MSc thesis in Geomatics

Spatial and semantic enrichment of utility networks data

Louis Dechamps

June 2024

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

The work in this thesis was carried out in the:

TU Delft 3Dgeoinfo

3D geoinformation group
Delft University of Technology

Supervisors:	Dr. Giorgio Agugiaro
	Prof.dr. Jantien E. Stoter
	Dr.ir. Léon L. olde Scholtenhuis
Co-reader:	Dr.ir. Bastiaan van Loenen

Abstract

Utility networks are critical components of urban infrastructure, providing essential services such as water supply, electricity, gas, and telecommunications. The traditional method for mapping these networks is typically two-dimensional (2D) schematic representations rather than topographically and geometrically correct maps. These representations lack the capacity to convey the complexity and vertical intricacies of urban infrastructures. This limitation hampers comprehensive planning, efficient management, and risk mitigation during construction activities because 2D maps do not effectively represent the multi-layered and interconnected nature of urban utilities, leading to potential oversights and inaccuracies. This thesis addresses the challenge of enriching utility network data by integrating detailed data from utility trench surveys. These surveys provide precise positional and attribute information about utilities that are often missing in standard maps, such as the exact depth, spatial configuration, and physical characteristics of the utility lines.

Data from utility trenches in three Dutch cities—Enschede, Rotterdam, and Amsterdam—was acquired and analyzed. Methodologies were developed to extract, standardize, and integrate this data into existing utility network maps, enhancing their semantic content and spatial accuracy. The research demonstrated that integrating trench data can reveal inaccuracies in traditional utility network maps at the utility trench locations.

Key findings include the development of algorithms for extracting and processing utility trench data, the identification of common challenges between cities such as cable/pipeline labeling inconsistencies, and the comparison of enriched utility data and that of traditional utility networks. The research also highlights the importance of standardizing data models and the potential of three-dimensional models to provide a more comprehensive understanding of utility networks. A resulting recommendation was to improve data collection by including all information found and providing properly geo-referenced data.

Acknowledgements

I would like to express my gratitude to everyone who has helped and guided me throughout the course of this research. I am extremely thankful for the support and guidance provided by my supervisors. Dr. Giorgio Agugiaro has been an exceptional mentor, consistently providing reliable insights and pushing me to achieve my best. Similarly, Dr. Jantien Stoter and Dr. Léon olde Scholtenhuis have greatly helped by facilitating the data acquisition and by providing thought-provoking questions and remarks that have significantly aided my process and product. I would also like to thank Dr. Bastiaan van Loenen for his in-depth review and remarks that helped me refine my work in the latter stages of the thesis.

Contents

List of Figures

List of Tables

List of Algorithms

Acronyms

2D two-dimensional . . . 1
3D three-dimensional . . . 1
CAD Computer Aided Design . . . 16
CRS Coordinate Reference System . . . 43
GIS Geographical Information System . . . 3
INSPIRE Infrastructure for Spatial Information in Europe . . . 6
KLIC Kabel- en Leiding Informatie Portaal . . . xiii
LVZK Leidingen Verzamelkaart . . . xiv
NAP Normaal Amsterdams Peil . . . 21

1. Introduction

In this chapter, the basic concepts and issues regarding utility network maps are introduced. Subsequent chapters will dive into related work (Chapter 2) that frames the current technology and methodology. Chapter 3, outlines the research methodology employed for data enrichment, and Chapter 4 discusses the implementation of these techniques. The results (Chapter 5) will be evaluated to determine the efficacy of the proposed methods, followed by a discussion (Chapter 6) that synthesises the findings and their implications for urban infrastructure management. Finally, the conclusion (Chapter 7) will summarise the research contributions and suggest future avenues for work.

1.1. Background

Utility networks are fundamental components of urban infrastructure, critical to the functioning and development of cities worldwide. These networks, including essential services such as water supply, electricity, gas, and telecommunications, not only support daily activities of urban populations but also support economic activities and ensure the welfare of communities. The role of these networks in urban planning and infrastructure development cannot be overstated, as they influence decisions related to the expansion, maintenance, and sustainability of urban environments.

The traditional methods used for mapping and representing these utility networks employ two-dimensional (2D) schemes. These 2D representations provide a simplified overview of the networks, mapping out the routes and connections of various utilities at a basic level. However, they significantly lack the ability to convey the complex, multi-layered nature of urban infrastructures, especially the vertical aspects. [McGuffin et al., 2022]

This limitation poses substantial challenges in comprehensive planning and effective management of utility networks [Chundeli, 2017]. For instance, without accurate three-dimensional data, urban planners and engineers may struggle to assess the spatial conflicts and integration possibilities between new developments and existing utilities [Chundeli, 2017]. This often results in inefficient use of space, increased costs, and potential hazards during construction phases due to unintended disruptions of utility lines.

Given these challenges, there is a pressing need for a shift towards more dynamic and detailed representations that can enhance the planning, implementation, and maintenance of utility networks. This shift involves not only moving from 2D to three-dimensional (3D) models but also integrating richer semantic information that provides a more holistic view of the urban infrastructure [Becker et al., 2011]. This would facilitate more informed decision-making, better risk management, and ultimately, the creation of more resilient and sustainable urban environments [Chundeli, 2017].

1.2. Problem Statement

The core challenge addressed in this research revolves around the limitations in data of current utility networks, particularly the lack of detailed semantic information that undermines comprehensive urban planning and infrastructure management.

One of the main challenges is the lack of or incomplete data in utility networks, both in terms of geometry (such as exact locations and dimensions) and semantic aspects (such as type of material, capacity, or condition). These gaps not only hinder the accuracy of utility representations but also pose risks during urban development projects, where unexpected encounters with incorrectly mapped utilities can lead to costly delays and damages [Döner et al., 2010].

Despite these challenges, some of this valuable data is available in the form of a utility trench survey, which capture precise information about utility attributes and positions. However, this surveyed data is rarely integrated into broader utility network datasets, resulting in potential missed opportunities for enhancing utility network data.

To address these issues this thesis proposes to develop methodologies to integrate data from utility trenches with existing 2D utility network maps. This integration aims to enrich both the semantic and geometric content of the existing utility network maps. The semantic enrichment will consist of incorporating information such as the attributes and operational data of utility components, which can be neglected in standard 2D representations [Becker et al., 2011]. Additionally, the 3D information provided by the utility trenches, in the from of individual cable depths, will be extracted and added to the existing information. This is crucial for understanding the vertical distribution of underground utilities. This will be valuable for planning in densely populated or geographically constrained areas [Döner et al., 2010].

By addressing these challenges, the research aims to enhance the quality of utility network maps, making them more accurate, informative, and useful for urban planners and engineers. This enhancement will facilitate better planning, reduce the risk of utility damage during construction, and contribute to more sustainable urban development practices.

1.3. Utility Trenches

Utility trenches (see Figure 1.1), known as 'proefsleuven' in Dutch, are investigative excavation techniques commonly used in the fields of archaeology, geology, and civil engineering, particularly for the preliminary exploration of underground utilities and other buried structures. This method involves digging trenches at strategic locations to directly observe the subsurface conditions. The primary purpose of these trenches is to gather accurate data regarding the presence, position, and condition of subterranean features, which is critical for a variety of construction and planning processes.

In the context of utility networks, utility trenches are employed to verify the accuracy of existing records, such as those found in utility network maps. This verification is essential because historical data can often be incomplete or inaccurate due to changes over time or initial data collection errors. Through utility trenching, engineers and planners can obtain real-world evidence of the network's layout, which includes the precise location, depth, and type of utilities present. As current 2D maps are often old and their reliability may

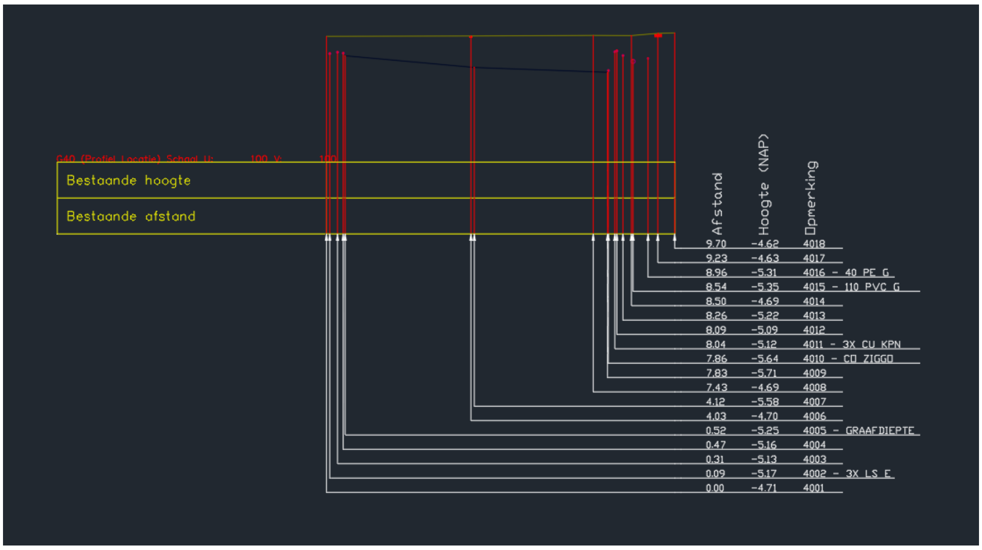


Figure 1.1.: Proefsleuf provided by the municipality of Rotterdam

be reduced, this surveyed information could prove valuable for updating Geographical Information System (GIS) maps in terms of accuracy, semantics and 3D information, and for ensuring that planning decisions are based on the most current and comprehensive data available.

Moreover, the data collected from utility trenches can provide detailed semantic information. Attributes of the utilities such as material type, size, and condition, which are often missing from older or more basic mapping systems. The absence of such detailed semantic information can limit the utility networks' ability to function beyond basic visualization. If these details were recorded, several significant benefits would be realised, such as, improved maintenance planning, enhanced risk management and informed decision-making.

However, adding enriched information to existing 2D utility network models presents several scientific challenges. Firstly, integrating detailed semantic and geometric data requires advanced techniques to accurately map and align the new information with existing models. This process is complicated by the varying levels of detail and accuracy in current datasets. Additionally, older models may lack the necessary frameworks to accommodate enhanced attributes, necessitating significant modifications to both the data structures and the visualization algorithms. These complexities highlight the need for innovative methods and tools to seamlessly enrich 2D models with comprehensive, precise utility trench data. [Vishnu and Sameer, 2020]

1.4. Research Question

Building upon the understanding of utility trenches (proefsleuven) and their importance in collecting utility network data, this thesis research will address the questions listed hereafter. The overarching goal is to explore the feasibility and the depth to which 2D GIS models of utility networks can be enhanced both semantically and geometrically into three dimensions, utilizing the data derived from utility trenches. Therefore, the main research question can be described as:

To what extent can existing 2D utility network models in the Netherlands be enriched using surveyed utility trench data?

Subsequent questions may be defined as:

- To what extent can the information be integrated either semantically and/or geometrically?
- To what extent can a common methodology be implemented for the 3 different cities?
- Which strategies can be developed to automate the data extraction and integration process?
- How far can current standards help reduce issues relating to data integration?

2. Related work

2.1. Theoretical Background

Underground utility networks play a crucial role in urban infrastructure, impacting city management service delivery, and urban planning. With the growing development of underground infrastructure, there is a need for monitoring and mapping these infrastructures for better usage and planning [Jaw et al., 2018]. Currently, the existing utility data describing pipe distributions, types, and dimensions is not well recorded and managed [Ward et al., 2017]. Additionally, since different city government agencies often oversee various utilities, there is frequently a lack of coordination between these utility administrations [Fossatti et al., 2020]. This is described by Fossatti et al. [2020] as horizontal or inter-disciplinary fragmentation, which arises because utility networks that transport various commodities are owned by different entities, and their construction involves the collaborative efforts of multiple trades, such as design, piping, surveying, and systems, among others. Additionally, Fossatti et al. [2020] describes what they deem longitudinal fragmentation, where ”[...] knowledge about a utility network and its components does not flow seamlessly through different life phases and stakeholders that manage the network”.

To combat these issues, developing and enforcing standards for modeling utility networks is necessary. These standards aim to facilitate the use, reuse, and sharing of utility networks, promoting interoperability. Implementing unified standards across different utility networks promotes seamless communication and coordination among various service providers. It also facilitates the integration of data and systems, which is critical for efficient management and operation. Moreover, standardised models help in reducing the complexity and cost of infrastructure development by providing clear guidelines for design, construction, and maintenance.

While this research does not directly focus on developing new standards, it acknowledges their importance and strives to align the collected data from different sources with one another to ensure consistency and interoperability. By organizing the data in a standardised format, this study attempts to facilitate future processes and contribute to better data integration and management practices.

The related work section of this thesis explores relevant studies and developments related to the research topic. This section will discuss existing data models and relevant case-studies in the field of utility networks.

2.2. Relevant Data Models

2.2.1. CityGML Utility Network ADE

The CityGML Utility Network Application Domain Extension (ADE) is an extension of CityGML, an open data model and XML-based format for 3D city models, designed to enhance the modeling and interoperability of utility networks such as water, gas, electricity, and telecommunications. This ADE allows for detailed 3D representations of utility networks, including their physical components, properties, and functional relationships. This capability addresses the limitations of 2D representations found in other standards, which often fail to capture the depth, spatial relationships, and complexity of urban utility networks. By providing a standardised way to model these elements in 3D, the Utility Network ADE facilitates the integration and analysis of utility networks within broader urban infrastructure models. Its alignment with CityGML ensures compatibility with other CityGML-based models, enhancing data exchange and collaboration across various domains. This interoperability is essential for projects focused on utility networks and standards, as it ensures better data accuracy and comprehensive urban management [Becker et al., 2011].

2.2.2. INSPIRE network

The Infrastructure for Spatial Information in Europe (INSPIRE) initiative plays an important role in standardizing and enhancing the accessibility of spatial data across Europe. It aims to create a unified framework that facilitates the sharing and interoperability of spatial information for various applications [Lieberman, 2019]. The relevance of INSPIRE to the this topic lies in its establishment of consistent data models and standards that ensure spatial data can be easily integrated, shared and utilised.

INSPIRE has set specific themes, some of which include critical information on administrative services, and are designed to follow interoperability rules, making it easier to integrate diverse datasets. The standardised data models provided by INSPIRE, based on UML, support the detailed representation and integration of utility networks into existing infrastructure. This includes providing guidelines for the consistent recording of spatial relationships and attributes of utility elements [Lieberman, 2019].

However, such models face several unresolved challenges. One significant issue is that utility networks are often treated as isolated entities, failing to consider their interactions with other networks. This leads to an incomplete understanding of the relationships between underground utilities. Additionally, achieving interoperability remains a major concern, particularly regarding the use of standards for legal references, contact information, and the harmonization of temporal data [Pavlidou, 2022]. Additionally, one of the greater issues is the ability to integrate the utility networks in 3D is lacking. As of right now, 2D relationships between networks are supported but not 3D.

2.2.3. Kabel- en Leiding Informatie Portaal (KLIC)

KLIC, is a system in the Netherlands originally designed to prevent excavation damage to underground utilities. Established in 1967 following a series of damaging incidents, KLIC facilitates the exchange of information regarding the location of underground cables and

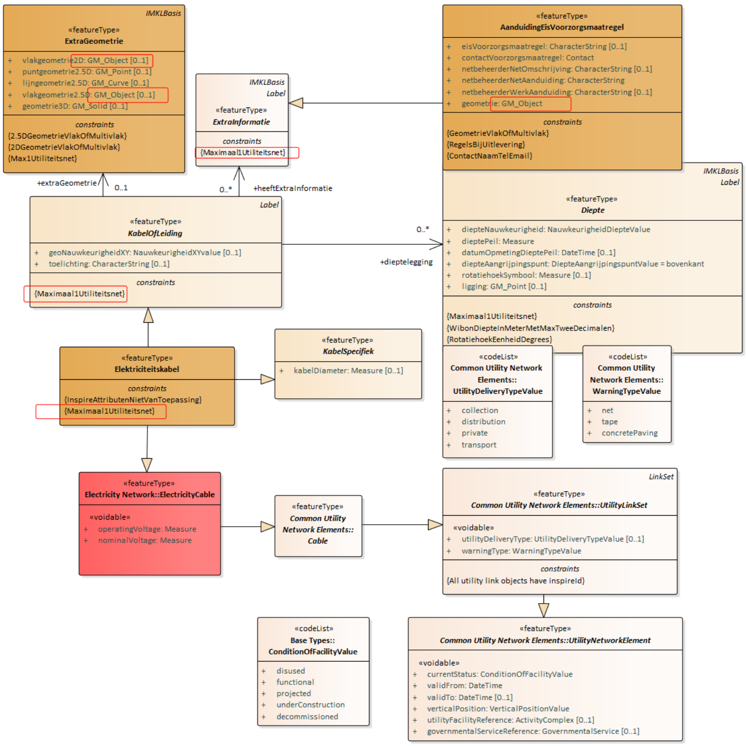


Figure 2.1.: IMKL example UML diagram for electric cables [van den Berg and Janssen, 2021]

pipelines [Koppens, 2014]. The primary purpose of KLIC is to ensure that any excavation activities do not inadvertently damage underground utilities such as electricity cables. When an excavator plans to dig, they submit a request to KLIC, which identifies relevant network operators. These operators then provide detailed maps and instruction about their infrastructure in the affected area. A law (WION) was adopted in 2008 that obligated the operators to do so using a standard information model to register their infrastructure [Koppens, 2014]. The Information Model for Cables and Piplines (IMLK), an extension of the INSPIRE model, serves as a set of naming and classification standards that underpin the KLIC system, ensuring that all data pertaining to underground networks is consistently formatted and easily interoperable [van den Berg and Janssen, 2021]. These standards specify how different types of utility lines should be identified and documented in the shared database. The adoption of these standards enhances safety and contributes to the efficiency of construction and maintenance projects. Figure 2.1 depicts an example of the UML diagram used by KLIC for electric cables. The IMKL calls for the mandatory use of 2D geometries, however, as 3D (depth) location is not always known, the IMKL includes the option to model objects in 3D however it is not always commonplace [van den Berg and Janssen, 2021].

2.2.4. MUDDI Data Model

The MUDDI (Model for Underground Data Definition and Integration) data model is designed to standardise and enhance the representation and interoperability of underground

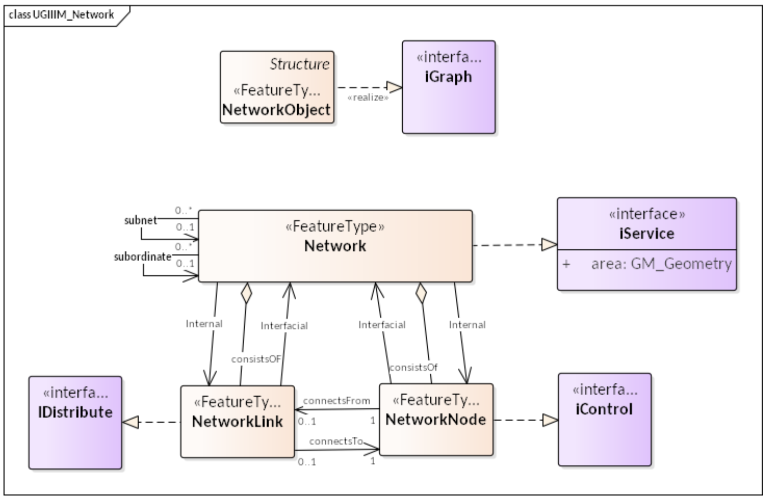


Figure 2.2.: MUDDI network entities and relationships [Lieberman, 2019]

utility data. MUDDI provides a comprehensive framework for modeling the spatial and semantic aspects of underground utilities, including water, gas, electricity, and telecommunications. This model supports detailed representations of utility components, such as pipes and cables, their properties, and their functional relationships. By using a standardised approach, MUDDI facilitates the integration of underground utility data into broader infrastructure models, enabling better planning, maintenance, and management. The MUDDI data model supports varying levels of network representation complexity and detail. The MUDDI data model supports different levels of network complexity and detail. It highlights the relationships between network entities, showing how they can interrelate and connect. Figure 2.2 illustrates these interconnections, emphasizing how nodes and links can serve multiple functions within and between networks [Lieberman, 2019]. Its emphasis on detailed and accurate 3D representations addresses the limitations of traditional 2D models, offering improved insights into the spatial relationships and interactions between different utilities. MUDDI's compatibility with other data models and standards ensures seamless data exchange and collaboration across various domains, making it an essential tool for projects focused on utility networks and urban infrastructure [Lieberman, 2019].

2.3. Relevant Case Studies

2.3.1. Representing Geographical Uncertainties of Utility Location Data in 3D

The paper "Representing Geographical Uncertainties of Utility Location Data in 3D" by olde Scholtenhuis et al. [2018], addresses the challenge of accurately representing the uncertainties associated with utility location data. The study emphasises the importance of 3D visualization to avoid ambiguities in utility data interpretation. Existing models and 3D solutions typically handle uncertainties using textual attributes or requires complex stochastic

Property	INSPIRE	KLIC	MUDDI	CityGML (CGML)
Purpose	Spatial data infrastructure for Europe	Utility location information center	Utility data integration	3D city models
Data Model	Multi-theme data specifications	Utility network data	Multi-utility data integration	XML-based 3D models
Standardization Body	European Commission	Kadaster (Netherlands)	MUDDI Consortium	Open Geospatial Consortium (OGC)
Geometric Representation	2D and 3D	Primarily 2D	2D and 3D	3D
Semantic Content	High	Limited	High	High
Data Integration	Pan-European datasets	National level	Multi-utility, national level	City-level, comprehensive
Key Features	Interoperability, harmonised data	Utility locating, damage prevention	Utility data standardization	Detailed 3D city representations
Application Domains	Environmental, urban planning	Construction, excavation	Urban planning, utility management	Urban planning, infrastructure
Accessibility	Public and private	Mainly private for utility operators	Public and private	Public and private
Interoperability	High	Medium	High	High

Table 2.1.: Comparison of INSPIRE, KLIC, MUDDI, and CityGML

data and expert input, which are often impractical to obtain. The authors propose an innovative approach that integrates multiple available utility location datasets to represent geographical uncertainties more effectively. By extending the CityGML Utility Network ADE model, the study introduces parameters for location data—surveyed, standard, estimated, and unknown—and calculates 3D uncertainty buffer shapes based on these parameters. This method enables a more accurate and comprehensive representation of utility networks by considering the inherent uncertainties in location data. The authors identify several issues with current practices, such as the reliance on 2D plans and relative positioning, which lead to inaccuracies in utility data. Additionally, they highlight the challenge of incorporating the vertical dimension (z-coordinates) in utility records due to the uncertainties and potential liability issues [olde Scholtenhuis et al., 2018].

2.3.2. Underground Utilities 3D Data Model

"Towards an Underground Utilities 3D Data Model for Land Administration" by Jaw et al. [2018] explores the development of a comprehensive 3D data model to improve the management and representation of underground utilities. Their research is driven by the increasing urban density and the need to optimise land use, prompting a shift of infrastructure below ground. The authors propose a framework for utility data governance that manages the entire workflow from data capture to data usage. This includes integrating newly collected data with existing 2D and cadastral information to create a unified 3D map of underground utilities.

A key discovery of this study is the development of a conceptual 3D underground utility data model, which incorporates geometric, spatial, and physical information about utility networks. The model aims to connect underground utility data with cadastral parcels to enhance land administration processes. In a case study conducted in Singapore, the integration of Ground Penetrating Radar (GPR) data with existing utility records demonstrated the potential for improving the accuracy and completeness of underground utility maps. However, the study also identifies significant challenges, such as the limitations of current 2D data, the need for reliable data governance standards, and the difficulty of ensuring data accuracy and completeness. These challenges highlight the necessity for ongoing development and refinement of data models and integration techniques to fully support urban planning and land administration needs.

2.3.3. Data Modeling for Operation and Maintenance of Utility Networks

"Data Modeling for Operation and Maintenance of Utility Networks: Implementation and Testing" by Fossatti et al. [2020], investigates the application of the Operation & Maintenance (O&M) Domain Ontology, an extension of the CityGML Utility Network ADE, to manage utility networks. The goal is to create a consistent and comprehensive data model that can process, store, and exchange O&M-related utility network data effectively. By simulating a street reconstruction project, they evaluated the model's ability to support asset management tasks, such as retrieving maintenance history and performance data, as well as information on site conditions and valve locations.

Key discoveries include the demonstration that the O&M Domain Ontology can indeed support real-life asset management tasks by providing rapid and comprehensive access to utility data. The study highlighted the importance of systematic data collection and registration to enable data-driven asset management practices. However, the main issues encountered were the fragmentation of asset information across different organizations and data models, and the challenges in achieving interoperability. Additionally, the lack of detailed life-cycle data and surrounding soil and groundwater information in existing models was identified as a significant limitation.

2.3.4. Decision Support for Test Trench Location

The paper "Decision Support for Test Trench Location Selection with 3D Semantic Subsurface Utility Models" by Racz et al. [2017] explores the development of decision support systems for selecting test trench locations using 3D semantic models of subsurface utilities. This research focuses on enhancing the accuracy and efficiency of subsurface utility detection and management by leveraging detailed 3D models. The authors propose a methodology that integrates multiple data sources and advanced 3D visualization techniques to identify optimal test trench locations, considering factors such as utility density and spatial relationships. The study highlights the importance of comprehensive 3D models in improving decision-making processes, reducing excavation risks, and optimizing resource allocation for utility maintenance and infrastructure projects. This approach contributes to the field by demonstrating how 3D semantic models can enhance the precision and effectiveness of subsurface utility management [Racz et al., 2017].

2.3.5. Modelling Utility Network Features in Rotterdam

"Modelling Below- and Above-Ground Utility Network Features with the CityGML Utility Network ADE: Experiences from Rotterdam" by den Duijn et al. [2018] discusses a comprehensive approach to integrating below-ground utility networks with above-ground city objects using the CityGML Utility Network ADE. The goal of this research was to create a unified 3D data model that can manage and represent both below- and above-ground utility features for urban planning, city administration, and disaster management. The researchers utilised the Feature Manipulation Engine (FME) software to transform existing utility data into a format compliant with the CityGML Utility Network ADE and stored the data in the 3D City Database (3DCityDB) [den Duijn et al., 2018]. The test case involved the low-voltage electricity and sewer networks of Rotterdam, integrating data from streetlights, manholes, and other city objects.

Key discoveries include the successfully demonstrating the Utility Network ADE's capability to handle complex topological and topographical relationships. The research showed that the 3DCityDB, combined with pgRouting, could perform network analyses essential for utility management and urban planning. However, the study also identified several challenges. One major issue was the complexity of the ETL (Extract, Transform, Load) process, which heavily depended on the quality and format of the source data. Another significant challenge was the need for further detailing of the CityGML Utility Network ADE to accommodate additional types of analyses and relationships.

2.3.6. Summary

The studies reviewed highlight several common issues that are directly relevant to the challenge of semantically enriching utility network data using utility trench survey data. A key problem across these studies is the difficulty of integrating heterogeneous data sources due to the fragmentation of asset information and the lack of interoperability among different data models. This fragmentation leads to incomplete and inconsistent utility network representations, which can significantly hinder effective asset management and urban planning. Additionally, the reliance on 2D plans and the complexities of incorporating vertical dimensions (z-coordinates) are recurrent challenges that result in inaccuracies and ambiguities in utility data interpretation.

These studies also underscore the necessity for advanced data models that can comprehensively represent both topographical and topological aspects of utility networks in a unified 3D framework. The proposed solutions, such as the CityGML Utility Network ADE and the use of advanced GIS tools like FME and 3DCityDB, provide valuable methodologies for systematically incorporating diverse data sources into a cohesive and detailed utility network representation. These approaches can be leveraged to improve the completeness, reliability, and usability of utility network data, ultimately facilitating more effective urban planning and maintenance operations.

3. Methodology

The desired outcome of this thesis is a set of paired enriched datasets and visualisations constructed from the data provided of Amsterdam, Rotterdam and Enschede. This methodology chapter will be split into several parts; Data acquisition (Section 3.1), data analysis and preparation (Section 3.2), data integration (Section 3.4), and result analysis & validation (Section 3.5). The general methodology workflow can be seen in Figure 3.1, where a data analysis is performed on the input datasets of each city. Following the analysis, the data will be cleaned of any errors in the topology, geometry and/or attributes. With the errors corrected, the data can be extracted from the source files and the position of individual cables can be calculated, providing us with an enriched table for each trench. Following this, the data will be imported into QGIS for further processes and visualisation, and then an analysis of the results will take place.

3.1. Data acquisition

For the purpose of this research we were able to get in contact with three distinct cities, Amsterdam, Rotterdam, and Enschede. The data provided by these cities were acquired and utilised in this thesis. Each of these cities exhibits unique attributes, including distinct naming conventions and design decisions, which pertain to both their utility network maps and the cross-sections of their utility trenches. The following subsections will describe the data acquired from each of the cities respectively and will identify challenges encountered with both semantic and geometric information from the utility network, and utility trench data for each city.

3.1.1. Enschede data acquisition

The data received from Enschede is a combination of the KLIC utility network map and a set of utility trenches surveyed by the Siers Groep. Two sections of the city were acquired and their contents are displayed in Table 3.1.

Along with the contents listed in Table 3.1, several PDF files are included in the contents. The majority of these are simply prints of the individual utility trenches (Figure 3.2). However, some of them display examples of anomalies in the data in relation to real world positions, this is exemplified in Figure 3.3 where the red text informs us that the cable HDG is 160mm in diameter and not 200mm. Some of the files provided displayed the results of Ground Penetrating Radar (GPR), these provide valuable information regarding the positional differences between the digital map and reality (Figure 3.4). However, the GPR data did not have a trench visualisation that could be directly extracted and thus could not be used for future steps.

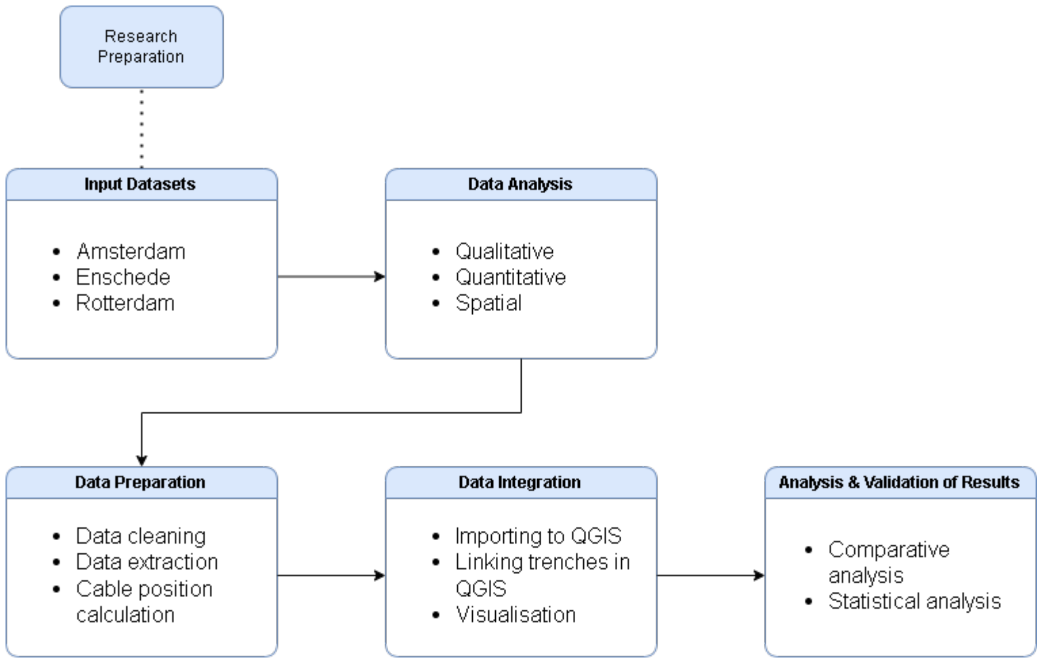


Figure 3.1.: Methodology workflow

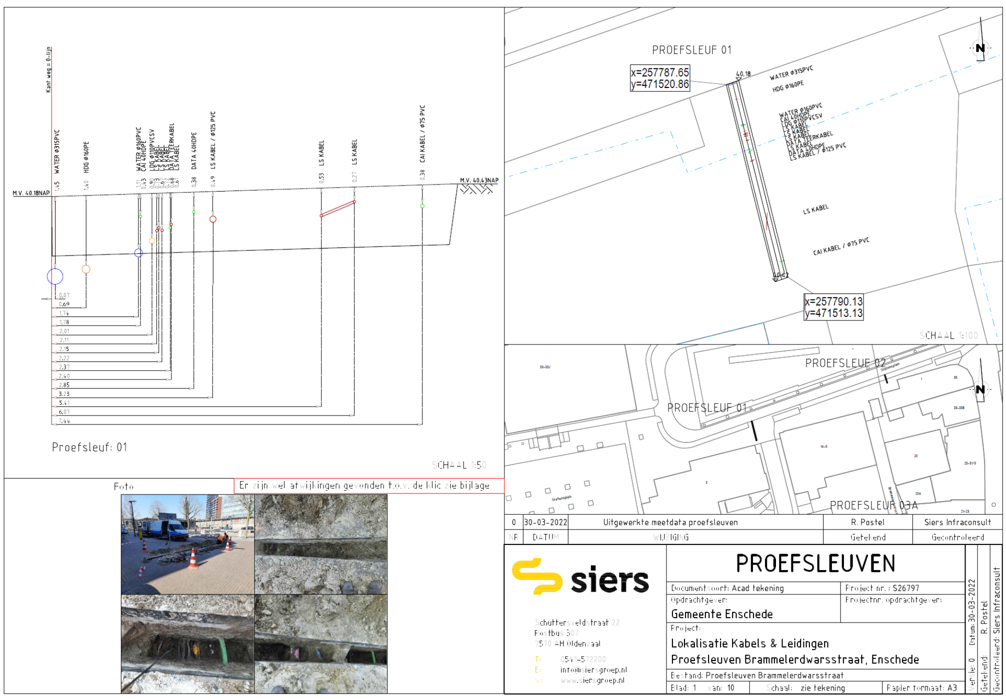


Figure 3.2.: Brammelerdwarsstraat, Enschede: Utility trench PDF file

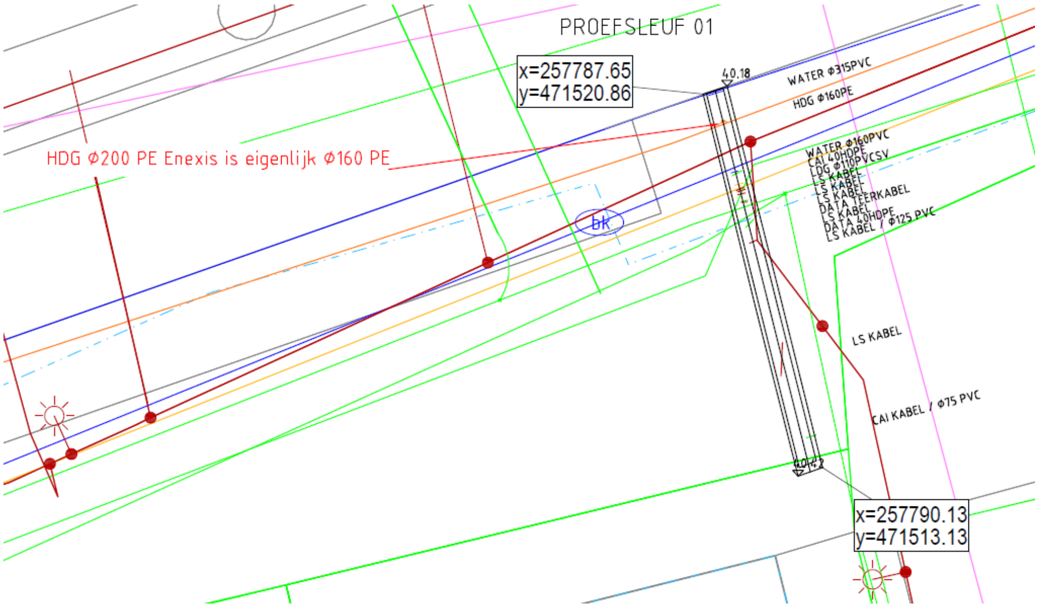


Figure 3.3.: Brammelerdwarsstraat, Enschede: Utility trench 'Afwijkingen' PDF file

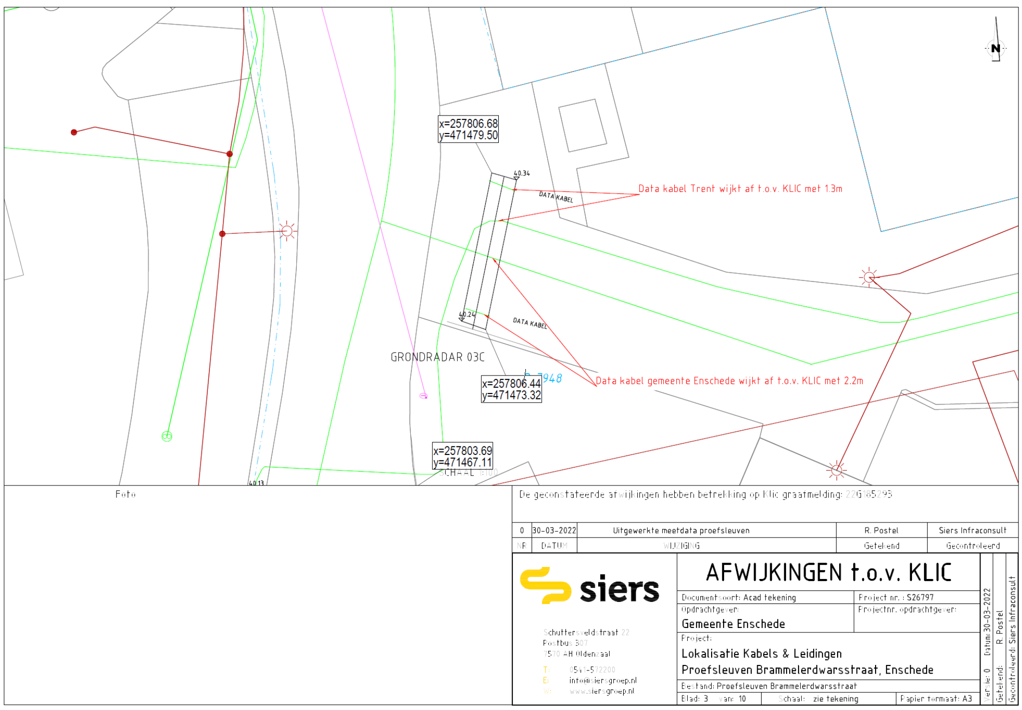


Figure 3.4.: Brammelerdwarsstraat, Enschede: Utility trench GPR PDF file

	File type			
	Utility trenches	GPR	CAD/.dwg	Images
Brammelerdwarsstraat	10	3	1	23
Deurningerstraat	25	3	1	58

Table 3.1.: Contents of data received from Enschede

CAD file

As discussed, the Computer Aided Design (CAD)/.dwg files obtained consist of KLIC utility network maps with added utility trench cross-sections from the contractor. Figure 3.5 illustrates the positions of each trench (proefsleuf) on the KLIC map. These trenches provide coordinates for the opposite corners of their bounding boxes, and each cable within is distinctly labeled (Figure 3.6). Notably, each utility trench containing anomalies is annotated with red text to highlight the discrepancies, such as the incorrectly labeled pipe in Figure 3.3.

All provided files originate from the Netherlands and are referenced using the Amersfoort / RD New Netherlands coordinate system, detailed as follows:

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

The CAD files from Enschede are geo-referenced, facilitating their integration into mapping software. This integration process will be explored further in a subsequent section of this thesis (Section 3.4).

Utility trench data

Each utility trench depicted on the map is accompanied by a corresponding cross-section. These cross-sections provide crucial elevation data, facilitating the enhancement of 2D maps to more detailed representations. Additionally, they offer precise positional data of cables within the utility trenches, with each cable distinctly labeled according to its function and diameter. An example of the utility trenches provided by Enschede can be seen in Figure 3.8. In this figure, one can observe the depth of each cable below the surface, the elevation at the start and end of the trench, the distance from the side of the utility trench as well as the names and diameters of each cable.

In the case of the CAD files from Enschede, a significant challenge with the utility trench cross-sections is that they are integrated in the same layout as the utility network maps itself. Extracting data from a CAD file that contains multiple utility trenches, let alone an entire map, poses considerable difficulties for automated extraction and integration processes, as envisioned in this thesis. These difficulties will be addressed in Section 3.2.

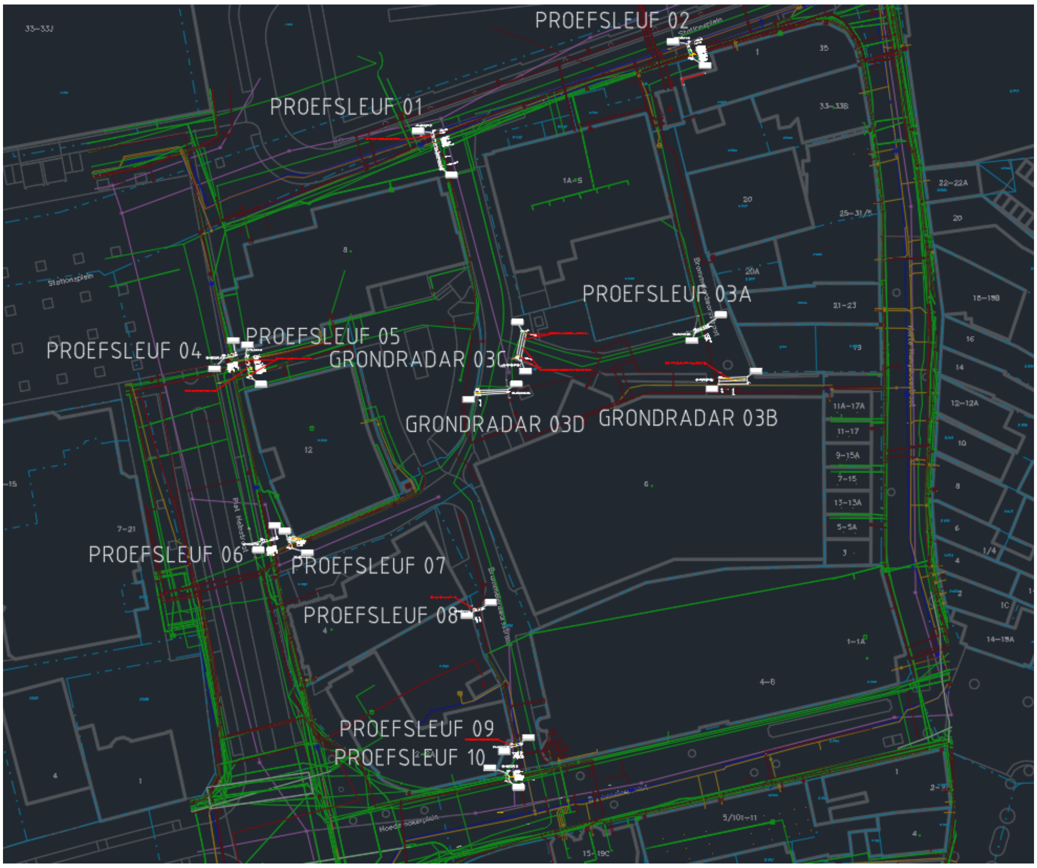


Figure 3.5.: Enschede, Brammelerdwarsstraat: Trench positions on overview map

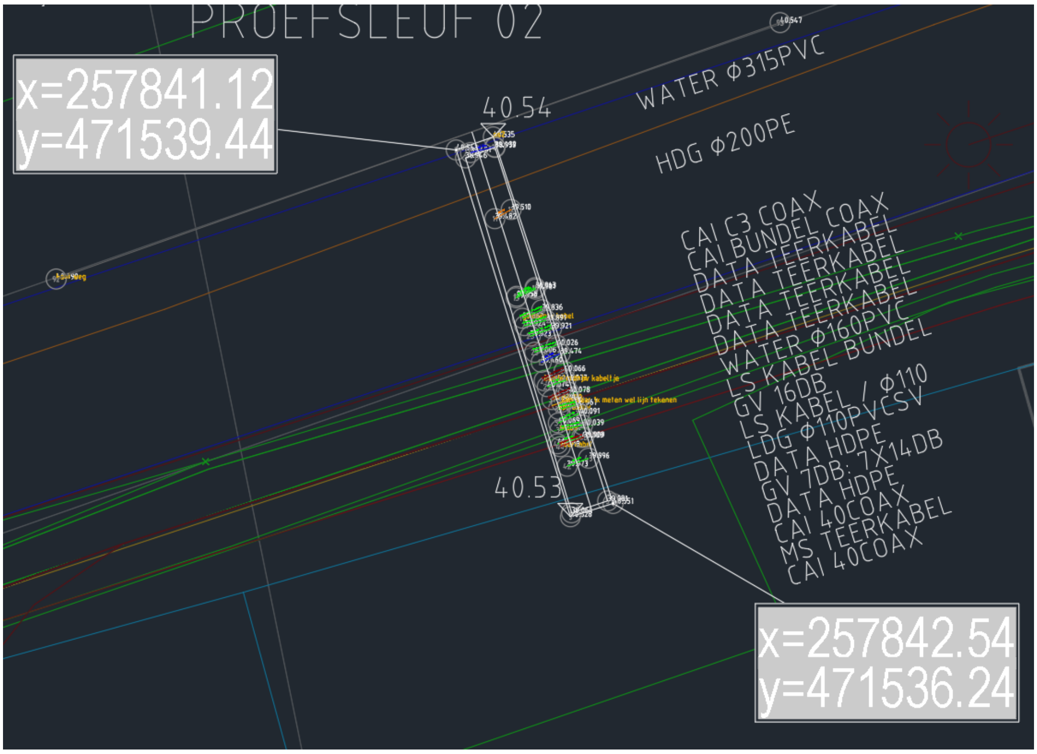


Figure 3.6.: Enschede, Brammelerdwarsstraat: Zoomed view on 'Proefsleuf 2'

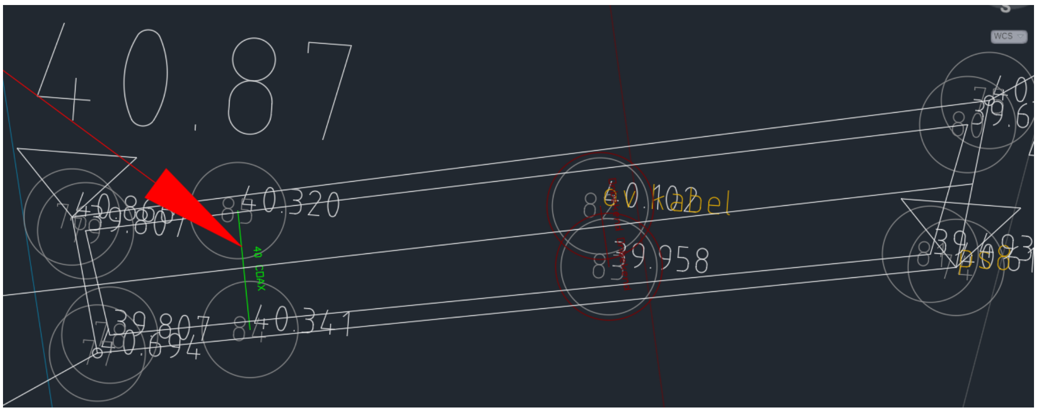


Figure 3.7.: Enschede, Brammelerdwarsstraat: Closeup view of 'Proefsleuf 8'

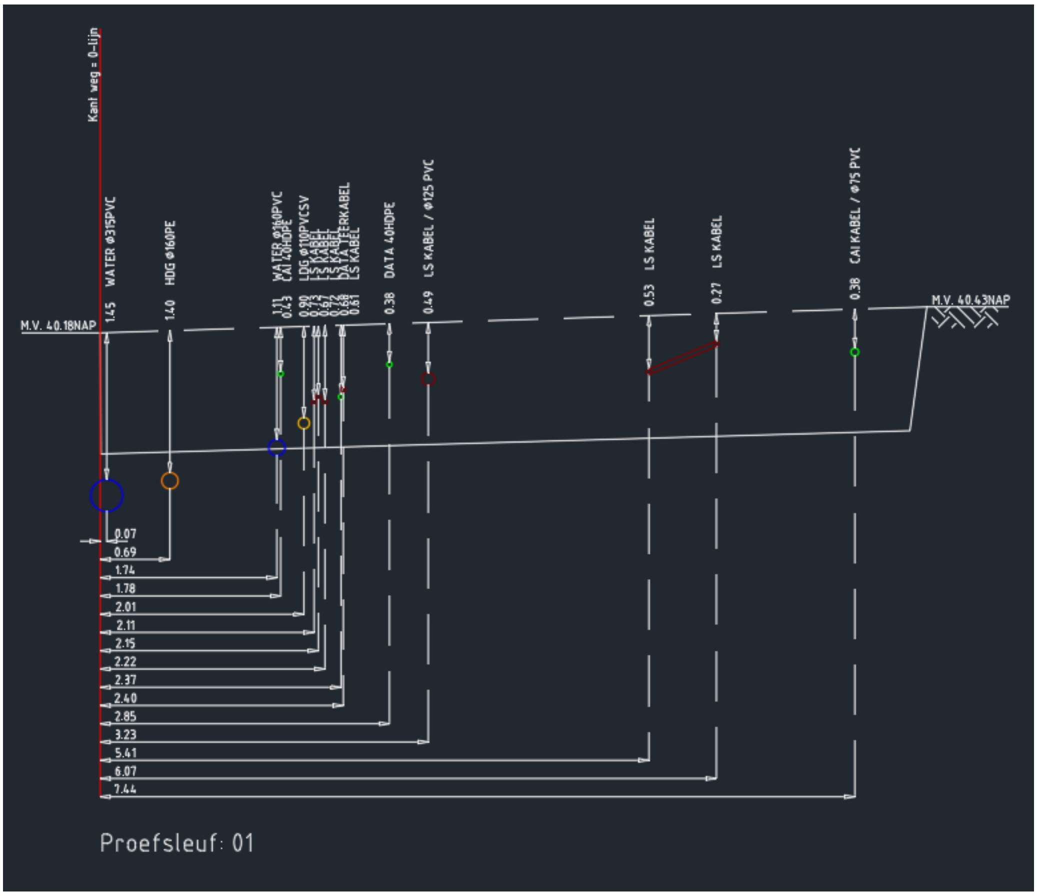


Figure 3.8.: Example utility trench on Brammelerdwarsstraat in Enschede

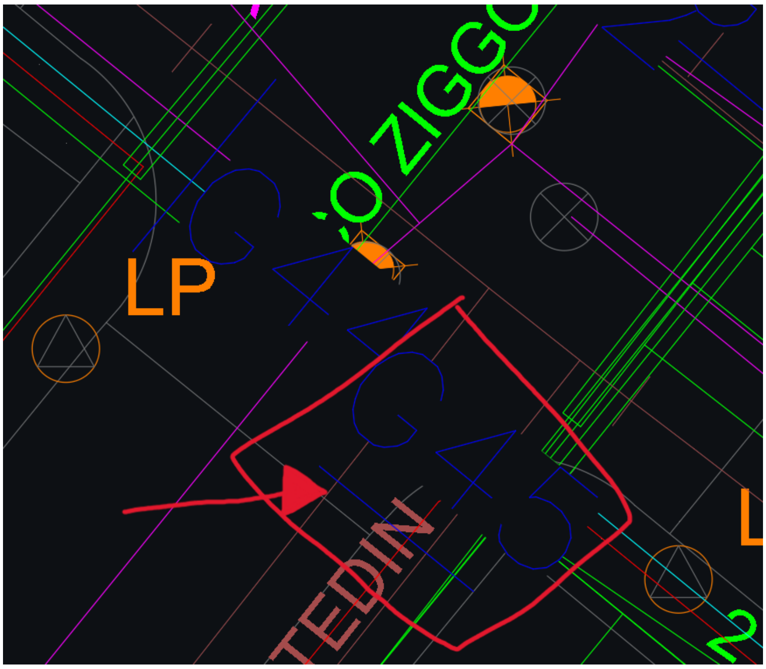


Figure 3.9.: Trench positions on Rotterdam utility network map

3.1.2. Rotterdam data acquisition

The Rotterdam data is much like that of Enschede's, however, there are a couple of key differences that will need to be taken into account for when developing the data extraction process. In terms of the contents received from Rotterdam, the files include:

- 1 overview .dwg
- 1 utility trench positions .dwg
- .dwg files containing 99 utility trench cross-sections (proefsleuven)

CAD file

Unlike Enschede, Rotterdam uses their own utility map, LVZK, instead of the KLIC. It exhibits cables that are largely similar to those in that of Enschede; however, it contains significantly more labels within the drawing file. While this increased labeling enhances clarity, it necessitates extensive cleaning and removal of extraneous labels prior to implementation. Unlike the data from Enschede, the Rotterdam data lacks a visualisation of the trenches on the utility network map and instead uses a line to represent the start and end points of each trench. As can be seen in Figure 3.9, the blue lines represent the trench cross-section.

It can be seen in Figure 3.10 that there are issues with the projection of Rotterdam's data. The data provided is in fact using the Amersfoort / RD New Netherlands as expected, but the file that was provided had the data projected to WGS84.

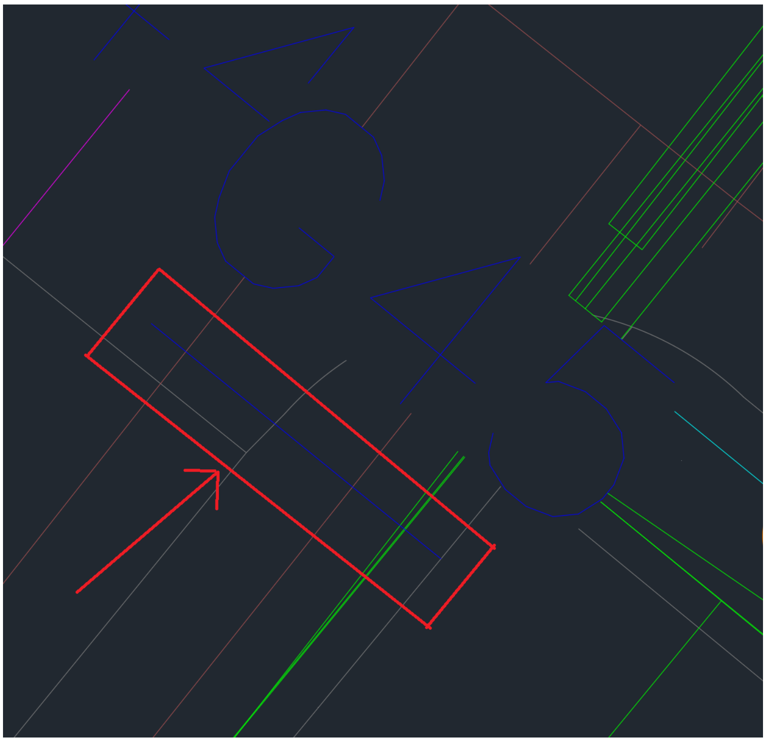


Figure 3.10.: Poorly projected data in Rotterdam's utility network map

Utility trench

The utility trench cross-sections provided by Rotterdam are much like that of Enschede in the sense that a distance from the origin and elevation is provided for each cable. Unlike Enschede, many of the line representing these cables are left unlabeled and thus, may affect how much information can be extracted.

One main distinction between the two cities is that the data from Enschede provides the depth of the top of the cable below the surface, whatever elevation that surface may be. In contrast, the Rotterdam data provides the actual elevation in Normaal Amsterdams Peil (NAP), which is the standard reference level used for altitude measurements across the Netherlands, originally based on the average sea level at Amsterdam. This will help reduce future calculations, but also is an example of why the extraction processes will have to be different for Enschede and Rotterdam.

3.1.3. Amsterdam data acquisition

Unfortunately, in the case of Amsterdam, issues were encountered with the extraction of the data. Initially the data provided was only in a PDF format. No vector information could be extracted from the PDF and the vector drawings from which the PDFs were generated could not be obtained. Eventually, additional data was provided (See Figure 3.12). This utility network map however, was not provided alongside utility trenches and thus the elevation

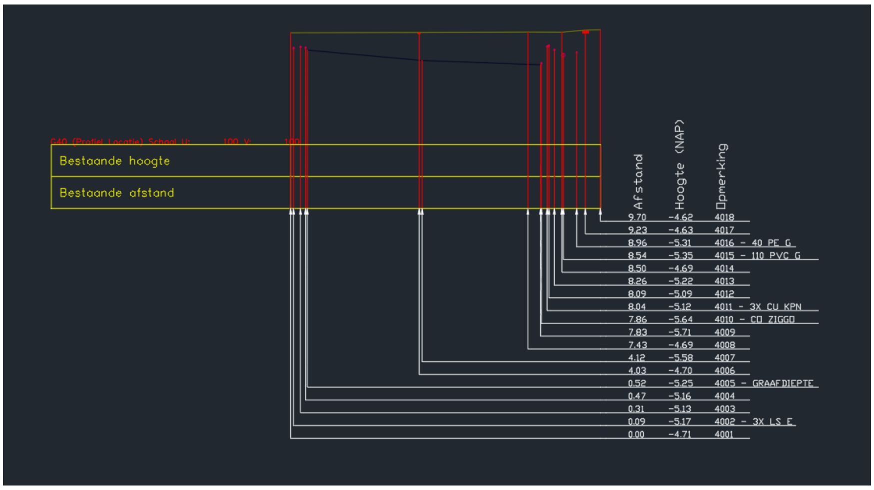


Figure 3.11.: Example of utility trench cross-section in Rotterdam

data was still not attainable. Issues as such, will be discussed in Chapter 6 however, in terms of methodology, further processing/extraction can not be performed and the data from Amsterdam will not be used in the implementation.

3.2. Data analysis

3.2.1. Enschede data analysis

The data analysis for Enschede focused on examining the alignment and integrity of the utility network maps and utility trenches. Key observations and steps taken to refine the data include:

- **Data organization:** In the Enschede CAD files, the utility network map and the utility trench drawings are both visualised in the same layer. This may cause issues in the extraction process as extracting the text from the utility trenches may result in the extraction of text from the map itself.
- **Elevation data:** The depths of individual cables in the utility trenches are measured as distance below the surface level at which the survey is conducted. The elevation of the surface is provided in the utility trench cross-sections at the start and end point of each trench. In order to acquire the elevation at any point along the trench, the trench surface is assumed to follow a straight line from the start NAP to the end NAP. This assumption is necessary to extract the elevation of individual cables, however, this may lead to some errors as it is unlikely that the surface between two points is exactly straight.
- **Alignment Errors:** A significant number of utility lines on the network map did not correspond accurately with the utility trenches, as illustrated in Figure 3.15. These annotations within the dataset highlighted the inaccuracies between the KLIC data and actual field observations, and need to be implemented in the data extraction process.

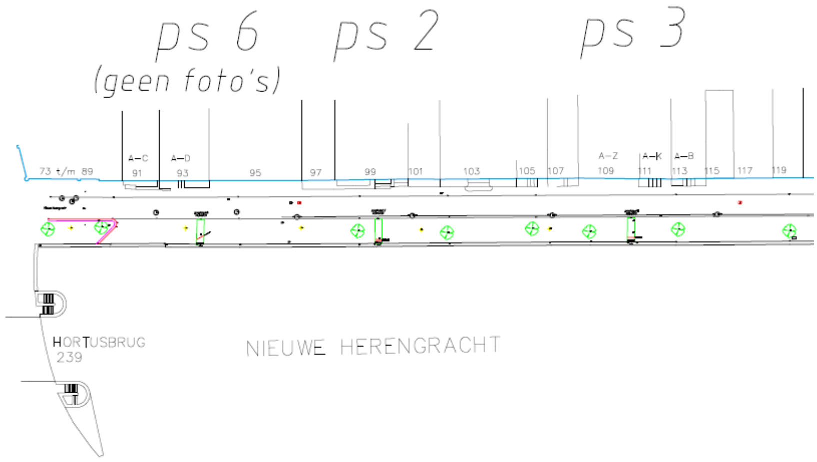


Figure 3.12.: Screenshot of Amsterdam utility trench map overview

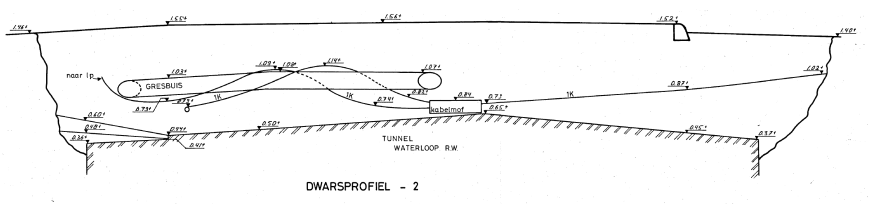


Figure 3.13.: Example of Amsterdam utility trench cross-section

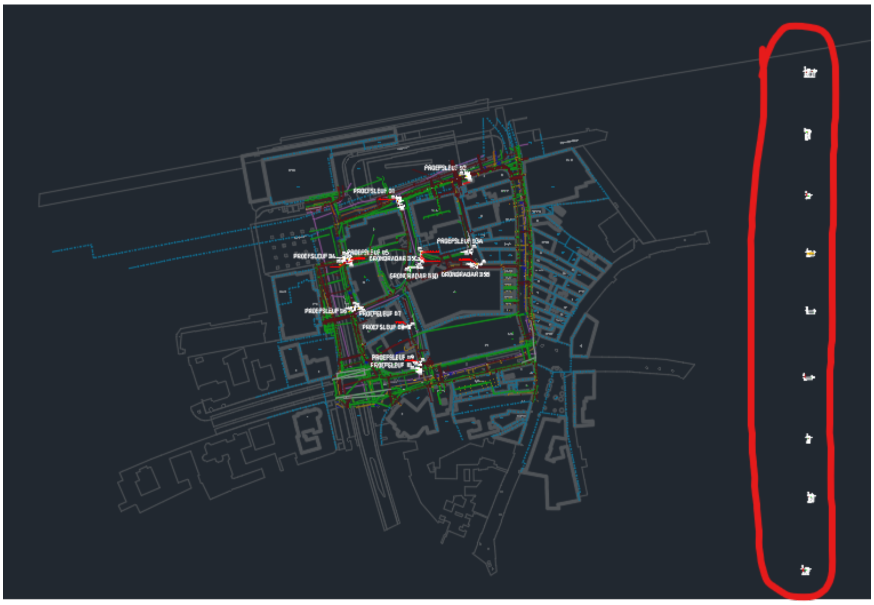


Figure 3.14.: Utility trenches cross-sections, circled in red, in the same layer as the utility network map

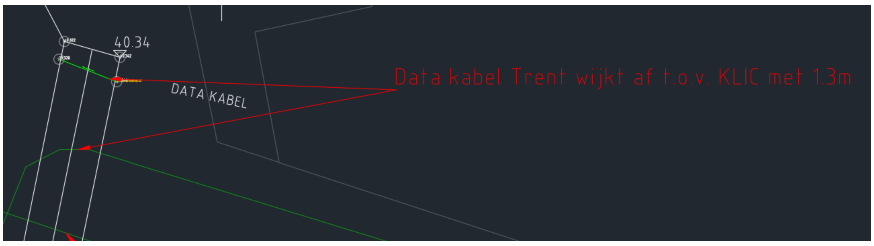


Figure 3.15.: Inaccuracies in Brammelerdwarsstraat utility network map

- **Unknown Labels:** Some cables in the utility network map were labeled as "Onbekend" (unknown), which indicated missing information about the cable type. These instances were catalogued for further investigation or possible exclusion from critical analyses until additional information could be obtained. Given that every cable in the utility trench surveys are identified and named, it is assumed that the unknown cables can be identified using the utility trench data.
- **Data Parsing:** The diameter of cables was often recorded in the same text field as the cable type. This will require the development of a parsing algorithm to separate these attributes into distinct fields to facilitate more structured queries and analysis.

3.2.2. Rotterdam data analysis

- **Data Organization:** Unlike the data from Enschede, the utility trenches in Rotterdam are organised in their own layer within CAD, which simplifies data management and reduces potential confusion during implementation.
- **Elevation Data:** The elevation information for cables is provided in terms of NAP rather than depth below the surface (Figure 3.16). This standardization significantly streamlines the process by eliminating the need for depth conversions, thus enhancing the efficiency of data utilization.
- **Unique Identifiers:** Each cable in the trenches is marked with a unique identifier that is embedded within the same text string as the descriptions, as seen in Figure 3.16. To standardise data for analysis, these identifiers must be parsed and removed to ensure uniformity across descriptions.
- **Separation of Attributes:** Similar to the approach taken with the Enschede dataset, the cable diameters in the Rotterdam data are combined with descriptions and need to be isolated into a distinct diameter attribute. This step is essential for data modeling and analysis.
- **Labeling Inconsistencies:** A significant challenge identified is that the majority of cables are unlabeled, thus lacking any information on the type a utility is. Figure 3.17, shows that only one cable in the trench is identified. Note that 'GRAAFDIEPTE' is just the depth of this specific trench and not a cable or pipe. This lack of labeling could complicate further analysis and modeling within the thesis. Steps need to be taken to either procure or infer the missing labels to ensure comprehensive data analysis. If this is not possible, the cables will not be able to be linked to one another between successive trenches.

3.3. Data preparation

To facilitate a semi-automatic integration procedure, it is essential that the utility network data from each city be standardised in structure. To do so, distinct scripts need to be developed for each city to extract and standardise their respective datasets. This preparatory step ensures that all subsequent processes can be executed automatically, independent of the city-specific data involved. This standardization is critical for maintaining consistency across the integration workflow.

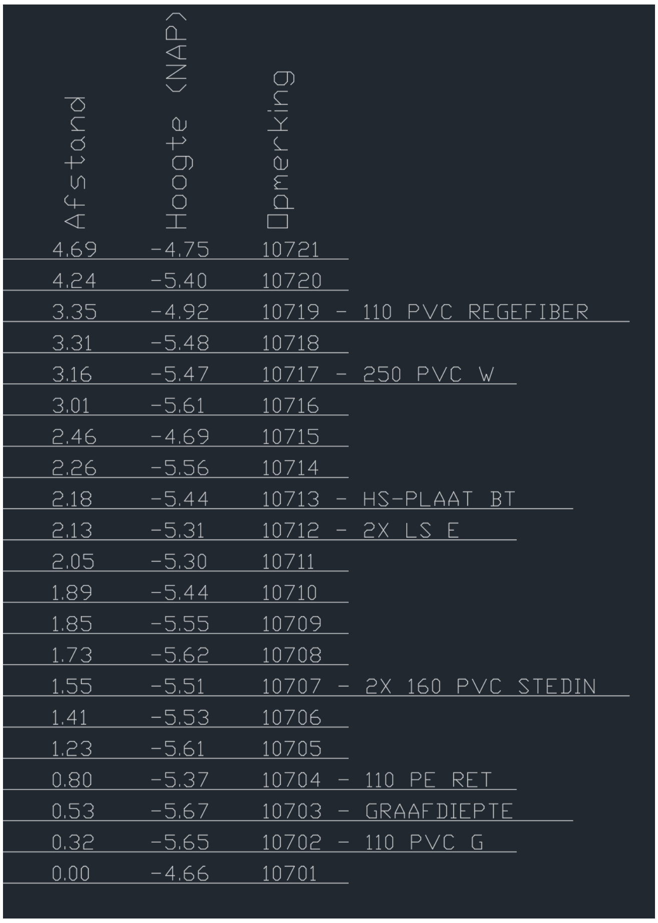


Figure 3.16.: Example of information stored in a utility trench provided by Rotterdam

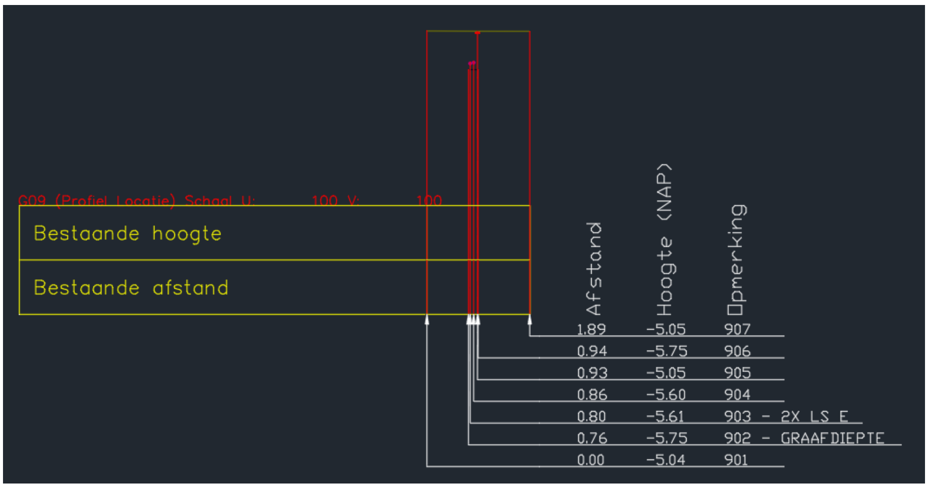


Figure 3.17.: Rotterdam: Utility trench G09

3.3.1. Extracting surveyed trench data

The first step of the process is to extract the points from the utility trench drawings. This means getting the distance, depth and description from the trench drawings. Algorithm 3.1 is a pseudo-code that shows how text and number layers in the DXF are extracted and split into chunks (distance, depth and description). This is done by searching for the prefix of each trench in the DXF file, in this case ”M.V.” and then relating all data in that trench to the trench number. In addition, information regarding the position of the trench itself is required in order to re-project these points on a map. Algorithm 3.2 shows that, by filtering by the ”Profielen” (trench line) layer, it is possible to extract the vertices of the start and end coordinates (including elevation) of each trench line.

3.3.2. Linear Interpolation for cable positioning

The process for calculating the exact geographic positions (Easting, Northing, and Elevation) of cables along the utility trench incorporates linear interpolation based on extracted coordinates. This method ensures precise positioning necessary for accurate mapping and further analysis.

Easting Calculation:

$$E_x = X_1 + \left(\frac{d_x}{l}\right) \times (X_2 - X_1)$$

Here, E_x is the calculated Easting, X_1 and X_2 are the Easting values at the start and end of the trench line, d_x is the distance along the trench line to the cable point, and l is the total length of the trench line.

Northing Calculation

$$N_x = Y_1 + \left(\frac{d_x}{l}\right) \times (Y_2 - Y_1)$$

Similar to the Easting, N_x is the Northing calculated for each point, using Y_1 and Y_2 as the Northing at the start and end respectively.

Elevation Calculation

$$NAP_x = Z_1 + \left(\frac{d_x}{l}\right) \times (Z_2 - Z_1) - h_x$$

NAP_x represents the adjusted elevation considering the depth (h_x) of the cable from the surface, interpolated from the start (Z_1) and end (Z_2) elevations of the trench line.

It is important to note that, the depth value provided by Enschede is the depth below the surface of the trench and thus the prior would be needed. However, for Rotterdam, the elevation value provided by the trench surveys is already in NAP and thus the elevation calculation will not need to be performed, only the easting and northing calculations.

Algorithm 3.1: Extract Text and Numbers from DXF Layout

```
Input: file_path, layout_name
Output: Excel file with extracted text and numbers
Function extract_text_and_numbers_from_layout(file_path, layout_name="Layout1"):
    doc ← Read DXF file(file_path);
    if layout_name in layout_names of doc then
        layout ← Get layout(layout_name);
        all_data ← Initialize empty list;
        extracted_data ← Initialize empty list;
        foreach entity in layout.query("TEXT MTEXT") do
            text ← entity.dxf.text;
            if text starts with "M.V." then
                if all_data is not empty then
                    Process all_data in chunks of 3;
                    Replace empty chunks with "NA";
                    Append DataFrame of chunks to extracted_data;
                    Append empty row to extracted_data;
                end
                Reset all_data;
            end
            Remove "
            A1;" from text;
            Append text to all_data;
        end
        if all_data is not empty then
            Process remaining all_data in chunks of 3;
            Replace empty chunks with "NA";
            Append DataFrame of chunks to extracted_data;
        end
        if extracted_data is not empty then
            final_df ← Concatenate extracted_data;
            Save final_df to Excel(output_file);
            Print "Data saved to " + output_file;
        end
    end
    else
        Print "Layout not found in the DXF file";
    end
return
Run extract_text_and_numbers_from_layout(file_path);
```

Algorithm 3.2: Extract Vertices from DXF

```
Input: dxf_file_path, output_excel_path
Output: Excel file with extracted vertices
Function extract_vertices(dxf_file_path, output_excel_path):
    doc ← Read DXF file(dxf_file_path);
    wb ← Create new Excel workbook;
    ws ← Get active sheet of wb;
    Append ["Layer", "Point X", "Point Y", "Point Z"] to ws;
    foreach polyline in doc.modelspace().query('LWPOLYLINE[layer=="Profielen"]') do
        foreach vertex in polyline.points() do
            Append ["Profielen", vertex[0], vertex[1], vertex[2]] to ws;
        end
    end
    Save wb to output_excel_path;
return
Run extract_vertices(dxf_file_path, output_excel_path);
```

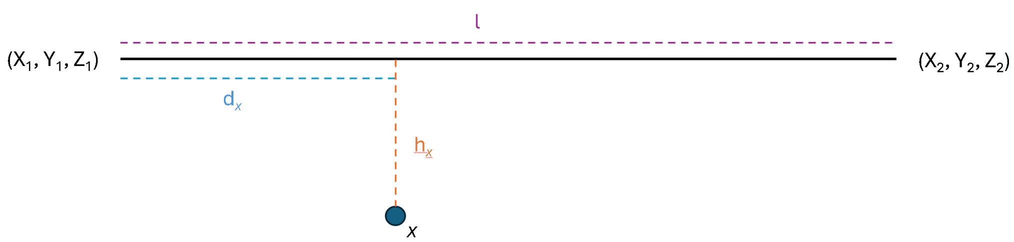


Figure 3.18.: Representation of trench information used in linear interpolation

3.4. Data integration process

3.4.1. Importing to QGIS

Once the datasets are prepared in the same format, they can be imported into QGIS to link trenches, visualise a map and the attributes and perform an analysis on the results. Therefore, the first step of the integration process will be to create a script that imports the trenches automatically. A pseudo-code example of this can be seen in Algorithm 3.3.

Algorithm 3.3: Add CSV Data as Layer

Input: file path, layer name, x field, y field, crs epsg, field types
Output: Layer added to the GIS project
Function *add_csv_as_layer*:
  uri ← Construct URI using input parameters;
  layer ← Add vector layer using uri and layer name;
  return *layer*;
return

3.4.2. Linking trenches

In order to have a proper visualisation of the enriched data, cables will have to be drawn, connecting trenches to one another. The main goal of this task is to do this in an automatic manner. This comes with its own issues as trenches that are not in-line with one another may create an inaccurate representation. The problem arises due to the fact that the surveyed trenches are identified in a random order and contain no topological information that describe how the trenches can be aligned consecutively. This means that, even on a straight road, it is not possible to tell what the order of the trenches are without manual input. Additionally, automating a connection process around corners can be difficult as the cable will likely take the shortest path rather than a more realistic path as represented by the red line in Figure 3.19. Therefore, for the purpose of this thesis it was decided that only linearly successive trenches will be used when creating connections, as non-linearly successive trenches would not be able to provide information regarding connections between trenches.

In the case of linearly successive utility trenches, the method employed for delineating cables involved utilizing the cable names to correlate points from one trench to the next. A significant challenge in this approach is the presence of unknowns or inconsistencies in cable names. The resolution of these issues involves making corrections and educated assumptions, which are elaborated upon in Chapter 4.

Once the inaccuracies have been taken into consideration and the labels of the cables/pipes have been corrected. The next step is to enable connections across successive trenches, the objective becomes to link them sequentially and automate this process. A decision needs to be made in terms of how the cables should be visualised. Evidently when there is a 1-to-1 connection, the best way to represent this is to draw a singular line between the two points. However, in the case that one cable from Trench "A" splits into many cables in Trench "B" (referred to as 1-to-many), visualisation options can vary. A few possible options are presented in Figure 3.20. Figure 3.20a, displays what would happen if one was to draw

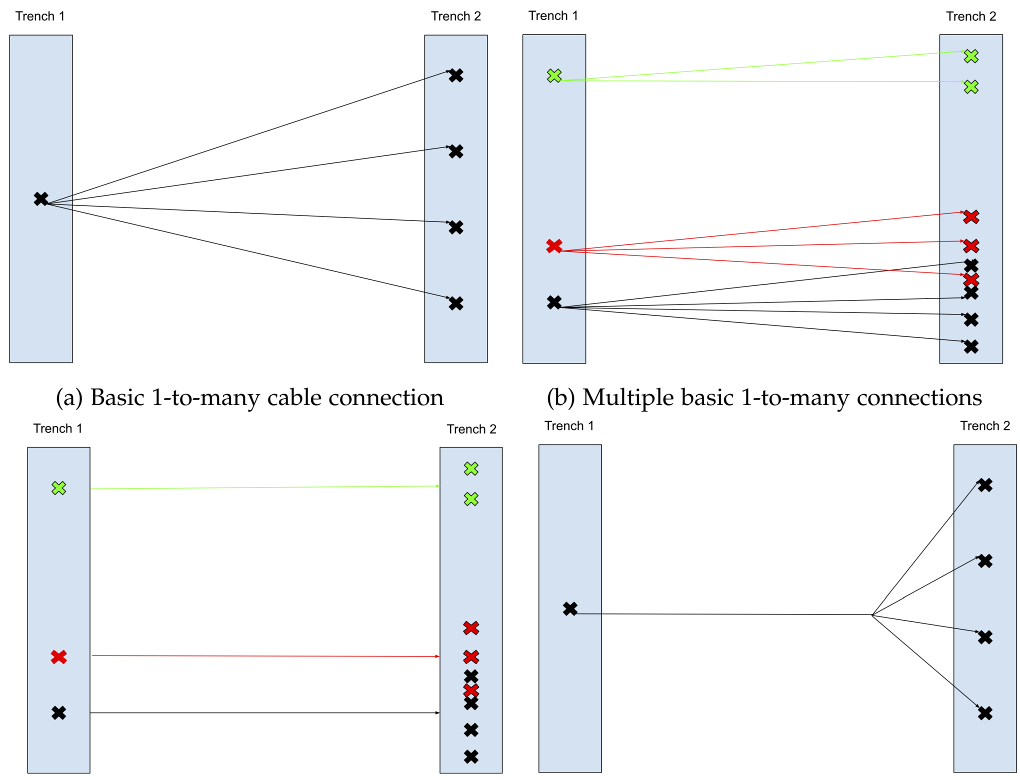
Figure 3.19.: An example of potential issues when automating trench connections around corners

connections between Trench 1 and Trench 2 in a basic iterative manner. This would be the simplest option to implement but also the least coherent as exemplified by Figure 3.20c. On the other hand, to reduce clutter, Figure 3.20d presents a simplified method to represent multiple lines with a single line. This option struggles to show the proper connections occurring between trenches. Therefore, the cable linking method decided upon is that of a "forked" cable that splits into many cables near the destination trench (Figure 3.20b). This will allow for the visualisation to remain relatively clean while also displaying all the existing connections involved between trenches.

Algorithm 3.4, displays the process for finding the centroid of a set of points pertaining to the same trench. This process makes it so that the line drawn from one trench goes to the center of multiple destination points, like Figures 3.20c & 3.20d, rather than to some arbitrary position. Algorithm 3.5, calculates an intermediate point from which the prongs of the 'fork' can split off from. Finally, Algorithm 3.6 shows the process for drawing a line from points in one trench to another.

Algorithm 3.4: Calculate Average Point

Input: List of points
Output: Centroid as QgsPointXY
Function *calculate_average_point*:
    xCoords, yCoords ← Extract coordinates from points;
    centroid ← Create point from mean of xCoords and yCoords;
    return *centroid*;
return



(a) Basic 1-to-many cable connection

(b) Multiple basic 1-to-many connections

(c) Simplified 1-to-many cable connections

(d) ”Forked” 1-to-many cable connection

Figure 3.20.: Four examples of how cable connections can be stylised.

Algorithm 3.5: Calculate Intermediate Point

Input: Points from trench 1, Points from trench 2
Output: Intermediate point as QgsPointXY

Function *calculate_intermediate_point*:
  avgPoint1 ← Calculate average of trench 1 points;
  avgPoint2 ← Calculate average of trench 2 points;
  biasRatio ← Determine bias based on count of points;
  intermediateX ← Weighted average based on biasRatio;
  intermediateY ← Weighted average based on biasRatio;
  return *Create new point intermediateX, intermediateY*;
return

Algorithm 3.6: Establish Connections Between Points

```
Input: Layers with points, Layer to store lines
Procedure establish_connections:
    foreach unique description in layer do
        match points from both layers;
        calculate intermediate point;
        foreach point in Trench 1 do
            create line to intermediate point;
            foreach matching point in Trench 2 do
                create line from intermediate to point;
            end
        end
    end
    commit changes to line layer;
end
```

3.5. Result Analysis & Validation

The following subsections discuss the steps that will be taken when validating and analysing the results of the integration. This consists of linking successive trenches to one another to determine labeling consistency, comparing the integration results to the existing utility network map, and creating visualisations to demonstrate this.

3.5.1. Cable Position Validation

In the data provided by the Enschede, there are markings to signify where the surveyors have located the individual cables to be along a trench line. These markings can be compared to our calculated cable positions in order to find the accuracy of our calculated points. This should allow us to validate the positions and future analyses that include comparisons to the calculated trench positions. Figure 3.21, depicts the surveyed positions, indicated with a red line, compared to our calculated positions. Using a distance calculation we will be able to find the average distance our calculated points are offset by.

3.5.2. Matching Cables Between Trenches

One key analysis involves matching points between trenches. There are three possible outcomes when matching points between two trenches:

1. **Exact Match**: The cable descriptions are identical, meaning they likely describe the same cable.
2. **Unmatched**: No similar descriptions between the two trenches.
3. **Assumed Match**: Informed assumptions are made to connect similarly named cables, common in a 1-to-many or many-to-1 situation (e.g., ”GV 25 DB BUNDEL” in Table 3.2).

Figure 3.21.: Surveyed cable positions (intersection of red line), compared to our calculated positions

Trench 4	Trench 6
WATER /160PVC	WATER /160PVC
CAI 40COAX	CAI BUNDEL COAX
LDG /160PVCSV	LS TEERKABEL
MS 2* BUNDEL	LDG /160PVCSV
LS KABEL	LS AL
GV 25 DB BUNDEL	CAI BUNDEL COAX
LS KABEL	GV 25DB
DATA HDPE	GV 25DB
DATA HDPE	DATA HDPE
RIOOL /200PVC	LS KABEL
	DATA TELEFONIE

Table 3.2.: Brammelerdwarsstraat, Enschede: Cable comparison between trenches 4 & 6

This matching process is visualised using pie charts (see Figure 3.22), showing the proportion of exact matches, assumed matches, and unmatched cables. For instance, in Figure 3.22, only 31% of cables would be visualised without assumptions, while including informed assumptions links an additional 48% of cables. This demonstrates the necessity of standardised naming for cables and pipelines.

For each trench, we count the number of matches, assumptions and unmatched cables. The results are represented in pie charts, showing the proportion of matched, assumed and unknown cables. This process is repeated for all trenches in each city, providing a comprehensive overview at both the trench and city levels. We link known and assumed cables between trenches and compute statistics, including the percentage of exact matches, assumed matches, and unmatched cables. Results are visualised in QGIS to show the links between trenches, highlighting areas with or without discrepancies. The processes above are repeated for all trenches within each case study. By aggregating the results, we obtain case-encompassing statistics, providing insights into the overall accuracy and reliability of utility network data. These aggregated results help identify city-wide trends and potential areas needing further investigation or improvement.

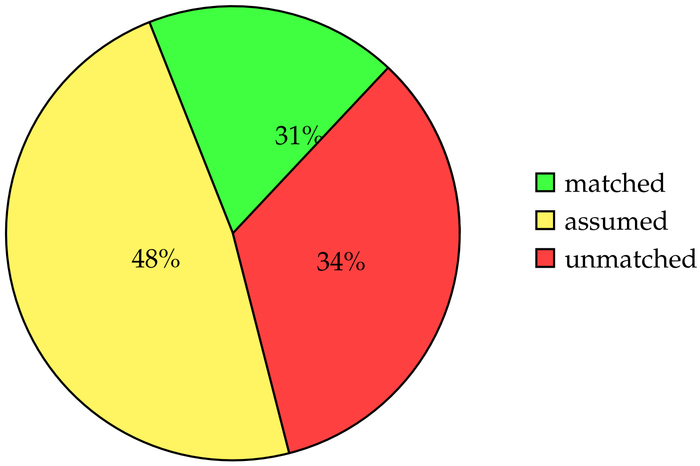


Figure 3.22.: Example distribution of cable status between two trenches

3.5.3. Comparative analysis

The original utility network data sourced from the two cities each employed distinct visualisations. Enschede utilised the KLIC data for mapping their utilities, while Rotterdam implemented its own unique system, referred to as the LVZK.

Given the nature of the data collection through trench surveying, it is a more reliable source of data. Trench surveyors provide highly detailed, localised insights into the subsurface utility layouts, often with a higher degree of accuracy and updated information compared to the pre-existing utility maps like KLIC and LVZK. This difference in data accuracy stems from the direct, physical observation methods used in trench surveying, as opposed to the often generalised data that inform utility network maps.

Therefore, the second analysis will employ a comparative approach, where the enriched, trench-derived data is juxtaposed against the existing datasets (KLIC and LVZK). This comparison not only tests the validity and accuracy of the existing utility maps but also serves to highlight any discrepancies that might affect urban planning and development processes. Specific instances where the trench data diverges from the mapped data could indicate areas in need of updating within the city's official data.

To perform this analysis the KLIC or LVZK points along the trench line must be found. Algorithm 3.7 displays a algorithm for processing the intersection of KLIC points against the trench line and then saving these points as a new layer. Next, as seen in Algorithm 3.8, the distances between the enriched points and their respective KLIC intersection points are calculated. It also contains a count for how many enriched points are unmatched in order to find how many cables are unrepresented in the KLIC or LVZK utility network data.

3.5.4. Visualization in QGIS

The results are represented in QGIS at multiple levels:

1. **Single Trench:** Shows the identification and matching of cables within a single trench.

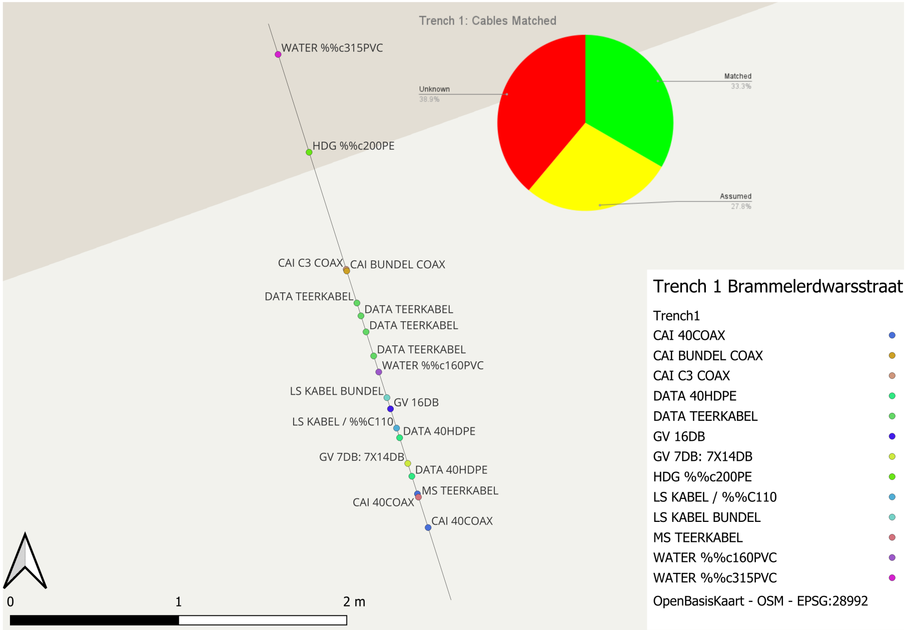


Figure 3.23.: Brammelerdwarsstraat: trench 1 imported points and cable matching visualisation

2. **Linked Trenches:** Visualises the connections and continuity between two successive trenches.
3. **Case Overview:** Provides a view of all trenches in an area.

3.5.5. Summary of analysis

In summary, the analyses will include:

1. Counting known and unknown cables and visualizing them with pie charts. To provide an understanding of how much real world data is unrepresented in the trench connections.
2. Linking cables between trenches and computing relevant statistics.
3. Comparing trench data with DWG files to quantify the accuracy of traditional datasets.
4. Performing statistical analysis of positional differences.
5. Representing results in QGIS for comprehensive visualization.

Algorithm 3.7: Process Intersection KLIC Points

```
Input: Geopackage path, list of line layer names
Output: Intersections layers added to GIS project
Function LoadVectorLayer(file path, layer name, type):
    Construct URI and load vector layer from URI;
    return Layer object or error;
return
Function CreateMemoryLayer(layer type, CRS code):
    Initialise memory layer with attributes and update fields;
    return Memory layer;
return
Procedure ProcessIntersections(base layer, other layers):
    For each layer in other layers, check validity and process intersections;
    Store and update intersection points in memory layer;
    Add processed layers to GIS project;
Main Execution:;
baseLayer ← LoadVectorLayer(gpkgPath, 'Profielen', 'type');
if baseLayer is valid then
    ProcessIntersections(baseLayer, layerNames);
end
```

Algorithm 3.8: KLIC offset distance calculation

```
Input: trench_layer_names, klicpoints_layer, csv_file_path
Output: CSV file with trench layer distances and unmatched counts
Function process_trenches(trench_layer_names, klicpoints_layer, csv_file_path):
    d ← Initialise distance calculator;
    all_results ← Initialise empty list;
    unmatched_counts ← Initialise empty dictionary;
    foreach trench_layer_name in trench_layer_names do
        trench_layer ← Get trench layer by name;
        d.setSourceCrs(trench_layer.crs());
        results ← Initialise empty list;
        unmatched_count ← 0;
        foreach trench_feat in trench_layer.getFeatures() do
            trench_point ← Get geometry as point;
            description ← Get trench feature description;
            matched ← False;
            request ← Create filter request for KLICPOINTS within 4 meters;
            foreach klic_feat in klicpoints_layer.getFeatures(request) do
                if klic_feat['layer_name'] == description then
                    klic_point ← Get geometry as point;
                    distance ← Calculate distance between points;
                    Append (trench_layer_name, description, distance) to results;
                    matched ← True;
                end
            end
            if not matched then
                unmatched_count += 1;
            end
        end
        unmatched_counts[trench_layer_name] ← unmatched_count;
        all_results.extend(results);
    end
    Sort all_results by distance;
    Write results to CSV: begin
        Open csv_file_path for writing;
        Write headers ['Trench', 'Description', 'Distance (m)'];
        foreach result in all_results do
            Write result to CSV;
        end
    end
    Print "Results saved to " + csv_file_path;
    foreach trench, count in unmatched_counts do
        Print "Number of unmatched " + trench + " points: " + count;
    end
return
Run process_trenches(trench_layer_names, klicpoints_layer, csv_file_path);
```

4. Implementation

This chapter describes the practical application of the methodologies discussed in the previous section. The implementation chapter details the processes employed in transforming individual trench models into actionable insights through a series of structured tasks including, data extraction, manipulation, and analysis. Using specific examples from both cities, Enschede and Rotterdam, the chapter explores the use of QGIS in conjunction with programming tools that facilitate the handling and visualisation of enriched utility data.

4.1. Tools used

1. **QGIS 3.34.6** [QGIS, 2024]: QGIS (Quantum GIS) is an open-source geographic information system that will be used for data management, analysis, and visualization. QGIS provides a wide range of functionality, including spatial data processing, geoprocessing. It will be employed for tasks such as data integration, geo-referencing, and creating basic visualizations of the utility network data.
2. **AutoCAD 2024** [Autodesk Inc., 2024]: AutoCAD is a widely used CAD software that is particularly well-suited for creating 2D and 3D models. It will be utilized for working with the original 2D utility network maps, extracting relevant data.
3. **Python 3.9.7** [Python Software Foundation, 2024]: Python is a powerful, general-purpose programming language widely employed for various applications, including data science, scripting, and automation. In this thesis, Python will be utilized for scripting and automating tasks, providing efficient data processing and analysis.

In addition to these main tools, several libraries were used in python scripts to either export, import or calculate information with given data. These additional libraries can be seen in Table 4.1.

4.2. Enschede

The municipality of Enschede provided two different areas of study for this thesis, given all the data is structured similarly between the two, this implementation section works for both datasets. As mentioned in Section 3.2, the data provided by Enschede needed not only to be extracted from the DXF file but additionally calculations had to be made to find the actual elevation value in NAP. Using Algorithm 3.1, the cable data for each trench in the DXF file was extracted, this data is shown in Table 4.2. Additionally, the start and end coordinates of each trench line have been extracted using Algorithm 3.2, as seen in Table 4.3. The following sub-section discusses how the coordinates of the line are combined with the coordinates of the individual cables in order to calculate the position of each cable.

Programming Language	**Library**	**Functionality**
Python	ezdxf	Reading and extracting DXF information, providing tools for manipulating DXF drawings.
	NumPy	Used for storing positions as numpy arrays and performing array-wise operations.
	openpyxl	Creates and saves an Excel workbook containing extracted coordinates.
	pandas	Structures extracted data into a DataFrame and exports it to an Excel file.
	qgis.core	Manages and manipulates spatial data, facilitating the addition of CSV data as vector layers, calculation of geometric points, and the dynamic creation and editing of line features.
	SciPy	Used to compute distances between two collections of points.

Table 4.1.: Overview of Python libraries used and their functionalities

Proefsleuf	Distance (m)	Depth (m)	Description
1	0.07	1.45	WATER PVC 315
1	0.69	1.4	HDG PE 160
1	1.74	1.11	WATER PVC 160
1	1.78	0.43	CAI 40HDPE NA
1	2.01	0.9	LDG PVCSV 110
1	2.11	0.73	LS KABEL NA
1	2.15	0.67	LS KABEL NA
1	2.22	0.72	LS KABEL NA
1	2.37	0.68	DATA TEERKABEL NA
1	2.4	0.61	LS KABEL NA
1	2.85	0.38	DATA 40HDPE NA
1	3.23	0.49	LS KABEL PVC 125
1	5.41	0.53	LS KABEL NA
1	6.07	0.27	LS KABEL NA
1	7.44	0.38	CAI KABEL PVC 75

Table 4.2.: Utility trench 1: extracted DXF data

Start X	Start Y	End X	End Y	Length	Start Z	End Z
257787.9	471520.94	257789.88	471513.04	8.14653	40.177	40.433

Table 4.3.: Utility trench 1: extracted geographical coordinates of the trench line

4.2.1. Calculating cable positions

Following the methodology set in Chapter 3, the extracted DXF data (Table 4.2) and the extracted positional data (Table 4.3) need to be used to calculate the positional data of each cable in a trench. Using the interpolation formulae from Chapter 3 and the process described in Algorithm 4.1, we are able to interpolate positional values for every cable and pipeline. These interpolated values enable us to create a precise three-dimensional mapping of each cable within the trench, crucial for subsequent steps. An example of the result of these calculations can be seen in Table 4.4. These values are exported to CSV files that are later imported in QGIS.

Algorithm 4.1: Combine trench and positional data

```
Input: input_file_path, output_file_path
Output: Excel file with calculated E, N, and Z values
Function transform_data(input_file_path, output_file_path):
    df_start ← Read Excel file(input_file_path);
    Function calculate_e(start_x, end_x, distance):
        return E ;
    return
    Function calculate_n(start_y, end_y, distance):
        return N ;
    return
    Function calculate_z(start_z, end_z, depth):
        return Z ;
    return
    foreach row in df_start do
        row['E'] ← calculate_e;
        row['N'] ← calculate_n;
        row['Z'] ← calculate_z;
    end
    Save df_start to Excel(output_file_path);
    Print "Data has been transformed and saved to " + output_file_path;
return
Run transform_data(input_file_path, output_file_path);
```

4.2.2. Matching cable names between successive trenches

As described in Chapter 3, it is necessary to ensure the names of cables match one another in order to connect successive trenches. This sub-section describes the decisions made when it comes to matching the trench descriptors.

An example of two successive trenches is depicted in Table 4.5. It can be seen in the table that there are some cables names that are exactly the same (in green), some similar (in yellow) and some that have no match (in red). If one were to connect the tables as is, that would result in fewer connections than in reality. For example, in Table 4.5, the cable "LS KABEL BUNDEL" is a bundle of grouped cables, it can be seen that there are 6 iterations of "LS

Proefsleuf	Distance	Depth	Description (Material)	Diameter (mm)	E	N	Z
1	0.07	1.45	WATER PVC	315	257787.92	471520.8761	38.7292
1	0.69	1.4	HDG PE	160	257788.0701	471520.2745	38.7987
1	1.74	1.11	WATER PVC	160	257788.3244	471519.2558	39.1217
1	1.78	0.43	CAI 40HDPE	NA	257788.3341	471519.217	39.8029
1	2.01	0.9	LDG PVCSV	110	257788.3898	471518.9938	39.3402
1	2.11	0.73	LS KABEL	NA	257788.414	471518.8968	39.5133
1	2.15	0.67	LS KABEL	NA	257788.4237	471518.858	39.5746
1	2.22	0.72	LS KABEL	NA	257788.4407	471518.7901	39.5268
1	2.37	0.68	DATA TEERKABEL	NA	257788.477	471518.6446	39.5715
1	2.4	0.61	LS KABEL	NA	257788.4843	471518.6155	39.6424
1	2.85	0.38	DATA 40HDPE	NA	257788.5932	471518.1788	39.8866
1	3.23	0.49	LS KABEL PVC	125	257788.6853	471517.8102	39.7885
1	5.41	0.53	LS KABEL	NA	257789.2132	471515.6951	39.8170
1	6.07	0.27	LS KABEL	NA	257789.3731	471515.0547	40.0977
1	7.44	0.38	CAI KABEL PVC	75	257789.7049	471513.7255	40.0308

Table 4.4.: Utility trench 1: enriched data table

Trench 1	Trench 2
WATER /315PVC	WATER /315PVC
HDG /200PE	HDG /160PE
CAI C3 COAX	WATER /160PVC
CAI BUNDEL COAX	CAI 40HDPE
DATA TEERKABEL	LDG /110PVCSV
DATA TEERKABEL	LS KABEL
DATA TEERKABEL	LS KABEL
DATA TEERKABEL	LS KABEL
WATER /160PVC	DATA TEERKABEL
LS KABEL BUNDEL	LS KABEL
GV 16DB	DATA 40HDPE
LS KABEL /110	LS KABEL /125 PVC
DATA HDPE	LS KABEL
GV 7DB: 7X14DB	LS KABEL
DATA HDPE	CAI KABEL /75 PVC
CAI 40COAX	
MS TEERKABEL	
CAI 40COAX	

Table 4.5.: Comparison of cable descriptions: Brammelerdwarsstraat trench 1 & trench 2

KABEL" that can be assumed to belong to this bundle. Therefore, for the purpose of this thesis, the similar cables names will be matched in order to correct such issues.

As part of the analysis, the matching for each pair of trenches will be visualised as pie charts, an example of this is seen in Figure 5.2. Figure 5.2 shows that without matching cables, only 31% of 33 cables and pipelines between the two trenches can be matched. If it is assumed that the similarly named cables match one another, 79% of cables are linked and will be displayed in the final visualisation.

4.2.3. Importing and connecting trenches

This section outlines the process implemented in QGIS using python scripting. The process involves loading the enriched data from CSV files, calculating intermediate points based on the centroid of same-named cables, and subsequently creating and modifying line features that represent connections between two sets of trench points.

Loading CSV data as spatial layers

The first task to perform is to import the enriched data into QGIS. This can normally be done manually by adding a comma delimited layer in QGIS. However, for the purpose of efficiency, it is necessary to create a script that can handle importing all trenches.

Spatial layers are created from CSV files using the 'add_csv_as_layer' function (Algorithm 3.3), which constructs a URI that includes the file path, delimiters, coordinate fields, Coordinate

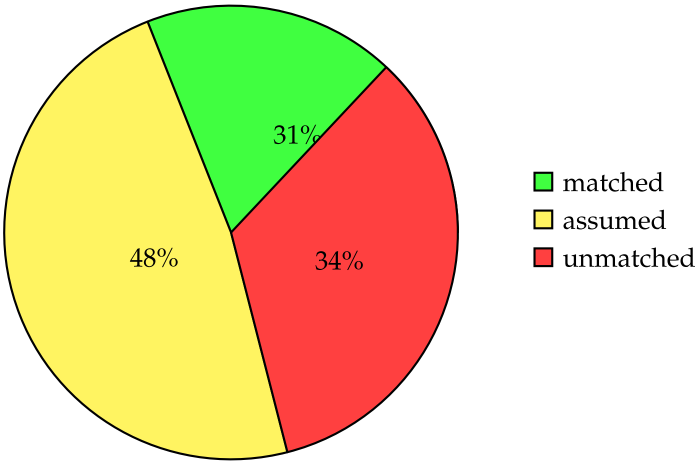
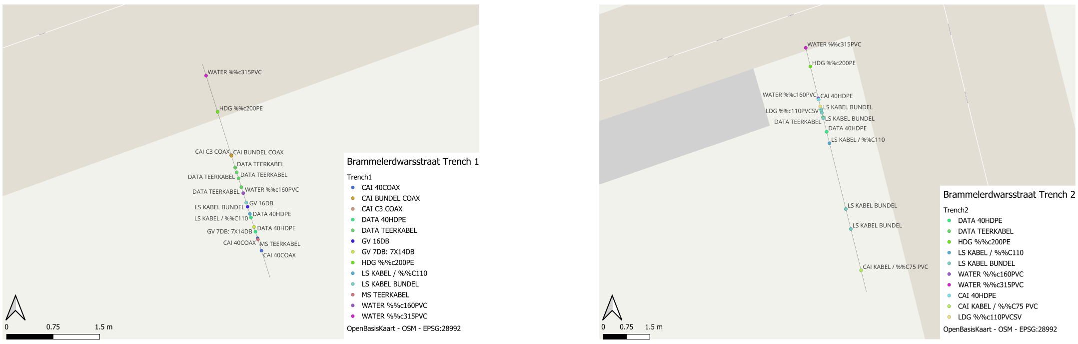


Figure 4.1.: Distribution of cable status for trenches 1 & 2

Reference System (CRS) ID, and field types. These layers are then added to the QGIS project for further manipulation. The outcome of this process can be seen in Figure 4.2.



(a) Brammelerdwarsstraat: Trench 1

(b) Brammelerdwarsstraat: Trench 2

Figure 4.2.: Enriched data trenches imported in QGIS

Validation of enriched data

Specifically for the data provided by Enschede, we were able to perform a validation of the enriched data by comparing the points with those provided by the surveyors. The validation was simply performed by calculating the distance of the enriched points calculated by us and the points drawn on the utility trench cross-section by the surveyors. This is exemplified by Figure 4.3, where the red lines were drawn by the surveyors.

Figure 4.3.: Surveyed cable positions (intersection of red line), compared to our calculated positions

Calculating average points

As discussed in Chapter 3, the chosen method of visualisation is using a ”forked” cable design. This requires some additional calculations and the creation of intermediate points from which the individual cables split off from/. The 'calculate_average_point' (Algorithm 3.4) function computes the centroid (average point) of the destination points. This function is important for determining the position of the intermediate points, which as mentioned, act as the transitional nodes in a 1-to-many situation.

Determining intermediate points

Between two trenches, there can be a 1-to many in either direction. Therefore, it is necessary to adjust the script so that the intermediate points are created closer to the destination trench in order to reduce clutter. The 'calculate_dynamic_intermediate_point' function (Algorithm 3.5), calculates an intermediate point between the starting trench and the centroid previously created. It contains a bias ratio that is determined by the number of points on either side. This dynamic adjustment allows for the intermediate points to be placed closer to the trench with many cables (in a 1-to-many situation). The bias ratio can then be adjusted to increase or decrease the distance from the intermediate point and the destination trench. In Figure 4.4, the intermediate points from which the cables split off from are highlighted for the connections between trench 1 and 2.

Establishing connections

With the intermediate point in place, the cable connections are visualized by creating line features that link each point in one trench to the calculated intermediate point and then to the corresponding points in the second trench. This multi-step connection emphasizes the relational aspect of the cable connections, and provides a cleaner visualisation of the complex relationships between trenches (Figure 4.4).

Figure 4.4.: Connections between Brammelerdwarsstraat trench 1 & 2

Automating the process

The final step of is to automate the process as much as possible. In order to achieve this, all scripts are combined with the additional option of adding as many successive trenches as needed. The algorithm works in a way where all the trenches are loaded in and then are linked one-by-one. A pseudo-code describing the full script can be seen in Algorithm 4.2. Both sets of linear cable connections in Figure 4.5 were generated in a semi-automatic manner. The script takes the names of the trenches as an input and outputs cable connections through two or more desired trenches. Note that, this algorithm only works for linearly successive trenches and in the example of Figure 4.5, the script was utilised twice, one time for each side of the street.

4.3. Rotterdam

The utility trench data provided by Rotterdam is largely similar to that of Enschede. Some minor changes make it so that the extraction process will require a different script than that of Enschede. Once the data is extracted and structured in the same format as the enriched Enschede data, the process for importing and drawing cables in QGIS uses the exact same code.

Again, like Enschede, the the data collected is in two parts at first. The positional data of the actual trench, as seen in Table 4.7, and the individual cable data (Table 4.6). It can be seen in Table 4.6 that some cleaning is required before proceeding to match trenches to one another. For example, the entry ”102 - GRAAFDIEPTE” is in fact not a cable but the measurement

Algorithm 4.2: Dynamic Point Connection in GIS

```
Input: Trench file paths, field names (x, y, z), CRS EPSG code, field types
Output: Connected lines between trenches stored in a GIS layer
Function AddCSVAsLayer:
    Input: file path, layer name, x field, y field, z field, CRS EPSG, field types
    Output: Layer object
    Construct URI with file path and parameters;
    return Add vector layer to GIS project using URI;
return
Function CalculateAveragePoint:
    Input: List of points
    Output: Average point as QgsPointXY
    Calculate x, y coordinates mean;
    return Point created from averages;
return
Function CalculateDynamicIntermediatePoint:
    Input: Points from two trenches
    Output: Intermediate point as QgsPointXY
    Calculate average points for each trench;
    Determine bias ratio based on point counts;
    Calculate intermediate point using bias ratio;
    return New intermediate point;
return
Procedure ConnectTrenches
    Input: List of trench file paths, field settings
    Initialize GIS lines layer with attributes;
    Load each trench layer sequentially;
    For each pair of trench layers;
    Get matching descriptions;
    For each description;
    Calculate intermediate point;
    Create lines from points in trench 1 to intermediate;
    Create lines from intermediate to points in trench 2;
    Commit changes to lines layer;

Main Execution:;
field names, CRS EPSG, field types ← Define field settings;
trenchPaths ← Define list of trench file paths;
ConnectTrenches(trenchPaths, field names, CRS EPSG, field types);
```

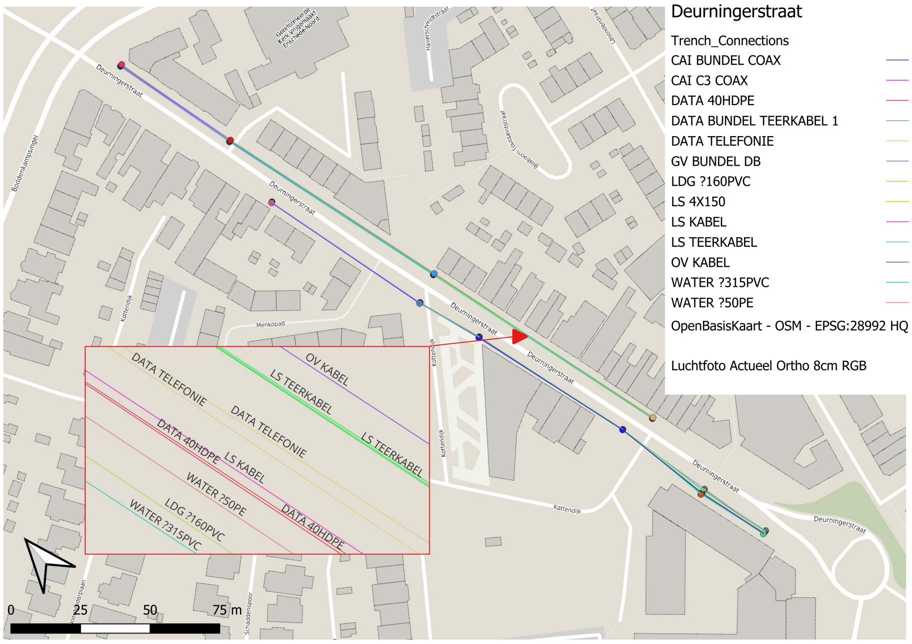


Figure 4.5.: Result of semi-automatic trench connection script in Deurningerstraat

of the trench depth. Additionally, in the description, every cable has a unique identifier preceding the actual description information. These IDs need to be removed in order to match two cables of the same name. Once the data cleaning is completed, the positional data and the cable data can be combined. It is important to note that, for the Rotterdam data, the linear interpolation for the cable positions is identical to that of Enschede **except** for the Z value which is already provided as NAP in the surveyed data. Table 4.8 presents an example of enriched trench data post cleaning.

4.3.1. Matching trenches

Following the same procedure as Enschede, the Rotterdam trenches need to be checked for matching issues and potential assumptions. Fortunately, the data provided by Rotterdam more organized with its naming. Meaning, that if there are two identical cables in either trench they will be given the same name. One of the issues, however, is that only a small percentage of the cables in each trench have been identified. This means that, for the majority of cables, there is not enough information to make a visualised link between trenches. Table 4.9 displays the links between two trenches and Figure 4.6 shows that only 26% of the total cables between the two trenches could be matched. More examples of this will be displayed in Chapter 5.

4.3.2. Importing and connecting trenches

Given that the enriched data is in the same format for both Enschede and Rotterdam, importing said data uses the same python scripts. The only changes that need to be made

Group	Distance (m)	Depth (NAP)	Description
G01A	4.39	-4.68	116
G01A	4.24	-5.78	115
G01A	3.96	-5.48	114 - 50 PE W ONBEKEND
G01A	3.51	-5.64	113 - 160 PVC W
G01A	3.47	-4.67	112
G01A	1.88	-4.72	111
G01A	1.85	-5.68	110 - 316 GY G
G01A	1.72	-5.92	109
G01A	1.58	-5.15	108 - 1 HDPE G
G01A	1.38	-4.74	107
G01A	1.35	-5.29	106 - 1 OV E
G01A	0.98	-6.28	105
G01A	0.87	-6.56	104- 400X600 T R GEPRIKT
G01A	0.84	-4.8	103
G01A	0.36	-6.36	102 - GRAAFDIEPTE
G01A	0.00	-4.9	101

Table 4.6.: Tabular representation of data

Trench	Start X	Start Y	End X	End Y	Length (m)
G01A	92341.66	92339.02	442108.19	442110.35	3.41

Table 4.7.: Rotterdam: utility trench G01A positional data

Proefsleuf	Width	Depth	Description	E	N	Z
G99	3.65	-5.89	9914	92289.27006	441710.545	-5.89
G99	3.60	-5.17	9913	92289.2384	441710.5063	-5.17
G99	3.46	-5.17	KUNSTOF MAT ZIGGO	92289.14974	441710.398	-5.17
G99	3.30	-5.24	1 CU KPN	92289.04842	441710.2741	-5.24
G99	3.03	-5.12	9910	92288.87744	441710.0652	-5.12
G99	2.98	-5.45	160 PVC W	92288.84578	441710.0265	-5.45
G99	2.90	-5.12	KABELBED REGGEFIBER	92288.79512	441709.9645	-5.12
G99	2.44	-5.38	160 PVC G	92288.50382	441709.6085	-5.38
G99	1.94	-5.09	1 LS E	92288.1872	441709.2216	-5.09
G99	1.76	-4.59	9905	92288.07321	441709.0822	-4.59
G99	0.41	-4.62	9904	92287.21832	441708.0374	-4.62
G99	0.29	-5.79	400X600 BT R	92287.14233	441707.9445	-5.79
G99	0.00	-4.67	9901	92286.95869	441707.7201	-4.67

Table 4.8.: Rotterdam: Enriched utility trench data

Trench G19	Trench G21
1922	2123
1921	CU KPN
CO ZIGGO LOS	2121
AFDEK ZIGGO KUNSTSTOF	NIET GEGRAVEN
1918	2119
1917	NIET GEGRAVEN
CU KPN	2117
1915	2116
1914	2115
1913	AFDEK ZIGGO KUNSTSTOF
1912	2113
KABELBED KPN ZIGGO	KABELBED KPN ZIGGO
1910	BUNDEL REGEFIBER
BUNDEL REGEFIBER	2110
500 BT R GEPRIKT	2109
1907	2108
1906	2107
HDPE KPN	2106
1904	HDPE KPN
1903	2104
1901	500 BT R GEPRIKT
	2102
	2101

Table 4.9.: Rotterdam: Matching Trench G19 and Trench G21

to the script are the input files. Figure 4.7 shows a singular trench imported in QGIS and labelled, Figure 4.8 displays the connections between two trenches.

4.3.3. Analysis

In order to compare the enriched data to that of the KLIC and Rotterdam's LVZK, these both need to be imported into the respective QGIS projects. This can be done simply by using the 'DWG/DXF Import' tool in QGIS. The best measure to compare the enriched data with the provided data is to calculate the distance between points of the same cable. As the enriched data was calculated along a trench line, it was decided to compare those points to the KLIC and LVZK data along that same line.

In order to do this, another python script (Algorithm 3.7) was developed. This script takes all the lines of the imported DXF file and places their intersection with the trench line as points (visualised in Figure 4.9). Unfortunately, as can be seen in Figure 4.9, the distance calculation process is hindered by the fact that the cables do not have the same naming conventions as the cables identified in the trenches. There is too much of a difference between the names to create a fully automatic process for matching the trenches. An example of enriched points that can be associated with the KLIC points can be seen in Figure 4.10. Here, it is relatively safe to assume that 'B-PV-KL-WATER_315_VITENS-G' can be associated to the trench point 'WATER 315PVC'. Figure 4.11, shows an enriched point 'LS KABEL BUNDEL' that has been

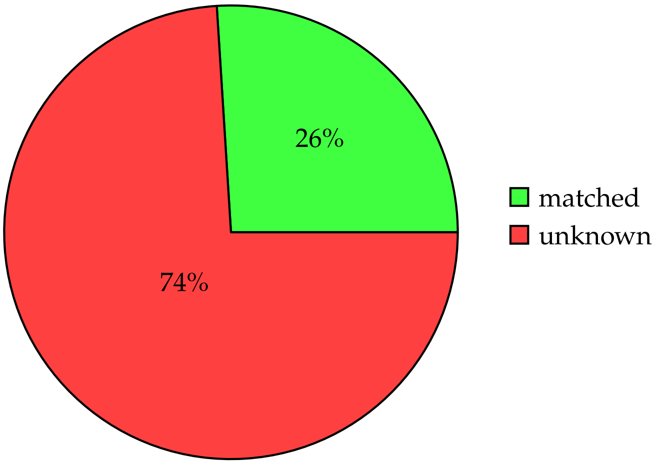


Figure 4.6.: Distribution of cable status for trenches G19 & G21

marked by the surveyors as 'Onbekend' or 'Unknown', this is because the surveyors were unable to find this cable on the KLIC map.

Given this, we can attempt to associate as many cable points together as possible and then calculate the distance between these points. Table 4.10 contains these results for 'Trench 1' in Brammelerdwarsstraat, Enschede. It also contains the percentage of enriched points that were able to be connected to KLIC points. The same process was able to be applied on the Rotterdam data, resulting in Table 4.11. This analysis was repeated for all trenches in Enschede and Rotterdam.

Trench	Description	Distance (m)	Points compared
1	WATER / 315PVC	0.0341	27.8%
	LS KABEL / C110	0.0838	
	WATER / c160PVC	0.206	
	HDG / 200PE	0.2417	
	CAI C3 COAX	0.7401	

Table 4.10.: Enschede: Trench 1 distances

Trench	Description	Distance (m)	Points compared
G76	110 PVC G	0.1582	23.5%
	CU KPN	0.5297	
	400X600 BT R	0.8065	
	160 PVC W	0.8523	

Table 4.11.: Rotterdam: Trench G76 distances

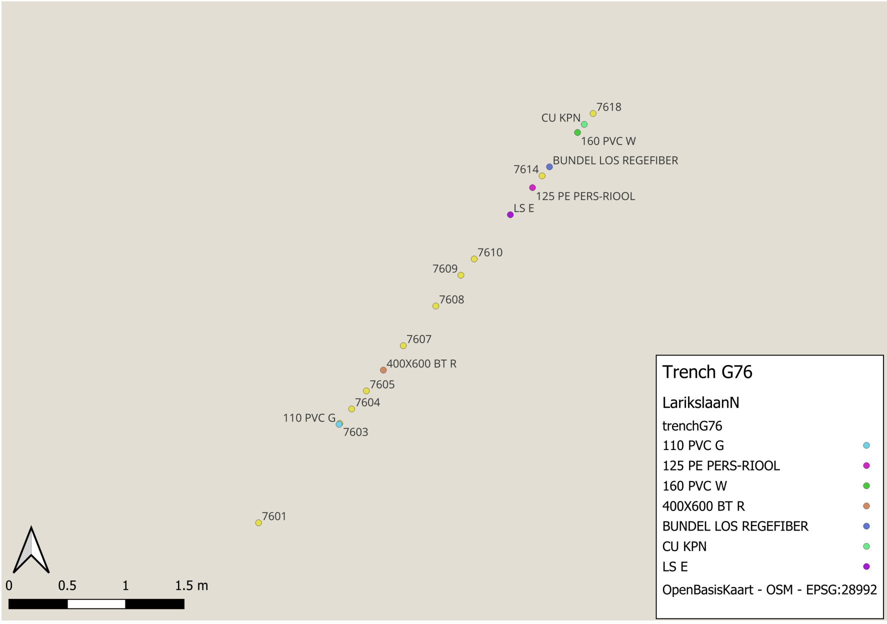


Figure 4.7.: Rotterdam: trench G76 imported

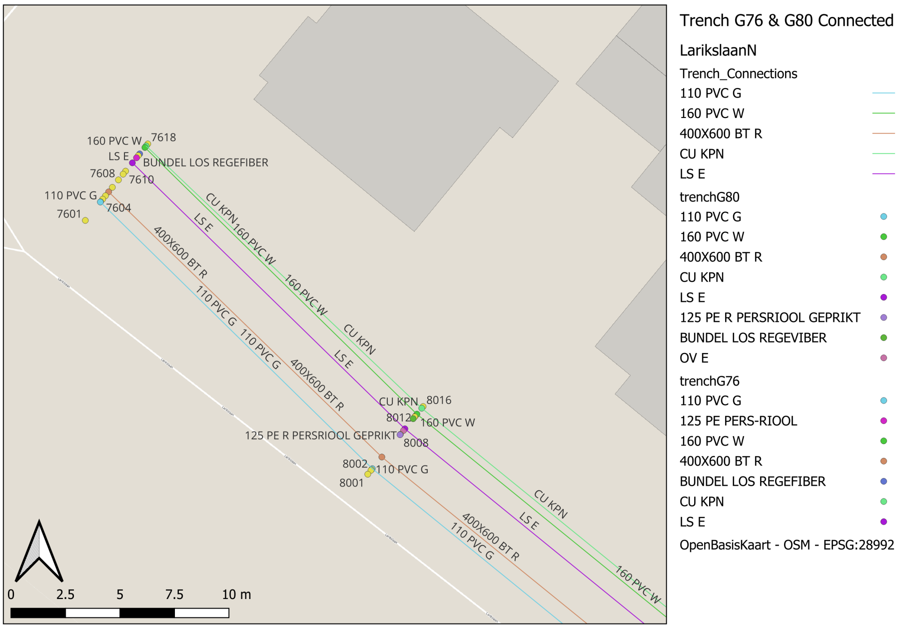


Figure 4.8.: Rotterdam: trenches G76 & G80 connected

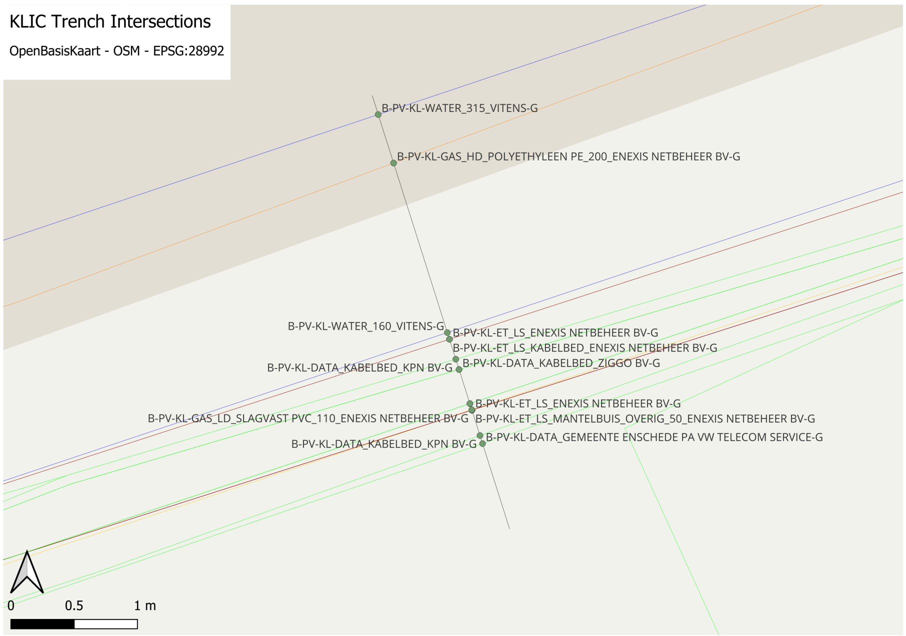


Figure 4.9.: Enschede: Intersection points with trench line and KLIC cables

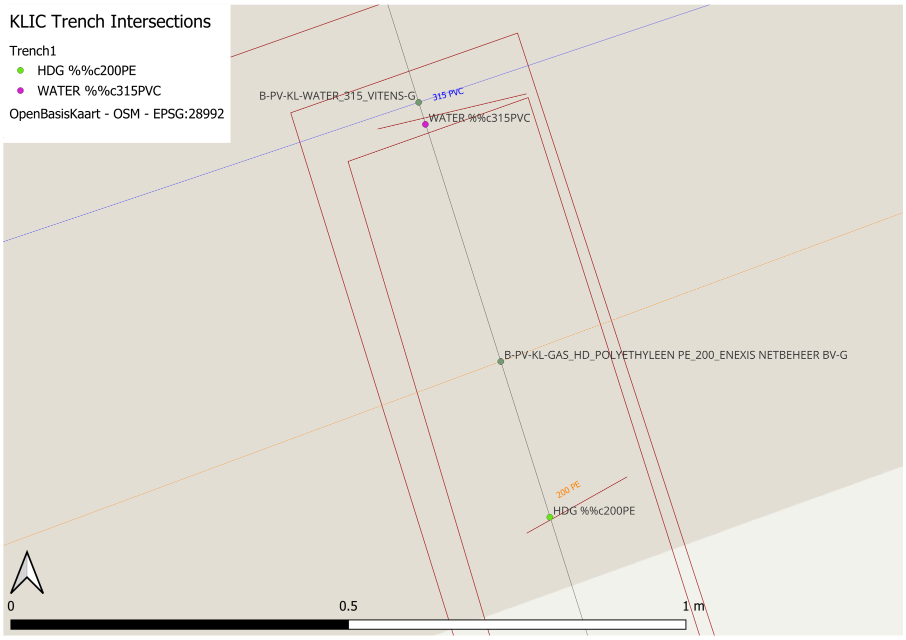


Figure 4.10.: Enschede: Example of KLIC and enriched data that can be matched

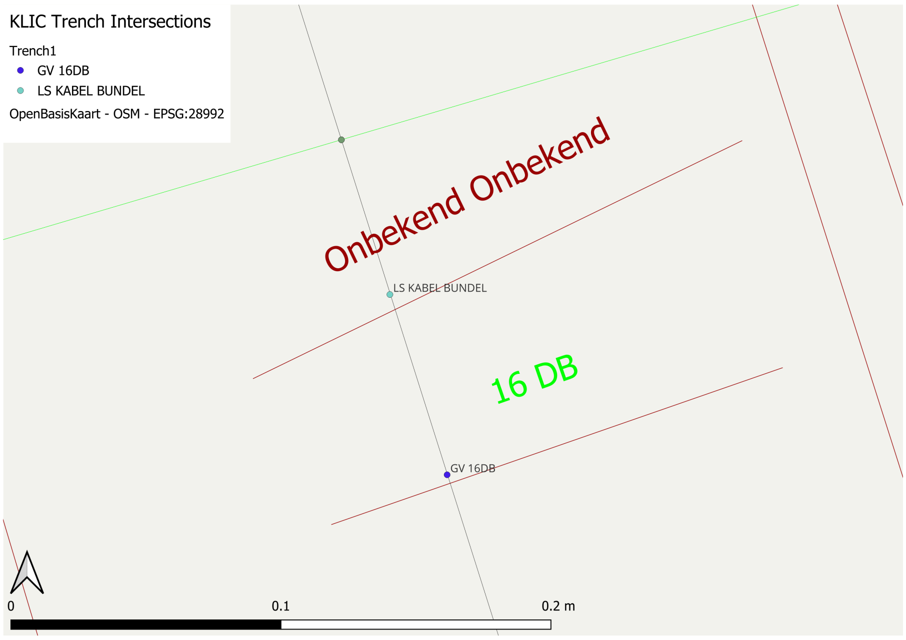


Figure 4.11.: Enschede: Example of KLIC and enriched data that can't be matched

5. Results

This chapter presents the research results in a summarised manner. Each section will describe the results of a case study. The results will contain: the enriched utility data, visualisations of trench connections, pie charts describing the matching analysis, KLIC or LVZK offset data and finally a visualisation of the offsets. Additional relevant figures and scripts will displayed in the appendix. The fourth section evaluates how the enriched data affects the usability and reliability of utility network information for urban planning and management purposes.

5.1. Brammelerdwarsstraat, Enschede

This section presents the results regarding Brammelerdwarsstraat in Enschede. Tables 5.1-5.6 represent the enriched data for each trench. These tables contain the coordinates for each individual cable. The validation of the enrich data found that the average distance between the enriched data points and those drawn by the surveyors was 0.007 m, meaning that our points on average have less than a centimeter of offset when compared to the surveyed data. Figure 5.1 depicts the connections between each of the trench pairings, followed by Figure 5.2-5.4 that displays the matching status between two trench pairs and the total of Brammelerdwarsstraat. Table 5.7 contains the enriched data comparison to the KLIC cables. This table shows the percentage of enriched points that found a match in the KLIC data along with the distance of offset from their respective cables/pipes. For Brammelerdwarsstraat the average distance offset for all matched trenches was found to be 0.30 m with an average of 47.2% of points being used. Visualisations of the offsets can be seen in Figure 5.5 and in additional figures in the appendix (Fig. A.1-A.5).

Trench	Description	E	N	Z
1	WATER 315PVC	257787.92	471520.8761	38.7292
	HDG 200PE	257788.0701	471520.2745	38.7987
	WATER 160PVC	257788.3244	471519.2558	39.1217
	CAI 40HDPE	257788.3341	471519.217	39.8029
	LDG 110PVCSV	257788.3898	471518.9938	39.3402
	LS KABEL BUNDEL	257788.414	471518.8968	39.5133
	LS KABEL BUNDEL	257788.4237	471518.858	39.5746
	LS KABEL BUNDEL	257788.4407	471518.7901	39.5268
	DATA TEERKABEL	257788.477	471518.6446	39.5715
	LS KABEL BUNDEL	257788.4843	471518.6155	39.6424
	DATA 40HDPE	257788.5932	471518.1788	39.8866
	LS KABEL / 110	257788.6853	471517.8102	39.7885
	LS KABEL BUNDEL	257789.2132	471515.6951	39.8170
	LS KABEL BUNDEL	257789.3731	471515.0547	40.0977
	CAI KABEL / 75 PVC	257789.7049	471513.7255	40.0308

Table 5.1.: Brammelerdwarsstraat: Trench 1 enriched data

Trench	Description	E	N	Z
2	WATER 315PVC	257841.3145	471539.4259	39.4239
	HDG 200PE	257841.4989	471538.8445	39.4868
	CAI C3 COAX	257841.7197	471538.1486	39.9502
	CAI BUNDEL COAX	257841.7227	471538.1391	39.9703
	DATA TEERKABEL	257841.7832	471537.9485	39.8412
	DATA TEERKABEL	257841.8074	471537.8722	39.9016
	DATA TEERKABEL	257841.8376	471537.7769	39.9121
	DATA TEERKABEL	257841.883	471537.6339	40.0128
	WATER 160PVC	257841.9132	471537.5386	39.4633
	LS KABEL BUNDEL	257841.9616	471537.3861	40.0540
	GV 16DB	257841.9828	471537.3194	40.0744
	LS KABEL / 110	257842.0191	471537.205	40.0449
	DATA 40HDPE	257842.0372	471537.1478	39.9652
	GV 7DB: 7X14DB	257842.0856	471536.9953	40.0760
	DATA 40HDPE	257842.1098	471536.919	40.0463
	CAI 40COAX	257842.1431	471536.8142	40.0369
	MS TEERKABEL	257842.1491	471536.7951	39.9070
	CAI 40COAX	257842.2066	471536.614	39.9879

Table 5.2.: Brammelerdwarsstraat: Trench 2 enriched data

Trench	Description	E	N	Z
4	WATER /160PVC	257744.759	471473.9824	38.58291251
	CAI BUNDEL COAX	257744.9343	471474.0232	39.14396102
	LS TEERKABEL	257745.1291	471474.0686	39.28512602
	LDG /160PVC	257745.207	471474.0867	38.53559202
	LS AL	257745.2362	471474.0935	38.92291251
	CAI BUNDEL COAX	257745.4797	471474.1502	38.88722303
	GV 25DB	257745.5673	471474.1706	39.17774728
	GV 25DB	257745.6258	471474.1842	39.17809679
	DATA HDPE	257745.8401	471474.2341	39.01937829
	LS KABEL	257746.0251	471474.2772	39.31048505
	DATA TELEFONIE	257746.0348	471474.2795	39.3005433

Table 5.3.: Brammelerdwarsstraat: Trench 4 enriched data

Trench	Description	E	N	Z
6	WATER /160PVC	257753.0299	471435.6082	38.40721632
	CAI 40COAX	257753.3232	471435.6714	39.1230007
	LDG /160PVCSV	257753.4014	471435.6883	38.4445432
	MS 2* BUNDEL	257753.5578	471435.7221	38.5876282
	LS KABEL	257753.5871	471435.7284	38.76820664
	GV 2* BUNDEL DB	257753.636	471435.7389	39.05917071
	LS KABEL	257753.7142	471435.7558	38.89071321
	DATA HDPE	257753.8608	471435.7874	38.6736054
	DATA HDPE	257754.0465	471435.8275	38.71726884
	RIOOL /200PVC	257754.5646	471435.9393	38.52748792

Table 5.4.: Brammelerdwarsstraat: Trench 6 enriched data

Trench	Description	E	N	Z
8	CAI 40COAX	257796.9391	471421.0307	40.37796365
	OV KABEL	257797.6067	471421.16	40.05725122

Table 5.5.: Brammelerdwarsstraat: Trench 8 enriched data

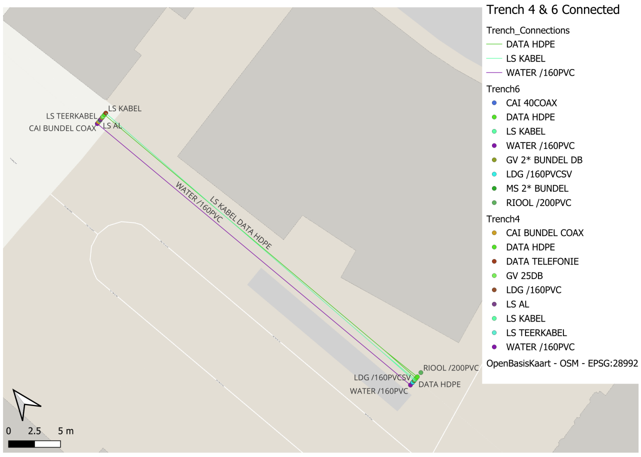
Trench	Description	E	N	Z
9	DATA HDPE	257804.642	471392.7206	39.57346683
	OV KABEL	257804.945	471392.7862	39.18998828
	6* LS BUNDEL	257805.1893	471392.8392	39.46847332
	CAI 40COAX	257805.3555	471392.8752	39.17624314
	CAI C3 COAX	257805.3946	471392.8836	39.30160075
	WATER /54CU	257806.0689	471393.0297	38.70901945
	LDG /110PVC	257806.3132	471393.0827	38.63750449

Table 5.6.: Brammelerdwarsstraat: Trench 9 enriched data

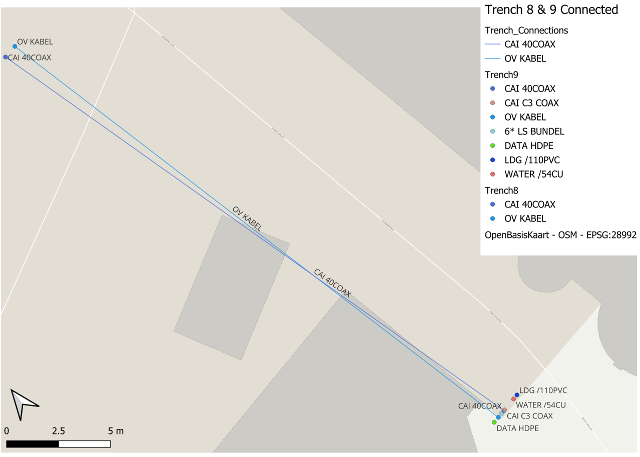
Trench	Description	Distance (m)	Points used
Trench 1	WATER 315PVC	0.03	27.8%
	LS KABEL / 110	0.08	
	WATER 160PVC	0.21	
	HDG 200PE	0.24	
	CAI C3 COAX	0.74	
Trench 2	WATER 315PVC	0.02	42.9%
	HDG 200PE	0.05	
	WATER 160PVC	0.07	
	LDG 110PVCSV	0.10	
	DATA 40HDPE	0.13	
	DATA TEERKABEL	0.14	
	CAI KABEL / 75 PVC	0.45	
Trench 4	GV 25DB	0.002	61.5%
	GV 25DB	0.01	
	GV 25DB	0.06	
	GV 25DB	0.07	
	LDG /160PVC	0.11	
	WATER /160PVC	0.36	
	CAI BUNDEL COAX	0.42	
	CAI BUNDEL COAX	0.98	
Trench 6	DATA HDPE	0.01	58.3%
	LDG /160PVCSV	0.01	
	DATA HDPE	0.01	
	CAI 40COAX	0.03	
	DATA HDPE	0.18	
	DATA HDPE	0.18	
	RIOOL /200PVC	0.20	
Trench 8	CAI 40COAX	0.74	50.0%
Trench 9	CAI C3 COAX	0.07	42.9%
	CAI 40COAX	0.11	
	WATER /54CU	0.12	
Average		0.30	47.2%

Table 5.7.: Brammelerdwarsstraat enriched data compared to KLIC data

(a) Brammelerdwarsstraat: trenches 1 & 2 final connections



(b) Brammelerdwarsstraat: trenches 4 & 6 final connections



(c) Brammelerdwarsstraat: trenches 8 & 9 final connections

Figure 5.1.: Trench connections between enriched cable points. Brammelerdwarsstraat, Enschede.

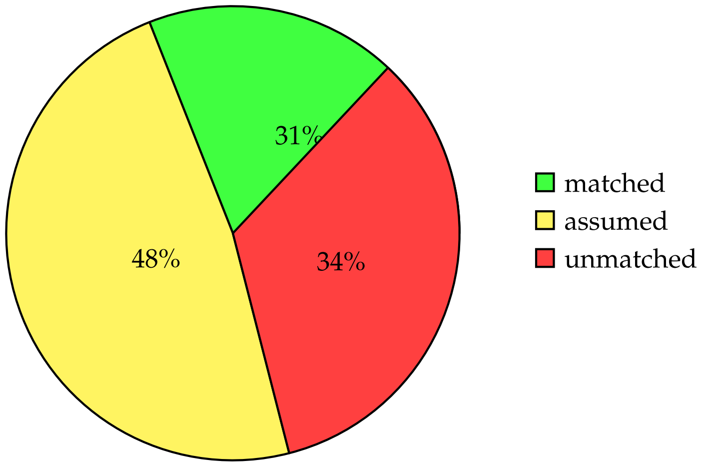


Figure 5.2.: Brammelerdwarsstraat: Cable status for trenches 1 & 2

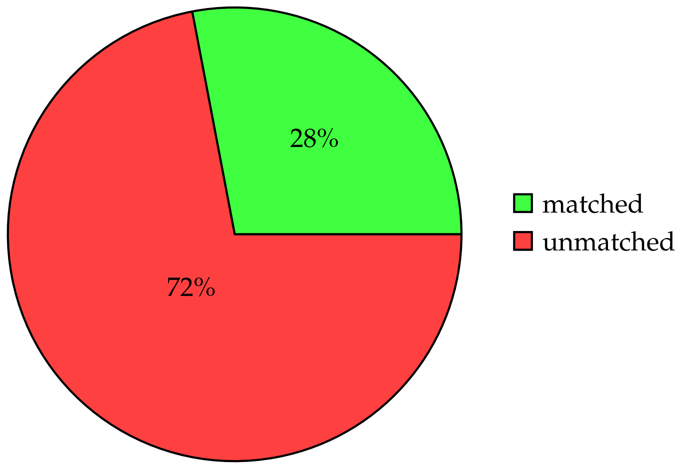


Figure 5.3.: Brammelerdwarsstraat: Cable status for trenches 8 & 9

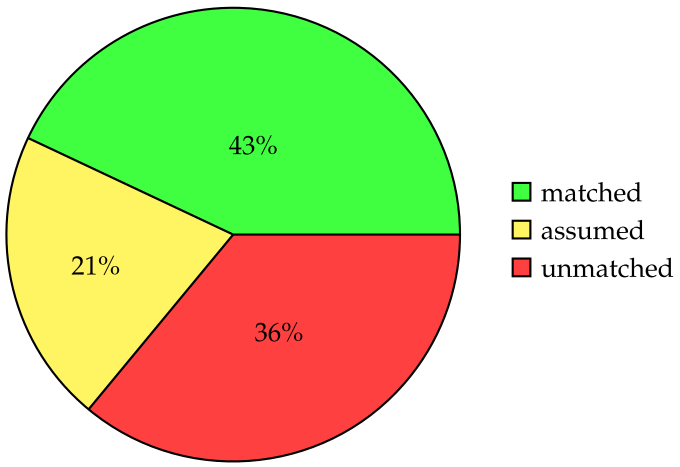


Figure 5.4.: Brammelerdwarsstraat: Total Distribution of cable status

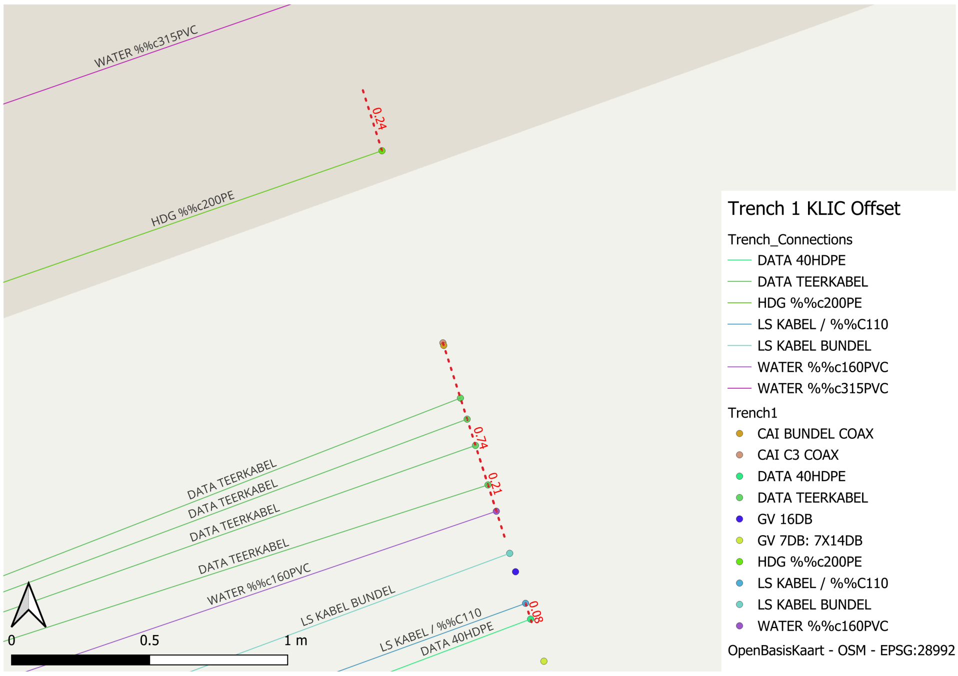


Figure 5.5.: Brammelerdwarstraat: KLIC offset, trench 1

5.2. Deurningerstraat, Enschede

In this section, Figures 5.8-5.17 display all the utility trench enriched data in Deurningerstraat. Figure 5.8 depicts all connection in the case study. As with Enschede, this is followed by pie charts representing the distribution of matched cable statuses. It can be seen in Figure 5.11, that 36.3% of cables in all of Deurningerstraat were able to be matched without any assumptions or edits. This value is increased to 78.9% by incorporating assumptions. Tables 5.18-5.19 show that, for this case study, an average KLIC offset distance of 0.33 m was found with 64.0% of enriched points being used in the comparison. Again, visualisations of the point offsets are seen in the following figures (5.12-5.13).

Trench	Description	E	N	Z
13	LDG ?110GY	257746.8656	471922.8638	38.97827121
	DATA BUNDEL TEERKABEL	257746.681	471922.8188	39.116626
	DATA BUNDEL TEERKABEL	257746.2535	471922.7147	39.17597394
	CAI C3 COAX	257745.7385	471922.5892	39.11927941
	OV KABEL	257745.6996	471922.5798	39.12103832
	CAI C3 COAX	257745.6511	471922.5679	39.11323695
	DATA 40HDPE	257745.5248	471922.5372	39.03895338
	WATER ?160PVC	257745.515	471922.5348	38.68939311
	DATA 40HDPE	257745.4859	471922.5277	39.03071229
	LDG ?160PVC	257745.1653	471922.4496	38.67522324
	CAI COAX ?110	257745.1458	471922.4449	39.02610269
	GV BUNDEL DB	257744.7864	471922.3573	39.07237255
	CAI 2xC3 COAX	257744.4171	471922.2674	39.11908214
	LS KABEL BUNDEL	257744.3491	471922.2508	38.92216022

Table 5.8.: Deurningerstraat: trench 13 enriched

Trench	Description	E	N	Z
15	LDG 160PVC	257742.26	471980.03	38.48
	WATER 315PVC	257742.26	471980.03	37.78
	WATER 50PE	257742.33	471980.06	37.95
	DATA 40HDPE	257742.37	471980.07	38.45
	DATA 40HDPE	257742.40	471980.08	38.43
	LS KABEL	257742.42	471980.09	38.67
	DATA TELEFONIE	257742.45	471980.10	38.44
	DATA BUNDEL TELEFONIE	257742.50	471980.12	38.52
	DATA TELEFONIE	257742.60	471980.15	38.46
	LS TEERKABEL	257742.62	471980.16	38.50
	LS TEERKABEL	257742.63	471980.16	38.52
	LS TEERKABEL	257742.66	471980.17	38.56
	CAI BUNDEL COAX	257742.69	471980.18	38.57
	OV KABEL	257742.70	471980.19	38.63
	LDG 100GY	257742.71	471980.19	38.31

Table 5.9.: Deurningerstraat: trench 15 enriched

Trench	Description	E	N	Z
16	DATA TEERKABEL	257739.5333	471948.3798	38.97572619
	CAI 32COAX / ?125 ST	257739.2544	471948.3006	38.90657017
	WATER ?160PVC	257738.8215	471948.1777	38.28339704
	LDG ?160PVCSV	257738.6098	471948.1176	38.4016235
	DATA TEERKABEL	257738.5329	471948.0957	38.72461495
	DATA 40HDPE	257738.2058	471948.0029	38.85732858
	CAI KABEL	257737.4362	471947.7844	38.87724301
	GV BUNDEL HDPE	257737.0803	471947.6833	38.88107844
	LS TEERKABEL	257736.7628	471947.5932	38.78341814
	DATA KABEL	257736.6955	471947.574	39.02603566
	LS TEERKABEL	257736.6186	471947.5522	38.7890271
	LS TEERKABEL	257736.4358	471947.5003	38.78613178
	LS TEERKABEL	257736.253	471947.4484	38.77323646
	LS AL	257736.1087	471947.4074	38.57884541

Table 5.10.: Deurningerstraat: trench 16 enriched

Trench	Description	E	N	Z
17	LDG ?160PVC	257731.419	471983.8108	38.47749835
	LS KABEL	257731.2009	471983.738	38.84702435
	LS KABEL	257731.1534	471983.7222	38.80344304
	CAI 40HDPE	257731.0965	471983.7032	38.87114548
	DATA TELEFONIE	257731.0396	471983.6842	38.95884791
	CAI 40HDPE	257731.0301	471983.681	38.88013165
	DATA BUNDEL TEERKABEL	257731.0112	471983.6747	38.84269913
	DATA BUNDEL TEERKABEL	257730.9163	471983.643	38.80553652
	CAI C3 COAX	257730.8973	471983.6367	38.828104
	GV BUNDEL DB	257730.8404	471983.6177	38.70580643

Table 5.11.: Deurningerstraat: trench 17 enriched

Trench	Description	E	N	Z
18	WATER ?150GY	257713.3565	472042.4476	37.86747013
	DATA TEERKABEL	257713.2694	472042.4252	38.45329351
	DATA TEERKABEL	257713.1241	472042.3877	38.32633247
	CAI BUNDEL COAX / ?75 PVC	257712.9111	472042.3328	38.51612295
	DATA TELEFONIE / ?110 PVC	257712.9111	472042.3328	38.15612295
	LS KABEL	257712.9014	472042.3303	38.29565888
	LDG ?160PVC	257712.8239	472042.3104	38.04194632
	DATA TELEFONIE / ?110 PVC	257712.8143	472042.3079	38.52148226
	LDG ?32PVC	257712.8046	472042.3054	38.38101819
	LS KABEL	257712.4172	472042.2055	38.44245542
	GV BUNDEL DB / ?75	257712.301	472042.1756	38.58688659
	GV BUNDEL DB / ?75	257712.2042	472042.1506	38.61224589

Table 5.12.: Deurningerstraat: trench 18 enriched

Trench	Description	E	N	Z
19	OV KABEL	257704.9091	472065.5199	38.82883332
	DATA TEERKABEL	257704.5655	472065.4127	38.44378332
	MS KABEL	257704.5082	472065.3949	38.35960832
	LS KABEL	257704.2313	472065.3085	38.58776248
	GV BUNDEL DB	257703.9354	472065.2162	38.4178583
	DATA TELEFONIE	257703.9068	472065.2073	38.6807708
	LS KABEL	257703.8113	472065.1775	38.62047913
	LS KABEL	257703.7636	472065.1626	38.6053333
	LDG ?160PVC	257703.7158	472065.1478	38.19018746
	LS KABEL / ?75	257703.7158	472065.1478	38.53018746

Table 5.13.: Deurningerstraat: trench 19 enriched

Trench	Description	E	N	Z
20	WATER 315PVC	257715.10	472070.12	37.65
	LS KABEL / 75	257715.19	472070.15	38.29
	DATA TELEFONIE	257715.28	472070.18	38.26
	DATA BUNDEL TEERKABEL 1	257715.30	472070.18	38.27
	DATA BUNDEL TEERKABEL 2	257715.32	472070.19	38.28
	DATA BUNDEL TEERKABEL 3	257715.35	472070.20	38.27
	LS KABEL BUNDEL	257715.35	472070.20	38.07
	DATA BUNDEL TEERKABEL 4	257715.39	472070.21	38.27
	LS TEERKABEL	257715.44	472070.22	38.15
	GV BUNDEL DB	257715.46	472070.23	38.39
	DATA BUNDEL TEERKABEL 5	257715.51	472070.24	38.17
	GV BUNDEL DB	257715.52	472070.25	38.39
	OV KABEL	257715.53	472070.25	38.40
	LS KABEL	257715.56	472070.26	38.38
	WATER 50PE	257715.58	472070.26	37.82
	LDG 160PVC	257715.63	472070.28	38.04
	LS KABEL	257715.67	472070.29	38.41
	DATA 40HDPE	257715.71	472070.30	38.35

Table 5.14.: Deurningerstraat: trench 20 enriched

Trench	Description	E	N	Z
21	WATER ?150GY	257687.1064	472127.2867	37.48871468
	DATA TEERKABEL	257686.8432	472127.1651	38.24898523
	DATA TEERKABEL	257686.707	472127.1022	38.23946999
	DATA TEERKABEL	257686.6071	472127.056	38.21715882
	DATA KABEL	257686.5708	472127.0393	38.45995476
	MS KABEL	257686.4256	472126.9722	38.02113851
	LS KABEL	257686.4165	472126.968	38.16183749
	CAI BUNDEL COAX	257686.2622	472126.8967	38.20372022
	CAI BUNDEL COAX	257686.2077	472126.8715	38.31791413
	LS KABEL BUNDEL	257686.1351	472126.838	38.26350601
	LS KABEL	257686.0625	472126.8044	38.30909788
	GV BUNDEL DB	257686.0352	472126.7918	38.27119483
	LS KABEL	257685.9808	472126.7667	38.35538874
	LDG ?160PVC	257685.9535	472126.7541	38.49748569

Table 5.15.: Deurningerstraat: trench 21 enriched

Trench	Description	E	N	Z
22	WATER ?315PVC	257690.1958	472153.9743	37.38117534
	DATA TEERKABEL	257690.4529	472154.0567	37.89169702
	DATA TEERKABEL	257690.4624	472154.0598	38.05208671
	DATA TEERKABEL	257690.729	472154.1453	38.15299809
	DATA TEERKABEL	257690.7481	472154.1514	38.02377747
	DATA TEERKABEL	257690.8052	472154.1697	38.01611562
	DATA KABEL	257690.929	472154.2094	38.20118161
	DATA TEERKABEL	257690.9766	472154.2247	38.22313007
	DATA TEERKABEL	257691.0433	472154.246	38.20585791
	DATA TEERKABEL	257691.1195	472154.2705	38.20897545
	DATA TEERKABEL	257691.1766	472154.2888	38.0413136
	WATER ?110PVC	257691.1766	472154.2888	37.5713136
	LS KABEL	257691.2052	472154.298	38.35248268
	DATA TEERKABEL	257691.3385	472154.3407	37.75793836
	DATA TEERKABEL	257691.4147	472154.3651	38.30105589
	OV KABEL	257691.4337	472154.3712	38.33183528
	LDG ?160PVC	257691.4813	472154.3865	37.68378374
	CAI C3 COAX	257691.5765	472154.417	38.09768066
	LS KABEL	257691.6337	472154.4354	37.83001881
	LS 4X150	257691.7384	472154.469	38.08430542
	LS 2x TEERKABEL	257691.8146	472154.4934	37.82742295
	DATA BUNDEL 40HDPE	257692.0431	472154.5667	37.83677555
	GV BUNDEL DB	257692.1098	472154.588	38.0895034
	CAI BUNDEL COAX	257692.3098	472154.6522	38.12768692

Table 5.16.: Deurningerstraat: trench 22 enriched

Trench	Description	E	N	Z
25	DATA TEERKABEL	257677.2739	472199.7829	37.95246413
	DATA TEERKABEL	257677.3883	472199.8193	38.00144979
	WATER ?200PVC	257677.4168	472199.8283	37.5036962
	LS KABEL	257677.6074	472199.8889	38.13867229
	DATA BUNDEL TEERKABEL	257677.8743	472199.9738	37.89963881
	CAI C3 COAX	257677.9219	472199.9889	38.18338283
	LS KABEL	257677.9696	472200.0041	38.15712685
	LDG ?200PVC	257678.0554	472200.0313	37.79386609
	DATA TEERKABEL	257678.0649	472200.0344	38.1346149
	LS KABEL	257678.0935	472200.0435	38.33686131
	LS KABEL BUNDEL	257678.2364	472200.0889	38.21809338
	DATA BUNDEL 40HDPE	257678.3031	472200.1101	38.16333501
	LS KABEL BUNDEL	257678.3222	472200.1162	38.23483262
	DATA 40HDPE	257678.4175	472200.1465	38.18232066
	LS KABEL	257678.4651	472200.1616	38.24606468
	DATA 40HDPE	257678.4937	472200.1707	38.3683111
	LS KABEL	257678.5509	472200.1889	38.23280392
	DATA 40HDPE	257678.6748	472200.2283	38.37253838
	DATA 40HDPE	257678.7415	472200.2495	38.34778001
	CAI BUNDEL COAX	257678.7701	472200.2586	38.30002643
	GV BUNDEL DB ?110	257678.9035	472200.301	38.13050969
	GV BUNDEL DB ?110	257679.0084	472200.3343	38.09874654
	CAI BUNDEL COAX	257679.0369	472200.3434	38.31099295
	CAI BUNDEL COAX	257679.056	472200.3495	38.31249056
	CAI BUNDEL COAX	257679.199	472200.3949	38.33372262
	LS KABEL	257679.3514	472200.4434	38.26570349
	LS KABEL	257679.38	472200.4525	38.26794991
	LS KABEL	257679.4563	472200.4767	38.25394034
	LS KABEL	257679.4944	472200.4888	38.26693556

Table 5.17.: Deurningerstraat: trench 25 enriched

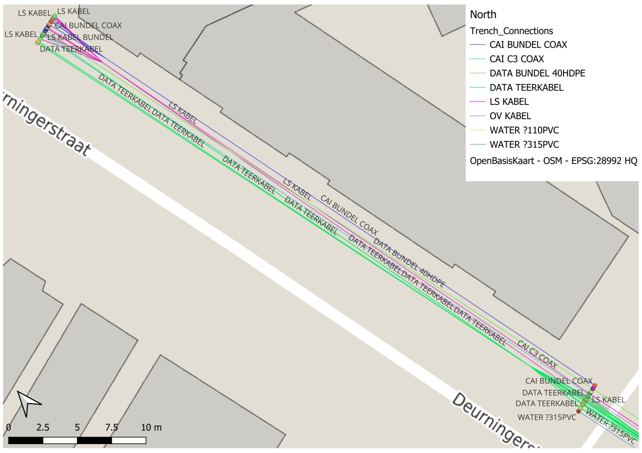


Figure 5.6.: Deurningerstraat: trenches 3 & 4 final connections

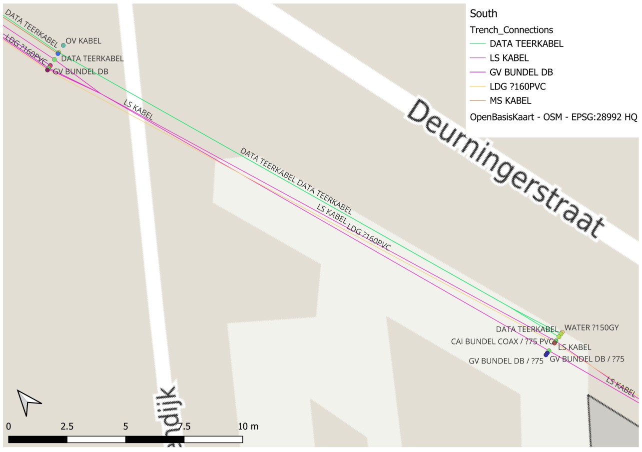


Figure 5.7.: Deurningerstraat: trenches 18 & 19 final connections

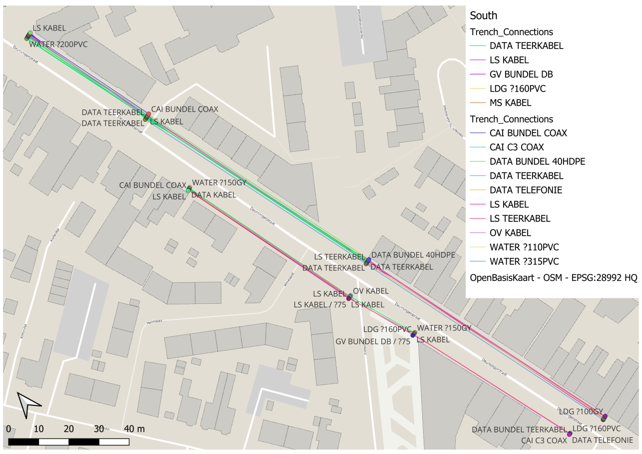


Figure 5.8.: Deurningerstraat: overall connections

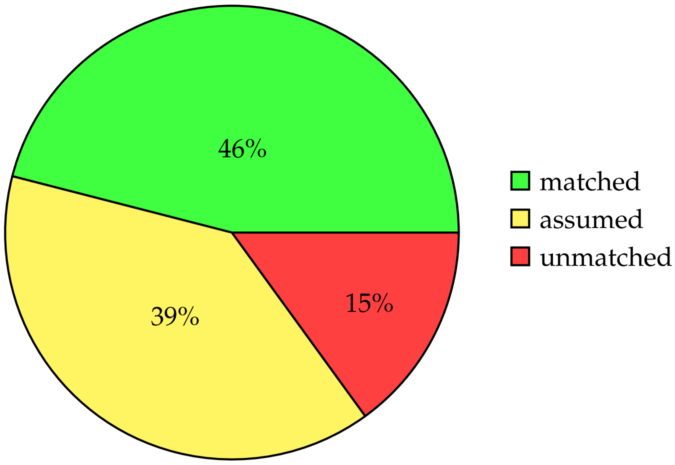


Figure 5.9.: Deurningerstraat: Cable status for trenches 19 & 21

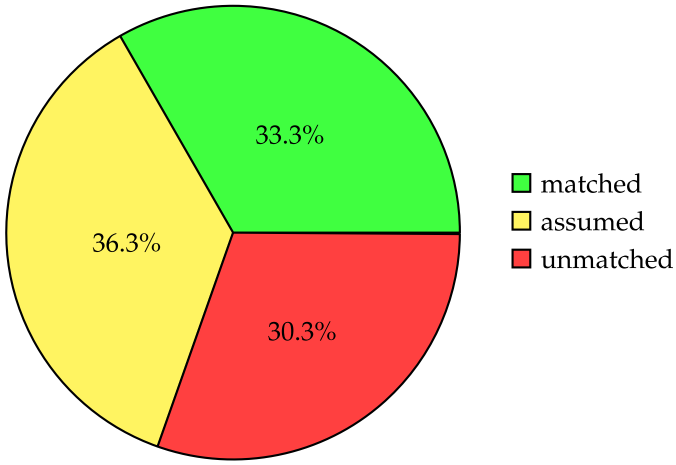


Figure 5.10.: Deurningerstraat: Cable status for trenches 15 & 20

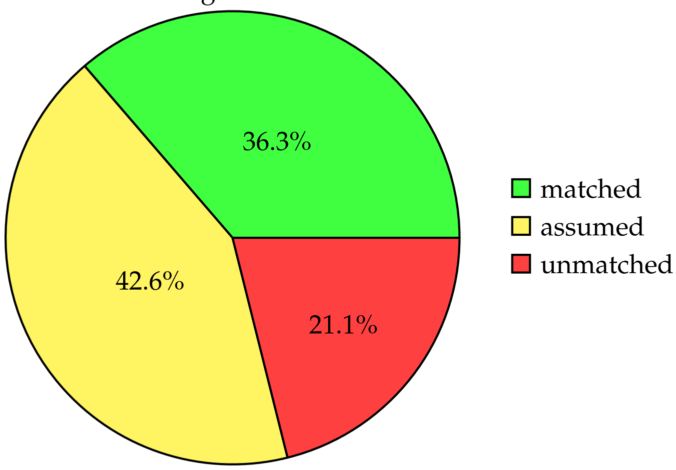


Figure 5.11.: Deurningerstraat: Total Distribution of cable status

Trench	Description	Distance (m)	Points used
15	WATER 50PE	0.001	66.7%
	LDG 160PVC	0.10	
	CAI BUNDEL COAX	0.12	
	WATER 315PVC	0.19	
	LDG 100GY	0.31	
20	DATA TEERKABEL	0.03	40.9%
	LDG 110PVC	0.04	
	DATA TEERKABEL	0.05	
	WATER 315PVC	0.10	
	WATER 110PVC	0.10	
	DATA TEERKABEL	0.12	
	DATA TEERKABEL	0.13	
	LS TEERKABEL	0.23	
	DATA TEERKABEL	0.49	
22	LS 4X150	0.01	75.0%
	CAI BUNDEL COAX	0.03	
	GV BUNDEL DB	0.07	
	LS 2x TEERKABEL	0.07	
	WATER 315PVC	0.10	
	LDG 160PVC	0.36	
25	WATER 200PVC	0.01	83.3%
	DATA 40HDPE	0.01	
	DATA 40HDPE	0.02	
	DATA TEERKABEL	0.02	
	DATA 40HDPE	0.06	
	LS KABEL	0.08	
	DATA 40HDPE	0.08	
	DATA BUNDEL TEERKABEL	0.09	
	DATA TEERKABEL	0.10	
	CAI BUNDEL COAX	0.13	
	LS KABEL	0.14	
	LDG 200PVC	0.17	
	LS KABEL	0.17	
	DATA 40HDPE	0.21	
	LS KABEL	0.25	
	DATA 40HDPE	0.27	
	CAI BUNDEL COAX	0.28	
	DATA 40HDPE	0.28	
	LS KABEL	0.29	
	CAI BUNDEL COAX	0.30	
	LS KABEL	0.30	
	CAI BUNDEL COAX	0.31	
	DATA 40HDPE	0.35	
	LS KABEL	0.43	
	CAI BUNDEL COAX	0.58	
	CAI BUNDEL COAX	0.59	
	CAI BUNDEL COAX	0.61	
	LS KABEL	0.71	
	DATA TEERKABEL	0.73	
	CAI BUNDEL COAX	0.76	
	LS KABEL	0.80	

Table 5.18.: Deurningerstraat KLIC comparison for trenches 15-25

Trench	Description	Distance (m)	Points used
17	GV BUNDEL DB	0.004	40%
	LDG 160PVC	0.01	
	LS KABEL	0.10	
	LS KABEL	0.15	
18	GV BUNDEL DB / 75	0.08	71.4%
	GV BUNDEL DB / 75	0.08	
	LDG 32PVC	0.1692	
	GV BUNDEL DB / 75	0.18	
	GV BUNDEL DB / 75	0.18	
	DATA TEERKABEL	0.20	
	LS KABEL	0.23	
	WATER 150GY	0.26	
	LS KABEL	0.27	
	DATA TEERKABEL	0.35	
19	LS KABEL	0.05	50.0%
	LDG 160PVC	0.08	
	LS KABEL	0.10	
	LS KABEL	0.39	
	MS KABEL	0.59	
21	DATA TEERKABEL	0.003	84.6%
	CAI BUNDEL COAX	0.02	
	LS KABEL	0.03	
	LDG 160PVC	0.04	
	LS KABEL	0.06	
	CAI BUNDEL COAX	0.08	
	WATER 150GY	0.09	
	LS KABEL BUNDEL	0.14	
	DATA TEERKABEL	0.15	
	DATA TEERKABEL	0.26	
	LS KABEL	0.45	
	MS KABEL	0.54	
Average		0.33	64.0%

Table 5.19.: Deurningerstraat KLIC comparison for trenches 17, 18, 19, and 21

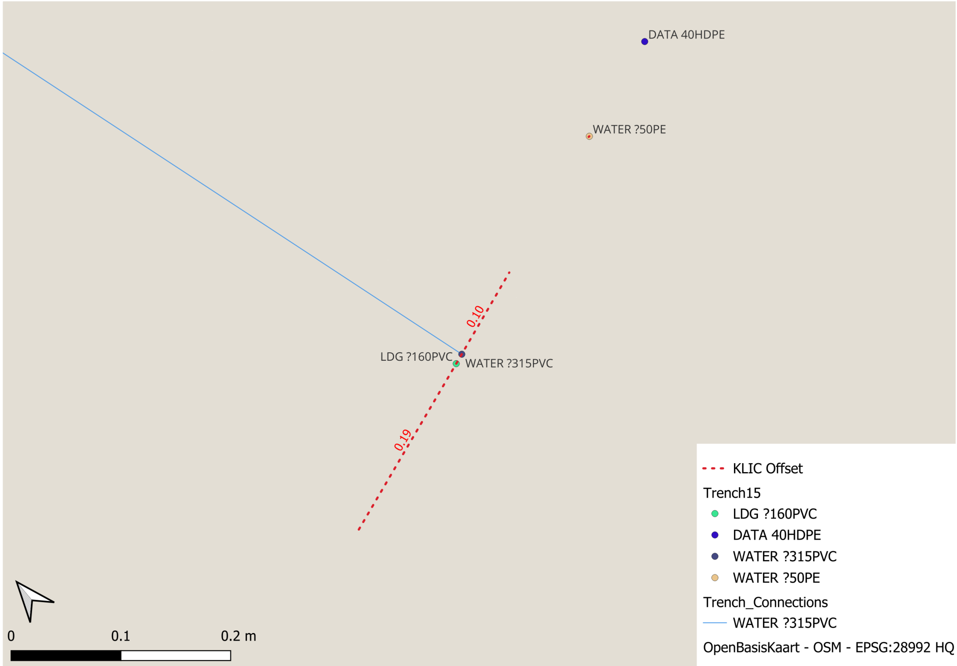


Figure 5.12.: Deurningerstraat: KLIC offset, trench 15

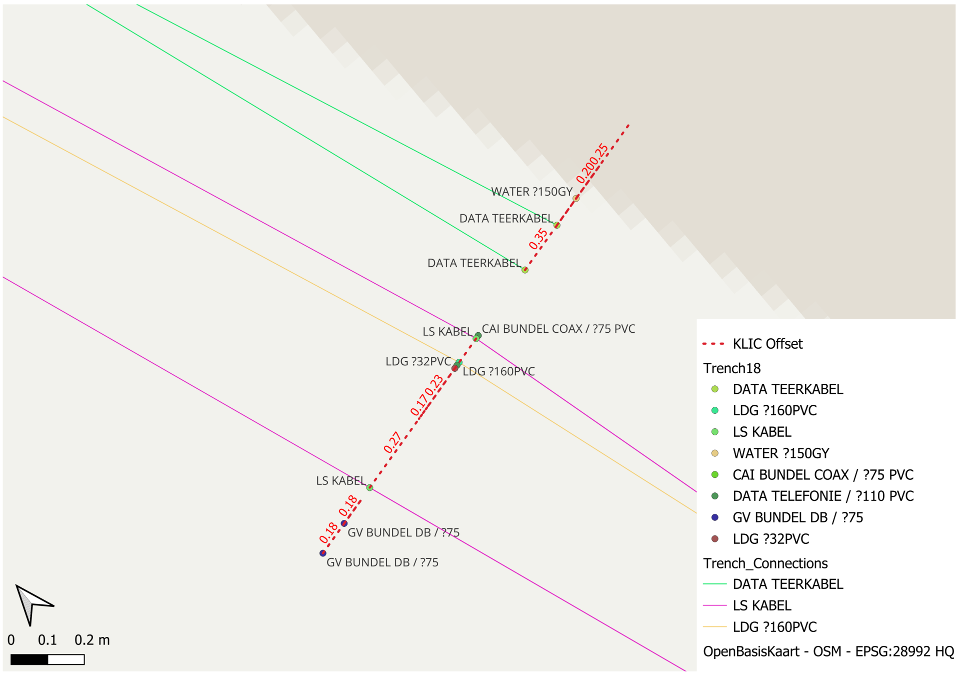


Figure 5.13.: Deurningerstraat: KLIC offset, trench 18

5.3. Rotterdam

This section displays the results for the Rotterdam case. As with Enschede, the enriched utility data is displayed followed by some visualisations of trench connections (Figures 5.14, 5.15 & 5.16). The pie charts display the matching cable status of a couple trenches in Rotterdam. Figure 5.19 displays the status of all cables, even those that were not able to be identified by the surveyors. To account for this, Figure 5.20, shows the status of only the labelled cables to provide a result that can be better compared to those of Enschede. Tables 5.28 and 5.29, show the comparison of the enriched cable points to the LVZK intersection points for two different streets, Larikslaan and Kastanjesingle, respectively. Finally, the LVZK offsets are visualised in Figure 5.21 and additional figures contained in the appendix.

Trench	Description	E	N	Z
G10	1016	92475.57439	441956.4378	-5.2
	LS E	92475.16555	441955.9454	-5.63
	1013	92474.9739	441955.7146	-6.13
	160 PVC G	92474.89724	441955.6223	-5.83
	1011	92474.82697	441955.5376	-6.1
	1010	92474.76309	441955.4607	-5.16
	1009	92474.4884	441955.1299	-6.07
	170 GY W	92474.43091	441955.0606	-5.81
	1007	92474.29675	441954.8991	-5.76
	1006	92474.29675	441954.8991	-6.11
	BUNDEL LOS REGEVIBER	92474.21371	441954.7991	-5.77
	1004	92473.95818	441954.4913	-6.17
	1003	92473.93901	441954.4682	-5.82
	AFDEK ZIGGO KUNSTSTOF	92473.87513	441954.3913	-5.78
	1001	92473.77931	441954.2759	-5.14

Table 5.20.: Rotterdam: Trench G10 enriched data

Trench	Description	E	N	Z
G20	2011	92325.57112	442002.437	-4.83
	2010	92325.64805	442002.3731	-5.68
	CO ZIGGO	92325.69421	442002.3348	-5.56
	1X CU KPN NTV	92326.6636	442001.5299	-4.83
	40 PE W ONBEKEND	92326.75592	442001.4532	-5.53
	2006	92326.78669	442001.4277	-5.68
	2005	92327.04827	442001.2105	-6.19
	2004	92327.30985	442000.9933	-4.83
	1X CU KPN 1X LS E NTV	92328.17153	442000.2778	-4.83
	2001	92329.48712	441999.1854	-4.81

Table 5.21.: Rotterdam: Trench G20 enriched data

Trench	Description	E	N	Z
G30	3006	92489.94325	441859.4039	-5.02
	OV E	92490.48106	441858.9716	-5.51
	3003	92490.62915	441858.8526	-5.01
	3002	92491.03446	441858.5268	-5.6
	3001	92491.14358	441858.4391	-5.02

Table 5.22.: Rotterdam: Trench G30 enriched data

Trench	Description	E	N	Z
G40	4018	92348.48995	441904.1728	-4.62
	4017	92348.19213	441903.8092	-4.63
	40 PE G	92348.02104	441903.6003	-5.31
	110 PVC G	92347.75491	441903.2754	-5.35
	4014	92347.72956	441903.2444	-4.69
	4013	92347.57748	441903.0588	-5.22
	4012	92347.46976	441902.9273	-5.09
	3X CU KPN	92347.43808	441902.8886	-5.12
	CO ZIGGO	92347.32402	441902.7493	-5.64
	4009	92347.30501	441902.7261	-5.71
	4008	92347.05155	441902.4167	-4.69
	4007	92344.95416	441899.856	-5.58
	4006	92344.89713	441899.7864	-4.7
	4004	92342.64132	441897.0323	-5.16
	4003	92342.53993	441896.9085	-5.13
	3X LS E	92342.40053	441896.7383	-5.17
	4001	92342.3435	441896.6687	-4.71

Table 5.23.: Rotterdam: Trench G40 enriched data

Trench	Width	Depth	Description	E	N	Z
G50	7.02	-4.45	5038	92159.97936	441983.6103	-4.45
	6.73	-4.42	5037	92159.79651	441983.3852	-4.42
	6.57	-4.4	5036	92159.69563	441983.261	-4.4
	6.51	-4.41	5035	92159.6578	441983.2144	-4.41
	6.48	-4.44	5034	92159.63888	441983.1911	-4.44
	6.46	-5.47	5033	92159.62627	441983.1756	-5.47
	6.26	-5.19	CO ZIGGO	92159.50017	441983.0204	-5.19
	5.53	-5.03	CU KPN	92159.0399	441982.4538	-5.03
	5.11	-5.06	6X HDPE 2X ZIGGO 4X KPN	92158.77508	441982.1278	-5.06
	5.03	-5.1	5029	92158.72464	441982.0657	-5.1
	5	-5.11	5028	92158.70572	441982.0424	-5.11
	4.98	-5.12	5027	92158.69311	441982.0269	-5.12
	4.95	-5.12	5026	92158.6742	441982.0036	-5.12
	4.9	-5.12	5025	92158.64267	441981.9648	-5.12
	4.71	-4.86	5X KABEL REGEFIBE	92158.52288	441981.8173	-4.86
	4.47	-5.38	5023	92158.37155	441981.631	-5.38
	4.35	-5.1	160 PE G	92158.29589	441981.5379	-5.1
	4.23	-5.44	5021	92158.22023	441981.4447	-5.44
	4.14	-4.4	5020	92158.16348	441981.3749	-4.4
	3.71	-5.43	5019	92157.89236	441981.0411	-5.43
	3.51	-5.31	250 PVC W	92157.76626	441980.8859	-5.31
	3.34	-5.51	5017	92157.65907	441980.7539	-5.51
	3.15	-5.52	5016	92157.53928	441980.6065	-5.52
	3.02	-5.15	5015	92157.45731	441980.5056	-5.15
	2.81	-5.08	5014	92157.3249	441980.3426	-5.08
	2.75	-4.96	OV E	92157.28707	441980.296	-4.96
	2.22	-4.44	5012	92156.9529	441979.8846	-4.44
	2.06	-4.45	5011	92156.85202	441979.7604	-4.45
	2.02	-4.54	5010	92156.8268	441979.7294	-4.54
	1.97	-5.05	CU KPN	92156.79527	441979.6906	-5.05
	1.6	-5.5	LS E BD ONBEKEND	92156.56198	441979.4034	-5.5
	1.31	-4.52	5006	92156.37913	441979.1783	-4.52

Table 5.24.: Rotterdam: Trench G50 enriched data

Trench	Description	E	N	Z
G60	6012	92129.43818	441916.2452	-4.67
	110 PVC REGEFIBER	92129.91461	441915.8484	-5.12
	2X CU KPN	92130.16051	441915.6436	-5.5
	6009	92130.26041	441915.5604	-5.44
	250 PVC W	92130.49863	441915.3621	-5.6
	6007	92130.62158	441915.2597	-4.62
	6005	92130.92127	441915.0101	-5.57
	6004	92131.1979	441914.7797	-5.65
	HS-PLAAT BT	92131.35159	441914.6517	-5.37
	6002	92131.52833	441914.5046	-5.36
	6001	92131.91255	441914.1846	-4.58

Table 5.25.: Rotterdam: Trench G60 enriched data

Trench	Description	E	N	Z
G70	7018	92394.95715	441697.3429	-4.89
	7017	92394.9885	441697.3818	-6.3
	7016	92395.01359	441697.413	-5.69
	KUNSTOF MAT ZIGGO	92395.08884	441697.5065	-5.69
	1 CU KPN	92395.24563	441697.7012	-5.71
	7013	92395.30834	441697.7791	-5.16
	KABELBED REGGEVIBER	92395.35851	441697.8414	-5.16
	7011	92395.42123	441697.9193	-4.86
	VERZAMEL BAK REGGEVIBER	92395.43377	441697.9349	-4.97
	160 PVC W	92395.43377	441697.9349	-6.2
	1 OV E	92395.69717	441698.262	-5.57
	7007	92395.9543	441698.5814	-6.3
	7006	92396.02955	441698.6748	-4.88
	7005	92396.07345	441698.7294	-5
	7004	92396.69432	441699.5005	-4.96
	110 PVC G	92396.70059	441699.5083	-5.82
	7001	92397.36536	441700.3339	-4.92

Table 5.26.: Rotterdam: Trench G70 enriched data

Trench	Description	E	N	Z
G80	8016	92108.97	441858.54	-4.62
	CU KPN	92108.90	441858.46	-5.38
	160 PVC W	92108.68	441858.19	-5.61
	8013	92108.60	441858.09	-5.88
	8012	92108.58	441858.07	-5.11
	BUNDEL LOS REGEVIBER	92108.51	441857.98	-5.12
	LS E	92108.13	441857.52	-5.32
	OV E	92108.07	441857.44	-5.4
	8008	92107.92	441857.26	-4.6
	125 PE R PERSRIOOL GEPRIKT	92107.91	441857.25	-6.32
	400X600 BT R	92107.07	441856.22	-5.45
	110 PVC G	92106.64	441855.69	-5.42
	8003	92106.60	441855.64	-4.59
	8002	92106.57	441855.61	-4.68
	8001	92106.43	441855.44	-4.67

Table 5.27.: Rotterdam: Trench G80 enriched data

Trench	Description	Distance (m)	Points used
G76	110 PVC G	0.16	23.5%
	CU KPN	0.53	
	400X600 BT R	0.81	
	160 PVC W	0.85	
G80	110 PVC G	0.47	33.3%
	400X600 BT R	0.60	
	LS E	0.74	
	160 PVC W	0.74	
	CU KPN	0.91	
G88	110 PVC G	0.28	28.6%
	400X600 BT R	0.28	
	160 PVC W	0.54	
	CU KPN	0.64	
G93	160 PVC W	0.04	31.3%
	110 PE G	0.08	
	LS E	0.09	
	110 PVC ZIGGO	0.14	
	400X600 BT R	0.16	
	CU KPN	0.43	
G99	400X600 BT R	1.33	42.9%
	160 PVC G	1.34	
	160 PVC W	1.41	
	KUNSTOF MAT ZIGGO	1.48	
	1 LS E	1.53	
	1 CU KPN	1.64	
G101	110 PVC O	0.07	44.4%
	110 PVC E	0.60	
	1 OV E	1.08	
	160 PVC G	1.09	
Average		0.69	34.0%

Table 5.28.: Rotterdam: Larikslaan enriched data compared to LVZK data

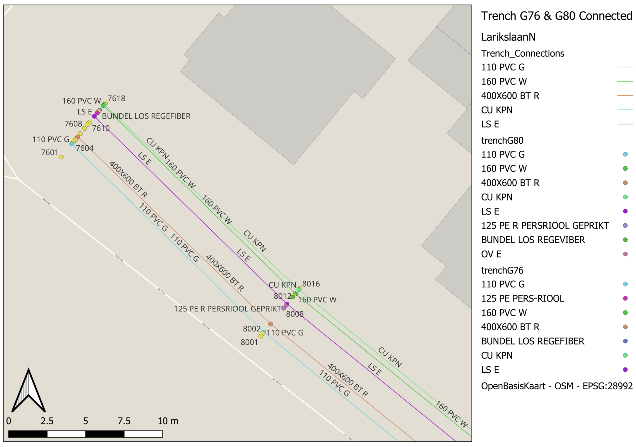


Figure 5.14.: Rotterdam: trenches G76 & G80 final connections

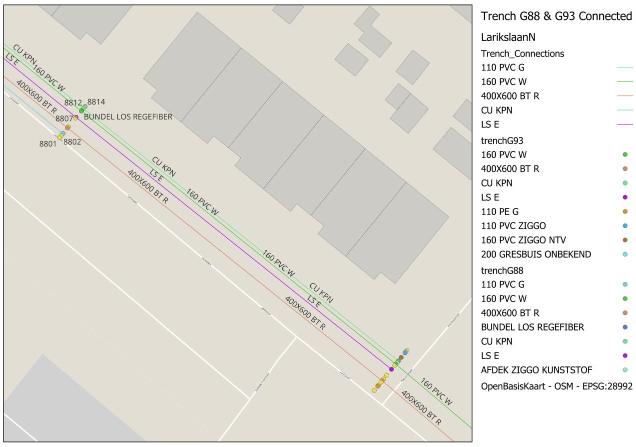


Figure 5.15.: Rotterdam: trenches G88 & G93 final connections

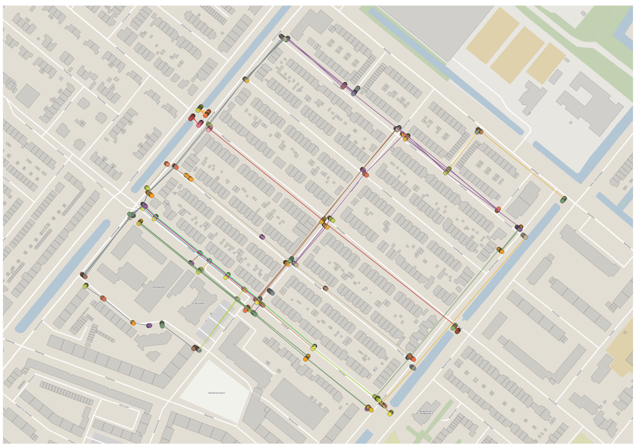


Figure 5.16.: Rotterdam: All trench connections

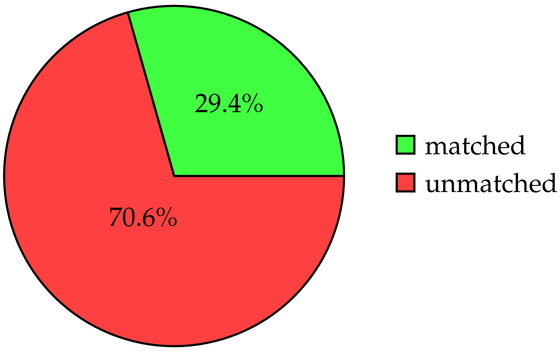


Figure 5.17.: Rotterdam: Cable status for trenches G76 & G80

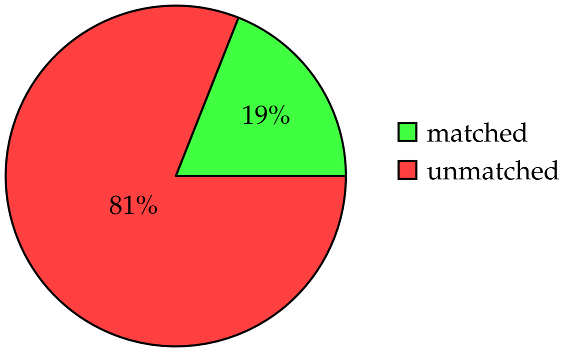


Figure 5.18.: Rotterdam: Cable status for trenches G101 & G104

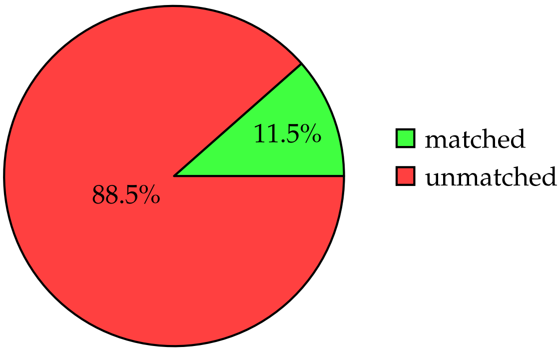


Figure 5.19.: Rotterdam: Total Distribution of **labelled & unlabelled** cable status

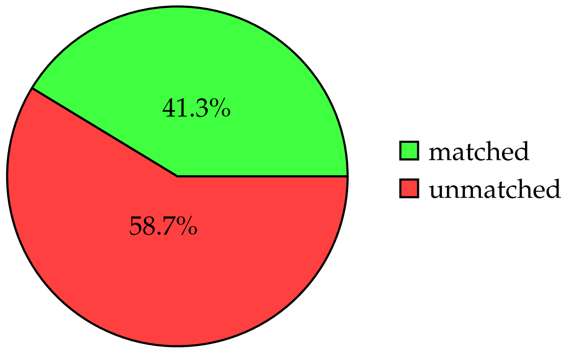


Figure 5.20.: Rotterdam: Total Distribution of **labelled** cable status

Trench	Description	Distance (m)	Points Used
52	3X LS E	1.40	11.50%
	CU KPN	1.47	
	118 GY W	4.84	
53	110 PVC ZIGGO	0.62	27.80%
	200 PVC KPN	0.65	
	2X LS E	1.30	
	2X CU KPN	1.56	
	160 PVC W	1.96	
54	250 PE W	0.003	27.30%
	CU KPN NTV	0.58	
	110 PVC REGEFIBER	0.75	
55	250 PVC W	0.15	25.00%
	110 PVC REGEFIBER	0.25	
	2X CU KPN	0.85	
56	LS E	0.31	20.00%
	2X LS E	1.30	
	250 PVC W	4.48	
	2X CU KPN	4.89	
57	250 PVC W	0.26	21.10%
	CU KPN	1.20	
	4X LS E	1.34	
	CU KPN	2.96	
58	LS E	1.58	19.00%
	CU KPN	1.76	
	250 PVC W	2.40	
	400 BT R	3.15	
59	110 PVC REGEFIBER	0.95	21.70%
	110 PE RET	1.09	
	110 PVC REGEFIBER	1.15	
	2X 160 PVC STEDIN	1.23	
	110 PVC REGEFIBER	1.35	
Average		1.54	21.7%

Table 5.29.: Rotterdam: Kastanjesingle enriched data compared to LVZK data

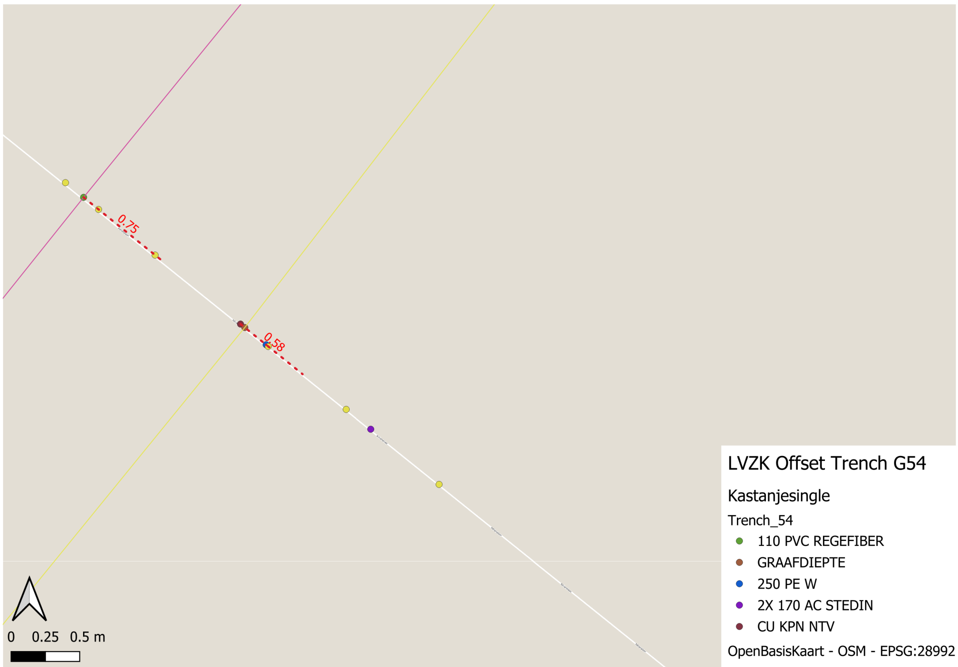


Figure 5.21.: Rotterdam: LVZK offset, trench G54

5.4. Impact on Data Usability

The utilization of enriched utility data derived from direct measurements in utility trenches offers some insight into the spatial accuracy of utility network datasets, in this case, KLIC and Rotterdam's LVZK. Specifically, at the locations where trenches have been excavated, this enriched data provides a more accurate reference that shows discrepancies in these traditional datasets. Instances of misplaced or missing cables in KLIC and LVZK maps have been identified through comparisons with the trench-derived data, as can be seen in tables 5.28, 5.29, 5.7, 5.18, and 5.19. Where the percentage of points used signifies the percentage of trench-enriched points that were matched to either the KLIC or LVZK cables.

For example, the average distance from each area suggests that the LVZK of Rotterdam Larikslaan has an average cable offset error of 0.69 m, with only 34% of cables properly identified. At Kastanjesingle, the results show an average offset of 1.54 m with 21% of points identified. Note that, the low percentage of points used in Rotterdam is due to the high number of unlabeled cables in the utility trench survey. In Enschede, comparisons to the KLIC data reveal that Deurningerstraat has an average offset of 0.33 m, with 66.3% of cables identified properly, while Brammelerdwarsstraat shows an average offset of 0.30 m and 47.2% of cables identified properly. These findings indicate that, on average, cables will be offset by at least 0.3 m, and over 30% of cables are not accounted for in traditional datasets, highlighting significant limitations for urban planning and management.

While these discrepancies are important in urban planning where the precise location of underground utilities can impact construction safety and infrastructure development, it is important to acknowledge the limitations and challenges associated with the data quality. The enriched data pinpoints inaccuracies but is limited by the quality and completeness of the trench surveys. These limitations reduce the added value of the trenches, making it challenging to exploit the data fully for practical applications.

Moreover, the improvement in positional accuracy provided by the enriched data, though informative, is minimal under current conditions. The benefits are constrained by the extent of data enrichment and the inherent issues in the existing datasets. Given the current limitations, it is questionable whether the added value of the trenches can be fully utilized in real-world urban planning scenarios.

In conclusion, while traditional datasets like KLIC and LVZK provide broad coverage, they lack accuracy at least at the trial trench locations. The enriched data from utility trenches offers a view of issues in traditional utility network mapping at specific sites, but its usability is limited by the data quality and completeness. In the course of this research, an issue that was faced time and time again was that of labeling, data quality and completeness. In order to make further progress, improved labeling and documentation of cables in trench surveys and utility network maps are needed in order to increase the percentage of cables/data used. This would hopefully help bridge the gap between the current limitations and the potential benefits of using enriched utility data, potentially contributing to safer and more efficient urban infrastructure development.

6. Discussion

This chapter discusses the outcomes of the semantic and spatial enrichment of utility network data. It aims to explain the challenges faced during the data enrichment process, the solutions applied to address these challenges, and the benefits gained from integrating enriched data into urban utility management. This discussion is critical for understanding the impacts of data accuracy improvements on urban planning and infrastructure development.

6.1. Challenges and proposed solutions

The integration of enriched utility data was confronted with several challenges, primarily relating to discrepancies in data quality and the complexity of aligning new data with existing datasets. A notable issue was the integration of trench-derived data, which often contradicted existing records from databases like KLIC and LVZK, revealing missing or incorrectly positioned cables. This issue is consistent with findings in previous studies, which also highlight the fragmentation and inaccuracy of utility data [Fossatti et al., 2020].

6.1.1. Handling different data formats

One significant challenge encountered during the enrichment process was the variation in data formats provided by different cities, combined with the complexity of the DWG files used. DWG files, primarily designed for CAD software, are not inherently GIS formats, which complicated the automation of data integration. The diversity in data formats worsened the issue, as each city's data required unique handling. This hindered the development of a streamlined, automatic process for data integration, echoing the difficulties noted by Lieberman [2019] in integrating heterogeneous data sources.

To address this issue, a multi-faceted approach was adopted leveraging Python scripting and the 'ezdxf' library, designed to interact with DWG/DXF files. This approach involved developing separate scripts for each city's data format to first standardize the data extraction process. By parsing relevant layers and entities within the DWG files that contained the utility data, the scripts converted the extracted data into a uniform format, specifically focusing on the coordinates and adding the third dimension into the originally 2D data. This standardization greatly simplified the process for the data integration process, consistent with the benefits of standardized models discussed by Becker et al. [2011].

6.1.2. Inconsistencies in Naming Standards

Another obstacle in the data integration process was the inconsistency in naming standards between the data derived from the utility trenches and the cables and the cables in the KLIC and LVZK datasets. Discrepancies in naming conventions hindered proper geometrical and semantic connections, a problem highlighted in various studies [van den Berg and Janssen, 2021]. Firstly, the discrepancies between individual trenches impeded proper geometrical and semantic connections (ex. "OV E" and "1 OV E"). Without an effective solution, the preliminary matches were exceedingly low, with only 43% of all cables in Brammelerdwarstraat and 36.3% in Deurningerstraat being correctly matched based on names alone. To overcome this, a semi-automatic matching process was developed, where the cables were matched manually for one trench and the process was repeated for other trenches using that information.

Once the connections between trenches were successful and all the trenches were visualised, the comparative analysis had to be performed. This was done by comparing the enriched points to those where the KLIC and LVZK intersected the trench line. Again, issues arose in regards to the naming standards. The KLIC dataset used a completely different naming standard for the cables than for the trench points. This was remedied by performing another matching similar to the first. However, this highlights the critical need for standardized naming conventions across datasets. This finding aligns with the recommendations of Pavlidou [2022] for the adoption of unified standards to enhance interoperability and data quality.

6.2. Role of standards in facilitating semantic enrichment

Throughout the research, numerous challenges were encountered that stemmed from inconsistencies in, or the complete absence of, standardized naming conventions for cables and pipelines. These issues underscore the critical role that standards play in the semantic enrichment of utility networks. Standards provide a robust framework that not only ensures consistency and accuracy but also facilitates the interoperability of data across different systems and geographic regions Becker et al. [2011].

In utility management, standards are essential for defining and guiding the collection, storage, processing, and sharing of data related to various utility components. They help to eliminate ambiguity and reduce errors in data handling by ensuring that everyone, across different departments, cities, or countries, uses the same language and formats when dealing with utility data.

1. **Adoption of nationwide or international standards:** The case of Rotterdam's LVZK provides a compelling example of how localized standards can enhance data management within a specific city. The LVZK had fewer issues related to matching names to trenches compared to Enschede. However, the specificity of the LVZK to Rotterdam limits its interoperability with other datasets across the Netherlands. To overcome such limitations, it is recommended to adopt nationwide or even international standards for naming and managing utility data. This would ensure uniformity and facilitate easier data integration and comparison across different regions [Pavlidou, 2022].
2. **Alignment of surveyor practices with existing data standards:** Surveyors play a crucial role in the collection of utility data. One potential improvement is for surveyors to align their naming conventions with those used in the existing datasets of the cities

they operate in. This alignment would decrease the occurrence of mismatches and discrepancies in data collected from field operations compared to the data already present in city databases [Fossatti et al., 2020].

6.3. Benefits of Data Integration

The integration of utility data presents potential benefits for urban planning and infrastructure management. However, we must acknowledge that the current state of data quality and completeness significantly limits these benefits. This section discusses benefits and the necessary steps to fully realize the potential of data integration.

6.3.1. Identifying dataset errors

Using data from utility trenches demonstrates a potential to correct the spatial representation of underground utilities. The comparison of traditional utility network datasets with the enriched trench data has shown some mismatches and inaccuracies in existing records. However, integrating this enriched data into urban planning frameworks would offer minimal, or even no, improvements in planning accuracy. The utility data includes slightly more precise positioning, but its usability is constrained by the limited coverage and quality of trench surveys. The biggest benefit of this research is that it demonstrates that there are significant errors in the traditional utility network maps. However, as this information is only accurate along a trench line and only to a few points along that line, this method cannot be recommended to automatically enrich utility network maps without improved documentation and labeling of utility cables.

While the integration of utility data from trenches with utility network maps has the potential to offer some improvements in accuracy, the current reality is far from this ideal. The findings of this research indicate that the benefits are limited by the quality and completeness of the data. Moving forward, several steps are necessary to bridge the gap between the current state and the ideal scenario. Firstly, improved labeling and documentation of cables and an improved interoperability between individual utility networks. Additionally, improved surveyor practices would help with the data integration process greatly. For example, to align the surveyed data between Enschede and Rotterdam, the surveyors could each provide both the depth (below surface) and elevation of the cables so regardless of the purpose, users of the data have all information needed. Additionally, instead of providing cable positions locally (per trench), the surveyors could also measure the actual coordinates in the relevant CRS, thus facilitating many steps in regard to data integration.

6.4. New Contributions to the Field

This research introduces a couple approaches in attempts to enrich utility network data. Firstly, we developed a methodology for integrating detailed trench data with existing utility maps, which was adapted to a couple different cities and data formats. The study provides a practical example of handling and integrating differing data formats, with the use of Python scripting and GIS tools to automate and streamline the data enrichment process.

This approach demonstrates the feasibility and benefits, however minimal they may be, of enhancing utility network maps with spatial and semantic data.

Like the related work, the research underscores the importance of standardization in utility data management. Specifically, the alignment of surveyed data with existing standards in order to enhance data consistency and interoperability, contributing to better data integration and management practices.

In conclusion, this thesis not only addresses the technical challenges of integrating utility trench data with existing maps but also continues to highlights the importance standardized data sets and collection of data in agreement with previous research papers.

7. Conclusion

This thesis aimed to tackle the intricate challenges faced in enhancing the accuracy and detail of utility network maps through spatial and semantic enrichment using trial trench data. This chapter provides a summary and discusses limitations faced in this work.

7.1. Summary

To re-iterate, the main research question posed in the introduction is:

To what extent can existing 2D utility network models in the Netherlands be enriched using surveyed utility trench data?

With the subsequent questions being defined as:

- To what extent can the information be integrated either semantically and/or geometrically?
- To what extent can a common methodology be implemented for the 3 different cities?
- Which strategies can be developed to automate the data extraction and integration process?
- How far can current standards help reduce issues relating to data integration?

The research presented in this thesis tackles the challenge of improving the accuracy and detail of 2D utility network GIS maps through a semi-automatic approach that provides spatial and semantic enrichment with data from utility trenches. The methodology adopted includes the acquisition of data from multiple cities and locations, followed by data analysis and preparation to enhance the semantic and spatial accuracy of existing utility maps. The core of the methodology is the integration of trial trench data with the 2D data to introduce a third dimension—depth—which is crucial for accurate urban infrastructure planning and risk management.

The main contribution of this thesis is the development of a semi-automatic approach that enhances the semantic and spatial dimensions of utility network maps by incorporating detailed data from trial trenches. In regards to the main research question, this research has shown that it is possible to enrich the existing utility networks using utility trench data but the results show that data quality and inconsistency issues still play a dominant role that hinders a proper and really successful exploitation of the added value of such data integration. That being said, the research shows that it is possible to minimally enrich the utility networks by providing semantic and elevation information extracted from utility trench surveys. The work done, contributes to solve this question by presenting enriched datasets of each city in question. As for the sub-questions, the information was integrated to the extent that one can calculate positional error on the behalf of the KLIC and LVZK utility maps. As for developing a common strategy, each DXF file had to be individually extracted using a Python script; along with naming and matching issue, this process can not be deemed automatic, but perhaps semi-automatic.

7.2. Limitation

Despite the innovative approach and contributions of this thesis, several limitations have been identified that impact the overall scope and applicability of the research. First and foremost, data availability and quality presented substantial challenges. The utility network data, in the case of Amsterdam, was not necessarily available or extractable, and where it was available, it often contained issue relating to the naming standards of cables and pipelines. Additionally, limitations arose from the inconsistency and format in which the surveyed data was delivered. For example, the positions of the individual cables were provided in a local coordinate system, relative to the trenches they pertained to, if the cables and pipeline positions were provided in a proper CRS when surveyed, the extraction procedure would be greatly simplified. These limitations were particularly evident when attempting to connect successive trenches in QGIS.

Issues as such, constrained the scope of the data enrichment along with the automation of said enrichment. While the processes used were able to automate most of the data integration, they still required some manual oversight and intervention, particularly in the cases of data inconsistency or when manually selecting the trenches to be used as there is no structure to how trenches themselves were named. This dependency on manual intervention could have introduced human error and limits the scalability of the approach.

Limitations in the methodology could stem from certain assumptions, such as that the surface between the start and end of a trench is assumed to be linear. This is more than likely not the case and could slightly inaccurate elevation calculations if the actual elevation is not measured on-site, as was did in Rotterdam. Another limitation in the methodology is that only trenches that were in-line with other trenches were selected for processing. In Brammelerdwarsstraat, this means that of 10 trenches only 8 were used for this research. Not incorporating this data leaves us with a smaller sample set to test from and may reduce the accuracy of the findings, compared to if they were taken into consideration.

Additionally, if one were to attempt to apply this procedure to other cities, it would require a different extraction process and implementation process depending on the way in which the data is provided from said cities.

7.3. Future research directions and recommendations

Looking forward, there are several promising options for further research that can build upon this thesis. One key area is exploration of additional semantic attributes that could enrich the utility maps even further. Additionally, one can perhaps find a way to create an algorithm that routes cables in a more successful manner, and allows for the connection of trenches around corners.

Another direction for future research is the integration of larger and more diverse datasets. This could involve expanding the scope of the data collection to include more varied urban environments or integrating data collected from newer technologies. However, these would require for a better system of cables and pipelines naming standards be set in place. Perhaps, even taking cable names from different cities across the Netherlands and creating a process that makes these interoperable.

Additionally, automating the data integration process to reduce the reliance on manual interventions could dramatically increase the efficiency and applicability of the enrichment process. Developing advanced algorithms that can handle data discrepancies and integrate multiple data sources autonomously would be a crucial step forward in the evolution of utility network GIS applications.

Throughout the thesis process, we ran into many issues that pertain to the extraction of the surveyed data. One recommendation, if implemented, that would greatly improve the process would be to find the positions of each cable in the relevant CRS rather than locally per trench. In this case, if the individual positions of cables and pipelines were provided in EPSG:28992, the extraction of information would require minimal processing. Additionally, as exemplified by the two cities studied, the depth of the cables can be provided as either actual elevation or as depth below the surface. It is recommended in this thesis that surveyors provide both the elevation and depth below the surface. Elevation will help the use of this information for GIS purposes and the depth below the surface will aid in on-site construction purposes.

In conclusion, this thesis has demonstrated the feasibility of enhancing traditional 2D utility network maps through a semi-automatic method that integrates trial trench data to provide both spatial and semantic enrichment. Despite the methodological advances and the improvements in utility data accuracy, challenges remain in terms of data quality, availability, and the scalability of manual processes involved. Recommendations include a more holistic surveying and data storage practice. Future work could address issues by refining automation techniques, expanding data integration, and further enriching semantic attributes, potentially transforming the utility network management landscape significantly.

7.4. Reflection

Over the course of this thesis, we have ran into many challenges and obstacles that have allowed us to gain new knowledge in a few topics. The goal of enriching traditional utility network data using surveyed trench data has mainly required a focus on the extraction, integration and processing of data from varying sources.

A significant portion of the work involved handling and integrating different data formats, as highlighted in the methodologies used for both Enschede and Rotterdam. This required the use of Python scripting and QGIS to streamline the data enrichment process. Specific libraries, as seen in Table 4.1, were utilised for calculating cable positions and matching cable names between successive trenches.

The quality of the utility network data received posed considerable challenges. Issues such as inconsistent labeling and disconnected components within the datasets complicated the integration process. Extra time and effort were dedicated to cleaning and matching the data, as well as developing semi-automatic processes to handle these inconsistencies. Despite these efforts, the need for manual intervention in some cases highlighted the limitations of current datasets and the methodologies implemented.

Additionally, the thesis discusses the importance of standardization in utility data management. Aligning surveyed data with existing standards is one of the recommendations for enhancing data consistency and interoperability.

One of the key reflections from this research is the realisation of the potential and limitations of semi-automatic data integration processes. While enriching utility maps was made possible using the spatial and semantic data provided by utility trenches, the reliance on manual work and intervention indicated the need for change in how data is stored when surveyed. The feasibility of enriching 2D utility network models was demonstrated to a small extent, but the quality and inconsistency issues in the data remained a dominant challenge.

In conclusion, the journey of this thesis has been both challenging and rewarding. It has highlighted the complexities involved in utility data enrichment and the necessity for continued advancements in data processing and standardization. It is our hope that the contributions and suggestions made through this research will provide assistance in future work within the realm of automating and enhancing the integration of utility network data.

A. Relevant Figures

A.1. Enschede

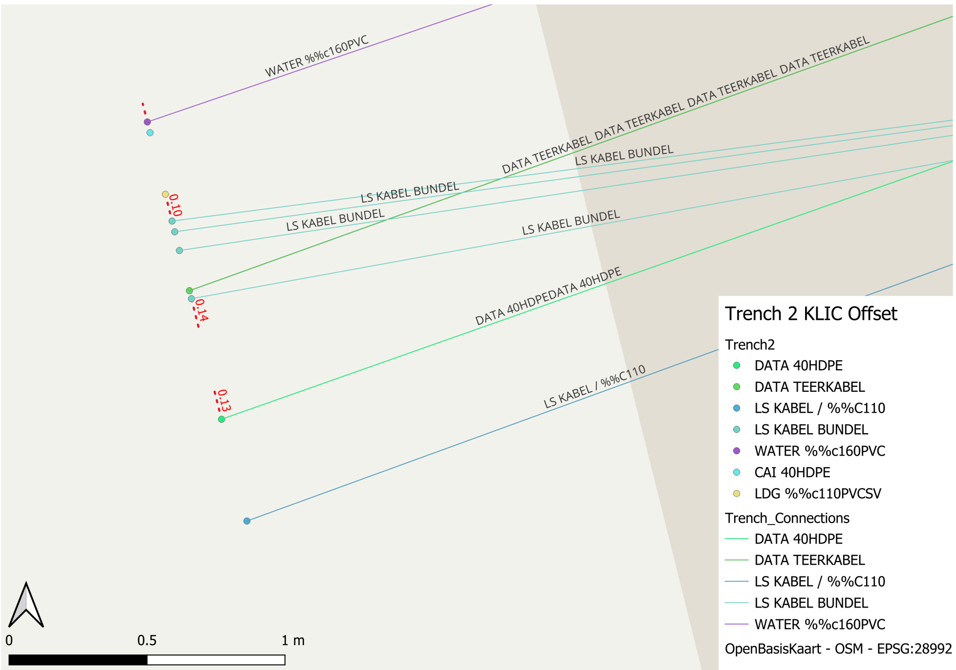


Figure A.1.: Brammelerdwarstraat: KLIC offset, trench 2

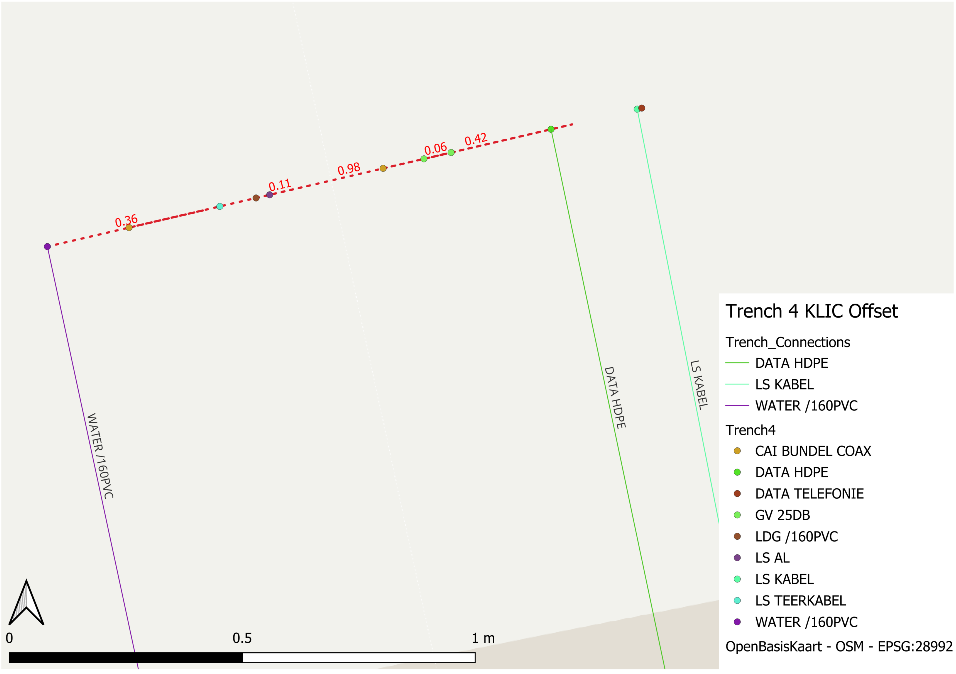


Figure A.2.: Brammelerdwarstraat: KLIC offset, trench 4

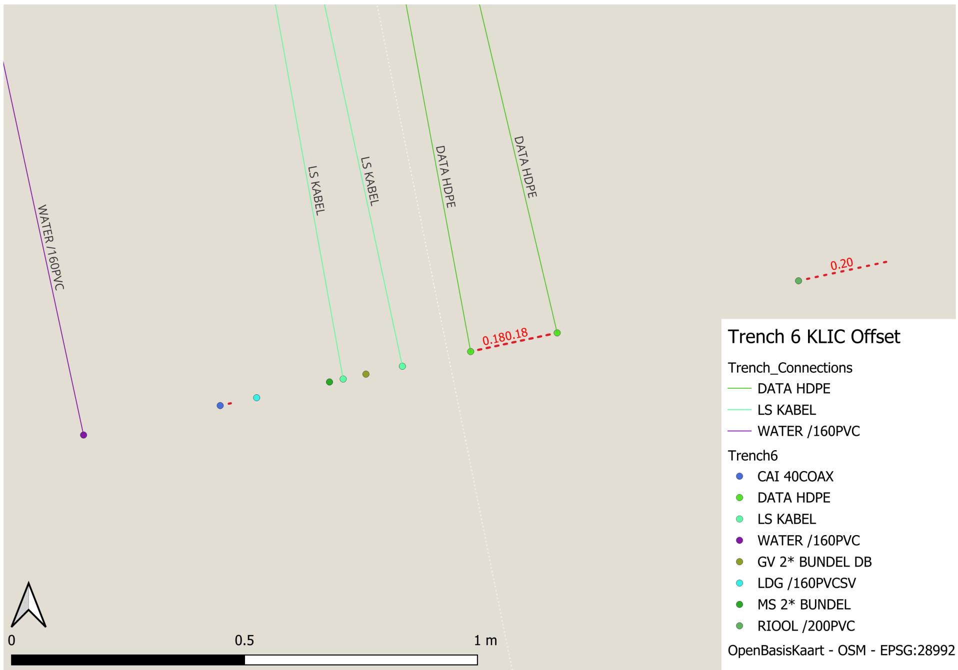


Figure A.3.: Brammelerdwarstraat: KLIC offset, trench 6

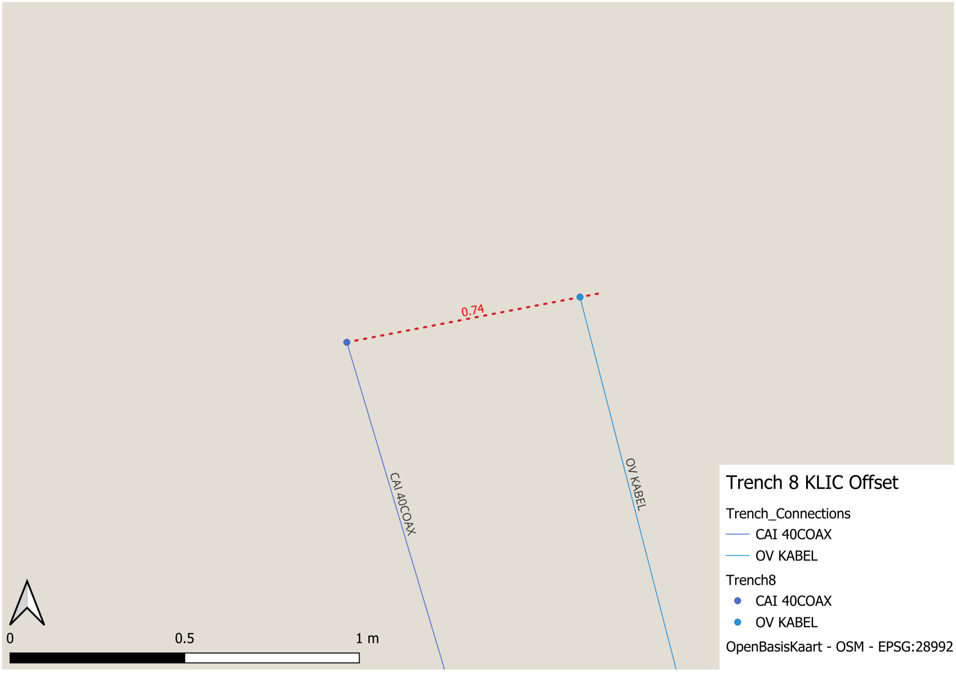


Figure A.4.: Brammelerdwarstraat: KLIC offset, trench 8

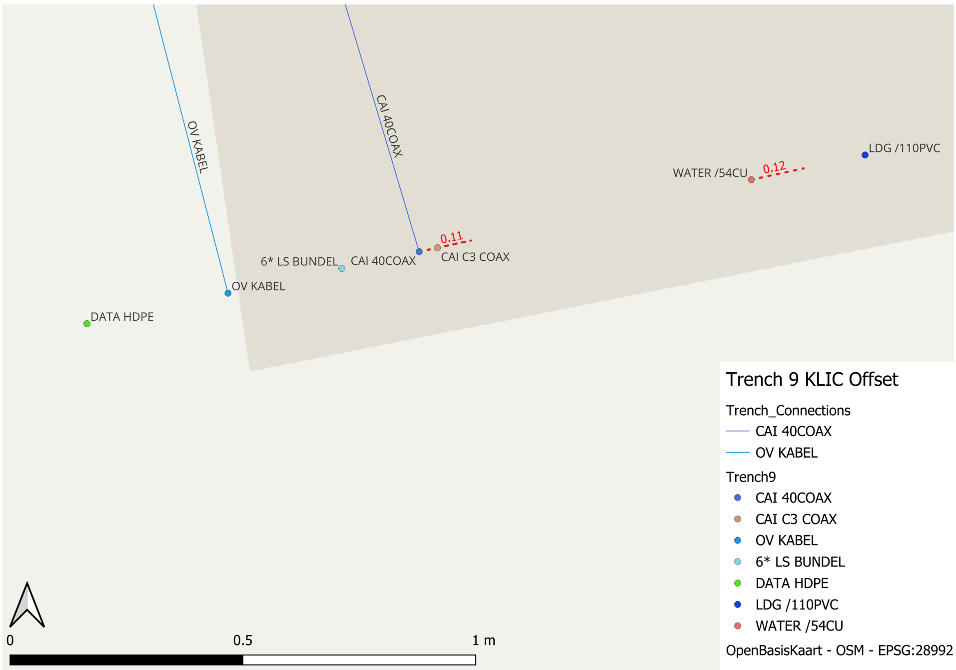


Figure A.5.: Brammelerdwarstraat: KLIC offset, trench 9

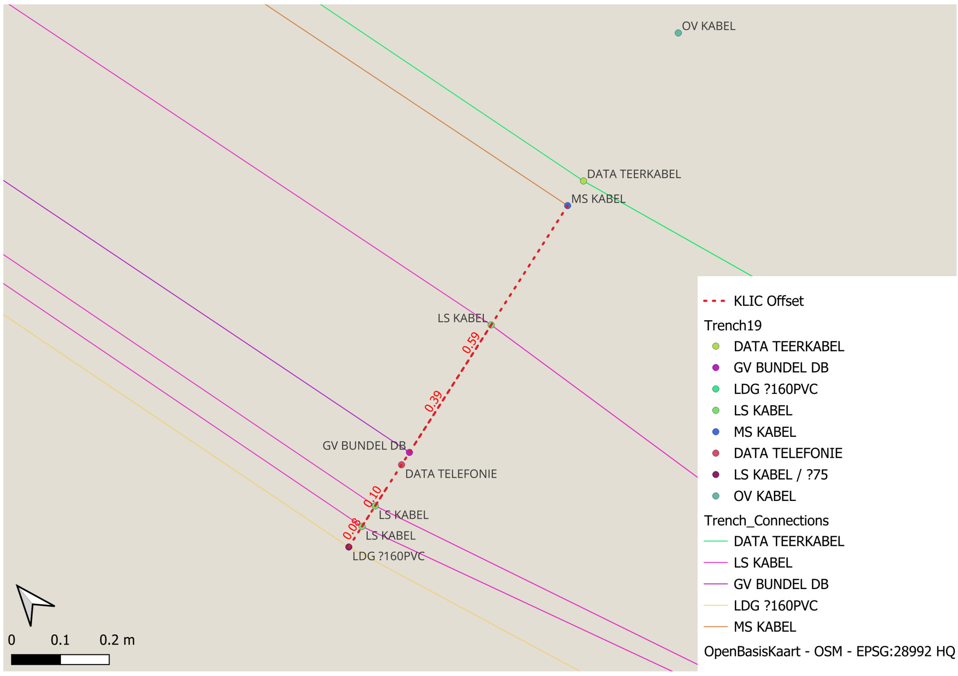


Figure A.6.: Deurningerstraat: KLIC offset, trench 19

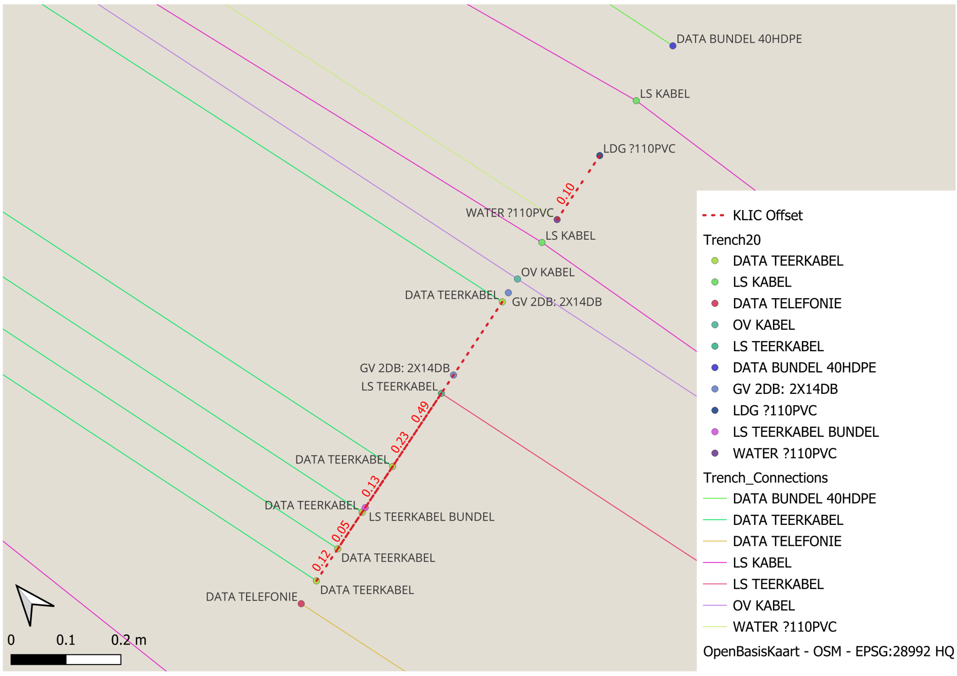


Figure A.7.: Deurningerstraat: KLIC offset, trench 20

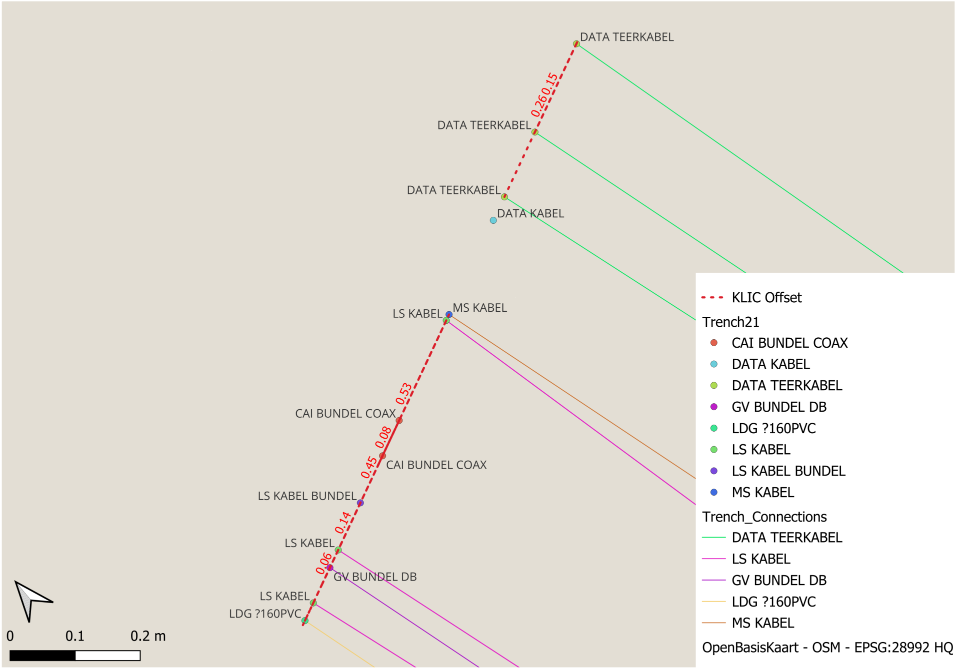


Figure A.8.: Deurningerstraat: KLIC offset, trench 21

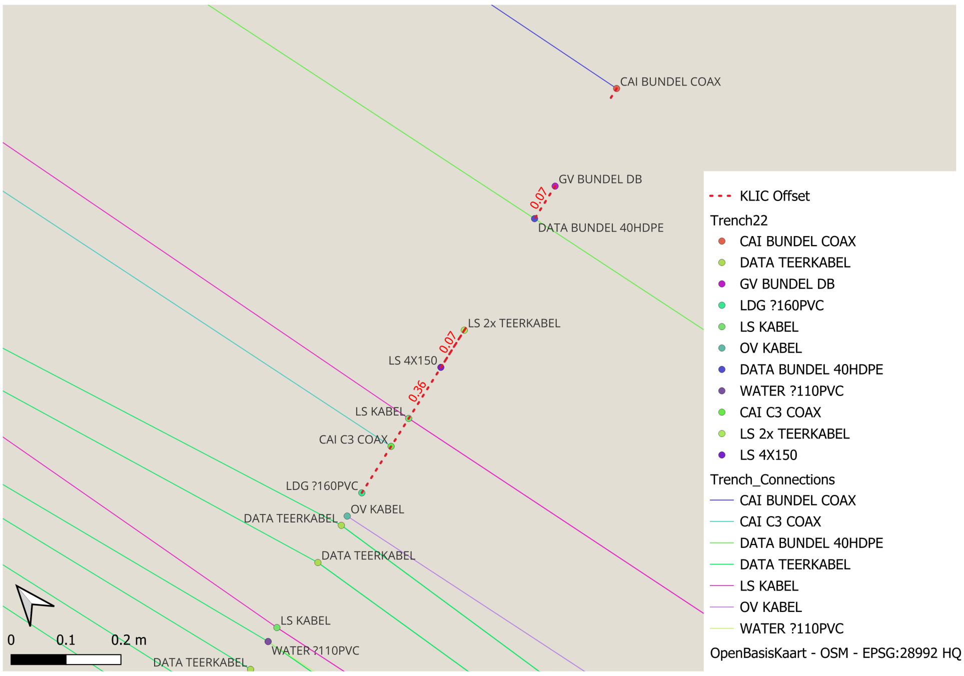


Figure A.9.: Deurningerstraat: KLIC offset, trench 22

A.2. Rotterdam

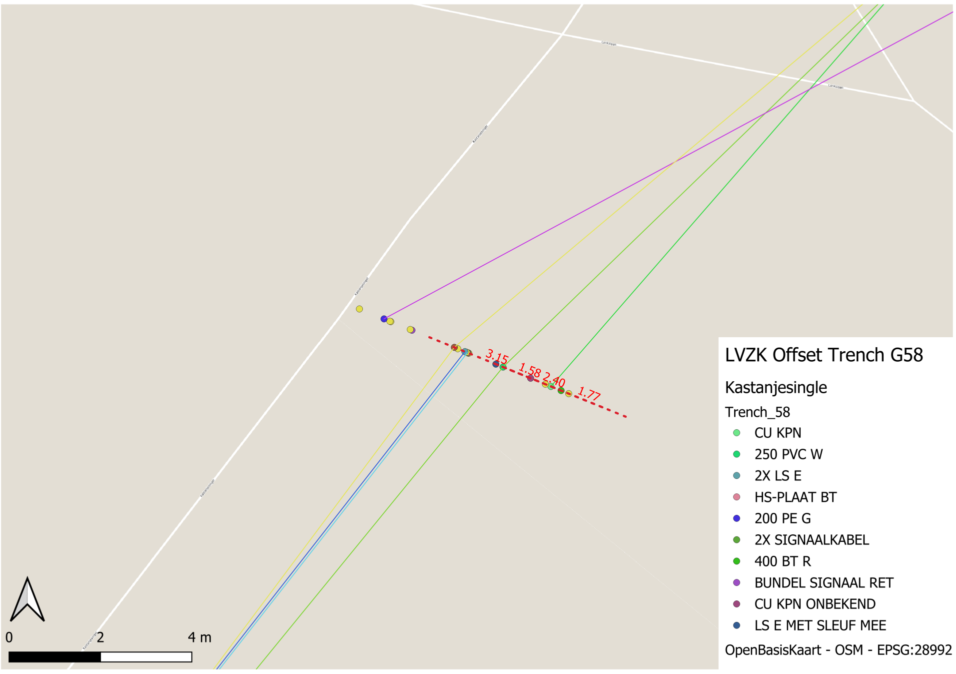


Figure A.10.: Rotterdam: LVZK offset, trench G58

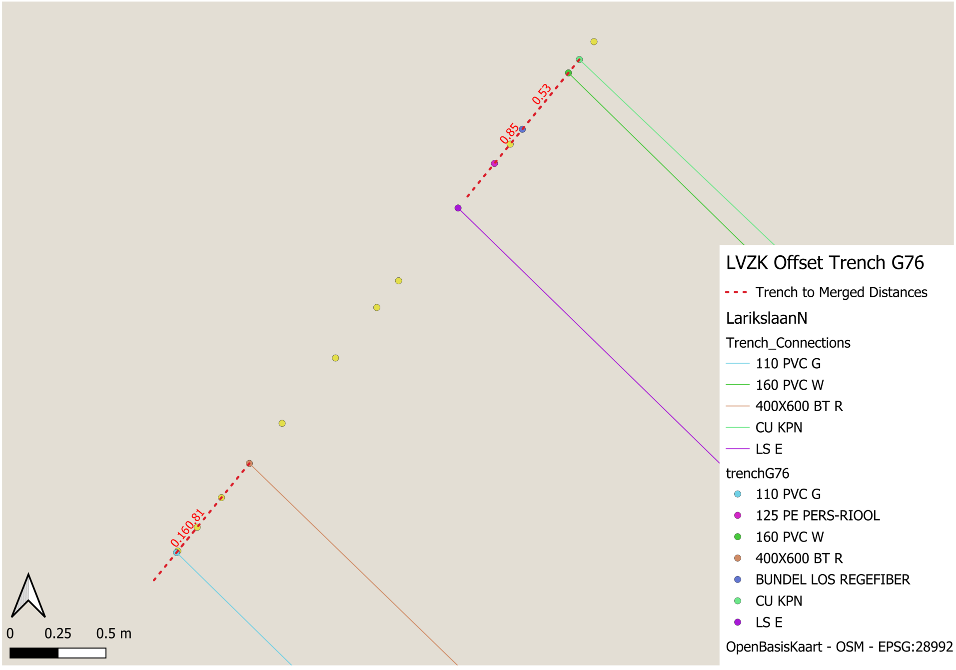


Figure A.11.: Rotterdam: LVZK offset, trench G76

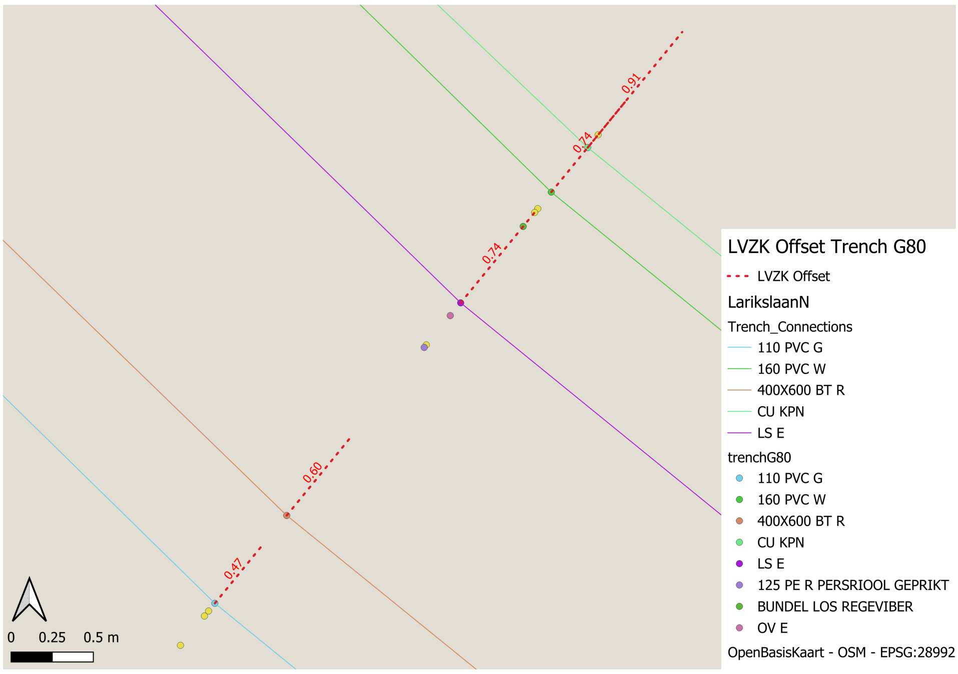


Figure A.12.: Rotterdam: LVZK offset, trench G80

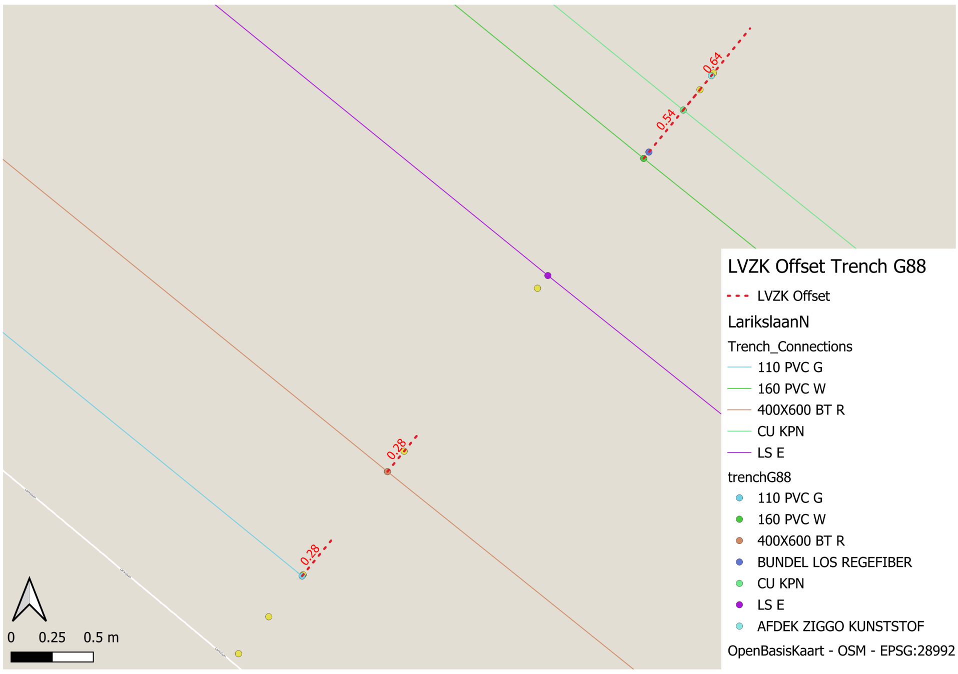


Figure A.13.: Rotterdam: LVZK offset, trench G88

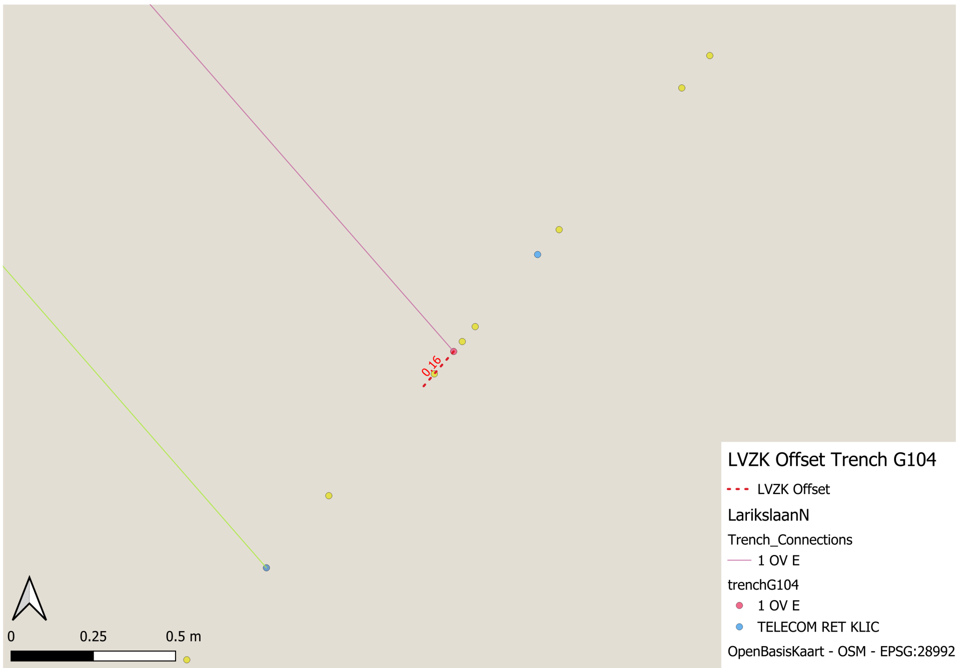


Figure A.14.: Rotterdam: LVZK offset, trench G104

B. Relevant Scripts

```python
import ezdxf
import csv
from scipy.spatial.distance import cdist
import numpy as np

def extract_dxf_data(filepath):
    doc = ezdxf.readfile(filepath)
    msp = doc.modelspace()

    # Lists to store Trench Labels and their positions
    trench_labels = []
    trench_positions = []

    # List to store cleaned Cable information
    cable_info_cleaned = []
    cable_positions = []

    # Extract Trench Labels from B-PR-PROFIELNR layer
    for text in msp.query('TEXT[layer=="B-PR-PROFIELNR"]'):
        label = text.dxf.text.split()[0]  # First word of the text content
        position = text.dxf.insert
        trench_labels.append(label)
        trench_positions.append([position.x, position.y])

    # Extract and clean Cable Information from B-NW-MAATVOERING layer
    for mtext in msp.query('MTEXT[layer=="B-NW-MAATVOERING"]'):
        info = mtext.text.replace('^I', ' ').strip()  # Clean the text
        position = mtext.dxf.insert
        cable_info_cleaned.append(info)
        cable_positions.append([position.x, position.y])

    # Find the nearest trench label for each cable
    trench_positions_np = np.array(trench_positions)
    cable_positions_np = np.array(cable_positions)
    distances = cdist(cable_positions_np, trench_positions_np, 'euclidean')
    nearest_trench_indices = np.argmin(distances, axis=1)
    nearest_trench_labels = [trench_labels[index] for index in
        nearest_trench_indices]

    # Splitting cable information into distance, height, and description
    cable_data_split = []
    for info in cable_info_cleaned:
        parts = info.split(maxsplit=2)
        if len(parts) > 2 and '-' in parts[2]:
            description_parts = parts[2].split(' - ', 1)
            if len(description_parts) == 2:
                parts[2] = description_parts[1]
        cable_data_split.append(parts)

    # Combine nearest trench label with split cable information
    combined_data = []
    for trench_label, cable_data in zip(nearest_trench_labels, cable_data_split)
        :
```

```
        if len(cable_data) == 3:
            distance, height, description = cable_data
        else:
            distance, height = cable_data
            description = ''
        combined_data.append([trench_label, distance, height, description])

    return combined_data

def write_to_csv(filepath, data):
    with open(filepath, 'w', newline='') as file:
        writer = csv.writer(file)
        writer.writerow(['Trench Label', 'Distance', 'Height', 'Description'])
        writer.writerows(data)

# Replace 'filepath' with the path to your DXF file
dxf_filepath = "filepath"
combined_data = extract_dxf_data(dxf_filepath)

# Specify the CSV file path where you want to save the output
csv_filepath = "..../extracted_cable_info.csv"
write_to_csv(csv_filepath, combined_data)
```

Listing B.1: Python script for extracting and processing DXF data and writing to CSV

```
from qgis.core import QgsFeature, QgsGeometry, QgsVectorLayer, QgsField, QgsProject, QgsPointXY
from PyQt5.QtCore import QVariant

def add_csv_as_layer(file_path, layer_name, x_field, y_field, crs_epsg, field_types):
    uri = f"file:///{file_path}?delimiter=,&xField={x_field}&yField={y_field}&crs=epsg:{crs_epsg}&fieldTypes={field_types}"
    return iface.addVectorLayer(uri, layer_name, "delimitedtext")

def calculate_average_point(points):
    x_coords, y_coords = zip(*[(point.x(), point.y()) for point in points])
    centroid = QgsPointXY(sum(x_coords) / len(x_coords), sum(y_coords) / len(y_coords))
    return centroid

def calculate_dynamic_intermediate_point(t1_points, t2_points):
    # Calculate average points for each trench
    avg_t1_point = calculate_average_point(t1_points)
    avg_t2_point = calculate_average_point(t2_points)

    count_t1 = len(t1_points)
    count_t2 = len(t2_points)
    bias_ratio = 0.1 if count_t1 > count_t2 else 0.9  # Adjust the bias ratio as needed for your use-case

    # Calculate the intermediate point using the determined bias
    intermediate_x = avg_t1_point.x() * (1 - bias_ratio) + avg_t2_point.x() * bias_ratio
    intermediate_y = avg_t1_point.y() * (1 - bias_ratio) + avg_t2_point.y() * bias_ratio

    return QgsPointXY(intermediate_x, intermediate_y)

# List of trench file paths
trench_files = [
    # Add Trench CSV Paths
```

```
]

x_field, y_field, crs_epsg, field_types = "E", "N", "28992", "integer,double,
    string,double,double,double,string"

# Create lines layer to store connections
crs = "EPSG:28992"
lines_layer = QgsVectorLayer(f"LineString?crs={crs}", "Trench_Connections", "
    memory")
pr = lines_layer.dataProvider()
pr.addAttributes([QgsField("description", QVariant.String)])
lines_layer.updateFields()
QgsProject.instance().addMapLayer(lines_layer)
lines_layer.startEditing()

# Load and connect each trench sequentially
for i in range(len(trench_files) - 1):
    trench_1_layer = add_csv_as_layer(trench_files[i], f"Trench{i+1}", x_field,
        y_field, crs_epsg, field_types)
    trench_2_layer = add_csv_as_layer(trench_files[i+1], f"Trench{i+2}", x_field
        , y_field, crs_epsg, field_types)

    if trench_1_layer is None or trench_2_layer is None:
        continue

    descriptions = set(feat['description'] for feat in trench_1_layer.
        getFeatures())
    for desc in descriptions:
        t1_points = [QgsPointXY(feat.geometry().asPoint()) for feat in
            trench_1_layer.getFeatures() if feat['description'] == desc]
        t2_points = [QgsPointXY(feat.geometry().asPoint()) for feat in
            trench_2_layer.getFeatures() if feat['description'] == desc]
        if t2_points:
            intermediate_point = calculate_dynamic_intermediate_point(t1_points,
                 t2_points)
            for t1_feat in [feat for feat in trench_1_layer.getFeatures() if
                feat['description'] == desc]:
                t1_point = QgsPointXY(t1_feat.geometry().asPoint())
                line_to_intermediate = QgsFeature()
                line_to_intermediate.setGeometry(QgsGeometry.fromPolylineXY([
                    t1_point, intermediate_point]))
                line_to_intermediate.setAttributes([desc])
                pr.addFeature(line_to_intermediate)
                for t2_feat in [feat for feat in trench_2_layer.getFeatures() if
                     feat['description'] == desc]:
                    t2_point = QgsPointXY(t2_feat.geometry().asPoint())
                    line_from_intermediate = QgsFeature()
                    line_from_intermediate.setGeometry(QgsGeometry.
                        fromPolylineXY([intermediate_point, t2_point]))
                    line_from_intermediate.setAttributes([desc])
                    pr.addFeature(line_from_intermediate)

lines_layer.commitChanges()
```

Listing B.2: Python script for importing multiple trenches and drawing connections

```
from qgis.core import (
    QgsProject, QgsFeature, QgsGeometry, QgsVectorLayer,
    QgsField, QgsWkbTypes, QgsPointXY
)
from PyQt5.QtCore import QVariant

```

```python
gpkg_path = "[...]/Proefsleuven Deurningerstraat.gpkg"

# Layer names that should intersect with 'Profielen'
layer_names = [
    #....Add KLIC CABLE LAYERS
    ]

# Load 'Profielen' layer
profielen_layer = QgsVectorLayer(f"{gpkg_path}|layername=polylines|subset=\"layer\"='Profielen'", "Profielen", "ogr")
if not profielen_layer.isValid():
    print("Failed to load 'Profielen' layer")
else:
    # Create a memory layer for storing intersection points, adding a 'layer_name' attribute
    intersection_layer = QgsVectorLayer("Point?crs=EPSG:28992", "Intersection Points", "memory")
    dp = intersection_layer.dataProvider()
    dp.addAttributes([
        QgsField("id", QVariant.Int),
        QgsField("layer_name", QVariant.String)
    ])
    intersection_layer.updateFields()

    id_counter = 1  # Counter for intersection points

    # Process intersections with other specified layers
    for layer_name in layer_names:
        other_layer = QgsVectorLayer(f"{gpkg_path}|layername=polylines|subset=\"layer\"='{layer_name}'", layer_name, "ogr")
        if other_layer.isValid():
            for feature1 in profielen_layer.getFeatures():
                for feature2 in other_layer.getFeatures():
                    intersection = feature1.geometry().intersection(feature2.geometry())
                    if intersection and not intersection isEmpty():
                        if intersection.wkbType() == QgsWkbTypes.PointZ:
                            point = intersection.asPoint()
                            new_point = QgsPointXY(point.x(), point.y())
                            new_feature = QgsFeature()
                            new_feature.setGeometry(QgsGeometry.fromPointXY(new_point))
                            new_feature.setAttributes([id_counter, layer_name])
                            dp.addFeature(new_feature)
                            id_counter += 1
                        elif intersection.wkbType() == QgsWkbTypes.MultiPointZ:
                            points = intersection.asMultiPoint()
                            for point in points:
                                new_point = QgsPointXY(point.x(), point.y())
                                new_feature = QgsFeature()
                                new_feature.setGeometry(QgsGeometry.fromPointXY(new_point))
                                new_feature.setAttributes([id_counter, layer_name])
                                dp.addFeature(new_feature)
                                id_counter += 1

    # Add the new layer to the project
    QgsProject.instance().addMapLayer(intersection_layer)
    print("Intersection points have been added to the map.")
```

Listing B.3: Finding KLIC Points Script

Bibliography

Autodesk Inc. (Accessed 2024). *AutoCAD*. Autodesk Inc., San Rafael, CA, USA.

Becker, T., Nagel, C., and Kolbe, T. H. (2011). Integrated 3d modeling of multi-utility networks and their interdependencies for critical infrastructure analysis.

Chundeli, F. A. (2017). Using 3d gis as a decision support tool in urban planning. *Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance*.

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/W7:43–50.

Döner, F., Thompson, R., Stoter, J., Lemmen, C., Ploeger, H., van Oosterom, P., and Zlatanova, S. (2010). 4d cadastres: First analysis of legal, organizational, and technical impact—with a case study on utility networks. *Land Use Policy*, 27(4):1068–1081.

Fossatti, F., Agugiaro, G., olde Scholtenhuis, L., and Dorée, A. (2020). Data modeling for operation and maintenance of utility networks: Implementation and testing. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, VI-4/W1-2020:69–76.

Jaw, S. W., Van Son, R., Soon, V. K. H., Schrotter, G., Kiah, R. L. W., Ni, S. T. S., and Yan, J. (2018). The need for a reliable map of utility networks for planning underground spaces. In *2018 17th International Conference on Ground Penetrating Radar (GPR)*, pages 1–6.

Koppens, B. (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. [Accessed 2024].

Lieberman, J. (2019). Model for underground data definition and integration (muddi) engineering report. visited on 11/27/2021.

McGuffin, M. J., Servera, R., and Forest, M. (2022). Path tracing in 2d, 3d, and physicalized networks. *IEEE transactions on visualization and computer graphics*, PP.

olde Scholtenhuis, L. L., den Duijn, X., and Zlatanova, S. (2018). Representing geographical uncertainties of utility location data in 3d. *Automation in Construction*, 96:483–493.

Pavlidou, A. (2022). Integrated modeling of utility networks in the urban environment. Accessed: 2023-05-14.

Python Software Foundation (Accessed 2024). Python programming language. https://www.python.org/.

QGIS (Accessed 2024). QGIS Geographic Information System. https://qgis.org/.

Racz, P., Syfuss, L., Schultz, C., van Buiten, M., Scholtenhuis, L. O., Vahdatikhaki, F., and Dorée, A. (2017). Decision support for test trench location selection with 3d semantic subsurface utility models. pages 68–75.

van den Berg, H. and Janssen, P. (2021). Information model cables and pipes (imkl) - version 2.0. visited on 12/02/2021.

Vishnu, E. and Sameer, S. (2020). Ogc citygml 3d city models enriched with utility infrastructures for developing countries. *Journal of the Indian Society of Remote Sensing*, 49:813 – 826.

Ward, B., Selby, A., Gee, S., and Savić, D. (2017). Deterioration modelling of small-diameter water pipes under limited data availability. *Urban Water Journal*, 14:743 – 749.

Colophon

This document was typeset using LaTeX, using the KOMA-Script class `scrbook`. The main font is Palatino.