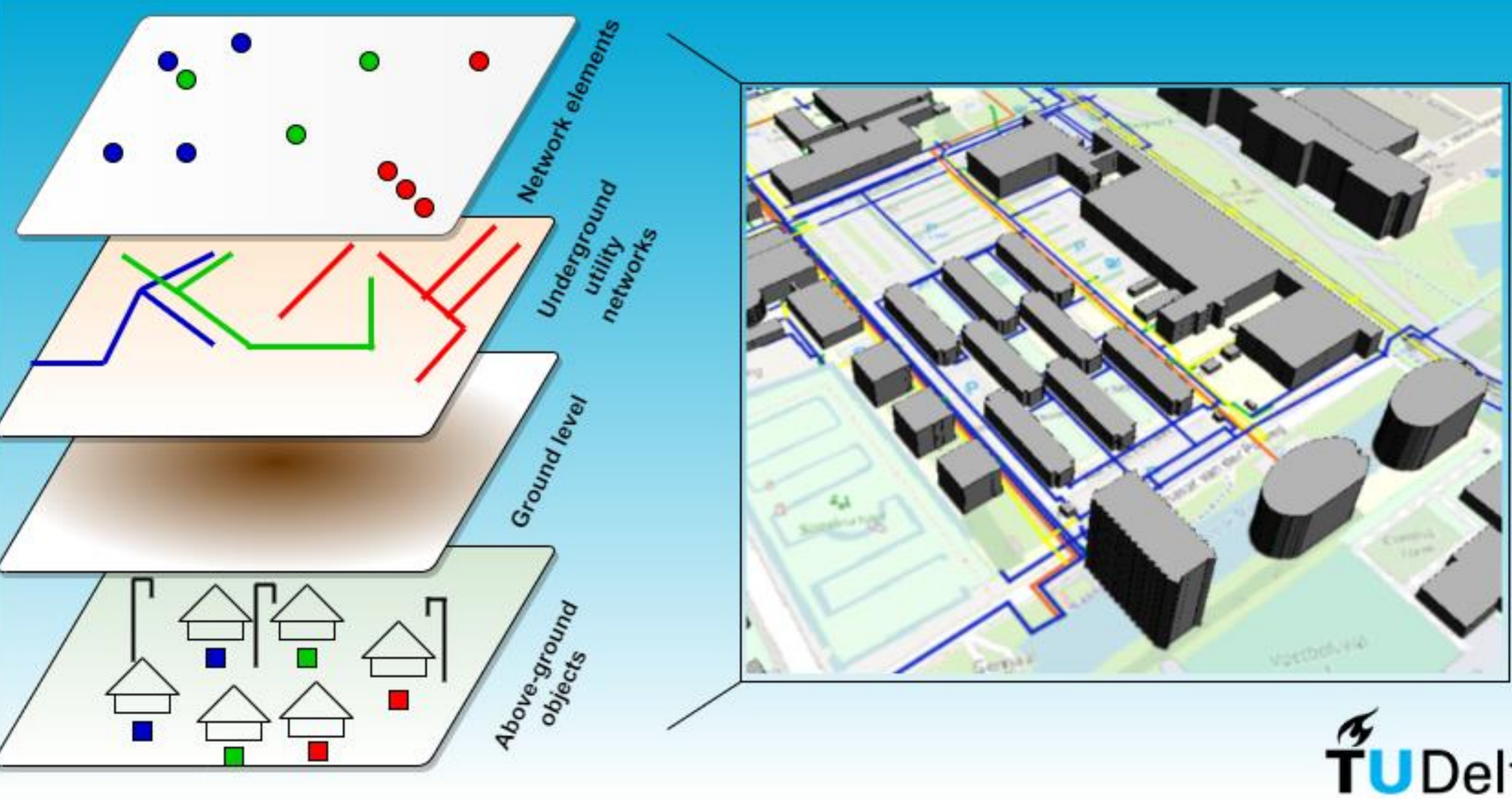
MSc thesis in Geomatics for the Built Environment

Integrated modeling of utility networks in the urban environment

Androniki Pavlidou

2022



TU Delft

INTEGRATED MODELING OF UTILITY NETWORKS IN THE URBAN ENVIRONMENT

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

Master of Science in Geomatics for the Built Environment

by

Androniki Pavlidou

June 2022

Androniki Pavlidou: *Integrated modeling of utility networks in the urban environment* (2022)

The work in this thesis was made in the:

3D Geoinformation group
Urbanism
Faculty of Architecture & the Built Environment
Delft University of Technology

Supervisors: Dr. Giorgio Agugiaro
Prof.dr J.E. Stoter
Co-reader: Dr.ir. B. van Loenen

ABSTRACT

In recent years, there has been an exponential growth in the need for 3D spatial information, particularly 3D models representing the geographic information of the urban environment. The existence of a comprehensive integrated 3D model representing the condition of the underground utility networks in accordance with the above-ground city objects is crucial in the ever-increasing infrastructure demands. To acquire such models the related spatial information (geo-information) is required. This information and its quality determine the quality of the models since it describes the functionality, physical properties, semantic details of the urban features as well as the between them relationships (if exist) and interconnections. Currently, there are available various 3D models that, however, are limited in the representation of objects at a city level, with the corresponding underground information about the utility networks supporting cities' functionality being restricted and/or underdeveloped. Although the modeling of underground utility networks is under development, there are available models that allow for the mapping of the existing information to predefined schema. However, these models have some limitations that restrain the display of the integrated condition with the above-ground city objects as well as they do not support (completely) the interdependencies between networks of different uses.

This thesis focuses on how to develop a three-dimensional model of underground utility networks integrated with above-ground objects, in order to utilize the model in real-world scenarios (disaster management, cost-effective routes). These networks concern sewage networks (and the stored sub-networks), while the objects are related to the buildings of the study area. For the research, vector datasets were used, related to the existing underground utility networks of a limited area of the Delft University of Technology campus, as well as network elements related to them.

For the examination of the proposed model creation, a methodology was developed that could constitute a general approach for relevant applications, considering the condition of the available vector datasets of the similar and/or relevant content. This methodology is divided into seven stages that concern the preparation of the research and data collection process, the statistical and spatial analysis of their content, the integration and cleaning approaches to reconstruct invalid information, and the creation of a relational database to store them. Finally, based on the results from these steps model, routing analysis and functionality were implemented to examine the effectiveness of such model in real-world scenarios. According to the results and limitations that arose, suggestions for the harmonization of the data are addressed as well as future application proposals.

Experimentation demonstrates that the developed methodology lead to the creation of a model that, although it represents a simplification of the existing geographical information, it can successfully be used for the implementation of the proposed case studies (disaster management, cost-effective routes). However, taking into account the poor quality of the input data, the model necessitates improvements in order to be used out of the box for other applications, as well as to verify its compliance with available standards supporting the mapping of the including information.

Keywords: Underground utility networks, PostgreSQL, PostGIS, Routing, pgRouting analysis, OGC standards, CityGML Utility Networks ADE, KLIC

ACKNOWLEDGEMENTS

I would like to warmly thank my parents and friends, who stood by my side throughout my master's degree, as well as all of my supervisors for the guidance and advice they gave me during the thesis implementation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 INTRODUCTION

In recent years, an exceptional effort has been made to utilize geospatial technology in order to model reality. This modeling concerns the simulation of entire cities, individual installations, large physical systems, accompanied by their dimensions. This simulation, provided in digital form, can be characterized as the "digital twin" of the real/natural condition, in terms of spatial (geographic) information, and it aims to represent, virtually, real-world objects, processes, relationships, behaviors as well as interdependencies [Andrews, 2021]. The reason for modeling the existing information along with data as well as reports is their contribution to improving business processes, optimizing operational efficiency, reducing risk, improving decision-making using automation to predict results [Campbell and Wachal, 2021], and moving to evolve into a comprehensive digital era. In the context of the digital era, the representation of reality is accomplished through the use of geographic information systems (GIS), which capture, store, control, and display data related to positions in the urban environment [Evers and Editing, 2017].

With the help of GIS, data from different sources can be combined to create a complete model that allows people to better perceive, analyze and understand existing patterns and relationships of real-world objects. With the development of technology, tools, and methods that collect various data, the visualization of reality is digitally enriched more and more, while the demands for more reliable models/simulations accompanied by attributes that characterized them (e.g. dimensions) are increased. Most of the current, complete, GIS representations are available in 2D, thus simplifying the actual aspects of the objects. However, significant upgrades have been made to the current 2D information, adding the 3rd (and/or other dimensions) dimension to the displayed objects.

Indeed, there has been a significant development in the mapping and displaying of city objects and this visualization has been enriched to such an extent that large-scale city simulation models are now available, including important detail of various levels. The transition to higher levels of information demands, however, the existence of sufficient and correct information at the basic levels (2D), in order to constitute, then, the main geometric unit for the creation of a three-dimensional system of geographic information (GIS) or of higher dimension (e.g. 4D-time) [Kocaman et al., 2022]. This evolution, in terms of mapping various dimensions, is currently available, mainly, at a city level, including above-ground city objects, with the related information concerning below-ground condition being absent or quite limited. Taking about 3D city models means the existence of the digital information about all the objects from which the cities consist of, such as buildings, infrastructures (below and above-ground), and other physical systems (e.g. water systems). Having a complete city model should entail the existence of detailed information for both conditions (above-below ground) regarding the location of the existing objects, their components, materials and costs, and/or other information about the construction. Not only the spatial information itself but also the parameters and attributes (e.g. materials, costs) related to it are necessary for the practical use of such a model. Considering that a city is made up of the infrastructure ensuring its normal urban activity, the costs accompanied by these infrastructures should be stated. Especially in the design phase of a project, accurate cost estimation is vital [Du et al., 2006].

Given the rapid development of smart cities and the need for further sustainable development of the urban environment, the need for an accurate location and mapping of the existing underground state of utility networks is revealed. The expansion of the urban fabric involved in excavations and trenches, results in the demand for knowledge of existing underground utility networks, in order to avoid interruption of services and other costly damage [Tabarro et al., 2017]. Given the complexity of underground networks, there is a need to map them in such a way as to allow changes to be made and to keep their spatial and thematic information up to date [Tabarro et al., 2017]. As the space of the city develops, the need for the existence of maps depicting the underground space seems greater, since this space will need to be developed to the same extent in order to serve the needs of the cities. Mapping accurately the underground condition and the current state of the existing networks of different uses leads to a better understanding of the aspects of the underground space planning [Yan et al., 2018] and allows for more efficient management of it.

Underground utility networks provide infrastructure services related to the urban environment. These facilities, located beneath the ground surface, are responsible for the proper function of the urban fabric as they offer services for the transmission of electricity, telecommunications and data, sewerage and water, gas, and/or other public goods [Pedia, 2021]. Their maintenance is vital, as any disruption will cost normal urban activity. The knowledge of their exact location can prevent harmful accidents while at the same time, assist in a more efficient installation of new utility lines along with existing ones. Therefore, the need for detailed utility network maps as well as digital data models is revealed to represent, exchange, and store their spatial information, with the ultimate goal to manage their relationship to other network systems as well as to integrate them, and better understand their interaction between city entities [Hijazi and Ihab, 2017]. Moreover, apart from the mapping and integration of the available geographic information, the related topology is a crucial component that should not be neglected. The existence of a coherent topology of geographical objects implies the representation not only of the individual objects but also of the relations between them, ensuring the communication of the relations of interaction and/or interdependence.

1.2 PROBLEM STATEMENT AND MOTIVATION

The needs of the urban environment in combination with the constantly evolving technologies are the basis for the development of an accurate and integrated model of urban space. This model will contribute to a better understanding of existing and/or future relationships while also being a source of reference for new environmental projects, urban administration as well as disaster management. The modeling of the underground infrastructure aims at the transfer from the geographical information to its topological representation including the relations that exist between real-world physical entities.

The third dimension in the simulation of underground utility networks facilitates the separation of different uses, highlighting the different levels at which they should be placed, reducing their complexity, while their connection to city objects, contributes to a better understanding of the interdependence between the below and above-ground infrastructure. In recent years, the interest in a comprehensive 3D model is increased significantly [den Duijn X. et al., 2018]. However, despite the development that has been made to 3D modeling of the urban environment, relevant information about the networks existing beneath the ground surface is very often still rather poor or incomplete. More specifically, although some 2D representations are available, the vertical information corresponding to the depth of the pipes/cables is often still missing. This absence makes it difficult to separate them both in terms of height and use, as well as for their topological analysis, accounting for the existing relationships (e.g. joining water pipes for deposition in a natural recipient- sea). However, before focusing entirely on 3D models, the available 2D models must be complete both in terms of content (features) as well as in terms of quality and quantity (content related analysis and number-based/statistical analysis respectively). Looking at the relevant literature, it was found that there are a number of common "typical" problems in today's datasets that still hamper the immediate upgrade to the third dimension. Specifically, some of the common problematic points are:

1. Insufficient or partial availability of metadata describing the current state of available data and its content.

2. The low quality of data, both in terms of content (e.g. empty records, incomplete feature information) and of quantity. Especially when it comes to infrastructure, emphasis should be given to all the objects that make them up (e.g. networks accompanied by the respective network connection points-networks elements).

3. The incomplete and/or totally absent information of the most important attribute of the beneath ground surface networks concerning their location (depth-3rd dimension).

4. Digitization errors exist in many of the available vector datasets. The current state of vector data, for the most part, contains errors that make it difficult to use them directly to develop any application that requires the creation of a correct topological model.

5. Lack of relationships that reflect reality. Here the link between the networks of different use and/or with other objects (city objects) is missing, while their embedding into 3D urban space is not supported.

6. The non-compliance with available standards. In this case, the data can not be exchanged immediately and thus their interoperability is deprived.

In the city of Delft in the Netherlands, although a 3D city model is available, the corresponding underground model as well as the connection between the above-ground condition, is missing. Creating a holistic 3D model might be beneficial for a number of applications, (e.g. where the 3rd dimension plays a vital role). Given that underground networks serve terrestrial installations and facilities, knowledge about their uses and internal connections is very important. In particular, taking into account the strong interdependencies between a network and the corresponding service, the destruction of one (e.g.) network can directly affect, other networks and/or the related above-ground facility, in a so-called cascading effect [Becker et al., 2011]. In the case of a disaster, natural or human-made, a comprehensive 3D model will contribute to an immediate estimation of the extent of the disaster as well as to the decision-making process that will lead to the development of critical solutions. Additionally, for future installations and planning, these models are necessary, since they allow for risk analysis and a-priori preparation of possible emergency situations, and suggestions for potential solutions, something that makes provision for efficient disaster management [Becker et al., 2011].

It is noted that the data concerning the underground infrastructure of the test case area, were not available online but they were provided by the Delft University of Technology and specifically from the Directie Campus & Real Estate Department.

1.3 OBJECTIVES AND RESEARCH QUESTIONS

The current thesis will be guided by the following main research question:

How is it possible to model underground utility networks in 3D, integrated with the above-ground objects, such that they can be suitable for multiple uses? - routing analysis

The aim of the present research will be to study the possibilities of "upgrading" the underground networks in 3D, using available tools. At the same time it will be examined if and how it is possible to follow one of the international standards that have been developed for this purpose, considering the content and the quality of the provided data. In particular, the desired 3D map will be obtained either by directly using the data of the underground networks in the third dimension (if available) or by specifying a way of extracting the 3D information by combining data from different sources (e.g. usage of a Digital Terrain Model of the study area). Having the three-dimensional information, the possibility of the direct connection of the underground network with the above-ground objects will be examined. (e.g. water pumps with the corresponding water pipe).

In order to implement this integration, certain steps will be followed regarding the understanding of the existing information and the physical connectivity of the networks/ sub-networks with each other, clarifying their hierarchy and their interdependencies with the above-ground condition. In order to achieve the last option, the following sub-research questions were defined:

- *How to represent a direct connection with the above-ground condition?*
- *Is it possible to achieve that connection?*

The research aims to highlight the importance of a complete mapping of underground utility networks accompanied by their interconnection with above-ground objects. Such a model will allow for the development, for example, of disaster prediction techniques and possible alternative scenarios (one-pipe), plans for the installation of new underground utility networks, or extension of the existing ones. Attention is paid to the creation of a three-dimensional georeferenced map of a subset of existing utility networks (e.g. sewage) in an area of the Delft University of Technology campus, to analyze the current situation and propose a harmonization strategy, by means of existing data models.

1.4 THESIS OUTLINE

The core content of the present thesis is divided into six main chapters. Specifically, Chapter 2 provides an outline of the scientific study that was conducted as part of the current project, and attention was paid to the proposed techniques used for the development of utility network models and their integration. Chapter 3 is reviewed the methodology adopted to mitigate the thesis research questions and they have described the particular steps that constitute it. More details about the technical implementation of the proposed methodology are given in Chapter 4, where also the case study area is determined, while in Chapter 5 are presented the results of the examined application. Finally, in Chapter 6 are formed the general conclusion and the research questions are addressed, that are followed by constructive ideas and recommendations for future research.

2 THEORETICAL BACKGROUND & RELATED WORK

2.1 THEORETICAL BACKGROUND

Underground utility networks have concerned the engineering community for years, as they constitute an integral part of the urban environment and its infrastructure. The best possible representation of the current situation contributes to the improvement of various business processes, reduction of future risks, optimization of operational efficiencies, as well as enhancement of the decision-making procedure, with the goal to automate the prediction of possible outcomes [ESRI, 2021]. In general terms, a utility network can be characterized as a hybrid geographically diverse service that delivers subscribers with energy, water resources, telecommunications facilities,and other services. It encompasses local councils, economic entities, and customers, as well as the nature of their relationships [Toktoshov et al., 2018]. Given the degree of importance of utility networks for the smooth operation of cities, the knowledge of underground location and digitization is becoming more and more necessary. In the last decades, significant efforts have been made for the location of manufactured objects using GIS technologies, which are evolving rapidly to meet current needs.

The simulation of utility networks is more complex compared to other types of networks (e.g. transportation), as they are interacting with more than one object (below or above ground) and they cannot be characterized as unique units. Thus, their location below the ground (depth) in combination with their interdependencies with the above-ground objects necessitates detailed digital performance [Toktoshov et al., 2018]. Nevertheless, the nature of utility networks in terms of their use is constant, in the sense of patterns (e.g. depth per network, the distance between pipes, etc), and therefore their modeling allows for the development of certain standards capable of delivering their information. In this regard, different models have been developed with common features and *'roots'* but developed by different organizations that follow customized standards (see Chapter 2.2).

Developing and enforcing standards for network modeling is the basis for developing models that can be used, shared, and reused over time while allowing for their interoperability. However, considering utility networks exceptional structure, to obtain a uniform output, they must be presented and stored using ontologies [Fossatti et al., 2020] that can be regarded as a specific domain data modeling standard (e.g. INSPIRE, Industry Foundation Classes (IFC), CityGML Utility Network ADE). Based on the ever-changing and growing needs of the modern urban environment, these standards need to be adapted to support all available real-world information that needs to be digitized. This becomes more clear when it comes to utility networks, since the majority of times they are treated as a two-dimensional entity on a plane. However, this does not perfectly reflect reality, because of the interrelation with the other objects (below and/or above-ground).

In order to simulate utility networks, predominantly graphs and/or other network objects are used as vector data structures. A graph data structure consists of a unique and dynamically modifiable set of vertices (nodes/points), but also a set of ordered or unordered pairs of edges (links/lines). The vertices of the graph could be either inner graph components or foreign elements indicated by integer or pointers [JavaTPoint, 2021]. The main goal is to transfer the real geographical information in the form of a graph capable of capturing the functional, structural, and topological properties of the simulated objects [olde Scholtenhuis et al., 2018]. At this point, the necessity of models able to represent the different levels of detail per network as well as the relationships between them (e.g. intersection between networks or with the above-ground objects) is revealed. Based on the above, models have been developed that serve this purpose (see Chapter 2.2), as efforts are constantly being made to optimize their providers, for the best possible representation of reality.

2.2 RELATED WORK: OVERVIEW OF EXISTING DATA MODELS

2.2.1 INSPIRE network

For the representation of utility networks, the Infrastructure for Spatial Information in Europe (INSPIRE) has developed an individual spatial data theme (1/34). The aim of that theme is to support the easiest accessibility of the available spatial and/or geographical information for various purposes in terms of sustainable development [Lieberman, 2019]. Specifically, the theme *Utility and Government Services* contains information concerning primary aspects on a range of administrative and social services of public interest. All the provided information follows the INSPIRE rules for interoperability of spatial datasets and services. These datasets are accompanied by data specification guidance documents that are based on UML data models developed by the INSPIRE Thematic Working Groups. In the case of underground utility networks, the UML model provides information describing the process of designing and integrating networks into existing infrastructure. The Utility networks model belongs to the *family* of Generic Network Model but it has some extra information for each network element (e.g., UtilityLinkSequence, UtilityNode and UtilityNetwork) that is accompanied with its specific application schema [Utility and governmental services, n.d.]. However, this kind of model has some open issues that need to be confronted. One issue is that utility networks are considered as separated objects and they do not take into account the other networks. This results in partial development of the underground relationships. Additionally, given the main concern for achieving the interoperability of the provided datasets, the use of standards for legal reference or contact information as well as the harmonization of the temporal information are other issues that should be covered. Finally, the most important issue trying to be covered is the need to host 3D data, which means the ability to integrate the utility networks in 3D [Utility and governmental services, n.d.] (Figure 2.1). Currently only 2D topological relationships between the networks are supported.

2.2.2 CityGML Utility Network ADE

As it is stated from [Becker et al., 2011], it is of vital importance to integrate the utility network models into their 3D urban context in order to make them functional for multiple uses. For that purpose they proposed a CityGML application domain extension, Utility Network ADE. This extension empowers CityGML by representing supply and disposal networks in 3D city models. Starting with the CityGML, it is an Open Geospatial Consortium (OGC) standard that defines a con-

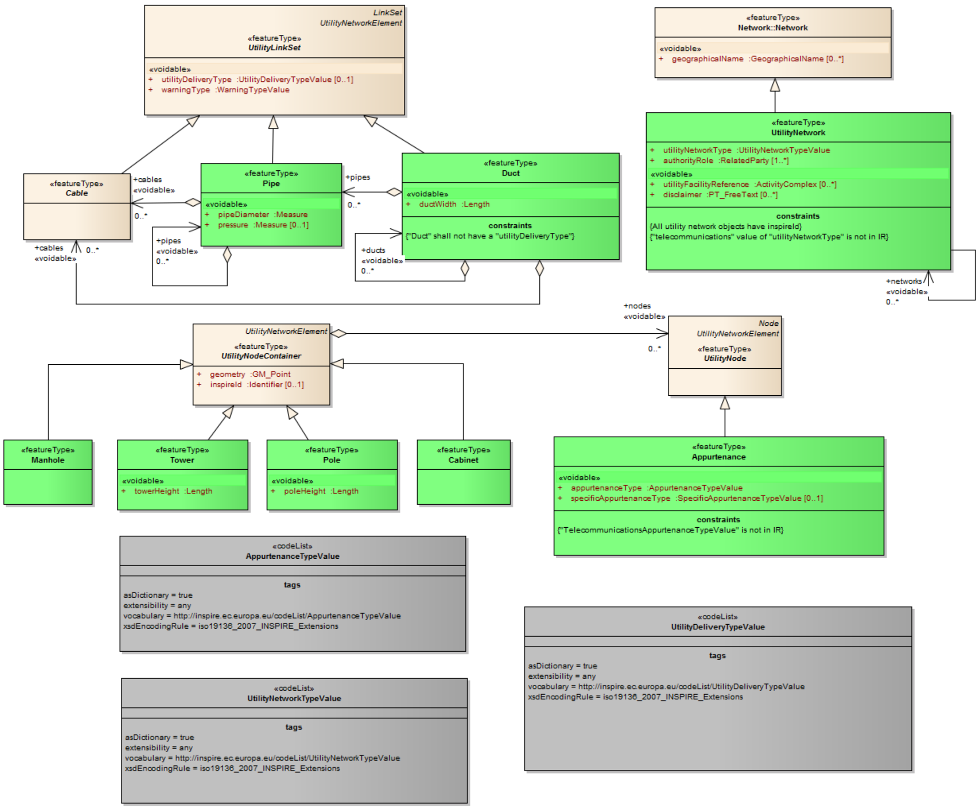


Figure 2.1: INSPIRE Consolidated UML Model

Source:`https://inspire.ec.europa.eu/Themes/136/2892`

ceptual model as well as an exchange format (XML-based) for the representation, storage and exchange of virtual 3D city models[Consortium, 2021]. The corresponding Utility Network extension follows the same standards and has contributed to the addition of new properties to existing CityGML classes that allow their integration This integration lead to an urban model that is spatially and semantically consistent, considering that all the urban objects are characterized by the related geometry, semantics, topology as well as hierarchies [Widl et al., 2021]. Utility Networks ADE, supports this representation and specifically both 3D topographic and topological representation (dual representation) as well as functional modeling of hierarchies [Kutzner, 2019]. In this model, there are specific classes and relationships representing the network entities used in the various types of utility networks and commodities [Adolphi et al., 2013]. Specifically, five modules have been developed for the Utility Network ADE structure [Kutzner, 2019]:

1. **Network Core**: Defines the topological and/or functional model (graphs) as well as the topological model (feature and network)

2. **Network Components**: Defines the individual network components based on use (e.g. distribution objects for transport, protection elements relevant for network security)

3. **Feature Material**: Defines the material of the above mentioned module. Specifically, it is defined the exterior, interior and the filling materials of them.

4. **Network properties**: Defines the product transported by the networks and their corresponding characteristics.

5. **Functional characteristics**: Defines the different functional concepts (e.g. supply area, functional role).

The main goal of the modules is be applicable to all types of utility networks. However, some of them are currently under development given their complexity (e.g. specific packages for district heating, electricity, gas, freshwater) [Kutzner, 2019].

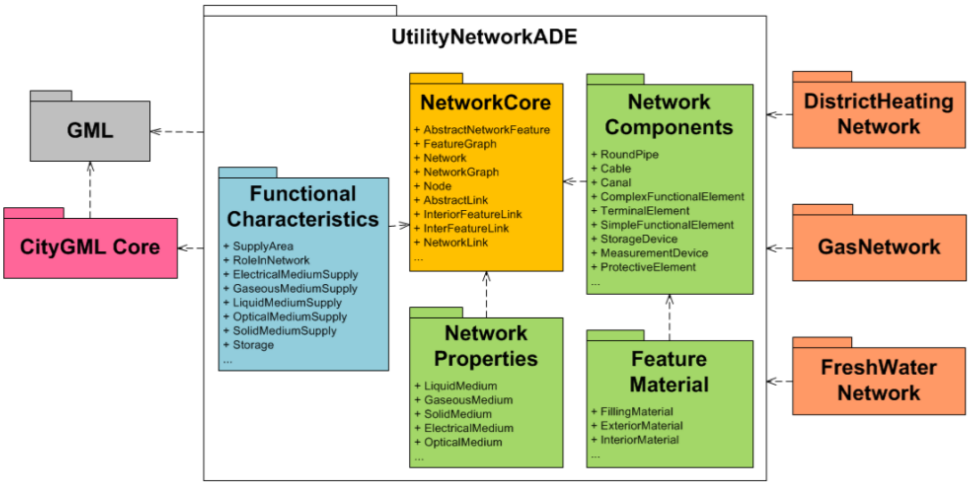


Figure 2.2: Utility Network ADE

*Source:*https://www.asg.ed.tum.de/en/gis/projects/citygml-utility-network-ade/

2.2.3 Industrial Foundation Classes -IFC

Industrial Foundation Classes is an extra open international schema developed for exchanging data concerning building information model (BIM) [TC, 2018]. This standard contains definitions about the cover data needed for buildings over their life cycle. IFC represents building structures accompanied by their properties, 2D and/or 3D geometry and utility networks [van Oosterom, 2008]. However, given their main functionality, these networks are limited only to the level of buildings, considering as networks a limited part of the whole network. That part is related to the building service system (inside building networks) [van Oosterom, 2008]. IFC includes a connectivity concept, which allows the representation of the relationship between the different network objects that are categorized according to their role in the network. This connectivity corresponds to both the physical as well as the logical connection between the elements and the service items (on the level of their ports) of a building, accordignly. Its structure consists of several layers [Borrmann et al., 2018]:

1. **Core Layer**: Defines the primary classes of the IFC data model. It is used as a reference for all the layers and it defines the basic structures, relationships and concepts that can be used (and re-used) by the classes.

2. **Interoperability Layer**: Defines the connection between the core of the data model and the domain-specific schemes. Other classes are defined, deriving from the classes of the Core Layer and they are used to support the various application schemes (e.g. walls, windows of the building- IfcWall and IfcWindow respectively).

3. **Domain Layer**: Defines more specialized classes and it forms the leaf nodes in the hierarchy of inheritance. The classes specified in this layer cannot be referenced by another layer and/or another domain-specific schema.

4. **Resource Layer**: Defines schemes that detail basic data structures that can be used throughout the entire IFC data model.

The limitation of this model, as far as the integration of city scale network is concerned, is the fact that it is restricted to model utilities at the building level [Hijazi and Ihab, 2017] and thus when it comes to a larger scale (city level), the underground condition cannot be supported by its classes.

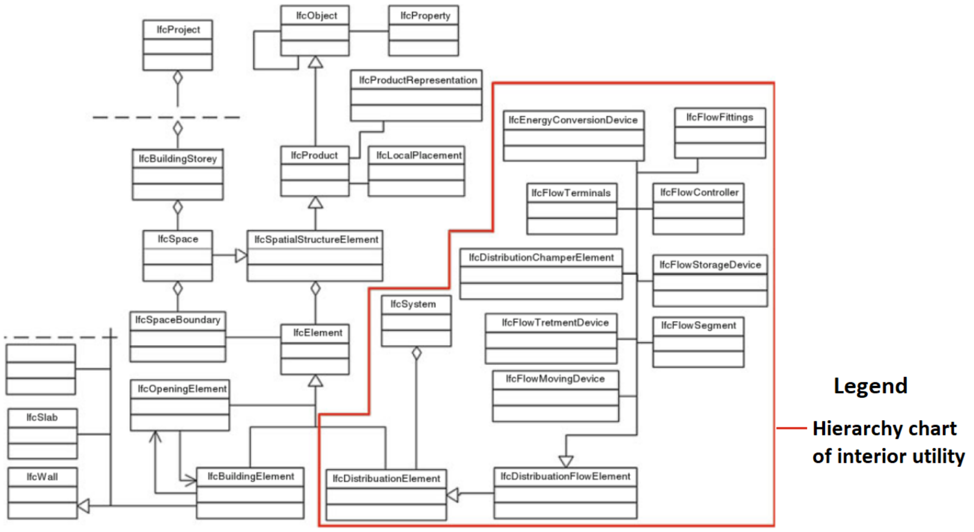


Figure 2.3: IFC inside building utility network schema

Source: Advances in 3D Geo-Information Sciences edited by Thomas H. Kolbe, Gerhard König, Claus [Hijazi and Ihab, 2017]

2.2.4 ESRI Geometric Network

ESRI has developed its own model for the utility networks. In this case, the defined model is based on a core concept, known as *Geometric Network*, which includes a set of connected edges and junctions in order to represent and model the behavior of a typical network infrastructure (above and underground). In more details, there are geo-database feature classes that support the definition of the geometric network and the corresponding rules that provide information about the way the resources flow through the geometric network (physical connectivity and logical relation of the network) [ESRI, 2017]. In the case of utility networks, ESRI offers a scalable network data model (Figure 2.4), where there are only two types of networks [JOAQUIN MADRID, 2017]:

1. **structure Network**: Supports the infrastructure of one single network. It includes three feature classes based on the type of the geometry (e.g. point, line, boundary).

2. **Domain Network**: Supports the modelling of the assets delivering the commodity. It includes five classes separated, as in the previous case, by the feature geometry. There are three classes for point geometry and the rest for lines.

The above structure derives from the geometry of a network that is constructed from edges as well as junctions (points) in 2D. It supports the representation of a real world utility object as one feature, while similar kinds of features are possible to be represented by the corresponding feature classes [ESRI, 2007]. The logical connectivity between the features is implemented by a set of tables. The limitation of this model is the fact that it is semantically rich only for the 2D topography representation of the networks, thus corresponding representation of these objects in 3D is missing [Hijazi and Ihab, 2017].

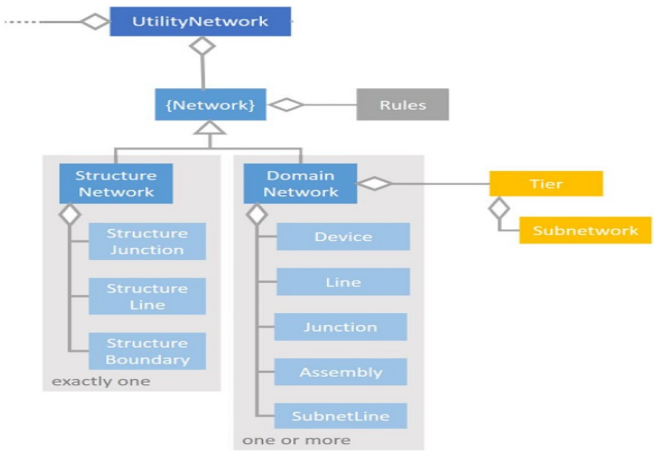


Figure 2.4: ESRI Geometric Network

Source: Esri Utility Network – Simple and Scalable Network Data Model https://sspinnovations.com/blog/esri-utility-network-simple-and-scalable-network-data-model/

2.2.5 Model for Underground Data Definition and Integration -MDDI

In addition to the above mentioned models, MUDDI (Model for Underground Data Definition and Integration) model is another conceptual model that is intended to serve as a basis for integrating underground data from multiple sources, systems, and schemes [J. and Roensdorf, 2020]. As the previous models, MUDDI consists of core entities that serve different cases. The aforementioned entities are further specialized and utility networks are supported, as well as their attributes [Lieberman, 2019]. MUDDI core model consists of various packages and features including the ones concerning networks, where network entities and relationships are be described (Figure 2.5). In more details, these entities, depending on the application, are enhanced in order to support certain types of (e.g.) input data, geometry types (e.g. 2D, 3D, voxels), network connectivity, and functionality. In this case, the basic network (IGraph) works as a container for other network elements and provides their topological relationships, which helps later on to represent the connectivity structure between the links and nodes [Lieberman, 2019]. The current status of this model is not the final one, since it is active development and focuses on the extension of the original classes and not on the replacement of existing models [Lieberman, 2019].

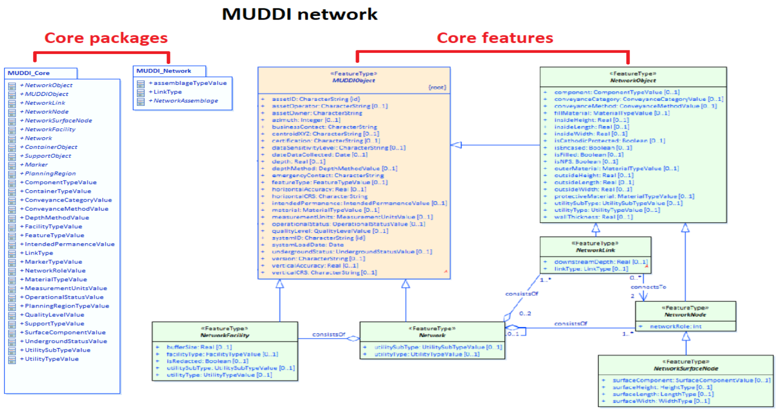


Figure 2.5: MUDDI utility network environment

Source: Model for Underground Data Definition and Integration Engineering Report `http://www.opengis.net/doc/PER/MUDDI`

2.2.6 Kabel- en Leiding Informatie Portaal -KLIC

In the Netherlands, another domestic system for the presentation of underground networks has been developed, which follows certain standards set by the country's cadastre. In more details, Dutch cadastre manages the national Kabel- en Leiding Informatie Portaal (KLIC) (online) application that includes information about the location of the existed cables and pipelines (IMKL), aiming to prevent excavation damage to the sub soil infrastructure [Koppens, 2014]. KLIC application was designed for building and maintaining a system for supporting the digital information exchange. In 2008 a law was adopted -WION- that obligates the networks operators to use a standard information model in order to register their infrastructure [Koppens, 2014]. The application apart from the networks' location, provides the possibility to visualise them through an online environment, after applying for it, allowing this way to obtain a clear view of the current condition [Zeiss, 2014]. This is achieved within the framework of the Information Exchange of above-ground and underground Networks and Networks Act (WION) [GEONOVUM, 2020a]. As the previous models, also here there is a common conceptual framework (IMKL) used for data exchange that works on both cables as well as pipes(pipelines). The information model is based on NEN3610 and therefore part of the NEN3610 family. NEN3610 is a basic geo-information model that has been developed for semantic coordination in the Netherlands and connection to international standards [GEONOVUM, 2020b]. In this case, there is not an individual model and the main purpose of this tools is to illustrate the exact location of the pipes without focusing on the integration of the detected networks. In general terms, KLIC model can be considered as an extension of INSPIRE model. The semantic core of IMKL, however, includes extra information for the Dutch user applications, thus only the information have been added extra to the INSPIRE Utility networks model is shown (geometry and network topology of INSPIRE Utility are excluded) [Herman van den Berg and Paul Janssen, 2021]. It is noted that 3D information is not always and immediately available. The existence of the current application is more about to prevent any danger or economic harm, which is caused by damage to underground cables or pipes (water, electricity and gas lines, telephone lines and oil and gas pipelines) and focuses on the Dutch utility system. An example of the UML diagram used in KLIC case is illustrated in Figure 2.6.

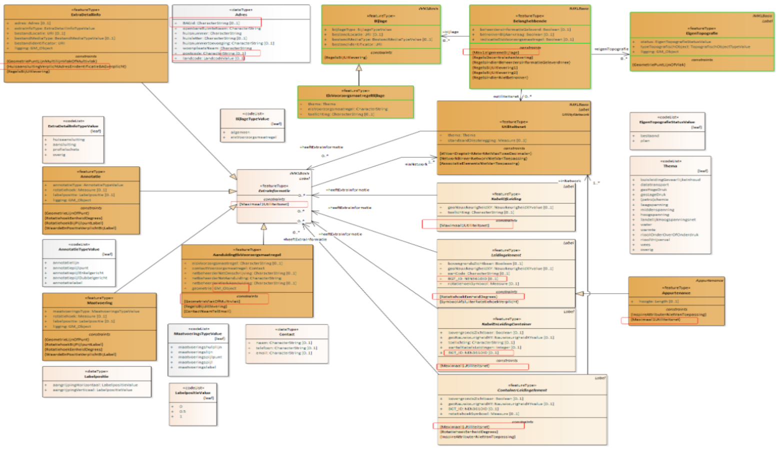


Figure 2.6: IMKL semantic core for WION application.The elements have attributes and a link with additional information.

Source: https://geonovum.github.io/imkl2015-review/IMKL2.x/1-dataspecificatie/#uml-diagrammen

2.3 RELATED WORK IN THE FIELD OF UTILITY NETWORK-SÍNTEGRATION

All the models mentioned above are the basis for the simulation, integration, and analysis of the underground utility networks existing in the fields of the urban environment. There are several types of research that attempt to check the maturity and potentiality of the existing models (e.g. CityGML Utility Network ADE, INSPIRE) as appropriate data models for underground utility network features related to above-ground city objects. From 2018 to 2020, relevant research has been done in the field of the (3D) modeling of utility networks. Below are presented examples of these research focusing on the practical application of underground network modeling, using disparate data sets applied in different areas (study cases).

2.3.1 Rotterdam case

In this research, the aim was the three-dimensional modeling of underground utility networks for integrated management applications, considering networks' relationships with the city objects [den Duijn X. et al., 2018]. Specifically, the city of Rotterdam, the Netherlands, was selected as a case study area and then a further selection was made regarding the types of utility networks used for the practical application. These utility networks are associated with electricity and sewer data, and city objects directly related to the use of the networks were also used, such as streetlights and manholes/covers. The methodology followed in this case has as a starting point the available data in ESRI shapefile format, which were then examined and evaluated in terms of their content. Based on the existed inconsistencies different software and tools were used in order to manipulate spatial operations and handle the data structure as well as their content. In the end, an XML-based file was generated that included the information needed in order to follow the CityGML Utility Network ADE model and the corresponding UML diagram rules and relationships. Additionally, in terms of the datasets, a 3D city model of Rotterdam was also used, given in CityGML format. Once the data were processed, they were stored in a database, 3DCityDB (https://www.3dcitydb.org/3dcitydb/) , to which some routing functionalities were implemented. The result of this research was the successful representation of utility networks in a third dimension, integrated with the corresponding network elements. In conclusion, it is noted that the CityGML Utility Network ADE model is suitable for supporting such applications, but the existing data need to have the structure supported by it. However, there is a need for further investigation since there are several other types of relationships that did not consider in this research [den Duijn X. et al., 2018].

2.3.2 Singapore case

The main objective of this research was the proposal of a framework capable for the underground utility networks workflow organization from the survey phase to data use [Yan et al., 2018]. In order to achieve this goal, the city of Singapore was used as a study area. The paper lists the problems limit the possibility of the direct visualization and management of underground utility networks in the third dimension. Some of the problems are the two-dimensional state of the utilized data, the absence of information regarding the location of the pipes/cables and the respective depths, while also in many cases only the very initial information was available, not the as-built condition. Based on these drawbacks, it was proposed the development of a 3D model which tries to fill existing gaps regarding the availability of this information to different users (e.g. stakeholders, government agencies) as well as the support to cadastral management of the underground utility networks. In order to create a reliable model, it was emphasized that accurate data collection is needed, to replace the existing ones (of poor quality) and then the use of a model that cannot only represent the 3D utility network but also its corresponding topography. CityGML Utility Network ADE was proposed as a suitable model, with the only limitation being the fact that the surveying method followed for data collection is not taken under consideration [Yan et al., 2018], while IFC model was mentioned as a second suitable model, with the limitation this time being the lack of spatial information. On the contrary with the above mentioned research, this was more a theoretical research, proposing only the basic steps needed in order to develop the desired 3D model of the underground utility networks.

2.3.3 University of Twente case

In this research, the attention was paid to the demonstration of the underground utility networks in order to show if it is supported and how a typical street reconstruction project, using CityGML Utility Network ADE with the extra support of the Operation & Maintenance Domain Ontology [Fossatti et al., 2020]. The examined are in this research, is a part of the University of Twente. The datasets used concerned two types of utility networks the gas low pressure (GLP) and district heating supply and return the corresponding (DH) networks, as well as a Digital Elevation Model (DEM) of the study area. The methodology followed in this case consists of four basic steps, which concern:

- The extraction of the relational database that was used *"as the backend of the system using the Unified Modelling Language (UML) class diagram of the O&M Domain Ontology"* [Fossatti et al., 2020].
- The processing of the data to fix existing inconsistencies (digitization errors), in order then to draped them over the DEM.

- Additionally to the previous step, a semantically transformation on the data took place and they were loaded in the end into the database.
- The last step concerned the visualization of the results with the help of a GIS application.

In each of the above mentioned steps further processed was implemented based on the complexity of the problem in combination with the existed information. From this research it was made clear the possibility of the O&M Domain Ontology to support a typical and real asset management tasks (in combination with the CityGML Utility Network ADE) and among others, it was highlighted the importance of having a large amount of data in order to support network operation and maintenance.

2.3.4 Nanaimo (Canada) case

In this research, it was presented the works implemented for the examination of the state of maturity and suitability of the CityGML Utility Network ADE. The orientation of this research differs slightly from the others since attention was paid to the interaction with the extended 3DCityDB and to Routing Analysis. However, in the methodology followed in order to test the suitability of this ADE, they were highlighted the advantages of the model as well as the way that it supports utility networks. The datasets used for that purpose concerned the water network of the city of Nanaimo (Canada), available in ESRI shapefile format, and the Digital Terrain Model (DTM) of the study area. The different layers in this case were stored in the 3DcityDB (`https://www.3dcitydb.org/3dcitydb/`) and thus the data had to be implemented as a database schema following the existing rules available in that geo-database. Apart from the direct mapping, they were tested some routing functionalities with the help of PostgreSQL/PostGIS (`https://postgis.net/`) extension pgRouting. Here it was highlighted the advantage of the CityGML Utility Network ADE to explicitly define the topography and topology of the available data and the fact that it is a versatile model capable to be used in multi-network modelling [Boates et al., 2018].

2.3.5 Summary

From all of the above, it is observed that all cases shared some information about:

- The existing problems related to the data (qualitative-quantitative).
- The main model used to implement the mapping, which was the CityGML Utility Network ADE.

The common denominator of all the data used was their non-interoperable format, their incomplete content, the existing inconsistencies both in terms of geometry and topology, and finally, the fact that it was tried to be mapped the underground condition following CityGML Utility Network ADE rules. Based on the above, the techniques followed were similar and concerning conversions to other -compatible- formats, critical decisions regarding the content of the datasets and the information that should be discarded as well as spatial operations that needed to be implemented in order to obtain topological correct datasets. All the changes that took place, in the datasets, were guided by the UML diagram accompanied by the CityGML Utility Network ADE model. The conclusion from all the research is that in order to implement a complete or potentially complete model it is necessary to gather a lot of information and data and combine them appropriately to achieve the final goal (modeling of underground networks). The common problems mentioned in the above research were addressed in the present thesis as well. The challenging situation of the data was expected to a certain extent due to the neglect of the characteristics of the underground utility networks, both in theory and in practice, since the attention was paid mainly to the above-ground simulations [Emgard and Zlatanova, 2007]. The methodology developed in this thesis was inspired by the above research.

3 METHODOLOGY

The research approach utilized in this MSc. thesis is divided into seven steps, described in seven respective sections (Figure [3.1]), starting with the preparation of the current research and the data collection process, which are followed by their analysis in terms of quality, quantity as well as spatial nature. Based on the results of this analysis, the data are to be integrated and cleaned at the same time in order to reconstruct their content in terms of attributes, geometry and topology. This step is followed by the creation of a relational database used to store the spatial information, as well as to further analyze them. Finally, the case study application is implemented and further discussion of the results and recommendations for future studies are to be formulated. Each module's thorough description is covered in its own section below.

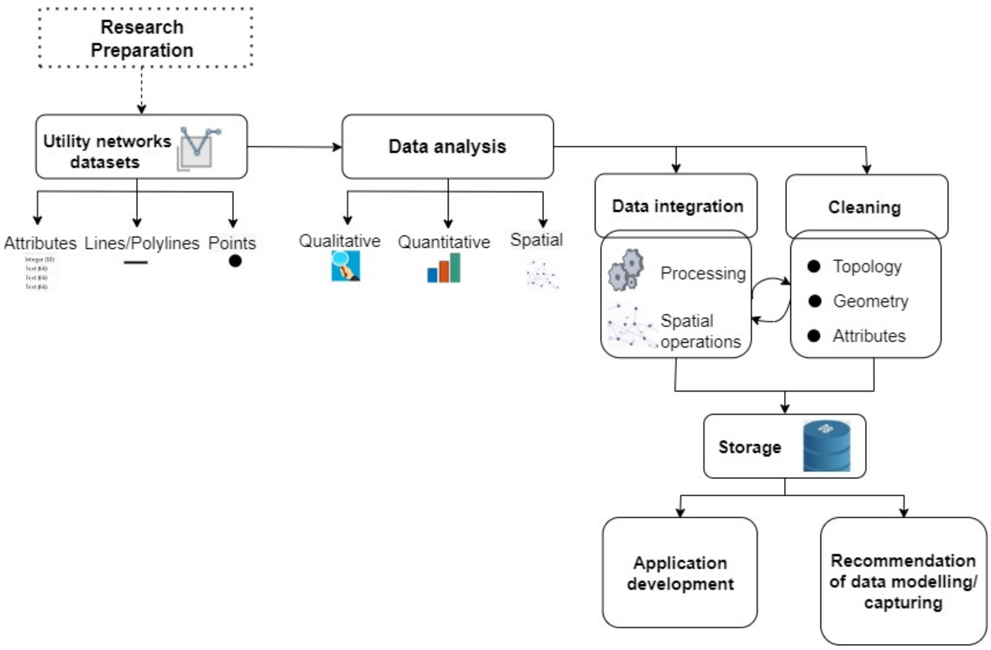


Figure 3.1: Overview of the proposed methodology

3.1 RESEARCH PREPARATION

The first step of the research implementation was the definition of a clear research question, as well as sub-questions accompanied by the identification of the study's broad scope. Specifically, starting with a literature study (see Chapter2) the above-mentioned points were defined, which constitute the guide of the whole research. Based on them, the data required to construct the desired model, as well as the approaches that must be implemented to modify and make them usable were identified. For the data handling, different tools and software were used that were both open sources and/or required licenses. Regarding the former ones, they were

QGIS, Python programming, PostgreSQL, and PostGIS, while as far as the latter one is concerned, was the Feature Manipulation Engine (FME).

3.2 DATA COLLECTION

To start implementing any data-based application there is the need for gathering a set of data relevant to the application, which would be capable of meeting its needs. In the case of the current thesis, where the attention is on the underground utility networks, a corresponding dataset used, provided by the *Directie Campus & Real Estate Department* of the Delft University of Technology. In more detail, the data concern underground utility networks of different types (pipes/cables) existing in the wider area of the TU Delft Campus, that constitutes the study area/case study. For further analysis, however, it was selected a limited part (subset of the underground facilities), taking into account the amount of the available information and its content. All the information is stored in vector data structure that comprises lines/edges and points/nodes, to which are attached a set of attributes/entities. The vector data are geo-referenced in the Dutch reference system, which is the Amersfoort/RD New Netherlands. The attributes of that reference system are:

- **Geodetic CRS:** Amersfoort
- **Datum:** Amersfoort
- **Ellipsoid:** Bessel 1841
- **EPSG code:** 28992

In addition to the provided data, raster data was used concerning the Digital Terrain Model (DTM) as well as a 3D city model of the study area in geo-package format (derived from (`https://www.pdok.nl/introductie/-/article/actueel-hoogtebestand-nederland-ahn3-`) and (`https://3dbag.nl/en/download`) accordingly). In the raster case, each pixel corresponds to a continuous elevation value, which was necessary for enriching the data with the 3D information, while the second dataset was used for the integration of the underground information with the terrestrial (city) objects.

3.3 DATA REQUIREMENT ANALYSIS

Considering that the current thesis is focused on the representation of the underground utility networks, a specific piece of information is needed for the reflection of the existing condition of the networks. A different application would demand distinct requirements based on the related needs. These requirements concern their semantics, attributes, dimensions, geometry, relationships as well as other information describing the real-world object that is to be transferred in digital form. For underground utility networks, the requirements are determined by the objects that make up the different networks. For example, in sewage network (thesis case study) case, is comprised of the pipes, manholes as well as the destination network element (e.g. appurtenances) they are connected to (e.g. building). For these objects, it is necessary to obtain the complete (as much as possible) knowledge of their characteristics (attributes), in order to acquire a complete model. These attributes should be able to describe the network elements (pipes, connection points) in terms of their usage, geometry, topology as well as the between them relationships (Figure [3.2]). In more detail, a network topology or infrastructure, as well as the function for which the network is planned or constructed, are two essential "components" of every network [Mieghem et al., 2010]. A network topology describes how elements, known as nodes, are linked together to form a network. On the other hand, network service (functionality) employs network infrastructure to transfer goods between a collection of nodes. This transportation (of the included on the pipes/cables pipes) depends on the framework that defines the parameters for each application [Mieghem et al., 2010]. Both network components (topology, service) is possible to have their own particular qualities and requirements. Thus it is necessary to have the information related to these features for being able to construct and represent the existing underground networks condition (reality simulation).

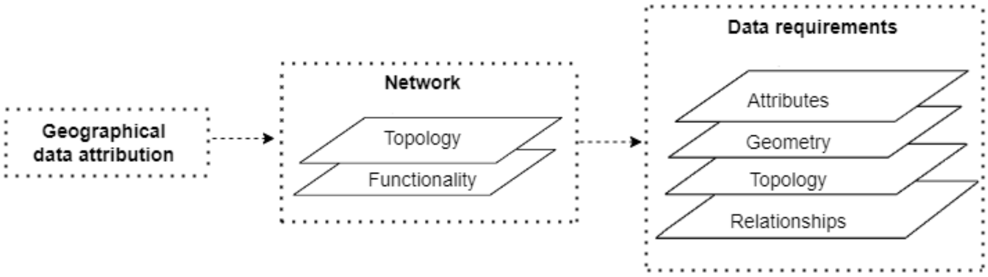


Figure 3.2: Data specification/requirements

Inspired by: [Mieghem et al., 2010]

Specifically, regarding the attributes, is required for the type (usage) of the network to be clarified as well as the types of sub-categories (if exist). Spatio-temporal information about the construction date accompanied by the current status of the network (in use or out of operation) is an additional requirement, since networks' topology and functionality are (typically) time-variant and is important to keep track of changes. In addition to this, a general textural description about functional characteristics is considered crucial information (metadata). Other attributes essential to be known are the capacity of the pipes, the dimensions of the available objects, as well as the flow direction or relevant information about gravity that would allow for the detection/calculation of ground inclination thus, the flow direction could be defined. Another critical point is the information related to the location where the pipes are placed beneath the ground (depth- Z information for underground objects) as well as the related reference surface used for that determination (X, Y, Z - 3D information). In addition to the location, the coordinates of the significant points (surface points that the pipes/cables are connected with) are important in order for the whole network to be correctly geo-referenced. Referring the geometry of the network elements, it is vital for the shapes and the corresponding length information to be recorded. In this way, by visualizing the data, using available tools/software, a general view of the size and volume of this infrastructure is obtained. Finally, as far as the topological information must be stored in the data attributes is concerned, should include both pipes connectivity as well as the relationships they have with neighboring pipes and/or other network elements (connection points). The connectivity between different networks and/or network points (e.g. manholes) can be achieved with the use of specific fields that will internally connect the related information (e.g. in terms of the primary and foreign keys).

Considering the above-mentioned points related to the data specifications/requirements that utility network data should have, through the analysis that made on the datasets used in the present thesis, attention was paid to the existence, degree of completeness, quality and quantity of them.

3.4 DATA ANALYSIS

To understand the data and their content it is necessary to traverse through their stored information (e.g. attributes). Any set of data is distinguished for its completeness, both qualitatively and quantitatively. Starting with the quality of the data, we refer to the elements that characterize them; metadata (transcription) that give a description of the content as well as to its completeness (e.g. clean data free of useless/ redundant and/or irrelevant information- or even wrong) [Lacey and Luff, 2007]. The qualitative analysis of the data includes, also, the organization based on their content or internal connection, in order to be able to retrieve them. Having completed these steps, the process of getting acquainted with the material has essentially begun and therefore it is possible to start doing some preliminary coding in order to process them so that they serve our purpose and make the dataset reliable and valid according to existed standards [Lacey and Luff, 2007]. As far as the quantitatively analysis of the data is concerned, we refer to the process followed to collect and analyze numerical data (see Tables 3.1, 3.2). In more details, the attention this time is paid mostly on finding patterns, predictions, as well as cause-effect relationships between the variables to be studied [VoXco, 2020].

During both stages of the analysis (quantitatively and qualitatively), it was made clear that the provided datasets were not ready for use as their content was vague and/or incomplete. Specifically, the data were given in a geo-package format, containing the vector data themselves, which were not accompanied by metadata describing them. For this reason, their contents were analyzed, traversing through the related attribute table, in combination with the visual comparison in the environment of the software: QGIS version 3.16 and FME Workbench version 2021.1.1. During this examination, attention was paid to all of the included units accompanied by their units of measurement (if available). In more detail, the provided data concern sewage pipes, and cable networks concerning electricity, gas transmission, telecommunication network (signal cables) as well as other particular elements that are relevant to the use of the networks and constitute connection points between them (e.g. water pumps connected with fire pipes). Considering the desired integration of the underground utility networks with the above-ground objects, the analysis of their content was divided into three parts:

1. Attributes: All the attributes stored in the datasets were examined and attention was paid to the existence of the three-dimensional information regarding the Z coordinate (the depth at which the networks are located -and/or mapped/registered), or relevant semantic height information. Additionally, they were reviewed in terms of their completeness and logical values.

2. Geometry: The geometry of the datasets were examined to make a clear distinction between the stored features and their shapes.

3. Topology: The topology of the provided datasets was the most crucial part of the initial analysis since it provides a detailed overview of the connectivity between the features stored in the vector data.

A visual example of the provided datasets and the study area is given in Figure [3.3], [3.4] as well as in Figure [3.5].

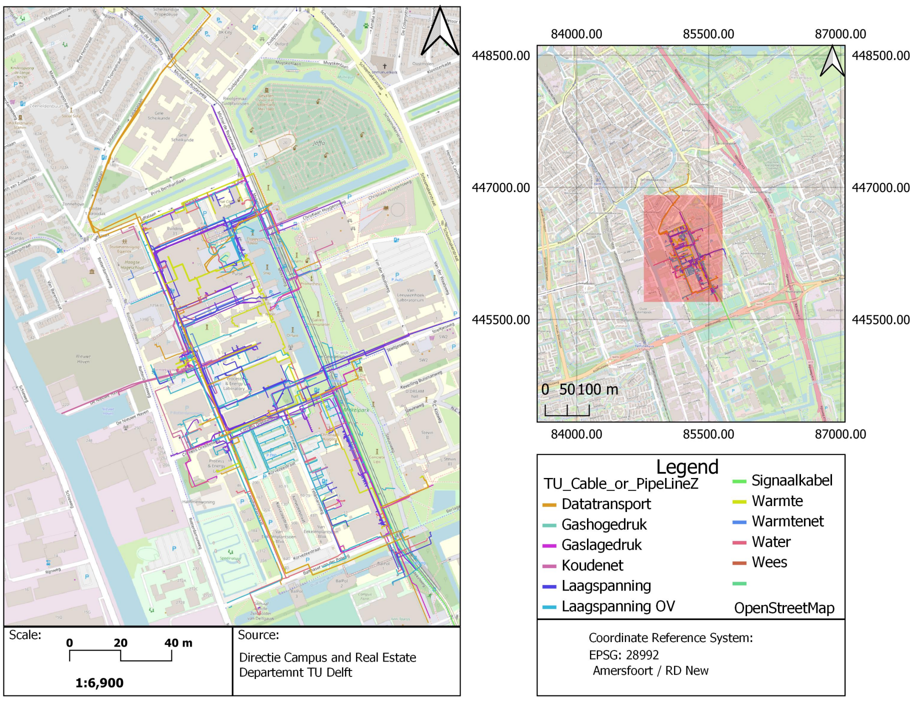


Figure 3.3: TU cable/ pipes network (e.g gas, electricity)

In Figure [3.3] are depicted the different types of cables and pipes located in the area of interest. Specifically, there are thirteen (13) types concerning (e.g.): *data transport, gas (high and low pressure), cooling and heating network, low and medium voltage, petrochemical, sewage network, drainage, drink water and 'other' (e.g. left over) pipes*.

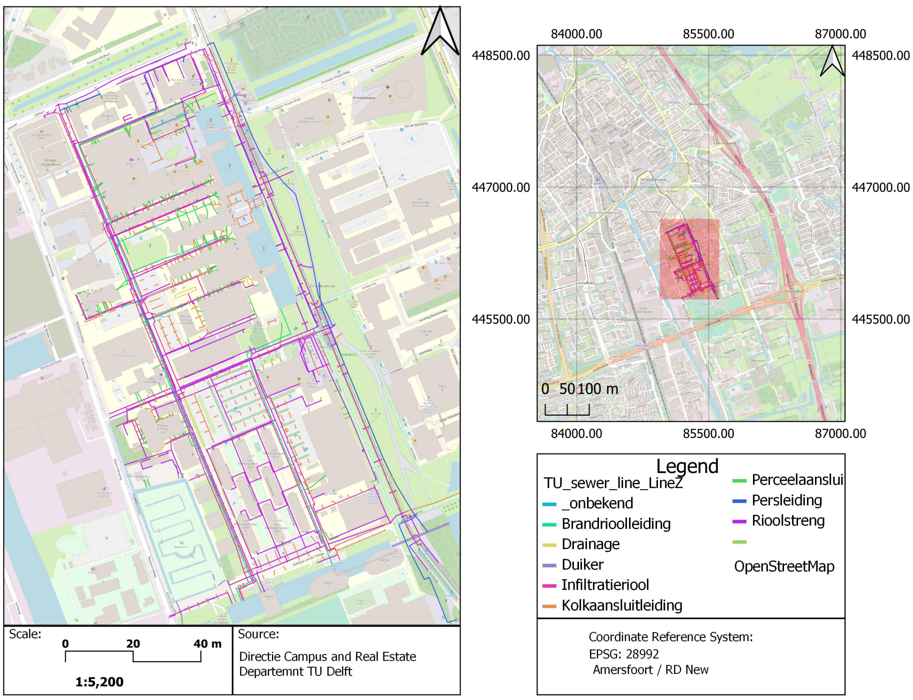


Figure 3.4: TU Sewage network pipes

In Figure [3.4] is illustrated the sewage network existing in the study area. This network includes nine (9) sub-networks concerning *sewer pipes, drainage network, infiltration sewer, fire sewer pipe, field/plot pipes, pressure and overflow pipes as well as diver and unknown use pipes*.

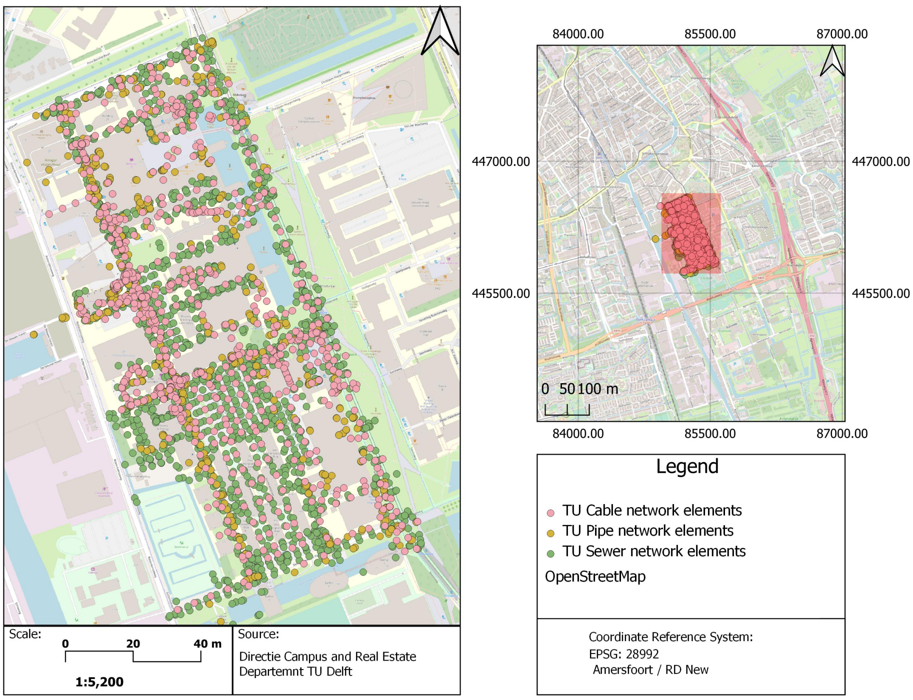
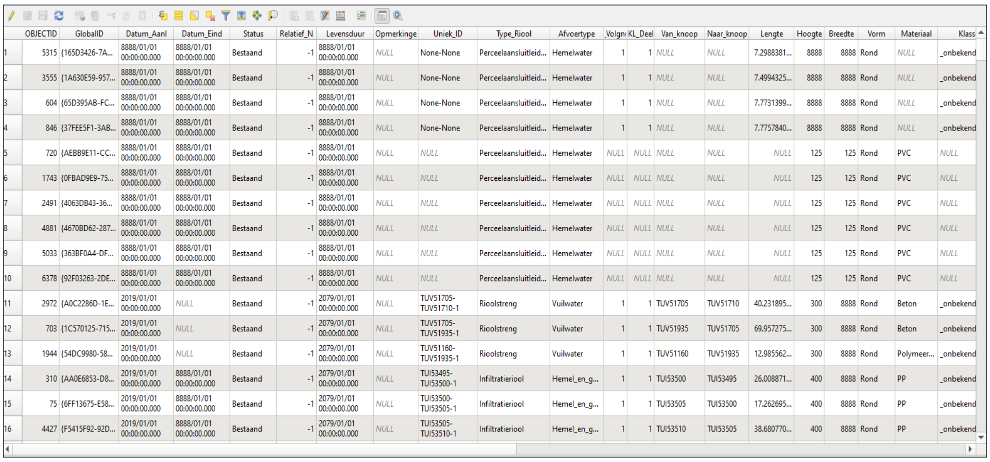


Figure 3.5: TU Sewer network appurtenances

Finally, in Figure [3.5] is given an example of the (sewer) network element points/nodes with whom the sewer pipes of the same use are connected and/or come across with sewer pipes of different use.

3.4.1 Attributes analysis

Considering the format of the data structure of the datasets (vector data), information for their content can be extracted from the associated attribute table since no metadata was available. The absence of metadata affects the immediate acquaintance with the data since they are not accompanied by a description of their characteristics and their management through time. Therefore the only source of information for the data and its content is the features stored in the corresponding attribute table. Reviewing this table, it was observed that in all of the available datasets, there was a list of fields that describe the current state of the utility networks, in terms of their type -usage-, a relative height in relation to a reference surface, status (if they still exist or not), length and width as well as other construction information (e.g. cable/pipe dimensions and material, date of construction etc.). An example of the form of an attribute table is given in Figure [3.6].



	OBJECTID	GlobalID	Datum_Aanl	Datum_Eind	Status	Relatief_N	Levensduur	Opmerkinge	Uniek_ID	Type_Riool
1	5315	{165D3426-7A...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	None-None	Perceelaansluitleid...
2	3555	{1A630E59-957...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	None-None	Perceelaansluitleid...
3	604	{65D395AB-FC...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	None-None	Perceelaansluitleid...
4	846	{37FEE5F1-3AB...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	None-None	Perceelaansluitleid...
5	720	{AEBB9E11-CC...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	NULL	Perceelaansluitleid...
6	1743	{0FBAD9E9-75...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	NULL	Perceelaansluitleid...
7	2491	{4063DB43-36...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	NULL	Perceelaansluitleid...
8	4881	{4670BD62-287...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	NULL	Perceelaansluitleid...
9	5033	{363BF0A4-DF...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	NULL	Perceelaansluitleid...
10	6378	{92F03263-2DE...	8888/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	8888/01/01 00:00:00.000	NULL	NULL	Perceelaansluitleid...
11	2972	{A0C2286D-1E...	2019/01/01 00:00:00.000	NULL	Bestaand	-1	2079/01/01 00:00:00.000	NULL	TUV51705-TUV51710-1	Rioolstreng
12	703	{1C570125-715...	2019/01/01 00:00:00.000	NULL	Bestaand	-1	2079/01/01 00:00:00.000	NULL	TUV51705-TUV51935-1	Rioolstreng
13	1944	{54DC9980-58...	2019/01/01 00:00:00.000	NULL	Bestaand	-1	2079/01/01 00:00:00.000	NULL	TUV51160-TUV51935-1	Rioolstreng
14	310	{AA0E6853-D8...	2019/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	2079/01/01 00:00:00.000	NULL	TUI53495-TUI53500-1	Infiltratieriool
15	75	{6FF13675-E58...	2019/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	2079/01/01 00:00:00.000	NULL	TUI53500-TUI53505-1	Infiltratieriool
16	4427	{F5415F92-92D...	2019/01/01 00:00:00.000	8888/01/01 00:00:00.000	Bestaand	-1	2079/01/01 00:00:00.000	NULL	TUI53505-TUI53510-1	Infiltratieriool

	Afvoertype	Volgn	KL_Deel	Van_knoop	Naar_knoop	Lengte	Hoogte	Breedte	Vorm	Materiaal	Klass
1	Hemelwater	1	1	NULL	NULL	7.2988381...	8888	8888	Rond	NULL	_onbekend
2	Hemelwater	1	1	NULL	NULL	7.4994325...	8888	8888	Rond	NULL	_onbekend
3	Hemelwater	1	1	NULL	NULL	7.7731399...	8888	8888	Rond	NULL	_onbekend
4	Hemelwater	1	1	NULL	NULL	7.7757840...	8888	8888	Rond	NULL	_onbekend
5	Hemelwater	NULL	NULL	NULL	NULL	NULL	125	125	Rond	PVC	NULL
6	Hemelwater	NULL	NULL	NULL	NULL	NULL	125	125	Rond	PVC	NULL
7	Hemelwater	NULL	NULL	NULL	NULL	NULL	125	125	Rond	PVC	NULL
8	Hemelwater	NULL	NULL	NULL	NULL	NULL	125	125	Rond	PVC	NULL
9	Hemelwater	NULL	NULL	NULL	NULL	NULL	125	125	Rond	PVC	NULL
10	Hemelwater	NULL	NULL	NULL	NULL	NULL	125	125	Rond	PVC	NULL
11	Vuilwater	1	1	TUV51705	TUV51710	40.231895...	300	8888	Rond	Beton	_onbekend
12	Vuilwater	1	1	TUV51935	TUV51705	69.957275...	300	8888	Rond	Beton	_onbekend
13	Vuilwater	1	1	TUV51160	TUV51935	12.985562...	300	8888	Rond	Polymeer...	_onbekend
14	Hemel_en_g...	1	1	TUI53500	TUI53495	26.008871...	400	8888	Rond	PP	_onbekend
15	Hemel_en_g...	1	1	TUI53505	TUI53500	17.262695...	400	8888	Rond	PP	_onbekend
16	Hemel_en_g...	1	1	TUI53510	TUI53505	38.680770...	400	8888	Rond	PP	_onbekend

Figure 3.6: Example of existing fields and attributes

In the Figure above are illustrated, among others, the different types of the sewage network (Type_Riool e.g. Rioolstreng= Sewage pipes, Infiltratieriool= Infiltration sewer) and the two types of identifiers (*OBJECTID*, *GlobalID*), which are unique identifiers per attribute. However, these recordings are not all up-to-date, while others are generally not available or even unknown. Specifically, the existence of empty records (NULL values) can be observed or of invalid information (e.g. "8888", which is not a valid value). In addition, the unit of measurement of the data that are physical quantities (measurable) is not distinguished from the table, something that makes it more difficult to perceive their real size, while in the case of different units between the different quantities, confusion is created. Traversing all the attribute tables from the provided datasets it is noted that only 2/6 datasets contain some fully completed information while the rest are incomplete, with missing records (Table 3.1).

Table 3.1: Data statistics

Dataset/ completeness	Total number of rows	Rows with complete information	Rows with incomplete information
TU Cable or Pipe line Z	1407	-	1407
TU Cable element point Z	912	34	878
TU Cable pipeline Z	1407	-	1407
TU Sewer knot point Z	1986	-	1986
TU Sewer line Z	2080	19	2061
TU pipe element point Z	1038	-	1038

It is noted that the above table strictly presents the information that is available or not in the datasets, without separating its correlation with information related to the current application. The information that is considered directly related to the present application- underground utility network 3D integration- is the length-/shape of the underground pipelines, the coordinates of the corresponding point elements, their date that indicates if the data is up to date, their usage (type) as well as semantic height information. As mentioned above the provided datasets concern both edges (lines/polylines) as well as nodes (points) of different uses. These two geometries seem to be connected with each other depending on their use. For example, in the sewage network, there are different sub-categories of pipes such as the fire pipe network, plot connection pipes, gully pipes and others. Fire pipe network is connected, among others, with some nodes corresponding to water pumps (-extinguishers).

This connection is a logical/physical connection between the fire pipe network and the corresponding nodes and it can be seen in Figure [3.7]. However, in terms of the attributes stored in their attribute tables, these two datasets are not connected, internally, as far as primary and foreign keys (shared attributes) are concerned. In more detail, if the two datasets included information that would allow for their internal association, the logical connectivity would be able to get implemented topologically, and thus the analysis, as well as their management, would be more efficient and representative. The initial condition of the datasets did not allow for that interconnection and thus a different approach had to be implemented (see Chapter 3.5).

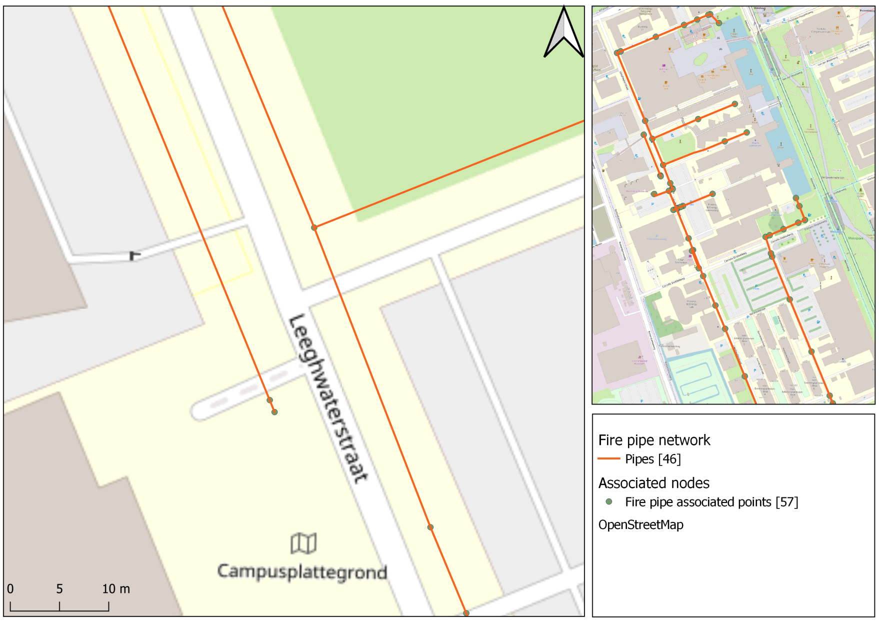


Figure 3.7: Connectivity between nodes and edges of same/relevant use

In addition to the above remarks about the data, there were errors that have a direct impact on their understanding. An example of this situation is shown in Figure [3.8], where the location of the pipe does not correspond to reality. The point that stands out (lower level) and corresponds to the incorrectly written coordinate Z = 32 (m) cannot be located at this level, taking into account the overall form of the network. At this point, this incorrect information existing in the data was either skipped (manual modifications) or fixed based on the frequency, in which this information appears on them.

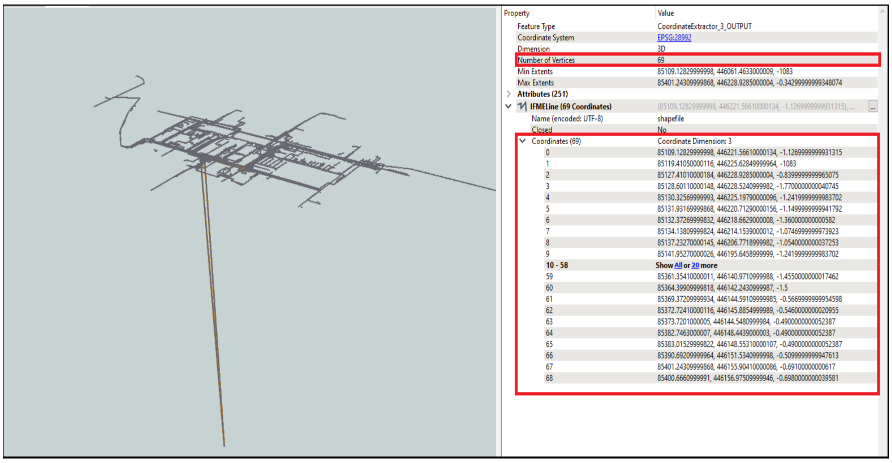


Figure 3.8: Incorrect topology derived from incorrect, input, attribute

3.4.2 Geo-spatial data

Moving forward the analysis, to the extraction of the 3D information, it was observed that only a few records of the datasets provided indeed 3D information (depth- Z-coordinate). Specifically, with the help of Feature Manipulation Engine (FME) software, the coordinates included in the geometry of the datasets were extracted and by implementing a statistical analysis, Table 3.2 was created.

Table 3.2: 3D vs 2D information

Dataset/ completeness	Total number of rows	Missing Z coordinate	Zero Z coordinate	Existed Z coordinate	Percentage of existing 3D information	Percentage of existing 2D information
TU Cable or Pipe line Z	1407	0	1361	46	3.27 %	96.73%
TU Cable element point Z	911	0	909	2	0.22%	99.78%
TU Cable pipeline Z	1407	0	1361	46	3.27 %	96.73%
TU Sewer knot point Z	1986	0	1954	32	1.61%	98.39%
TU Sewer line Z	2080	2	2048	30	1.44%	98.46%
TU pipe element point Z	1037	0	1005	33	3.18%	96.91%

To make the comparison clearer, in the below chart [3.9] it can be noticed that less than 4% contains 3D information in terms of the Z coordinate, which corresponds to the depth of the pipes. That means that a direct 3D modeling of the underground utility networks is not possible and thus ancillary data should be used in order to assign the third dimension information to the existing objects.

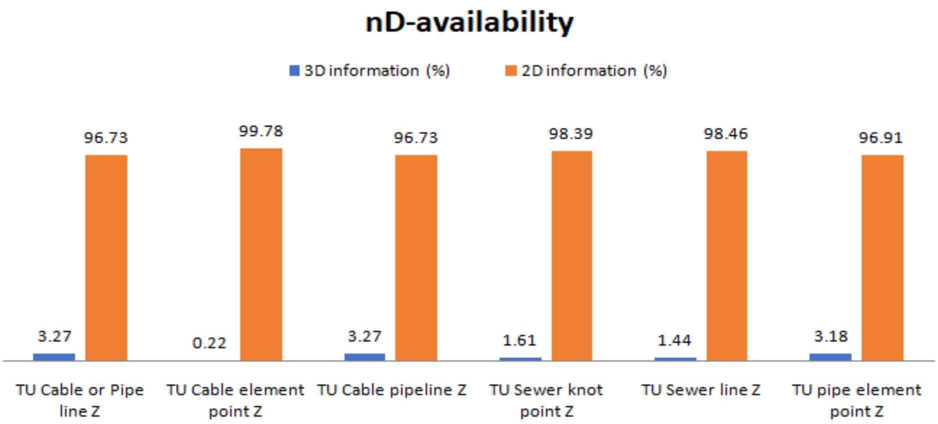


Figure 3.9: Availability of nD-information

From Table 3.2, however, it can be observed that datasets, provided by the TU Delft, cannot be characterized as 3D. They are mainly 2D with some exceptions. Still, given the lack of information about the units of measurement, the reference surface as well as their concurrent validity, the information provided as Z coordinate cannot be considered completely reliable. Continuing the research regarding the available height information, it should be mentioned that apart from the Z-coordinate corresponding to the depth specification, on the datasets some semantic height information was apparent. This information relates to a reference surface (which, however, is not known), as well as to nodes concerning starting and ending points (Table 3.3).

Table 3.3: Semantic Height information

Dataset/ height informaion	Total number of records	Relative level	Height (Z-coordinate)	Height bottom	Height top	Height begin	Height end
Unknown information							
TU Sewer knot point Z	1986	553	1953	1780	803	-	-
TU Sewer line Z	2080	-	2048	-	-	1431	1499
Known information							
TU Sewer knot point Z	1986	1433	34	206	1183	-	-
TU Sewer line Z	2080	2080	32	-	-	649	581

Table 3.3 presents the amount of available information about height. However, it cannot be used without concerns since a description of the reference surface or reference point is absent. Moreover, in the same data set, this information cannot be combined to reach a conclusion about the exact third dimension as at first glance they do not show any correlation or it does not make sense. An example of the limited 3D information is depicted in Figure [3.10], where the differentiation of networks in terms of depth layers can be seen. However, considering that the majority of the rest pipes are located at the same level, this single part may be 'left-over' or is placed, indeed, at the correct underground location. In any case, while missing the knowledge about the reference surface (after personal discussions with the data provider) and the exact length of the pipes (that could be a means of distinction between actual pipes and left-overs), assumptions were made.

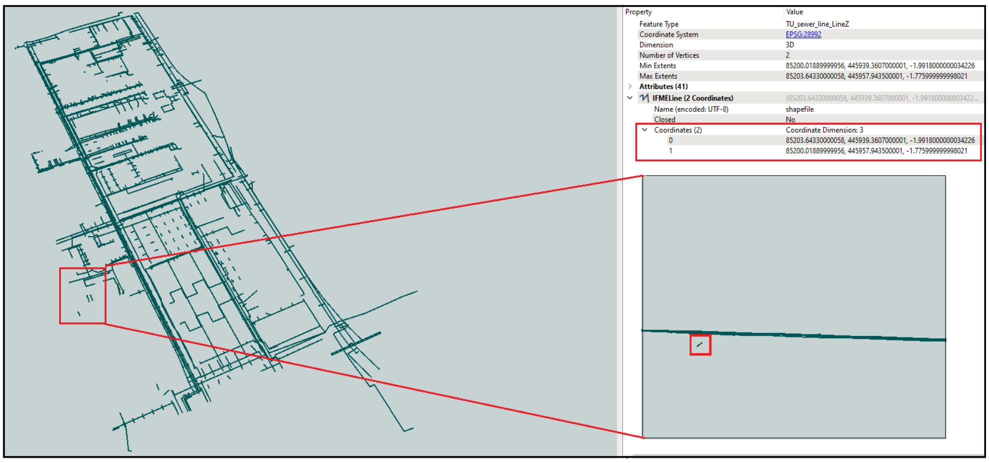


Figure 3.10: Visualization of the existing 3D information

3.4.3 Topology

As far as the topology of the existing networks is concerned, we refer to the way in which both the networks are arranged to relate to each other and to the relevant points of interest (nodes) and thus, how this relation can be depicted in terms of physical and logical connections. More analytically, topology defines a set of data integrity rules that, for example, do not allow gaps between lines and points of the same use (connectivity), it allows for the creation of relationship queries and navigation as well as supports feature construction derived from no-structure geometry [FME, 2020]. The available networks can be characterized by a set of lines (polylines) that are interconnected either continuously or distinctly. However, these lines are not accompanied by information about their starting and ending point -topological nodes-, while information about the crossing points (if existing), with pipes of different use, are also absent. Here was made a distinction between the nodes and their use. In more detail, when talking about *topological nodes* we are referring to those points that are located at the beginning and end of the network edges (did not exist in the provided data), and have been extracted using the available transformers from the FME software. On the other hand, we will refer to *survey/measured points* as those points present in the provided datasets and concern the nodal network elements that are connected with the existing networks based on their use (see Chapter 3.5)- appurtenances.

Analyzing further, the underground networks in terms of the connectivity with each other (networks of the same use and/or not) as well as with the respective survey nodes, a series of inconsistencies were observed. An example is given in Figure [3.11], where it can be noticed that the lines represent the *Plot connection pipes* are not connected with the survey nodes crossing the main *Sewer pipe*.

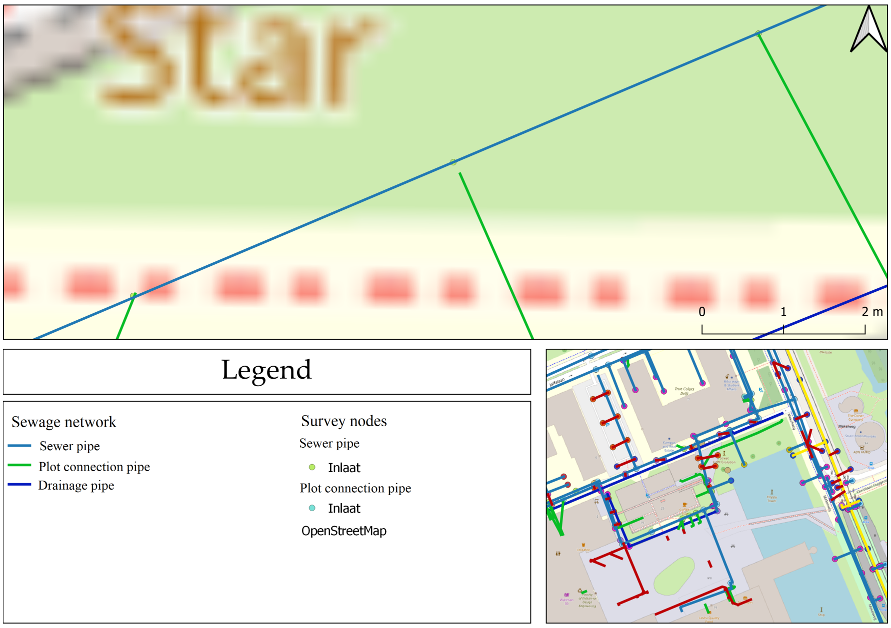


Figure 3.11: Example of topological inconsistency- overshoots/undershoots

In this case (Figure [3.11]), they have presented two types of topological errors that violate the (connectivity) relationships between the datasets. Specifically, starting from the left line and moving to the right one, there are three different cases. The first case concerns a line that ends beyond the point it should connect to (overshoot); in the middle case, the line does not meet perfectly at the node located in the sewer pipe (undershoot -including a gap); and finally, the last case (right line) is presented the connection as it should be. These errors arose from the digitization process followed during the creation of the original data and since they have not been corrected the error is transmitted from user to user making it difficult to understand and analyze them immediately. Another concrete example of an error is illustrated in Figure [3.12], where it can be seen the discontinuity of the lines of the same use (electricity- purple lines), as well as in Figure [3.13], where it can be observed the off-set of the pipes in accordance to the survey nodes that they should connect with.

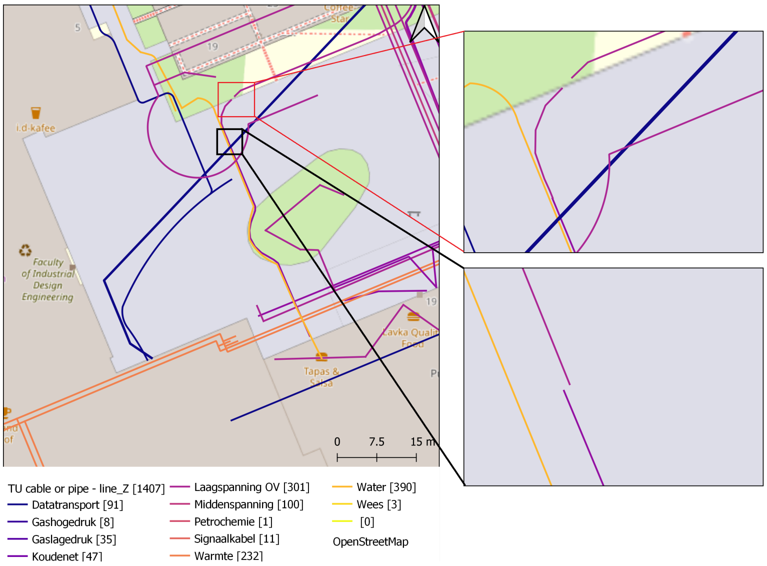


Figure 3.12: Example of topological inconsistency- discontinuity

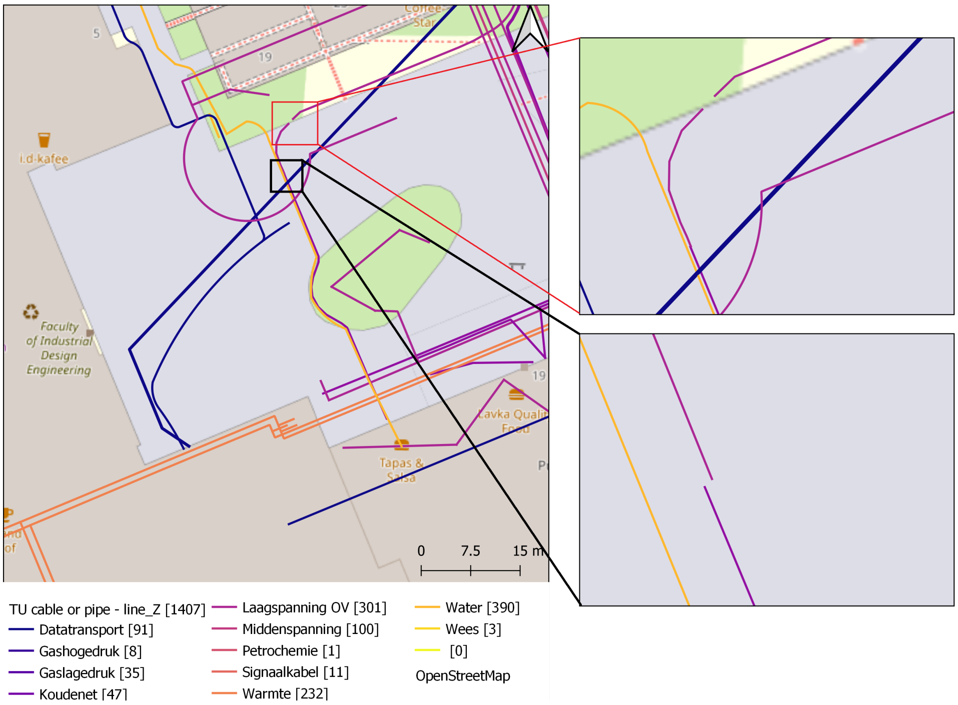


Figure 3.13: Example of topological inconsistency- dis-junction

During the thorough analysis of the data content by network category, it was discovered that there was a significant amount of discrepancy. For this reason, for further analysis, cleaning, and use for the present application, it was chosen to use one of the available categories of underground utility networks and in particular the sewerage network, including its sub-networks (see Chapter 4).

3.5 DATA PROCESSING AND CLEANING

3.5.1 Attributes management

Once the existing inconsistencies in the content of the data were identified, the approaches followed for their processing and cleaning were based on the different categories of the stored information. These categories relate to the attributes of the datasets and the corresponding geometry and topology. Starting from the first family of problems, it is noted that no modifications were made to the existing attributes, on the contrary, some additions were made to separate the included information into subcategories based on the types of networks. The additional attributes created are related to unique indicators, which characterize the different sub-categories of the main sewage network as well as elements that belong to them (edges, network elements -nodes-). The ultimate goal of creating unique identifiers (id/s) per network category is to develop structures capable to be supported in a database environment. In our case, where the datasets are interdependent (networks- network elements), a relational database structure is to be developed for storing, managing, and accessing data information. The construction of well-organized datasets is critical for further spatial analysis, since the access to their content as well as the creation of links between them become more efficient and effective, while with the creation of interdependencies (nested id/s), all the information about connectivity, encapsulation, and continuity is recorded [Lloyd, 2010].

The steps followed to uniquely identify the networks are shown in the below flowchart (Figure 3.14):

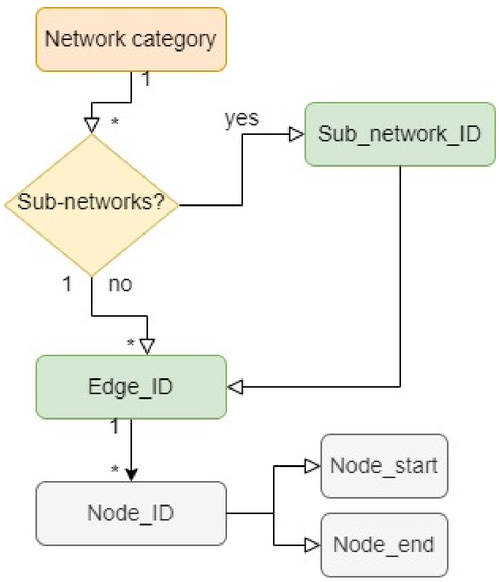


Figure 3.14: Identification of unique ids per network element

The flowchart [3.14] consists of a graphical representation of the newly added fields on the datasets that depicts the relationships among the datasets' different entities [Wang et al., 2019] and it can be considered as an entity-relationship diagram. This is the starting point for the design of a database and a more efficient data handling on them. It can be noticed that there appear different cardinalities between the fields that are:

1. A **one-to-one** relationship: In this case, one network is associated with a unique ID, that is not shared with other networks.

2. A **one-to-many** relationship: In this case, since there are many sub-networks corresponding to one network, the relationship is one-to-many. The same goes for the relationship between the sub-networks and their edges. Here many edges consist part of one sub-network, something that does not work the other way round [Biscobing, 2019].

With the addition of these fields (see Chapter 4.4), it is achieved the creation of an integrated table that includes the information from all the different, in terms of use, entities. In this way, there is no need to create extra tables per network since all the information is stored in the same layer. In addition to the differentiation of the data based on their use, through the introduction of the mentioned fields, the basis for the enrichment of the data with characteristics that are deemed necessary for the observance of the existing standards and their specifications, is built. Likewise spatial information, their coding serves to transform them into a more formal format (standard format) capable of being used by different users, helping them to perceive existing relationships more directly. However, this is a starting point for improving the content of the data, as other changes and additions are also necessary (see Chapter 6), for the existing information to be useful and immediately usable.

3.5.2 Repair geometry

Moving forward to the second family of inconsistencies concerning the geometry of the provided datasets, both datasets (networks, network elements) had to be modified. As it has been mentioned above, the representation of the available geographic features has been made using points and lines. Regarding the former one, they are used in order to represent network elements, for example, fire hydrants that should be connected at a specific point, on the earth's surface, with the corresponding fire pipe networks. The position of this element is determined by its coordinates in space and its connection to the corresponding pipe is achieved by the respective user who created the data. However, the connection of the attributes included in the one dataset as well as the connections between the different ones (lines-points) was not captured precisely and thus the inconsistencies mentioned above arose. To tackle that problem, judgments about the better representation of the available geographical features information were made, based on what is inherently right in the case of utility networks. It should be mentioned that all the modifications that took place concern, mainly, the re-location of the existing features in order to cross/touch each other at the assumptive correct locations, while no simplification techniques were implemented, keeping the original geometry of the datasets immutable. Additionally, considering the nature of the errors, it was not possible to find a generalized solution/algorithm capable of remedying all of them. For this reason, manual modifications were implemented at a certain extent, since the overall improvement of the datasets is beyond the scope of the current thesis. Nevertheless, a significant degree of change has taken place in order to produce a coherent network that includes logical connections between its components.

All the inconsistencies mentioned above (Chapter 3.4) became distinct, also, in the case of the selected network (including sub-networks) with a higher frequency of occurrence of the wrong representation of the existing topological relationships. The objects that should be connected to one another are displayed separately or many overshoots and undershoots are present and thus the information about their connectivity is missing. Thus the geographical information has not been encapsulated precisely. With the help of QGIS and FME software, the datasets were visualized and cross-referenced to detect existing inconsistencies. To made the self-made changes clear, an extra field - *Comments* - was added to the attribute table of original datasets, noting each time the type of change was made (creating a new object or re-locate it). Some examples of the changes that were implemented are given in the Figures below (Figure 3.15, 3.16) and involve modifications to both the network data and the data contained in the survey points. The changes were not made abruptly, but after consultation and internal discussions with the data provider of the underground utility networks. In particular, several meetings were held to clarify the correctness of certain changes. However, some changes were made on the basis of assumptions and following some existing patterns in the pipeline connections, as the provider was not fully aware of the as-built condition of the datasets. In order to identify the various inconsistencies to the selected for further analysis network, the main (sewage) network was divided into its subcategories, from which an additional separation was made, where were maintained only networks that contained, as already mentioned, the most information (quantitative). Each network was studied separately (Figure [4.4, 4.5] - Rioolstreng network), modified to some extent, and finally, the networks were combined with each other, based on their physical connectivity (geographical information), and changes were implemented again where necessary.

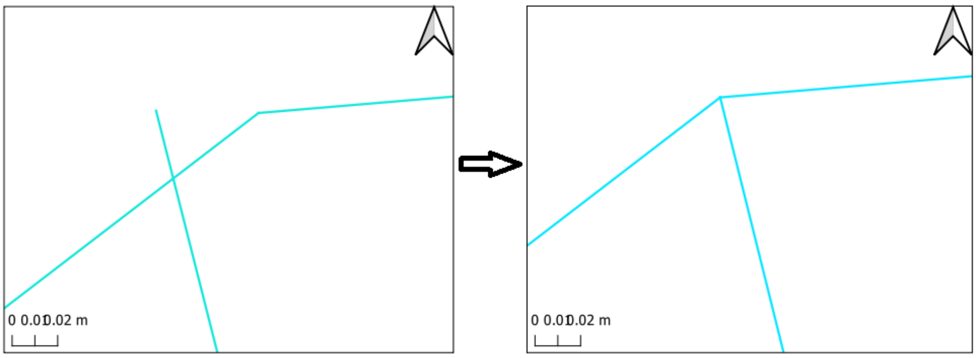


Figure 3.15: Example of fixing digitization errors- network

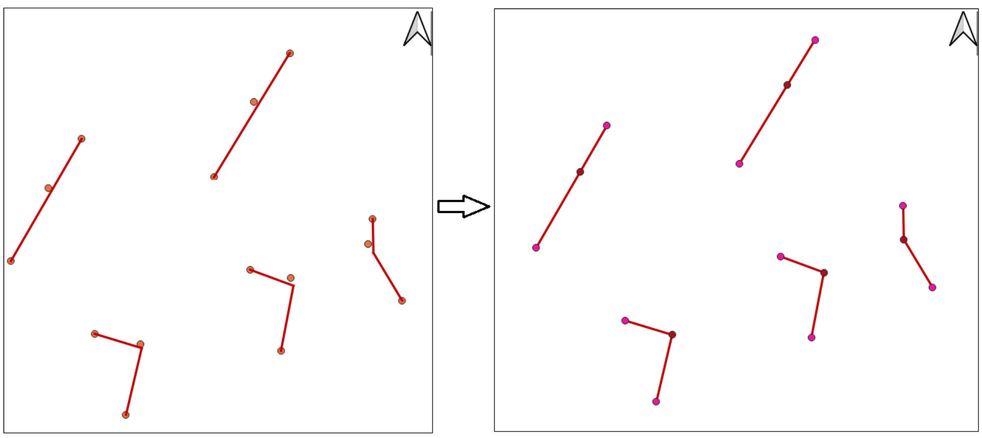


Figure 3.16: Example of fixing digitization errors- network elements

In the above Figure, it can be observed the digitization mistake existed in the original datasets (left image), while in the right one it can be noticed the modification made in order to achieve the connectivity of the network and the correct geographical representation.

3.5.3 Topology reconstruction

The reconstruction of topology includes, substantially, all the modifications related to the attributes, geometry, and finally topology (in terms of connectivity) in the data sets. As in the previous families of inconsistencies, also here the data were modified to match the needs of a topologically coherent network. The procedure followed for this purpose is illustrated in the following flow chart (3.17). In the first stage, the two datasets were processed separately and eventually integrated with the development of internal relationships. The goal is to create a single network that will be enriched with new attributes that will allow for a more efficient use for further applications. To do so, the information about both the connections between the entities and nearby relationships must be preserved in the existing topology. Taking into account the different nature of the available datasets, distinct approaches were followed for their topology optimization. Specifically, for the point dataset, the coordinates per point are sufficient for their representation as well as their spatial association with the other objects (e.g. neighboring lines/pipes). Things get more challenging when it comes to the representation of line topology and its relationship to these points (survey points). The nodes that lie between the available arcs must be recorded in a compatible format (Figure 3.18) for line topology representation that will allow for understanding the continuity in the network as well as the direction of its edges/arcs.

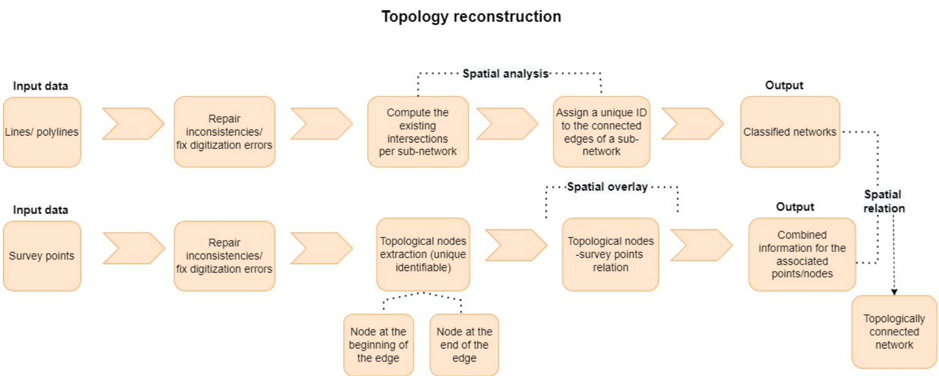
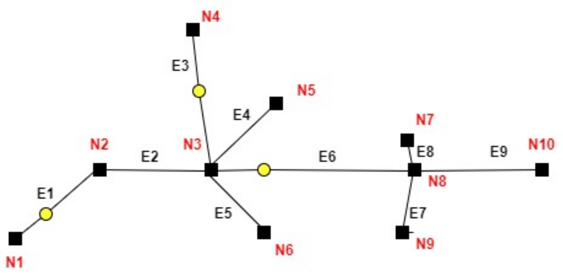


Figure 3.17: Topology reconstruction flowchart

In Figure [3.18], below, it is given an example of how the connectivity between the available arcs of a network should be presented in an attribute table. All the edges are characterized by a unique identifier (*primary key*) and they are connected with at least two nodes that illustrate the direction. The latter information can be extracted by reading the *Node From* and *Node To* fields that indicates the start and end points per edge. It can be observed, however, that there are some extra points (vertices- yellow circles) in the middle of some edges that do not correspond to the line representation. Taking into account the second set of data, containing points, these may/ or are not be located at different places than the nodes that characterized the edges themselves. This addition was made in order to make clear that the topological relationships existing between the same dataset may differ from the ones that should be reconstructed to meet the needs per application.



Edge ID	Node From	Node To	Length (m)
1	1	2	15.3
2	2	3	14.5
3	3	4	17
4	3	5	12.3
5	3	6	11.9
6	3	8	19.4
7	8	7	8.5
8	8	9	5.5
9	8	10	17.6

Figure 3.18: Line topology representation

The reason for which the optimization and cleaning of the data were sought, is the possibility to be transferred from the geographical representation to a topological graph. In the former case, the data are represented as separate physical entities, as they are presented in reality (regardless of designing errors), while in the latter case it is obtained a coherent network in which the intermediate connections (e.g. with sub-networks and/or survey points) are achieved with nodes, which represent the physical objects [Rawat, 2021]. In a topologically cohesive network, its extremities for every individual edge that make it up will be linked to a single node, as well as to additional edges that are also connected towards that unique node (if not a dead end) [Obe et al., 2017]. Once this connectivity is achieved then the full transmission from the geographical representation to a topological graph is possible to be made. In the case of the utility network data used in the current thesis, both the connectivity (e.g. disconnected pipes) was absent as well as the nodes related to their edges. As a result, the process of understanding the relationships and interdependencies (if any) as well as their differentiation in terms of their use becomes more difficult. A noded network consists the basis for the construction of a meaningful topology since it allows for routing services that are critical in applications such as disaster/emergency management as it makes it possible to predict, for example, the lowest risk path for a given pair of source and destination. This actually indicates that where two or more utility network subcategories connect, there must be a node at the intersection, and all network segments must be broken at the intersection, assuming that navigation (e.g. flow direction) from any of these segments to any other segment via that intersection is possible.

However, at this point it should be mentioned that although the creation of a noded graph is of vital importance, in the case of underground utility network its creation becomes more complex. There are different circumstances where the creation of a fully noded network is not the optimal course action. A concrete example is the existing sewer network, which include pipes that overpass and/or underpass and intersection and thus the network ended up with redundant information for those crossings. Here it becomes more apparent the value of having the 3D information available. In the before-mentioned case, the pipes are treated as if they are flat (in the 2nd-dimension). Therefore, regardless their use, they are presented to intersect, which points to a completely wrong representation of reality. To tackle this problem, 2D data should be enriched with the corresponding 3D information to differentiate the levels that the pipes are located.

The procedure for extracting these significant nodes is described below (see Chapter 4). Once this information was also stored in the data attribute table, it was possible to move forward to the creation of the 3D model, which includes both the topological nodes associated with the points dataset.

3.5.4 Digital Terrain Model & 3D BAG data process

To make the transition of the underground utility networks from 2D to 3D, a Digital Terrain Model of the study area was used. This model, in raster format, was obtained from PDOK (https://www.pdok.nl/introductie/-/article/actueel-hoogtebestand-nederland-ahn3-) and includes information about the height values of the wider area of the TUD Campus (Figure [3.19]). Specifically, two tiles of 0.5m resolution were downloaded (M_37N2 and M_37N1) and then they clipped based on the bounding box of the utility network datasets (Figure [4.13]).

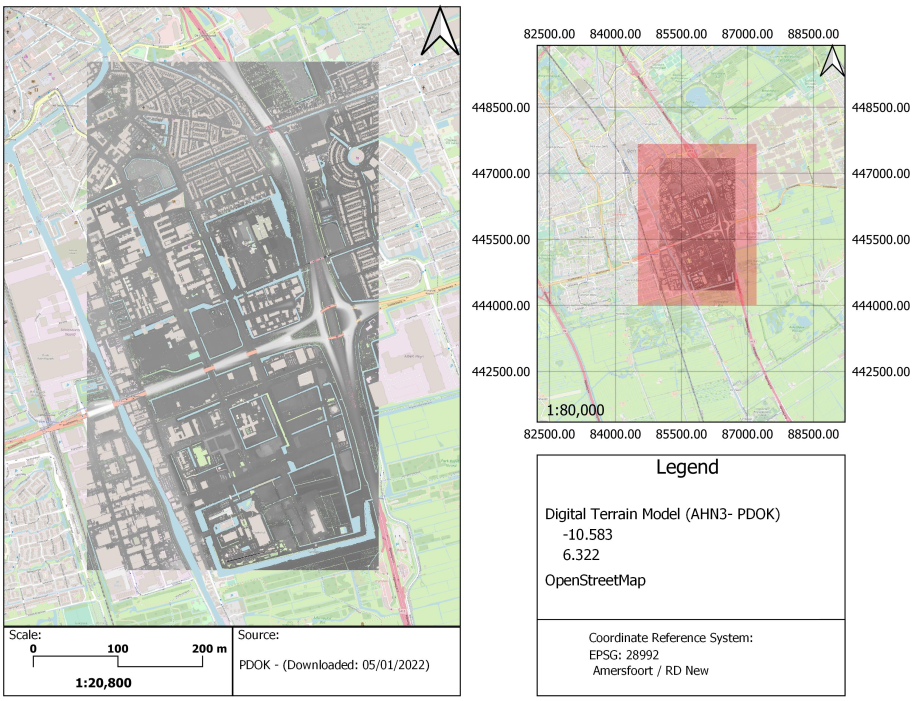


Figure 3.19: Digital terrain model of the study area

Once the topology of the existing underground utility networks is restored, the next phase of the thesis on their connection to above-ground objects (eg buildings, city furniture - if any) has progressed. Information on the latest reported objects comes from the 3D BAG platform (https://3dbag.nl/en/download), which provides open data information on existing buildings in the Netherlands, at different levels of detail (LoD 1.2 to 2.2.). In our case, the specific tiles covering the area of interest were selected, given in GeoPackage format (Figure [3.20]). In order to combine this information with the underground utility networks, only the footprints of the buildings were preserved, as they are the connecting link with the underground installations (Figure [3.21]).

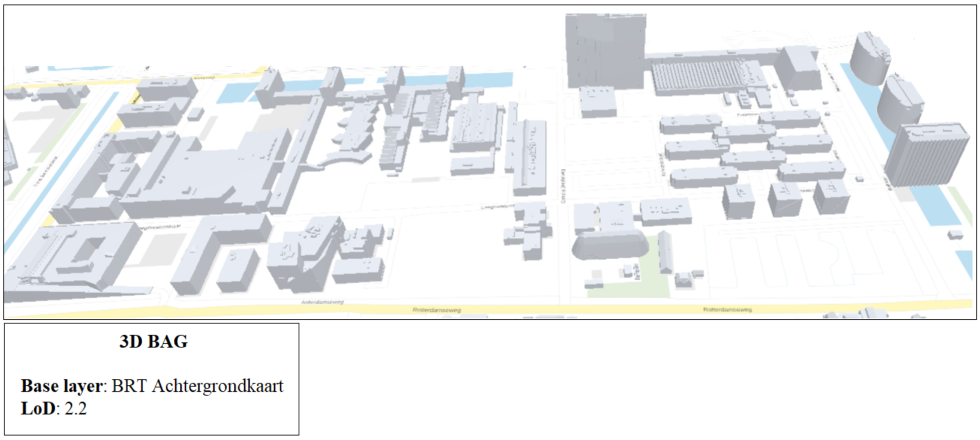


Figure 3.20: Part of the 3D BAG - assigned to the study area

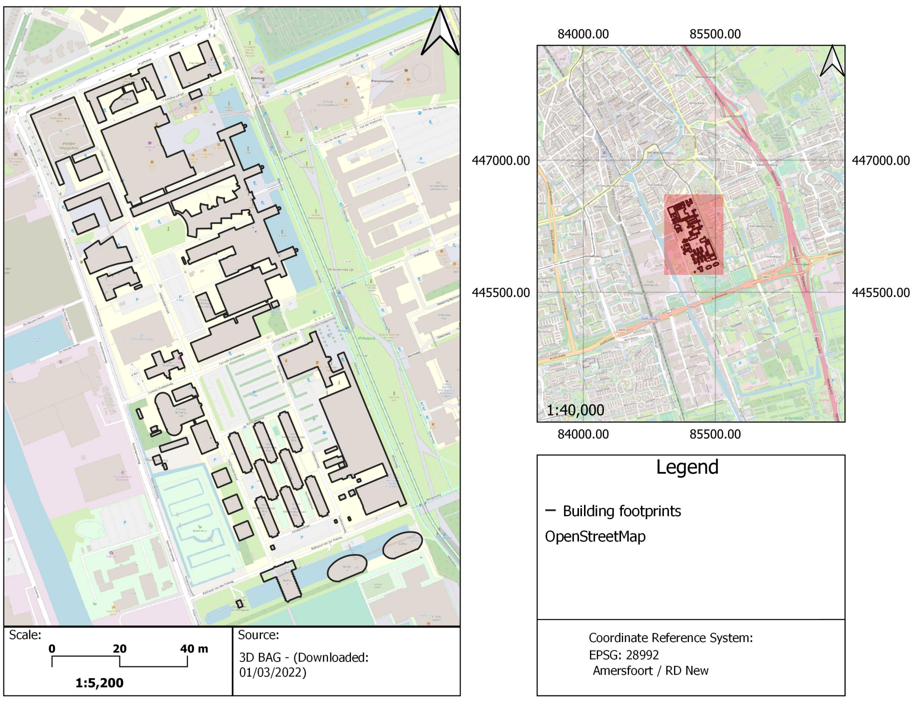


Figure 3.21: Footprints of the buildings of interest

Source: 3D BAG

To create a complete model and to achieve the connection of the underground utility networks with the above-ground buildings that they serve, the geographical information from the latter had to be transformed into a topological representation. In the former case, the building is represented by a point corresponding to its center of gravity (centroid). It should be noted that for the connection of the networks with the respective building they served, the topological nodes were used instead of the survey points. This distinction was made due to the incomplete and/or inaccurate description of the survey data. When processing the data, the topological nodes were correlated with the survey points based on proximity techniques (spatial overlay application), and this implies the transmission of errors, as there are cases where correlated points actually serve different purposes or the correlation failed while they should be identified.

4 CASE STUDY- TECHNICAL IMPLEMENTATION

4.1 CASE STUDY– DELFT UNIVERSITY OF TECHNICAL

To develop a use case application, it was necessary to select a research area (case study), which was determined by the provided data, concerning the underground utility networks. This area is part of the Delft University of Technical. Delft, is a municipality of the South Holland province in the Netherlands (Figure [4.1]). In the study area is located a variety of underground utility networks including pipes, cable, and assets. A practical application for which the existence of such an integrated 3D model is the practice of effective prevention and response to emergencies/disasters and/or the proposed planning for the expansion of existing networks to provide services to new subscribers. As a disaster is considered to be any significant disruption to the functioning of a community or society that has far-reaching human, material, economic, or environmental consequences that go beyond the capacity of the affected community or society to cope using its own resources [UNDRR, 2016]. Specifically, given the direct connection of underground utility networks to the objects of the ground surface, damage to the underground condition directly affects the regular operation of the objects connected to them, causing damage not only functional but economical as well. Therefore, the knowledge of the existing network as well as its connectivity both with the terrestrial situation and with neighboring networks is of particular importance. Having as a starting point an integrated model that represents the physical geographical information allows for both the prediction of possible damage as well as an immediate action for its restoration. In particular, in case of pipe failure, it will be examined which pipes are directly affected, suspending their operation and consequently the service of the building they are connected simple topological routing acting as an example for more advanced and sophisticated development.

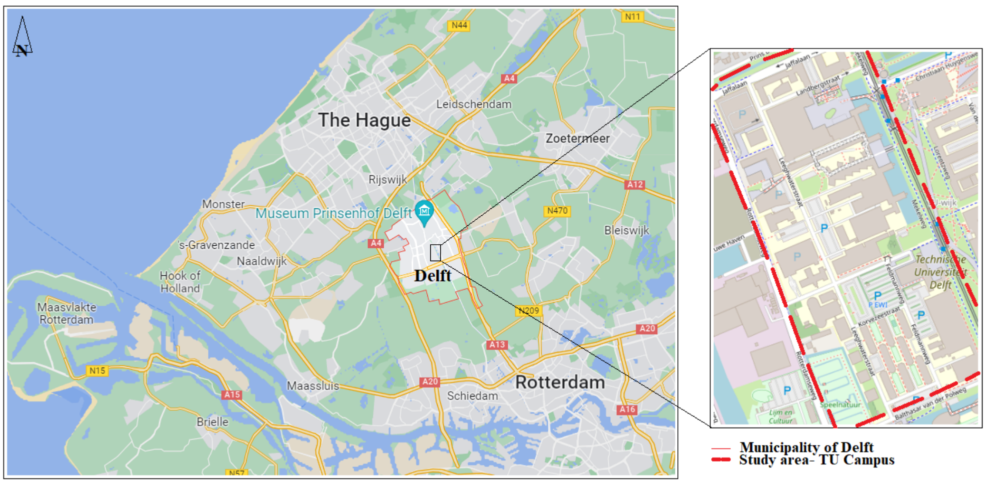


Figure 4.1: Delft, the Netherlands

Source: Google maps

Taking into consideration the amount of the input data in combination with their existing inconsistencies as well as their quality/quantity, it was decided to use only one category for further analysis and processing. Specifically, the selected network concerns the underground sewage network, as well as the sub-networks that are included, accompanied by the respective nodal network elements, such as appurtenances. Below, are given the two datasets used in the present thesis that concern the sewage network itself as well as the corresponding network elements that are related to sewage use, (Figure [4.2, 4.3] respectively). Later on, an additional separation was made between all of the sub-networks, in order to examine those who had the most information both in terms of features, quantity, and quality.

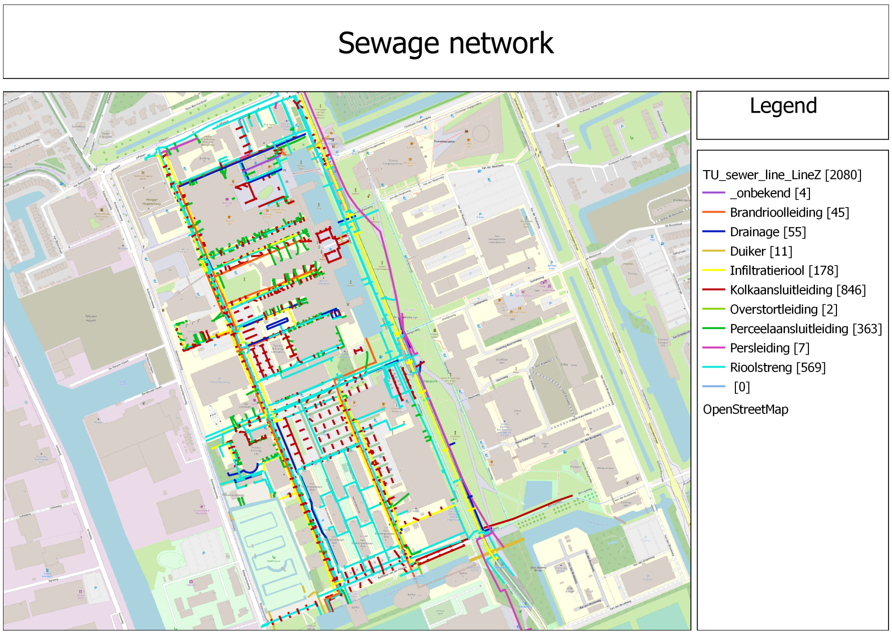


Figure 4.2: Sewage networks

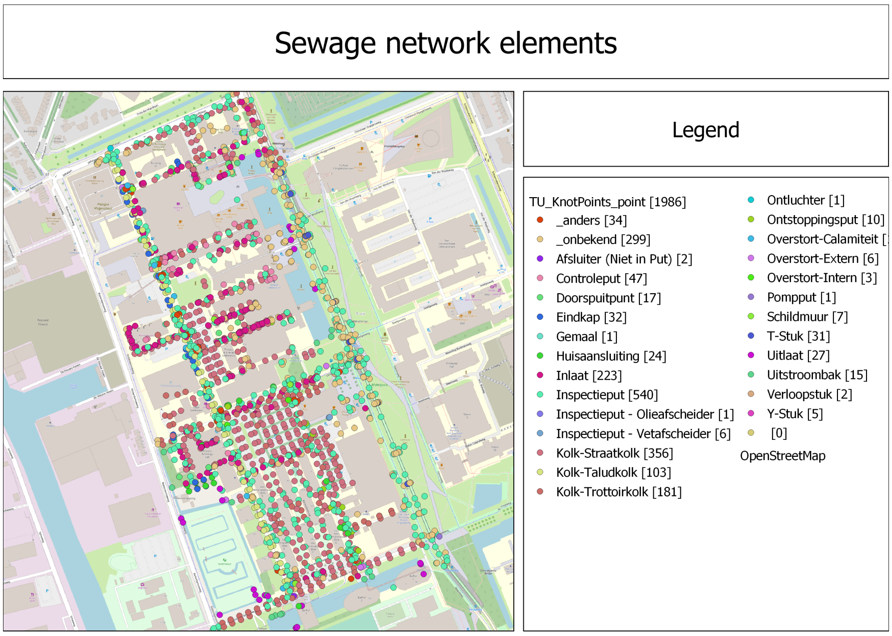


Figure 4.3: Sewage network elements

4.2 SOFTWARE AND TOOLS

Various software and tools were used to initiate the evaluation and further utilization of the data provided. Specifically, Python was used for the primary analysis (statistical analysis) of the different data sets, and QGIS and FME software were utilized to get familiarize with their visualization. The former software was used extensively during the data cleaning process, while the latter was used to further process the data in order to correct the defects - geometry, topology, and attributes. Furthermore, the PgAdmin platform, an open-source administration, and development tool for PostgreSQL used, with the extension of PostGIS and pgRouting as well. The PostGIS extension supports the storage and processing of spatial data in the object-oriented PostgreSQL database, enabling geolocation queries in a structured query language (SQL). On the other hand, pgRouting provides location - based routing as well as other network analysis capabilities, such as finding the shortest route (e.g. Dijkstra), determine dynamically the "cost" by using SQL (the quantity can come from numerous fields or columns) and flow calculations. During the elaboration of the various stages of the methodology that was developed, a combination of the aforementioned tools was made, each time intermediate elaborations and changes were deemed necessary.

4.3 DATA FAMILIARISATION

As mentioned in the previous section, given the amount of the existing inconsistencies and lack of metadata, a subset of the available utility networks were selected for further processing and analysis. In the process of familiarization with the data and their content, the data provider (*Directie Campus & Real Estate Department* staff- Jan van Voorst Vader) made a significant contribution. A series of personal meetings were held to explain the use of the networks and some of their features, contributing to a broader understanding of their content, as the information was provided in Dutch, making it difficult to understand their features immediately and accurately.

From all the different categories of sewerage networks (10), it was chosen for further analysis to use six (6) of them, the ones that contained the most information, and for which their initial topology, in terms of connectivity, was less deficient.

The selected (6) networks concern:

1. **Rioolstreng = Main sewer network**: This is the main sewer pipe existing between two sewer wells. Its purpose is to transport wastewater and/or sewage from buildings for treatment or disposal. The length of a sewer line can vary from one meter to many tens of meters and is only limited by the need for cleaning and inspection. The maximum length, therefore, depends on the length that can be achieved with modern means for cleaning and inspection. Usually, a maximum length of 60 to 90 meters is applied. The diameter of the pipe is in many cases the same over the entire length of the string and usually changes in sewer wells. The design is always based on a constant diameter between sewer wells (Figure 4.4).

2. **Drainage**: Drainage pipes can be considered as a system used in order for water or other liquids to be drained from a place. It ensures, actually, the removal of excess water from the root zone of the soil [ELLINGSON, 2019], and thus the corresponding liquid can be used for different purposes (Figure 4.5).

3. **Brandrioolleiding = Fire sewer pipe**: This sub-network consists of permanently filled sewer lines, connected to open water, for supplying fire extinguishing water (Figure 4.6).

4. **Infiltratieriool = Sewer infiltration**: In an Infiltration and Transport sewer or IT sewer, the rain infiltrates underground through a geotextile (fabrics used in geotechnical applications) wrapped, perforated, horizontal pipe in the ground. To absorb heavy showers, the IT sewer can lead to an overflow facility on surface water or the rainwater sewer (Figure 4.7).

5. **Kolkaansluitleiding = Gully connecting pipe**: The gully connection pipes, which are present in large numbers in the network, direct rainwater running from paved surfaces (such as roads and car parks) from the ravine to the waste or rainwater drain (Figure 4.8).

6. **Perceelaansluitleiding = Plot connection pipe**: The main use of these pipes in the TU Delft campus is to connect the existing buildings with the main wastewater sewer (Rioolstreng) (Figure 4.9).

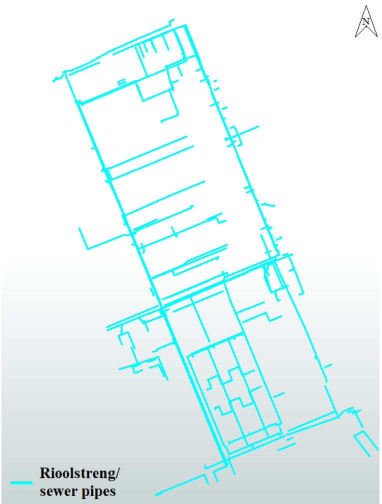


Figure 4.4: Rioolstreng/ sewer network

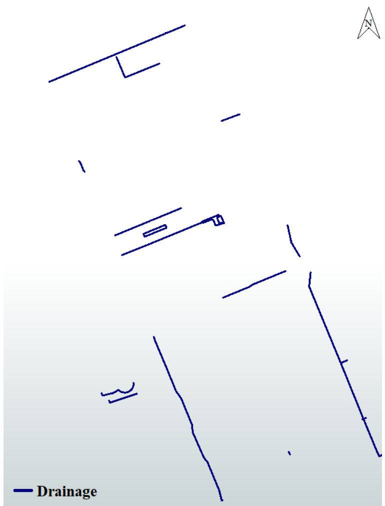


Figure 4.5: Drainage network

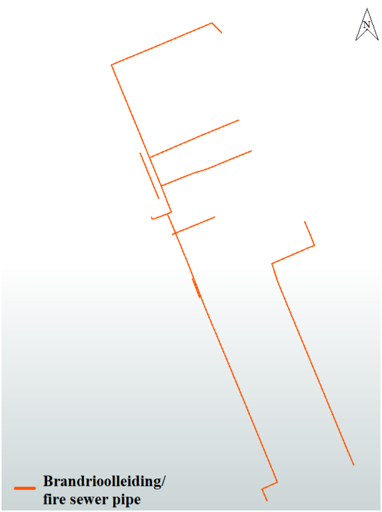


Figure 4.6: Brandrioolleiding/ Fire sewer pipe

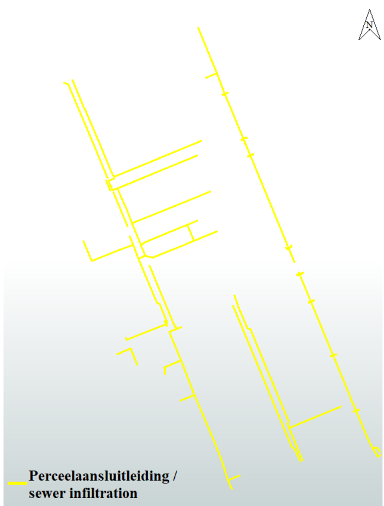


Figure 4.7: Infiltration sewer pipe

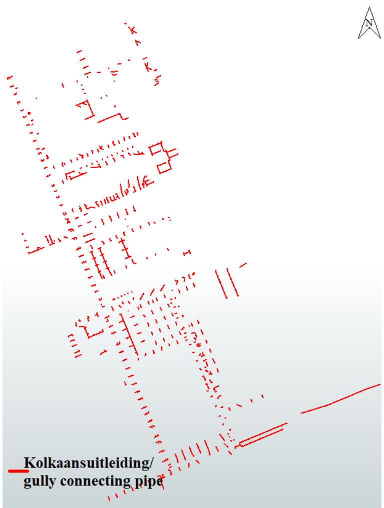


Figure 4.8: Kolkaansluitleiding/ Gully connecting pipe

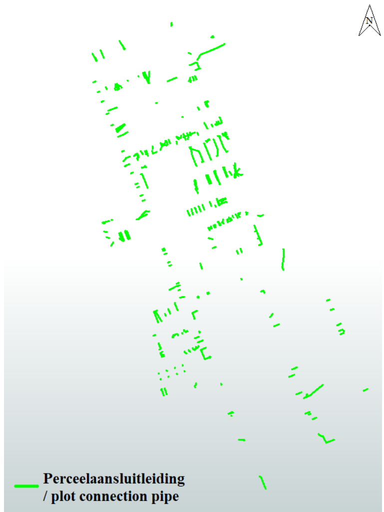


Figure 4.9: Perceelaansluitleiding/ Plot connection pipe

In addition to the above mentioned networks, for the creation of the 3D model as well as their integration with the above ground objects, a Digital Terrain Model (raster) and a 3D city model of the study area used accordingly (see Chapter 3.2).

4.4 3D MODEL CREATION

4.4.1 Topology restoration

The creation of a coherent model of underground utility networks requires, in addition to the existence of their third dimension corresponding to the depth of the available pipes/cables, a correct topological model that will reflect reality (geographic information) and the existing connections between the various networks. A coherent topology, as mentioned above, demands logical connections both in terms of geometry and attributes, which will allow for the development of internal connections between data sets (e.g. by means of primary-foreign key relations in linked tables representing different datasets).

Based on the geometric and topological inconsistencies found in the input datasets, some manual edits had to be made in the first stage. These changes mainly concern the re-positioning of existing pipes in places connected to other objects of the network, whether they are pipes or survey points (see 3.5.1), as well as the addition (by manual digitisation/drawing) of edges that contribute to the creation of a cohesive network that supports the relation of pipes of the same use. It is noted that some additional edges were created in the cases where pipes of the same network were disconnected but they were located close to each other, thus their connectivity was reasonable to be done. The necessity of a fully connected network serves the development of navigation algorithms that help the direct identification of the immediate affected networks and sub-networks, in case of failure and/or the development of new network components (find connection points with the existing pipes). An example of the re-location of wrongly located pipes is given in Figure [3.15] and [3.16], while an example of the creation of the supported edges is given in Figure [4.10].

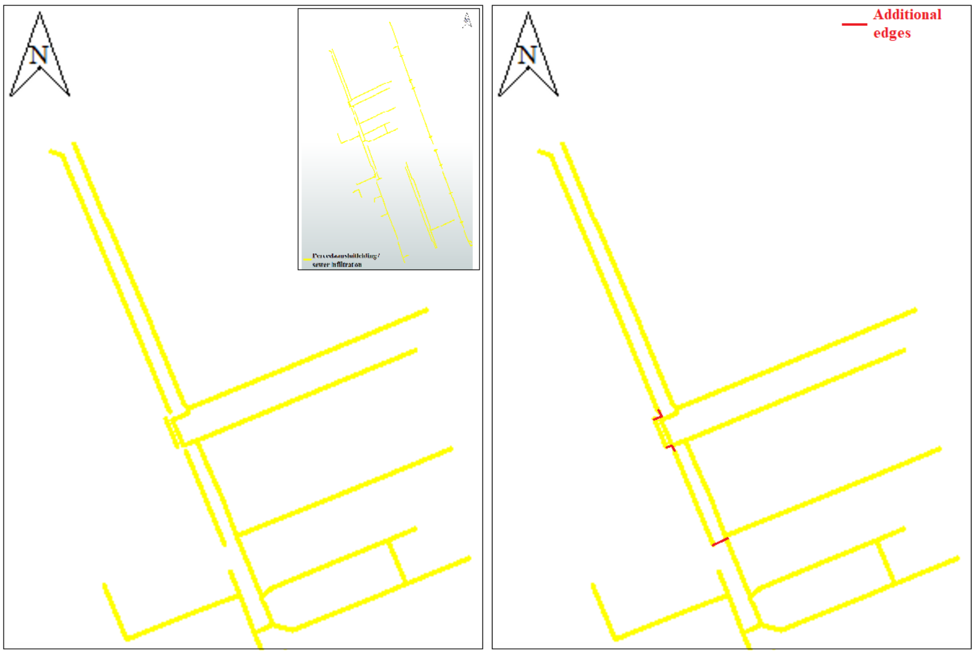


Figure 4.10: Example of manually digitized edges- infiltration network

Taking into account that in the initial attribute table these edges would be a new entry, an additional column - *Comments*- was created to easily distinguish them from the initial dataset. In addition to separating the newly added edges from the existing ones, the networks themselves should be separated based on their uses (Figures [4.4, 4.5, 4.7, 4.6, 4.8, 4.9]) and therefore based on the various sub-networks they contain (disconnected pipes of the same use). In order to achieve this hierarchical distinction, additional fields were created concerning first the different type of networks. To create these additional fields, a combination of the transformers available in FME software were used (e.g. *TopologyBuilder, NetworkTopologyCalculator*). The former fields are related to the network category, the edges/arcs belong to it accompanied by the significant nodes concerning their starting and ending point (node_start, node_end). This information was absent from the provided dataset and it was extracted during the data process phase (see Chapter 3.5.3). The unique identifier created per network, constitute the primary key (connection) key to which the intersecting edges and the corresponding junctions (nodes) belong to. The code list for these ids is given below and was created using a conditional query related to the available types of networks, an information stored in the attribute table (*Riool_Type* field):

- Rioolstreng = Main sewer network: **id=1**
- Infiltratieriool = Sewer infiltration: **id=2**
- Brandrioolleiding = Fire sewer pipe: **id=3**
- Perceelaansluitleiding = Plot connection pipe: **id=6**

Then, as each network consists of independent sub-networks, each of them was assigned again an ID, which is related to the main ID of the network. From the attribute table shown above [Table 4.1], the Network_ID fields correspond to the unique identifier of each network (see Chapter 4.3). It is noted that for the reconstruction of the network topology it was necessary to separate all the available edges in order to add a unique identifier so that there is continuity in the internal consistency of the network in terms of their attributes. This identifier corresponds to column Edge_ID. Based on the enumeration given above, all the edges belonging to a specific network share the same id that is further inherited to the related significant nodes (Table 4.1). The relationship between the interdependence and the connection of the indicators is more perceptible in the flowchart (Figure [3.14]).

Table 4.1: Example of id=2 determination

	Network_id	Edge_id	Subnetwork_id	Node_start	Node_end	
. . .	2	2	130	157	55	. . .
	2	1	174	73	56	
	2	2	167	173	57	

It should be mentioned at this point that the ids correspond to the node per edge were added once the topological nodes were extracted. In the data used, the topological nodes were not apparent, and thus these specific nodes were entered per edge using the FME transformer *TopologyBuilder*. The last transformer received all input features to compute topologically significant nodes and edges, which then output with supplementary attributes that define the topological relationships. It should be noted that with this process the topology has not yet been repaired but this is the first step for its reconstruction. The separation between the provided point data and the extracted topological nodes is exemplified in Figures [4.11, 4.12], where in the first case [4.11]) it can be observed that all lines that make up the sewage network have a node in their start and endpoint, while in the second case [4.12] are present all the available survey points, some of which are related to this network. After *TopologyBuilder* transformer that inserted the information about the topological nodes and their unique identifiers, *NetworkTopologyCalculator* transformer along with attributes transformers, was used to locate all the sub- edges in the network graph that belong to that specific network. By combining the outcomes, the obtained result is a topological connected network having the basic attributes needed to characterize its components (Table 4.1).

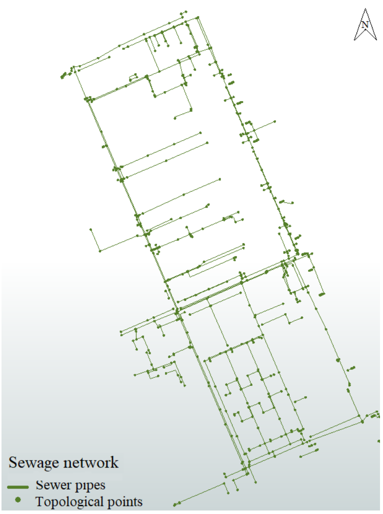


Figure 4.11: Topological nodes

Figure 4.12: Survey points

Once the topological nodes were extracted, the next step, as shown in the flowchart (3.17), was to relate them to the respective survey/measurement points. This association was aimed as these points are connected (or should be connected) to the underground utility networks, serving a purpose. However, given the quality of the data, this connection was not immediately discernible in all cases and therefore spatial overlays approaches had to be implemented. From the set of the survey points, in each network corresponds a certain number of them, while there are cases where one survey point belongs to more than one network (e.g. points where pipes of different uses are connected). Examining the data, it was found that most of the survey points are located close and/or at the exact same place to a topological node and for this reason, their spatial relation was searched for. To achieve this, a spatial overlay was carried out starting from a topological node and defining a 0.2 m buffer zone surrounding it. The buffer zone was determined after a trial and error procedure and further discussions with the data provider, and it was concluded to keep these survey points located quite close to the edges belonging to a certain network category (and/or sub-network), as logical- physical- connection points.

The main aim was to keep the radius short to avoid error propagation of incorrect selections, as large radius (>0.5 m) implies the inclusion of survey points that may not be related to a network, while a very small radius (<0.2 m) implies the possible omission of points that are actually connected to a network. The survey points associated with the topological nodes were stored in a common attribute table to share their attributes. In terms of attributes, topological nodes contain information that corresponds to unique integers that characterize each point. Specifically, three new fields, as mentioned above, have been created that add information about the nodes' connection, from where the line starts and where it ends, as well as the specific edge they belong to. Regarding the corresponding attributes of the survey points, the unique identifier of the respective topological node was assigned. In this way, the connection of the points between them is achieved and the topological information that will be used later for the integration of the networks (edges) with the respective nodes/points is stored. An example of the association between the topological nodes and the survey points is given in Table 4.2. Once the selection of the nearest neighbor was implemented, a "virtual" edge between the associated points was created using the *GeometryReplacer* transformer that replaces the geometry of the stored features according to the line geometry encoding parameter's configuration (LineString creation). These virtual edges were created for conceptual reasons and for a better visual understanding of the between them connectivity. It should be noted that these edges do not exist in reality, while they were not given in the initial datasets. This visualization technique was also later on for the integration of the underground utility networks with the above-ground objects (see section 4.4.3). However, these virtual edges were not stored as extensions to the existing edges per network, considering also their negligible length. Despite this, in case, a larger radius was chosen, then the lines would be of vital importance since they would contribute to the extension of the connectivity of the network.

Table 4.2: Attributes of associated topological nodes-survey points

Node_id	Network_id	Sub_network	OBJECTID
1	2	1	2563
3	2	1	3347
4	2	1	4272
5	2	2	3808
6	2	2	4492
. . . . .			

In the table above, the last column (*OBJECTID*) corresponds to the index of survey points located within the search radius used. In case a survey point is not related to a topological node then the corresponding record is empty. In addition, in case a survey point is associated with more and topological nodes, then the corresponding *OBJECTID* is shared between the different topological nodes, which have their own node_id. Given the small search radius used, it was ensured that there was exactly one node_id for each topological node per edge, per network that corresponds to exactly one survey point, avoiding duplication (overlapping points). The above-mentioned procedure was followed, in the first phase, for each network separately to avoid errors, and in the end, they were merged into a single table. The final product consists in two tables that store the two geometries (edges, nodes), respectively. These tables can, now, successfully be linked for further spatial analysis (e.g. routing functions) due to the creation of the identifiers per element and the between them interdependencies.

4.4.2 3D integration

The reconstruction of the topology is followed by the creation of the three-dimensional model of the underground networks. As the provided data was lacking 3D information, the DTM of the study area was used as starting point. This DTM is a 0.5 meter grid that covers the study area, and is designed to be a ground-level file, with all "ground level" points re-sampled to the grid using the Squared IDW technique. The interpolation did not include points classified in a distinct category (off-ground features such as trees, buildings, bridges, water, and other things). The used digital height map (AHN3- Actueel Hoogtebestand Nederland (AHN)) has the characteristics given in the Table 4.3 (table source: https://www.opendem.info/opendemeu_meta_netherlands.html) and has released in 2019.

Table 4.3: AHN3 details

Data URL	https://app.pdok.nl/ahn3-downloadpage/
Metadata URL	https://www.ahn.nl/
Horizontal Reference System	EPSG: 28992
Vertical Reference System	EPSG: 5709
Resolution [m]	0.5
Distributor	Actueel Hoogtebestand Nederland (AHN)
License	CC0 1.0
EVRF 2000 Offset [cm]	-1

From the chosen networks (six categories 4.3) a further selection was made due to the high amount of missing information and the existing inconsistencies in terms of topology. In the end, only four networks were maintained, specifically those that after the processing and cleaning procedure, had their topology been reconstructed to a greater extent. The selected networks are:

- **Rioolstreng network**
- **Infiltration network**
- **Plot connection pipes**
- **Fire sewer pipes**

In the existing data, given their disorganised way of designing and the absence of 3D information, several assumptions had to be made about the connections between the networks in order to finally create a network divided into its subcategories accompanied by the attributes that characterise it. The intersections implemented in the utility networks are based on their hierarchy and the logical connectivity that results from their content (use) and from their visualization by combining the existing information with the available buildings in the area. It is noted that since there was no accurate information and verification of the hypotheses that were made (even after the personal discussions with the data provider), this is an experimental approach. The networks that logically are connected concern the main sewage networks, the plot pipes as well as the infiltration pipes. The fire sewer pipes constitute a separate network that does not intersect with the others due to different content. Having this hierarchy as a reference, the networks were draped on the 3D surface of the DTM. To do so, the FME transformer *SurfaceDraper* was used, which accepts as input the 2D breaklines (network edges) and creates a Delaunay triangulation using the DTM. This creates a triangular surface (at the top of the digital terrain model) in which each line forming a triangle has known heights. For the utility networks that are essentially draped on this surface, the triangle to which they belong and/or intersect is identified and the elevation at this point is determined by applying mathematical algorithms. In this way, the received drape features that are layered on the surface model lead to the draped features, in 3D (Figure [4.15]).

Taking into account that the bare soil under the buildings has no height value, that gap was filled by interpolating the no-values raster cells using QGIS (Figure [4.14]). In this way, the values for the no-data zones are estimated using inverse distance weighting from the neighboring pixel values. Following the interpolation, the findings from the previous procedure are flattened out. This method is effective for interpolating missing portions of rasters that vary reasonably regularly [testing QGIS, 2022].

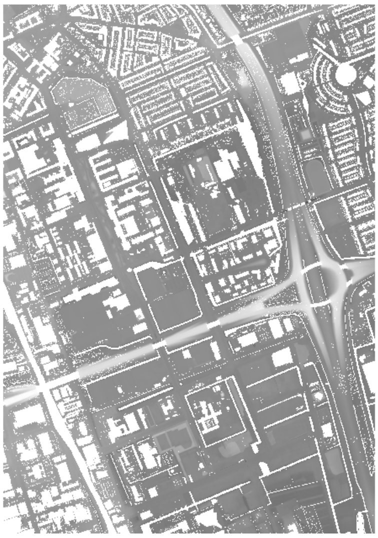


Figure 4.13: Initial DTM including null values

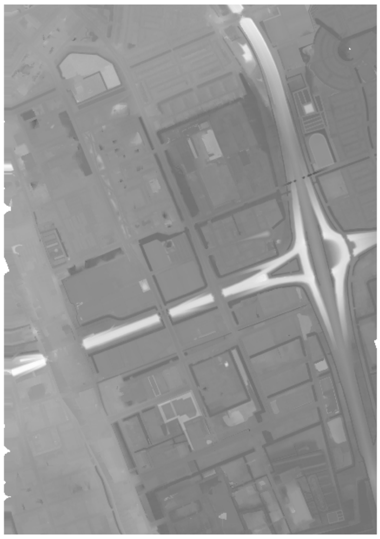


Figure 4.14: Filled DTM replacing values

Then, the underground utility networks were draped onto the 3D surface model (Figure [4.15]). However, since the networks are located below the ground, they had to be offset below that surface to better highlight their differentiation in terms of their position (based on their use) as well as in order to make them more perceptible, visually. Considering the different type of networks and their use, different offsets (Figure [4.16]) were implemented. However, it should be noted that this is an arbitrary decision that neither reconstruct the 3D geometry of the available pipes nor their 3D modeling. On the contrary, it was only used to further facilitate data exploration and visualization, since the 3D information was absent from the used datasets. The selected offsets are:

- **Sewer pipes**: -2.5 m
- **Infiltration pipes**: -1.5 m

- **Fire pipes**: -2 m
- **Plot pipes**: -0.40 m

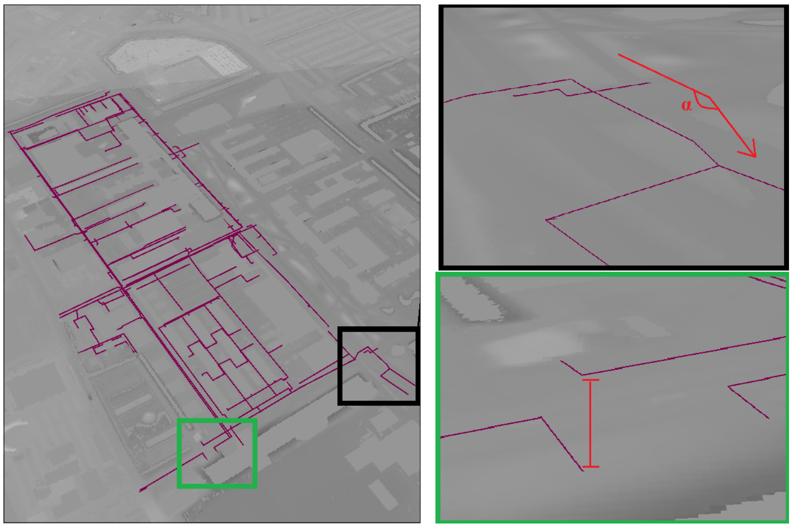


Figure 4.15: Draped 2D features to the 3D surface model

In the above Figure [4.15] it can be observed in the highlighted areas (black and green squares) how the 2D features were draped on the 3D surface information. In the first case (top right) α symbolizes the slope formed by the pipes following the natural slope of the ground, while in the second case (bottom right) the position of two sub-networks on different height levels is presented.

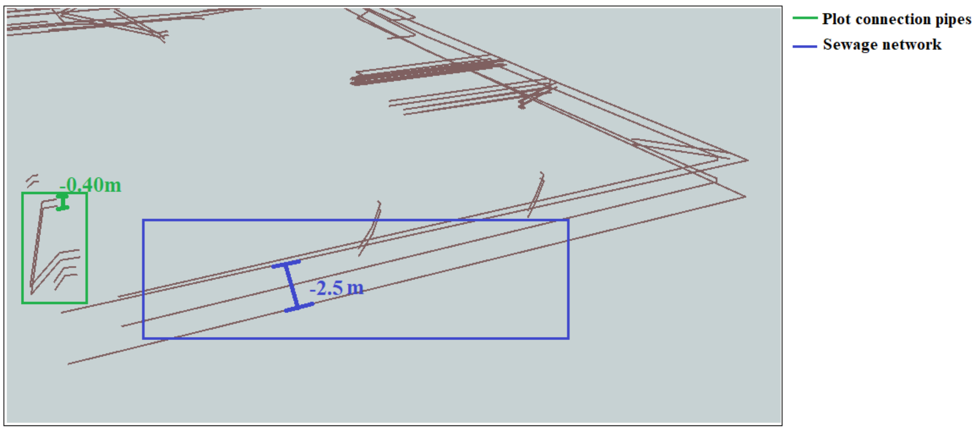


Figure 4.16: Offset of draper features

Finally, the tables created during the topology restoration process were updated with the 3D information. Note that for the point table (storage of both topological nodes and search points), the three-dimensional information per node/point is stored as a new field per coordinate (e.g. column X, Y, Z). On the other hand, this was not immediately possible with regard to the geometry of the lines, as for the full rendering of the 3rd dimension the corresponding information would have to be rendered for each point on the line. This whole information was stored using lists containing the coordinates of all the points that constitute a line, saved as a new field in the attribute table. However, this extra information was stored temporally for visualization purposes, since it does not correspond to the real depths of the networks.

4.4.3 Integration with the 3D city model- buildings

The next step in was the integration of the model of the underground utility network with the above-ground objects. In particular, the three-dimensional city model of the wider study area (from https://3dbag.nl/en/viewer) was used, from which the buildings are included in the bounding box surrounding the available utility networks.

For the integration of the selected networks with the above-ground buildings, spatial overlay techniques were used. In more detail, the first step was to switch from the geographical information to the topological representation, in terms of the buildings. This was achieved by representing the buildings, and specifically their footprints (footprint geometry: polygon) as a point that later on would be connected with the nearest topological node, per network. This point was initially selected to be their centroid (gravity center calculated from the coordinates of the footprint), but given the shape of the buildings, it was possible for this point to be placed outside the footprint. For this reason, the *RealCentroid* QGIS plugin was used that generates an internal point of the polygons (footprints) (Figure 4.17).

Figure 4.17: Topological representation of buildings - centroid

Due to the unclear situation of the networks regarding the representation of reality, we had to ensure the intersection between the edges of the networks with the outline of the buildings, which represents their service from the respective pipes. Therefore, a buffer zone of 1 meter was chosen to be created surrounding the building footprints. In this way, it is ensured that the topological nodes that were quite close to the buildings but did not cross their boundary, will be considered that they serve them. This value was selected as the most appropriate, after a trial and error process, where it was observed that a smaller buffer zone did not serve our purpose (zero modifications) while with a larger one, many duplicates nodes seem to serve more than one building. The latter case was not desirable, as the one-to-one service of buildings was sought, given the existing topology and the available pipes. With a buffer zone greater than one meter the footprints of the buildings overlapped significantly and there were pipes (including the topological nodes) that seemed to serve more than one building, which is not correct, as one end pipe in a network serves exactly one building. Once the buffer zone was created, a spatial overlap was applied between the extended outlines and the topological nodes in order to maintain those closest to the center of the buildings, assuming that the edges they belong to, serve the respective building. In more detail, a spatial join was executed that seeks to connect the features of the point dataset to the properties of the nearest feature, which is the centroid of the building parts. The resulted product is a new vector dataset- point dataset- that combines the attributes of both the input feature as well as the nearest feature from the testing set.

Subsequently, in order to visualize this connection, virtual edges were created again, following the same procedure mentioned in previous section (see section 4.4.1). Also in this case, these edges do not exist in reality but they were considered necessary for the visual part of the thesis. Here, these edges are connecting the topological nodes serving a building with its centroid were attempted to be made in order to clean later on the redundant information. Taking into account that one edge corresponds to two nodes, then in case, a pipe is entire "inside" (underground) a building, two virtual edges will be created (one per node), serving the object, creating the redundant information. Once the virtual edges were created, the duplicate information was manually removed, maintaining only the nodes closest to the building centroid (Figure [4.18]).

Figure 4.18: Example of topological node simplification

Once the connections between the topological nodes and the centroid of the buildings were established, the point table was updated, this time adding information about the buildings they serve. It is noted that during the transition from the footprint (geometry: polygon) of the buildings to a point, all the features were transferred to the new geometry (point). The table of points, after all the process stages, includes all the topological nodes of the edges per network, the corresponding survey points related to them (in case a node is not associated with any survey point the corresponding record is empty), as well as the relevant information about the buildings they serve (a field containing the unique identifier per building).

After processing the different geometries stored in the datasets, the final integrated result combines all the attributes captured in the model. An example of the combined 3D model is shown in Figure [4.19], where are illustrated the virtual edges created to connect the topological nodes of the pipes per network, to the center of the buildings.

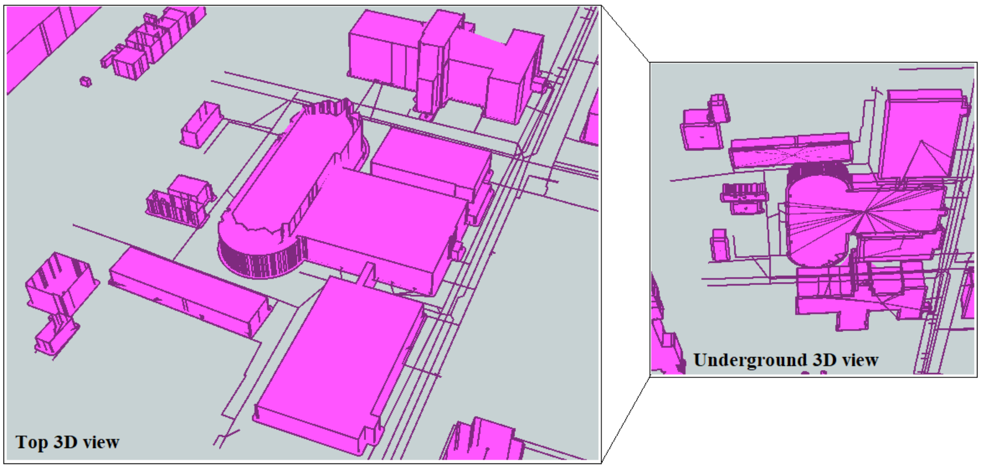


Figure 4.19: 3D representation of buildings and underground utility network

4.4.4 Integration with the 3D city model- building addresses

Using the centroid per building object it was achieved the creation of a direct connection of the underground utility networks with the above-ground object. All the pipes nearest to that specific centroid were considered that serve the buildings in their entirety. However, given the volume of the buildings and the fact that each building consists of different compartments and/or rooms identified by a single/unique address, a building can be considered an aggregation of these sections. For that reason, it was examined the connection of the underground utility networks with the available addresses per building object. In more detail, I used the addresses of the study area (available in PDOK - `https://www.pdok.nl/introductie/-/article/adressen-inspire-geharmoniseerd-`) (Figure 4.20). From this dataset, which includes information about the addresses in the municipality of Delft, only those addresses were extracted inside the bounding box of the study area. This information was combined with the available building footprints to have a complete and clear view of the addresses' location and match with the available buildings (Figure 4.21). As in the previous datasets, an attribute analysis was performed to understand the content of the data as well as a data cleaning process that this time concerns the exclusion of addresses located at a significant distance from the available network.

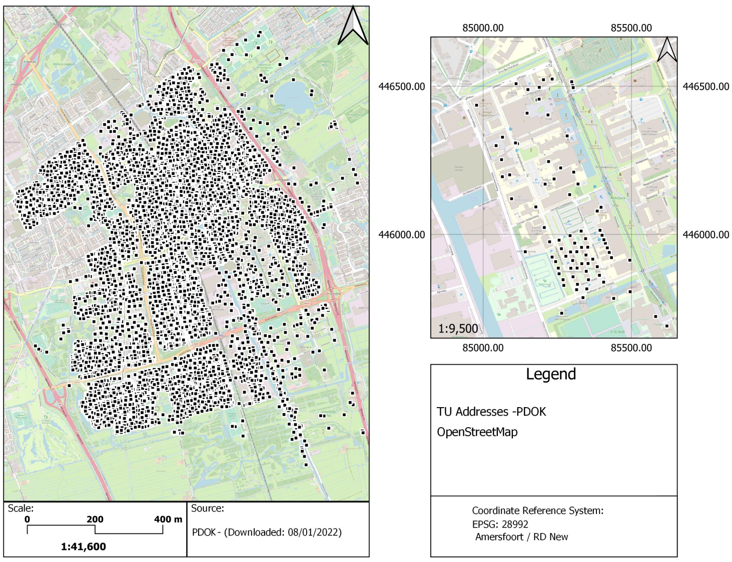


Figure 4.20: Addresses point dataset

Figure 4.21: Associated addresses with the buildings of the study area

Addresses point dataset analysis

As mentioned above, the dataset concerns the addresses available in the wider area of Delft. Specifically 59605 records were stored, in geojson format, from which only the 1471 located inside the study area were further utilized. During the extraction the new limited dataset were exported in shapefile format to keep the same structure with the rest of the datasets. To understand the content of the dataset the corresponding attribute table was traversed. Information was available about the: *Street Name, House Number, Postcode, Status of the address/building, Function of the buildings they identify, Year of construction, and unique identifiers per House Number, Function, Building*. An example of the attribute table is given in Figure [4.22].

	openbareru	isnumm	uislette	isnumm	postcode	oonplaa	emeenten	rovincie	verblijfso	opervlal	verblijf_1	peadre:	adresseerb	pandid	andstatu	ndbouv	nummeraand
1	Rotterdamseweg	139	B	11	2628AL	Delft	Delft	Zuid-H...	woonfu...	33	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...
2	Rotterdamseweg	139	B	46	2628AL	Delft	Delft	Zuid-H...	woonfu...	30	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...
3	Rotterdamseweg	139	B	51	2628AL	Delft	Delft	Zuid-H...	woonfu...	30	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...
4	Rotterdamseweg	139	B	53	2628AL	Delft	Delft	Zuid-H...	woonfu...	30	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...
5	Rotterdamseweg	139	B	14	2628AL	Delft	Delft	Zuid-H...	woonfu...	29	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...
6	Rotterdamseweg	139	B	47	2628AL	Delft	Delft	Zuid-H...	woonfu...	31	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...
7	Rotterdamseweg	139	B	17	2628AL	Delft	Delft	Zuid-H...	woonfu...	31	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...
8	Rotterdamseweg	139	B	28	2628AL	Delft	Delft	Zuid-H...	woonfu...	29	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...
9	Rotterdamseweg	139	B	5	2628AL	Delft	Delft	Zuid-H...	woonfu...	29	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...
10	Rotterdamseweg	139	B	44	2628AL	Delft	Delft	Zuid-H...	woonfu...	30	Verblijfsob...	VBO	NL.IMBAG.Verbl...	NL.IMBAG.Pand....	Pan...	2010	NL.IMBAG.Nummer...

Figure 4.22: Example of addresses point data attribute table

In the above Figure, the unique identifiers have been highlighted and used to distinguish the different addresses and building entity numbers related to them. By building entity(/ies) we refer to any type of building such as educational buildings (TUD buildings), church, residential buildings, stores or other uses available in the area of interest. The existence of these identifiers was of vital importance since they were used to associate and create interdependence relationships with the different datasets. These three columns, highlighted with a green, blue, and red border, provide information about the residence id (adresseerb), building id (pandid), and number designation id (nummeraand) per address respectively. The residence, as well as the number designation id, are unique per address, while the building id appeared many times (duplicate ids) since in one building more than one address can be present. While processing the data it was observed that one point (point geometry) representing an address did not correspond to only one record row in the attribute table. There were cases where at a certain place more than one point (individual points) was (were) located that correspond to more than one house number id (Figure [4.23]). In this case, the different house numbers should be distinguished since in case of a disaster all the buildings entities would be affected.

	huisnummer	huisletter	postcode	pandid	pandstatus	nevenadres	gid	openbareru	huisnumm_1	woonplaa
1	181	NULL	2628DJ	NL.IMBAG.Pand....	Pand in gebruik	NULL	8293152	Korvezeestraat	NULL	Delft
2	182	NULL	2628DJ	NL.IMBAG.Pand....	Pand in gebruik	NULL	7354978	Korvezeestraat	NULL	Delft
3	183	NULL	2628DJ	NL.IMBAG.Pand....	Pand in gebruik	NULL	4867736	Korvezeestraat	NULL	Delft
4	184	NULL	2628DJ	NL.IMBAG.Pand....	Pand in gebruik	NULL	6222750	Korvezeestraat	NULL	Delft
5	185	NULL	2628DJ	NL.IMBAG.Pand....	Pand in gebruik	NULL	4710192	Korvezeestraat	NULL	Delft
6	186	NULL	2628DJ	NL.IMBAG.Pand....	Pand in gebruik	NULL	1227784	Korvezeestraat	NULL	Delft
7	187	NULL	2628DJ	NL.IMBAG.Pand....	Pand in gebruik	NULL	6211336	Korvezeestraat	NULL	Delft

Figure 4.23: Example of overlapped points with different house numbers

Moving forward to the connection of the underground utility networks with the addresses, spatial overlay techniques were applied. This time, the address points were associated with the nearest topological node belonging to a certain pipe. The distance between the origin features (address point) and their closest destination one (topological node of a pipe) was calculated and the resulting datasets were enriched with additional fields indicating the identifier of the nearest destination feature (topological node unique id) and the distance to it. Having this internal connection between the addresses and the topological nodes dataset, it was made possible to connect the network (in terms of pipes) with the related address point. As in the previous case (building centroid), virtual edges were created (temporally) to visualize these connections and better understand the new network connectivity (Figure [4.24]).

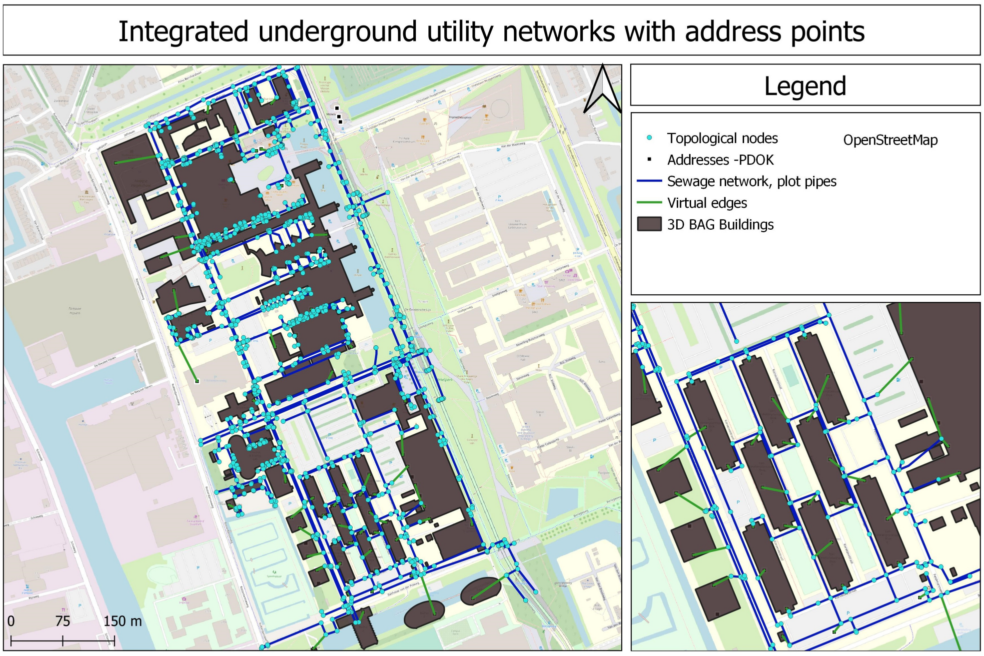


Figure 4.24: Example of virtual edges between address points and topological nodes

4.5 CASE STUDY SCENARIOS

Moving forward to the case study scenarios related to disaster/cost management, the relational database PostgreSQL was used via the PgAdmin client. Specifically, the derived from the above-mentioned processing procedure tables were stored in a self-defined relational database. Since the current condition of the data their format do not correspond to any of the available standard models (see Chapter 2.2), a customized data structure in the database environment was created, to examine if the reconstructed network is able to be used for routing analysis, using pgRouting extension and its available functions.

These tables are related to each other as the extra attributes have been added to them that allow for the development of the primary foreign keys relationship. This allows the immediate detection of the correlation between the various objects, and therefore it is now possible to look for adjacent relationships and connections. To do so, the pgRouting extension will be used, which adds geospatial routing capability to the PostGIS/PostgreSQL geospatial database.

Regarding the topological graph used for the implementation of the routing, it should be noted that its topology was adapted to the specifications of the pgRouting tool. An example of such an adaptation is to convert the geometries of the stored objects to a type compatible with the pgRouting algorithms. Specifically, the geometries stored in the edge table were MultiLineString, while the needed one for pgRouting was LineString. Additionally, a new enumeration in terms of unique identification of the stored features was necessary to match the connectivity of the automatically created pgRouting tables.

Routing through the edges of the networks, we refer to the possibility of moving from point A to point B, following a direction either from A to B or vice versa. The ability to move in both directions depends on the type of network. In the case of the current condition of the utility network graph, the absence of information about the gravity factor related to the flow of the product carried by the pipes, as well as the total ignorance of the ground slope, implies the creation of a two-way graph. Both information about gravity as well as the ground inclination are fundamental factors that in combination with the related pumping stations (vague information of their location- in our datasets) would allow for a more realistic hydraulic simulation. However, the current bi-directional graph corresponds to reality in cases of services such as internet, telecommunications, etc. where the transmission time does not affect the operation of both of them and the terrestrial objects they serve. However, in the case of sewage networks, this is not the case, as the flow direction of the liquid elements plays an important role in the smooth operation of the pipes and consequently of the objects they serve. Nevertheless, in the present work, in order to develop and test the routing algorithms, the graph was considered to be bi-directional.

Following the steps needed to reach the pgRouting schema needed for the implementation of its algorithms, a further selection between the networks should be made. These steps address a series of functions that modify the original table by adding an extra field (unique ids per record-row) as well as creating support tables that are internally linked to the original one (see Appendix 7). One of these functions (*pgr_nodenetwork*), however, created an incorrect relationship between objects. In more detail, the stored networks in the topological edges table were four, and as mentioned before they are not all connected with each other (see Chapter 4.4.2). By applying the function *pgr_nodeNetwork*, nodes were created at each intersection between the existing edges of all networks. This does not correspond to reality since the networks are either located on different levels or are not connected with each other given their type, in terms of their usage. Thus to further process the networks utilizing the routing algorithms, different combinations based on their physical connectivity were created. For example, plot connection pipes were combined with both the main sewage as well as the infiltration network, since their use is to connect these two networks with buildings. On the other hand, fire pipes constitute a single network that does not cross other pipes from the whole sewage network, and thus were handled separately.

After the execution of the pgRouting functions, the relations between the tables with respect to their primary and foreign keys are represented in Figure [4.25] for the centroid case, and in Figure [4.26] for the addresses point dataset one. Based on these internal relationships of the tables were created two scenarios are detailed discussed in the next Chapter (see Chapter 5). In these scenarios, attention is paid to the implementation of functional analysis on the corrected -in terms of topology-network that allows allows for the detection of the shortest path either in terms of the shortest route or the lowest cost in case a high cost needed to be invested to serve a building. To do so internal connections between the generated from pgRouting tables should be created as well as the initial ones derived from the reconstruction of topology (see Chapter 4.4). It should be mentioned here that in the case of the addresses point dataset, which was analyzed at a second phase during the thesis implementation, it was added a new table in the database including all the information of the addresses accompanied by the related with them topological nodes id. In this way it was made possible to join the different tables between them to extract the information needed per case study. However, as mentioned above, there were cases of overlapping points that correspond to different house numbers (points located at the same building entity). During the spatial overlay approach followed it was implemented one-to-one matching and so in order not to lose the rest of information (topological node ids were existing more than once based on the house numbers)the duplicates were first identified and then stored in a database view. In PostgreSQL, a view can be retrieved as a virtual table and it was used to simplify the complexity of the queries used during the routing analysis.

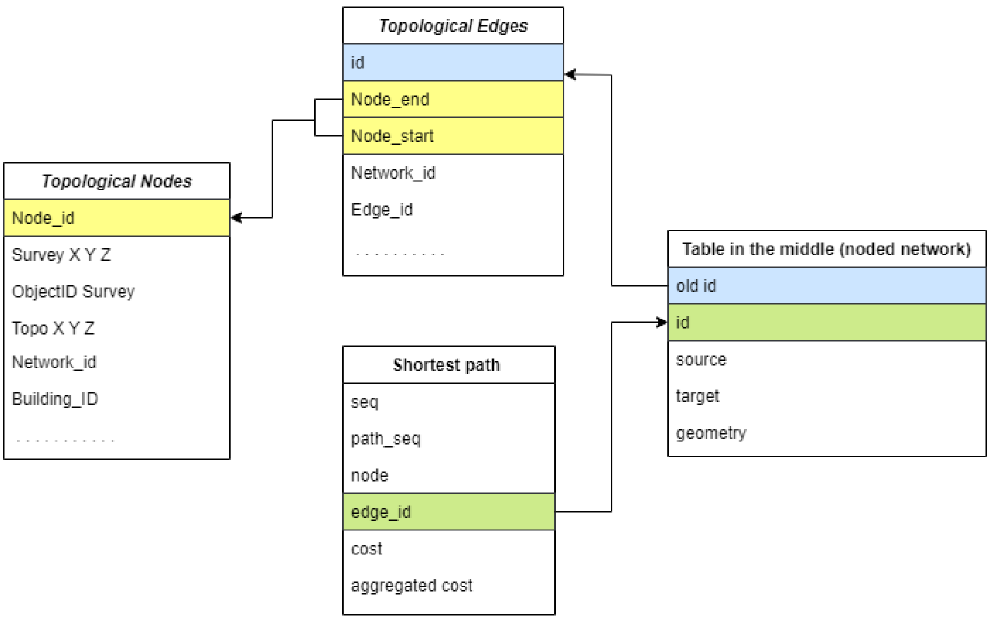


Figure 4.25: Table relations- using only the building centroid

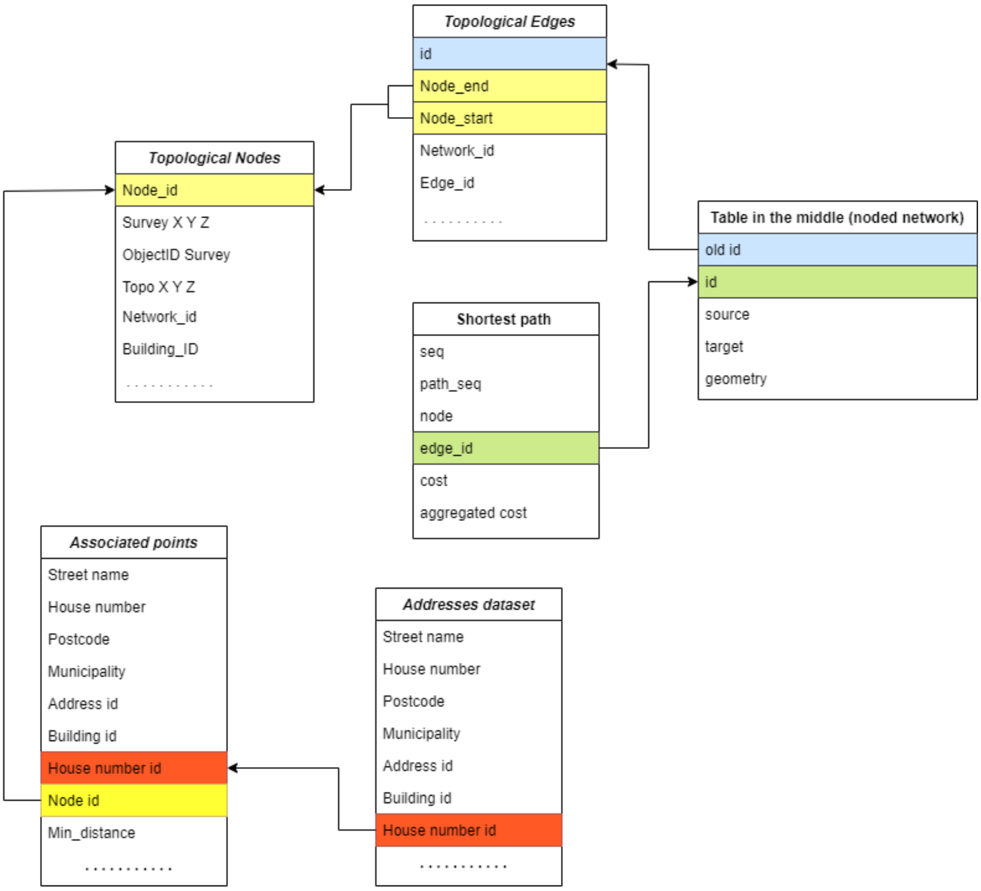


Figure 4.26: Table relations- adding the addresses information

Based on these relations, it is possible to combine the spatial information needed to be extracted for different cases/scenarios. Once the tables were joined using the related keys *pgr_routing* functions were executed to detect the optimum route in different cases. The outcomes of the tested cases are given in the next Chapter (5).

5 RESULTS & ANALYSIS

5.1 RESULTS

This chapter present the results derived from the process described in the previous chapter that answers the research questions and evaluates the usefulness of developing the proposed model in possible real-world applications. In addition, the results from the scenarios, as they arose from the utilization of the limited information of the data sets used, will be further analyzed, preparing the ground for the next chapter that concerns the overall limitations and conclusions about the current work.

Starting from the research questions of this thesis (see Chapter 1), it is noted that the main objective was to consider the possibility of creating a three-dimensional model of underground utility networks, capable of being integrated with above-ground objects in order to be used for real-world applications. Additionally, attention was paid to the evaluation of the effort required for the transition from reality (geographical information) to the corresponding digital twin (dual representation) taking into account the quality/quantity of available information. The use of a combination of tools contributed to the gradual analysis, processing, and management of the provided datasets in order to synthesize the final product.

5.1.1 Data quality & quantity

One of the most important points of the research was the familiarity with the data and their content. Due to the absence of metadata or corresponding information that would describe the data and their content, both quantitative and qualitative analysis had to be done in order to be able to assess their condition and the effort needed to reconstruct their inconsistencies in terms of topology (in the context of the thesis needs). In addition to this analysis (quantitative, qualitative), the personal meetings with the data provider (Jan van der Voorst) contributed importantly to a better understanding of the content of the data, who, among others, gave a more accurate interpretation of the networks and their uses, as the data was given in Dutch.

Traversing the attribute table per datasets it was made clear that they were of poor quality and quantity while important information needed for the implementation of the suggested application was also absent. Additionally, in terms of their compliance with any of the available standards for underground utility networks mapping, they could not directly either enrich or convert (e.g. to CityGML format), due to both their disorganized content as well as the lack of relation to the schema of a standard (see Chapter 2.2). The analysis carried out concerned both the completeness of the data regarding the available fields in the attribute tables and the existence (or absence) of the third-dimensional information. The results are shown in the graphs below (Figure [5.1,5.2]).

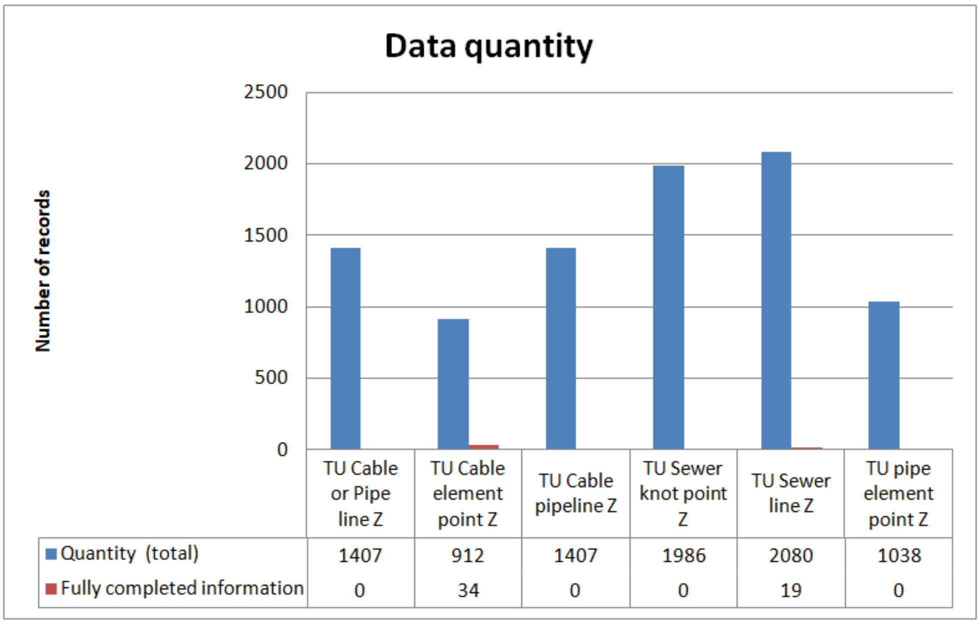


Figure 5.1: Data quantitative analysis results

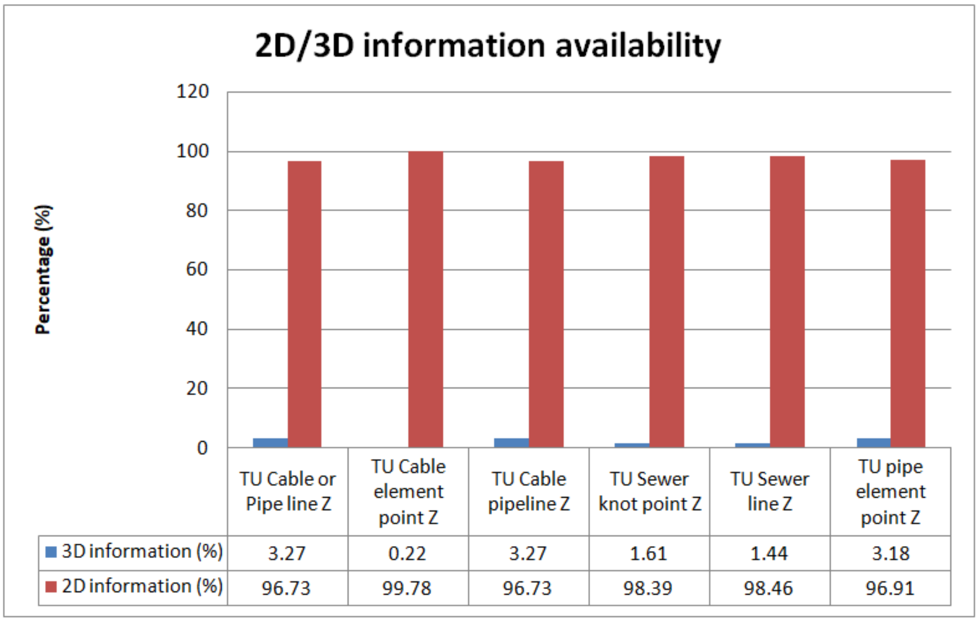


Figure 5.2: 3D information availability

It can be observed that in both cases the content of the information is insufficient and therefore several assumptions and simplifications were required to be able to utilize it. With regard to three-dimensional information, the necessity lies in the effort to integrate the underground utility networks with the ground objects, in order to identify those that are served by the respective pipes (or cables). However, in the case of line geometry, the information was limited to 2D and the 3D conversion was performed mainly for visual purposes, to facilitate the differentiation of the available network layers. Finally, based on the statistics and the corresponding content of the data sets, as mentioned, a limited number was used to apply the methodology (see Chapter 3) and to produce the desired result, which appears in the next sub-chapter.

5.1.2 3D integrated model of utility networks

To create the final model, the data topology was reconstructed, to obtain a better representation of reality. Then in order to detect and connect the networks with the above-ground objects they serve, they were integrated successfully with the use of the 3D building model of the study area. The coherently reconstructed network is shown in Figure [5.3] below, while the integrated set is illustrated in Figure [5.4].

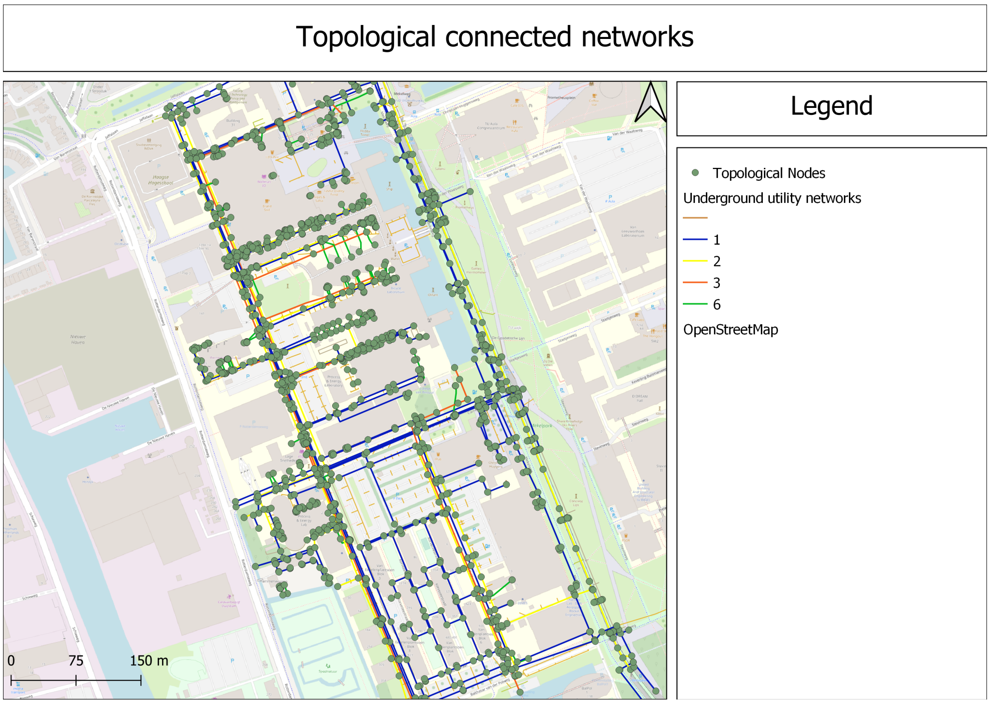


Figure 5.3: Topologically reconstructed network

The above map depicts the studied networks united with the respective (topological) nodes that make them up (start-end point). This information, as well as the relationships between them, have been stored in a relational database, and therefore if more information about the networks was available, for example, the flow of the products in the pipes, a more detailed simulation of their function could be performed by utilizing the existing graph.

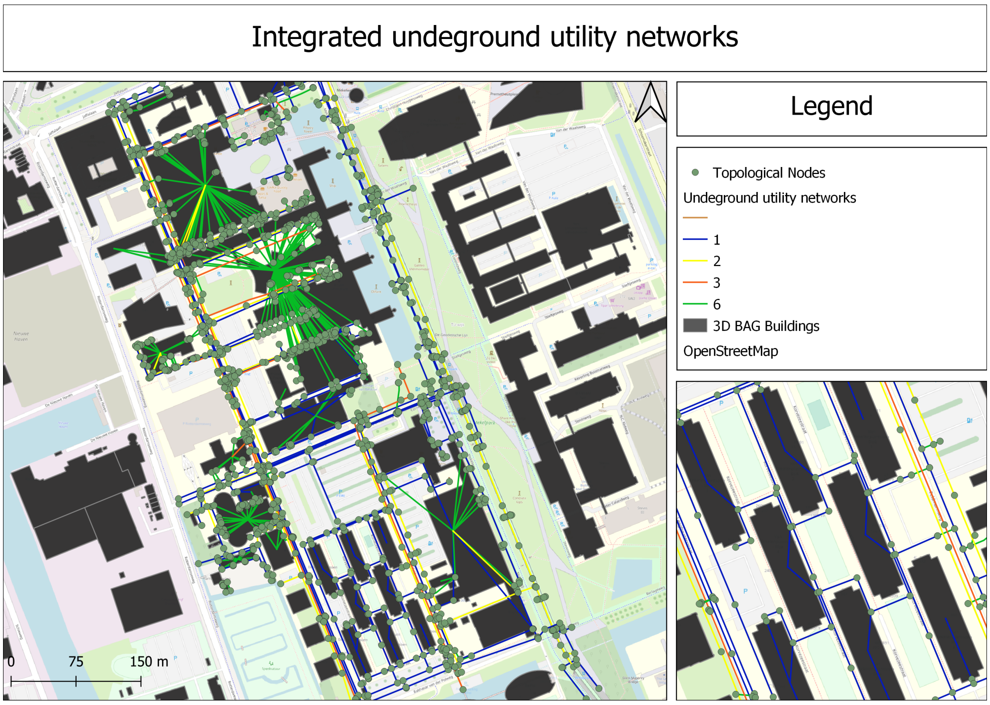


Figure 5.4: Underground utility networks integration

By introducing the above-ground objects, which in our case regard the buildings of the study area, the underground utility networks were integrated with them, as shown in the Figure [5.4], through the creation of virtual edges between the center of gravity of the buildings and the nearby (dead-) ends of the edges per network. The consequence of this union is the direct connection of the networks to the respective nearest building, thus separating the provision of services per network and building. The result is limited to the service of the nearest neighbor (building) as there was no accurate information about the connection of the networks with any of the ground objects (e.g. appurtenances), building, or not. However, this hypothesis made it possible to make a logical distinction between the core networks that serve the city objects (buildings), avoiding cases of unreasonable connections between pipes and objects that are at a great distance from each other (and/or in a completely different location) and are not served by them. A similar approach could be used in the case of an electricity network. In this situation, the (e.g.) street lamps nearest to the electricity branches of the network would be looked for.

5.2 ANALYSIS

To be able to implement any network functionality using the pgRouting extension to PostGIS/PostgreSQL relational database, a set of modifications was made to the stored edge dataset, as mentioned in the previous Chapter (4) in order to follow the structure needed. The modified results in terms of the quantity differentiation (creation of extra edges per intersection) are given in below. It is noted that the number of edges initially stored in the edge table (Topological Edges) was 1416 in which all the selected network categories (4) were included. In each network were included:

- *Main sewer network:* number of edges: 702

- *Infiltration network:* number of edges: 199
- *Fire network pipes:* number of edges: 52
- *Plot connection pipes:* number of edges: 463

After combining the networks based on their hierarchy, for further testing, as has been mentioned, were used only the main sewer network and the plot connection pipes (1165 edges in total). By following the procedure needed to modify this combined networks according to pgRouting structure the altered table contained 336 extra edges, derived from the computed intersections (see Figures [5.5, 5.6]).

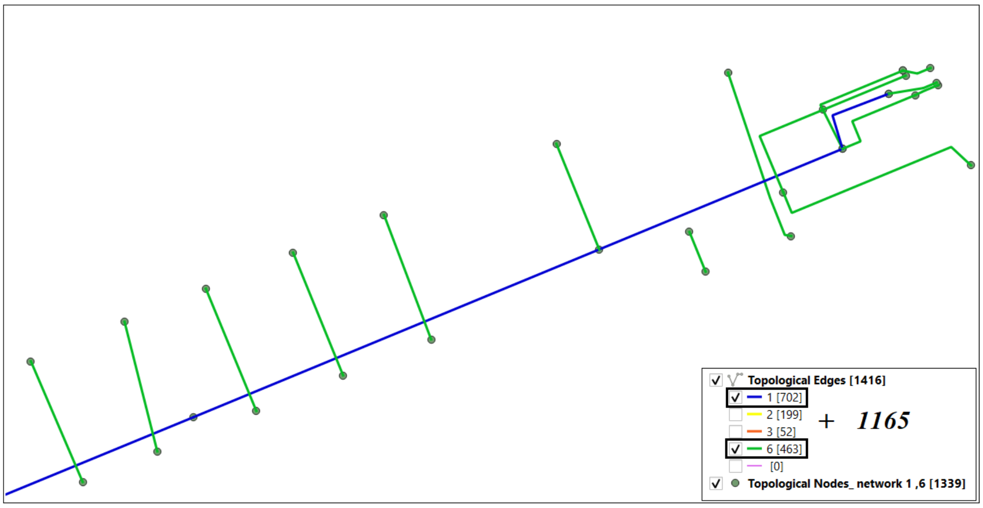


Figure 5.5: Topological edges (sewage-plot pipes networks) **before** pgRouting algorithms implementation

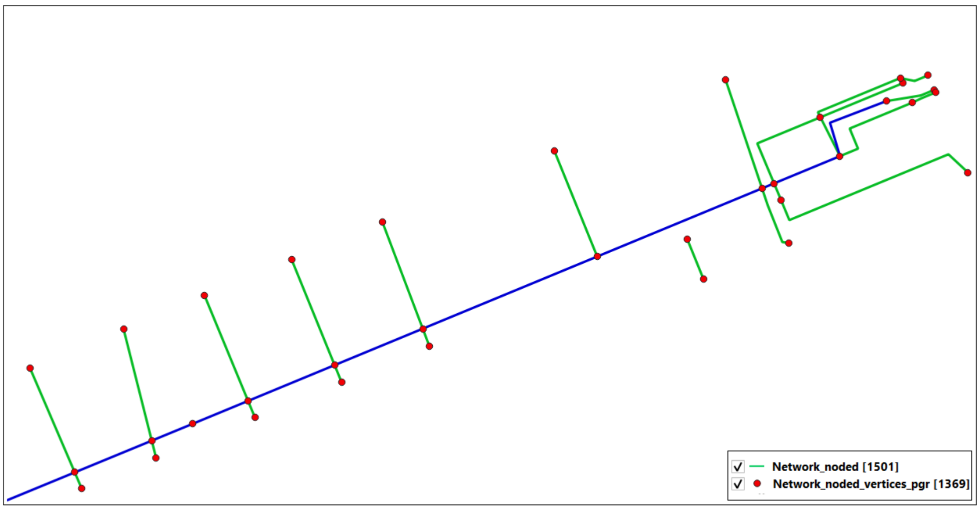


Figure 5.6: Topological edges (sewage-plot pipes networks) **after** pgRouting algorithms implementation

In Figures [5.5] and [5.6] it becomes more clear the difference in the number of edges. Specifically, in the first case [5.5] they only considered the topological edges with the related topological nodes, without taking into account the intersections between the pipes. On the other hand, in the second Figure [5.6] they can be observed the extra (red) nodes created at every intersection between the pipes, which was the desirable graph, in order for the routing algorithms to be executed successfully. These adjustments become more clear in the numerical analysis given in the below lists.

```
  ...
Analyzing for intersections -- 1165 initial edges

ANALYSIS RESULTS FOR SELECTED EDGES:
  Isolated segments: 151
  Dead ends: 687
  Potential gaps found near dead ends: 105
  Intersections detected: 149
  Ring geometries: 0
  ...
```

```
  ...
Creation of a fully noded graph

  Split Edges: 202
  Untouched Edges: 963
  Total original Edges: 1165
  Edges generated: 538
  Untouched Edges: 963
  Total New segments: 1501
  ...
```

```
  ...
Create topology for the noded graph --1501 edges

ANALYSIS RESULTS FOR SELECTED EDGES :
     Isolated segments: 34
     Dead ends: 584
     Potential gaps found near dead ends: 2
     Intersections detected: 1
     Ring geometries: 0
  ...
```

The final network/graph that consists of 1501 edges, fully noded per intersection can now support the implementation of any routing algorithm, based on the available datasets' information. Having ensured that the final model is in a form suitable for use both in terms of geometry and topology, two different scenarios were considered. The first concerns a disaster management case, where a pipe breaks and there is a need to detect the ones directly connected to it since their use will be affected by the disaster and therefore they will stop serving the corresponding city object. Depending on the application, the type of network as well as the availability of the data for that specific type of network, this connection is related to the flow (e.g. downstream) of the product that the pipes (or cable) carried. The second case is related to economic aspects and specifically, the lowest cost route is aimed to be found, considering the amount of the investments that must be allocated for a specific project (e.g. repairing a burst pipe- it may be more expensive to restore one pipe, depending on its size and/or material - rather than following another path).

5.2.1 Case 1: Shortest route, for building service, detection in case of pipe disaster

A case where calculating the lowest-risk route for a particular source-destination combination is in disaster/emergency management. More specifically, in case of damage to a pipe, it must be possible to detect the pipes that are directly connected to it, as well as the corresponding above-ground object that it serves. After reconstructing the network topology, it is possible, now, to find adjacent connections/relationships either with the same or with different pipes. Knowing the nodes that determine the beginning and end of each edge, the information about the adjacent ones that are directly connected to these nodes is automatically extracted and therefore distinguishes their influence or not from possible damage to the intermediate pipe. In particular, taking into account the flow of the product inside the pipe from a point A to a point D (above-ground object) and their intermediate connections (eg A -> B -> C -> D), if pipe A (representing by an edge in the topological network) breaks then all subsequent, up to the destination point, pipes will go out of operation. An example of this case is illustrated in Figure [5.7].

Figure 5.7: Pipes out of operation

In the above case, if the pipe represented by the *edge= 394* collapsed, then all the pipes connected to it and carrying the related content (drainage and/or drinking water) will cease to operate and the specific side of the building will stop being served by them. In this case, it is not possible to find an alternative route since the connected pipes reach dead ends, which are connected directly to the building. However, as can be seen in the sub-image, the tubes located prior to this edge are not affected by the existing fault, as they do not intersect at the starting point (starting node) nor at any point in the middle of it (Figure 5.8).

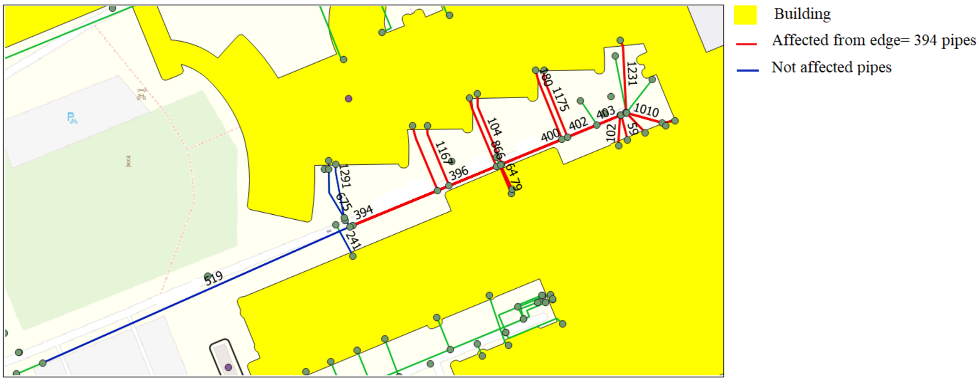


Figure 5.8: Pipes on operation

In this case, the route between *edge= 519* to *675, 1291, 241*) can be used effectively and thus the building is still served by them. To calculate the optimum distance between two (or more) selected nodes the algorithm evaluates only the edges with positive costs, and it returns a result only if the path exists (connected network). A positive cost edge is considered the one that allows for the computation of an optimal path by adding its value to the next one (next edge) until reaching the destination point, by keeping at the same time the nodes that have been visited. In the case of negative weight edges, the algorithm terminates and the optimal path has been successfully defined. It is noted that the algorithm returns a result only in case there is a continuity between the pipes, thus unconnected/ single pipes are not taken into account and an empty table is returned.

Another example of finding the optimal path using one to many relation is given in Figure [5.9].

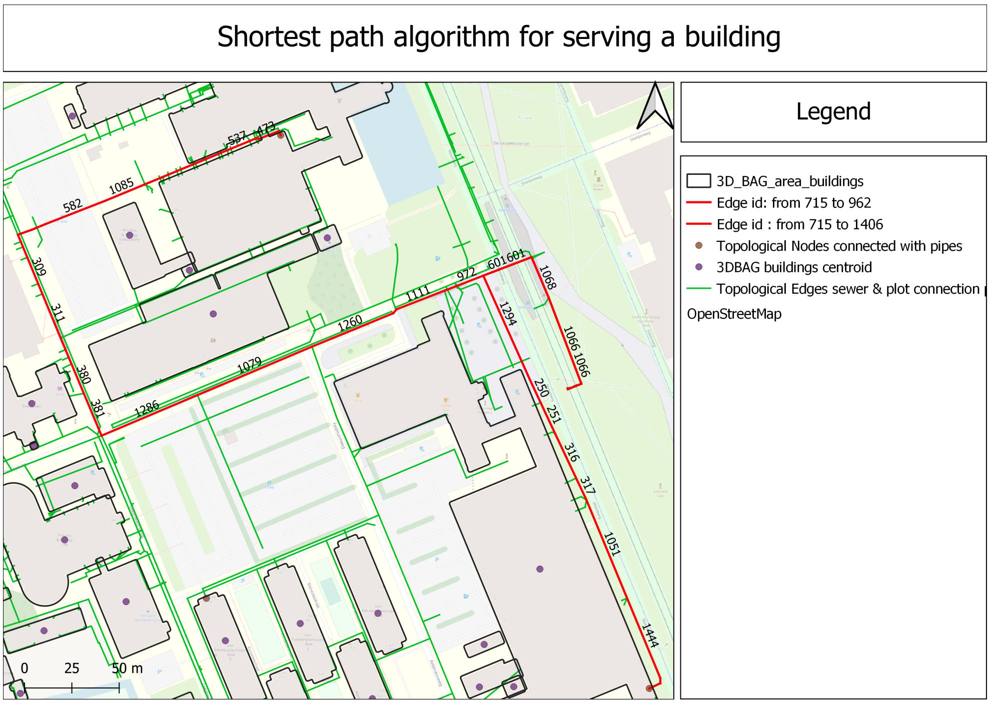


Figure 5.9: Shortest path between nodes

In this situation, the routes for the service of several buildings are placed from a single starting point. In circumstances when the flow of products transported by the pipes depends on certain starting locations, then it is very helpful to know the paths that are directly influenced by this point. This information is needed as in case of failure at this point, all routes will be affected, taking note that the flow has a direction in the sewer systems (in the case of bidirectional networks -electricity- other routes will be determined). Both paths will be completely out of order in case of failure of the starting point, considering that the flow of the pipe product follows a specific single direction. Therefore, both the marked red line will be affected as well as the sub-networks connected to it and extended to the building. Through this analysis, it is possible to predict the pipes directly affected by a failure and to start scenarios for the restoration of any damage. It is noted that the buildings are served by multiple pipes that are connected at different points around them, so if a route is affected then a specific part of them will cease to have facilities.

The same approach but with a different impact in terms of the affected objects was followed using the addresses point dataset of the examined area. Here, in a possible pipe disaster, it is not affected the whole building but instead the specific building entities located at a certain address. If in an address there are more than one building entity (house numbers), then all of these entities are detected as immediately affected objects from the disaster. In this case, it becomes more specific the amount of the affected objects and by further combining this information with the registrations per building entity it would be possible to identify the number of the affected households in case of a simple residential (building) use or to a greater extent the damage in public service building (e.g. hospital). An example of the shortest path routing algorithm and the affected of a possible disaster building entity is presented in Figure [5.10].

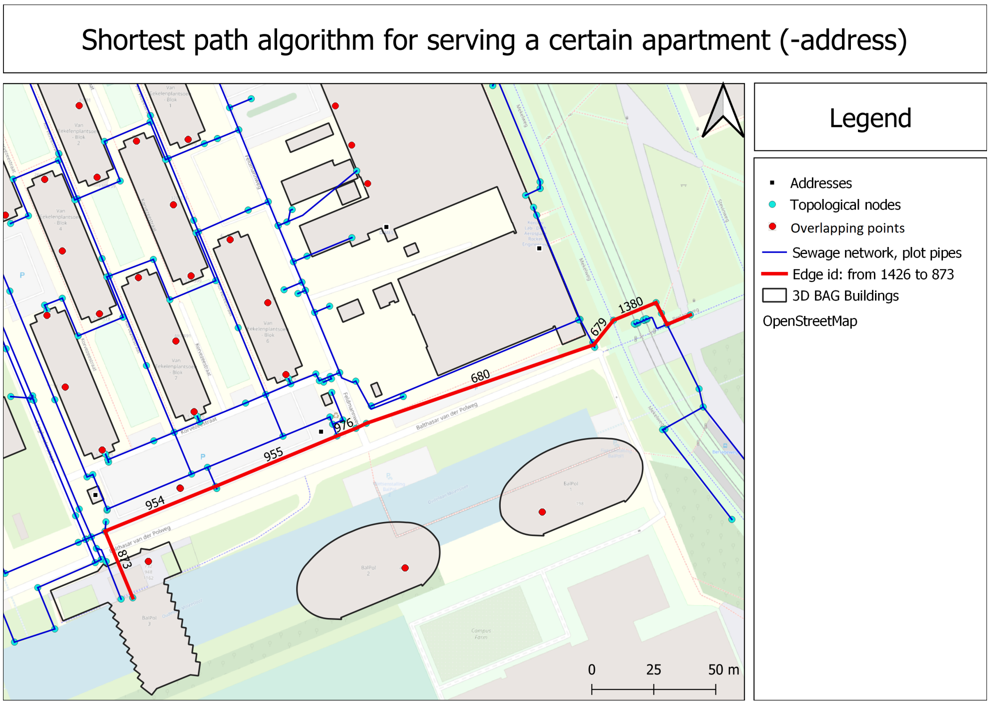


Figure 5.10: Example of optimum path and the affected addresses in case of disaster

In the above Figure, they have been selected two nodes *source: node_id = 1346* to *target: node_id = 926* that correspond to edges *source: edge_id = 1426* and *target: edge_id = 873* and the shortest path between them is highlighted with the red line. As mentioned in the previous section all of the available addresses are associated with their nearest topological node corresponding to a certain edge, which represents a pipe, and so if a topological node (or both nodes bounding the edge) of an edge is associated with an address and this pipe collapses, then all the corresponding building entities will be affected. To check if the associations as well as the data structure in the relational database works effectively also in the case of this dataset, different tests were implement to cover all the possible cases, based on the complexity of the stored records in the addresses point dataset (see Figure 4.23).

Starting with the selected edges, along the path there are more than one topological nodes that are (internally) connected (serve) an address (Figure [5.11]). Based on the available information about the sewage network only the destination topological node is considered accurate for the service the associated with its address since the middle one, serving the building on the other side of the canal is not possible to be served from that network, but from other pipes that are not included in the current dataset. However, for the testing purposes, these situations did not exclude and in case a topological node is connected with these two buildings located on the opposite side of the canal, they remained in the routing analysis. It is noted here that in Figure [5.11] the red points represent the addresses where are located more than one house (>1 house number) and thus in a possible pipe collapse, the whole complex will stop being served by it.

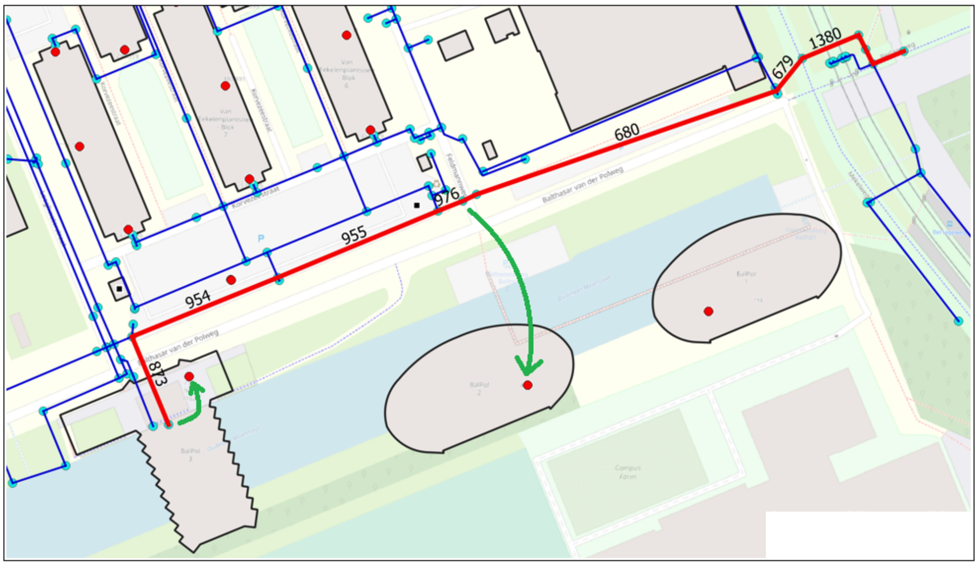


Figure 5.11: Addresses points associated with the nearest topological nodes

The resulting table this time apart from the information related to the graph and its connectivity (source- target nodes), returns the address information that is associated (if it is associated) with the topological nodes along the computed path. If this address belong to the duplicate points, then also this information is retrieved. An example of the returned tables is given below:

Table 5.1: Shortest path algorithm results

seq	path_seq	node	edge	cost	agg_cost
1	1	1346	1426	10.222	0
2	2	1342	1501	5.001	10.222
3	3	1364	1470	4.863	15.223
4	4	1321	1380	18.581	20.087
5	5	713	679	12.786	38.668
6	6	712	680	97.039	51.454
7	7	714	1476	4.649	148.494
8	8	1026	976	8.190	153.144
9	9	1006	955	53.0521	161.334
10	10	647	954	48.186	214.386
11	11	751	873	29.131	262.573
12	12	926	-1	0	291.705

The above table provides the information related to the shortest path followed to connect the selected nodes. The route is calculated by the execution of the Dijkstra algorithm (*pgr_dijkstra* function) and the returned information concern the starting node per traversed edge as well as the corresponding edge id, accompanied by its cost value (concept of cost proportional to the length of the edges). The last column stores the aggregate cost each time one edge is successfully crossed and the algorithm proceeds to control the next one. Once the route has been calculated its information was combined with the related information stored in the other tables concerning the edges of the network and the associated to them addresses that they serve. The returned table and its content is shown in Table 5.2:

Table 5.2: Optimum path combined with the served addresses

Address id	House number id	Building id	Network id	Source node	Target node	Edge id
NULL	NULL	NULL	1	533	607	1470
NL.IMBAG. Nummeraanduiding. 0503200000106044	NL.IMBAG. Verblijfsobject. 0503010000007568	NL.IMBAG. Pand. 0503100000004242	1	601	323	1476
NULL	NULL	NULL	1	323	92	680
NL.IMBAG. Nummeraanduiding. 0503200000123570	NL.IMBAG. Verblijfsobject. 0503010000054830	NL.IMBAG. Pand. 0503100000033026	1	182	183	873
NULL	NULL	NULL	1	607	530	1380
NULL	NULL	NULL	1	117	326	955
NL.IMBAG. Nummeraanduiding. 0503200000106044	NL.IMBAG. Verblijfsobject. 0503010000007568	NL.IMBAG. Pand. 0503100000004242	1	326	601	976
NULL	NULL	NULL	1	455	456	1426
NULL	NULL	NULL	1	182	117	954
NULL	NULL	NULL	1	533	455	1501
NULL	NULL	NULL	1	530	92	679

In the above table, is presented the information derived from the internal join of the Table 5.1 and the topological edges table (the modified one following the pg_routing rules) stored in the relational database. With this connection, it is extracted the information related to the addresses that are served from the edges (pipes in reality) constitute the optimum path. It can be observed that again not all pipes serve an address and thus an empty record is returned (NULL value). The edges bounded by nodes that are associated with an address store the information of the address on the related field. It can be noticed that there are two records with the same information, which mean that two edges share one node (source in one case, target in the other) and so this association is presented twice. The common node is the one with node_id = 601 that serves the building on the other side of the canal (see Figure [5.11]). The process continues with the examination of the inclusion of these addresses in the overlapping points table. Once the served addressed are detected the building id is used to verify its existence (or absence) in the overlapping point table. If a certain address exists in the before mentioned table then all the different house numbers (building entities -e.g. building apartments) returned as they will be immediately affected by a possible disaster in the pipe they are connected with (Table 5.3). On the other hand, if this address is single (no more than one house number) then an empty table is returned.

Table 5.3: Identification of no-single address point

count	nummeraand	networkid	nodeid	pandid
1	NL.IMBAG. Nummeraanduiding. 0503200000105986	1	601	NL.IMBAG. Pand. 0503100000004242
1	NL.IMBAG. Nummeraanduiding. 0503200000105987	1	601	NL.IMBAG. Pand. 0503100000004242
1	NL.IMBAG. Nummeraanduiding. 0503200000105988	1	601	NL.IMBAG. Pand. 0503100000004242
. . . .				
1	NL.IMBAG. Nummeraanduiding. 0503200000123499	1	183	NL.IMBAG. Pand. 0503100000033026
1	NL.IMBAG. Nummeraanduiding. 0503200000123500	1	183	NL.IMBAG. Pand. 0503100000033026
1	NL.IMBAG. Nummeraanduiding. 0503200000123501	1	183	NL.IMBAG. Pand. 0503100000033026
. . . .				
Total number of apartments (unique house number ids) = 441				

The above table gives an example of the returned information from the query executed in order to detect overlapping points, which means more than one building entities located at the same address. In total the were detected 411 records. It is noted here that in these records there are no duplicate building ids although one of two buildings appeared twice in Table 5.2. This distinction was made to keep only the information needed avoiding the redundant one. The above example concern one-to-one connectivity between two nodes, but the algorithm tested with one-to-many connections as in the building centroid case (see Figure 5.8).

5.2.2 Case 2: Cost optimum route detection

A second example of using the same algorithm is to determine an economically optimal route (cost optimization). Note that if the relevant cost information had been stored in the data attributes, then the *pgr_dijkstraCost* function could be used to calculate the aggregated cost per route. However, since this information was not available, an assumption was made. This assumption considers that the cost is proportional to the length of the edges (pipes) per network, ans thus the length value is used instead of the unknown real cost. To detect a low-cost route, the real length value was modified with an extreme value (e.g. from 12,5m to 100000m). In this case, other routes sought to be found in order to reach the selected destination point. Also here, only the positive value edges were utilized and once the negative one is detected the algorithm returns the result. The negative value is generated automatically at the point where the target point is reached (dead end of the destination edge).

To test the functionality of this hypothesis, some auxiliary structures were created, which include the information (attributes) of the original table containing the edges, with the only exception being an additional column (modified_length), where are defined the costs per edge. An example of the query used to modified a particular edge, as well as the query used for the detection of the lowest-cost path is given in the Appendix (7):

The visual example of the corresponding query is shown in Figure [5.12]

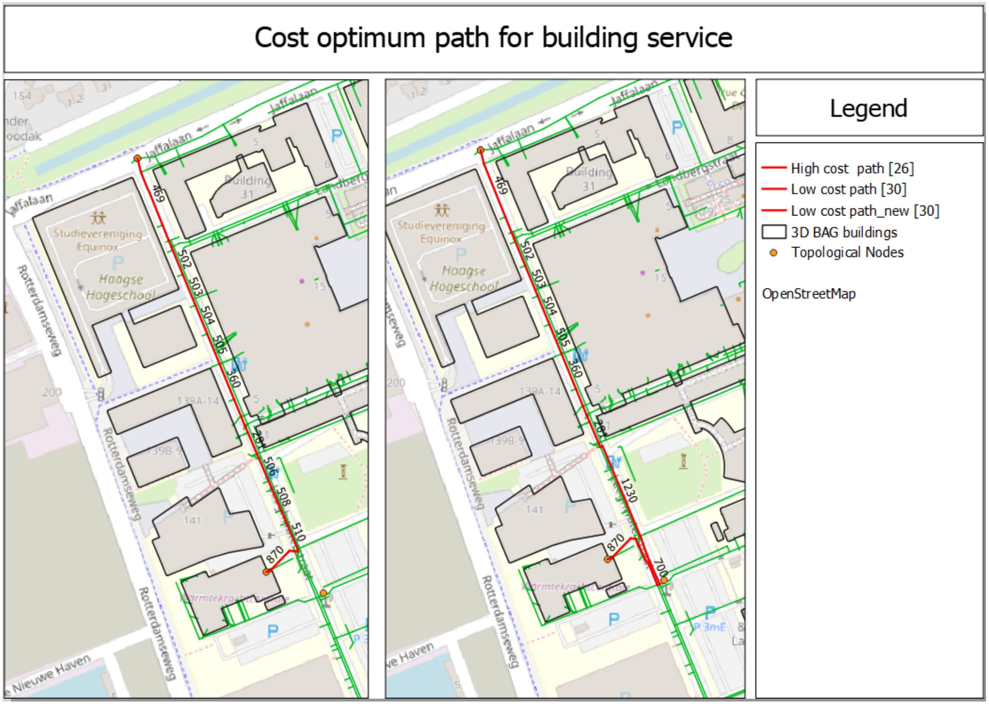


Figure 5.12: Cost-optimum path detection

In the picture above, the initial shortest route followed in order to serve the selected building is shown in red, accompanied by the corresponding edge IDs. By modifying the lengths of the multiple lines/ edges, increasing them significantly, and re-applying the algorithm, it is observed that the path (automatically) changed (right sub-figure), following the path that in the end gives a lower total cost (aggregated lengths of the traversed edges).

The following tables present the information of the edges that are traversed in the case of the economically optimum route, after the modification of costs:

- *Edge: 508* - length 100000.00m
- *Edge: 807* - length 100000.00m
- *Edge: 1306* - length 100000.00m

The physical significance of replacing the original length of an end with an extreme value is to approach a real-world scenario where a high amount of money must be invested to follow a path (e.g. construction of a new pipe and/or extending the network with starting point one of the selected topological nodes). In this case, alternative routes are sought, if they are available and accessible, in order to serve the corresponding above-ground object with which the selected nodes are connected. In the first table (5.4) are presented the records derived from the execution of the Dijkstra algorithm (same as the results in the previous example -see table 5.1). All this information does not contain the edge geometry and thus it should be related with the initial edge table to extract it. This can be, now, implemented effectively since the tables are internally connected in terms of primary and foreign keys.

Table 5.4: Cost-effective route computation

seq	path_seq	Node	Edge	Cost	Aggregated cost
1	1	321	1460	9.473446425792797	0.0
2	2	467	469	38.28246923214117	9.473446425792797
3	3	468	470	11.203377038251343	47.75591565793397
4	4	469	917	4.495574098592385	58.959292696185315
5	5	495	501	2.200004634384923	63.4548667947777
6	6	496	502	17.285455657800174	65.65487142916263
7	7	497	503	22.002865106390114	82.9403270869628
8	8	145	504	20.49926845392435	104.94319219335291
9	9	226	505	21.49996330294493	125.44246064727727
10	10	396	360	25.199975697894235	146.9424239502222
11	11	397	361	12.455136142233453	172.14239964811645
12	12	46	362	6.210597319166775	184.5975357903499
13	13	398	1000	1.152473290549216	190.80813310951666
14	14	339	280	4.1509460479223605	191.96060640006587
15	15	340	281	15.163708280669722	196.11155244798823
16	16	341	1131	2.9708326457109977	211.27526072865794
17	17	498	1419	1.6557047635222704	214.24609337436894
18	18	823	1230	65.04225868638667	215.9017981378912
19	19	678	654	1.3976908906136516	280.9440568242779
20	20	676	653	12.925083249050816	282.34174771489154
21	21	677	700	20.429629588716164	295.26683096394237
22	22	735	1465	1.7067720664359822	315.69646055265855
23	23	880	824	15.430298636132958	317.4032326190945
24	24	542	554	3.7690604911801375	332.83353125522746
25	25	541	1277	10.09801312344755	336.6025917464076
26	26	392	1372	0.4054910749318553	346.70060486985517
27	27	352	300	4.962866430633872	347.106095944787
28	28	282	225	3.4434513085424556	352.0689623754209
29	29	283	870	17.61904671754125	355.51241368396336
30	30	239	912	3.471252839117845	373.1314604015046
31	31	964	-1	0.0	376.60271324062245

The following table shows the link between the previous table results with the table of the topological edges of the combined networks (main sewer network and plot connection pipes). Once their connection has been achieved, the geometry stored in the latter table that allows for the visualization of the result is extracted (see Appendix 7 for the completed table including geometry column). Apart from the geometry information, also information about the buildings is obtained that are served from the available networks. In case an edge stops at a point that is not related to any of the available ground information then the particular building id will be absent.

Table 5.5: Cost-effective route detection including building's service

Building ID	Network ID	seq	path_seq	Node (source)	Edge ID	Cost	Aggregated cost
None	1	1	1	321	1460	9.473446425792797	0.0
None	1	2	2	467	469	38.28246923214117	9.473446425792797
None	1	3	3	468	470	11.203377038251343	47.75591565793397
None	1	4	4	469	917	4.495574098592385	58.959292696185315
None	1	5	5	495	501	2.200004634384923	63.4548667947777
None	1	6	6	496	502	17.285455657800174	65.65487142916263
None	1	7	7	497	503	22.002865106390114	82.9403270869628
None	1	8	8	145	504	20.49926845392435	104.94319219335291
None	1	9	9	226	505	21.49996330294493	125.44246064727727
None	1	10	10	396	360	25.199975697894235	146.9424239502222
None	1	11	11	397	361	12.455136142233453	172.14239964811645
None	1	12	12	46	362	6.210597319166775	184.5975357903499
None	1	13	13	398	1000	1.152473290549216	190.80813310951666
None	1	14	14	339	280	4.1509460479223605	191.96060640006587
None	1	15	15	340	281	15.163708280669722	196.11155244798823
None	1	16	16	341	1131	2.9708326457109977	211.27526072865794
None	1	17	17	498	1419	1.6557047635222704	214.24609337436894
None	1	18	18	823	1230	65.04225868638667	215.9017981378912
None	1	19	19	678	654	1.3976908906136516	280.9440568242779
None	1	20	20	676	653	12.925083249050816	282.34174771489154
None	1	21	21	677	700	20.429629588716164	295.26683096394237
None	1	22	22	735	1465	1.7067720664359822	315.69646055265855
None	1	23	23	880	824	15.430298636132958	317.4032326190945
None	1	24	24	542	554	3.7690604911801375	332.83353125522746
None	1	25	25	541	1277	10.09801312344755	336.6025917464076
None	1	26	26	392	1372	0.4054910749318553	346.70060486985517
None	1	27	27	352	300	4.962866430633872	347.106095944787
None	6	28	28	282	225	3.4434513085424556	352.0689623754209
None	6	29	29	283	870	17.61904671754125	355.51241368396336
NL.IMBAG. Pand. 0503100000031391	6	30	30	239	912	3.471252839117845	373.1314604015046

The same procedure was followed to check the algorithm with the addresses point dataset. Again, they have been selected two nodes and the optimum path followed to get connected was determined. Then, the lengths of some edges were modified to extract the adopted to the shortest length path. The results are shown in Figure [5.13].

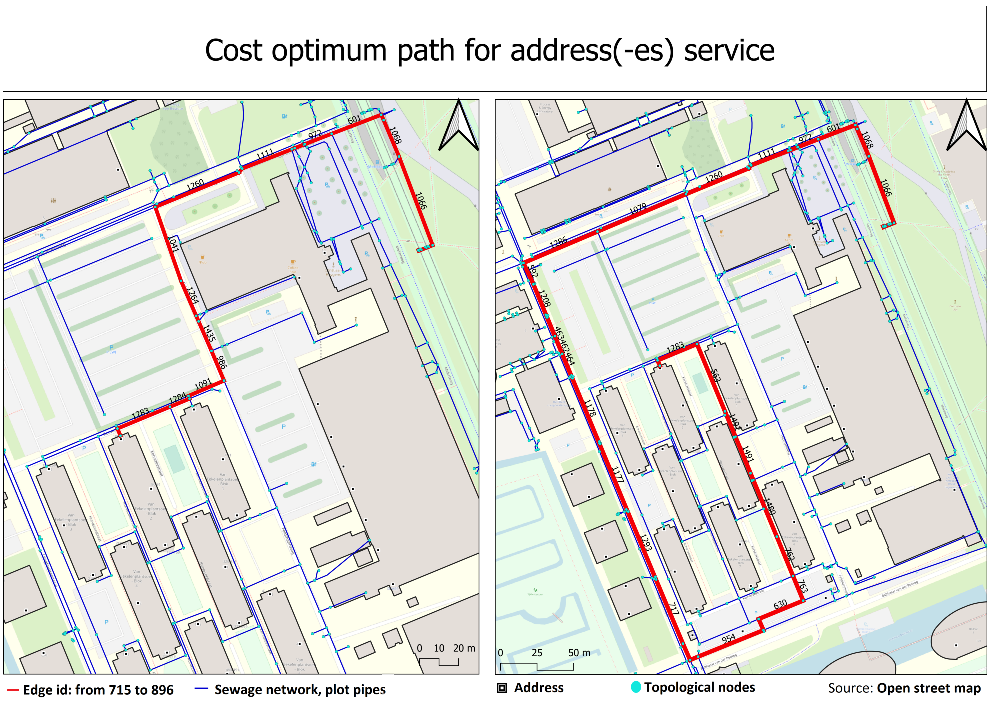


Figure 5.13: Cost-optimum path detection- addresses dataset

From the above examples, as well as summarizing the overall results of intensive analysis and processing of the original data, it is concluded that following the basic principles of topology and logical connections between networks, it was possible to create a coherent network capable of different applications. However, it is noted that to achieve the final product, a significant number of assumptions were made as important information was absent. For example, with the existing reconstructed network, the maximum flow of the pipes could be calculated (using *pgr* functions), based on their spatial capacity. Especially for sewage networks, their capacity accompanied by the dimensions and material of the pipes contribute significantly to the development of more detailed models and simulations of their operation, which is a suggestion for future research. Based on the latter, both the summary of the present work and future work are presented in the next chapter.

5.3 comparison between the building-centroid and addresses approach

The completion of the tests per case study using the two approaches, the building-central and the address points, resulted in different results in terms of servicing the ground objects from the underground utility networks. In more detail, in the first case, which was considered as the connection point the centroid of the buildings, a simplified version of the connection of the underground networks with the ground objects is presented. This may be a strong simplification, which was based on estimations but since there was not available a clear connectivity link between the existing entities, this approach presents a logical representation of reality.

On the other hand, using the address point data set, a more accurate connection was achieved between the utility underground networks and the above-ground entities. This accuracy stems from the fact that a better spatial distribution of connection points between networks and ground objects was evident in this dataset. Unlike the centroid that represents an entire building, the addresses are scattered around it and therefore there is better spatial information. Having multiple connectivity points (available in the addresses dataset) more accurate integration is achieved. This results from the first case study, where, in a pipe failure, the first approach generalizes the damage at the building level, while using the addresses only a part of the building appeared to be affected. However, in both cases, assumptions have been made as an expert, who can verify these links, and their reliability was not apparent.

5.4 results discussions/limitations

Completing the analysis of the data in terms of their content as well as the study/efficiency of the produced product (topological graph representing the underground utility networks), various shortcomings were noted. These shortcomings are due both to the content of the available resources and to the various assumptions that had to be made to examine whether and how it is possible to create the desired model and its usefulness. Below are listed the explained the limitations per assumption.

1. **Topology reconstruction:** Considering the state of the original datasets, a set of assumptions should be made to reconstruct their topology and create a

connected network. These assumptions are related to modifications in the topology and/or geometry of the data, such as the relocation of some edges representing the pipes/cables and/or the addition of new ones to connect network elements. These network elements correspond to both the lines and points available in the datasets. With respect to the last-mentioned elements corresponding to the nodal network points (appurtenances) that connect the existing disposable pipes and/or auxiliary pipes, due to their incomplete and vague content, they were replaced by the topological nodes bounding the edges of the network graph. To have a fully topological connected network, both geometrically and in terms of attributes, these points (survey points see section 3.4.3) were associated (spatial overlay) with the topological nodes extracted during the data processing procedure, with the ultimate goal to maintain their attributes and they were not further utilized. The limitation, in this case, is that due to their vague topology and not direct connection to the networks, it was not possible to consider them, as non-negligible changes had to be made to the entire dataset, something that went beyond the scope of this thesis. However, the relationship with the topological nodes was essential to preserve the original information in the final model, since they represent the initial geographical information.

2. **Underground utility network 3D information:** Although the provided dataset were considered to include the third dimension (depth), this information was quite limited and/or absent. A complete transformation from a 2D line/edge to 3D means that each point that forms the line must be assigned a depth (Z information), so the corresponding geometry will include all dimensions. So to extract insert the desired 3D information the digital terrain model of the study area was used. This solution worked temporarily as geometric 3D information was rendered at each point of the pipes that make up a network, but this affects the exact representation of reality, as the lines are draped on the ground model following the corresponding slopes. Especially in the case of the underground networks where we talk about concrete pipes/cables that are placed at fixed locations (based on the standards and specifications per facility) this technique is limited, while for the sewage network, in particular, which is the only one has to follow an inclination (in order for the pipes product to be able to flow) further information was needed. As a result, the 3D information added to the data does not correspond to reality and it was retained for data exploration reasons, for understanding the differentiation of network levels as well as for the optimization of their visualization.

3. **Underground utility network integration with the above-ground objects:** For the integration of the available below-ground networks, it is noted that it is limited in the connection with the buildings of the study area (as a unit) and in the available addresses. Other city objects related to the network elements that theoretically correspond to the ground points that should be connected and/or serve, were not sufficient, either in terms of their exact location (inaccurate position) or based on their connectivity to any of the available and thus, were not utilized. For that reason, it was assumed that the buildings (represented by a point), as well as the addresses, are connected to the nearest topological nodes of the pipes of different uses. In this way, the derived integrated model concern a limited part of the above-ground condition, which is based on logical connections between the available entities (nearest neighbor relationships).

4. **Relational database modifications:** Once the model was created, it was stored in a customized relational database. In order to use the database and its capabilities for further network applications, the data have been modified to follow the principles of internal linking between different tables containing the geo-

metric information of different elements (lines, points). The logic of modifying attributes of the data is the development of interdependence relationships, so for the available information to be combined and each time depending on the application the corresponding result to be extracted, while in case of a change in a table the associated information on the related one, will automatically be updated. For example, in the formats developed in the present work to identify the optimum path between two (or more) nodes, information stored in the different tables had to be joined in order to acquire the route in terms of the edges as well as the related to the attributes. This internal connectivity was successfully achieved with the help of a table in middle the allows the join of the different tables stored in the database.

During the research of the most appropriate functions and queries needed to be executed to implement the real-world scenarios, some problems with 3D information had arisen. More analytically, as the data with line geometry did not contain all the three-dimensional information, but only at the starting and ending points (nodes), a restriction on a set of properties related to the Z (3D) information was revealed. Although the column that stores this information included the third dimension, the corresponding geometry did not and so, due to the unconventional storage of this relevant data, the geometry was forced back into 2D, in order to be able to execute the routing algorithms. Through the return to 2D, and while utilizing the pgRouting topographic functions, other issues related to the loss/absence of 3D information detected. Specifically, when the nodes were to be created for the development of a topological connected network, the algorithm could not automatically separate the differentiation of the networks based on their use and level. The consequence of this is the wrong correlation and further exploitation of underground networks. Auxiliary structures, therefore, had to be created, which increases the complexity of developing and studying the various functions and their corresponding effects.

From the above statements, despite the lack of information and the assumptions made, it was possible to move from the geographical information to the topological graph. Of course, the initial constraints on the data related to the status of their content, imply the creation of chain constraints, which were solved each time by making assumptions and/or simplifications of the original data, keeping those that were in a more efficient condition and could be used (better quality/quantity in terms of geometry, topology, attributes).

Summarizing the condition of the data and the existing limitations that directly affect the quality of the produced model can be mentioned and suggested the following:

The complete and accurate representation of reality suffered from shortcomings in terms of quality and quantity of data provided. In addition to the lack of information about the features as well as their disorganized topology and geometry, the data lacked basic metadata that would facilitate their understanding and reduce (if not completely avoid) assumptions. Specifically, concerning the first category (attributes), all units of measurement were absent and therefore the perception of volumes of physical (measurable) quantities (network edges and their length) was limited. Since the fields (filled or not) are related to the length of the pipes, their dimensions, and the reference surface used for the depths, it was important to determine their units of measurement to communicate the values of the presented quantity to the user. Another obstacle was the discrepancy between their content and any of the existing standards and therefore the impossibility to use already developed schemes/structures that support the storage, representation, and analysis of underground utility networks. International standards presuppose the existence of data that follow the current rules of compliance with content and formats with the goal of avoiding making decisions based on bad data that lead to potentially undesired consequences. A direct mapping of the available data to any of the standards supporting such geometries and attributes is not a straightforward procedure. Considering the amount of the existing inconsistencies in terms of topology-geometry-attributes the content of the data apart from a detailed cleaning needs to be enriched with attributes corresponding to the standard that is to be used for the mapping. In our case, that was not immediately possible, since although the data were cleaned and their topology was reconstructed, they cannot be considered completely reliable/realistic due to the assumptions have been made. However, the mapping process would be more valuable if it was possible to enrich the data with all the necessary elements (for the current standard) from the beginning of the data creation, otherwise, many assumptions need to be made again to suit the needs of the networks. Finally, the information that was missing for the creation of the requested model was that of the third dimension. As efforts are made to create representative models and simulations of reality, the third dimension that defines the size, location, and orientation of objects (terrestrial and underground) and its inclusion in vector data are crucial. Nevertheless, it is, also, of vital importance for the corresponding topology to represent reality since it stores all the information about the objects and the between them relationships. Therefore, prioritizing the needs for creating a model, we need first, a sufficient amount of data, a correct topology that corresponds to reality, and then the addition of the third dimension that clarifies the different levels in the model.

6 CONCLUSIONS AND FUTURE WORK

6.1 RESEARCH OVERVIEW

The overall goal of this thesis was to develop a three-dimensional model that would combine underground utility networks with above-ground objects (city objects) to depict a holistic 3D model of our urban environment. The underlying idea that was the driving force of the current thesis was the fact that simulation is an extremely valuable approach for understanding such complex systems (combination of below with above-ground objects - e.g. in the case of underground utility networks). As a result, their modeling became a standard tool in many engineering domains and currently, efforts are made by the GIS community, to extend and improve their digital representation. Combined with optimization simulations, in particular, these models progressed from a troubleshooting tools to standard virtual validation tools and, eventually, design driving tools. The main idea of the thesis (starting point), was the creation of a 3D model combining the underground with the above-ground information (utility networks with city objects) that would be mapped according to CityGML Utility Network ADE rules, in order to simulate reality and examine its usefulness in real-scenarios applications. However, the key concern during the different stages of thesis implementation was to get a more comprehensive model that required to account for topology. Although 3D information plays an important role, it is not possible to consider it while having a "broken" topology. It was made clear that these two key factors are of equal dignity for the development of the desired model. So in the first place, the provided data was inspected in the form of depths, as we are dealing with underground settings, or through pertinent height data (semantic height information), examining at the same time the condition of the existing topology (evaluating the amount of inconsistencies) and the utility of having a coherent model in various applications. To accomplish this, it was necessary to employ qualitative and quantitative analysis (statistics) to split their content into useful/usable and incomplete/irrelevant categories in order to address both the core research questions and other questions that occurred during the familiarization with the data. The main research questions addressed by this thesis are:

- *How is it possible to model utility networks in 3D, integrated with the above-ground objects, such that they can be suitable for multiple uses?*

  - **Answer:** The answer to the question is: *partially achievable, using the available tools/software*. It is indeed possible to have an integrated model (graph) from the underground utility networks with the above-ground objects (in our case the buildings) but this was implemented by making a set of assumptions related to the topology of the provided data. Regarding the 3D information, it was added to the data with the use of the digital terrain model, but yet there is a considerable distance between the current model and a comprehensive one that would include a fully correct topology, as well as would include the 3D information. The critical points for the creation of the desired model were both the existing attribute information and the topology of the initial dataset that constitute the basis for the implementation of any network application. The reconstruction made allows for routing analysis in any network for which the 3D information does not affect its use. For example in the case of cables and pipes where the gravity and/or ground inclination do not affect the transportation of the cable/pipes product the graph can effectively be used. However, in the case of sewage networks, 3D information plays an important role when it comes to the simulation of their use (e.g. flow direction).

- *How to represent a direct connection with the above-ground condition?*
  - **Answer:** The answer to this question is: *by converting the geographical information to a topological graph*. A direct connection between the available objects means the creation of relationships based on their function. These relationships are related to both the connection between geometries as well as the creation of attributes allowing for their internal connectivity (primary-foreign key relationships). In the case of the buildings used in the current thesis, their geometry had to be modified and from polygon representation they were replaced by point representation. This was done, in order to create the connections between the topological nodes and the building point (centroid) used to represent it (point-to-point connection- "virtual edge" creation- same approach for the addresses case).
- *Is it possible to achieve that connection?*
  - **Answer:** The connection between the underground utility networks with the above-ground objects used (buildings) was *possible*. This connectivity is now apparent in the attributes of the different datasets, after associating them internally utilizing the new added attributes that allow for the creation of interdependence relationships.

Based on the results of this preliminary analysis and the first contact with the data, other questions arose, such as:

- *Is the 3D information useful? -for the current as well as for other applications*
  - **Answer:** The available in the provided Tu Delft datasets 3D information was *not useful directly*. It is noted here that the 3D information was not useful since its amount was quite limited and/or could be considered reliable due to metadata absent. Thus, the 3D information extracted using the DTM of the study area only facilitated a better understanding of the underground networks and the differentiation of the levels they are located. However, these layers were created based on assumptions, and thus the 3D information cannot be considered reliable. In the cases studied, which are applied to a 2D graph, the 3D information does not directly affect the results. However, if 3D information and/or other information was available regarding the gravity and the slope of the ground, then it would be crucial for sewage simulation. This simulation would be a more reliable model that could represent the flow direction of the transported product. The significance of 3D information depends on the application and the purpose for which it is to be used.
- *Can 3D information be replaced by relevant information?*
  - **Answer:** This question cannot directly be answered, since in our datasets although there was available (some) information about semantic height, it was not known the reference surface used and the exact meaning of the available values, and thus these details were not utilized. In case both in-

formation was known another process could be followed (offset from a specific surface -that would lead to a more accurate depth determination). Depending on the accuracy and reliability of this information (semantic height information) the derived model will have the corresponding reliability.

- *Is it possible a limited 3D information to be extended to a larger network?*
  - **Answer:** Same as the previous question, there is no concrete answer. In case there was an accurate 3D information on the datasets (even partially), other approaches would be followed for its extension to the whole dataset (e.g. spatial interpolation techniques).

According to the above-mentioned questions, the methodology to be developed for the desired modeling, which was applied to a part of the Delft University of Technology in the Netherlands, was clarified, by utilizing the vector datasets provided by Directie Campus & Real Estate Department. The main phases developed in this process concern the detailed study of the datasets, seeking their full understanding in terms of their content (attributes, geometry, topology), and their process (cleaning, processing, integration) in order to modify them to cover the needs of the proposed application. The very first stage is one of the most important steps in identifying all the constraints as well as the possibilities of exploiting the data by developing the appropriate approaches.

To move forward to the 3D model creation, the original geographical data had to be replaced with their digital graph, which is nothing more than their topological representation in the form of a graph, by utilizing various tools (software) and approaches. Several constraints were revealed throughout this transition (see Chapter 5.4), which were temporarily overcome with key assumptions, simplifying the complexity of the network representation and the associated information. Final step before the implementation of real-scenarios applications was the integration of the underground utility networks model with the above-ground city objects, which in the current thesis were considered only the buildings of the study area. This integration was aimed in order to be able to detect the direct connection between the utility networks and the corresponding buildings they serve, as well as with the available addresses, acquiring this way a complete view of reality and the interdependencies between below and above ground objects.

6.2 Discussions & Limitations

Completing the research of the present dissertation and combining the results with the existing ones in the available literature (see Chapter 2.3), it was observed that their common point is the low quality of the vector data used in each application. The development of any application requires the existence of a satisfactory data set and/or combination of data from different sources to meet the needs of each application and achieve the desired result. In this thesis, where the desired model aims to represent underground utility networks and their relationship to above-ground city objects, data of the existing cables and pipes accompanied by other network elements related to them were utilized.

For the creation of the 3D model and its utilization in various applications, as mentioned above, the data were examined in detail in terms of their content (qualitative and quantitative). The results of this analysis confirmed that the same restriction pattern observed in the relevant study studied was also evident here. These limitations concern the characteristics of the data in terms of their completeness and reliability, the network topology, and the existence and/or not of the desired 3D information. No matter the application a dataset should be accompanied by a description of their content, while their compliance with one of the available standards will allow for the data further utilization. In more detail, starting with the first category, a significant shortcoming of the provided data is related to the metadata and the limited and/or missing information related to data attributes. The lack of a textual description of the data makes more difficult to understand their content, while the limited information leads to either the production of incomplete results or to hypotheses that limit the quality of the final result. When it comes to spatial data of any category, the user who is not always directly related to the object and content of the data should be able to understand their content before processing it. In all the research that has been done this description is absent in most of the available data, something that increases complexity in the data processing procedures.

Another important limitation of datasets was some missing features and/or invalid values. In all datasets, the existence of several empty records limited the applications that could be developed due to their inadequacy. Therefore, simplifications had to be made, and/or their use should be excluded. In addition, the way in which they are stored follows a customized format, which deviates significantly from any of the existing standards that support the storage and representation of relevant data. The consequence of this is the fact that the produced result followed this form and therefore their mapping to a standard is not immediately possible. Finally, regarding the stored information in terms of the topology of the networks, a considerable amount of inconsistencies were detected. These discrepancies cost the network connectivity as well as the creation and representation of interdependence relationships between the network elements. The reconstruction of topology in this thesis was achieved up to a certain extent, after the personal discussions with the data provider as well as to a set of assumptions.

As far as the availability of the 3D information is concerned, it was negligible. Although the data were considered to be in the third dimension of the total only a few records included the relevant information. This information, however, could not be considered reliable as information relevant to the reference surface from which it is measured was missing and in combination with the quality of data as a whole was not considered reliable. The absence of three-dimensional information combined with the poor quality topology of the available networks resulted in the reconfiguration and redefinition of the research questions. From the original idea of simulating the sewage underground utility network and representing the flow of the product inside the pipes, the focus was on creating a topological model which could then be enriched with the third dimension. An important observation that emerged after the overall processing of the data is the need to improve them in terms of their topology and their 2D information. Indeed, three-dimensional information is important both for the better visualization and perception of the depicted object, as well as for the development of applications that require 3D knowledge (e.g. simulation of pipe inclination). In the present work, it was deemed necessary to improve and reconstruct the existing topology as without it it would not be possible to develop any application. Given the lack of the third dimension, it was chosen to develop applications that are not directly affected by its existence. The algorithms that are created are efficient for any 2D network, provided that there is a correct (topologically) network.

The significance of the 3D information can be considered application-based. In more detail, depending on the application and the respective needs, the existence or not of the third dimension is a critical point. As mentioned above, in the case of creating a 3D model of underground networks, the knowledge of the relevant information is of great necessity. However, taking into account the nature of the networks, of the same importance appear to be the existence of a coherent topology, as without the latter the first cannot be created. For example in the current thesis, where the initial goal was to create a simulation of the flow direction through the pipes, knowledge of the three-dimensional information was crucial. However, the state of the data reveals more significant problems than the absence of the third dimension, which concerned the creation of a correct and coherent topology between the networks and their components. On the other hand, in an application such as 3D simulation of individual city objects, knowledge of the third dimension plays a key role. Therefore, the significance of the third dimension is defined by the application and the respective needs.

6.3 FUTURE WORK

Considering that experiments with real data are a time-consuming procedure, in terms of the time required to understand their content, structure, internal relationships, and interdependencies with other datasets, various adaptations, tests, and experiments have been postponed for future work. Future research should focus on a more thorough examination of the complexity of utility network systems and seek to harmonize and integrate existing information with available data models. To achieve this, the content of the relevant information should be enhanced in terms of quality and quantity corresponding to their attributes, geometry, and topology. This thesis focused mainly on examining the possibility of creating a three-dimensional model of underground utility networks integrated with the relevant (above-ground) city model objects. The goal was to explore its usefulness and utilization, obtaining the optimal result, in real-world applications, inspired by existing literature.

Taking into account the current outcomes as well as the limitations that affect a detailed representation of reality, attention should be paid to the development of data-driven strategies that focuses on improving data quality and data governance in an attempt to optimize the particular network performance. Considering the different types of errors/inconsistencies, these data-driven strategies should be accompanied by manual inspections since, from the data cleaning/processing procedure followed in the current thesis, it was made clear that not all challenges could be solved automatically. Once the available 2D data of the underground utility networks are repaired in terms of the quality of their attributes, then it will be possible to start the adaptation to the 3rd dimension. The recommendations below provide some possible directions for future work.

6.3.1 Data quality improvement

The available data (either public or private), are the basis for the creation of a model as well as determine the quality of the final result. Based on these starting points, and according to the difficulties that have arisen and the obtained outcomes of the present work, the most important point is to improve the quality of their content. During the analysis of the data, in terms of their semantic information, and the more we delved into the details of the attributes, it was perceived that the provided information was incomplete and/or absent (e.g. metadata). This fact undermines the understanding of the data and their content and affects at the same time their correspondence with reality. This derives from the fact that the data did not appear to follow any standards in which would be documented certain agreements and specifications on how they should be represented, formatted, defined, structured, manipulated, and managed [EPA, 2020]. The ultimate goal of adopting a standard is to improve the quality of the data as well as to promote interoperability, sharing, and the ability to use them in various situations. More specifically, having as a reference point the data provided in this thesis, it was realized that in terms of their semantics they were of low quality, which makes their use complicated and reduces the likelihood of their reuse. Therefore, the immediate use of them for any application is not a direct process and they need to be modified in advance, which increases the cost in terms of time and thus, they become ineffective.

Considering the advantages and disadvantages of the existing data, as they appear from the literature study and from the results of the present thesis, in order to improve them, it is necessary to combine certain methods and techniques. In particular, the available data should be used firstly, and the above steps (see Chapter 3, Figure 3.1) of restoring their topology should be followed. Then, where necessary, and in case of missing records, additional field measurements should be made so that the data correspond to the current state of the networks. For example, for the improvement of data quality Ground Penetrating Radar (GPR) technology is widely used. Taking into account the nature of the underground networks, it is not possible to excavate unwisely. GPR, on the other hand, allows for the detection of man-made objects, gaps, and layers of different combination or water content in the underground, by measuring the delay and power of electromagnetic (EM) signals scattered and reflected at permittivity discontinuities. These irregularities are attributed to differences in materials or differences in material properties, allowing for their detection. This technique is quite useful for the determination of the depths of the existing underground networks. However, before moving to extra field work, the available information should be utilized, per data category. In mode detail, considering the complexity of the stored features and the attributes that characterize them, the different data types and values stored in the vector data structure must be valid. In order for existing data to be ready for use, it must, first, be valid [Warner, 2021]. This validity is determined by specific parameters (Figure [6.1]) concerning:

1. **Schema verification**: When creating data, the primary goal should be to enforce its correctness. This implies the control of their fields, both in terms of their content and their representative name, the types of attributes (e.g. string, integer, double, varchar a.o.) accompanied by their units of measurement, especially when it comes to geometric objects and the correct definition of the coordinate system. In this check, the existence of metadata is of vital importance, since it contributes to the immediate familiarization with data by providing information about its aspect and summarizing their content, while they contribute to the reduction of redundancy between systems.

2. **Data values verification**: In this case, the control becomes more specific and concerns the actual values of the fields. Before sharing the data, it is necessary to verify that their content is true, complete, and reasonable to use. For example, both in our case as well as in the relevant research publications, empty records were present, which either had the value "NULL" or another value without and/or with the same meaning (e.g. "8888"). When the user comes in, takes the data, and comes and sees such values, problems arise about how it can be used. Therefore, these values should be described (e.g. in case of their forced existence there should be a description) or completely absent (if possible).

3. **Geometry validation**: One of the most important problems addressed during the research is of a geometric and topological nature. In particular, as we use geometric entities that represent networks, their connectivity should follow their true state (as-built condition). This is something that is missing from

relevant data to a significant extent. They are, as mentioned above, inconsistencies prevented the correct topology of the networks from being achieved. Consequently, the data should be modified by the user itself before their actual use. Thus the creator of the datasets must make sure that the network topology is correct, there are no duplicates consecutive lines and/or overlapped parts and/or left-overs, and generally degenerated or corrupted geometries. The correct datasets lead to correct decision making procedures.

4. **Internal relationships between datasets**: In this case, we are referring to those data fields that are used as a way to restrict or link related data. More specifically, in this case (of the underground networks) there were connections that concern both the networks (internal connections with sub-networks) as well as with other networks elements (survey/measured points). Given the amount of data and current needs in GIS management, an effective way to store and further manage them is a relational database. The usefulness of a database management system for storing this information is determined, among other things, by the amount of data that is or will be stored, the type of use that will be made of it, and the number of users who may be involved [ITC and Observation, 2015]. Databases are critical in the case of the such models, which contain a set of various records because they allow for keeping track of all the data by querying the relational database according to user needs. By querying a relational database, the data stored in the different tables can be associated, in case they share common attributes (primary-foreign key relations). The existence of these keys allows for the investigation of dependence, which facilitates the immediate tracking down of the data and improves the performance of the data retrieval operations [Slawinska, 2021]. Additionally, if a change to one dataset takes place then the corresponding associated table of the related datasets will be updated automatically, which saves time and prevents further errors.

5. **Verification of compliance with standards**: As mentioned before, in order for the data to be usable and reusable, they have to follow and comply with a set of existing specifications. Standards enable integrity and comprehension and their adoption provides a shared understanding of data that are frequently reused, while at the same time their mapping to a standard allows for comparisons with relevant information. Finally, they also provide consistent outcomes during data retrieval maintaining the initial information stored to them. The most common standards supporting utility networks are the Open Geospatial Consortium (OGC) CityGML, INSPIRE as well as other international (ISO) standards.

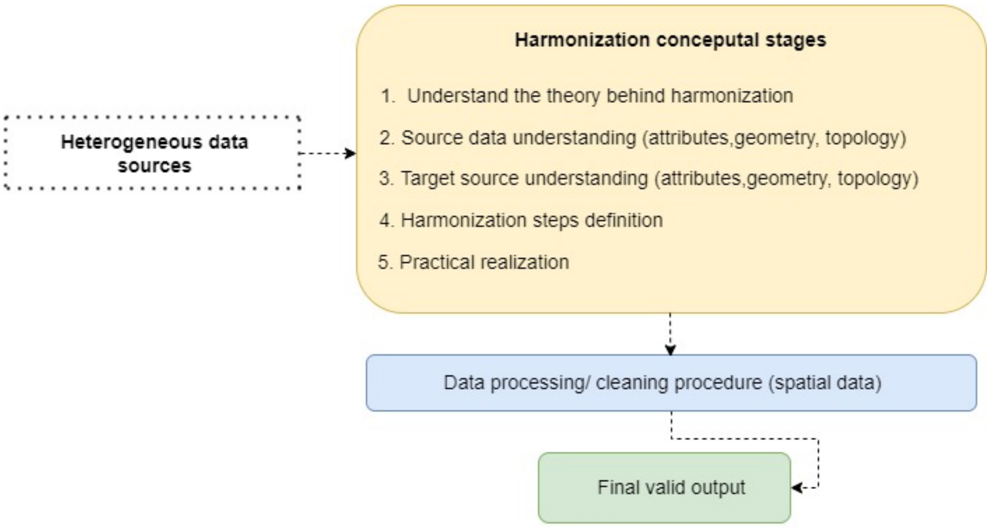


Figure 6.1: Suggested harmonization and integration methodology for spatial planning data harmonization

Inspired by: https://www.interreg-central.eu/Content.Node/CE906-BOOSTEE-CE-D.T1.1.2-Methodology-to-harmonize-and-integ.pdf

In addition to the previous verifications, what can become particularly critical for the "recycling" of available datasets is the correction as well as the reporting of incorrect and/or ineffective data. A specific example is the data used in this thesis, which can be characterized as a miniature of the current state of the majority of relevant datasets, related to underground networks, but also a wider range of data. Once the inconsistencies are identified, the next step should be to record and improve them (as much as possible, if not completely) in order to maintain, as well, a history of change. Especially in the case of utility networks, a history change is valuable, since it may concern not only the improvement of the data at a certain time but also the modification of them based on real-time changes in networks' as-built conditions. Depending on the data type and the number of inconsistencies, schema mapping should be applied to match the destination data model, while attention should be paid to the network topology reconstruction. Regardless of the degree to which datasets are poorly compatible, the user should create a report on their content, as the underground utility networks play a vital role in the proper operation of cities and the need for their digital representation use is growing exponentially.

6.3.2 Enrich data content

Underground utility networks are characterized by particular complexity. Although there are many independent networks, there are relationships of interaction or even interdependence between some of them. Therefore, the ability to differentiate them in terms of their location under the ground (depth layers), as well as the identification of the aforementioned relationships, are critical prerequisites, so for a digital model to reflect reality and to be utilized for the development of projects concerning urban environment. Based on the continuously efforts made for putting the available geographical information across to its digital representation, the existence of the 3D knowledge is invaluable. The 3rd-dimension does not correspond only to the Z-coordinate (depth value determining the location of pipes/cables) but also in the dimensions of the stored in the vector data structure objects (pipes/cables - e.g. geometry: polygon). By enriching the data with the information of these dimensions and/or the material of the cable/pipe in use, it will be derived a more detailed simulation concerning a 3D topological model accompanied by its functionality. In more detail, apart from the topological connectivity between the networks and the above-ground city objects, the ability to define the carrying capacity of the networks (based on their use) is of utmost importance. Having this information, or associated values it will be possible to calculate and evaluate the carrying capacity, and therefore the number of the city objects they can serve. This extra information adds detail to existing model and thus the more information is added the more the model is enriched with different levels of detail (LoD). In addition, it would be beneficial to have the information about the ground points where the cable/pipes are connected to (destination point of service). Considering that these points are located on the ground surface, they can be collected by doing a survey (fieldwork) and measuring their location with a GPS (Global Positioning System) accompanied by a textual description of their use, and dimensions. In this way, it would be achieved a direct connection with the datasets include the relevant point information. Finally, the information mentioned above, regarding the history change, is also necessary since time is an important dimension in geospatial visualizations and analyses. Time contributes to even more detailed models - spatio-temporal models that allows for more organized representation of space and time in a GIS system. Regarding the new data of the underground utility networks, it is noted that they should be digitized according to the proposals mentioned above, as from the research of the present research, it was found that the available data shows the same errors and inaccuracies. Again, not all attributes in a data set are necessary for all types of applications, and storing many different attributes in a custom way does not solve existing problems. New datasets should be compatible with available standards and the relevant features should be complete for further use.

6.3.3 Future applications

The way users perceive, expect, and consume spatial data in their applications is heavily influenced by the current context, projects, and market needs. Interactions among these three components serve as the foundation for geoinformation system requirements at any given time. At the moment, these interactions translate into systems that must operate in a networked environment. Based on that, a useful application for more efficient handling of such data models would be the development of a 3D WebGIS applied to utility networks. Considering the privacy of the providers of the underground utility networks, the interested user (e.g. construction company) will be able to access the online platform upon request, where he will be able to browse the available information. This partially has started to be developed in the Netherlands, using KLIC services (see Chapter 2). A 3D WebGIS would support and extend the current KLIC services and in combination with the actions required prior to excavation, it would be beneficial for any constructor (related to utility networks) to enter the web map and be able to visualize (and/or access and store) third party network datasets by area of interest. So far this service is limited to providing 2D portable document format information, which does not provide sufficient data about the current state of the underground networks. This form of information makes it difficult to be used immediately. Given the constantly evolving forms/shapes of technology, it is crucial to store and distribute spatial information to the competent bodies (e.g. Land Registry), so that a relevant user can receive the information he needs but also enrich it with the changes that are going to implement in the field. In order for the application to be practically useful, it is necessary to meet technical specifications and guides that will address interoperability issues among geo-resources [ITC and Observation, 2015]. Here it is made clear the necessity of the existence of correct, in terms of quality and precision, datasets, which contain the fundamentals for any application. Thus the attention should be paid first to the reconstruction of their content, in order to make them compatible with the available standards, and then, more advanced technological applications and models can be derived. All of the above mentioned actions will lead to the improvement and development of comprehensive 3D models that combine the above with the underground information and will contribute to the development of security and risk management, architectural planning, utility network planning and will improve decision making procedures.

6.4 REFLECTIONS

Completing my studies at the MSc. of Geomatics at the Delft University of Technology has laid the foundations for knowledge and hands-on experience in obtaining, analyzing, modeling, managing, storing, and visualizing existing spatial data. Throughout the current thesis and depending on the stage of the methodology developed, the specialized knowledge gained from the various courses proved to be particularly useful. It became clear that the creation/simulation of a model that represents reality requires a combination of various datasets, the appropriate tools/software as well as the related technical knowledge in order to properly merge the available information and finally produce -depending on the application- the desired result. Apart from that, throughout the cleaning/processing of the data it was understood the remarkable value of having a correct topology, representing reality, since based on that we can move forward to data integration and addition of the third dimensions. Although the main idea of the thesis was the creation of a 3D model combining the below with above-ground information, it shown up other difficulties that bereft of the integration of the existing data. On the contrary, I came up with a more complex problem concerning the topology of the underground utility networks, the way it was stored in the vector data structure and how it was possible to reconstruct it in order to create usable data. Throughout this process and during the implementation of all the different phases of the current thesis, was gained experience as well as it was improved my theoretical and practical knowledge of the urban and natural environment. Finally, it clearly indicated the importance of having high-quality datasets capable to be utilized immediately, avoiding cost-effective procedures for modifications.

BIBLIOGRAPHY

Adolphi, Thomas, Claus Nagel, and Thomas Kolbe

2013 "Semantic 3D Modeling of Multi-Utility Networks in Cities for Analysis and 3D Visualization," in, pp. 41-62, ISBN: 978-3-642-29792-2, DOI: 10.1007/978-3-642-29793-9_3.

Andrews, Chris

2021 *ArcGIS: A Foundation for Digital Twins*, https://www.esri.com/arcgis-blog/products/arcgis/aec/gis-foundation-for-digital-twins/, [visited on 03/14/2022].

Anthony Tasca, Virginia Vergara

2018 *pgRouting Manual*, https://docs.pgrouting.org/pdf/en/pgRoutingDocumentation-2.6.0.pdf, [visited on 04/04/2022].

Becker, Thomas, Claus Nagel, and Thomas H. Kolbe

2011 "Integrated 3D Modeling of Multi-utility Networks and Their Interdependencies for Critical Infrastructure Analysis," in.

Biscobing, Jacqueline

2019 *Entity Relationship Diagram (ERD)*, https://www.techtarget.com/searchdatamanagement/definition/entity-relationship-diagram-ERD, [visited on 03/05/2022].

Boates, I., G. Agugiaro, and A. Nichersu

2018 "NETWORK MODELLING AND SEMANTIC 3D CITY MODELS: TESTING THE MATURITY OF THE UTILITY NETWORK ADE FOR CITYGML WITH A WATER NETWORK TEST CASE," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4, pp. 13-20, DOI: 10.5194/isprs-annals-IV-4-13-2018.

Borrmann, André, Jakob Beetz, Christian Koch, Thomas Liebich, and Sergej Muhic

2018 "Industry Foundation Classes: A Standardized Data Model for the Vendor-Neutral Exchange of Digital Building Models," in.

Campbell, Christa and David Wachal

2021 *Digital Twins Bring Value to Water Utilities*, https://www.esri.com/arcgis-blog/products/arcgis/water/digital-twins-water-utilities/, [visited on 03/14/2022].

Consortium, Open Geospatial

2021 *CityGML*, https://www.ogc.org/standards/citygml, [visited on 01/29/2022].

Den Duijn X., Agugiaro G., and Zlatanova S.

2018 "MODELLING BELOW- AND ABOVE-GROUND UTILITY NETWORK FEATURES WITH THE CITYGML UTILITY NETWORK ADE: EXPERIENCES FROM ROTTERDAM," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-4/W7, pp. 43-50, DOI: 10.5194/isprs-annals-IV-4-W7-43-2018, https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-4-W7/43/2018/.

Du, Yunqiang, Sisi Zlatanova, and Xianglin Liu

2006 "Management and 3D visualisation of pipeline networks using DBMS and AEC software," 36 [Jan. 2006].

ELLINGSON

2019 *WATER MANAGEMENT 101*, https://ellingsoncompanies.com/wp-content/uploads/2019/11/Ellingson_H20nomy.pdf, [visited on 03/04/2022].

Emgard, L. and S. Zlatanova

2007 *Design of an integrated 3D information model*, http://www.gdmc.nl/publications/2007/Design_integrated_3D_information_model.pdf.

EPA, United States Environmental Protection Agency-

2020 *Data Standards*, https://www.epa.gov/data-standards/learn-about-data-standards, [visited on 03/07/2022].

ESRI

2007 *GIS Technology for Water, Wastewater, and Storm Water Utilities.Systems Research Institute, Redlands, CA. Environmental*, http://www.esri.com/library/brochures/pdfs/water-wastewater.pdf, [visited on 01/30/2022].

2017 *What are geometric networks?* https://desktop.arcgis.com/en/arcmap/10.3/manage-data/geometric-networks/what-are-geometric-networks-.htm, [visited on 01/30/2022].

2021 *Digital Twin Technology & GIS — What Is a Digital Twin?* https://www.esri.com/en-us/digital-twin/overview.

Evers, Jeannie and Emdash Editing

2017 *GIS (Geographic Information System)*, https://www.nationalgeographic.org/encyclopedia/geographic-information-system-gis/#:~:text=GIS%20can%20show%20many%20different,and%20understand%20patterns%20and%20relationships., [visited on 03/14/2022].

FME, Feature Manipulation Engine

2020 *Tutorial: Working with Topology and Network Transformers*, https://community.safe.com/s/article/working-with-networks-topology-and-networktransfor, [visited on 02/20/2022].

Fossatti, F., G. Agugiaro, L. olde Scholtenhuis, and A. Dorée

2020 "Data modeling for operation and maintenance of utility networks: Implementation and testing," English, *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 6, 4/W1 [Sept. 2020], 3rd BIM/GIS Integration Workshop and 15th 3D GeoInfo Conference ; Conference date: 07-09-2020 Through 11-09-2020, pp. 69-76, ISSN: 2194-9042, DOI: 10.5194/isprs-annals-VI-4-W1-2020-69-2020.

GEONOVUM

2020a *Information Model Cables and Pipes*, https://www.geonovum.nl/geo-standaarden/informatiemodel-kabels-en-leidingen, [visited on 11/27/2021].

2020b *NEN 3610 basic model for information models*, https://www.geonovum.nl/geo-standaarden/nen-3610-basismodel-voor-informatiemodellen, [visited on 11/27/2021].

Hartmann, Dirk and Herman Van der Auweraer

2021 "Digital Twins," in *Progress in Industrial Mathematics: Success Stories*, ed. by Manuel Cruz, Carlos Parés, and Peregrina Quintela, Springer International Publishing, Cham, pp. 3-17, ISBN: 978-3-030-61844-5.

Herman van den Berg, Land Registry and Geonovum Paul Janssen

2021 *Information Model Cables and Pipes (IMKL) - version 2.0*, https://geonovum.github.io/imkl2015-review/IMKL2.x/1-dataspecificatie/#uml-diagrammen, [visited on 12/02/2021].

Hijazi and Ihab

2017 "Use cases and their requirements on the semantic modeling of 3d supply and disposal networks," in.

ITC, Faculty of Geo-Information Science and Earth Observation

2015 *ITC Corebook 2013 - Chapter 08 - Spatial data modelling*, https://issuu.com/itc-utwente/docs/corebook2013_08_spatialdatamodellin, [visited on 04/28/2022].

J., Lieberman and Carsten Roensdorf

2020 "MODULAR APPROACH TO 3D REPRESENTATION OF UNDERGROUND INFRASTRUCTURE IN THE MODEL FOR UNDERGROUND DATA DEFINITION AND INTEGRATION (MUDDI)," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIV-4/W1-2020 [Sept. 2020], pp. 75-81, DOI: 10.5194/isprs-archives-XLIV-4-W1-2020-75-2020.

JavaTPoint

2021 *Types of Graph in Data Structure*, https://www.javatpoint.com/types-of-graph-in-data-structure, [visited on 04/07/2022].

JOAQUIN MADRID, KONSTANTIN KORCHAKOV

2017 *Esri Utility Network – Simple and Scalable Network Data Model*, https://sspinnovations.com/blog/esri-utility-network-simple-and-scalable-network-data-model/, [visited on 01/30/2022].

Kocaman, Sultan, Devrim Akca, Daniela Poli, and Fabio Remondino

2022 *3D/4D CITY MODELLING FROM SENSORS TO APPLICATIONS*, WHITTLES PUBLISHING.

Koppens, Drs. Bart

2014 *Abstract KLIC the Dutch one call system*, https://www.concawe.eu/wp-content/uploads/2017/01/s3-4_summary_klic_the_dutch_one_call_system-2014-01257-01-e.pdf, [visited on 01/31/2022].

Kutzner, Dr. Tatjana

2019 *CityGML Utility Network ADE*, https://www.asg.ed.tum.de/en/gis/projects/citygml-utility-network-ade/, [visited on 01/30/2022].

Lacey, A. and D. Luff

2007 *Qualitative Research Analysis*, The NIHR RDS for the East Midlands/ Yorkshire & the Humber, [visited on 11/28/2021].

Lieberman, Josh

2019 *Model for Underground Data Definition and Integration (MUDDI) Engineering Report*, http://www.opengis.net/doc/PER/MUDDI-ER, [visited on 11/27/2021].

Lloyd, C.

2010 *Spatial Data Analysis: An Introduction for GIS Users*, Knovel Library, OUP Oxford, ISBN: 9780199554324, https://books.google.nl/books?id=7WGcAQAAQBAJ.

Mieghem, Piet Van, Christian Doerr, H. Wang, J. Martin Hernandez, David Hutchison, Merkourios Karaliopoulos, and Robert E. Kooij

2010 "A Framework for Computing Topological Network Robustness," in.

Munikoti, Sai, Kexing Lai, and Balasubramaniam Natarajan

2021 "Robustness assessment of Hetero-functional graph theory based model of interdependent urban utility networks," *Reliability Engineering & System Safety*, 212, p. 107627, ISSN: 0951-8320, DOI: https://doi.org/10.1016/j.ress.2021.107627, https://www.sciencedirect.com/science/article/pii/S095183202100168X.

Obe, Regina O., Leo S. Hsu, and Gary E. Sherman

2017 *PgRouting: A Practical Guide*, Locate Press, Chugiak, AK, ISBN: 0989421732.

olde Scholtenhuis, Léon L., Xander den Duijn, and Sisi Zlatanova

2018 "Representing geographical uncertainties of utility location data in 3D," *Automation in Construction*, 96, pp. 483-493, ISSN: 0926-5805, DOI: https://doi.org/10.1016/j.autcon.2018.09.012, https://www.sciencedirect.com/science/article/pii/S0926580518301997.

Pedia, Trenchless

2021 *Underground Utilities*, https://www.trenchlesspedia.com/definition/3702/underground-utilities-excavation#:~:text=Underground%20utilities%20are%20any%20public,and%20gas%20and%20oil%20pipelines., [visited on 03/15/2022].

Rawat, Soumyaa

2021 *Introduction to Network Graph and Network Topology*, https://www.analyticssteps.com/blogs/introduction-network-graph-and-network-topology, [visited on 03/05/2022].

Singh, Puyam, Rosly Lyngdoh, Dibyajyoti Chutia, Victor Saikhom, Bhargav Kashyap, and Sunkari Sudhakar

2015 "Dynamic shortest route finder using pgRouting for emergency management," *Applied Geomatics*, 7 [July 2015], DOI: 10.1007/s12518-015-0161-4.

Slawinska, Martyna

2021 *Why Do Relational Databases Use Primary Keys and Foreign Keys?* https://learnsql.com/blog/why-use-primary-key-foreign-key/, [visited on 03/07/2022].

Tabarro, P. G., J. Pouliot, R. Fortier, and L.-M. Losier

2017 "A WEBGIS TO SUPPORT GPR 3D DATA ACQUISITION: A FIRST STEP FOR THE INTEGRATION OF UNDERGROUND UTILITY NETWORKS IN 3D CITY MODELS," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W7, pp. 43-48, DOI: 10.5194/isprs-archives-XLII-4-W7-43-2017.

TC, Technical Committee ISO/TC 59/SC 13

2018 *Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries — Part 1: Data schema*, https://www.iso.org/standard/70303.html, [visited on 01/30/2022].

Testing QGIS, Documentation

2022 *FILL NODATA*, https://docs.qgis.org/2.8/en/docs/user_manual/processing_algs/gdalogr/gdal_analysis/fillnodata.html, [visited on 04/04/2022].

Toktoshov, G. Y., Anastasiya N. Yurgenson, and Denis A. Migov

2018 "On a Problem of the Utility Network Design ?" In.

UNDRR

2016 *United Nations Office for Disaster Risk Reduction*, https://www.undrr.org/#letter-p, [visited on 04/04/2022].

Utility, INSPIRE Thematic Working Group and governmental services

n.d. *D2.8.III.6 INSPIRE Data Specification on Utility and governmental services – Draft Guidelines.*

Van Oosterom, Petrus Johannes Maria

2008 *Advances in 3D geoinformation systems*, http://site.ebrary.com/id/10266979.

VoXco

2020 *What is Quantitative Research?* https://www.voxco.com/blog/quantitative-research/, [visited on 11/28/2021].

Wang, Mingzhu, Yichuan Deng, Jongsung Won, and Jack C.P. Cheng

2019 "An integrated underground utility management and decision support based on BIM and GIS," *Automation in Construction*, 107, p. 102931, ISSN: 0926-5805, DOI: https://doi.org/10.1016/j.autcon.2019.102931, https://www.sciencedirect.com/science/article/pii/S0926580518312457.

Warner, Tiana

2021 *The Ultimate Geospatial Data Validation Checklist*, https://www.safe.com/blog/2014/11/data-quality-checklist/, [visited on 03/07/2022].

Widl, Edmund, Giorgio Agugiaro, and Jan Peters-Anders

2021 "Linking Semantic 3D City Models with Domain-Specific Simulation Tools for the Planning and Validation of Energy Applications at District Level," *Sustainability*, 13 [Aug. 2021], p. 8782, DOI: 10.3390/su13168782.

Yan, Jingya, Siow Wei Jaw, Rob van Son, Kean Soon, and Gerhard Schrotter

2018 "Three-Dimensional Data Modelling for Underground Utility Network Mapping," XLII-4 [Sept. 2018], pp. 711-715, DOI: 10.5194/isprs-archives-XLII-4-711-2018.

Zeiss, Geoff

2014 *Map of underground utilities available within one day of a one-call information request*, https://geospatial.blogs.com/geospatial/2019/12/map-of-underground-utilities-available-within-one-day-of-a-one-call-information-request.html, [visited on 01/31/2022].

7 APPENDICES

7.1 FME WORKSPACE

7.1.1 Example of the FME workspace for data processing

In the Figures below are presented some examples related to the data processing phase of the thesis. In the first one are presented some of the *FME transformers* used for the reconstruction of topology, while in the second one is presented the approach followed to create the virtual edges between the entities of interest.

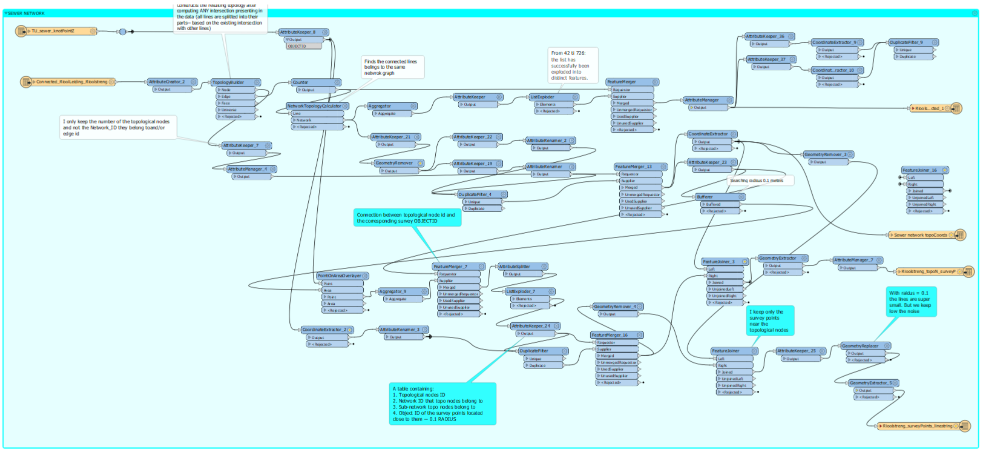


Figure 7.1: FME workspace example- sewer network processing

7.1.2 Example of the FME workspace for virtual edges creation

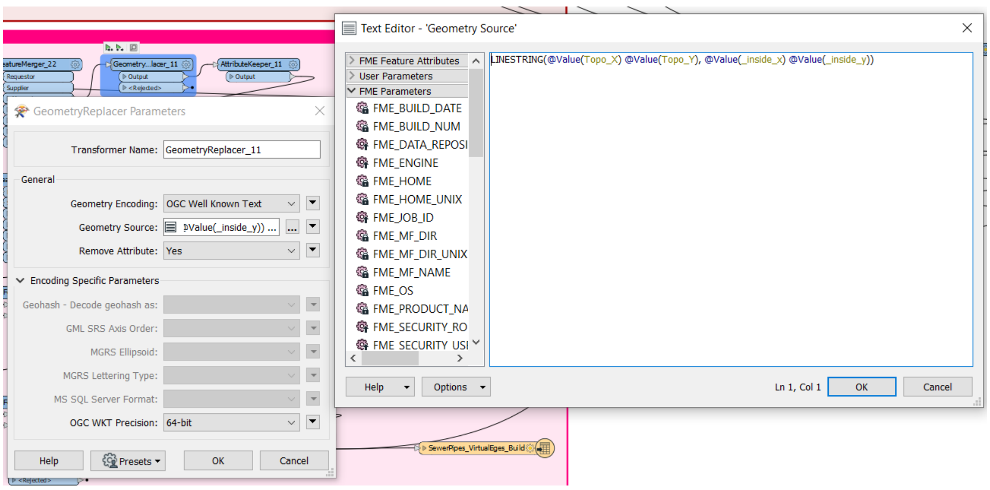


Figure 7.2: FME workspace example- virtual edges creation

7.2 CASE STUDY: SCENARIO 1

An example of the queries used in order to detect the pipeline affected by a possible damage -building-centroid approach. The tables $'Network_noded','Network'$ correspond to the modified according to pg_routing rules underground utility networks that derived from the latter table (Network), where all the available sewage networks are stored. The tests were examined using the main sewage network and the associated with it plot connection pipes.

```
...
SELECT net16.building_id, net16.network_id, route.*,
edge_geometry.geom
FROM Network_noded AS  edge_geometry

JOIN (
    SELECT * FROM pgr_dijkstra('SELECT id, source,
             target, st_length(geom) AS COST
                    FROM Network_noded',508, array[738,1260,304],
false)
             ) AS  route
ON edge_geometry.id = route.edge
INNER JOIN Network AS   net16
ON net16.id = edge_geometry.old_id
...
```

An example of the queries used in order to detect the pipeline affected by a possible damage -addresses approach. It is noted that the table *'duplicateadr'* stores all the overlapping points correspond to the different addresses placed at the same location in a building polygon.

```
...
DROP VIEW IF EXISTS CASE_1

CREATE VIEW CASE_1 AS  (
SELECT DISTINCT(topo.addressn), topo.housen, topo.pandid, topo.geom,
 topo.network_id, topo.node_start, topo.node_end, edge_geometry.id
FROM (Network_node AS  edge_geometry

JOIN (
    SELECT * FROM pgr_dijkstra('SELECT id, source,
                          target, st_length(geom) AS COST
                          FROM Network_node',753, 878 ,false)
             ) AS  route
ON edge_geometry.id = route.edge
JOIN Network AS  topo ON topo.id = edge_geometry.old_id))


SELECT (COUNT(DISTINCT(dup.nummeraand)), dup.nummeraand, dup.
networkid, dup.nodeid, dup.pandid
FROM duplicateadr AS  dup LEFT JOIN CASE_1 ON CASE_1.pandid = dup.
pandid
WHERE EXISTS (
```

```
    SELECT CASE_1.addressn from CASE_1
    WHERE CASE_1.pandid = dup.pandid AND (CAS E_1.node_end = dup.
nodeid or CASE_1.node_start = dup.nodeid) )
    GROUP BY dup.nummeraand, dup.networkid, dup.nodeid,dup.pandid

...
```

7.3 CASE STUDY: SCENARIO 2

An example of the queries used to detect a cost effective route.

```
...
UPDATE Network_node
SET modified\_length = st_length(geom)


UPDATE Network_node
SET modified\_length = 1000000
where id = '508';

...
```

```
...
SELECT net16.building_id, net16.network_id, route.*, edge_geometry.
geom
FROM Network_noded AS  edge_geometry

JOIN (
    SELECT * FROM pgr_dijkstra('SELECT id, source, target,
     modified_length AS  COST
                                    FROM Network_noded',420, 432,
false)
             ) AS  route
ON edge_geometry.id = route.edge
INNER JOIN Network AS  net16
ON net16.id = edge_geometry.old_id    ...
```

7.3.1 Completed table for cost-effective route including geometry column

Table 7.1: Cost-effective route detection including geometries and building's service

Building ID	Network ID	seq	path_seq	Node (source)	Edge ID	Cost	Aggregated cost	Geometry
None	1	1	1	321	1460	9.473446425792797	0.0	0102000020407100000200000004E315E2319C2F44031BE102971401B4100567DAE3DC2F440C00968624C401B41
None	1	2	2	467	469	38.28246923214117	9.473446425792797	01020000204071000002000000000567DAE3DC2F440C00968624C401B41C409508922C3F4407CB92C58BE3F1B41
None	1	3	3	468	470	11.203377038251343	47.75591565793397	0102000020407100000200000C409508922C3F4407CB92C58BE3F1B418BF2C38265C3F440797CC1C6943F1B41
None	1	4	4	469	917	4.495574098592385	58.959292696185315	010200002040710000020000008BF2C38265C3F440797CC1C6943F1B4100AEB66280C3F44080ABAD18843F1B41
None	1	5	5	495	501	2.200004634384923	63.4548667947777	01020000204071000002000000000AEB66280C3F44080ABAD18843F1B419F371FCC8DC3F440211FCEF57B3F1B41
None	1	6	6	496	502	17.285455657800174	65.65487142916263	010200002040710000020000009F371FCC8DC3F440211FCEF57B3F1B417AFE702CF7C3F440D6E4A6083C3F1B41
None	1	7	7	497	503	22.002865106390114	82.9403270869628	010200002040710000020000007AFE702CF7C3F440D6E4A6083C3F1B41CC87EC4E7DC4F440043A40A9EA3E1B41
None	1	8	8	145	504	20.49926845392435	104.94319219335291	0102000020407100000200000CC87EC4E7DC4F440043A40A9EA3E1B410B90D646FAC4F4400F9264D99E3E1B41
None	1	9	9	226	505	21.49996330294493	125.44246064727727	0102000020407100000200000000B90D646FAC4F4400F9264D99E3E1B41003E79587DC5F440804F1E564F3E1B41
None	1	10	10	396	360	25.199975697894235	146.9424239502222	010200002040710000020000000020177B587DC5F440C12E1D564F3E1B41A5AD542F16C6F440F88D280FF23D1B41
None	1	11	11	397	361	12.455136142233453	172.14239964811645	0102000020407100000200000A5AD542F16C6F440F88D280FF23D1B41A2E3D9B961C6F440CE61F0F4C33D1B41
None	1	12	12	46	362	6.210597319166775	184.5975357903499	0102000020407100000200000A2E3D9B961C6F440CE61F0F4C33D1B4100F8C26487C6F4400010E9F7AC3D1B41
None	1	13	13	398	1000	1.152473290549216	190.80813310951666	01020000204071000002000000000F8C26487C6F4400010E9F7AC3D1B412DAC68A88EC6F4403E5636BBA83D1B41
None	1	14	14	339	280	4.1509460479223605	191.96060640006587	010200002040710000020000002DAC68A88EC6F4403E5636BBA83D1B41B4356FD2A8C6F4405B776078993D1B41
None	1	15	15	340	281	15.163708280669722	196.11155244798823	0102000020407100000200000B4356FD2A8C6F4405B776078993D1B410042CF6608C7F44080226CB8613D1B41
None	1	16	16	341	1131	2.9708326457109977	211.27526072865794	01020000204071000002000000000042CF6608C7F44080226CB8613D1B41B9DA181D1BC7F4406209EDCB563D1B41
None	1	17	17	498	1419	1.6557047635222704	214.24609337436894	0102000020407100000200000009A081B35C7F44040F38E13583D1B41B9DA181D1BC7F4406209EDCB563D1B41
None	1	18	18	823	1230	65.04225868638667	215.9017981378912	0102000020407100000200000009A081B35C7F44040F38E13583D1B4133285D0FC7C8F4402FD4E818683C1B41
None	1	19	19	678	654	1.3976908906136516	280.9440568242779	01020000204071000002000000033285D0FC7C8F4402FD4E818683C1B4100EA95B2CFC8F440000EBEF0623C1B41
None	1	20	20	676	653	12.925083249050816	282.34174771489154	0102000020407100000200000000EA95B2CFC8F440000EBEF0623C1B41008AB0E112C9F44080CC5D0B323C1B41
None	1	21	21	677	700	20.429629588716164	295.26683096394237	01020000204071000002000000008AB0E112C9F44080CC5D0B323C1B41F3975A4A98C9F440B58A1A71E73B1B41
None	1	22	22	735	1465	1.7067720664359822	315.69646055265855	0102000020407100000200000F3975A4A98C9F440B58A1A71E73B1B41008104C57EC9F44080A52C03E53B1B41
None	1	23	23	880	824	15.430298636132958	317.4032326190945	010200002040710000020000000036CD3B21C9F44080F7E4211E3C1B41008104C57EC9F44080A52C03E53B1B41
None	1	24	24	542	554	3.7690604911801375	332.83353125522746	01020000204071000002000000008E06F0F9C8F44080A98291293C1B410036CD3B21C9F44080F7E4211E3C1B41
None	1	25	25	541	1277	10.09801312344755	336.6025917464076	0102000020407100000200000000D712F2B8C8F44040F1638C4E3C1B41008E06F0F9C8F44080A98291293C1B41
None	1	26	26	392	1372	0.4054910749318553	346.70060486985517	0102000020407100000200000000D712F2B8C8F44040F1638C4E3C1B41912683F8B2C8F4401E7A90EA4D3C1B41
None	1	27	27	352	300	4.962866430633872	347.106095944787	0102000020407100000200000026F2FF7497C8F440B885B8889603C1B41912683F8B2C8F4401E7A90EA4D3C1B41
None	6	28	28	282	225	3.4434513085424556	352.0689623754209	0102000020407100000200000025F2FF7497C8F440B885B8889603C1B4100E10B9360C8F44000492EBF613C1B41
None	6	29	29	283	870	17.61904671754125	355.51241368396336	0102000020407100000200000000E10B9360C8F44000492EBF613C1B41B820BBEF9CC7F440A0C137012F3C1B41
NL.IMBAG. Pand.0503100000031391	6	30	30	239	912	3.471252839117845	373.1314604015046	0102000020407100000300000076711B6AC7F440C032C471293C1B410010E9B78EC7F440401CEB222D3C1B41B820BBEF9CC7F440A0C137012F3C1B41

COLOPHON

This document was typeset using LaTeX. The document layout was generated using the `arsclassica` package by Lorenzo Pantieri, which is an adaption of the original `classicthesis` package from André Miede.