GIMA

Geographical Information Management and Applications

Standardization of geo data exchange between network operators and contractors in underground utilities

MSc Thesis

Oscar Ribberink

Supervisors:

Prof. dr. Ir. P.J.M. van Oosterom (TU Delft)

Drs. R. Knook (Alliander)

June 5, 2017

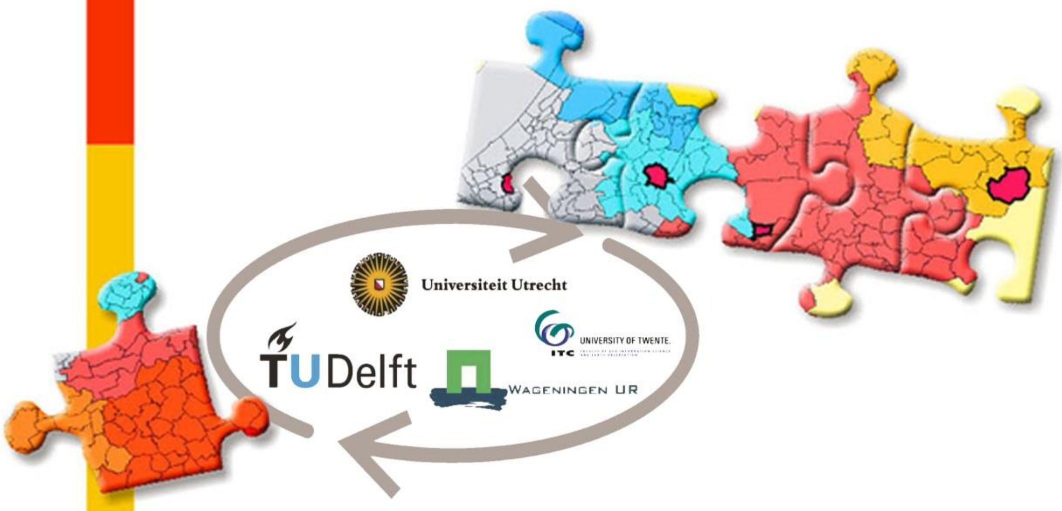

Table of Contents

Summary

In the utility sector the digitalization ‘in the trench’ set in only 5 years ago. Driving force behind it were the introduction of mobile smart device. These devices became unmissable with laws and regulations that demand the presence of lots of information about quality and work safety. The development speed of new applications for the mobile smart devices also plays its part.

In the Netherlands the major part of utility networks are buried underground. This makes them more resistant to weather and other threads, but brings other risks. Most important are the risks of excavation damage and invalid registration of the location of pipes and cables.

There has been several projects for standardization of information exchange of the utility network revision updates. Some of the standardization projects were successful, but in all the projects the standardization of Geo-Information was of less importance.

This research projects focusses on data exchange between network operator and the executing contractor in underground utilities. The scope of the research project is limited to the commodities electric power and natural gas and the focus is on residential service connections. Three scenarios help define the scope. Relocation pipes and cables for the creation of a roundabout, Adding multiple residence to the supply network and the last scenario is repairing excavation damage.

The research problem focusses on how standardization can help make information exchange in this application more efficient and effective. The standardization is twofold. A study is performed on what information the data model must comprise and the second part is to study the technical communication protocols. The goal is to create a model to enable standardized communication in a heterogenous world.

The research starts with a study into existing data models and communication standards. Several data models are compared and one of the data models, IMKL2015, is chosen as the main donor for the new data model. IMKL2015 is an INSPIRE compatible information model for cables and pipes in the Netherlands. The newly create model is a subset of IMKL2015 with some small additions to make it ready for asset information exchange. The communication protocol is based on the WFS-T protocol.

Also part of the project was the development of a prototype using recent web technologies. The prototype was used as Proof of Concepts and to evaluate the developed data model and WFS-T implementation. The evaluation is executed using the earlier introduced scenarios. The evaluation showed that the model is a useful, but that the model needs more research before adoption is possible.

The developed model is not perfect, but demonstration of the created prototype triggered interest and discussions by professionals in the field. It is those professionals who should take the next step by study the results ant conduct a business case analysis to prove the claimed advantages are a reasonable and feasible goal.

1 Introduction

From the beginning of Geo Information Science standardization has been one of its goals. For many government organizations the resulting standards enable publication and reuse of data on international, national and regional levels. In Europe this process of standardization is regulated by INSPIRE. Very soon the Dutch distribution network operators will also have to comply with the INSPIRE regulations and share information on their network assets using the INSPIRE standard.

The data exchange between the network operators and their contractors does not have to comply with the INSPIRE regulations, but there is the necessity and intention to standardize these information flows as well. Standardization of these information flows is a complex task, because within the Netherlands there are more than 30 distribution network operators with different corporate information systems and they supply several utility products: gas, electricity, district heating, water, cable and communication. There are more than 50 contractors involved. These companies have different ways of working especially when it comes to As-Built surveying of cables and pipelines. The information is handled by a large number of different software packages.

In the Netherlands the major part of utility networks are buried underground. This makes them more resistant to weather and other threads, but brings other risks. Most important are the risks of excavation damage and invalid registration of the location of pipes and cables.

The situation in the Netherlands is special because of another reason. The Netherlands have a very high percentage of residential connections to the natural gas supply network. In 1959 a natural gas reserve was found in Slochteren, which turned out to be one of the biggest natural gas reserves of Europe. The government decided that every Dutch house hold should be connected to the natural gas distribution network.

During that time there were a lot of mostly local or regional utility companies. Since than there has been merges into larger utility companies (and by regulation enforced division into network operators and energy supply companies). The history of the Dutch energy market has resulted in a divers landscape of utility companies and contractors. Each company has its own approach to reach its goals especially in case of Geo-Information management.

Several programs started to standardize information exchange between the operators and contractors. These programs focus either on process management or on geo information exchange. At both levels standardization is necessary. This research project focusses on both aspects and the hope is that the results of this project contribute to the process management and geo information management standardization programs.

The following two sections explain the context and relevance of this research project. Section 1.3 introduces three use cases that will help explain and scope the project. These use cases are also part of the research objects that will be introduced in the last section (1.4) of this chapter.

1.1 Context

In the previous section was stated that there is much attention to standardization in the field of distribution network operators. One of the reason for this attention is the Dutch WION law, which states that the data in the Geographical Information Systems (GIS) of network operators is supposed to be updated within 30 days after changes in a distribution infrastructure network (Ministry of Economic Affairs, 2008).

The WION was introduced to prevent excavation damage. All excavation projects must be announced, so that public utilities can send information on existing subsurface cables and pipes. This law and other laws have made the Netherlands a front runner when it comes to standardization and data quality.

Currently there are several projects being executed within the public utility sector which are trying to improve communication between the different parties. These projects can be grouped into top-down and bottom-up projects. They are either trying to get all parties in the sector to use the same standards or they work from the bottom-up starting with a few parties, trying to improve processing speed of the As-Built survey information on new or changed cables and pipelines.

An example of a top-down project was Digitale Rotonde (Digital Roundabout or DR). This national project included more than 20 distribution network operators. The goal of DR was to standardize the complete process starting from a consumer application for a public utility service connection to the billing and invoice handling at the end of the process. DR was succeeded by DSP (Digital collaboration portal) with a more bottom-up approach. Until now the focus of DR and DSP was on process management mainly. This is expected to shift to geo information management when the process management standardization is a more advanced stadium.

Another example of a national project was the DURF project (Digital exchange of cable and pipeline location data). It involved the 4 major distribution network operators for gas and electricity (Prins et al., 2013). Together they serve the large majority of the Netherlands. DURF was focusing on a shared standard for digital exchange of geo information. Process management was not within the scope of DURF.

There are also various bottom-up projects, were the focus is more on automation of processing steps than on standardization. The goals of these projects is to reduce the number of human actions needed to handle As-Built measurements and drawings. An example is the GeoDirect project (Geurts, 2013). GeoDirect is an ongoing project in the south of the Netherlands for exchange of standardized digital As-Built data. The aim of this project is to automate import of As-Built information from contractors automatically into the network operators' information systems. By mid-term 2014 all contractors have to deliver As-Built information (on home connections) in an XML based file format.

Due to the complexity and the number of stakeholders the top-down projects are bound to be long-term projects. The bottom-up projects have shorter lifetime, because they only focus on a few stakeholders. In these local projects only one network operator is involved per public utility product. This network operator decides what the exchange format is for his product (or products). The disadvantage of these local projects is that their solutions are not directly applicable to other network operators and contractors in other parts of the Netherlands.

There is a need for a project which studies the (intermediate) results of the different programs and combines their advantages in a new communication system. The goals of this new system must be to not only exchange standardized data, but also to create a standardized procedure for exchange of the data.

1.2 Relevance

If there is any evidence of the relevance it is the number and size of projects in this direction which in recent years have been or are being executed. A program for enabling standardization projects in this field is the CERISE-SG program (CERISE–SG, 2012). The goal of this program is coupling three different domains: Geo spatial standards, Government datasets and utility network standards (See Figure 1-1).

The relevance of this thesis project is related to the goals of the CERISE-SG program. When the goals of CERISE-SG are reached, the information on utility networks and Governmental Datasets like BAG (Dutch Key register for Addresses and Buildings) can be linked. This is only possible if at the level of data acquisition the datasets are linked or data transformation enables linkage.

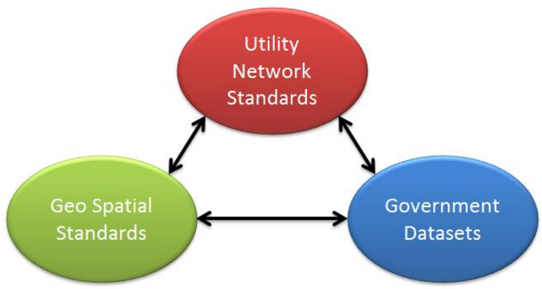

Figure 1-1 CERISE-SG program: linking domains
Source: CERISE–SG (2012) adapted

The link with Geo Spatial Standards is important as well. Government datasets can be distributed with Open Geospatial Consortiums (OGC) standard web services and in the previous section was explained that there is a need for standardized way for exchanging geo information in the acquisition phase.

An important related standardization program involves reference maps. The large scale topographic dataset of the Netherlands (Dutch: Grootschalige Basiskaart Nederland, GBKN) is currently the reference map for mapping underground utilities. This dataset is used as reference for planning and As-Built utility network information.

GBKN is in the process of being replaced by the Key register for Large Scale Topography (Dutch: Basisregistratie Grootschalige Topografie, BGT) and BGT will become the new reference model for utility networks in the future. The current state of technology used at the level of the contractors is not supporting the full benefits of the linkage. The gained knowledge in the CERISE-SG program and other projects can be used to improve technology in use, but research is needed to prove this hypothesis.

From the perspective of the contractor this may appear less relevant. Contractors are on the side of the data acquisition and generally work with flat data and sketches. Computerizing does have consequences for the contractor. If it is not accompanied by standardization it is possible that the contractor's field engineers in the future will have an application on his mobile device for each distribution network operator separately. Even more important is that contractors will have to keep up with technology, otherwise they risk being excluded from procurements in the future. When contractors have implemented a standardized exchange system this will increase implementation speed and reduce implementation cost when contracts with other utility companies are signed.

The social relevance can be found in the gains in the data registration speed and quality. The benefits range from reducing excavation damage to more efficient programs for management and maintenance services. Improved data quality also improve the applications that depend on the data, such as risk analysis and emergency response. There are also monetary benefits to be expected. When supply chain optimizations reduces network revision cost, this will lead to reduced costs for contractors and end users. Most network operators are 100% owned by governmental organization. An efficiency gain will also result in saving public sector money.

1.3 Use cases

An important part of this research are the use cases. In this chapter they will be introduced to explain the goal of the research and they support the scope definition and assessment of the end product. Three practical examples are chosen to link the theoretical research with practical situations. They are introduced this early in the report, because they are referenced by the research objectives.

1.3.1 Actors

As the title of this thesis project suggests, there are two main actors in this research: the network operator and the contractor. The use case diagram can be found in Figure 1-2. This very basic diagram presents the focus of this research project. It is simplified when compared to reality. In reality there may be more actors involved in a utility network project. There may for example be a third company responsible for location of assets or the work may be executed by a subcontractor. There are may be other companies involved in delivery of reference data. These other actors will not be part of this research project.

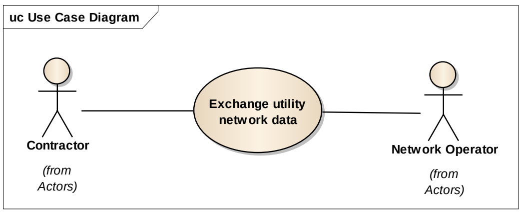

Figure 1-2 Research Use Case Diagram

Earlier was explained that in this thesis project the information exchange was focused on geo-information. Not all information exchange between the two actors is part of this thesis project. In this project we only take information of physical objects into account. All physical objects have a location. Administrative information like procurement information, work planning data and completion statement are not part of this research project.

In Figure 1-4 the basic sequence of delivery of geo information is shown. There are 3 basic data sets. The first 'As-Is' the current situation. The As-Planned dataset with the project plans. The contractors sends back the results of his activities called As-Built data.

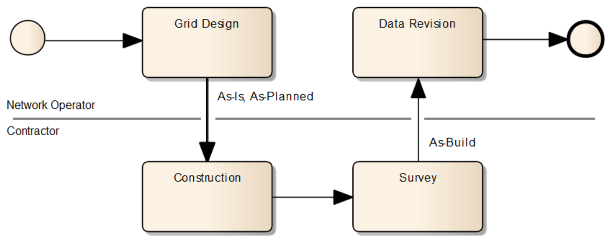

Figure 1-3 Network revision process diagram

The process is redrawn in a sequence diagram of Figure 1-4. At the moment the process in reality is far from optimal. This has several causes. On major cause for delay in the process is that the As-Built processing step of the network operator takes a lot of time.

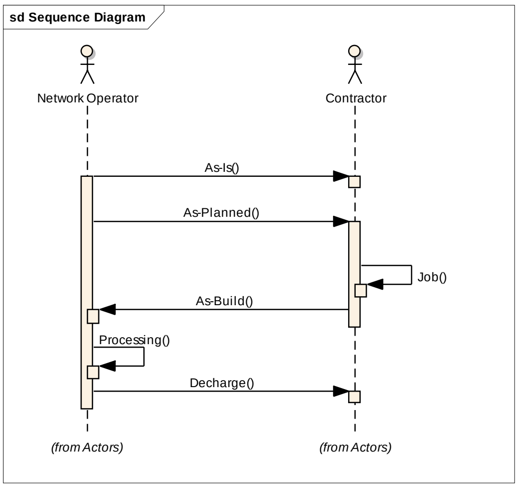

Figure 1-4 Prototype sequence diagram

Here is simplification applied as well. In reality more geo information may be send then just the current network configuration. Often a large scale topographic map is also send along with the network configuration. Depending on the scale of the work, these large scale topographic maps may act as reference model for location of network objects.

As a result of scope limitations there are only a limited number of objects defined that are part of this use cases: Cable, Cable joints, pipes, pipe joints and customer connections. With joints several network functions are possible: branche off, termination/seal or a connection between two pipes or cables.

Table 1-1 Most important object types

Objects
Cable
Cable joint
Pipe
Pipe joint
Customer connection

The objects will have a limited number of attributes. The attributes that are used during this thesis project will be limited by the attributes that are part of the current communication between model (and the earlier presented limitations).

Table 1-2 Most important attributes of pipes and cables

Attributes
(Geometry)
Material
ClampType
JointType

1.3.2 Use Case 1 Construction of a roundabout

Road infrastructure projects often raises the need for cables and pipelines relocation. The first use case is caused by the replacement of a road junction by a roundabout. In this use case the network operator develops a plan for new routes for the local cables and pipelines. The contractor of the project must have access to the plan drawings and he must be able to deliver the As-Built route and asset information when the project is finished.

Necessary objects: Cable, cable joint, pipe, pipe joint

1.3.3 Use Case 2 Connecting multiple consumers

The second use case is a combined order to add three properties to the gas and electricity network. The contractor will take on this job and the location of the main route of gas and electricity are available. During the job all changed objects are registered so that the location and attributes of all changed and new infrastructure assets can be sent to the operators.

Necessary objects: Cable, cable joint, pipe, pipe joint, costumer connection.

1.3.4 Use Case 3 Restore excavation damage

The third and final use case is an unplanned order. A technician is sent to a location to repair a reported leakage or broken cable. The technician sends a message with the location, to receive information about the location of the underground infrastructure. After repairing the damage, the technician can inform the operator exactly which portion of a pipe is replaced and what materials and joints are used.

Necessary objects: Cable, cable joint, pipe, pipe joint

1.4 Research Objectives

In the previous sections the importance and theory of standardization of exchange of information between contractors and operators was presented. In this section, the objectives of this thesis research project are discussed. First the research questions are defined followed by some statements to define the limits of the scope of the project.

1.4.1 Research Question

The central question for this research is:

How can geo data exchange between utility distribution network operators and contractors be standardized, both for information data models and for communication protocols?

The distribution network operators involved are the Dutch companies responsible for the infrastructure needed for the distribution of utility products of electricity, gas. There are more related utility networks, like district heating, water, cable and communication, but they are beyond the scope of this research project.

The contractor is the company responsible for execution of projects for extension or changes in the infrastructure. With geo data exchange is meant the network change plan drawings and As-Built drawings, so that the contractor knows what is expected of him and the GIS system of the network operators can be updated afterwards. The purpose of standardization is to provide a unified way of working and to enable automated exchange of geo information.

1.4.2 Sub Questions

The central research question is divided into four sub questions, which correspond to the phases of the project. The questions have been formulated to answer the central question step by step. There is a reference to use cases in the research questions. These use cases will be introduces in the next section.

1. ***Which standards are in use or available for use for data storage and data exchange in revision projects on pipelines and cables?***
2. ***How far has standardization come in case of standards for data storage and exchange in network revision projects similar to the use cases?***
3. ***How can these standards be used to create a prototype to standardize information data models and communication protocols.***
4. ***What is the result of a qualitative assessment of the created prototype for the use cases?***

1.4.3 Sub Question 1

To answer the first sub question research is needed that will focus on investigating relevant techniques and standards in general. It examines the current state of the information technology. It also looks at new developments and regulations in the sector.

1.4.4 Sub Question 2

This question could also be formulated by asking how far standardization has come already in practice. What is the status of current standardization projects. What is the interest of standardization for the various stakeholders? Are there conflicts of interest or policies which stand in the way of standardization? Which standards support communication related to the use cases.

1.4.5 Sub Question 3

When the previous questions are answered there must be sufficient knowledge to design and implement a prototype. How can the studied tools and techniques be combined to create a prototype for the exchange of geo information for the use cases?

1.4.6 Sub Question 4

With this final sub question the answers of the earlier question will be validated. Can the prototype be used for the proposed use cases. To what extent and under which conditions does the prototype solve the stated problem. What would be the impact of implementing the system?

2 Methodology

In the previous chapter the research objectives were introduced. In this chapter the methodology to reach the objectives is presented. This chapter starts with the introduction of the conceptual model (Section 2.1) and the scope (2.2). In the project plan of this research project the project structure was introduced. The project structure is recaptured in section 2.3. An extra section (2.4), especially for the design methodology, is introduced to explain the steps needed to ensure the quality assessment of the design. This chapter is finalized by the description of the excepted requirements on software tools (2.5) and (geo) data (2.6).

2.1 Conceptual model

Development of models or systems starts with defining the concepts and relationships. The objective of this research projects is to design a system for the exchange of geographical information between the network operators and contractors. An outline of the conceptual design is presented in Figure 2-1.

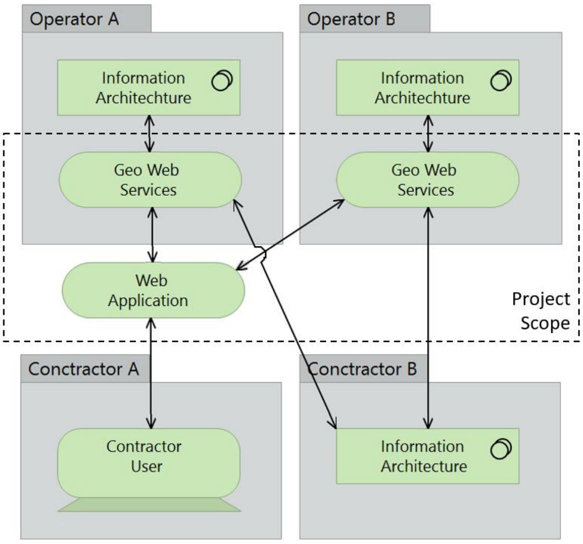

Figure 2-1 Conceptual model and scope definition

The information architecture of the contractors and network operators represent all the internal information systems needed by the organizations to function. This research project proposes to use Geo Web Services to integrate the information architectures of network operators and contractors. In this outline the network operators host their own geo web service.

The web services are standardized and interoperable to insure that contractors can use the web services of several operators in the same system or application. When this web services are hosted by the operators the contractors can integrate their information systems or use applications to connect to one or multiple geo web services. When the services apply common standards for data structures and exchange, the client applications of the contractor is software vendor and client platform independent.

The conceptual model also introduces a Web application. This application will be developed during this project to demonstrate, test and evaluate the design of the Geo Web Services. In the research objectives the prototype was introduced. This prototype will include one or more Geo Web Services and the web application. The design of the prototype has two major design component. The data model for information exchange and the technical communication model.

2.2 Scope

The scope of this project is largely based on the use cases. When the prototype can be functionally tested using the use cases, there should be enough information to evaluate the developed data model and technical communication model.

The focus of the research is on geographical information exchange. To make the prototype fully functional in real live there should be some sort of work flow management and planning involved. But in the prototype plan drawings and revision information is the main focus as well.

The number of network operators and utility products involved are very small. The proportions of the use cases are deliberately kept small as well. The impact analysis for implementing the system will not contain a quantitative cost-benefit analysis.

Connectivity is an important part of the network topology especially for network operators. Connectivity can be modeled in different ways and in different level of detail. End-to-end electricity connectivity will not be a part of the scope. It may be important for distribution network operators, but it will take too much effort and knowledge to include it in this project. Connectivity will still be an important subject and measurement of the

For the localization and registration of new cables and pipes several techniques from advanced GPS devices to tape-measure are used. In general for consumer connections the registration is measured relative to topographic elements like buildings. Through BAG the precise location of all Dutch buildings is public available. Together with relative measurements the location of the cables and pipes can be calculated quite accurate. The digital registration of this process is quite complex and is also beyond the scope of this project. For the success of this project only the registration of the absolute location of cables and pipes will be sufficient.

2.3 Project structure

In chapter 2 the research objectives were captured in a system architecture model. The next step is the development of a methodology for the development and implementation of the architecture. The methodology of this project consists of four succeeding work packages Figure 2-2.

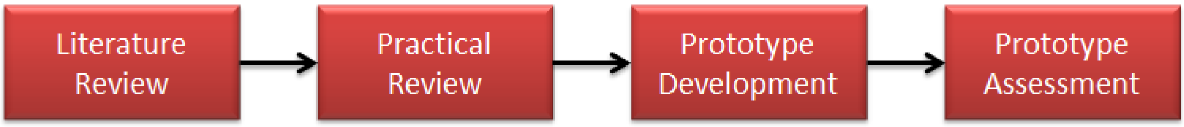

Figure 2-2 Methodology steps

The work packages cover the four sub questions as presented in the research objective chapter. When these work packages are completed successfully the central research question can be answered.

2.3.1 WP 1 Literature review

The major part of the first work package is a literature review. Relevant standards for data storage, transformation, exchange and visualization will be examined. This includes general Geo Information Science standards, industrial standards and software standards in use. The deliverable of this work package is a report on the relevant results of the literature review. It is presented in the next chapter.

2.3.2 WP 2 Practical review

This work package consists of an examination of the current way of working and the systems needed to do so. The use cases will be defined more elaborate using UML use case diagrams. This work package also includes a closer review of the current projects for standardization. Further will it analyze the needs of the stakeholders and analyze how existing tools can be modified or replaced.

The result of this work package is a report with the needs of different stakeholders. During this work package opportunities and risks for the project will be discovered. Recommendations for work package 3 are therefore an important part of the work package report.

2.3.3 WP 3 Prototype development

This work package includes the development of the information model for data exchange and the design and development of the system prototype. The development methodology will be chosen during work package 2. Given the limited resources an iterative development methodology is very likely. Testing of the product will be part of this work package if this is part of the chosen development methodology.

The outcome of this work package is a prototype system for demonstrating exchanging of geo data between two network operators and one contractor. The prototype will include a server and client application. The system must at least be able to exchange the plan and revision data for the three use cases introduced earlier in this report.

2.3.4 WP 4 Prototype Assessment

When the prototype is completed the system will be assessed using the use cases as presented in section 1.3. The results are presented to the stakeholders. An important part of this work package is a study of the applicability of the technology beyond the scope of work package 2 and 3. The last part of this work package is writing a business case to ground the financial feasibility of implementing the system. The conceptual design will be turned into a technical realization using a methodology loosely based on the agile development (Beck et al., 2001).

2.3.5 WP 0 Project management

The general project activities and deliverables are grouped under this work package. The deliverables include project meetings, writing progress reports, giving a mid-term presentation and the final thesis writing and defense. There is a change that based on the thesis report, there will be a request to publish the results in the Dutch journal Geo-Info. Therefore writing a Dutch paper on the results is also included in the last stage of the project.

2.4 Design methodology

The prototype that will be developed has two major design components with two different design methodologies. The first component is the data model for information exchange and the second component is the technical communication model. The methodology steps are alike but they differ enough to take a look at them separately. The methodology is summarized in Figure 2-3.

Figure 2-3 Methodology overview

The first step of the data model methodology is to gather the requirements of the data model. The source of the requirement are the use case, the literature review (WP1) and the interviews (WP2). The second step is to develop multiple data model designs based on these requirements. The most important design differences and their advantages and disadvantages will be described. The third step is to develop design options, compare them and to choose one of the designs options and to finalize the design in more detail. The last step is to evaluate the model by creating the physical model in the prototype and then to test for completeness (and redundancy) and to demonstrate the model to the stakeholders.

The second main component is the technical communication model. The first step of the design methodology of this component is to analyze the requirements by analyzing techniques and standards and which software must be supported. The following step is to analyze which tools and techniques are available to create the technical communication model. The third step is to choose the set of tools and techniques that will make up the model. The last step is to evaluate the technical communication model by creating the prototype and to analyze the result of the choices that were made.

2.5 Software

During the research project several software packages will be used for different purposes. Applications can be used because of their suitability and usability. In most ICT projects license costs are taken into account, but all software licenses are provided by the research institute. Sometimes user experience will also be taken into account.

These software packages needed for this research project can be divided into two groups. First the software packages that are needed to complete the research project. The second part of this section will discuss the software that may be needed by the prototype.

The software package that will be used extensively is Enterprise Architect by Sparx Systems. It is a UML analysis and design tool. It support several design modeling languages. It also includes tools to create technical solution/definition files from the designs or models.

Some application development is needed to create the prototype web application. An Integrated development environment (IDE) will be needed. This is usually and advanced Text editor which includes tools for scripting, deploying and testing applications.

Because the prototype will be a web application a web browser will be needed to test the application as well. Google Chrome will be used because it includes good developer tools to debug web applications and analyze web requests (and their response). The datasets must be transformed to fit the prototype. Depending on the complexity of the transformations ArcGIS Desktop by ESRI or FME Desktop by Save Software will be used. ArcGIS Desktop will also be used to test the compatibility of the prototype.

Prototype services

The web application will need a standard web server environment. The used software is less important, because the hosting will be done by an external provider. The properties of the hosting package will depend on choices made during prototype.

For the GIS web services installation, more software is needed. As explained earlier the interfaces with information architectures of the stakeholders are not part of the scope. The GIS web service will need geo data storage. GeoServer will be used to host the GIS web services. PostGIS enabled PostgreSQL is the GIS database server.

The prototype will include a web application to test the GIS web services. For compatibility testing ArcGIS and QGIS will be used.

2.6 Data

Earlier during the scenarios and scope definition was explained what kind of information is used in the project. The source of the data are the two network operators. Small subsets of the datasets are created and transformed to the prototype data model.

There is a risk that the data received by the network operators is not usable or the process to transform the data to the data model is to complex. There is a backup data source available in the open asset data set of Enexis (https://www.enexis.nl/over-enexis/open-data). This is currently a set of ESRI Shapefile with INSPIRE attributes.

Publieke dienstverlening op de kaart (PDOK) is the Dutch National SDI, a central facility for unlocking geodatasets of national importance (PDOK, 2017). Recently the BGT is added as free to use dataset. This is and other PDOK datasets will be used as reference data.

.

3 Literature

In the previous chapters the research objectives and methodology were introduced. The next step according to the methodology is the theoretical background. This chapter introduces basic knowledge on concepts and standards. This chapter is divided into two parts to serve the two main themes separately: Designing the system for communication and messaging (Section 3.1) and designing the data model (Section 3.2).

3.1 Web information exchange

To enable information exchange the parties must agree on a communication system. A communication system can be developed using a data model standards and a message model standard. The current research projects mainly focus on the data model level. Developments in information technology and geographical information science enable a potentially more efficient way of enabling the development of a common message model for exchange of as-built data.

3.1.1 Web communication architecture

The classic way of modeling the power behind the internet is through the OSI Model (See Figure 1-2). In this 7 layered model all basic web techniques are presented.

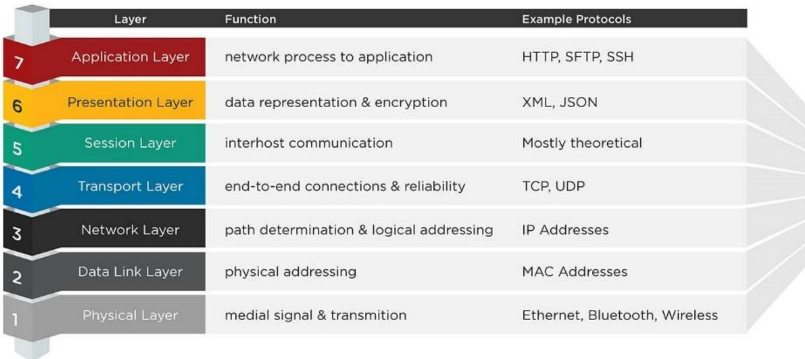

Figure 3-1 OSI Model for web communication. Image source: (Wordfence, 2015)

With the arrival of cloud technologies as web services, Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS), the OSI model became less visible. It is still the backbone of web communication, but web application developers do not have to worry about all the layers any more. Although the model became less visible in modern web technologies, it still helps to show the of certain standards for web technology. Most of the layers are far beyond the scope of this thesis, but the presentation layer and the application are important concepts as will become clear in the remainder of this chapter.

3.1.2 Data representation

In the presentation layer (6) of the OSI model is defined how data is represented, in other words what file formats or encryption is used. Because the goal of this research project is to design a communication system the logical choice would be to use Extensible Markup Language (XML).

XML and GML

The XML standard (W3C, 2008) is widely accepted as the base data structure for data exchange between information systems (IS). XML is a self-describing text based file structure. Self-describing means that file-structure of the documents also can be described using Schema definitions (W3C, 2012) that as well are encoded in XML. Even transformation between XML files can also be stored in XML using Extensible Stylesheet Language Transformations (XSLT). These properties and more advantages make XML a reasonable choice for information exchange.

JSON

In recent years JSON (Internet Engineering Task Force., 2014) gained popularity as alternative to XML as data format for information exchange. Critics say XML is less efficient than JSON (Nurseitov et al., 2009), because it takes more processing time and needs more overhead information. But especially in complex information exchange application XML advantages outweigh JSON compactness. The best XML ability for validation by XML Schema Definitions (XSD).

3.1.3 Web protocols

The most used protocol in the application layer of the OSI Model is HTTP (Hypertext transport protocol). It is well known for delivering webpages to internet browser, but it is also the basis for other web services used by other application.

Web Services

The term web services is a broad used term, but in this research project it is a method for integration of applications and systems based on XML and Web standards (Haas & Brown, 2012). It is a widely used technique for information exchange between information systems.

Web services are often used for point-to-point communication between information systems, but they also make it possible to develop multi-tier information architectures with Service Oriented Architecture (SOA) methodology. Instead of the classic 2-tier client-server architecture, more tiers are added. The information system is divided across different system with their own function. The most common application is the creation of a presentation tier, a business logic tier and a data storage tier, but the possibilities are endless.

3.1.4 Geospatial web protocols

The Open Geospatial Consortium (OGC) is an organization for standardization and interoperability of systems based on geospatial information. The OGC is supported by industry, government and academic members.

An important group of standards from the OGC are the OGC Web Services (OWS). This group of web service standards enable the exchange of geospatial information through web services. OWS are mainly used for one-way publication of geospatial information, but amongst other applications, it also enables feature editing and remotely invoking geospatial processing algorithms.

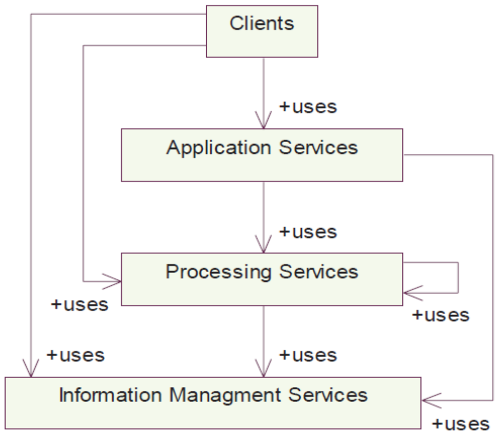

Figure 3-2 A possible tiers separation for an OWS architecture. Source: Open Geospatial Consortium Inc. (2005)

The most applied OGC Web Services are the Web Map Service (WMS), the Web Feature Service (WFS) and the Web Processing Service (WPS). WMS is mainly used for mapping applications, WFS is used when feature / vector data exchange is needed. WPS enables remote execution of geospatial algorithms.

The different OWS types can be combined and used to create or support one application architecture. When such an architecture is separated into three tiers (Figure 3-2), WFS is on the Information or data management level (feature/object handling). WMS is used for portrayal of geo data and thus part of the application tier. WPS is part of the processing tier in this model.

The Geography Markup Language or GML (Open Geospatial Consortium Inc., 2007) is a XML-based markup language for geographical information. The most used application of GML is the modelling of geographical features. It enables the development of systems for exchange of geo-information.

GML is used by WFS for the exchange of geospatial features. Most implementation of WFS are only to query objects, but the WFS protocol also includes the ability to create, update and delete objects. When these functions are implemented service is called a transactional Web Feature Service (WFS-T). The WFS-T protocol includes feature locking functionality. This system can prevent conflicts when multiple users are using the service.

Web services are examples of the classic server-client architecture, but there are two properties of OWS which enable them to make them part of modern SaaS applications. The first property is chaining. Web services can be used by other web services and clients can use data from different sources to create advanced applications.

3.2 Data modelling

In information exchange deciding on the file format is not enough to ensure efficient data exchange. There also has to be an agreement on what objects are exchanged and how objects are described. Data modelling is the process of creating an abstract model of real-world phenomena for a particular purpose. This process usually starts with a high level of abstraction with a conceptual model. The goal of conceptual modelling is to identify the objects that need to be modeled and which object relate to each other. The next abstraction level is logical data modelling: The unique identifier of objects and all needed attributes of the objects (and the relations to other objects) must be described. The last stage of modeling is physical data modeling. This is the technical description of the data encoding. Depending on the purpose of the model, this can result in database structure designs or xml schema definitions.

3.2.1 UML

The sections following this one will discuss existing data model and how data model can be derived from each other. To comprehend this section some basic knowledge of data modelling terminology and UML is needed. Some readers may want to skip this section.

An important tool for modelling is the Unified Modeling Language (UML). It enables system designers to create graphical representation of the functional or technical model for a software application for example. UML consists of many structure and behavior diagram type (Object Management Group, 2003). In this research project the class diagram, use case and sequence diagrams are used.

The most used diagram is the class diagram. All classes (object types) are represented by rectangular shapes and associations between them by arrows (Figure 3-3) . The most basic form of association is dependency.

Two other cases of association are aggregation and composition. Aggregation is the 'part-of' relationship (e.g. wheel to a car). Composition is the 'consists-of' relationship (e.g. room to a house). Composition implies that an object of the child class cannot exists if the parent class does not or stops existing. Aggregation and composition are forms of containment.

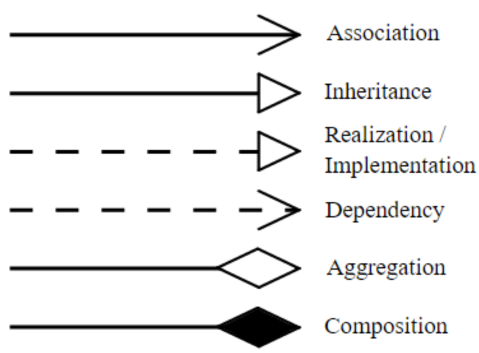

Figure 3-3 UML Association types

Inheritance is the 'is-a' relationship between classes (e.g. cat to an animal). Inheritance implies that the child class inherits all properties (e.g. associations, attributes and operations) of the parent class.

There are two important techniques to improve the quality of a class diagram: generalization and specialization. Generalization is the process of extracting shared properties from several classes and combining them into a superclass. Specialization is the creation of subclasses from an existing class. Some properties of a superclass may not be needed for all objects of a class. Abstract classes are classes that exist in a data model, but which are only needed because other classes inherit its properties. There will be no objects of this class in the physical form of the model.

The result of data modelling is a set of classes and their relations with each other. The classes are often grouped to a domain or package. It is also possible to reuse classes from other domains or packages.

3.2.2 Modeling phases

The process of data modelling is usually split into different phases. In the first stage only the objects and their relations are defined. This stage is called conceptual data modeling. In the next stage attributes and methods are added resulting in the logical data model. In the final stage information is added for encoding or storage of the data. The physical data model contains enough information to create the needed database table or how an XML message is build up.

3.2.3 Data model re-use techniques

Re-use of existing data models has great advantages. During data model development it takes a lot of time and effort to define and describe the meaning of objects and their associations. Acceptance of existing data models or data models based on existing data model is more likely.

There are different techniques for reuse of data models. Classes can be substituted by creating a new class in the new domain which inherits all the characteristics of the donor class. If necessary this class can be extended to suit the needs of the new domain by also associating local classes for example. It is also possible for an object to inherit properties from two or more different classes even from different domains. This is called Mixin. The term Application domain extensions (ADE) is often used when a data model is extended for a certain application.

3.2.4 Data model architecture

Data models are often designed for a specific purpose or domain, but some data models are created to become standards in a specific field. These data models are used in multiple domains or contexts. It is possible to create subsets of a data model for specific applications. These subsets are also called be context profiles. Using tools like Enterprise Architect can be used to create the physical data model or message syntax from a context profile.

3.2.5 Data model role models

Three existing data models are investigated to find out their suitability for reuse in this project. The three data models have three different backgrounds. The IMKL data model is mostly used by contractors, the CIM framework is used by stakeholders in the energy market. CityGML (Open Geospatial Consortium Inc., 2012) has its roots in Geo-Information science. Depending on the background of a stakeholder a good guess can be made what kind of data model he or she will be accustomed to. In following sections these three data models will be explained.

3.2.6 IMKL2015

To reduce the number of excavation damage incidents, the Dutch government introduced a law in 2008 (Ministry of Economic Affairs, 2008) that all excavation must be announced so that public utilities can send information on existing subsurface cables and pipes. Excavation may not be started unless this information is on site. With this law a new information model was put into service called the Information Model for Cables and Pipes (IMKL, Dutch: Informatiemodel Kabels en Leidingen).

The first version of IMKL the information was exchanged in raster format, which makes reuse of the model for other purposes almost impossible. Currently this system is being modernized in the KLIC-WIN program (Kadaster, 2016). The purpose of this project is amongst others to introduce feature based communication. The used data model IMKL2015 (Geonovum, 2017) can also be used to generate INSPIRE (European Commission, 2016) compliant data sets and to comply with other open data regulations (Figure 1-2).

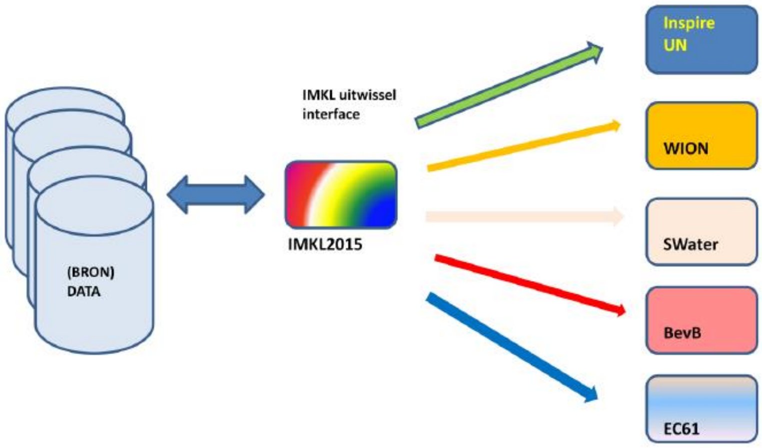

Figure 3-4 IMKL 2015 Data model can be used to comply with several open data regulations

In the previous sections several techniques were introduced for data models reuse. The IMKL2015 standard uses almost all off of them. It is basically based on two data models. The representation of physical network features is based on the INSPIRE data specification for the theme Utility and governmental services. The other administrative information was already stored in the predecessing IMKL data model. The core objects of IMKL 2015 use multiple inheritance to combine these two data models into an application extension (Figure 3-5).

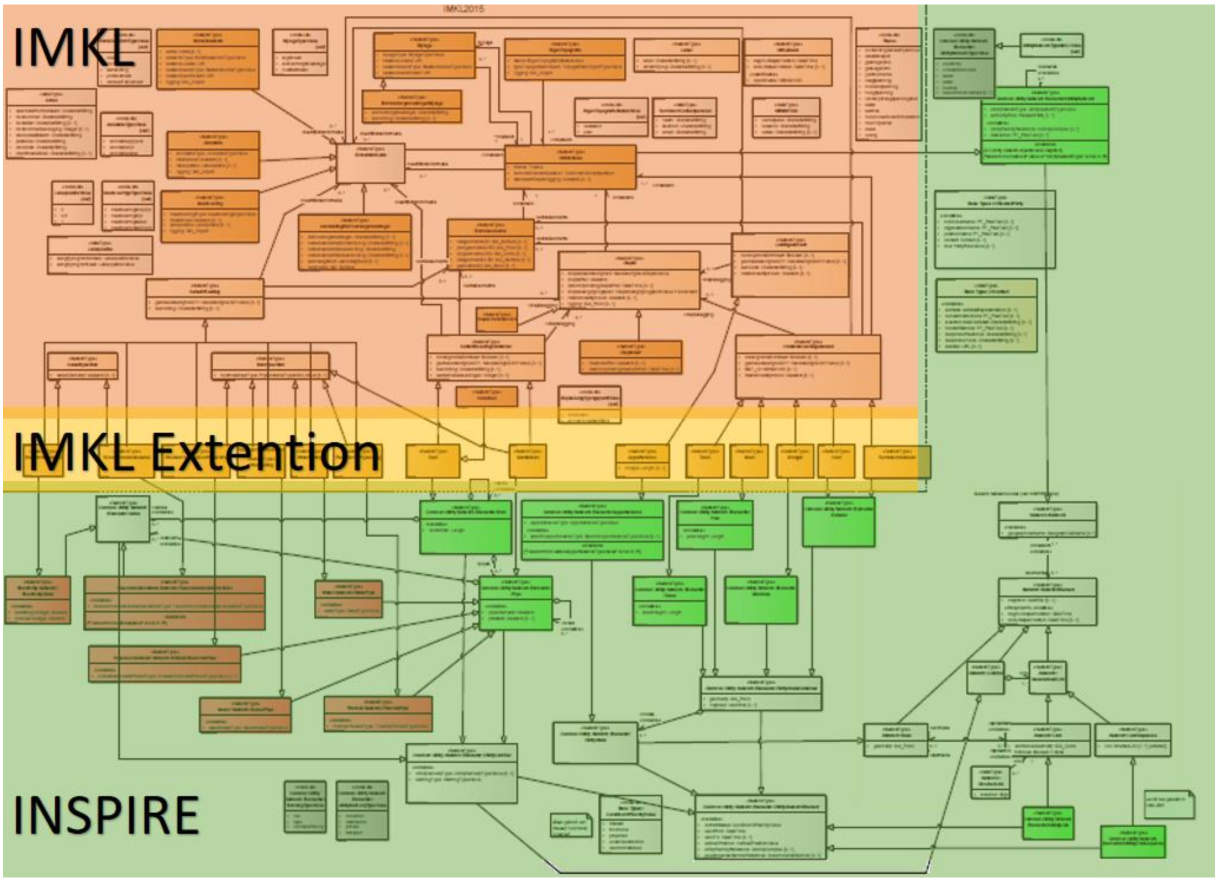

Figure 3-5 IMKL 2015 uses multiple inheritance to combine data models

To elaborate on this we zoom in to the Electricity Cable (Dutch: Electriciteitskabel) object diagram, which is presented in Figure 3-6 on page 20. The Electriciteitskabel object has three Superclasses. The first is

ElectricityCable from the INSPIRE data model. The other two, KabelOfLeiding (English: CableOrPipe) and KabelSpecifiek (English: Cable Specific), are from the IMKL data model.

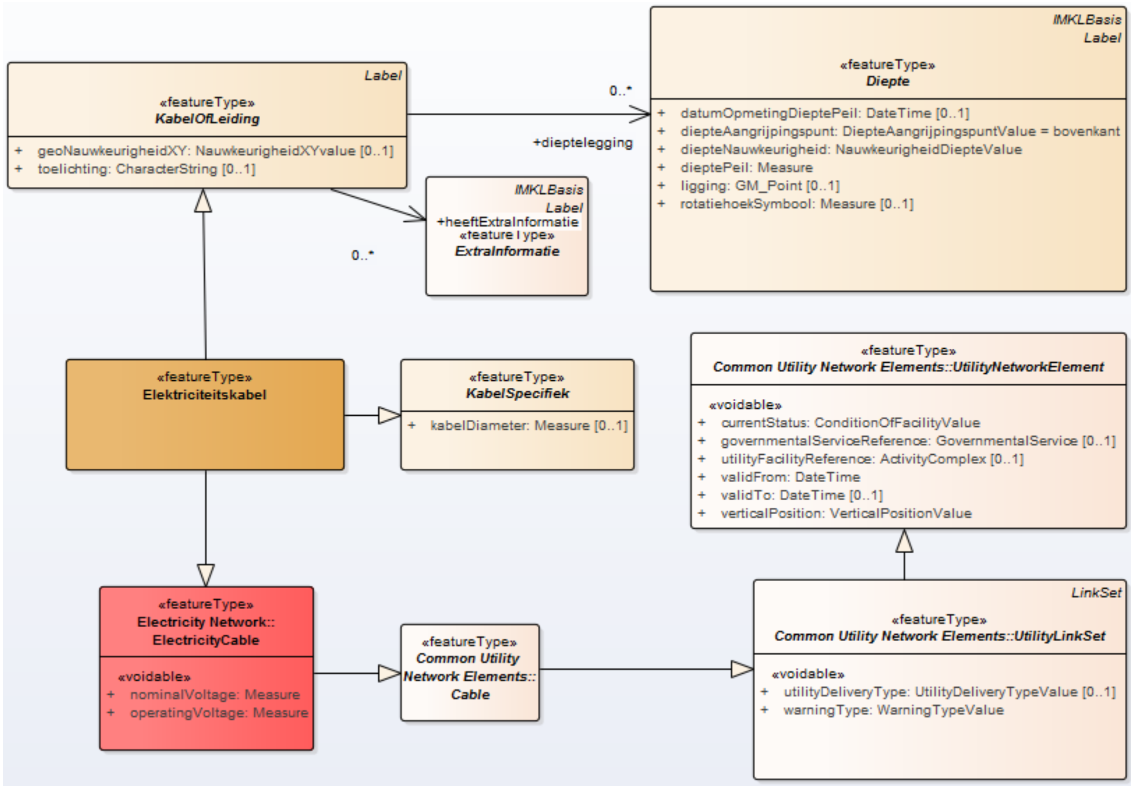

Figure 3-6 Extract from IMKL data model: Elektriciteitskabel (Some objects are hidden for readability purposes)

The geometry of all network elements is inherited from the INSPIRE network model (Figure 3-7) through the LinkSet element. All network elements are part of the topology and described as nodes and links. The connectivity of elements is based on arc/node topology. If the endpoint of a cable network is a building, which is described as by a Node. The cable may end at the wall of the building, will the point feature of the building lies in the center of the building.

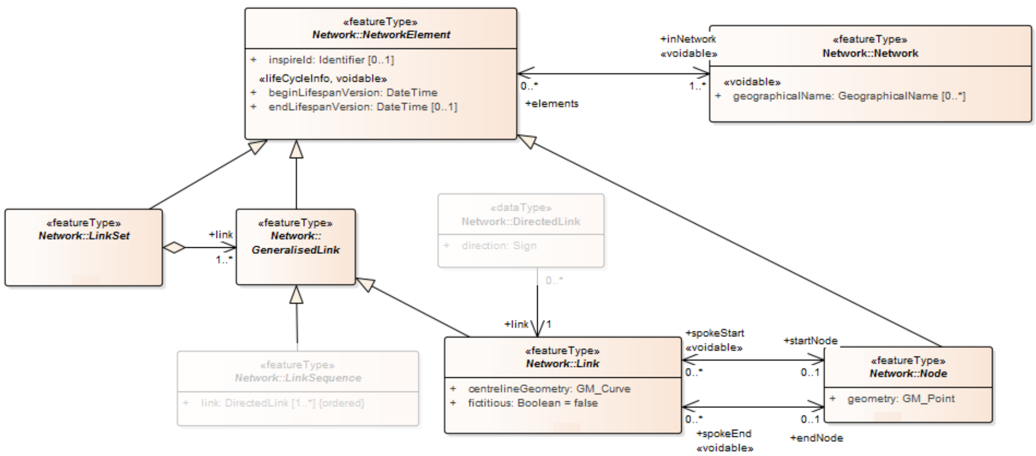

Figure 3-7 INSPIRE Generic Network Model

All Line features in IMKL2015 are formed by the LinkSet element which is a set of GeneralisedLinks. A GeneralisedLink object can be a Link or a LinkSequence of DirectedLinks. The lather are not used in IMKL2015 and therefor grayed out in Figure 3-7. In INSPIRE multiple use of the same Link for different NetwerkElements is possible, but in IMKL2015 it is not allowed.

3.2.7 CityGML

The second information model that is introduced is CityGML (Open Geospatial Consortium Inc., 2012). This information model is designed to represent urban objects in 3D. It is an application schema for the Geography Markup Language (GML). In 2008 it was adopted as official OGC Standard and it is widely used by a big variation of organizations. In the Netherlands, CityGML is used for the earlier introduced BGT, the new national map for large scale topography. This amongst others shows the relevance of CityGML.

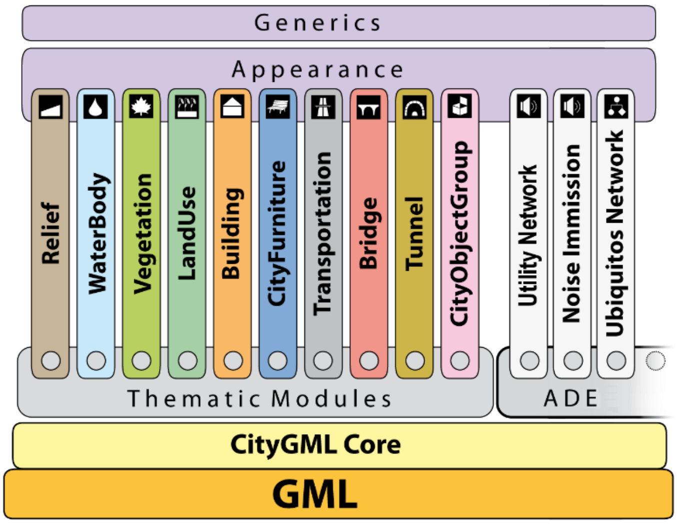

Figure 3-8, Schematic view on CityGML base component (Adapted from virtualcitysystems.de)

The CityGML Core (see Figure 3-8) were a common CityObject is defined. This CityObject is Abstract Class for classes in different thematic models and application domain extensions (ADE). One of those ADEs is the Utility Network ADE. This extension introduces network nodes edges. Figure 3-9 shows the network model of CityGML ADE. It is inspired by INSPIRE Generic Network Model (Figure 3-7).

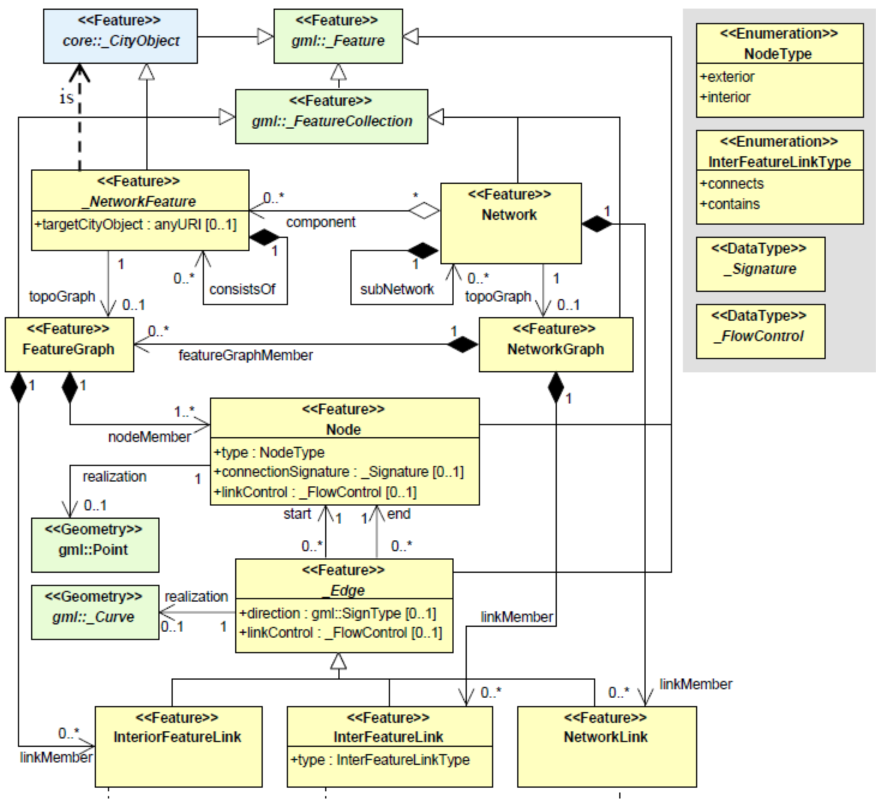

Figure 3-9 CityGML UtlityNetworkADE network model. Source: (Becker et al., 2013)

For many users of CityGML, the rendering of a CityGML model is very important. CityGML offers the possibility to render an object in different levels of details. From 2D features (Lod0) to textured 3D shapes even interior (LoD4). For this research project the rendering is of less importance. More important is are the object classes and how network topology is modelled.

An advantage of this model is that all type of public utilities can be modelled using the same basic objects (Figure 3-10). The network model is focused on network topology and the product or commodity of the network is of less importance.

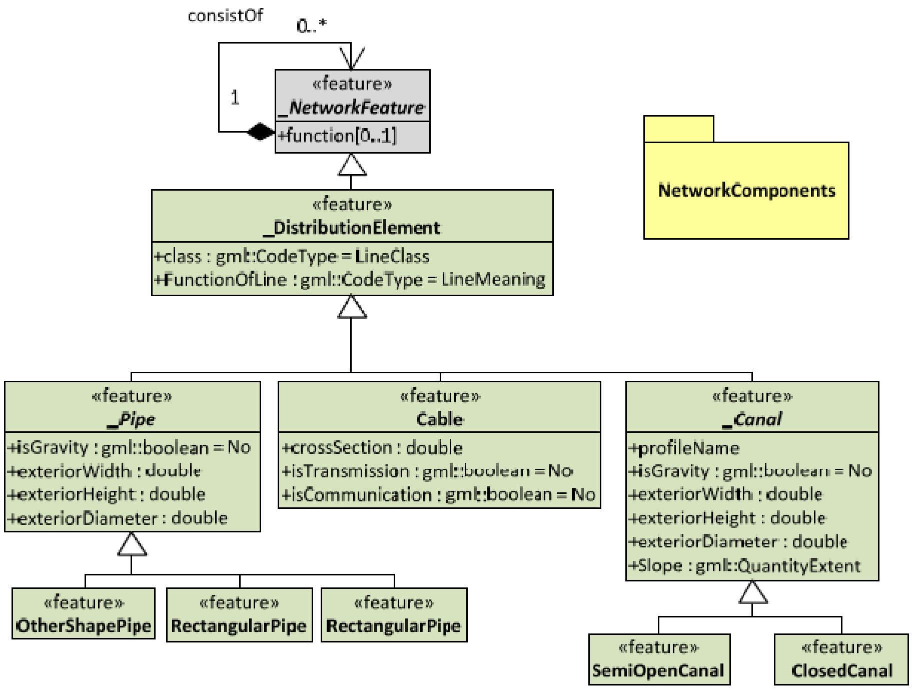

Figure 3-10 CityGML Utility Network object classes. Source: (Becker et al., 2013)

The commodity or product distributed by the network is modelled in the transportedMedium association. The transportedMedium object has 4 subclasses which are presented in Figure 3-11.

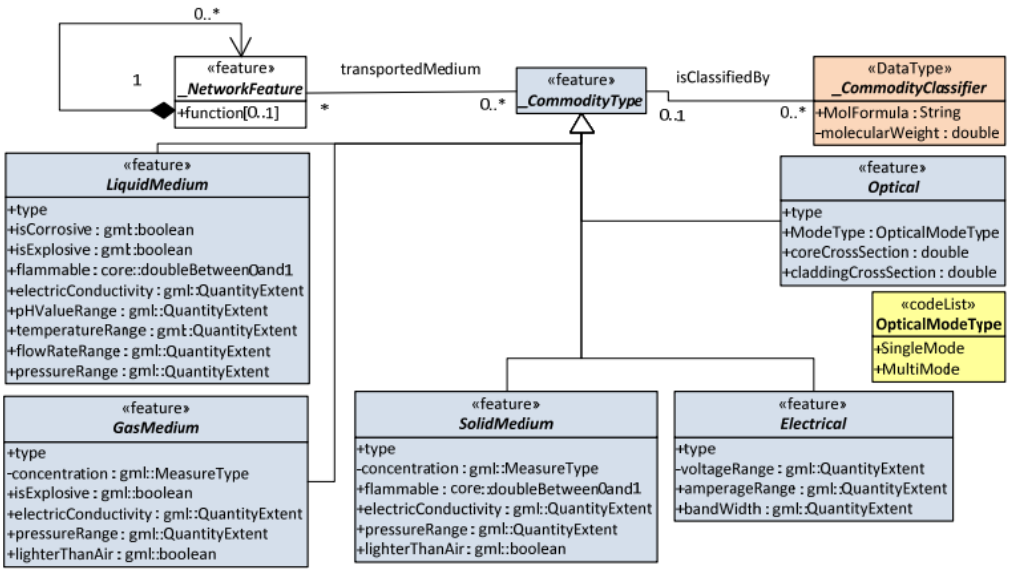

Figure 3-11 The commodity class types. Source: (Becker et al., 2013)

Previous research has shown that using the Utility Network ADE it is possible to store utilities in CityGML. The focus of the data model is on network topology first. The last 5 years there is less research on the Utility Network ADE. Recently the CityGML community is more active again now that the CityGML 3.0 is being developed. The plan is to intergrate UtilityNetork ADE as a module into CityGML 3.0 (Kutzner & Kolbe, 2016). These developments are not part of this project.

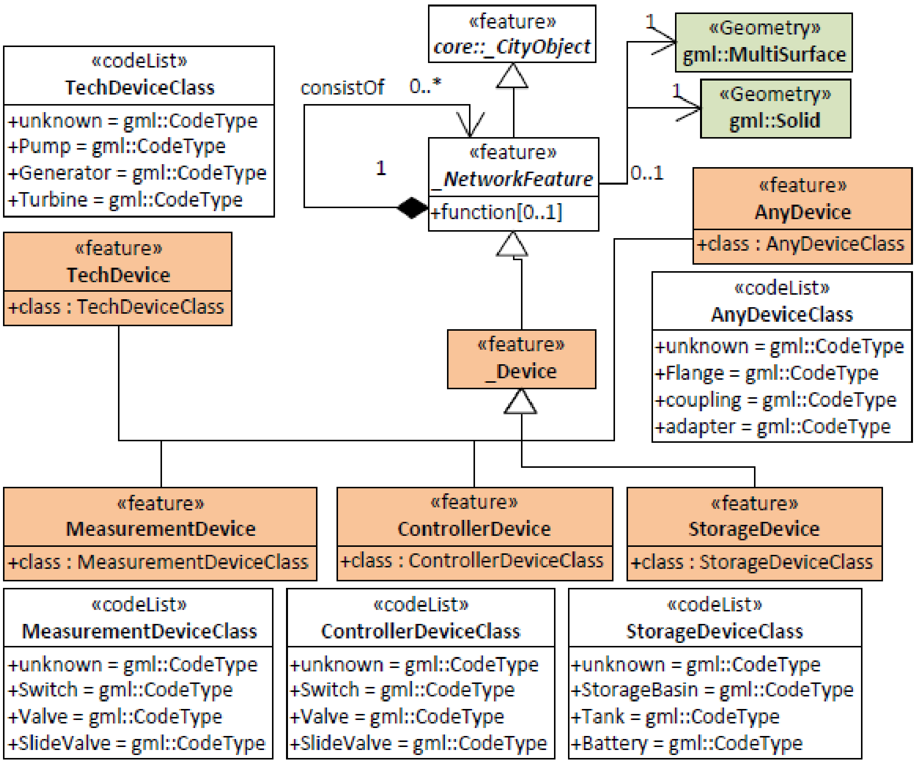

Figure 3-12 CityGML NetworkUtilityADE's Network Devices Overview. Source: (Becker et al., 2013)

Most nodes of the utility network can be modeled by the abstract class _Device (see Figure 3-12). To give an example: When during physical revision of the network joints are needed, these joints can be modelled using an AnyDevice object of type ‘coupling’. The last type of NetworkFeature of interest is the TerminalElement element. This element is used to model the end or service connection of the network.

The CityGML Utility Network ADE objects that are of interest to this project are: RoundPipe, Cable, ControllerDevice, AnyDevice and TerminalElement.

3.2.8 CIM

The final data model that will be discussed is the Common Information Model (CIM). CIM is not an information model like the previous discussed data models. CIM is a conceptual framework to create conceptual data models. There are hundreds of CIM classes with thousands of attributes. If one wants to understands all the concepts off CIM it would take a very long time. It is out of the scope of this document to explain all the concepts of CIM. Only the concepts needed for this project will be explained here. For interested readers that want to know more about CIM, the CIM Primer (Electric Power Research Institute, 2015) is good place to start.

CIM is maintained by the International Electrotechnical Commission (IEC). The Technical Comity 57 to be precise. The CIM is a collection of standards of which two can be seen as the core. IEC standard 61970-301 describes the objects of electrical power supply system and the object relations. IEC standard 61968-11 extends the core object description with data exchange and distribution information.

In Figure 3-13 is visualized how CIM can be seen as a layered architecture. The core objects and semantic model are defined in the top layer. Using profiles a specific business or application context can be defined. In the bottom layer this context can be encapsulated into message syntax for information exchange.

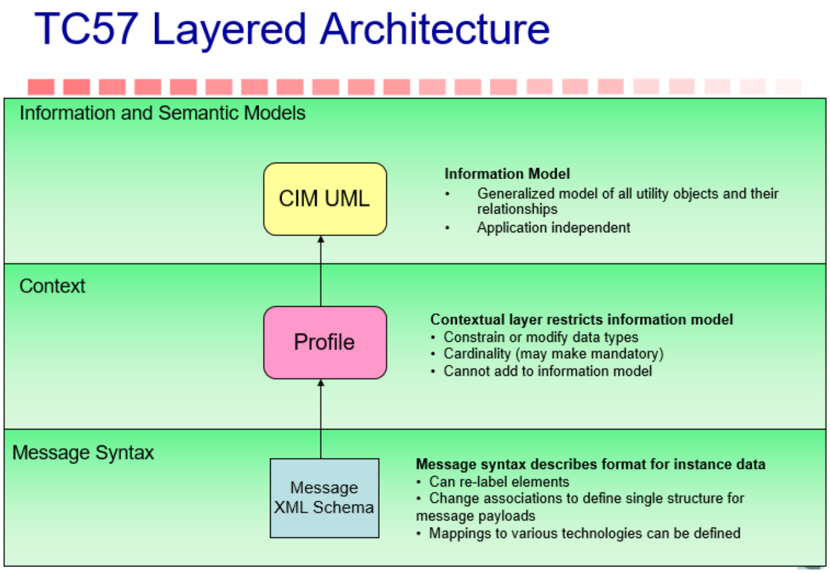

Figure 3-13 The layered architecture of CIM. Source (CIM Users Group, 2010)

The basic objects that make up an electric power network in the data model are presented in Figure 3-14. All objects that make up the physical network share the properties of the abtract class PowerSystemResource. The master abstract object class in CIM is IndentifiedObject.

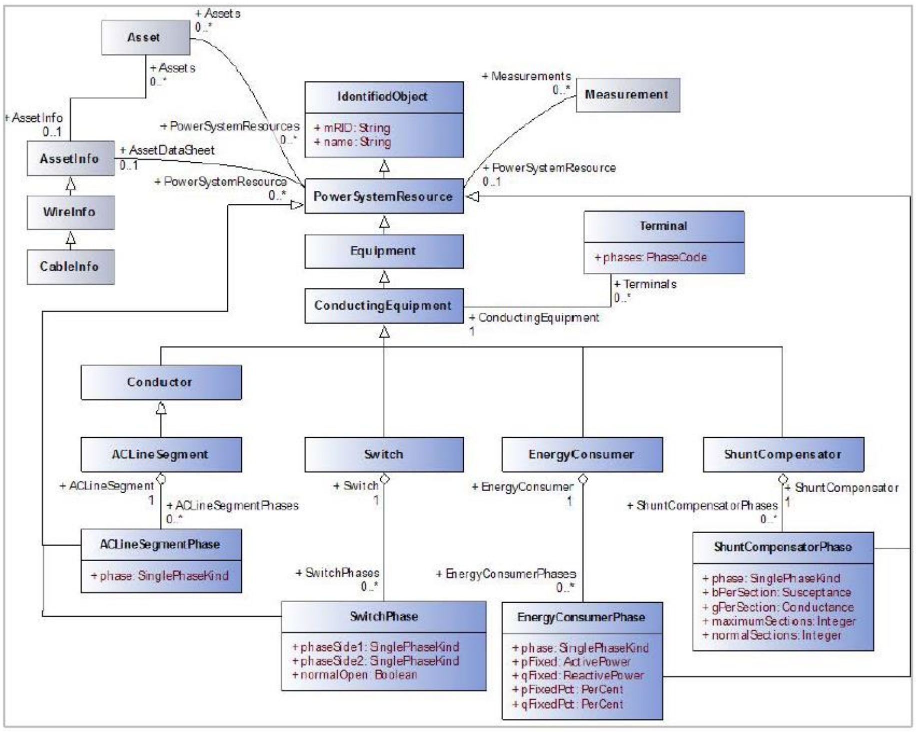

Figure 3-14 CIM basic objects

From this object diagram can the most important objects of the network can be identified. The ACLineSegment, Switch, EnergyConsumer are identified as the objects that can be used in this research project. During the design of CIM one of the objective of CIM was to be able to model the complete utility network from power station to power consumer. It is widely used as information model for information exchange, but mainly between systems within companies not between companies.

The core objects of CIM are presented in Figure 3-14. This is only a small excerpt of the hundreds of CIM objects available. It has to be said that the majority of objects is not for modelling physical objects of network assets. In Figure 3-14 the Asset object is of particular interest. When a device in the network is replaced by another device with the same function. The network layout does not have to change. Only the asset specific information will be replaced.

Connectivity between objects in CIM are applied by using the Terminal and Connectivity class. All objects that inherit from the ConductingEquipment class can have Terminals, which can be connected to ConnectivityNodes. These objects are used to create the power network topology.

The geographical information of Assets is stored in the Location and PositionPoint objects (Figure 3-15). Both PowerSystemResources and Assets can have a Location associated with it, because when an Asset is not part of a network, it still has a location (in a warehouse for example).

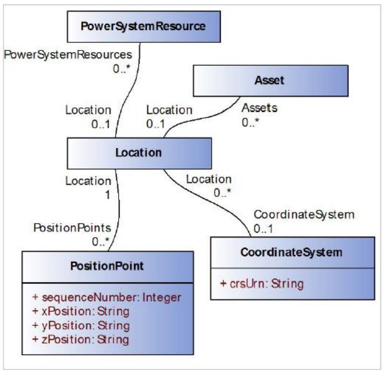

Figure 3-15 CIM Object diagram with Location object to store geographical location

The GIS component of CIM is not very complex. The only feature type that can be represented in Cim are point features or lists of point features. The location of cables is not part of the information model.

CIM is a versatile framework for data modelling and information exchange. It was developed in North America but is used extensively around the world. It is built for reuse and there are a lot of tools to create a CIM based data exchange model. CIM can not only be used for the representation of physical network objects but also administrative objects.

A major disadvantage of CIM is that is designed for power supply networks and not for other commodity types. In the Netherlands most public utilities for electricity also supply natural gas. These companies have extended the CIM standard to create a data model which is able to serve natural gas supply network.

3.2.9 Reflection

The complexity of the CIM framework is the result of years of development by the electric power industry. It results in a usable model but only for the electric power networks.

In recent years, the Dutch government made progress in the field of standardization. With the ongoing implementation of the Dutch key registers a great efficiency gain is reached, because different government bodies use each other’s datasets (Geonovum, 2015). The principle of register once and use multiple times applies. The large scale topographic dataset of the Netherlands (Dutch: Grootschalige Basiskaart Nederland, GBKN) is a base map for mapping underground utilities. This dataset is used as reference for planning and As-Built utility network information. GBKN is in the process of being replaced by the Key register for Large Scale Topography (Dutch: Basisregistratie Grootschalige Topografie, BGT) and BGT will become the new reference model for utility networks in the future.

3.3 Visualization

Visualization is a prominent part of Geographical Information Science. With Object-oriented data exchange, the attribute information and the visualization of objects are separated. For example, someone who uses a WFS, can decide how a line element is visualized in his application. Standardization can ensure that a line of a certain properties looks the same everywhere. Not all utility companies have the same visualization rules, but it is possible to create different rule sets. The discussion if utility companies should harmonize visualization rules is not within the scope of this document.

Style Layer Descriptor

Style Layer Descriptor (SLD) is an OpenGIS (Open Geospatial Consortium Inc., 2005) standard to specify the appearance of features on map layers. It is designed for use with WMS, but with SLD it is also possible for WFS clients to apply the symbology rules to WFS layers. Since SLD version 1.1 the language for symbology encoding (SE) is separated from the implementation schema.

4 Data model results

The second chapter explained the methodology of this research project. This chapter will contain the data model results for this research project. It will start with describing how the requirements are gathered. These requirement are used to create three preliminary data model options. After evaluation of quality and comparison of the options, one of the data models is chosen to become the data model for this project. This will be turned into an logical data model.

4.1 Requirements

The next section (4.2) will describe the data model options. These model options must be compared using requirements. From the use cases a list of requirements is abstracted. This list will be discussed in this section.

Class completeness (RD1)

From the use case discussion functional network components can be derived that need to be modelled. The most important components are the summarized in Table 4-1. The functional components may be grouped in the data model, if technical component can have the multiple functions. The related scenario and network feature type are also presented.

Table 4-1 List of required functional objects

Description	Scenario	Network feature type
Main cable/pipe	S1,S2,S3	Link
(Service) supply cable/pipe	S2,S3	Link
Cable branch off joint, pipe tap saddle	S2	Node
Repair joints	S3	Node
Cable/pipe termination	S1	Node
Supply point	S2	Node

Relations and attributes completeness (RD2)

A detailed description of the relations and the attributes is not part of the conceptual modeling phase. Nonetheless an examination off the attributes and associations between the objects will be executed.

GIS Compatibility (RD3)

One of the primary objectives of the data model will be to represent objects and their geolocation. The model must support the modeling of geoinformation according to geo standards.

Extensibility (RD4)

Because the requirements for this small research project only cover the use cases a requirement is added for the extensibility of the data model. The intended extensions may cover extra classes for other utility product for example fiber optics or water supply networks.

Simplicity (RD5)

The model should not be more complex than necessary. The end user must be kept in mind. There must not be more classes and attributes than needed.

Model well supported/accepted (RD6)

Further research and adoption of the model will benefit if the model is already supported, accepted or even implemented by experts.

Reuse of international standards (RD7)

When international standards like INSPIRE are reused, better integration with other standards is expected. Compliance to OGC standards and ISO/TC 211 is for many organizations a must.

Easy integration with other models (RD8)

The possibility to integrate the model with other models will become important when the reference model like BGT will become available.

4.2 Model options

Creating data models takes a lot of effort especially when many stakeholders are involved and want to take part in the process of modelling and decision making. Implementing the data models and creating transformations are time consuming as well. Reuse of existing data model for different purposes saves time and money. Less discussion about the standard is needed, because for existing data models the meaning of elements is usually documented well.

In the previous chapter some existing data models were discussed. In this section subsets of these model are used to create three new data models to meet the requirements discussed in the previous section. One of the models will be chosen to become the new data model.

4.2.1 Model A

The first proposed data model, Model A, is an adapted subset of IMKL2015. The resulting data model looks complex Figure 4-1. There are a lot of classes, but most of the classes are abstract. They are needed to make common attribute inheritance to other objects possible.

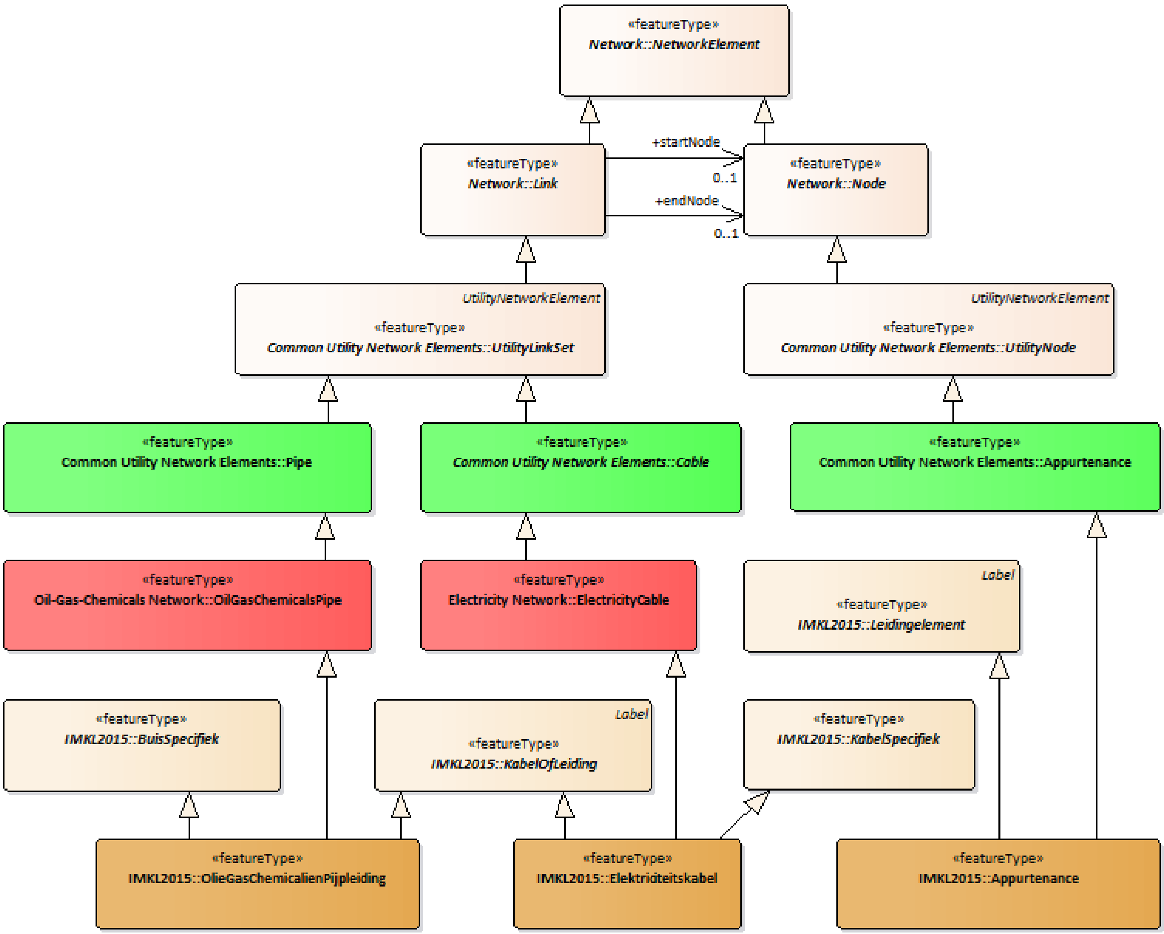

Figure 4-1 Conceptual data model A

The changes made to the IMKL2015 subset are limited. The biggest change is the removal of some subclasses between objects. The reason for this adaptation is simplicity. IMKL2015 is INSPIRE compliant, but some classes are not needed for the application of this research project. The functions of the UtilityLink, the GeneralisedLink and the LinkSet classes are replaced by the Link class. The main function of these classes was to create the possibility of reuse of UtilityLink objects, but IMKL2015 has agreed on not making use of this possibility. The GeneralisedLink object and LinkSet object created the possibility to create network link from several links. This function is also removed because of this adaptation.

The class completeness of Model A is high because all relevant objects are present in the donor models for both the utility products power and natural gas (See Table 4-2).

Table 4-2 The required objects and their classes in model A

Description	**Model object type**
Main cable/pipe	Elektriciteitskabel / OlieGasChemicalienPijpleiding
cable branch off joint, pipe tap saddle	Appurtenance
(Service) supply cable/pipe	Elektriciteitskabel / OlieGasChemicalienPijpleiding
Repair joints (one on one)	Appurtenance
Cable/pipe termination	Appurtenance
Supply point	Appurtenance
Geometry (link)	UtilityLink
Geometry (node)	Part of Appurtenance (inherited from Node)

All the node features can be modelled using the Appurtenance class. The relation and attribute completeness (RD2) scores a plus. The GIS compatibility (RD3) is good, because it is INSPIRE based.

The extensibility (RD4) is very good. A lot of objects that may be needed in the future model versions are already modelled in the used donor model IMKL2015. On complexity (RD5) the score is positive because objects can be modelled using single objects. The scores are summarized in Table 4-3.

Table 4-3 Model A requirement score

Requirement	Score
RD1 Class completeness	++
RD2 Relations and attributes completeness	+
RD3 GIS Compatibility	+
RD4 Extensibility	+
RD5 Complexity	++
RD6 Model well supported/accepted	+
RD7 Reuse of international standards	++
RD8 Easy integration with other models	++

The acceptance (RD6) is based on the fact that all stakeholders will be acquainted with INSPIRE and/or IMKL2015. The network operators more with INSPIRE and both operators and contractors with IMKL2015 when it is introduced in 2018.

Reuse of international (RD7) standards is insured by the integration of INSPIRE in IMKL2015. Integration with other models (RD8) with BGT is by definition arranged because both IMKL2015 and BGT are part of the Dutch national base system for geo-information (Nederlands Normalisatie-instituut, 2011).

4.2.2 Model B

Data model B is based on CityGML Utility Network ADE (Application Domain Extension). As explained before all CityGML objects have associated geometry classes for 3D representation, but these geometries are not sufficient for network topologies and connectivity. The Utility Network ADE delivers the classes to do so. This dual geometry system enforces a complexity as was explained in the previous chapter.

To create data model B, from the CityGML Utility Network ADE a subset is taken. In model B the geometry classes are used to model the objects. The InterFeatureLinks and all the connectivity classes are removed to create model that complies with the minimum requirements. The result model is presented in Figure 4-2.

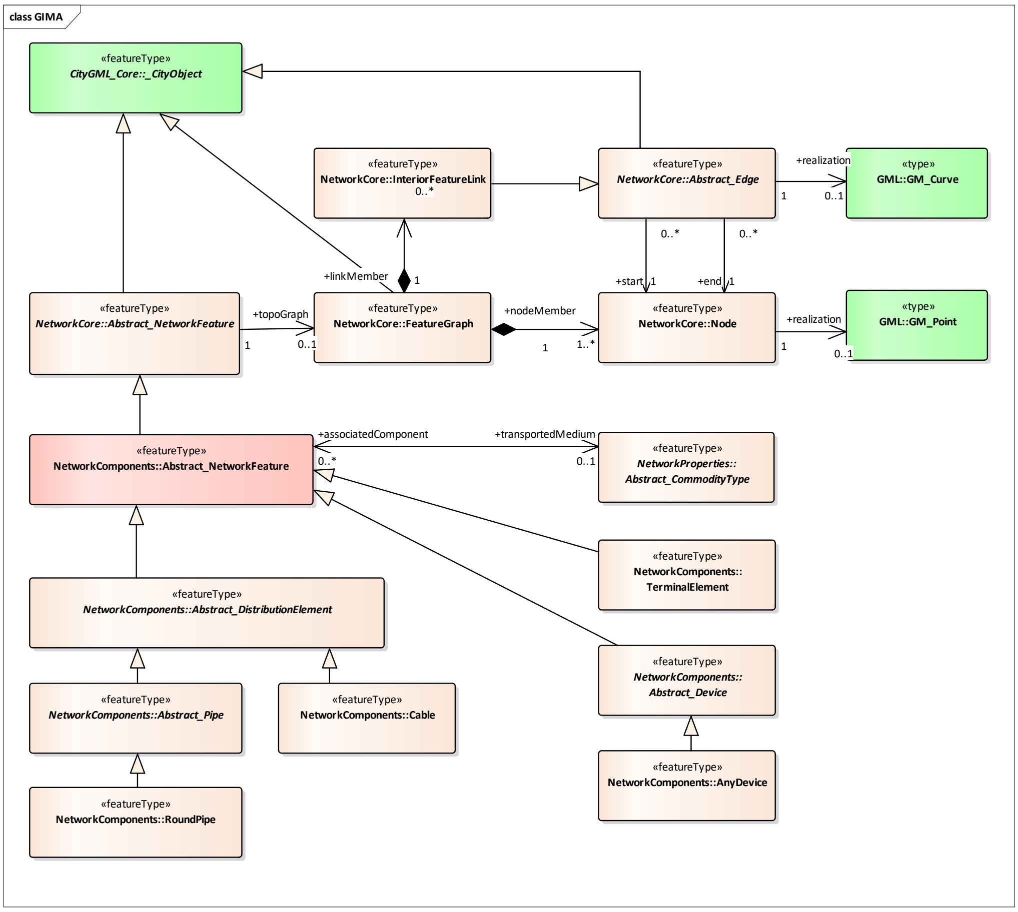

Figure 4-2 Conceptual data model B

The resulting model is compatible with a range of utility products. There are quite a lot of classes but just as with Model A most classes are abstract. All the required objects (RD1) have their representing class (See Table 4-4).

In this model the service supply point has its own class, the TerminalElement, because this supply point is modelled as the 'exit' from the network. This class must not be confused with a cable of pipe termination which are not supposed to be supply points.

Table 4-4 The required objects and their classes in model B

Description	**Model object type**
Main cable/pipe	RoundPipe / Cable
cable branch off joint, pipe tap saddle	AnyDevice
(Service) supply cable/pipe	RoundPipe / Cable
Repair joints (one-on-one)	AnyDevice
Cable/pipe termination	AnyDevice
Supply point	TerminalElement
Geometry	FeatureGraph

Model B scores well on most of the other requirements as well. The relations and attribute completeness (RD2) is good. The GIS Compatibility (RD3) is very good. It is based on GML and has INSPIRE based network geometry model. The extensibility (RD4) to accommodate other utility products is better than model A, because the commodity is stored on even a higher abstraction level. In model A the class for an electricity cable is different from the class of an communication or optic fiber cable. In model B they are all of type 'Cable'.

Model B (RD5) is not more complex than needed despite the dual representation (3D and network).

The model is not supported by the stakeholders (RD6) yet, because it is not a well-known standard in the public utility business. CityGML is a standard from the GIS world. Adoption without enforcing laws is less likely for Model B.

There is reuse of international standards (RD7) INSPIRE and CityGML. But this may change in the future. BGT will become the new reference model for utility networks in the future and BGT is CityGML based. This will make integration with BGT easier (RD8). The results of the requirement analysis are summarized in Table 4-5.

Table 4-5 Model B requirement scores

Requirement	Score
RD1 Class completeness	++
RD2 Relations and attributes completeness	+
RD3 GIS Compatibility	+
RD4 Extensibility	++
RD5 Complexity	+
RD6 Model well supported/accepted	-
RD7 Reuse of international standards	+
RD8 Easy integration with other models	+

4.2.3 Model C

In the literature chapter the CIM model (IEC61970) was introduced as well. A subset is used to create the conceptual data model C (Figure 4-3). In the donor model the connectivity between the objects is modelled by terminal and connectivity nodes. These classes are removed from the model because connectivity is not part of the requirements.

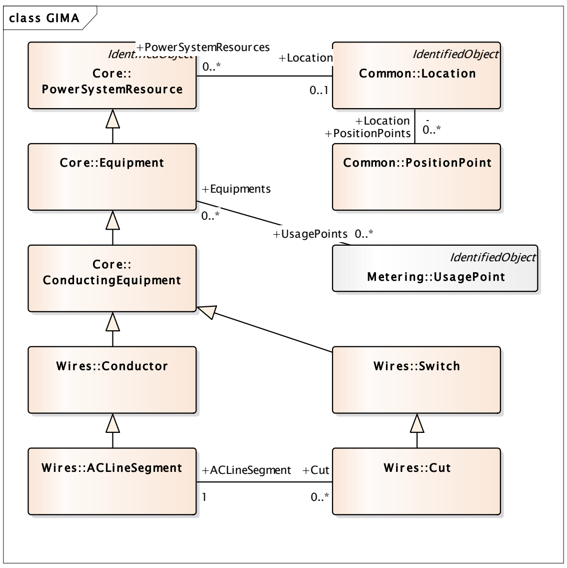

Figure 4-3 Conceptual data model C

This model is also assessed using the requirements. The classes completeness (RD1) requirement is not met, because the gas commodity is not yet represented in CIM. These classes must be added, when this model becomes the base for the data model. In Table 4-6 for the other objects and their model classes. For power networks all objects have a representative class in the model.

Table 4-6 The required objects and their classes in model C (only for power network)

Object description	**Model class**
Main cable	ACLineSegment
cable branch off joint	Cut
(Service) supply cable	ACLineSegment
Repair joint	Cut
Cable termination	Cut
Supply point	UsagePoint
Geometry	Location / PositionPoint

Most of the attributes (RD2) in the donor model are related to the electric properties of the network components. This is logical given the background of the model, but it means not all attributes are available in the donor classes. During logical model design the modeling of physical attributes will require more attention.

The model is clearly not designed with GIS (RD3) in mind. The geographic location of objects can be stored, but CIM is designed for schematic visualization of network components for example for SCADA (Supervisory control and data acquisition) systems for network and load management.

Model C is not very complex (RD5), but extending it with other utility products will make it a lot more complex. The extensibility (RD4) of the model for extra functionality within the power supply networks is possible, because the donor model (CIM) is focused on power networks. Extension towards other utility products that are not part of CIM will take much effort.

The acceptance (RD6) depends on the stakeholder. Utility companies will be used to the CIM model, but most contractors will need more resources to implement the model. The scores per requirements are summarized in Table 4-7.

The reuse of international standards (RD7) is partly met. CIM is an international standard. All integrations with other models are yet to be modelled (RD8).

Table 4-7 Model C requirements score

Requirement	Score
RD1 Class completeness	--
RD2 Relations and attributes completeness	-
RD3 GIS Compatibility	++
RD4 Extensibility	-
RD5 Complexity	+/-
RD6 Model well supported/accepted	+/-
RD7 Reuse of international standards	+
RD8 Easy integration with other models	-

The scores of CIM are very low compared to the other models. This can be explained by the absence of ISO/TC 211 compatible geographical objects to store the location of the network components. When stakeholders want to use CIM, it is possible to extend it. An option would be to apply Mixin with INSPIRE for example.

4.3 Resulting data model

The IMKL subset scores overall highest in the model. This model is chosen to become the base of the data model. The next step is to create the logical data model.

4.3.1 Logical data model

The most important part of the logical data model is the attribute assignment. The donor model IMKL2015 has a lot of attributes that are not part of the scope of this project and may be unnecessary. For now these attributes are left untouched. Reason for this is that the test/evaluation data also is according to the INSPIRE data model.

During the evaluation all these attributes may distract from the purpose of the data model within this project, but the prototype that will be build has a demonstration function. The presence and reorganizability of these attributes will increase the interest of stakeholders. It will also make the transformation and loading of the test data easier. The resulting data model can be found in Figure 4-4.

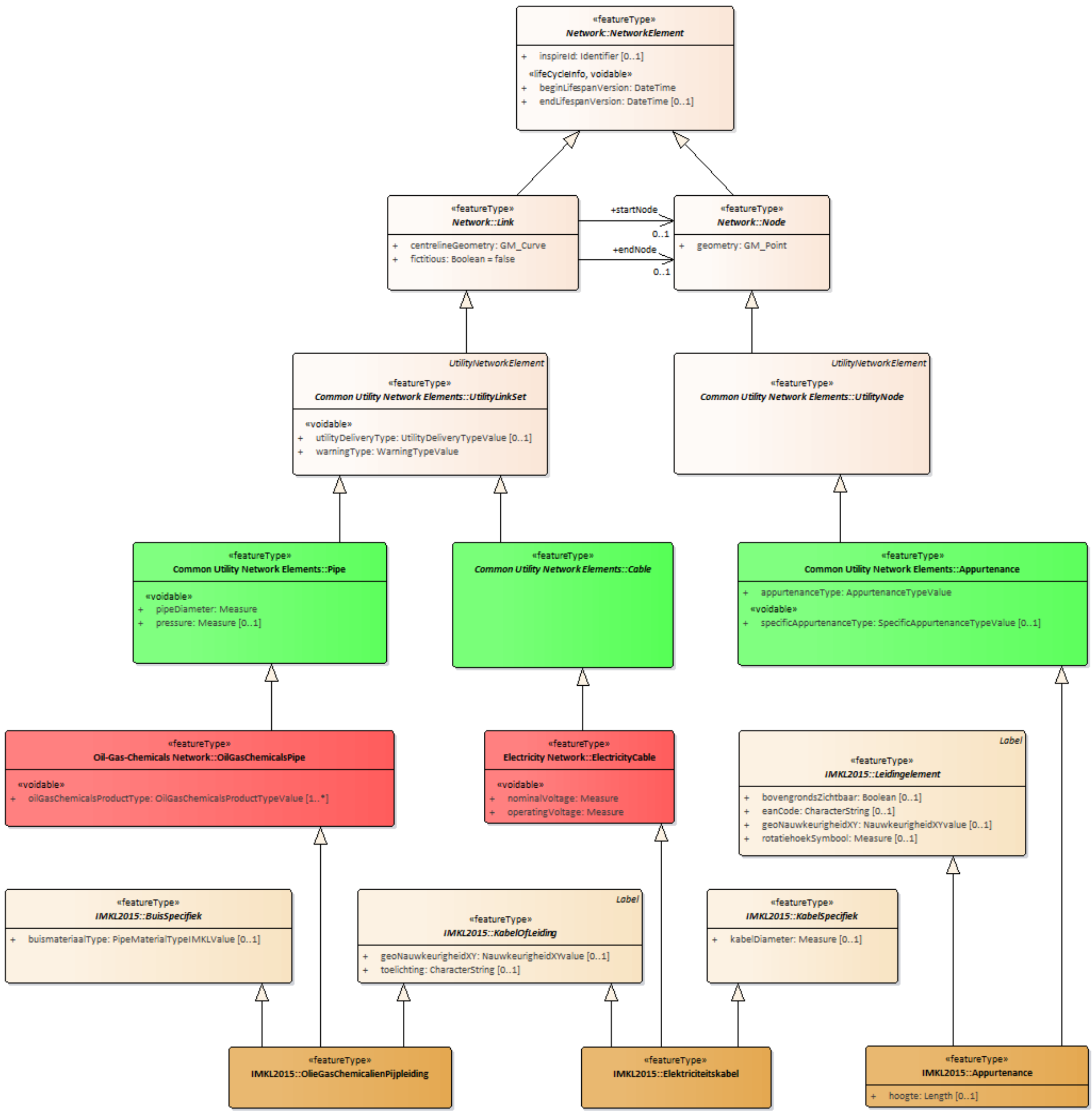

Figure 4-4 Logical data model adapted imkl2015 subset

4.3.2 Physical data model

The next step is to create the physical data model to enable data storage in tables of a GIS enabled database. This step also takes a lot of design choices. The most important design choice may be the number of tables that are going to be used. Usually a trade-off will be made between usability, integrity and non-redundancy.

WFS-T server software often does not support the exchange of complex features. This enforces the data model to be denormalized. There are three non-abstract classes in the data model. These classes are presented in Figure 4-5 with all their inherited attributes.

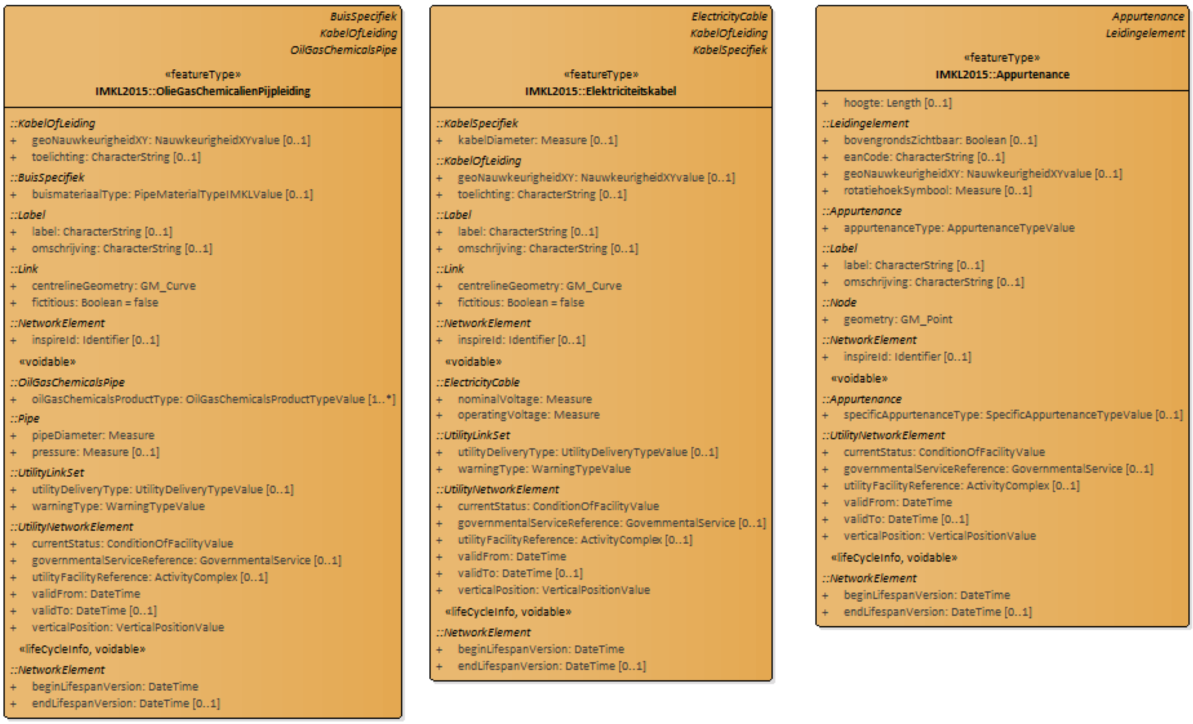

Figure 4-5 Base classes with inherited attributes.

These three classes form the structure of the physical data model. The next step is to create the database table structure. This is done manually. In Figure 4-5 the endNode and startNode fields are still missing. These must not be forgotten when creating the database tables. They represent the association between link objects and node objects.

4.4 Conclusions

During this chapter a data model was developed. From three donor information models IMKL2015 was chosen to become the main donor. This information model was turned into a data model for exchange of asset data for utility network revisions. In the following chapter the communication model will be developed. In chapter 6 they will be evaluated together.

5 Results Technical model for communication

The previous chapter introduced the data model. This chapter will introduce the technical model communication. This chapter starts with defining the requirements. This section is followed by describing available tools to comply with the requirements. Section 5.3 describes the system design. Section 5.5 describes the resulting web application. The chapter is finalized by the evaluation of the system using the scenarios.

5.1 Requirements

The starting point of the requirements analyses is the conceptual model as defined earlier in the methodology and recaptured in Figure 5-1. First the requirements on the prototype are discussed. In this model two major model components are defined. The geo web services and the web application.

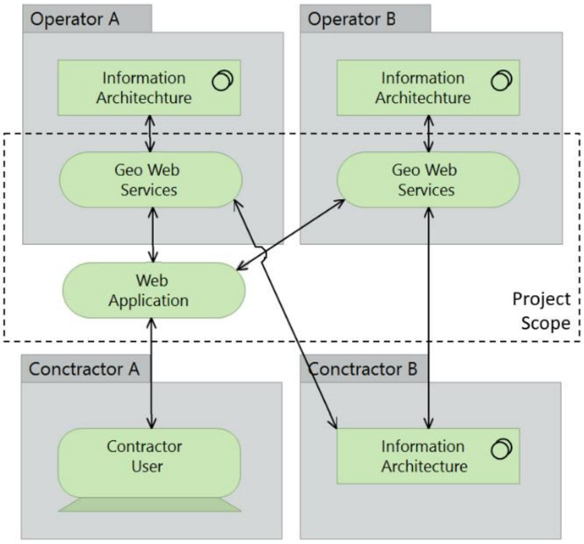

Figure 5-1 Conceptual model communication system

The operators information architectures is not part of the scope. As a result the storage of data must also be part of the geo web services provider. The resulting preliminary design is presented in Figure 5-2.

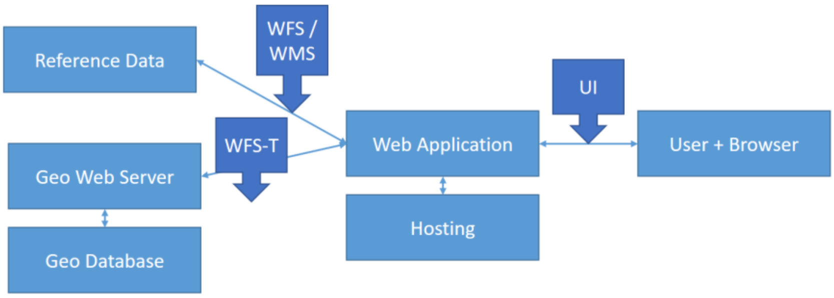

Figure 5-2 Preliminary prototype design based on conceptual model

The geo webserver and the geo database depend heavily on each other. The same goes for the web application and the hosting. As a result the requirements analyses is split in these two parts.

5.1.1 Geo web server and storage

The first requirement and most important requirement is that the server software must be able to handle WFS-T requests as have been explained in the literature chapter (Ch. 3). Only WFS-T enabled software will be compared and tested against the requirements. This will ensure the prototype can be used by multiple users in a heterogeneous environment.

RC1 Compatibility

Standards are not always implemented to full extend or in the way they were intended. The WFS-T must be implemented in such a way that multiple sorts of clients can access and revise the features on the server.

RC2 Configurability

There are not enough resources to do a complete course on how the software must be installed and configured. There must be enough optional settings, but not to much.

RC3 User community

An active community may help overcome issues which other people have run into before.

RC4 Proven technology

With extensively used software products problems should be less likely to occur.

The requirements analyses would be more extensive if the resulting system was more than a prototype. Extra requirements could be e.q. performance, scalability and license costs. The license costs are not taken into account because most software licenses are freely available to universities. There is no 'Open Source' requirement, but mature Open Source products usually score high on the stated requirements.

5.1.2 Web application

The web application can be installed on a standard web server. The hosting will be done by a webhosting provider. For the GIS web services more software is needed. The prototype will include a web application to test the GIS web services.

5.2 Software options

The number of software packages that can handle WFS-T is quite extensive. There is only time to validate a limited number of applications against the requirements. These software packages are presented in Table 5-1. They are preselected because they are well known for their proven technology.

Table 5-1 Geo Server possible software packages

WFS-T software	Database/Data storage
GeoServer	PostgreSQL + PostGIS
MapServer + TinyOWS	PostgreSQL + PostGIS
ArcGIS Server	ArcGIS Server

GeoServer

The GeoServer community-driven open source server for sharing geospatial data is supported by the Open Source Geospatial Foundation (OSGeo). GeoServer dates back to 2001 and have since than grown to an mature software package server that supports the most used OWS services (Open Source Geospatial Foundation, 2014).

MapServer

MapServer is even older than GeoServer. The earliest release dates back to 1994. It was one of the founding projects of OSGeo (Open Source Geospatial Foundation, 2017). With the TinyOWS extension MapServer is also capable of handling WFS-T request, but only when data is store in a PostgreSQL/PostGIS database.

ArcGIS Server

ESRI is the biggest GIS software company in the world (Schutzberg, 2011). It is known for its Desktop software package ArcGIS, but it has a complete software range and also includes cloud bases software solutions. The ArcGIS Server software is able to serve WFS-T request.

5.2.1 Results

The three software packages are validated against the requirements. The results can be found in Table 5-2. Not all score will be explained. Most of the scores are based on internet research.

Table 5-2 Geo server requirement scores

Requirement	GeoServer	MapServer + TinyOWS	ArcGIS Server
RC1 Compatibility	++	++	-
RC2 Configurability	++	+	++
RC3 User community	++	+	-
R4 Proven technology	++	++	++

5.3 Messages

The most important WFS-T methods are *GetFeature* and *Transaction*. With the *GetFeature* request method a list of features can be retrieved. With the *Transaction* method features can be created, updated or deleted using one transaction. In WFS-T all features are encoded using GML. Because of WFS-T, the message types are a given fact. The focus of data modelling can be on the objects or features to be modelled.

5.4 System design

In the previous section the software options were described. In this section the design is presented. Figure 5-3 presents the schematic system overview of the prototype. The user needs a web browser to use the web application. The web application is hosted by the web application server. The application is downloaded and runs in the web browser.

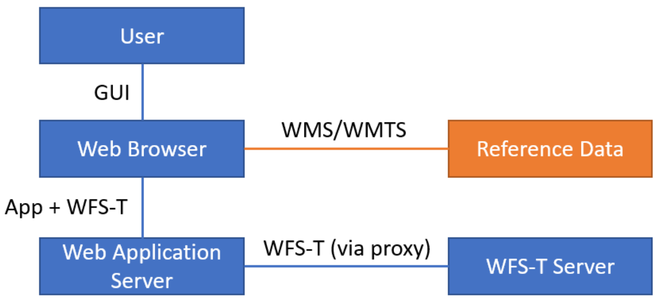

Figure 5-3 Prototype system design including proxy

The application connects to the WMS background services for background maps and reference data. For security reasons web applications cannot request XML messages from other domains than the web application is hosted on. A proxy script on the web application server handles the WFS-T request to the web application server by routing them to the real WFS-T server.

Jetty Web Server
GeoServer
PostGIS
PostgreSQL

Figure 5-4 Geo web server (WFS-T) software stack

The geo web server (WFS-T Server) is build up using four component (Figure 5-4). GeoServer was chosen over the other two options. It scores highest on all requirements.

The database will be a PostgreSQL server with the PostGIS extension. The data model is used to create PostGIS database tables. This database tables are published as WFS-T layers using GeoServer. Jetty is a light weight web server that handles the HTTP requests for GeoServer. Jetty ships with GeoServer, but could be replaced by another web server.

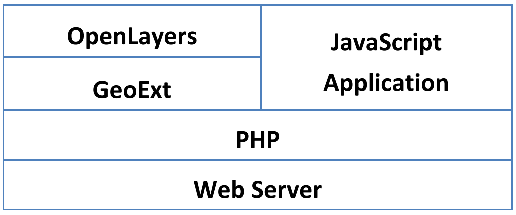

Figure 5-5 Geo web client software stack

The resulting client side application can also be viewed as a stack of component (Figure 5-5). The bottom component is a webserver. The webserver component could be further split in separate components, but this will create unnecessary complexity. The second component in the stack, PHP, is the programming language that is used to add basic security, to supply proxy functionality and to create the HTML page to clue the top components together. The rest of the stack is formed by the earlier introduced GeoExt and OpenLayers libraries. A custom JavaScript application uses these libraries to create the User Interface and to enable communication with the Geo Services.

5.5 System results

The versions of the used software packages are presented in Table 5-3. These versions were chosen because these were the stable release version at time of downloading. There were no know compatibility issues between these software packages related to the used technology.

Software package	Function
PostgreSQL 9.3.1	Database
PostGIS 2.1.1	GIS extension for Database
GeoServer 2.4.1	OpenGIS Web Services Server

Table 5-3 Database and server software packages

The installation and configuration off the three software packages was relatively straight forward. GeoServer was installed and configured using three configuration levels. A new prototype environment was created with WFS-T technology enabled. The PostGIS enabled database was added as source to this environment. The last step is to create layers from data source tables. During this last step GeoServer analyses the table data and proposes to publish the table columns as WFS-T feature attributes.

5.5.1 Web application

The web application was developed using the software versions presented in Table 5-4. These versions are a bit outdated. The version were chosen at the start of development three years ago. At that moment these version were the best interoperable choses.

Table 5-4 Web application building blocks

Software package	Function
Openlayers 2.13.1	JavaScript OGC client library
Ext JS Library 3.4.0	JavaScript UI library
GeoExt 1.1	JavaScript Ext and OpenLayers integration toolkit
Custom PHP + JavaScript	Application development

These libraries are brought together using the custom application development.

Table 5-5 Web application custom code efficiency

Description	Language	Lines of code
Main page	HTML	23
Prototype application (UI)	JavaScript	802
WFS visualization styles	XML/SLD	168
Proxy server (code-reuse)	*PHP*	*180*

The code is added to GitHub and can be found at https://github.com/ORibberink/gima-prototype.

The main reason for the efficiency of the custom code is the use of standards and the use of open source libraries. A small majority of the code consists of configuration of parameters. The remaining code adds interface logic, functions and event handling.

The main HTML page is really small because it only includes the libraries and initializes the JavaScript application. The JavaScript application is the backbone of the application user interface (UI). It creates all the panels, the map, the buttons and logic to make the UI interactive and the ability to create WFS-T requests. A code example is given in Code 5-1 to show how easy a WMS background layer is added to the map.

```
layers[] = new OpenLayers.Layer.WMS(
      "PDOK Cadastral map",
      "//geodata.nationaalgeoregister.nl/kadastralekaartv2/wms",
      {      layers: 'kadastralekaart',
             transparent: 'true'},
      {      isBaseLayer: false,
             minResolution: resolutions[16] ,
             maxResolution: resolutions[14]})
```

Code 5-1 Javascript code example. Adding cadastral map background layer with OpenLayers.

The proxy server is needed because of security related restrictions. The WFS request the UI executes in the background must be targeted at the same host as the UI. The WFS request are targeted at the local proxy server, which performs the same request to the real WFS server and relays the answer back to the UI. The proxy server is an adapted version of a script found on the web.

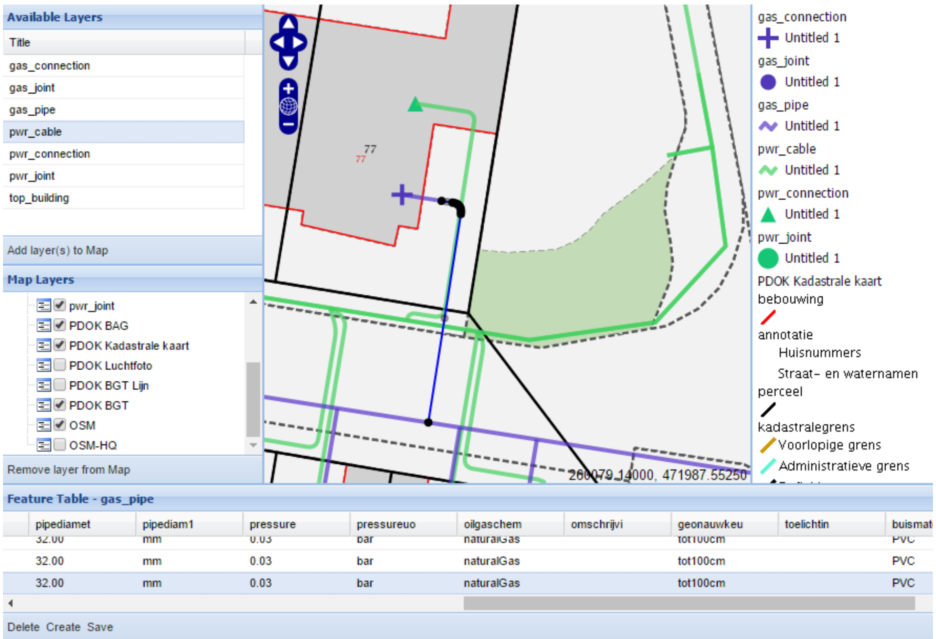

Figure 5-6 Web application user interface screenshot

A screenshot of the resulting web application is presented in Figure 5-6. The UI has 5 panels. In the center there is the map. At the right side is the map legend. At the left side there are two panels with available and current map layers. At the bottom a table or grid is presented if a WFS layer is selected. The WFS features on the map are presented with their attributes.

The map and the table grid can be used to select features. The location and attributes of selected features can be edited. There are three extra buttons for feature manipulation: create, delete and save. The create button enables the user to create new features. The delete button deletes the current selected feature. All the modification are stored locally. Only when the save button is clicked, all the locally made changes are send to the WFS server by WFS-T protocol.

There are 6 background layers available in the web application. They are presented in Table 5-6 with their names and purposes. Two background layers have a very similar copy. The Open Street Map is added as base map for large scale navigation. The HQ version is added for mobile devices. With that version enabled the street name labels are easier to read on small screens. The large scale topography maps are used as reference background. There is a version with colors and a version without colors. The visibility of features is better without colors on the reference map. The cadastral map is transparent. The administrative parcel borders can be plotted on top of the other refence maps.

Name	Function	Standard
Open Street Map	Small scale navigation	WMS
Open Street Map (HQ)	Small scale navigation for mobile devices	WMS
PDOK BGT	Large scale topography	WMTS
PDOK BGT Lijn	Large scale topography (lines only)	WMTS
PDOK luchtfoto	Aerial photography (25cm res)	WMTS
PDOK Kadastrale kaart	Cadastral map	WMS
PDOK BAG	Buildings	WMS

Table 5-6 Web application background layers

In the application not all background layers are available on all zoom levels. This has two reasons. The first is service availability. Some WMS layers simply do not provide tiles for all zoom levels. The other reason is usability. Because OSM background data is used for small scale and PDOK reference data for large scale OSM data is only enabled for low zoom levels and PDOK layers can only enabled on high zoom levels. The resulting default reference maps per zoom level are presented in Table 1-1.

Table 5-7 Default reference map per zoom level

Zoom level	Resolution	Scale (approx. with 90 DPI)	Reference map
1 – 14	3441 – 0.42 m/px	1:12.000.000 – 1:1.500	Open street map
15 – 18	210 – 26.3 mm/px	1:750 – 1:90	PDOK BGT + PDOK Cadastral data
19 – 21	13.1 – 0.82 mm/px	1:40 – 1:5	PDOK BAG

The scale column is not relevant for digital screens anymore, but for many users mm/px is not very practical. In practice the paper reference map GBKN is still used often with scales between 1:100 and 1:5000. The 90 DPI factor is chosen here because it is the OGC standard output resolution. In the prototype application you can zoom in even further, but only reference map BAG is available at this zoom level.

5.5.2 Conclusions

A system has been developed to serve the purpose of the research. A GeoServer provides the WFS-T services of several natural gas and electricity network layers. A web application has been built to evaluate the design choices and to demonstrate the technology to stakeholders.

There are some properties of the resulting system that could influence the evaluation. There are limitations to the usability as a result of design choices and limited resources. A major drawback of the web application is that only one layer can be edited at once. When changing layers all changes must be saved before the next layer can be edited. This is could be solved if more development time was available. The WFS-T technology supports multiple feature mutations over different layers in one Transaction message. An WFS-T example message can be found in Appendix A.

Due to limited resources some advanced features like snapping to existing objects when adding new objects are not implemented. The user-friendliness of the end results was of less importance during development. On the other hand were some design choices influenced by the fact that the prototype is also a demonstration system to show stakeholders the possibilities of the used technologies.

That concludes the description of the prototype. In the following chapter the prototype is used to evaluate the developed data model and communication system by evaluating it against the scenarios.

6 Evaluation

The last two chapters described the research results. This chapter will evaluate the results by validating the model using scenarios. First the test data is described in section 6.1. Section 6.2 describes how the scenarios are executed. The results are presented in section 6.3. The last section is dedicated to compatibility testing.

6.1 Data

The source of the data imported in the prototype is the open data program (Enexis, 2017) as introduced in the methodology (Section 2.6). The data sets contains the asset data of the complete Enexis region, which compromises the larger part of 5 out of the 12 provinces of the Netherlands.

The dataset contains both aboveground and underground asset data. The aboveground assets are amongst others substations, street cabinets and charging points for electric cars. None of the aboveground assets from the dataset are imported in the prototype. The underground assets are of type Cable, Pipe and Appurtenance. The layers from the dataset that are imported are presented in Table 6-1.

Table 6-1 Object layers from Enexis open dataset

Utility product	Name	Target object type
Electrical power	Mid voltage main cable	Elektriciteitskabel
Electrical power	Low voltage main cable	Elektriciteitskabel
Electrical power	Low voltage connection cable	Elektriciteitskabel
Electrical power	Low voltage service connection	Appurtenance
Natural gas	Main pipe	OlieGasChemicalienPijpleiding
Natural gas	Connection pipe	OlieGasChemicalienPijpleiding
Natural gas	Service connection	Appurtenance

The prototype described in the previous chapter has a limitation in the number of features it can handle per layer. Therefor using bounding boxes only two small sections from the dataset are selected to be imported in the prototype. One section in Oss and one section in Enschede. The section in Oss is selected because the BGT reference layer quality very good. Not all municipalities are at the same level in implementing BGT. The municipality of Oss is one of the frontrunners.

The section in Enschede is selected based on a visual inspection of the Enexis data. The cables and pipes are nicely laid out in this section. It may even be that for the service connection cables, the as-planned drawings are promoted to as-build drawings.

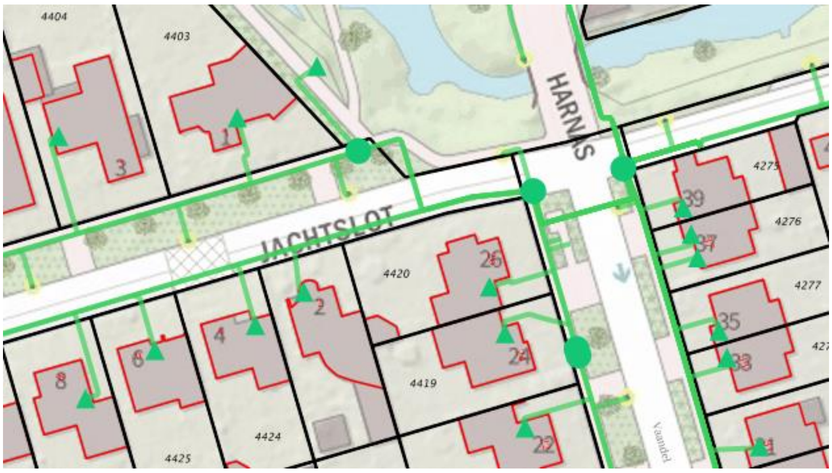

Figure 6-1 Result of dataset import into prototype

The visual quality of the input data is very good. Figure 6-1 is a screenshot of the prototype with the BGT as reference data and the imported electrical power assets. The cables all lay within the expected zones mostly beneath footpath as expected in the Netherlands. The service connections lay inside the BGT target buildings and the connection cables to street lighting line up with the street lights on the BGT topographic map. The assessment of other quality aspects of the dataset will performed in the following sections.

There is an attribute in the appurtenance class called rotatiehoekSymbool (English: symbol rotation angle). This attribute can be used to rotate the appurtenance symbol. In the prototype this virtualization option is not implemented.

6.2 Scenarios

In the methodology chapter the scenarios were introduced. In this section the scenarios are used to evaluate the executed using the prototype. During executing the prototypes several points of enhancement were discoved. They will be described and discussed in this section.

6.2.1 Scenario 1 Construction of roundabout

The construction of the roundabout requires removal and replacement of cables and pipes on a road junction. The network components will be replaced with a greater distance from the center of the junction to make space for the roundabout.

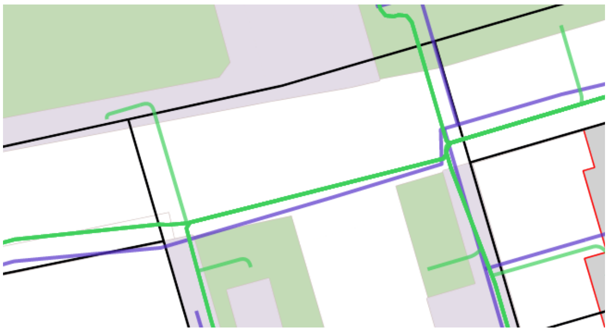

Figure 6-2 Scenario 1 before changes

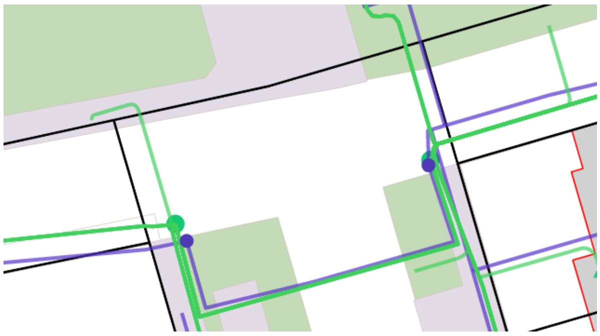

Figure 6-3 Scenario 1 after changes

The sections of cables and pipes that are too close to the roundabout are removed and new sections of pipes and cables are created with a greater distance from the center of the intersection. The new pipes and cables are connected to the existing network using seal joints. The results of the evaluation are summarized in section 6.3.

6.2.2 Scenario 2 Connecting multiple consumers

An often occurring scenario is that in a project multiple residences, must be connected to the energy supply network. Connecting consumers to the energy supply network often takes two stages. The first stage is to create the main cables and pipes. In the Netherlands these are often buried beneath a footpath along the street. From this main line branch off service cables or pipes are create to serve every house hold separately.

Figure 6-4 Scenario 2 before changes

Figure 6-5 Scenario 2 after changes

Using the prototype first the main power line is added to the network. A sealing joint is used to connect it to the existing network. A termination joint is added at the end of the cable. A cable should have a starting point and an endpoint.

In the second phase, for each house hold two branch off joint are created. From this branch off joints the service pipe and cable are drawn to the residence. There a service supply point is added. This is the end point of the network. The results of the evaluation are summarized in section 6.3.

6.2.3 Excavation damage (S3)

The third scenario involves excavation damage. A field engineer is send to the location to repair the damage. After the work is complete the technician must enter the updated and new network components. Sometimes a repair joint is enough to fix the problem. In this scenario a piece of pipe and a piece of cable were replace by a new piece and attached using joints.

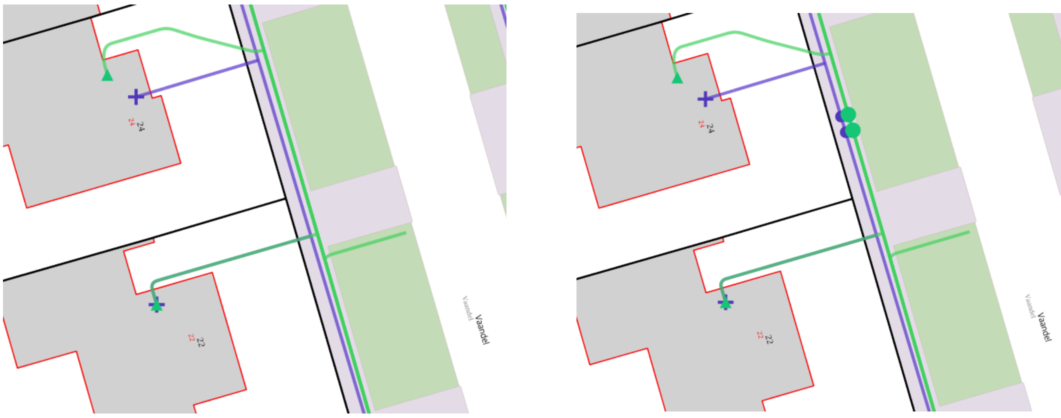

Figure 6-6 Scenario 3 before (left) and after (right) changes.

During simuliation of the scenario some issues occurred. Some bigger than others. As in the previous scenarios cable selection was one of the issues. When multple cables are on the same location selecting the right one is hard. The results are summerized in the following section.

6.3 Results

The scenarios are executed and the prototype is able to handle the requests as expected. The results of the execution of the scenarios can be grouped into two types of results. The data model will not be perfect. This can be caused by scope limitation, but also by unforeseen situations. The other type of drawbacks are caused by the prototype. There was only a limited time to develop the web application. Not all wishes could be implemented. During evaluation two major data model drawback where found. The drawbacks are a result of the scope limitations: There is no reuse of cables and pipe location data in a 'trench' object and the data model also does not provide for the storage of tape-measurements.

The most important prototype drawback is the fact that only one layer can be edited at a time. For a prototype this is not a problem, but if the prototype will be updated with new features, this should be a top priority. The second issue related to the prototype more than the data model is that objects cannot be split in the prototype. When a section of a pipe or cable is replaced. The existing cable section must be split at least once to be create space for a replacement section and the repair fittings. The prototype does not contain all functionality that can be expected from modern GIS applications.

Another function that would add much value is snapping to existing features. New features often start or end at existing features. When drawing new features on the map. It would be helpful when their location would snap to existing objects. This is also related to the fourth issue. Features are related to each other, but the user can only manually add this relation by entering the ID of the related feature.

Cables are often buried together in a trench. Their location is so close that their stored location is the same. When a cable must be edited or is used to create a branch off point, it must be selected. This selection is difficult if multiple cables are on the same location. The only way of selecting the right cable is using the feature table in the bottom of the screen. The fifth and last issue is therefor that the web application could provide tools to make this selection using the map easier.

6.4 Compatibility

The web application that was developed is platform independent. It works on different devices like PCs, tablets and mobile phones. It works on different operating systems like Windows, android and iOS. It is also browser independent.

To further test the interoperability of the communication system. Other software is used to edit the features hosted by the web service. QGIS 2.18 is used to make a connection to the WFS-T server and features are retrieved. The features added and updated in QGIS are then retrieved in the prototype web application and the changes are noticed. The compatibility test succeeded. The compatibility tests show that the developed system can be used in a heterogeneous environment.

7 Discussion and Conclusions

In the previous chapters the results were presented. This chapter will discuss the results and tries to put them in perspective by discussing the approach. Some statement will be about what steps are needed for adoption of the standards. The chapter and thesis is ended by some concluding statements.

7.1 Research questions

In chapter 1 the research questions were introduced. Now that the results are evaluated, the research question should have been answered. First the sub questions are recapitulated:

1. ***Which standards are in use or available for use for data storage and data exchange in revision projects on pipelines and cables?***
2. ***How far has standardization come in case of standards for data storage and exchange in network revision projects similar to the use cases?***
3. ***How can these standards be used to create a prototype to standardize information data models and communication protocols.***
4. ***What is the result of a qualitative assessment of the created prototype for the use cases?***

7.1.1 Sub question 1

During literature research (Chapter 3) standards were reviewed for both data storage and data communication. The conclusions from this review were that there are available standards possibly useful for the purpose of this project. Three data models were identified as being possibly a starting point. IMKL2015, CityGML UtilityNetwork ADE and CIM.

7.1.2 Sub question 2

The implementation of standardization in cases similar to the use cases is slim. In the Netherlands there were multiple projects to increase standardization. The results were divergent. The information exchange between the network operators and constructors improved the last years and is still rapidly improving. But the exchange of as-built drawing was not the main focus thus far and spatial data is not exchanged using Geo-Information standards.

Examples of standardization in other countries are hard to find. As earlier explained the Netherlands are quite unique in the percentage of underground utilities. For other countries standardization may have been of less importance.

7.1.3 Sub question 3

The available data models were compared and IMKL was chosen to become the donor model for the new standard. The communication system was developed after comparing software packages for WFS-T service hosting. GeoServer was chosen to act as the WFS-T server and PostgreSQL/PostGIS as database server. Using existing JS libraries a client interface was built to complete the prototype.

7.1.4 Sub question 4

The prototype was used to evaluate the developed data model and communication system. The three scenarios were used to perform the evaluation. The results were positive. The prototype has shown that the model supports data exchange in a heterogenous environment. The data model supports the information exchange for the scenarios, but there are points left to perform for further research.

7.1.5 Main research question

The sub questions were steps to make it possible to answer the main question:

How can geo data exchange between utility distribution network operators and contractors be standardized, both for information data models and for communication protocols?

This research has introduced a data model and system for communication for the communication of As-Built records between network operators and contractors in underground utilities. A prototype was build and the results were evaluated.

The results were positive and the proposed system can be used to exchange information within the scope of the research project. The application of the developed model outside the scope of the research project is not proven. There are some points for improvement and before implementation of the model is possible more research is recommended.

7.2 Discussion

Although validation was not part of the scope of this project is an important part of information exchange. The WFS-T technology allows for validating each transaction. The current version of the prototype only supports validation to a certain extend. Due to the prototype disadvantages relations between objects cannot be validated for example, because a transaction in the current prototype state only involves one object sort at a time. Validation of transmitted transaction must be performed on objects from different layers in one transaction.

Validation can be performed on different levels of the architecture. A database management system (DBMS) is able to validate database transaction against the data model to prevent inconstant data. Modern information system architecture often separates storage, business intelligence and presentation of data in different architecture layers. Because business process management (BPM) often services multiple systems with multiple functions validation on DBMS alone is not enough.

Validation of transactions could also be performed before the transactions are submitted. In chapter 2 WPS was introduced. WPS technology can be used to execute processes on geo-information remotely. The validation business intelligence used by the WFS-T server, could also be made available through a WPS server.

Feature locking is provided by the WFS-T standard to prevent multiple users updating the same features at the same time. It was not part of the scope of this project, but it should be implemented to prevent users from overwriting each other's changes without notice. Client developers should prevent editing of features that are locked.

Added value of the technology used in this project is quality feedback on the used reference data. When users of BGT and other services offered by PDOK discover data quality issues there may be even a reporting obligation. The use of OGC web services offers the possibility to client developers to integrate quality feedback in their application.

Related to these issues is the fact that security and authentication was not part of the scope. Authentication and user management is of course an important part of a system design. There are multiple technical solutions possible for user authentication.

7.3 Further research

There are several topics that need more research before the model can be applied. The scope of this research projects was wide-ranging and not all details could be addressed. Both the data model and the communication system need more attention.

Digital registration of tape-measurement location got less attention than it deserves. Although GPS technology is advanced enough, location registration of cable and pipes for residential service connections with GPS technology is still too expensive. GPS reception in buildup areas is also of influence on the location quality. Usually tape-measurements are used to calculate the distance of the infrastructure relative to the building being served. Further research is needed to model the data registration of those measurements. It is strongly related to earlier comments on how the future topographic reference model BGT is to be incorporated in the model.

3D is a very popular topic in Geo Information Science, a specially when it comes to Standards. The CityGML standards fulfills a need for modeling city objects in 3D, but in the industry, especially in utilities, 3D does not have many benefits. There are national standards for how deep cables and pipes should be buried. Drawing the third dimension only makes things unnecessary complex and therefor unnecessary expansive.

Visualization was only a small part of the project. More research into this subject could help in adoption of the standard. The symbolic encoding rules of WMS services can be applied to WFS-T data. Using this technology features can be mapped to look always the same.

Further research could also include update the prototype by solving the issues discovered during evaluation. The earlier discussed issues include editing multiple layers at the same time, snapping to existing features, cutting features and advanced object selection.

During the execution of this project, there were also developments in the IMKL2015 program. The visualization rules of IMKL2015 were published recently. Further research could be performed if these visualization rules are useful for the purpose of this research subject and if these rules could be added to the prototype.

7.4 Adoption

From the previous sections can be concluded that before implementation of the proposed model, some further research is advised. Nonetheless some statements can be made on what kind of steps, or hurtles, need to be taken before the system is widely adopted.

There are convincing properties of the model for adopting the model because the potential quality and efficiency gain. Firstly, if technicians have feature based information of the existing network they can make immediate feature based As-Built transactions. IF these can be send back directly to the network operator, this will increase network registration handling speed and the quality.

Secondly, BGT is around the corner. By using the proposed model, this reference data model can be optimally applied. The reference data will be feature based, and these feature relationships will prevent errors in processing the As-Built records transaction.

Thirdly because the OGC standards are used in the model all stakeholders are free to use any software that is OGC-enabled. This research project show that with minimum effort advanced exchange prototypes can be created by using advanced applications and libraries created in the last decades.

The fourth and last reason is the reuse of existing models prevents ontology and semantics discussions. The reuse of business logic is not evident, because especially with standards. They are often implemented only to the bare necessary minimum and often outsourced to external parties. The major part of the proposed data model is a subset from the well-documented existing data model IMKL2015.

There are also some expected threads for adopting the model. The model depends on agreements between the parties using it. These agreements are not only about what data is exchanged with the model, but what data is not exchanged with the model. The data model will depend on references to data like BGT and the cadastral map. These maps can be updated daily, so one of the agreements should be how to communicate the version of the map, or the version of the referenced objects for example.

Another thread is that the standard GIS software vendors of the network operators pretend to comply with OGC standards, but in practice they often do not. This could be expected, because they do not benefit from open standards. It threatens their revenue model. Network operators must be aware of this. The OGC performs compliance tests and the results are freely available.

From recent standardization projects can be learned that they are only successful when the implementation model start from a small number of stakeholders. When the model is a success with a small number of parties other parties may be invited to also help optimize and use the model.

Another approach could be to first finalize the model and then force the use of it by law or regulations. This will improve implementation speed, but this may have side effects. If there is need for this approach it would mean that the stakeholders do not see enough benefit in the model to implement it. Forcing standardization in favor of standardization itself cannot be enough reason to do so. This will only lead to standardization aversion.

7.5 Conclusions

This project has shown that there is a need for quality and efficiency improvement in the exchange of geoinformation between utility network operators and contractors and that this can be solved by using standardization.

A data and communication model has been developed. A working prototype has been developed and implemented. The prototype was used to evaluate the model. The prototype proved that with relatively small resources, great improvements are possible when OGC standards are used and existing data model are reapplied.

During evaluation the model proved to be useful and interoperable. Some research is needed to complete the model, but a business case is the logical next step to prove the claimed advantages are a reasonable and feasible goal.

8 References

Beck, K., M. Beedle, A.v. Bennekum et al. (2001), The Agile Manifesto.

Becker, T., J. Pouliot & S. Daniel. (2013). Semantic 3D modeling of multi-utility networks in cities for analysis and 3D visualization. pp. 41-62Springer Berlin Heidelberg.

CERISE–SG. (2012), Project Plan Combineren van Energie- en Ruimtelijke Informatie Standaarden als Enabler – Smart Grids (CERISE-SG).

CIM Users Group. (2010), Introduction to CIM and its Role in the Utility EnterpriseXtensible Solutions.

Electric Power Research Institute. (2015). Common Information Model Primer (Third ed.) < https://www.epri.com/#/pages/product/000000003002006001/>

Enexis. (2017). Open data van Enexis [online].Available on the world wide web: <https://www.enexis.nl/over-enexis/open-data>.

European Commission. (2016). About INSPIRE [online].Available on the world wide web: <http://inspire.ec.europa.eu/about-inspire>.

Geonovum. (2015), Raamwerk geo-standaarden (3.0th ed.)

Geonovum. (2017). IMKL2015, Informatiemodel Kabels en Leidingen [online]. Retrieved 04/30, 2017. Available on the world wide web: <https://register.geostandaarden.nl/imkl2015/>.

Geurts, H. (2013). GeoDirect, revisie binnen 24 uur. Paper presented at the Nationaal kabels en leidingencongres,

Haas, H. & A. Brown. (2012). W3C web services glossary.

Internet Engineering Task Force. (2014), The JavaScript Object Notation (JSON) Data Interchange Format (RFC 7159).

Kadaster. (2016). KLIC-WIN [online].Available on the world wide web: <https://www.kadaster.nl/klic-win>.

Kutzner, T. & T.H. Kolbe. (2016). Extending semantic 3D city models by supply and disposal networks for analysing the urban supply situation. Paper presented at the Lösungen Für Eine Welt Im Wandel, Dreiländertagung Der SGPF, DGPF Und OVG, 36. Wissenschaftlich-Technische Jahrestagung Der DGPF, , 25 pp.382-394. < http://www.dgpf.de/src/tagung/jt2016/proceedings/papers/36_DLT2016_Kutzner_Kolbe.pdf;>

Ministry of Economic Affairs. (2008), Wet informatie-uitwisseling ondergrondse netten. Staatscourant 4615

NEN 3610: 2011: Basismodel geo-informatie - Termen, definities, relaties en algemene regels voor de uitwisseling van informatie over aan de aarde gerelateerde ruimtelijke objecten, NENU.S.C. (2011).

Nurseitov, N., M. Paulson, R. Reynolds et al. (2009), Comparison of JSON and XML data interchange formats: A case study. Paper presented at the 22nd International Conference on Computer Applications in Industry and Engineering 2009, CAINE 2009.

Object Management Group. (2003), Unified Modeling Language (UML) Part 1: Infrastructure.

Open Geospatial Consortium Inc. (2005), OpenGIS® web services architecture description. OpenGIS® Best Practices Paper.

Open Geospatial Consortium Inc. (2007). Geography Markup Language [online].Available on the world wide web: <http://www.opengeospatial.org/standards/gml>.

Open Geospatial Consortium Inc. (2012). CityGML [online].Available on the world wide web: <http://www.opengeospatial.org/standards/citygml>.

Open Source Geospatial Foundation. (2014). What is Geoserver?[online].Available on the world wide web: <http://geoserver.org/about/>.

Open Source Geospatial Foundation. (2017). Mapserver - About [online].Available on the world wide web: <http://mapserver.org/about.html>.

PDOK. (2017). PDOK Organisation [online]. Retrieved 4/30, 2017. Available on the world wide web: <https://www.pdok.nl/en/about-pdok>.

Prins, A., H.v.d. Berg, M. Olij et al. (2013), DURF3 Het Functionele Spoor (CONCEPT/VERTROUWELIJK).

Schutzberg, A. (2011). Esri has 40+% of GIS Marketshare [online]. Retrieved 03/22, 2017. Available on the world wide web: <http://www.directionsmag.com/entry/esri-has-40-of-gis-marketshare/215188>.

W3C. (2008), Extensible Markup Language (XML) 1.0 (Fifth Edition).

W3C. (2012). XML Schema [online].Available on the world wide web: <https://www.w3.org/standards/techs/xmlschema>.

Wordfence. (2015). OSI Model [online]. Retrieved 4/1, 2017. Available on the world wide web: <https://www.wordfence.com/wp-content/uploads/2015/11/wordfence-osi-model.jpg>.

9 Appendixes

A. WFS Transaction example

Simple WFS-T Request without attributes. Namespaces are removed for readability.

```
<wfs:Transaction service="WFS" version="1.1.0">
	<wfs:Insert>
		<feature:gas_connection xmlns:feature="http://ribberink.com/prototype">
			<feature:geom>
				<gml:Point srsName="EPSG:28992">
					<gml:pos>260135.5659 472026.7983</gml:pos>
				</gml:Point>
			</feature:geom>
		</feature:gas_connection>
	</wfs:Insert>
	<wfs:Insert>
		<feature:gas_joint xmlns:feature="http://ribberink.com/prototype">
			<feature:geom>
				<gml:Point srsName="EPSG:28992">
					<gml:pos>260136.6705 472027.9029</gml:pos>
				</gml:Point>
			</feature:geom>
		</feature:gas_joint>
	</wfs:Insert>
</wfs:Transaction>
```

WFS-T Response

```
<wfs:TransactionResponse version="1.1.0">
	<wfs:TransactionSummary>
		<wfs:totalInserted>2</wfs:totalInserted>
		<wfs:totalUpdated>0</wfs:totalUpdated>
		<wfs:totalDeleted>0</wfs:totalDeleted>
	</wfs:TransactionSummary>
	<wfs:TransactionResults/>
	<wfs:InsertResults>
		<wfs:Feature>
			<ogc:FeatureId fid="gas_connection.8315"/>
		</wfs:Feature>
		<wfs:Feature>
			<ogc:FeatureId fid="gas_joint.135"/>
		</wfs:Feature>
	</wfs:InsertResults>
</wfs:TransactionResponse>
```

B. Prototype web application source code

Index.html

```
<!DOCTYPE html>
<html debug="true">
    <head>
        <title>GIMA Thesis Prototype</title>
        <link rel="stylesheet" type="text/css" href="ext/resources/css/ext-all.css">
        <link rel="stylesheet" type="text/css" href="geoext/resources/css/popup.css">
        <link rel="shortcut icon" href="images/GIMAlogo_cut.png" />
        <link rel="apple-touch-icon" href="images/GIMAlogo_cut.png" />

        <!-- Include Libraries-->
        <script type="text/javascript" src="//extjs.cachefly.net/ext-
3.4.0/adapter/ext/ext-base.js"/></script>
        <script type="text/javascript" src="//extjs.cachefly.net/ext-3.4.0/ext-
all.js"></script>
        <script src="//openlayers.org/api/2.11/OpenLayers.js"></script>

        <script type="text/javascript" src="geoext/lib/GeoExt.js"></script>
        <script type="text/javascript" src="proj4js/lib/proj4js.js"></script>

        <!-- Prototype Software -->
        <script type="text/javascript" src="inc/google.js"></script>
        <script type="text/javascript" src="inc/functions.js"></script>
        <script type="text/javascript" src="inc/application.js"></script>
    </head>
    <body>
        Deze applicatie is niet geschikt voor oude browsers...
    </body>
</html>
```

Main application (inc/application.js)

See https://github.com/ORibberink/gima-prototype/blob/master/src/inc/application.js

Functions (inc/functions.js)

See https://github.com/ORibberink/gima-prototype/blob/master/src/inc/functions.js

Proxy (proxy.php)

See https://github.com/ORibberink/gima-prototype/blob/master/src/proxy.php