way that is specific to the relevant domain [12]. To be able to manage the data that is created, various information management systems are used that are designed based on the needs and skills of the various stakeholders that are part of the construction processes [13].

The usage of different information management systems has led to the creation of various 'silos' of information, which are not well integrated with other systems. The currently used methods of data storage and communication often fail in the integration of data that is not specific to the construction industry [12]. These issues can also be found within the railway domain, where systems such as BIM are insufficient to manage the wide range of data used. A need exists to better utilize cross-domain data [13], with one of the main barriers being interoperability issues.

As stated by [14, p. 2], "*Interoperability refers to the ability of a technology to exchange information, communicate and cooperate with other systems without major modification of their structure*". These can emerge when different standards and software are used in relation to those of other stakeholders. Since open standards such as IFC are often not sufficient for the exchange of information outside of the domain specific to its development [13]. A different method of sharing relevant railway asset information is needed, where the usage of various data schemas are not limiting interoperability.

One of the main conclusions regarding interoperability issues in the construction industry is that, despite the many improvements being made with open standards, more efficient methods of integrating data are needed to prevent the forming of data silos [11]–[13], [15]. The forming of these 'silos' can mainly be explained by the lack of interoperability between the various standards in use.

It is possible to integrate multiple 'silos' of data that have been created based on different standards. However, this will always be a one-time solution to interoperability. Data integration is done on a case-to-case basis and does not serve as a sound basis for integration in future cases. Semantic web technologies such as Linked Data could offer a solution to this.

Semantic web technologies are mainly applied for improvements in data sharing, discovery, integration and reuse [16]. Linked data offers the opportunity for data integration by making use of well-connected data graphs [13], effectively connecting the storage of related data from the currently existing silos [17]. As stated by Curry et al. [13], linked data can be seen as a best practice in the sharing of different datasets that are based on web standards. Linked data could offer a solution to interoperability issues if multiple standards are in use, since it is a method of exchanging data without the usage of schemas. Linked data can thus also provide a solution to the data retrieval issues in the Dutch railway sector.

Linked data is about the creation of links between datasets, where every link is built up using a triple statement. Linked Data is built upon the following four principles [17]: (1) things are named by making use of a Uniform Resource Identifier (URI), (2) names are looked up by making use of HTTP URI's, (3) useful information should be provided when the URI's are looked up by making use of the standards and (4) when a URI is retrieved, links to other URI's should be provided.

Linked data allows for the sharing and merging of information originating from cross-domain silos of information [13]. As stated by [13, p. 3], "*Linked data offers a method of exposing, sharing and connecting data in a manner that is reusable and not a one-off integration solution*". Linked data has the potential of merging heterogeneous sources of data [16], mitigating the need for one standardized data schema.

Adoption of linked data could bring improvements in the communication of relevant data, as it removes the need to integrate data on a case-to-case basis and it can improve the data retrieval process as more accurate data can be presented [11]. Therefore, it could potentially be beneficial to the Dutch railway sector, in which the data retrieval process is currently sluggish.

2.2. Linked data

While making use of linked data, data can be merged based on 'directly linked data' or based on 'ontology linked data' [10]. Directly linked data makes use of matching of the URI's residing within the various RDF datasets in order to expand the available knowledge graph. Ontology linked data makes use of matching of the semantics to expand the knowledge graph. These types of data merging allow for the full usage of the available data, without the need for it to be adjusted or stored centrally. Linked data can be expanded by importing other datasets which are open and converted towards RDF, and afterwards by making connections between these datasets.

The research by Quinn et al. [10] and Corry et al. [12] showed how linked data could be used for the integration of data originating from the AEC-industry with other cross-domain data. It was shown that by exporting data stored in a digital model and converting this to RDF, semantic relations could be made with other RDF based data. Corry et al. [12] also showed that converting the datasets to RDF enabled the working with public data already in RDF format. By making use of RDF the various silos of data were efficiently matched and merged, allowing for a more holistic means of data management.

A similar approach was taken in the study of Lee et al. [11], in which data from an IFC file was converted towards linked data and combined with a self-developed defect-ontology. This allowed for improvements in detection of defects and statistical analysis of the models. The research by Quinn et al. [10], Corry et al. [12] and Lee et al. [11] did not reach semantic interoperability with the source of their data, which was the digital model. Converting the data towards linked data offered many opportunities and improvements in these studies, but it was not possible to convert back to the original data. Currently, no full ontologies exist that could be used in this research that could be used to create digital models. Therefore, further developments for ontologies specific to design data are required.

Just like [12], [11] and [10], Curry et al. [13] proposed a method for which digital models could be converted from IFC to RDF. After converting the digital models to RDF, non-AEC-domain data was linked. Different from these studies, is that Curry et al. [13] created a standard manner in which data was exported based on matching attributes, and the data was then exported towards other applications that allowed for an improved holistic management of assets. Therefore, showing that data originally residing within a digital model could also be managed without the existence of this digital model, where data is exported using standard methods. Therefore, non-AEC domain standards such as IFC can also be used in the developments of an ontology for design data.

2.3. Geospatial semantics

For the creation of the ontology for railway asset design and maintenance data, geospatial semantics are a requirement. As stated in section 2.2, no full ontology exists that is a two-way communication between digital models and linked data. Similar to the modelling of geospatial semantics, digital models were used as a one-way communication towards RDF. The ontology of GeoSPARQL and its incorporation of Geography Markup Language (GML) and Well Known Text (WKT) will be reused in this research. These ontologies will allow for relating the geospatial position of the various objects. Moreover, geospatial querying can be used in combination with the directly linked data method to relate the various instances from the various datasets.

The research by Karan & Irizarry [15] proposed a method in which semantic interoperability could be reached between two systems often used in the AEC-domain, namely GIS and BIM. The study showed how linked data could assist in the combining of the two datasets in RDF, which could then be converted back towards the standards used in the specific software. Using this method the main advantages of BIM and from GIS were combined, while having the advantage of using the original software applications. It was shown that geospatial semantics could be combined with design data without making adjustments to the original RDF design dataset.

Another approach can be found in the research by Stepien et al. [18]. BIM and GIS datasets were both converted to RDF and then by making use of GeoSPARQL, geospatial queries were created. GeoSPARQL is a RDF query engine specifically for geospatial information, allowing for semantic interoperability between linked data and geospatial reasoning. The study shows the possibilities of converting geospatial information to linked data. The usage of GeoSPARQL could also be used in the design of the railway asset design and maintenance ontology, as it can be used to relate the various instances from the various datasets together by making use of the directly linked data method.

This approach can also be found in the study by Ranjgar et al. [19], where ontologies were combined with the ontology of GeoSPARQL and Well Known Text (WKT) serialisation of geospatial data, in order to perform geospatial queries. One of the major benefits of this approach is that it is not needed to create data from a GIS environment, but that geospatial information can be directly linked to an ontology and thus also the design data. Therefore, no third party applications are required in the design of the ontology of this research and most requirements can be met by using RDF.

2.4. Ontology development

The usage of metadata in the semantic web can be done by making use of domain-specific ontologies, which allow for logic-based data interconnection [16]. As stated by [20, p. 1], an ontology is a "*formal, explicit specification of a shared conceptualization*". An ontology can be used to represent domain knowledge by modelling concepts, and identifying relationships, and can assist in the sharing and reusing of information that is stored in unstructured formats [11].

Since linked data is built upon shared understanding and common vocabularies, the integration of data is done at the information level instead of at the system level [13]. Therefore it is important that the used ontologies can be reused by others [16]. The engineering process of an ontology should be done using a systematic approach to ensure the quality and usability of the final product. Important to this is the involvement of domain experts, creators of the ontology and end-users [21].

As presented in sections 2.2 and 2.3, several ontologies have been developed for the communication of design and maintenance data for assets within the construction domain. However, no such ontologies are currently available for the railway sector that also make use of semantic web standards. For the geospatial information of the railway assets, multiple ontologies are readily available [22], [23]. These ontologies are generic and can be used in any kind of ontology that needs to include geospatial semantics. These ontologies were reused and incorporated in the ontology for railway design and maintenance data. These ontologies make use of open standards which are common, making it beneficial to be reused.

The development and publishing of ontologies using semantic web standards allow for nearly seamless integration with other ontologies and data. Reusability is a big part of semantic web principles and is actively recommended. The goal of this study is to develop an ontology for the exchange of railway asset design and maintenance data that originates from various sources based on linked data standards.

3. Research methodology

During this research, a design study research approach was adopted, based on the design cycle of Wieringa [24] (figure 1). This approach splits design science up into two activities: firstly the development of an 'artefact' which is used to help the client, and secondly the investigation of the performance of this artefact in a certain context. The decision for this approach was taken based on the scope of this research, which is the Dutch railway sector. At the end of this study, a tested ontology will be presented that can be directly implemented. Therefore, it was important that the artefact was tested in its intended context.

The design cycle is a subset of the engineering cycle. The engineering cycle also encompasses the treatment implementation and evaluation. However, these are outside the scope of this research. The design cycle makes use of design loops, which allow for improvements of the artefact based on the treatment validation. The developed ontology was designed and tested, and based on the outcomes of this test an implemented ontology was developed for the Dutch railway sector.

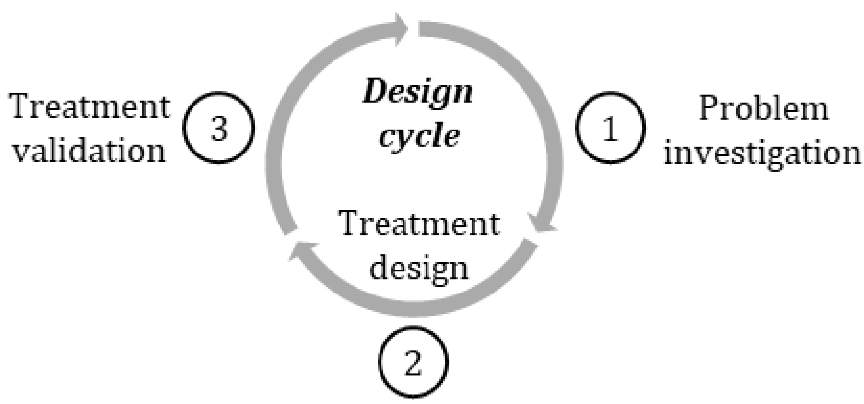


Figure 1 The design cycle (adapted from Wieringa [24])

The artefact of this research was the linked data framework that provides a method on how the various databases can be converted towards linked data standards and how the information can be efficiently retrieved by means of several queries. The context consists of the environments of design, development, maintenance and usage of the developed artefact. It is important that the study investigates the working of this artefact in the specific context to ensure the usability [24].

A technical action research was conducted, in which a newly created artefact was tested on a real case in order to assess how this artefact can help the client [24]. The newly created framework was tested on a case study, a track section for which data was available. This was done for both the conceptual and implemented ontology, where the conceptual ontology was tested using a synthetic dataset and the implemented ontology using a dataset with real data available from the railway branch. The case study and the setup are further elaborated on in section 3.3.

The author created a research strategy in which the standard knowledge questions and design problems have been adapted as described by Wieringa [24]. These were incorporated into the research strategy in combination with the research goal. An overview of the research strategy can be found in appendix A.

3.1. Approach for the problem investigation

The problem investigation was done based on an informal interview with the main stakeholders. The outcome of this were three use cases for the linked data framework, from which several functional requirements in the form of competency questions were derived. These are questions that the designed ontology should be capable of answering [25]. A validation session was held with the stakeholders to assess if the competency questions were sufficient enough to create an ontology from or that more competency questions were required. These competency questions were then used to develop the conceptual design of the railway asset design and maintenance ontology.

3.2. Approach for the treatment design

In the treatment design, a conceptual ontology was designed and afterwards implemented for the Dutch railway domain. Based on the earlier established competency questions, the conceptual design of the ontology was created and modelled, which is presented in section 5. Firstly, a schematic design was made of the ontology for all data and the type of semantic relationships that were required. Afterwards, this schematic design was modelled using RDF syntax. The modelling of the conceptual design of the ontology was done using Web Ontology Language (OWL), a semantic web standard developed by the W3C [26]. Based on this conceptual design an implemented design was created, which is presented in section 6.

In order to create the implemented ontology, the author analysed the currently stored design and maintenance data and analysed the data integration issues that could occur with the conceptual ontology. The data integration issues were analysed based on the method by Curry et al. [13], which describes four data integration challenges. Although linked data does not integrate data, these issues can also occur when trying to relate instances from the various datasets together. The following four challenges were used in the analysis: (1) **object identity and usage of separate schemas**, (2) **data mismatch,** (3) **abstraction level** and (4) **data quality**. These integration issues were analysed using the documentation as provided by the Dutch railway client ProRail and maintenance contractor Strukton.

After this, the conceptual ontology was implemented using Protégé, which is an open-source application for the modelling of ontologies in OWL [27]. The W3C standards Resource Description Framework (RDF), Resource Description Framework Schema and Web Ontology Language (OWL) were used as syntax. Syntax created in this research received the prefix 'St'. Moreover, for geospatial semantics the ontologies of Geography Markup Language (GML), Well Known Text (WKT) and GeoSPARQL were used in the ontology. These are indicated with the prefix 'Geo' in the ontology. Lastly, for the implementation of the ontology the author converted currently existing data structures to RDF and semantically modelled the relationship with the earlier developed conceptual ontology based on the data integration analysis.

In order to test the working of the implemented ontology, a case study was performed in this research using a synthetic dataset and a real dataset. Two datasets were used since the currently existing datasets are often incomplete and filled with errors. By making use of these two datasets, a distinction could be made in shortcomings related to the dataset or related to the implemented ontology. The data was uploaded to a triple store named GraphDB [28], forming the linked data framework. Using several SPARQL queries the author assessed if the implemented ontology worked as intended and if all competency questions could be answered when using the two datasets.

3.3. Approach for the treatment validation

In the treatment validation the author verified if the created artefact, the linked data framework, fulfilled the goals and derived requirements of the stakeholders that were involved in this research. This was done using two activities. First of all, the author verified if all functional requirements of the ontology were met by determining if the ontology can answer all of the competency questions. At the start of this research, three use cases were defined for which the linked data framework should offer its usability. After these use cases were defined, a number of competency questions were created that could verify if each use case was adequately implemented in the linked data framework. Secondly, the developed artefact was verified by using an expert-session with the stakeholders involved in this research.

In this expert-session, the created artefact was discussed as well as the outcomes of the case study. The ontology was schematically presented to the stakeholders. Several disciplines can be found within the developed ontology. Therefore, the stakeholders present during the presentation were invited based on their role within the organisation and were made responsible for the verification of their parts (e.g. an advisor for subsurface infrastructure was involved for the validation of the subsurface infrastructure aspects of the ontology). Moreover, the functioning of the artefact in the case study was addressed in this meeting in order to determine any shortcoming. Any shortcoming in the artefact were addressed using the design loops as presented in appendix A.

4. Use cases and competency questions

At the start of this research, an informal interview was held with the several stakeholders involved in this research. Together we formulated the main needs that the stakeholders have for data integration and how the development of the semantic web ontology could contribute to this. These three main needs are referred to as 'use cases' in this study. These use cases formed the basis from which the author derived several functional requirements in the form of competency questions. The derived competency questions were verified by the stakeholders and afterwards used to create the conceptual ontology.

The first use case of the ontology was the identification of design and maintenance information for railway assets. This would be the information that is available from all sources for the specific object. Considering that for a specific railway asset multiple stakeholders maintain datasets that contain unique data, this use case would allow for easier retrieval of all information that could be found for one asset. General information was required such as the geometry, location and unique identifiers of the object. This use case also encompassed the specific maintenance data for each railway asset.

The second use case was the identification of the interface of subsurface infrastructure and railway assets. A large part of railway asset management is in relation to the subsurface infrastructure and the interface that this infrastructure has to the railway assets on the surface. Using this use case, it could for example be possible to find the specific malfunctioning cable that was used to power a signal and quickly locate this. Again, general information was required for the subsurface infrastructure such as the geometry, location and unique identifiers. In addition, to create an integral overview of all asset data, relationships between the subsurface infrastructure and railway assets were modelled.

The third and last use case was the assessment of the validity of the data for railway assets. Considering that the ontology is designed to find and semantically match the instances from the various datasets, it is important to know the deviations in stored data and the reasons why matches were made. Multiple stakeholders maintain the data and it is often not clear what data is the most reliable or recent. Using this use case, it will be possible to always find the most recent data or the most accurate measurements. The following three use cases and accompanying competency questions were formulated (table 1):

Table 1 Use cases and competency questions for the linked data framework

UC1	**Identification of design and maintenance information for railway assets**
1.1	What are the unique naming conventions (ID) that correspond to the railway asset?
1.2	What is the object type of the railway asset (in accordance with the various OBS)?
1.3	What is the location of the railway asset (Using RD-coordinates and N.A.P.)?
1.4	What is the rotation of the railway asset (both horizontal and vertical rotation)?
1.5	What is the 3D object from the Object Type Library that corresponds to the railway asset (used to visualize the objects in a digital model)?
1.6	What are the super- and subtypes, as registered in the databases, for the railway assets?
1.7	What are the unique naming conventions that correspond to the subsurface infrastructure that interfaces with the railway assets?
UC2	**Identification of the interface with subsurface infrastructure and railway assets**
2.1	What are the unique naming conventions that correspond to the subsurface infrastructure?
2.2	What are the unique naming conventions that correspond to the railway assets that interface with the subsurface infrastructure?
2.3	What is the geometry of the subs. infrastructure?
2.4	What is the object type of the subs. infrastructure?
UC3	**Assessment of the validity of the data for railway assets**
3.1	What is the reason for the various data instances from the databases being matched with the detected object?

3.2 What is the spatial accuracy of the objects as registered?
3.3 What is the difference between the registered location of the railway asset when compared between the various datasets?
3.4 What is the most up-to-date information regarding the detected objects from the various datasets?

5. Conceptual design of the ontology

The railway asset design and maintenance ontology is modelled with the railway assets at its basis. In this ontology, a railway asset is seen as an object that is required for the functioning of the railway infrastructure that is situated in the immediate vicinity of the railway tracks, excluding the tracks themselves. Examples of these objects could be catenary portals, signals, relays cabinets and other related objects.

In figure 2 a simplified version of the ontology is shown. The assets were modelled using classes, where each distinct object type received a class type. This enabled the various railway asset instances to be related to a specific class, and for each class to have attainable object properties and data properties. This allows for each asset type to have specific attainable attributes. The modelling of the classes, data- and object properties is shown in sections 5.1. to 5.3.

The design and maintenance data and object properties were modelled in accordance to the competency questions from section 4. These properties were created with a specific domain, meaning that only a specific railway asset class can attain these properties. For example, data properties that were modelled for signals had the domain signal classes.

The ontology is modelled in a way that it could relate the information from the various datasets from third parties towards the correct railway asset. Each asset was modelled separately as an instance, and attainable design and maintenance information were created accordingly. In the implemented ontology, related data from multiple parties will be semantically modelled with the D&M information (marked in figure 2 as 'has relation'). OWL offers existing functionalities that make this possible. This allows for further data integration by third parties without altering the D&M ontology.

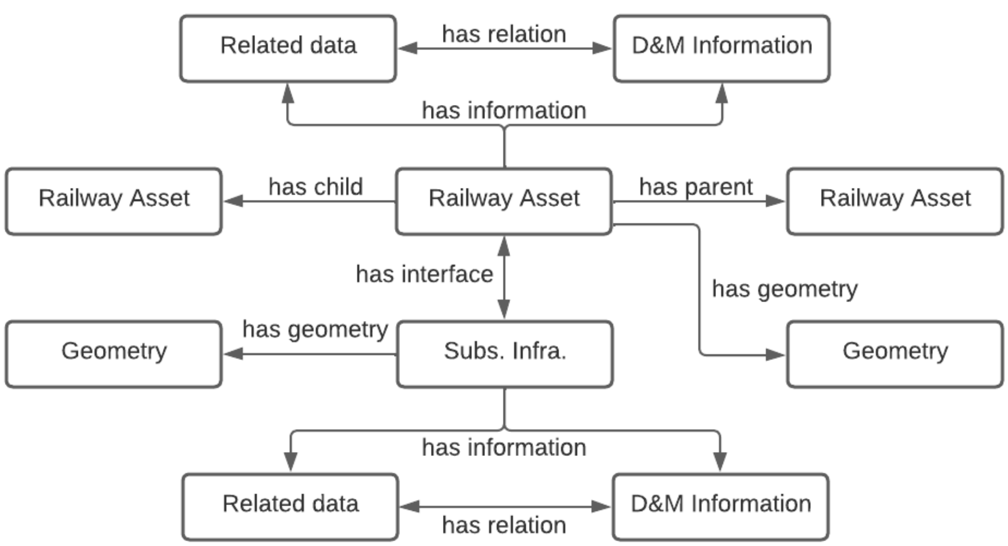


Figure 2 Simplified railway asset design and maintenance ontology

Interface with subsurface infrastructure and railway assets was modelled using a symmetric function, allowing for the interface to be semantically related to the subsurface infrastructure as well as with the railway asset it is connected to. Moreover, to indicate that a railway asset is part-of or consists-of other assets, semantic relationships between other assets in the form of child- and parent assets were made.

This was modelled as an inverse function for child- and parent objects. This assures that the ontology can provide an overview of the related railway assets.

Lastly, the property of geometry was modelled to relate the railway instances to a specific geometry. In the ontology, geometry is an instance similar to the various railway assets. The decision was made to split up the geometry from the railway asset instances in order to allow non-spatial data to also be linked to the developed ontology.

Below, the modelling of the classes, data properties and object properties is explained for the conceptual ontology. Annotation properties were created throughout the ontology, linked to the various classes, data properties and object properties. These explained what was contained within certain entities and how the inferencing functions. Moreover, general annotation properties were given to the ontology to indicate the date of the most recent version and links were made towards the imported ontologies.

5.1. Modelling of the classes

In the D&M ontology each of the types of railway assets and subsurface infrastructure were created as an individual class, where an object breakdown structure was used from general to specific as can be seen in figure 3 (e.g. railway asset signal with subclasses for each of the specific signal types). This allows for a branching out structure in which the properties of the supertypes is made inheritable to the classes that are subclasses. Most generic railway assets were modelled, and the specific types were modelled based on the known assets within the Dutch railway domain.

Also visible in figure 3 are the classes imported from the ontology GeoSPARQL, indicated with the prefix "geo". In the ontology, geometries were linked to the various railway asset instances by making use of the class geo:Geometry. GeoSPARQL can form literals to the GML syntax [29, p. 20], which was a requirement for the ontology. This also allowed for geospatial querying in section 6.3.3.

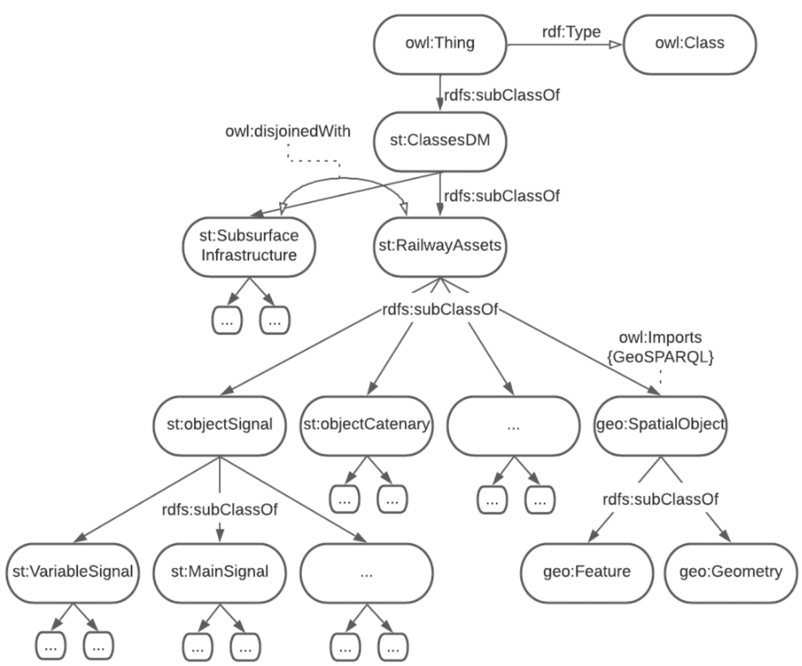


Figure 3 Railway asset D&M ontology classes

In order to infer sameness in the ontology between the various instances, each class received an *owl:hasKey* that enables the ontology to infer sameness. Each railway asset class received various object properties and data properties, which the ontology can use to match instances that are the same *owl:Class* or for instances that are inferred as *owl:equivalentClass*. The classes that did not have any matching instances (e.g. classes of signals and fix stops) were made to be disjoint (*owl:disjointWith*). This prevents any unwanted inferencing and matching of instances between these classes.

The exact 'keys' depended on the specific object class and its attainable attributes. Using this functionality, the same *owl:namedIndividual* instances from the various datasets will be automatically inferred as *owl:sameAs*. This functionality allows for third parties to link up data by simply semantically modelling their classes as *owl:equivalentClass*. The modelling of the data properties could be done using *owl:equivalentProperty*. This is further explained in chapter 5.2.

5.2. Modelling of the data properties

In the railway asset design and maintenance ontology the data properties were created by using *owl:dataProperty*. Data properties were then further divided into subsurface infrastructure and railway asset data properties, as can be seen in figure 4. This allowed for a distinction in the attainable data properties for each of the modelled classes by specifying the domain corresponding to the correct class. The classes received a number of generic data properties and data properties were created specific for each class. The modelling of the data properties was done based on the competency questions.

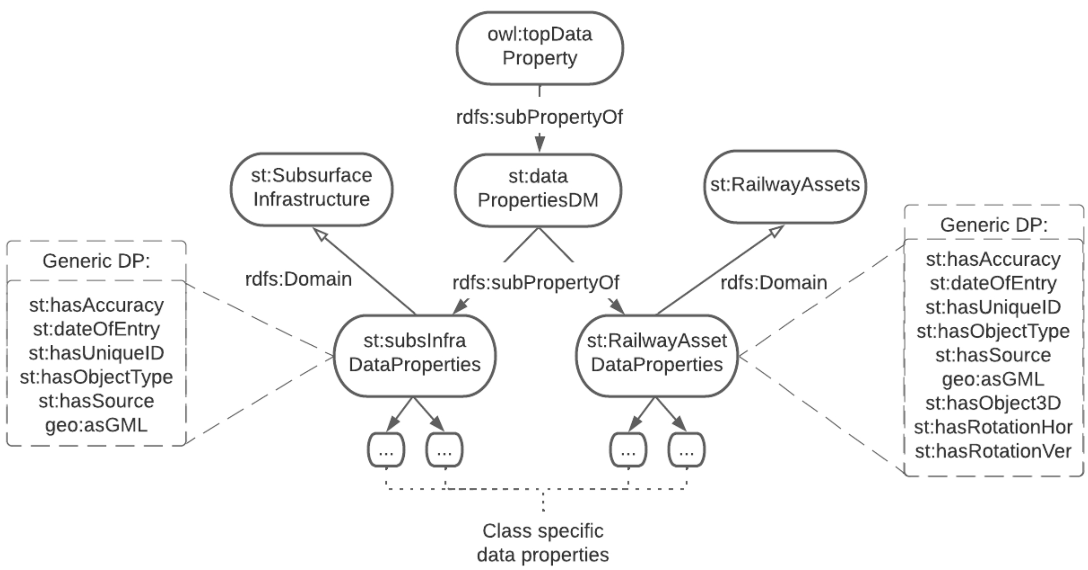


Figure 4 Railway asset D&M ontology data properties

Different data properties were required for the specific classes that were modelled in the ontology. For each of these classes a sub property was modelled and the specific data properties were created. This allows for each object to inherit the top data properties, but also to have specific data properties. The generic data properties are shown in figure 4, in which a distinction was made between properties for subsurface infrastructure and for the railway assets.

Using OWL and more specifically GML, it is possible for instances to attain multiple geometry types. The geometry for the railway assets can be split up into line geometries for the catenary buttresses and into point geometries for all other assets. The geometries for subsurface infrastructure was split up into polylines for the path of the cable and point geometries for the attachment points for the subsurface infrastructure.

5.3. Modelling of the object properties

Several object properties were modelled to indicate the relationship between various instances that can reside within the ontology. These properties can be found in figure 5. Using these object properties, it is possible to relate the parent- and child- relationships between railway assets, the interface between railway assets and subsurface infrastructure and to relate the geometries to the railway assets and subsurface infrastructure.

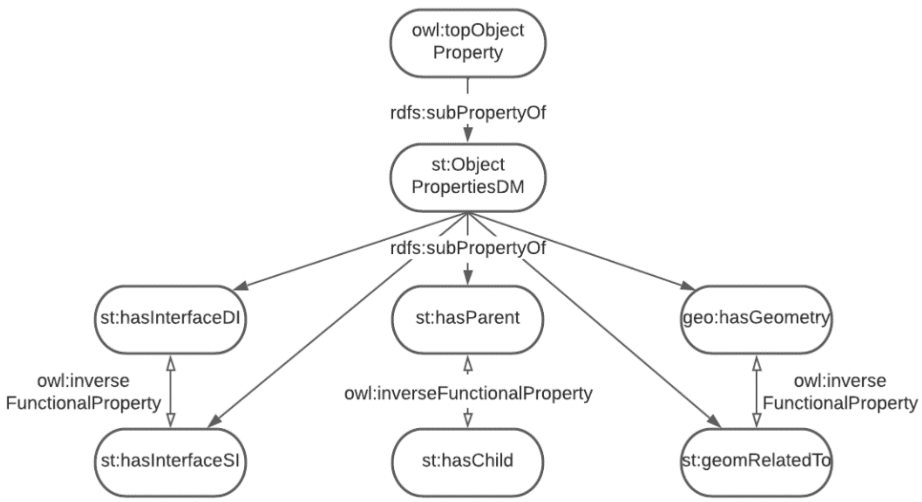


Figure 5 Railway asset D&M ontology data properties

Relationships between instances are given based on super-classes (e.g. object X has super-class object Y). These relationships were modelled using an object property named *st:hasParent*. Afterwards, the sub-class relationship *st:hasChild* was created by making an *owl:InverseFunctionalProperty* of the object property of *st:hasParent*.

To indicate the interface between railway assets and subsurface infrastructure, two object properties were created named *st:hasInterfaceRA* (railway asset) and *st:hasInterfaceSI* (subsurface infrastructure). These functions were also set to be *owl:InverseFunctionalProperty*. Lastly, the inverse property *st:geomRelatedTo* was created for the imported object property of *geo:hasGeometry*, which could be used to quickly relate the geometries to the instances.

6. Implemented design of the ontology

6.1. Available data

Multiple datasets are in use within the scope of this research and the ontology that was developed. As previously mentioned, linked data is more than a one-off solution to the integration of data. However, it is often difficult to transform the current data to semantic web standards and link the various databases by making use of URI's. Moreover, when linking data from multiple sources, the differences become apparent [30]. Two methods are possible to convert the currently used datasets towards RDF: either by directly using the developed D&M ontology, or secondly by semantically modelling the relationships which might be preferable if other object breakdown structures are used.

Within the Dutch railway domain, two datasets are mainly used to communicate design and maintenance data. These datasets have two different object breakdown structures. One of these datasets is tabular data and will be referred to using 'BID'. The other dataset is geospatial data and will be referred to using 'RP'. For the design and maintenance data (D&M) for the D&M ontology an extension is made on these two object breakdown structures. The railway asset D&M ontology is modelled using semantic relationships between these two OBS, allowing for better usability of the currently existing data. This also allows for further development of the ontology in railway domains, where other breakdown structures might be used.

Lastly, a third dataset was used in this research which was directly mapped using the developed railway asset D&M ontology. This dataset originated from the maintenance contractor Strukton, which included results from a survey of the as-is situation of railway assets. The data includes relevant D&M data, as well as geospatial positioning of the surveyed objects and subsurface infrastructure. By using the developed ontology, railway asset and subsurface infrastructure data was converted to RDF and semantically linked with the other RDF datasets in order to provide a holistic method of data management.

Before the data is prepared for the implemented ontology and case study, several data integration issues were explored. The outcome of this analysis is a method on how the currently available datasets can be converted to RDF and semantically modelled to work with the developed conceptual ontology. As previously mentioned, four data integration issues were explored: (1) object identity and usage of separate schemas, (2) data mismatch, (3) abstraction level and (4) data quality**.** The analysis showed that various schemas and standards were used and that the data is filled with many errors and shortcoming. Moreover, various abstraction levels were used, making it difficult to relate all classes. The full analysis can be found in appendix B.

6.2. Data preparation

In order to implement the currently used datasets with the developed railway asset D&M ontology, the datasets had to be converted to RDF. The approach for this was that the 'BID' and 'RP' datasets were converted to RDF with minimal changes. Each of the object types used within these datasets were modelled as an *owl:Class*, corresponding to the method of modelling of the conceptual ontology. This is shown in figure 6 as 'BID Asset' and 'RP Asset', which represent all modelled classes. Moreover, all properties were converted to RDF and the geometry from the 'RP' data was converted to a GML RDF geometry. After the conversion of these datasets to RDF, the relationship with the conceptual ontology was semantically modelled.

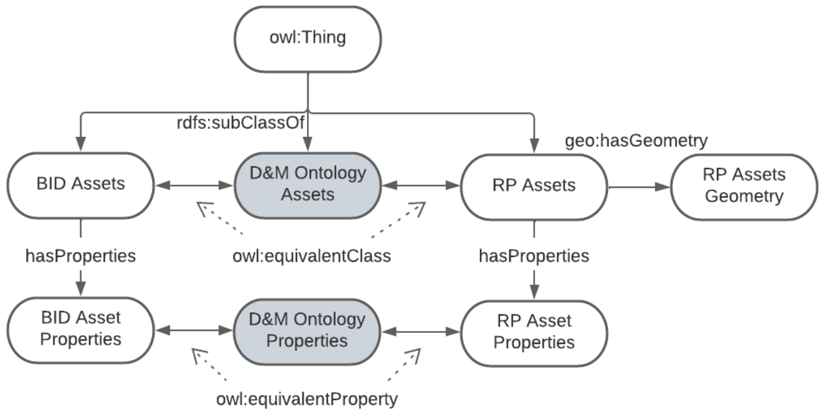


Figure 6 Converting the datasets to RDF

The relationships between the various datasets and the developed ontology were modelled using *owl:equivalentClass* and *owl:equivalentProperty*. This allowed for improved inferencing and querying of all available data. Each object type from the various object breakdown structures was semantically linked to the corresponding object type from the other object breakdown structure. Moreover, every object data property that corresponded was also semantically linked.

The next step in the data preparation was the conversion and mapping of the available datasets towards RDF. The maintenance contractor Strukton provided a dataset for railway assets which was directly mapped out using the developed railway asset D&M ontology. The datasets from the client ProRail were converted based on the simple structure as provided in figure 6. The conversion of the datasets was done using the application Protégé [27], which offers an RDF-conversion tool. Protégé allows tabular data to be transformed to RDF by mapping this out based on a provided ontology. The used data was asset data for a railway segment of approximately two kilometres within the Netherlands, which includes a large amount of subsurface infrastructure and several asset types within the case study area.

Each of the used datasets was split up into a 'real' and 'synthetic' dataset. Firstly, a 'real' dataset was converted to RDF. This was raw data directly used from the various datasets and converted to RDF, filled with errors and inconsistencies. Secondly, this 'real' dataset was cleaned up by me and afterwards also converted towards RDF, forming the 'synthetic' dataset. In the synthetic dataset the linear referencing system used was also adjusted, considering that the inferencing used within OWL works with exact values. Therefore, the accuracy was changed from meters to hectometres. This is due to *owl:hasKey* being restricted to values as defined in the ontology and not ranges of values [31]. Therefore, another solution was found in the real dataset in section 6.3.2. Lastly, inconsistencies and erroneous data was removed.

6.3. Case study

After implementing the ontology with the available datasets, the author performed a case study at a railway track segment managed by the Dutch railway client ProRail. The data that was available from this location from the 'BID' and 'RP' datasets were obtained and converted to RDF as explained in section 6.2. This was done as well for the dataset provided by the railway contractor Strukton using the developed D&M ontology. An overview of the case study area can be found in figure 7, in which sixteen railway assets can be seen as well as some subsurface infrastructure that interfaces with some railway assets. Using several queries the author verified if the linked data framework works as intended and if all of the data can be recovered for the railway assets and subsurface infrastructure.

A set of standardized queries was created that could answer each of the competency questions. These queries can be found in appendix D.

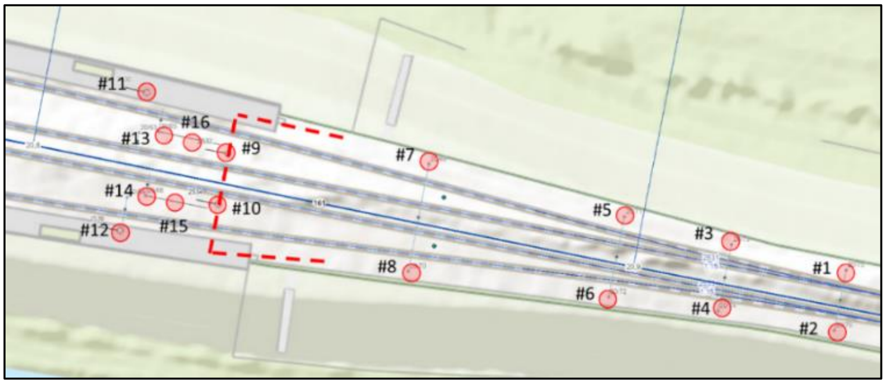


Figure 7 Overview of the case study area

6.3.1. Testing the synthetic dataset

Using the inferencing engine of GraphDB based on the *owl:hasKey* function as presented previously, semantic relationships in the form of *owl:sameAs* were made between the various railway assets present in the converted datasets. The correct functioning of the inferencing was verified in appendix C. All railway objects were correctly matched with their corresponding instance from the various datasets.

To match up the interface between the railway assets and the subsurface infrastructure a standardized query was created. No inferencing could be used since geospatial inferencing was not supported by GraphDB nor Protégé. The used SPARQL query for the creation of the interface can be found in figure 8.

Used query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
PREFIX gml: <http://www.opengis.net/ont/gml#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
insert {?SI st:hasInterfaceRA ?RailwayAsset .
  where{
    select distinct * where {
    ?SI rdf:type st:subsurfaceInfrastructure .
    ?RailwayAsset st:hasSource “Strukton” .
    ?SI geo:hasGeometry ?GeomSI .
    ?RailwayAsset geo:hasGeometry ?GeomRA .
    ?GeomSI rdf:type gml:MultiPoint .
    ?GeomSI geo:asGML ?GeomSIGML .
    ?GeomRA geo:asGML ?GeomRAGML .
    filter (geof:distance(?GeomSIWKT, ?GeomDIWKT) < 0.25) .
        }
}
```

Figure 8 Creating the interface object attributes between the railway assets and subsurface infrastructure

The query is based on the known 'attachment points' of the subsurface infrastructure from the dataset, in the query denoted as the *gml:MultiPoint.* Railway assets directly placed on top of these attachment points were then semantically linked using the interface object property. After using the query to find the matching railway assets with the subsurface infrastructure, an insert is used to create the semantic relationships in the RDF dataset. This can then be used to directly query for interfaces. It is also possible to create these relationships in the RDF knowledge graph based on an insert query with known relationships.

Now that all the various instances from the datasets are matched and the RDF knowledge graph is expanded upon, it is possible to query the information that can answer the competency questions. In figure 9 a query is shown that was used to find all the known relationships with other instances of the railway asset originating from the maintenance contractor dataset with the unique identifier 'ID-09'. As can be seen, the linked data framework matched the railway asset with two other instances that originated from the other two datasets. Moreover, it can be seen that the subsurface infrastructure 'st:Trench188' is shown to interface with the given asset.

Used query:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO #>
select * where {
    ?ID st:hasUniqueID "ID-09" .
    ?ID owl:sameAs ?SameAs .
    optional {?ID st:hasInterfaceSI ?Interface} .
}
```

Query result:

ID:	SameAs:	Interface:
st:ID-09	st:11411809	st:Trench188
st:ID-09	st:d75c7a9e-a899-48c1-8eda-8238b8673e3e	st:Trench188

Figure 9 Querying related instances and interface for asset ID-09

Now that all the instances from the various datasets are matched, we can query the object properties that any of these instances have. Shown in figure 10 is a query that was used to retrieve all the object properties that the railway asset with unique identifier 'ID-09' or its known related instances had. As can be seen, the railway asset has a known relation of *st:hasChild*, which originated from the 'BID' dataset. The ontology inferred this relationship with the instance based on the known relationship of *owl:sameAs* between the various instances.

Used query:

```
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
select ?ID ?Objectproperty ?Value where {
    ?ID st:hasUniqueID "ID-09" .
    ?ID ?Objectproperty ?Value .
    ?Objectproperty rdfs:subPropertyOf st:DMObjectProperties .
    filter (?Objectproperty != st:DMObjectProperties) .
}
```

Query results:

ID:	Objectproperty:	Value:
st:ID-09	st:hasChild	st:11411995
st:ID-09	st:hasInterfaceSI	st:Trench188

Figure 10 Querying all available object properties for asset ID-09

Now that all railway assets are inferred as *owl:sameAs*, it is possible to query all possible design and maintenance information that is available for a specific object. In figure 11 an example is given on how all available data properties for railway asset 'ID-09' could be queried. The result of the query gives an overview of all related information and indicates which of the used datasets was the source of the data property. The full excerpt of this query can be found in appendix D.

Used query:

```
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
select ?ID ?Dataproperty ?Value ?Source where {
   ?ID st:hasUniqueID "ID-09" .
   ?ID ?Dataproperty ?Value .
   ?Dataproperty rdfs:subPropertyOf st:DMDataProperties .
   ?Dataproperty rdfs:subPropertyOf ?Source .
   Filter ( ?Source = st:DMDataProperties ) .
}
```

Figure 11 Querying all available data properties for asset ID-09

A visual overview of the railway asset that was used in the previous queries in the form of RDF triples can be found in figure 12. GraphDB offers a triple viewer that could easily help a user to navigate throughout the RDF knowledge graph. The figure below is a cleaned up overview from the triple viewer from GraphDB.

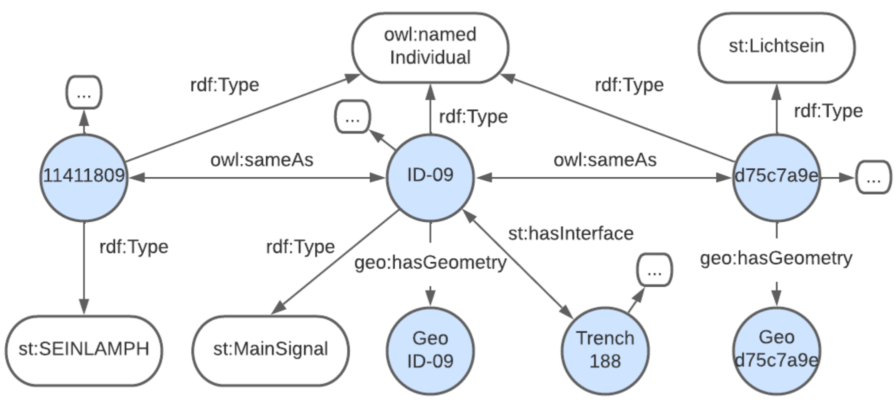


Figure 12 Railway asset ID-09 as triples

As can be seen, the railway asset with 'ID-09' was matched with two other railway assets marked as *owl:sameAs*. Moreover, it shows an interface with the subsurface infrastructure 'Trench-188'. Moreover, it is possible to navigate these triples and extend the overview by selecting the instances, which shows other relationships in the form of object properties and the available data properties of each of these instances.

The case study using the synthetic dataset showed that it is possible to facilitate large scale data integration without altering the various databases in use or by using various schemas. The semantic web ontology provided a clear and easy way to integrate the various datasets together by making use of inferencing. More datasets could be used by directly mapping these out on the developed railway asset design and maintenance ontology (such as the maintenance contractor dataset) or by semantically linking the RDF datasets (such as the client's datasets). All competency questions could be easily queried using the developed ontology, therefore the developed semantic web ontology is capable of fulfilling all three of its intended use cases.

6.3.2. Testing the real dataset

The process as described in section 6.3.3.1. was repeated for the real dataset, which was the unedited data directly from the maintenance contractor and client. The data was converted to RDF and the inferencing engine was used again to semantically relate the various instances as *owl:sameAs*. As expected, the inferencing engine did not match all railway assets as *owl:sameAs* due to the accuracy of same of the data properties (see also section 6.3.2. on the usage of *owl:hasKey* for the shortcomings).

The lack of capabilities of the inferencing engine was solved for the real dataset by making use of a SPARQL query in combination with the GEOSPARQL extension. The used query to match all railway assets can be found in figure 13. Using the query, all railway assets were successfully matched. It is also possible to relate the railway assets based on an insert query with known relationships between the various railway assets. These known relationships were not available in this study. Non-geospatial data was matched based on matching data properties.

Used query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO #>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
insert    {?RP owl:sameAs ?RailwayAsset .
           ?RailwayAsset owl:sameAs ?RP .}
where{
      select distinct ?RP ?RailwayAsset where {
      ?RP rdf:type ?ObjecttypeRP .
      ?RailwayAsset rdf:type ?ObjecttypeRA .
      ?ObjecttypeRP owl:equivalentClass ?ObjecttypeRA .
      ?RP geo:hasGeometry ?Geom2 .
      ?RailwayAsset geo:hasGeometry ?Geom1 .
      ?Geom2 geo:asGML ?Geom2GML .
      ?RailwayAsset geo:asGML ?Geom1GML .
      filter (geof:distance(?Geom2WKT, ?Geom1WKT) < 0.25) .
           }
}
```

Figure 13 Matching the railway assets using GEOSPARQL

The case study using the synthetic dataset showed that the inferencing that was incorporated in the developed railway asset design and maintenance ontology could not correctly infer all railway assets as *owl:sameAs*. Inferencing was not done incorrectly, but due to the data properties having a large deviation between the various instances or data properties completely missing matching railway assets were not identified by the ontology. Using a GEOSPARQL query it was possible to relate all these instances, and afterwards to query all related data. Despite inferencing not identifying all semantic relationships, it is still possible to provide one overview of all available information using the developed ontology in combination with queries.

The case study showed that it is still possible to answer each of the competency questions and fulfil the intended use cases. However, due to issues with inferencing not all data could be retrieved from the linked data framework. Not all relationships could be established between the various datasets and therefore it was not possible to query all of the available data for a railway asset. It is still possible to query this data without the usage of inferencing. This could be done by converting a dataset towards RDF in which the relationships between the datasets are available. By updating the triple store from the linked data framework, all inferencing should work as intended.

7. Verification

7.1. Competency questions

The competency questions formed the functional requirements for the ontology that guided this research. Therefore, to validate the ontology the author checked if all competency question could be answered and also if each could be queried using the linked data framework. The used queries and outcome can be found in appendix D.

All competency questions could be answered by the ontology and therefore all functional requirements of the ontology have been met. Moreover, the linked data framework was capable of querying each of the competency questions using a set of standard queries (appendix D).

7.2. Expert session

During the development of the railway asset design and maintenance ontology a group of experts was involved that helped with the formulation of the competency questions and gave expert input for the ontology. At the end of this research, an expert session was held to validate the developed ontology. The involved experts were asked to judge the developed ontology on shortcomings for data properties and object properties, clarity of the used syntax in the ontology and general remarks or wishes.

In this expert session, the developed ontology was discussed and feedback was noted. The feedback was afterwards incorporated in the ontology. The following experts were present in this meeting:

- Technical manager of the innovation project.
- Project manager of the innovation project.
- Planner & designer subsurface infrastructure.
- Planner & designer rail.

The expert session was organised as follows: first a general presentation was given on how the ontology was developed, an overview of the ontology itself and the outcomes of the case study. After this, an informal discussion was held and each attendee was asked for input.

No shortcomings were found in the used data properties and object properties of the developed ontology, also considering that the experts were involved from the early stages and that all competency questions could be answered. Several notes were made on the extendibility of the developed railway asset design and maintenance ontology. The wish is to further extend this ontology for the various lifecycle stages of an asset to provide a more elaborate dataset in RDF. Recommendations for further developments of this ontology are noted in the discussion.

An issue with the used syntax was that multiple languages were in use in the RDF datasets. Upon conversion of the case study data, the two datasets from the client were converted to RDF without altering any of the available syntax. This data was in Dutch, whereas the developed ontology was in English. To solve this, the used syntax from the Dutch RDF dataset were semantically linked to various object- and data properties residing in the developed ontology. Therefore, when querying the dataset it was possible to always get the English properties as a result, while the converted datasets where still in Dutch.

A general remark made was the usability of the SPARQL interface of the linked data framework. Considering that not everyone has experience with SPARQL or RDF, it might prove difficult to retrieve the required information. Although a list of standardized queries was provided in this research, it is still difficult to adjust these or create more queries. It was therefore advised to use an application with a visual query builder, which could assist the user in retrieving relevant data.

Most of the feedback noted during the expert session were was minor and did not have any significant impact on the development of the semantic web ontology. This was partly expected, since the stakeholders involved in the expert session were also involved during the research. This allowed them to express their expert judgement in more stages than only the expert session. It could however be that if different stakeholders were involved that other feedback would have been given. Qualitative feedback was taken into account. However, the study lacked quantitative feedback from many stakeholders. This was mainly due to the scope of the research and the limited time available. A recommendation was made in the discussion to perform a more extensive expert session and to involve more stakeholders in the further development of the ontology.

8. Discussion

The main scientific contribution of this research was the development of a semantic web ontology for the railway asset design and maintenance domain and the tested and implemented ontology specific for the Dutch railway domain. It was shown that by making use of the developed ontology large scale data integration is possible while still using decentralised data. Several datasets were converted towards RDF and mapped out based on the developed ontology, and by making use of a set of standardized SPARQL queries information could be retrieved. The SPARQL queries were able to relate all known information across each of the stakeholders to the correct railway asset and relate this information to the end user. Despite the working of the ontology, several limitations should be addressed and further explored.

One of the main limitations found while using the developed ontology is the lack of options for inferencing data that is not a one-on-one match. Using the synthetic dataset, it was possible to facilitate full scale data integration by using the *owl:hasKey* function from OWL 2 for the inferencing of *owl:sameAs* instances. However, since this function is restricted to values as defined in the ontology and not ranges of values [31], this could not be used for the unedited dataset. In the synthetic dataset this was solved by altering the accuracy of the data properties, which resulted in the one-on-one matching of the properties of the instances.

Another limitation found with the usage of the inferencing and semantically relating similar instances together is that missing identifiers present within the used datasets could lead to triples not being made. This could be resolved by making used of so called blank nodes, which could indicate that an identifier is missing but should have been present. This issue can become even more apparent when working with large scale real data, where errors occur more often. However, this was not implemented in the ontology and could thus be a valuable addition.

With large scale data integration, issues could occur from the usage of the various object breakdown schemas in use. When trying to semantically relate the various instances, it could happen that due to the used level of detail in these breakdown structures some object classes cannot be easily related. An extension could be made on this ontology by making use of the OWL semantics such as *owl:intersectionOf*, *owl:unionOf* and *owl:complimentOf*. This way it is possible to relate the various instances in more ways than just the functionality of *owl:sameAs* as used in this research. This issue was not experienced in this research with the used datasets.

Future developments of the ontology should include the expansion of the railway assets classes included. An extension should be made for the attainable data- and object properties for these classes. Moreover, more information relevant to the lifecycle stages of the railway assets could be modelled in OWL and added to this ontology, allowing for holistic management of railway assets. During this expansion, it is important to involve more stakeholders than were involved in this research to also address the quantitative outcomes. There are many stakeholders involved in the railway asset design and maintenance domain, so it was not possible to include a representative for each group of stakeholders in this research. The developed ontology could also allow for more cross-domain data integration by making extensions on the developed ontology or by using open linked data trough SPARQL endpoints.

9. Conclusion

The author implemented and tested the semantic web ontology using datasets from the Dutch railway domain. Using several functions from Web Ontology Language and by making use of an inferencing engine it was shown that it is possible to semantically relate the same instances from various datasets. This was possible for a synthetic dataset in which the data properties matched between the various instances in use. This was also possible for a dataset that was not cleaned up, but some limitations were found.

Issues occurred with the developed ontology when using a dataset that was not cleaned up, due to limitations in Web Ontology Language and its inferencing engine. Not all railway asset instances from the various datasets were correctly matched by the inferencing engine due to the accuracy of the data properties. To solve this, a query could be used in combination with the GEOSPARQL extension to find and match similar objects. However, this only occurred with some instances. It was also shown that by slightly cleaning up the data the inferencing worked correctly.

Several classes were modelled for the various railway assets in usage in the Dutch railway domain. Related data properties were modelled for each of these assets and a number of object properties were modelled to address the relationships between the various instances of railway assets. The developed ontology provided a manner in which datasets can be mapped towards RDF or a manner on how RDF data using a different structure can be semantically related to the developed ontology.

Using the developed railway asset design and maintenance ontology, it was shown that large scale integration of decentralised data was possible. By using an inferencing engine, various railway assets were marked as *owl:sameAs* and all relevant information could be easily queried by making usage of the identifiers that each party uses themselves. Moreover, the various silos of data could be integrated and more relationships could be made than with the data separately, such as the interface object attributes between railway assets and subsurface infrastructure.

10. References

[1] European Union, "Implementation of ERTMS by the EU." https://transport.ec.europa.eu/transport-modes/rail/ertms_en (accessed Dec. 21, 2022).

[2] ProRail, "ERTMS: het digitale spoorplatform." https://www.prorail.nl/programmas/ertms (accessed Aug. 07, 2022).

[3] RailTech, "EU Commission: deployment ERTMS is behind on schedule," 2021. https://www.railtech.com/digitalisation/2021/01/27/eu-commission-deployment-ertms-is-behind-on-schedule/?gdpr=accept (accessed Aug. 07, 2022).

[4] T. Cerovsek, "A review and outlook for a 'Building Information Model' (BIM): A multi-standpoint framework for technological development," *Advanced Engineering Informatics*, vol. 25, no. 2, pp. 224–244, Apr. 2011, doi: 10.1016/j.aei.2010.06.003.

[5] B. Succar and E. Poirier, "Lifecycle information transformation and exchange for delivering and managing digital and physical assets," *Autom Constr*, vol. 112, Apr. 2020, doi: 10.1016/j.autcon.2020.103090.

[6] Y. Jung and M. Joo, "Building information modelling (BIM) framework for practical implementation," *Autom Constr*, vol. 20, no. 2, pp. 126–133, Mar. 2011, doi: 10.1016/j.autcon.2010.09.010.

[7] B. Succar, "Building information modelling framework: A research and delivery foundation for industry stakeholders," *Autom Constr*, vol. 18, no. 3, pp. 357–375, 2009, doi: 10.1016/j.autcon.2008.10.003.

[8] R. Kenley, T. Harfield, and A. Behnam, "BIM Interoperability Limitations: Australian and Malaysian Rail Projects," in *International Building Control Conference*, 2016. doi: 10.1051/00102.

[9] A. Bradley, H. Li, R. Lark, and S. Dunn, "BIM for infrastructure: An overall review and constructor perspective," *Autom Constr*, vol. 71, pp. 139–152, 2016, doi: 10.1016/j.autcon.2016.08.019.

[10] C. Quinn, A. Z. Shabestari, T. Misic, S. Gilani, M. Litoiu, and J. J. McArthur, "Building automation system - BIM integration using a linked data structure," *Autom Constr*, vol. 118, Oct. 2020, doi: 10.1016/j.autcon.2020.103257.

[11] D. Y. Lee, H. lin Chi, J. Wang, X. Wang, and C. S. Park, "A linked data system framework for sharing construction defect information using ontologies and BIM environments," *Autom Constr*, vol. 68, pp. 102–113, Aug. 2016, doi: 10.1016/j.autcon.2016.05.003.

[12] E. Corry, J. O'Donnell, E. Curry, D. Coakley, P. Pauwels, and M. Keane, "Using semantic web technologies to access soft AEC data," in *Advanced Engineering Informatics*, Elsevier Ltd, Oct. 2014, pp. 370–380. doi: 10.1016/j.aei.2014.05.002.

[13] E. Curry, J. O'Donnell, E. Corry, S. Hasan, M. Keane, and S. O'Riain, "Linking building data in the cloud: Integrating cross-domain building data using linked data," *Advanced Engineering Informatics*, vol. 27, no. 2, pp. 206–219, Apr. 2013, doi: 10.1016/j.aei.2012.10.003.

[14] S. Shirowzhan, S. M. E. Sepasgozar, D. J. Edwards, H. Li, and C. Wang, "BIM compatibility and its differentiation with interoperability challenges as an innovation factor," *Automation in Construction*, vol. 112. Elsevier B.V., Apr. 01, 2020. doi: 10.1016/j.autcon.2020.103086.

[15] E. P. Karan and J. Irizarry, "Extending BIM interoperability to preconstruction operations using geospatial analyses and semantic web services," *Autom Constr*, vol. 53, pp. 1–12, 2015, doi: 10.1016/j.autcon.2015.02.012.

[16] P. Hitzler, "A review of the semantic web field," *Communications of the ACM*, vol. 64, no. 2. Association for Computing Machinery, pp. 76–83, Feb. 01, 2021. doi: 10.1145/3397512.

[17] T. Berners-Lee, "Linked data," Jul. 27, 2006. http://www.w3.org/DesignIssues/LinkedData.html (accessed Sep. 20, 2022).

[18] M. Stepien, A. Jodehl, A. Vonthron, M. König, and M. Thewes, "An approach for cross-data querying and spatial reasoning of tunnel alignments," *Advanced Engineering Informatics*, vol. 54, 2022, doi: 10.1016/j.aei.2022.101728.

[19] B. Ranjgar, A. Sadeghi-Niaraki, M. Shakeri, and S. M. Choi, "An ontological data model for points of interest (POI) in a cultural heritage site," *Herit Sci*, vol. 10, no. 1, Dec. 2022, doi: 10.1186/s40494-021-00635-9.

[20] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, no. 2, pp. 199–220, 1993, doi: https://doi.org/10.1006/knac.1993.1008.

[21] R. Falbo, "SABiO: Systematic approach for building ontologies," *CEUR Workshop Proc*, vol. 1301, Oct. 2014.

[22] Open Geospatial Consortium, "OGC - Geography Markup Language." https://www.ogc.org/standard/gml/ (accessed Mar. 23, 2023).

[23] R. Battle and D. Kolas, "GeoSPARQL : Enabling a Geospatial Semantic Web," 2011. Accessed: Jan. 16, 2023. [Online]. Available: https://www.semanticscholar.org/paper/GeoSPARQL-%3A-Enabling-a-Geospatial-Semantic-Web-Battle-Kolas/9ab92c14f104da13b67d9d5b87b8de324fb4ebf3

[24] R. J. Wieringa, *Design Science Methodology for Information Systems and Software Engineering*. Springer, 2014. doi: 10.1007/978-3-662-43839-8.

[25] M. Grüninger and M. S. Fox, "Methodology for the Design and Evaluation of Ontologies Analysis Methods for Urban Infrastructure Systems View project Callisto: An Intelligent Project Management System View project," 1995. [Online]. Available: https://www.researchgate.net/publication/2288533

[26] World Wide Web Consortium (W3C), "Web Ontology Language," Dec. 2012. https://www.w3.org/OWL/ (accessed Apr. 23, 2023).

[27] M. A. Musen, "The protégé project," *AI Matters*, vol. 1, no. 4, pp. 4–12, Jun. 2015, doi: 10.1145/2757001.2757003.

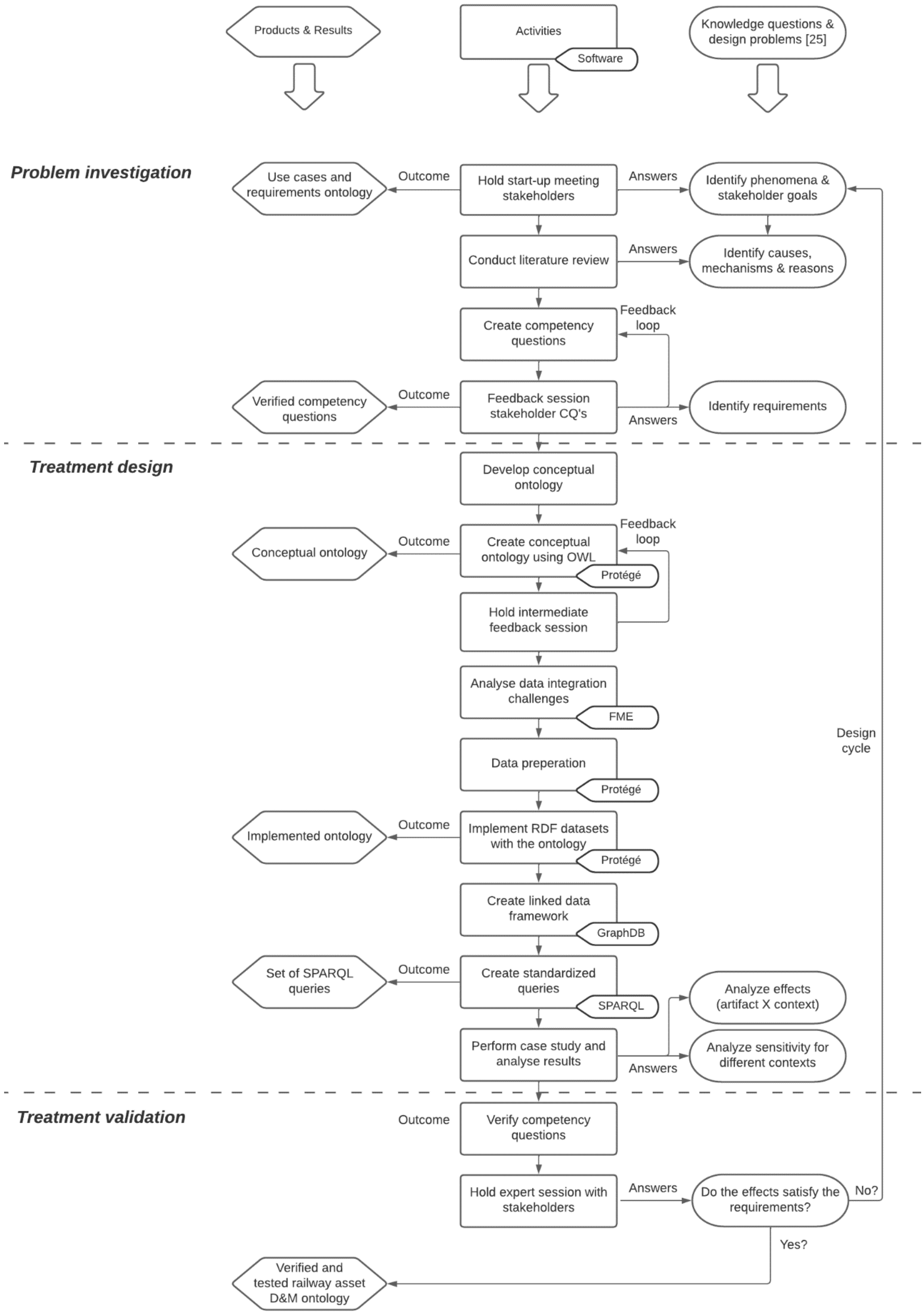
[28] Ontotext, “Ontotext GraphDB.” https://www.ontotext.com/products/graphdb/?ref=menu (accessed Apr. 23, 2023).

[29] Open Geospatial Consortium, “OGC GeoSPARQL - A Geographic Query Language for RDF Data.” Accessed: Jan. 11, 2023. [Online]. Available: https://portal.ogc.org/files/?artifact_id=47664

[30] D. Allemang and J. Hendler, “Chapter 6 - RDF and inferencing,” in *Semantic Web for the Working Ontologist (Second Edition)*, D. Allemang and J. Hendler, Eds., Boston: Morgan Kaufmann, 2011, pp. 113–123. doi: https://doi.org/10.1016/B978-0-12-385965-5.10006-8.

[31] W3C, “Easy Keys,” 2008. https://www.w3.org/2007/OWL/wiki/Easy_Keys (accessed Jan. 23, 2023).

Appendix A – Research strategy



Appendix B - Data integration analysis

The four Integration issue challenges were analysed using a document study and by exploring a dataset using Feature Manipulation Engine (FME). A script was created in FME capable of data integration for the specific scenario of this research and issues during the integration were highlighted. The goal of this section is not to integrate all data from the databases, but only to retrieve the manner in which can be referred between the various datasets.

The data integration script was tested on several railway assets for which the relationship between the various datasets was known. If the script matched these railway assets correctly, this could be used in the development of the ontology. The various instances of the railway assets were integrated for as far as possible. Meaning, that for some instances no conclusion could be made based on the available data. These limitations will carry over towards the linked data framework.

When analysing the data, it becomes apparent that not all objects from the datasets can be matched. These issues mainly occur due to the available data being lacking of some railway assets. The results of the data integration issues and the way object instances can be related was used in the creation of the ontology for railway assets. This was incorporated in the ontology using the *owl:hasKey* function.

Object identity and separate schema:

In the case study for the implemented ontology, three datasets were used. One of this datasets contained tabular data, whereas the other two datasets were XML-files based on the geospatial standard GML. The datasets provided by the maintenance contractor and the ‘RP’ dataset from the client both contain geospatial data, whereas the dataset ‘BID’ contains only railway asset maintenance information in tabular form. Using standard data integration techniques, it is difficult to integrate this data.

Although two of the datasets were both derived from the GML standard, the schemas did not match and made data integration difficult. Considering that the ‘BID’ dataset is also created using a different schema, data integration becomes difficult without the usage of capable software.

Linked data makes use of RDF for the exchange of data, which can be used regardless of the schema that was used to map out the data. However, at the start of this research, all data was still in tabular form. Using open software this tabular data can be converted to RDF based on the developed ontology, which can then be used in the linked data framework. Therefore, object identity and separate schema is not an issue in this research. This was explained in section 6.3.2.

Data mismatch:

The two datasets from the client and the dataset from the railway contractor each used different unique identifiers for the railway assets and subsurface infrastructure. There is no known or logical relation between the unique identifiers, which prevents simple data integration. For the datasets used there is also no logical link between the attributes that were used. Different names are used for classes, object- and data- properties although these had the same intention. In the ontology this was overcome using the OWL functionalities of *owl:equivalentClass* and *owl:equivalentProperty*.

When taking a closer look at the data, it can be noted that object attributes are stored using multiple naming standards, which are standardized but misused during the entry of the object information into the system. The datasets contained data that originated from various documents that did not limit its entry values. Using the developed railway asset design and maintenance ontology, this issue would be limited. This because the datasets can be directly mapped out using the developed ontology, leading to a standardized RDF dataset. Moreover, In the ontology this was also solved by allowing for the querying of all the used attribute naming conventions from the datasets including source.

Lastly, a mismatch occurs when an object is referred to using a linear reference system, as it is not specified which linear reference system should be used if two systems overlap. Therefore, the same object can be referred to using multiple linear reference system. This could not be solved in the developed ontology, considering that this is an entry error and not within the scope of this research.

Abstraction level:

Three different object breakdown structures were used in the three datasets from this research. The abstraction level of the railway assets depended on the type of data that was stored and how this could efficiently be stored. For example, the 'RP' dataset contained geospatial data based on GML. Thus, objects were split up into the geometries as also defined in GML such as points and lines. A similar approach was taken in the dataset of the maintenance contractor. However, a lower level of detail was used in this dataset. Lastly, the 'BID' object breakdown structure was derived for how maintenance is usually done, and can be seen as an object breakdown structure with functionality in mind.

An example for the various object breakdown structures can be found in figure 14. As can be seen, the maintenance contractor (MC) and 'RP' dataset are split up into either point (MC: Catenary pole, RP: pole) or line objects (RP: Beam and Arm). Also the 'RP' dataset has a higher level of detail. The 'BID' dataset used a different breakdown structure, which was done since based on the storage of maintenance data.

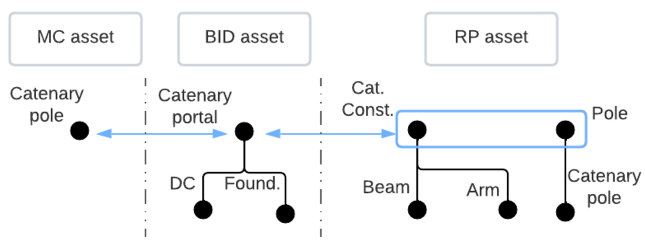


Figure 14 Various abstraction levels in the datasets

Data quality (DQ):

When analysing the various datasets in use, several errors were found related to the quality of the datasets. The case study data did not contain similar instances, but when using other locations some issues might be found. The following issues were found:

- Duplicates in the same dataset of instances.
- Duplicates in the same dataset of naming conventions.
- Entry errors by a human operator (typos).
- No consistency in the usage of the linear reference system.
- Inaccurate registration of location using the linear reference system, large differences in the datasets were found.

These issues could not easily be solved in the developed ontology, but were also outside the scope of this research. In chapter 6.3.2. the data preparation steps are explained. The case study data did not contain these errors, except for the inaccuracy in the linear reference system.

Appendix C – Validation of the inferencing

Table 2 validation of the inferencing using the case study datasets

#	Object type	Equipment	PUIC	Object ID	Correctly matched?
1	Catenary pole	11788084	14d8569b-9fc1-448f-92c0-3bfa0a1bbd81	ID-01	Yes
2	Catenary pole	11788085	10fc1fee-c125-4eec-bcef-6263fd2f5ce8	ID-02	Yes
3	Catenary pole	11788086	31ff428d-d511-4667-8a5c-0ee52bbd312b	ID-03	Yes
4	Catenary pole	11788087	6bbccf66-8725-4358-85af-dfd0c20ac82d	ID-04	Yes
5	Catenary pole	11788088	9fa5c8e1-ce24-4d81-b91d-2d8fa6b875f4	ID-05	Yes
6	Catenary pole	11788089	75a2f9fa-b3a4-4b76-9d52-8e13c5397430	ID-06	Yes
7	Catenary pole	11788090	c89861ec-d05a-4bbd-bf90-bcd282a5eb12	ID-07	Yes
8	Catenary pole	11788091	5047f07a-543b-4241-99c1-99ae75115942	ID-08	Yes
9	Signal	11411809	d75c7a9e-a899-48c1-8eda-8238b8673e3e	ID-09	Yes
10	Signal	11411811	d7f18861-99b2-4269-8b0a-75e4328538a0	ID-10	Yes
11	Signal	11411812	b9d4834e-a105-46e6-932c-87826278d8a3	ID-11	Yes
12	Signal	11411810	e75fa933-867f-4ae7-b89f-9e567e37a0dd	ID-12	Yes
13	buttress	Not available	b1626b61-dd9b-4145-b0da-775a5e1c9f87	ID-13	Yes
14	buttress	Not available	d149e1ff-ea3b-400b-b198-8297932a7a6c	ID-14	Yes
15	Catenary pole	11482044	a1baf870-2923-4bc2-99a8-31fbea0222b2	ID-15	Yes
16	Catenary pole	11482043	6de9c4a6-0352-48fc-b07e-1f3634a4176c	ID-16	Yes

#	Object type	Equipment	Has interface with	Object ID	Correctly matched?
17	Signal	11411809	Trench-188	ID-09	Yes
18	Signal	11411811	Trench-188	ID-10	Yes

Appendix D – Queries for the linked data framework

The railway asset with unique identifier ID-09 was used as an example for the following queries. It is also possible to use these queries for the other railway assets in the RDF dataset.

Querying all available data properties from ID-09:

Used query:

```
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
select ?ID ?Dataproperty ?Value ?Source where {
   ?ID st:hasUniqueID "ID-09" .
   ?ID ?Dataproperty ?Value .
   ?Dataproperty rdfs:subPropertyOf st:DMDataProperties .
   ?Dataproperty rdfs:subPropertyOf ?Source .
   Filter ( ?Source = st:DMDataProperties ) .
}
```

Query result:

ID:	Dataproperty:	Value:	Source
st:ID-09	st:has CoordinateX	“177852.851”	st:Data PropertiesDM
st:ID-09	st:has CoordinateX	”177852.523”	st:Data PropertiesRP
st:ID-09	st:has CoordinateY	”505219.271”	st:Data PropertiesDM
st:ID-09	st:has CoordinateY	”505219.296”	st:Data PropertiesRP
st:ID-09	st:has CoordinateZ	”3.52”	st:Data PropertiesDM
st:ID-09	st:has CoordinateZ	”0.0”	st:Data PropertiesRP
st:ID-09	st:hasDate	”30-01-2023”	st:Data PropertiesDM
st:ID-09	st:hasDate	”22-12-2022”	st:Data PropertiesRP
st:ID-09	st:has ObjectType	st:Sein	st:Data PropertiesDM
st:ID-09	st:has ObjectType	st:Lichtsein	st:Data PropertiesRP
st:ID-09	st:has ObjectType	st:SEIN	st:Data PropertiesBID
st:ID-09	st:hasStatus	“Vrij, INGB”	st:Data PropertiesBID
st:ID-09	st:has Startpoint	“208”	st:Data PropertiesRP
st:ID-09	st:has Endpoint	“208”	st:Data PropertiesBID
st:ID-09	st:has UniqueID	“ID-09”	st:Data PropertiesDM

st:ID-09	st:has UniqueID	“d75c7a9e-a899-48c1-8eda-8238b8673e3e”	st:Data PropertiesRP
st:ID-09	st:has UniqueID	“11411809”	st:Data PropertiesBID
		…	

Competency Question 1.1:

Used query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?UniqueNamingConv ?Source where {
    ?ID st:hasUniqueID "ID-09" .
    ?ID st:hasUniqueID ?UniqueNamingConv .
    ?ID ?hasProperty ?UniqueNamingConv .
    ?hasProperty rdfs:subPropertyOf ?Source .
    Filter ( ?Source = st:DMDataProperties ) .
}
```

Query result:

ID:	UNC:	Source:
st:ID-09	“ID-09”	st:Data PropertiesDM
st:ID-09	“11411809”	st:Data PropertiesBID
st:ID-09	“d75c7a9e-a899-48c1-8eda-8238b8673e3e”	st:Data PropertiesRP

Competency Question 1.2:

Used query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?Objecttype ?Source where {
    ?ID st:hasUniqueID "ID-09" .
    ?ID st:hasObjectType ?Objecttype .
    ?ID ?hasProperty ?Objecttype .
    ?hasProperty rdfs:subPropertyOf ?Source .
    Filter ( ?Source = st:RailwayAssets ) .
}
```

Query result:

ID:	Objecttype:	Source:
st:ID-09	st:Sein	st:Classes DM
st:ID-09	st:Lichtsein	st:Classes RP
st:ID-09	st:SEIN	st:Classes BID

Competency Question 1.3:

Used query:

```
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
select ?ID ?Geometry ?Source where {
    ?ID st:hasUniqueID "ID-09" .
    ?ID geo:hasGeometry ?GeomID .
    ?GeomID geo:asWKT ?Geometry .
    ?GeomID ?hasProperty ?Geometry .
    ?hasProperty rdfs:subPropertyOf ?Source .
    Filter ( ?Source = st:DMDataProperties ) .
}
```

Query result:

ID:	Geometry:	Source:
st:ID-09	POINT Z(177852.851 505219.271 0)	st:Data PropertiesDM
st:ID-09	POINT Z (177852.52299999818 505219.2960000001 3.140499999994063)	st:Data PropertiesRP

Competency Question 1.4:

Used query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?Rotation ?Source where {
    ?ID st:hasUniqueID "ID-09" .
    ?ID st:hasRotation ?Rotation .
    ?ID ?hasProperty ?Rotation .
    ?hasProperty rdfs:subPropertyOf ?Source .
    Filter ( ?Source = st:DMDataProperties ) .
}
```

Query result:

ID:	Rotation:	Source:
st:ID-09	“37”	st:Data PropertiesDM

Competency Question 1.5:

Used query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?Object3D ?Source where {
    ?ID st:hasUniqueID "ID-09" .
    ?ID st:hasRotation ?Object3D .
    ?ID ?hasProperty ?Object3D .
```

?hasProperty rdfs:subPropertyOf ?Source . Filter (?Source = st:DMDataProperties) . }
Query result:

ID:	Object3D:	Source:
st:ID-09	"https://docs.b360.autodesk.com/projects/0f23d881-224a-444b-8250..."	st:Data PropertiesDM

Competency Question 1.6:

Used query:

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?Child ?Parent ?Source where {
?ID st:hasUniqueID "ID-09" .
?ID st:hasChild ?Child .
?ID st:hasParent ?Parent .
}

Query result:

ID:	Child:	Parent:
st:ID-09	st:11411995	-

Competency Question 1.7:

Used query:

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?hasInterface ?IDinterface where {
?ID st:hasUniqueID "ID-09" .
?ID st:hasInterfaceSI ?hasInterface .
?hasInterface st:hasUniqueID ?IDinterface .
}

Query result:

ID:	hasInterface	IDinterface
st:ID-09	st:Trench-188	"Trench-188"

Competency Question 2.1:

Used query:
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#> select ?ID ?UniqueNamingConv ?Source where { ?ID st:hasUniqueID "Trench-188" . ?ID st:hasUniqueID ?UniqueNamingConv . ?ID ?hasProperty ?UniqueNamingConv . ?hasProperty rdfs:subPropertyOf ?Source . Filter (?Source = st:DMDataProperties) .

```
}
```

Query result:

ID:	UNC:	Source:
st:Trench-188	“Trench-188”	st:Data PropertiesDM

Competency Question 2.2:

Used query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?hasInterface ?IDinterface where {
    ?ID st:hasUniqueID "Trench-188" .
    ?ID st:hasInterfaceDI ?hasInterface .
    ?hasInterface st:hasUniqueID ?IDinterface .
}
```

Query result:

ID:	hasInterface:	IDinterface:
st:Trench-188	st:ID-09	“ID-09”
st:Trench-188	st:ID-10	“ID-10”

Competency Question 2.3:

Used query:

```
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
select ?ID ?Geometry ?Source where {
    ?ID st:hasUniqueID "Trench-188" .
    ?ID geo:hasGeometry ?GeomID .
    ?GeomID geo:asWKT ?Geometry .
    ?GeomID ?hasProperty ?Geometry .
    ?hasProperty rdfs:subPropertyOf ?Source .
    Filter ( ?Source = st:DMDataProperties ) .
}
```

Query result:

ID:	Geometry:	Source:
st:Trench-188	LINESTRING (177852.562 505219.298, 177844.972 505221.035, 177839.217 505198.462, 177848.197 505196.577)	st:Data PropertiesDM
st:Trench-188	MULTIPOINT (177852.562 505219.298, 177848.197 505196.577)	st:Data PropertiesDM

Competency Question 2.4:

Used query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?Objecttype ?Source where {
   ?ID st:hasUniqueID "Trench-188" .
   ?ID st:hasObjectType ?Objecttype .
   ?ID ?hasProperty ?Objecttype .
   ?hasProperty rdfs:subPropertyOf ?Source .
   Filter ( ?Source = st:DMDataProperties ) .
}
```

Query result:

ID:	Objecttype:	Source:
st:Trench-188	“Trench”	st:Data PropertiesDM

Competency Question 3.1:

Used query:

```
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?Matched where {
   ?ID st:hasUniqueID "Trench-188" .
   ?ID st:hasReasonForMatch ?Matched
}
```

Query result:

ID:	Matched:
st:ID-09	"Matched with RP dataset based on geospatial location (D<0,25)"
st:ID-09	“Matched with SAP dataset based on all object attributes”

Competency Question 3.2:

Used query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?Accuracy ?Source where {
   ?ID st:hasUniqueID " ID-09" .
   ?ID st:hasAccuracy ?Accuracy .
   ?ID ?hasProperty ?Accuracy .
   ?hasProperty rdfs:subPropertyOf ?Source .
   Filter ( ?Source = st:DMDataProperties ) .
}
```

Query result:

ID:	Accuracy:	Source:
st:ID-09	“10mm”	st:Data PropertiesDM

st:ID-09	“35mm”	st:Data PropertiesRP

Competency Question 3.3:

Used query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
select ?ID ?Object (geof:distance(?IDGML,?ObjGML) as ?Distance) where {
   ?ID st:hasUniqueID " ID-09" .
   ?ID owl:sameAs ?Object .
   ?ID geo:hasGeometry ?IDgeom .
   ?Object2 geo:hasGeometry ?Objgeom .
   ?IDgeom geo:asGML ?IDGML .
   ?Object2geom geo:asGML ?ObjGML .
}
```

Query result:

ID:	**Object:**	**Distance**
st:ID-09	“d75c7a9e-a899-48c1-8eda-8238b8673e3e”	“0,23”

Competency Question 3.4:

Used query:

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX st: <http://www.semanticweb.org/rakster1/ontologies/2023/0/DMO#>
select ?ID ?Date ?Source where {
   ?ID st:hasUniqueID "ID-09" .
   ?ID st:hasDateOfEntry ?Date
   ?ID ?hasProperty ?Date .
   ?hasProperty rdfs:subPropertyOf ?Source .
   Filter ( ?Source = st:DMDataProperties ) .
}  ORDER BY DESC(?Date)
```

Query result:

ID:	**?Date:**	**Source:**
st:ID-09	“30-01-2023”	st:Data PropertiesDM
st:ID-09	“21-12-2022”	st:Data PropertiesRP
st:ID-09	“22-12-2022”	st:Data PropertiesBID