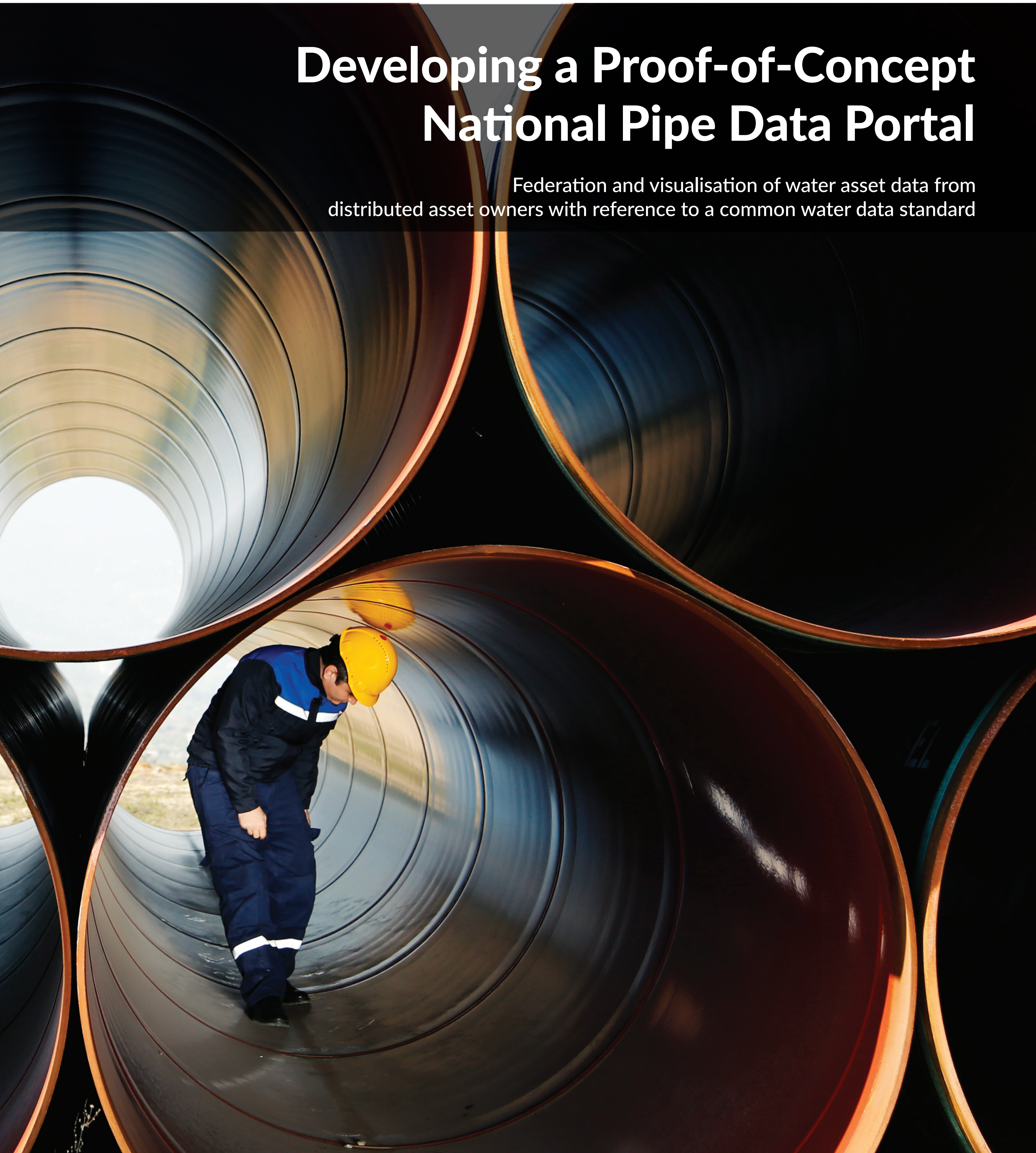
Developing a Proof-of-Concept National Pipe Data Portal

Federation and visualisation of water asset data from distributed asset owners with reference to a common water data standard



www.quakecentre.co.nz

Acknowledgements:

The Quake Centre would like to acknowledge the following for their valuable contribution to this project:

- Mark Johnson – Christchurch City Council
- Mayank Katira – Christchurch City Council
- Irmana Sampedro – Christchurch City Council
- Kim Webster – Tauranga District Council
- Lynda Ferguson –Tauranga District Council
- Dukessa Blackburn-Huettner – Auckland Council Healthy Waters
- Nina Sardareva – Auckland Council Healthy Waters
- Jean-Pierre (JP) Gallet – Auckland Council Healthy Waters
- Mark Thomas – Nextspace
- Alex Gnum – Nextspace
- Dr James Williams – University of Canterbury
- Jing Wu – Master of Applied Data Science Student, University of Canterbury
- Manveer Bhullar – Master of Applied Data Science Student, University of Canterbury
- Ziqi Gu – Master of Applied Data Science Student, University of Canterbury
- Graham Clark – Fulton Hogan

Quake Centre – Delivering Industry Solutions | Tauranga City | Christchurch City Council | Auckland Council – Te Kaunihera o Tāmaki Makaurau

Developing a Proof-of-Concept National Pipe Data Portal

Federation and visualisation of water asset data from distributed asset owners with reference to a common water data standard

Research Acknowledgement

This report is an output from the Quake Centre's Building Innovation Partnership programme (BIP), which is jointly funded by Industry and the Ministry of Business Innovation and Employment (MBIE).

Under the contract that the University of Canterbury' Quake Centre has with MBIE , this report forms partial completion of Research Aim 1.1 *Implementation of NZ Metadata Standards* and the complete fulfilment of Critical Step 1.1.3 *Pilot Portal*.

Executive Summary

The Quake Centre's Building Innovation Partnership (BIP) has developed the proof-of-concept for a National Pipe Data Portal. In this project, students from the Master of Applied Data Science (MADS) worked with three local authorities, namely: Christchurch City Council, Tauranga District Council and Auckland Council. The MADS students:

1. Undertook a quality assessment of the stormwater asset data of the three councils
2. Compared the as-built stormwater data with a beta-version of a national data standard
3. Mapped the councils' data to the beta-standard
4. Federated the three councils' data and visualised it as a single entity
5. Ran simple queries across the federated data.

In addition, this project:

- Informed and provided a testing ground to assist the development of Nextspace's *Bruce* tool for managing data
- Informed the development of the *3 Waters Metadata Standard*
- Initiated the development of a dictionary for 3 waters data in NZ
- Developed some preliminary tools for assessing 3 waters data quality.

This successful project developed, tested and confirmed a viable path towards developing a full National Pipe Data Portal, and the development of a full National Digital Infrastructure Model. Lessons learned include:

- Mapping infrastructure data to a national standard is technically viable and scalable.
- Federating infrastructure data from disparate sources can be carried out in a way that makes visualisation and analysis open and transparent.
- There are few technical barriers to implementing the sharing of infrastructure data based upon open data standards.
- There is value in undertaking the federation of national infrastructure data using an enterprise solution provided this is based on open standards.

Greg Preston,
Project Sponsor, Quake Centre

Angus Bargh,
Project Manager, Open Plan Ltd

Introduction

This report provides a summary of the work undertaken in support of the Building Innovation Partnership project to develop a proof-of-concept National Pipe Data Portal. This 12-week project was assisted by students from the Master of Applied Data Science (MADS) course at the University of Canterbury. The project sought to demonstrate the concept and benefits of federating infrastructure asset data using an agreed national data standard. The project also identified the challenges to overcome in moving beyond the proof-of-concept stage to a pilot trial with a wider range of stakeholders. The project also tested Nextspace's data federation and visualisation tool *Bruce.*

Background and Industry Need

The development of a national data standard for 3 waters infrastructure provides for a unified approach to assessing water infrastructure at a national level. Asset data currently held by the asset owners exists in unique data structures. These datasets have varied attributes which define and describe the assets and reside in many different types of data storage systems. Each council and water authority have developed their own unique approach to defining their asset base. This makes federation of data difficult across and between industry sectors.

A national data standard provides a framework for asset owners to standardise their data. There are two pragmatic reasons why standardisation at a national level is difficult to achieve. The first reason is that each asset owner is likely to have legacy systems, processes and protocols which are highly dependent on the uniqueness of the asset data which they support. If the base asset data is changed to meet a national standard, there will be a cost to replicate legacy processes which depend on the base data. The second reason is simply one of time. Pragmatically, an asset owner is unlikely to dedicate time to re-formatting their asset data and rebuilding associated legacy processes until there is a clear reason or change-moment for their organisation to invest in-demand resources to this task. There has been strong support amongst asset owners for a national 3 waters standard but, while the case for standardisation remains compelling, the decision to commit to a new standard and invest the effort to reformat data is often competing with other high priority tasks which the infrastructure sector is increasingly facing. Standardisation at a national level is possible but would take considerable time.

This proof-of-concept project provides an alternative approach to standardising asset data 'at-source'. It proposes a federation approach whereby asset data, provided by an asset owner, is not re-formatted to align with a standard but remains in its native format and is simply 'mapped' to a standard. In this process, the correspondence (or mapping) is defined at three levels; firstly between the asset class of the source data and the asset class of the national standard; secondly between the attributes for a particular asset class from the source data and the corresponding attribute data in the national standard and; thirdly between the permitted values for each attribute which exist in the source data and those which exist in the national standard.

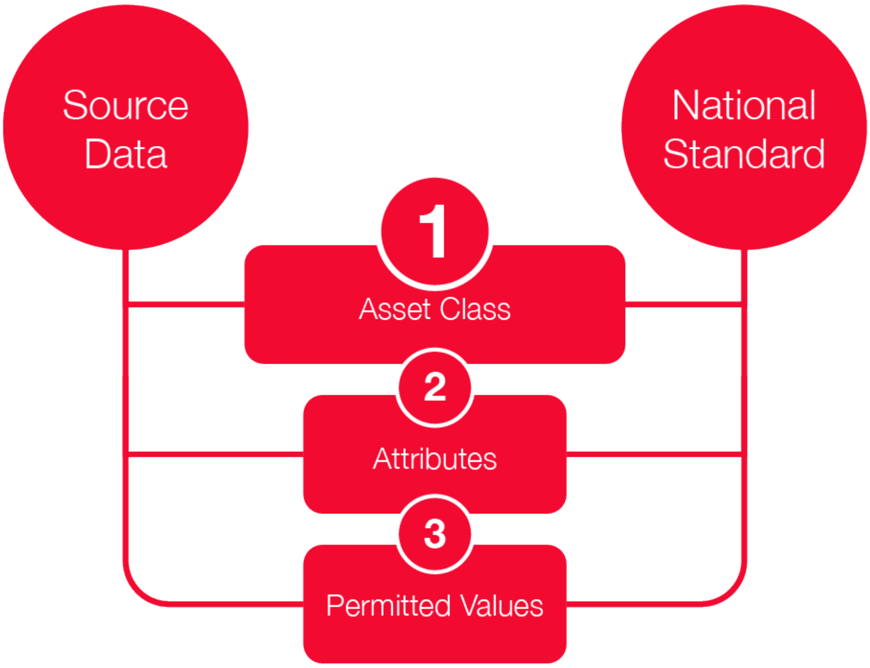


Figure 1: Data mapping levels between source data and national standard.
Note: Steps 1, 2 and 3 are the data mapping steps.

By defining and storing the data-mapping between the asset owner's data and the national standard, the asset owner isn't required to re-format their source data in order to realise the benefits which arise from federating their data with that from other asset owners. This approach achieves federation of data through a web application where data is either loaded dynamically (using a web service) or is stored within a cloud hosting environment. The overarching concept is that by mapping separate asset datasets to a national standard, all data can be considered to form a single dataset which can be visualised and analysed together *even though the individual datasets remain in their native hierarchy and format.*

A key consideration with this approach is ensuring there is completeness in the data-mapping between the source data and the national data standard. Completeness, in this context, means that all the asset classes, data attributes and data value types required by the data standard are contained within the source data from each council and asset owner. It is therefore very important for this proof-of-concept trial to also include a data quality checking process which reviews the completeness of each dataset with respect to the data standard. It is recognised that the resultant data quality report will provide an important source of feedback from which each asset-owner can assess the effort required to improve their data and create a comprehensive dataset. It is also important that any asset classes, data attributes or data value types which are contained within the data from each asset owner, but omitted from the national data standard, are also recorded as a means of determining the fitness of the data standard itself.

This approach also has the benefit of supporting dynamic standards. This means that, should the national standard change, only the mapping (correspondence) needs to be updated for each dataset for the federation to be preserved. By contrast, if each asset owner were required to maintain consistency with a national standard in order for federation to be achieved, this would require the re-formatting of source asset data every time a standard is updated. This would be prohibitively complex and time-consuming.

It is worth noting that in defining a national data standard for 3 waters infrastructure, there is an expectation that stakeholders will align their asset data with this standard over time – increased alignment will create better data interoperability and simpler mapping processes. However, this proof-of-concept project demonstrates that benefits can be captured now with limited requirements placed on asset owners to invest significant time or effort.

Development of national data standards for 3 waters

This project also feeds into the development of a national metadata standard for 3 waters. A previous project sponsored by Land Information New Zealand (LINZ), developed a draft metadata standard for 3 waters. This standard was published in 2017 and it was passed to Water NZ as custodians. However, for a number of reasons, the standard in its current form has not been adopted. On behalf of Water NZ, the Quake Centre is leading the development of an amended 3 waters metadata standard, to a point where it can be re-published and implemented nationally. As an interim step the Quake Centre has adopted a beta-version of a national standard for reticulation. This beta-version standard is limited to reticulation assets that are underground, e.g., pipes and chambers. It has been developed from the LINZ published standard taking account of industry feedback. This standard has been largely evolved by Graham Clark, formerly of Christchurch City Council and now at Fulton Hogan. The beta standard was used as the reference standard for this project. Further development of this standard is underway with the aim of publishing *3 Waters Metadata Standard V1.0 (Reticulation)* by the end of July 2019.

Project Overview

Goal and success criteria

The goal of this proof-of-concept data federation project was to demonstrate that asset data from a distributed sample of asset owners can be federated through a schema-mapping process that conforms to an agreed data standard. This process is best characterised as follows:

- Data is federated through an online application.
- Source asset data remains unchanged from its native format and structure.
- Federation is achieved by utilising a standardised data schema which includes asset classes, class attributes and asset value definitions.

The definition or demonstration of success for this proof-of-concept project was agreed between stakeholders at the outset of the project. The following success criteria are the specific steps which the project set-out to achieve:

1. Load each council's data into a central online repository (or enable data access through a web data service).
2. Develop mapping protocols for each council dataset to the agreed national beta-standard for stormwater with particular focus on pipes and chambers.
3. Assess the quality of each council's data in two parts. Firstly, assess the completeness, consistency and quality of the data. Secondly, assess the alignment of the data with the national beta-standard.
4. Visualise the federated data in an online map-based viewer. Ideally the visualisation would allow for individual assets to be identified visually and, when selected, provide a table view of the attributes and values of that asset (in the format described by the national beta-standard).
5. Undertake a quantitative query on the federated dataset. An example would be to assess the total length of pipe which is constructed from asbestos-cement across all council data. Another example would be to assess the length of pipe which is older than 30 years and is located at an invert level greater than 2m.

The proof-of-concept project considered that data should be federated from a minimum of three councils in order for the project to be considered a success.

Methodology

The following provides a brief overview of the methodology for the project.

Activity or element	**Proof-of-concept approach**	**Possible alternative approach which supports next-stage development**
Resourcing	The technical aspects of the project were achieved with the assistance of students from the Master of Applied Data Science at Canterbury University. Technical oversight was provided by the UC Quake Centre.	Asset owners to provide data directly and undertake schema mapping themselves with technical support provided by the Quake Centre. This approach would rely on a suitable suite of online data tools.
Data provision	Static data file provided at the start of project	Data provided through online service feed, or as regular data-downloads to a data repository.

Activity or element	Proof-of-concept approach	Possible alternative approach which supports next-stage development
Software	A software product called *Bruce* was provided by an Auckland-based company, Nextspace. This provided role/organisation login access, online data-upload, schema definitions, schema and data mapping functionality, data federation, data visualisation and simple querying of federated data.	At this early stage of development, it is unclear which software platform will prove the most effective and financially efficient. This is a future decision. In the meantime, *Bruce* will continue to play an enabling role in the development of a national infrastructure viewer, but Quake Centre consider that open-access to data and exportability of data/standard mappings are essential features for any software platform deployed.
Data mapping	Undertaken through the *Bruce* software using a 'click and drag' user-interface to map the source data to the beta-standard. This resulted in a schema mapping register which is stored in *Bruce.*	Next stage development will aim to allow for importing and exporting of schema mappings in addition to enhancements in the data mapping process to handle asset hierarchy and in-mapping functions and processes (such as the conversion of values in mm to m).
Data quality	Data quality was assessed by each student focusing on individual council's data. Data quality was scrutinised (e.g., analysis of null values) in addition to assessment of the data against the national beta-standard. This included an assessment of attributes which appeared in the source data but were not present in the beta-standard (and vice-versa).	There is currently a special project being progressed by the Quake Centre to standardise the quality reporting of asset datasets. This will provide a mechanism to automate and rank the quality of source asset data – both as a stand-alone dataset and against the national beta-standard.

Activity or element	Proof-of-concept approach	Possible alternative approach which supports next-stage development
Data visualisation	Data visualisation was undertaken through the *Bruce* software. This provides a GIS-type viewer which displays each asset layered on a base-map background. The viewer also has 3-D visualisation capabilities.	Data interoperability is a key tenet for the next stage of development. For data visualisation, it is expected that both source and mapped data will be in a consistent format for it to be displayed in the most common GIS systems. The *Bruce* software currently supports these formats.
Data analysis	Data analysis was undertaken separately by each student at the conclusion of the project. This was a manual process using either the basic analysis tools provided by the *Bruce* software or by exporting the federated data from *Bruce* and running queries in tools such as Excel (or Python scripts).	Next-stage development anticipates a standard set of data queries and analytical models will be developed for application across the entire federated dataset. This is part of the value proposition for a national infrastructure viewer in that standardised tools and processes can be developed once and re-used across data at a national level – bringing immediate value to regional asset owners and industry stakeholders.

Project outcomes

The proof-of-concept project concluded at the end of February 2019. Asset data representing stormwater pipes and chambers were successfully federated for Auckland Council, Tauranga City Council and Christchurch City Council. This represents approximately 0.5 million individual pipe sections across the three asset owners. This data was successfully visualised in a single viewer and analysis was undertaken on the federated data set.

The five success criteria for the project, as outlined above, were met and particular focus was placed on the data queries which were enabled, for the first time, by federation of the data across the three asset owners. This is a significant first step towards the improved future procurement and management of NZ's buried pipe infrastructure.

Figures 2 through 5 provide a summary representation of the outcomes of the project.

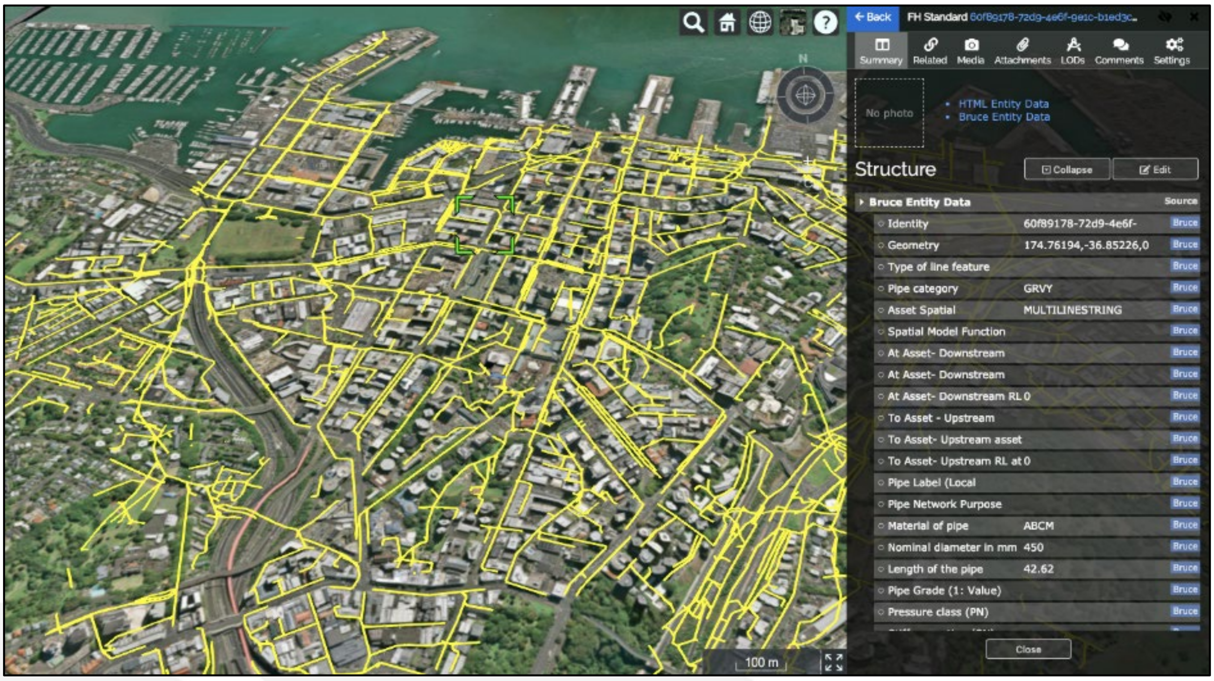


Figure 2: View of federated stormwater pipe assets in Auckland

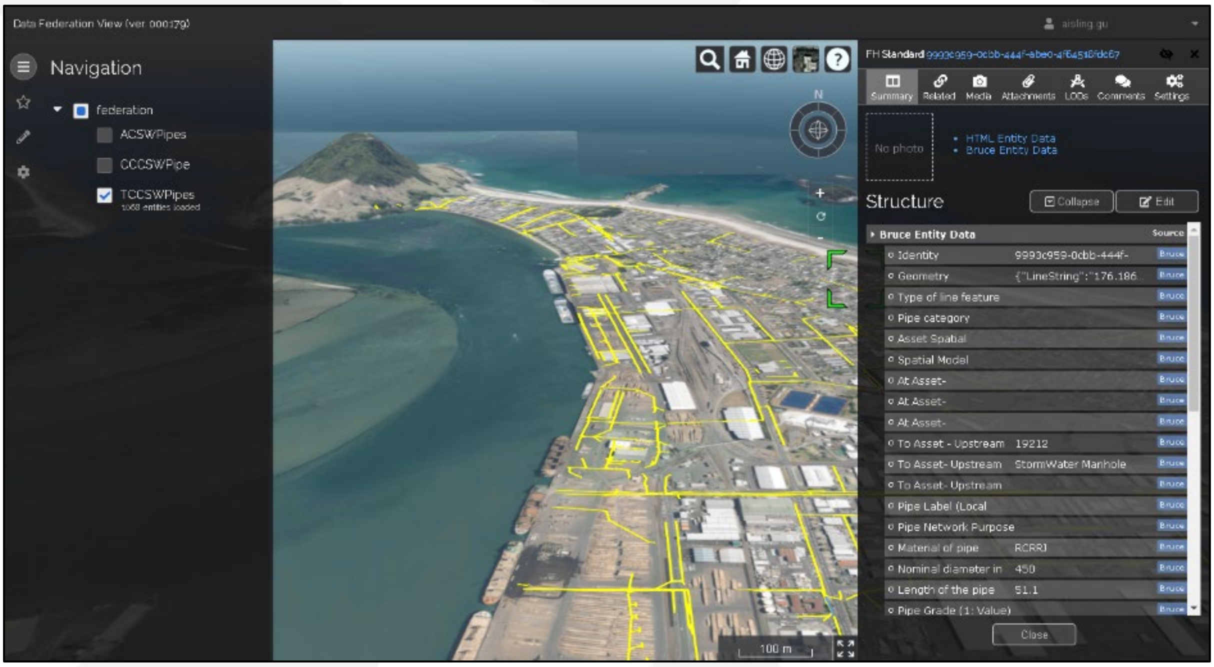


Figure 3: View of federated stormwater pipe assets in Tauranga

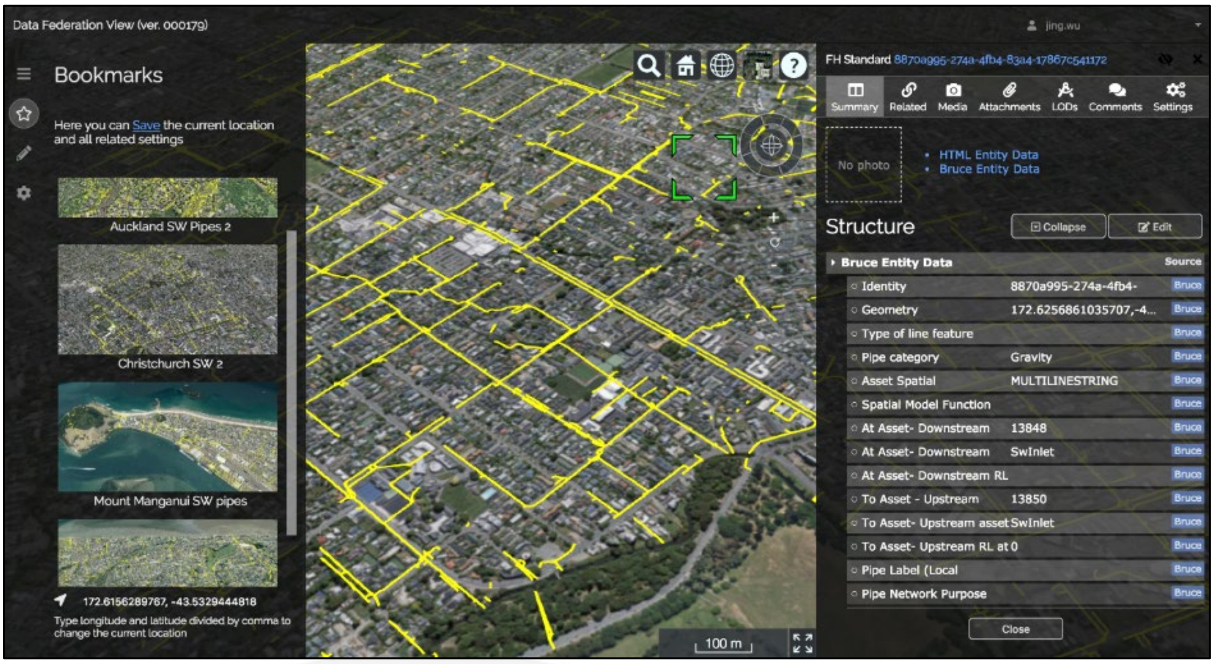


Figure 4: View of federated stormwater pipe assets in Christchurch

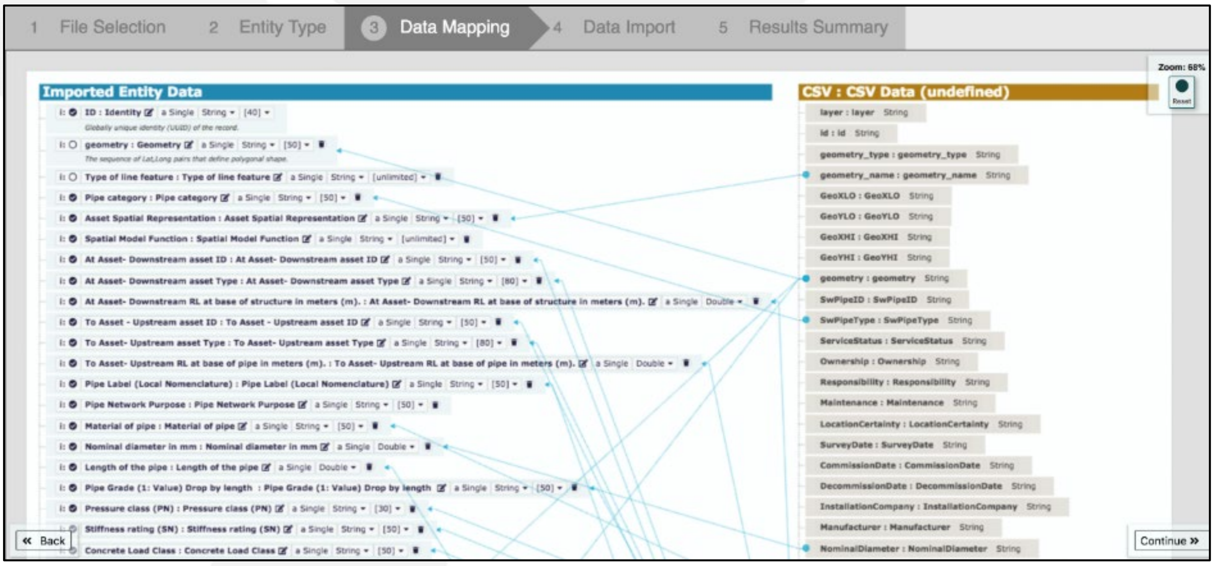


Figure 5: Example of schema mapping process between source data and the national beta-standard

Key findings from the project

This project was progressed with the specific intent that it act as a proof-of-concept trial which could support a scalable methodology and technical solution for providing coverage at a national level. The following is a summary of the findings which provide insight for scaling the approach applied in this project to a greater number of asset owners and a wider geographical coverage. Each summary point is annotated with the primary finding from the proof-of-concept trial.

Enterprise approach

Data federation across multiple asset owners was achieved using an online application. This type of online multi-tenanted environment requires a particular technical approach which is not necessary in non-federated environments. For example, it needs login and password access, assigned user roles, relationships defined between users and their organisations, password reset, user guides and software documentation. The requirement for enterprise-level software to provide these features means a minimum level of software-development is required before a wider stakeholder and asset owner community can be supported. This is one of the strengths of using *Bruce* as a platform.

Security

It is anticipated that asset owners will require confidence that their data can only be accessed by approved and vetted parties. Therefore, security and data access are key considerations of further development and expansion. This is an important factor when selecting the final software platform.

Data service feeds

The data used in this proof-of-concept project were provided as a static data file. For a scalable solution, it is anticipated that data would be provided by asset owners as a dynamic data service, thereby ensuring that federation and analysis are always undertaken on the up-to-date data 'at-source'. This means that when an asset owner updates their data, it is available immediately to the federated system.

Data quality assessment

A standard mechanism of reporting data quality would provide significant value to both asset owners and other industry and governmental stakeholders. The project team had visibility of the data quality framework developed by the NZ Transport Association/Local Government NZ-led Road Efficiency Group (REG) for the provision of roading asset data. This is considered a good starting point (template) for next-stage development. In addition to this, a standardised method of reporting against data alignment with the national beta-standard is also required.

Standards

Federation of asset data is achieved through a data-mapping process which, by definition, requires the choice of a standardised data schema. To achieve this in the project, the beta-version 3 waters metadata standard was adopted. Next-stage development will require early agreement on the standards to adopt for the data-mapping and federation process.

Dynamic standards

While data standards are necessary for the data federation process, it is possible to maintain a flexible data schema and therefore flexible standards. A dynamic standard is an important concept as it means that stakeholders can proceed with confidence knowing that any subsequent changes to standards will not undermine the federation process and the associated and supporting processes, analysis and models.

Data mapping and correspondence registers

The proof-of-concept project has revealed that a key component of the data-mapping process is the knowledge base which increases every time asset data is mapped to the standard. There is a requirement to maintain a register of the correspondences between the naming convention of asset classes, attributes and values. As an example, an attribute titled ‘Construction’ may have a correspondence with the attribute in the beta-standard titled ‘Material’. Similarly, an attribute value of ‘ACBM’ may correspond with a value of ‘AC’ within the beta-standard. By building a registry and/or dictionary of these correspondences, the federation process and system will become an important source for adding un-validated datasets by employing modern data science approaches.

Mapping access and export

The federation process creates a data-mapping process (also referred to as a mapping-schema or data-mapping model) which is unique for each dataset. This schema/model describes how each class, attribute and value in the source data corresponds to those within the beta-standard. A key requirement for any future development is that these schemas/models are exportable in machine readable format. This will enable these essential system components to be backed-up and modified without reliance on third party or proprietary software.

Data services (advice)

The three councils for this project (Auckland, Tauranga and Christchurch) have a level of internal capability which means that their data is well-defined and structured, including their metadata. For asset owners where this level of capability is not available, it is considered that data services and assistance will be necessary to source, define and prepare their data for the federation process.

Service monitoring (service):

A federated system which is used by a wide range of stakeholders, and which supports a wide range of ancillary services and analytics, requires a level of confidence and assurance that it remains available and current. As the system grows it is likely that a monitoring service will be required to ensure that the components of the system remain active and maintained enabling a timely response should an outage or disruption occur.

Visualisation

Visualisation of federated data provides a means to validate spatial datasets. However, it is unclear to what extent further meaningful analysis will be undertaken using map-based visualisation. Notwithstanding this, visualisation is likely to be an essential aspect for future developments such as asset location and works coordination.

Data analysis

A primary benefit of achieving federation of asset data is that standardised analytics and modelling can be readily applied across the entire dataset. A model, or analysis suite, commissioned for one region or by one asset owner can be readily applied to the data sourced in other regions and from other asset owners. This could provide a standardised and universal approach for asset valuations, criticality assessments and predictive asset condition assessments. While this proof-of-concept project has succeeded in meeting the success criteria agreed at the outset, there is further effort required for the next stage of development and the capture of the benefits afforded by asset data federation through dynamically mapping to agreed national standards.

Appendix A: Combined student report for proof-of-concept project titled ‘Gap analysis and mapping pipe asset data to a national metadata standard’

Note: This report was a component of the MADS students’ final assessment and has not been edited as part of this report.

Manveer (57771679) Jing (29696576) Ziqi (55088375)

Gap analysis and mapping pipe asset data to a national metadata standard

Contributors

This report was prepared by the University of Canterbury, at the request of the UC Quake Centre, and with data provided by Auckland Council, Western Bay of Plenty District Council and Christchurch City Council.

Please contact the authors for further information regarding this report or the supplementary material.

Name	**Manveer Kaur Bhullar**
Council	**Auckland Council**
Phone	**+64 27 588 8974**
Email	**Manveer.bhullar93@gmail.com**

Name	**Ziqi Gu**
Council	**Tauranga City Council**
Phone	**+64 0223468407**
Email	**aisling.gu@gmail.com**

Name	**Jing Wu**
Council	**Christchurch City Council**
Phone	**+64 21 149 6394**
Email	**jing.5@hotmail.com**

Contents

1. Executive Summary

1.a. Background

Each council in New Zealand essentially manages a network of three water systems, viz. Stormwater, Wastewater and Potable water. The asset management systems at each council are not nationally coordinated and there is a great disparity in data pertaining to assets held by each council in relation to the data held by other councils. Due to the age of the infrastructure and legacy issue relating to data providence, these asset data could be incomplete, inaccurate. This is a problem given the need to manage this ageing infrastructure.

The aim of New Zealand Metadata Standards for 3-water assets is to provide a consistent set of requirements for all digital data relating to assets and to define a common language to facilitate the sharing of asset information throughout New Zealand. The Beta Version National Asset Metadata Standard for 3 Waters is an evolution of the draft standard developed by LINZ. The main work on this evolution was done by Graham Clark, formerly of CCC and now at Fulton Hogan. The work done in this project will inform the next evolution of the standard which will be published as the NZ National Asset Metadata Standard for 3 Waters Version 1.0. The Beta Version National Asset Metadata Standard is split into the ‘Simple Reticulation’ document and the ‘All water assets’ spreadsheet.

The main objective of this project is to compare the metadata standards used in the three City Councils namely - Christchurch City Council, Auckland Council, Tauranga City Council. We were required to compare these Councils’ asset management Information System (SAP, Accela & GIS) against the Beta Version National Asset Metadata Standard for 3 Waters (NZ Beta Standard) and Land Information NZ (LINZ) Metadata Standards (for Christchurch City Council). This involved working through the NZ Beta Standard document for Pipe reticulation (FH70) and Access Chamber (FH36) for primarily the Stormwater network system of each council and comparing it to SAP and GIS metadata for each asset type managed by each Council. The alignment will need to be quantified in terms of SAP Object Types (Asset Types) and SAP Characteristics (attribute information including code lists).

Furthermore, the picklists as well as the valid value codes for attributes vary amongst different councils as well as in context to the Beta version of national data standard. This report will cover the assessment of alignment of the pick list and code lists to the NZ standard. Additionally, we will report on the population statistic of each asset layer for the geodatabases provided by each council to provide them a better understanding of their present data.

A further aim of the project was to federate the data from the three councils into a single view based on the NZ Beta Standard. This federated data should be capable of being:

1. Visualised
2. Analysed by running basic queries on the data

This part of the project was to be run in a software package called *Bruce* which is a data integrator and visualisation tool currently in development in NZ by Nextspace.
The purpose of this was to create a proof-of-concept for a National Digital Infrastructure Model (NDIM) which will federate and visualize all the horizontal infrastructure data in one.

All results and material used for this report are provided in the attached supplementary material which is a zipped file folder provided with the report.

1.b. Scope

1.b.1. Common Scope for Quake Centre

The common objective of this report was to work with Quake Centre to achieve the following common outcomes for the three territorial authorities, namely - the Auckland Council, Tauranga City Council and Christchurch City Council.

1. Identify gaps and missing values in data and check for completeness pertaining to NZ Beta Standard data standard and help council to understand the quality of their data.
2. Dynamically map council's data to a common standard for each attribute and values for the attribute for each asset class.
3. Federate data for each council for sharing or querying purposes.
4. Simplify the data transfer process between each council.
5. Visualize data using a prototype national infrastructure portal (Bruce).

1.b.2. Auckland Council Scope

The overall objective of this report is to assess the alignment, and to compare the data structure of Healthy Waters Data Standard for the asset classes – Pipe and Access Chambers to the NZ Beta Standard's Simple Reticulation document and to the 'Picklist Master" worksheet. Also, to achieve the same for dataset provided by Auckland Council. The NZ Beta Standard documents can be found in supplementary material [SA_1]. The healthy waters and 'AC Healthy Water Stormwater Network Asset layers.gdb' file is given in [SA_2].

The scope of the project for the Auckland Council is comprised of the following requirements:
1. To review the Healthy Waters (HW) stormwater data standard structure, classes and attribution.
2. To identify the asset classes, data fields and attributes in the HW standard that are relevant to the Beta national data standard for the assets - Pipe and Access Chambers.
3. To run a gap analysis to report on attribute value population statistics.
4. To produce the final mapping file for the AC's stormwater data standard (Healthy Waters) and NZ Beta Standard.
5. Final report on findings, outcome and recommendations.

The analysis in this report was primarily conducted using the attributes contained in Healthy Waters data standard and the gap analysis for attribute value population statistics was done on the extract of the Auckland Council's geospatial database only, and there may be other asset management databases that contain more asset metadata than what was available for Auckland Council authority.

1.b.3. Tauranga Council Scope

The main objective is to compare the asset metadata in the Tauranga District Council's Accela System synced with GIS against the NZ Beta Standard which involves mapping each attribute and values to a uniform national standard, identify gaps and missing values to check the completeness pertaining to standard and help the council to understand the quality of their data.

The main reticulations were firstly manually mapped to the standard in section 2.c.2 and analysed in section 3.b.1 include pipes of stormwater, wastewater and potable water and manholes of stormwater and wastewater. Fitting were excluded as they were not present in the data. Meanwhile, the visualization of main reticulations was achieved in Bruce and displayed in separate layers in section 3.b.2.

The automatic process for mapping was created by python code with several different distance calculation algorithms in section 3.b.1.iv) as discussed in section 3.b.2.

TCC's data was uploaded into Bruce and mapped with a shortened version of pipes of NZ Beta Standard schema then successfully federated with Christchurch council's data and Auckland council's data and achieved query within three councils as a whole dataset and visualization in section 3.d.

1.b.4. Christchurch City Council Scope

The objective of this report is the gap analysis between the current Christchurch City Council metadata standards and NZ Beta Standards (published as a draft standard by LINZ) of pipe reticulation assets for 3-waters.

The aims of Christchurch City Council were to:

• Evaluate the alignment the Christchurch City Council metadata standard and the NZ Beta Standard, focusing on pipe and access chamber attributes, picklists and picklist values.

• Evaluate the alignment the Christchurch City Council metadata standard and the LINZ metadata standard, focusing on pipe reticulation (pipe, access chamber, fitting and valve) attributes, codelists and codelist values. The section of LINZ is in the appendix.

The analysis in this report is based on the attributes and attribute values contained in the 3-waters network geodatabases of the Christchurch City Council. There may be other asset management systems that contain more attributes and attribute values that are not included in the project.

1.c. NZ Beta Standard attributes to be aligned by each territorial authority

The two asset classes focussed for the purpose of this report are - Pipe reticulation and Access Chambers and the following attributes were selected from the NZ Beta Standard 'Simple Reticulation' document. The related attributes for the two asset classes are as follows:

1.c.1. NZ Beta Standard Core Pipe attributes

Based on the analysis, we used the following NZ Beta Standard core attributes of pipe for each council:

- Pipe category
- Asset Spatial Representation
- Spatial Model Function
- At Asset- Downstream asset ID
- At Asset- Downstream asset Type
- At Asset- Downstream RL at base of structure in meters (m).
- To Asset - Upstream asset ID
- To Asset- Upstream asset Type
- To Asset- Upstream RL at base of pipe in meters (m).
- Pipe Label (Local Nomenclature)
- Pipe Network Purpose
- Material of pipe
- Nominal diameter in mm
- Length of the pipe
- Pipe Grade (1: Value) Drop by length
- Pressure class (PN)
- Stiffness rating (SN)
- Concrete Load Class
- Pipe Shape
- Average burial depth to invert of pipe
- Pipe external wrapping or sleeving
- Pipe Internal Treatment - Rehabilitation
- Internal Diameter Post Treatment
- Pipe Treatment Installation Date
- What is the pipe installation method

1.c.2. NZ Beta-Standard Core Access Chambers Attributes

Based on the analysis, we used the following NZ Beta Standard core attributes of access chamber for each council:

- Specific type of Access Chamber
- Asset Spatial Representation
- Spatial Model Function
- RL at access of structure in meters (m).
- RL at base of structure in meters (m).
- Structure Foundation Type
- Structure Treatment Type
- Chamber Construction Material
- Lids fitted - YES/No
- Lid Material

- Step irons fitted - YES/no
- Ladder fitted - YES/no
- External load capacity - tonnes
- Chamber Width (m)
- Chamber Length (m)
- Chamber Height (m)
- Chamber Diameter - (m)
- Type of security on access
- Chamber max working height -m
- Chamber Lid Shape
- Chamber Lid Style
- Network Flushing Point

Each territorial authority could use this report as a starting point to further investigate how their pipe asset data aligns with the NZ Beta Standard for wastewater, storm water, and potable water, developing the mapping from their asset attributes to the NZ Beta Standard attributes and improving their data to meet the asset attribute requirements. Also, this report could help to improve the NZ Beta Standard which could manage the 3-waters assets data at a national wide level. This should include documenting graphical data accuracy, and enforcing asset attribute requirements regarding attribute names, data type, data formats, code lists, and default values as described in Page 21 and Page 23 of the NZ Beta Standard simple reticulation document and further in the NZ Beta Standard 'Picklist Master Asset' worksheet by column H (NZ Beta Standard Master Code).

2. Analysis

2.a. Data Provided by the three Councils

2.a.1. Auckland Council Datasets

Auckland Council uses SAP as its Relational Database Management System and the spatial data is collected using ArcGIS and ArcMap software. The description for the two focussed asset classes as given by the online public *GeoMaps* portal of Auckland Council (powered by esri) is as follows:

For pipe asset class is stated as - "Pipelines form part of a reticulated stormwater network that includes pipelines, culverts and subsoil drains to drain stormwater runoff from roads, property and open areas to receiving environments. Purpose - Stormwater asset management Lineage - through the OnePlus project, all seven legacy systems were migrated to the new GIS-SAP environment where the creation of new assets and maintenance of existing assets are now being undertaken. Using asbuilts sent to the stormwater team from development engineers and/or internal projects, the geometry of stormwater assets is captured using standard ArcGIS editing functionality (updating the SAPGEO feature class) and its attributes are populated within SAP (updating the SAPGEO table). Whilst due care is taken to capture the assets as accurately as possible, the data is indicative and cannot be considered to align to any particular boundaries or features including cadastral."

For Manhole and Chamber asset class - “Manholes are an underground utility vault that can be physically accessed and used for making connections or performing maintenance on underground stormwater pipes. Manholes are provided at every change of direction, at every main junction, at every change of gradient and at distances approved by council. Chambers are intended to provide simple access for cursory inspection and access for drain rods or other maintenance equipment but no physical access. Purpose - Stormwater asset managementLineage - Through the OnePlus project, all seven legacy systems were migrated to the new GIS-SAP environment where the creation of new assets and maintenance of existing assets are now being undertaken. Using asbuilts sent to the stormwater team from development engineers and/or internal projects, the geometry of stormwater assets are captured using standard ArcGIS editing functionality (updating the SAPGEO feature class) and its attributes are populated within SAP (updating the SAPGEO table). Whilst due care has been taken to capture the assets as accurately as possible, the data is indicative and cannot be considered to align to any particular boundaries or features including cadastral.”

The link to above portal can be found in the reference section [4].

The data provided by Auckland Council includes Healthy waters metadata standard spreadsheet and geodatabase file (‘AC Healthy Water Stormwater Network Asset layers.gdb’). The Healthy Waters metadata standard spreadsheet consisted of six asset layers only for the stormwater network, namely:

1. Manholes
2. Pipes
3. Catchpits
4. CP Leads and Connections
5. Inlet Outlets
6. Soakholes

Additionally, the sample geodatabase (.gdb) file was provided for the stormwater network type. This gdb file consisted of the following 11 asset layers:

1. Catchpit
2. Channel
3. Connection
4. Erosion_And_Flood_Control
5. Inlet_And_Outlet
6. Manhole_And_Chamber
7. Pipe
8. Pump_Station
9. Soakage_System
10. Water_Treatment_Device
11. Watercourse

Each asset layer has its own characteristic attributes based on its properties. The spatial coordinates and the geometry type of the asset (point, polygon or multiline) are hidden and hence had to be extracted using external softwares.

2.a.2. Tauranga Council Datasets

The geospatial data provided by Tauranga City Council was an ArcGIS geodatabase (gdb) for each network with separate layers for each asset type extracted from Accela system. Due to the variation and ongoing improvement of TCC's three water's asset data structure, all the analyses and results are only based on the extracted data and may not consistent with the current council's data structure.

2.a.3. Christchurch City Council Datasets

Christchurch City Council provided the council metadata standard and geospatial data for stormwater, wastewater and water supply.

The geospatial data provided by Christchurch City Council was grouped by network and formatted as an ArcGIS geodatabase (gdb). Each network (stormwater, wastewater, and water supply) was divided into multiple layers by asset classes, with a separate layer for each specific asset class. Each layer contained the spatial data of the asset and its own attributes and attributes values. The geodatabases are provided in the supplementary material [SC_1].

The metadata standard provided by Christchurch City Council was formatted as a spreadsheet (xlsx). The council metadata standard is provided in the supplementary material [SC_2]. The University of Canterbury provided the LINZ metadata standard others compiled which was formatted as a spreadsheet. The LINZ metadata standard is provided in the supplementary material [SC_3].

2.b. Data Extraction and Statistical Analysis

The data extracted from each of the councils (.gdb extensions) was loaded into asset layers and the related geometry for each class was loaded. The structure is identified in Table 2.b-1.

Table 2.b-1. Layer columns and descriptions

Columns	Description
Layer Name	Layer Name
Asset ID	Index starting at 1
Geometry Type	Geometry type, which is defined by numbers
Geometry Name	Geometry name
GeoXLO	Downstream Longitude
GeoYLO	Downstream Latitude
GeoXHI	Upstream Longitude
GeoYHI	Upstream Latitude
Attributes extracted	Feature attributes

The attribute values themselves were also checked for consistency. We counted how many values were specified, what data types were used, and a range of other statistics specific to each data type. The statistics analysis is identified in Table 2.b-2.

Table 2.b-2. Statistics computed for each feature level attribute in the geospatial data provided.

Columns	Data Type(s)	Description
Count	All	Count of features in layer (assets in asset class)
Not Empty	All	Count of attribute values that are not empty
Empty	All	Count of attribute values that are empty
Unique	All	Count of unique attribute values (across all features)
Data Type	All	Data type (based on all attribute values)
Negative	Integer, Decimal	Count of negative attribute values
Zeros	Integer, Decimal	Count of attribute values that are zero but not empty
Positive	Integer, Decimal	Count of positive attribute values
Min	Integer, Decimal	Minimum attribute value
Max	Integer, Decimal	Maximum attribute value
Min Length	Alpha / Numeric	Minimum character length (excluding attribute values that are empty)
Max Length	Alpha / Numeric	Maximum character length (excluding attribute values that are empty)
Commas	Alpha / Numeric	Count of attribute values that contain commas
Date String	Date	Example date string

2.b. Inter-council assessment of consistency

2.b.1. Asset Class Consistency

The mapped result for each council's asset classes mapped to the NZ Beta Standard data standard is in Table 2.b-1. The Christchurch City Council had the layer 'Junction' for the three network systems, which was missing in Tauranga and Auckland Council data. Additionally, the Auckland Council only dealt with storing Stormwater network data.

Table 2.b-1. Asset layers mapping of three water networks for each council.

NZ Beta Standard Asset Code	NZ Beta Standard Asset Class	CCC Asset	TCC Asset	ACC Asset
FH70	Pipe Reticulation	SwPipe	SWEPipe	Pipes
		WwPipe	WWEPipe	NA
		WsPipe	WSEPipe	NA
FH36	Access Chamber	SwAccess	SWEManhloe	Manholes
		WwAccess	WWEManhole	NA
FH71	Junction	SwFitting	NA	NA
		WwFitting	NA	NA
		WsFitting	NA	NA

2.b.2. Attributes Consistency

2.b.2.a Stormwater Pipe Reticulation - Attributes Consistency

We achieved the attribute mapping for Pipe asset layer individually for each council. The shortened version of mapped results was then compared other councils for naming consistency and the result is in Table 2.b-2.

Table 2b-2. Stormwater NZ Beta Standard **Pipe attributes mapping to each council.**

NZ Beta Standard	CCC	TC	AC
Pipe category	SwPipeType		SAPGEO_MASTER_swPipeTable_SW_ASSET_TYPE
Asset Spatial Representation			
Spatial Model Function			
At Asset- Downstream asset ID	DownstreamFeatureID	ToID	
At Asset- Downstream asset Type	DownstreamFeature	ToType	
At Asset- Downstream RL at base of structure in meters (m).	DownstreamInvert		SAPGEO_MASTER_swPipeTable_SW_DEPTH_DOWNSTREAM_M
To Asset - Upstream asset ID	UpstreamFeatureID	FromID	
To Asset- Upstream asset Type	UpstreamFeature	FromType	
To Asset- Upstream RL at base of pipe in meters (m).	UpstreamInvert		
Pipe Label (Local Nomenclature)	PipeName		
Pipe Network Purpose			
Material of pipe	Construction	Material	SAPGEO_MASTER_swPipeTable_SW_MATERIAL
Nominal diameter in mm	NominalDiameter	Diameter	SAPGEO_MASTER_swPipeTable_SW_DIAMETER_MM
Length of the pipe	drvLength	Length	SAPGEO_MASTER_swPipeTable_SW_LENGTH_GIS_M, Shape_length
Pipe Grade (1: Value) Drop by length	Grade		
Pressure class (PN)	PressureClass		
Stiffness rating (SN)	StiffnessClass		
Concrete Load Class	LoadClass		

Pipe Shape	Shape	MainShape	
Average burial depth to invert of pipe	Depth		SAPGEO_MASTER_swPipeTable_SW_INVERT_LEVEL_DOWNSTREAM_M
Pipe external wrapping or sleeving			
Pipe Internal Treatment - Rehabilitation	Treatment		SAPGEO_MASTER_swPipeTable_SW_MATERIAL_LINING
Internal Diameter Post Treatment	TreatmentDiameter		
Pipe Treatment Installation Date	TreatmentDate		SAPGEO_MASTER_swPipeTable_SW_LINING_DATE
What is the pipe installation method	InstallationMethod		SAPGEO_MASTER_swPipeTable_SW_INSTALLATION_TYPE

2.b.2.b Stormwater Access Chambers - Attributes Consistency

We achieved the attributes mapping for Access Chambers asset layer individually for each council. The mapped results were then compared other councils for naming consistency and the result is in Table 2.b-3.

Table 2.b-3 Stormwater NZ Beta Standard Access chamber attributes mapping to each council.

NZ Beta Standard	CCC	TCC	AC
Specific type of Access Chamber	SwAccessType		*SW_ASSET_TYPE
Asset Spatial Representation			
Spatial Model Function			
RL at access of structure in meters (m).	LidLevel	LidLevel	*SW_COVER_LEVEL_M
RL at base of structure in meters (m).	BaseLevel	InvertLevel	*SW_INVERT_LEVEL_M
Structure Foundation Type			
Structure Treatment Type	SwAccessTreatmentType		
Chamber Construction Material	SwAccessConstruction		*SW_MATERIAL_LINING
Lids fitted - YES/No			*SW_HAS_STEPS_OR_LADDER
Lid Material			*SW_MATERIAL_COVER
Step irons fitted - YES/no			
Ladder fitted - YES/no			
External load capacity - tonnes			
Chamber Width (m)	PitWidth		
Chamber Length (m)	PitLength		
Chamber Height (m)	PitDepth	Depth	*SW_DEPTH_TO_INVERT_M
Chamber Diameter - (m)		Diameter	
Type of security on access	SwAccessSecurity		*SW_LOCKING_MECHANISM
Chamber max working height -m			
Chamber Lid Shape	SwAccessLidShape		
Chamber Lid Style	SwAccessLidStyle		*SW_COVER_TYPE
Network Flushing Point			

***all the attribute names in Auckland Council data for Access Chamber initiate with SAPGEO_MASTER_swManholeAndChamberTable_*, and for readability, the names mentioned in the above table 2.b-3 are the suffix of this.*

2.b.3. Pipe material picklists Consistency

The mapped results for each council's pipe material picklists mapped to the NZ Beta Standard data standard are in Table 2.b-4.

Table 2.b-4. Stormwater NZ Beta Standard pipe material picklists mapping to each council.

Beta standard Pipe Material	Description	Healthy Waters	TCC	CCC
ABS	Acrylonitrile Butadiene Styr			ABS
AC	Asbestos Cement	ABCM	AC	AC
ALK	Alkathene	ALKA		AL
API-STEEL	Amer. Petroleum Inst. Steel Pipe			API
BRICKB	Brick Barrel	BRCK		BB
BB	Black Brute			BLBRUTE
B-STEEL	Bitumen Lined Steel			
CI	Cast Iron	CAIR	CI	CI
CIPP	Cured In Place Plastic			
CL-DI	Concrete Lined Ductile Iron			CLDI
Cor-PE	Corrugated Polyethelene			CORRPE
DI	Ductile Iron	DUIR		DI
EARTH	Earthenware	ERWR		EW
GALV-STEEL	Galvanised Steel	GAIR		GALV
Glass Reinforced Polymer(code not exists)	Glass Reinforced Polymer			GRP
HDPE	High Density Polyethelene		HDPE	HDPE
LDPE	Low Density Polyethylene		LDPE	LDPE
MDPE100	Medium Density Polyethylene 100		MDPE?	
MDPE80	Medium Density Polyethylene 80		MDPE?	MDPE80
MLDI	Mortar Lined Ductile Iron			MLDI
M-PVC	Modified Polyvinyl Chloride		mPVC	
NOVA	NovaFlow		Novaflo?	NovaFlow
O-PVC	Oriented Polyvinyl Chloride			OPVC
PE	Polyethelene	PYTH	PE80?	PE
PE100	Polyethelene 100		PE100	PE100
PVC	Polyvinyl Chloride	PYVN		PVC

RCRR	Reinf. Concrete Rubber Ringed			RCRR
Rib-PVC	Ribloc- Ribbed PVC Pipe Liner			
SW-ALUM	Spirally Wound Aluminium			SPWAL
STEEL	Steel	STEL	Steel	STEEL
U-PVC	Unplasticised Polyvinyl Chloride		uPVC	UPVC
NOTCAP	Not Capture		Unknown	Unknown
CU	Copper	COPR		
TIMB	Timber		Timber	Timber

2.c. Assessment of alignment to NZ Beta Standard - Simple Reticulation

2.c.1. Auckland Council

2.c.1. (a) - Alignment of HW Attributes:

The alignment of the asset classes in Healthy Waters Data standard named as Pipes, and Manhole and Chamber to the NZ Beta Standard asset classes Pipe Reticulation (FH70) and Access Chambers (FH36) was completed following partially automated techniques.

Attribute grouping was done beforehand to avoid mapping the same attributes more than once for each asset class, and the results were provided in the same spreadsheet as the attribute grouping. The mapping identified is presented in tables 2.c.(i) and 2.c.(ii) respectively. These are the results of mapping the attributes of the two assets. These tables also depict which attributes could not be matched to the HW data standard.

These results are provided in the supplementary material [SA_3].

Table 2.c.(i). NZ Beta Standard Core Pipe attributes mapped to Healthy Waters standard Pipe attributes.

NZ Beta Standard Standard (Pipe Reticulation)	Healthy Waters Standard (Pipes)
Pipe category	Pipe Type
Asset Spatial Representation	
Spatial Model Function	
At Asset- Downstream asset ID	
At Asset- Downstream asset Type	
At Asset- Downstream RL at base of structure in meters (m).	Depth - Downstream (m)
To Asset - Upstream asset ID	
To Asset- Upstream asset Type	
To Asset- Upstream RL at base of pipe in meters (m).	Depth - Upstream (m)
Pipe Label (Local Nomenclature)	
Pipe Network Purpose	
Material of pipe	Pipe Material
Nominal diameter in mm	Diameter (mm)
Length of the pipe	
Pipe Grade (1: Value) Drop by length	
Pressure class (PN)	
Stiffness rating (SN)	
Concrete Load Class	
Pipe Shape	Shape of Pipe
Average burial depth to invert of pipe	Invert Level - Downstream (m)
Pipe external wrapping or sleeving	
Pipe Internal Treatment - Rehabilitation	Lining Material
Internal Diameter Post Treatment	After Lining Diameter (m)
Pipe Treatment Installation Date	Lining Date
What is the pipe installation method	Installation Type

Table 2.c.(ii). NZ Beta Standard Core Access Chamber attributes mapped to Healthy Waters standard Manhole and Chamber attributes.

NZ Beta Standard (Access Chamber)	Healthy Waters Standard (Manholes)
Specific type of Access Chamber	Manhole type
Asset Spatial Representation	
Spatial Model Function	
RL at access of structure in meters (m).	Lid/Cover Level (m)
RL at base of structure in meters (m).	Invert Level
Structure Foundation Type	
Structure Treatment Type	
Chamber Construction Material	Manhole Material
Lids fitted - YES/No	
Lid Material	MH Lid/Cover Material
Step irons fitted - YES/no	
Ladder fitted - YES/no	
External load capacity - tonnes	
Chamber Width (m)	Manhole Width (mm)
Chamber Length (m)	Manhole Length (mm)
Chamber Height (m)	Depth to Invert (m)
Chamber Diameter - (m)	Manhole Diameter (mm)
Type of security on access	MH Lid/Cover Locking Mechanism Type
Chamber max working height -m	
Chamber Lid Shape	
Chamber Lid Style	MH Lid/Cover Type
Network Flushing Point	

2.c.1. (b) - Alignment of HW Picklist and Code list Values:

Additionally, the assessment of alignment of HW attribute code list to beta standard code list in tables 2.c.(iii) and 2.c.(iv). The spreadsheet of these results is provided in the supplementary material [SA_4].

Table 2.c.(iii). NZ Beta Standard Core Pipe Reticulation Code list mapped to Healthy Waters standard Pipe Code list with Picklist and Description.

NZ Beta Standard Attributes	MAPPED HW ATTRIBUTE	NZ BETA STANDARD VALUE PICKLIST	NZ BETA VALID VALUE DESCRIPTION	NZ BETA VALID VALUE CODE	HW MAPPED VALID VALUE CODE	HW VALID VALUE DESCRIPTION
Pipe Category	Pipe Type	MAX_PIPECLASS	PIPE / GRAVITY		GRVY	Gravity Main
			PIPE / SIPHON		SYPN	Syphon
Material of Pipe	Pipe Material	MAX_MATERIALPIPE	Asbestos Cement	AC	ABCM	Asbestos Cement
			Alkathene	ALK	ALKA	Alkathene
			Brick Barrel	BRICKB	BRCK	Brick
			Cast Iron	CI	CAIR	Cast Iron
			Ductile Iron	DI	DUIR	Ductile Iron
			Earthenware	EARTH	ERWR	Ceramic / Earthenware
			Galvanised Steel	GALV-STEEL	GAIR	Galvanised Iron or Steel
			Polyethelene	PE	PYTH	Polyethylene
			Polyvinyl Chloride	PVC	PYVN	Polyvinyl Chloride
			Steel	STEEL	STEL	Steel
Pipe Shape	Shape of Pipe	MAX_PIPESHAPE	Circle		CIRC	Circular
			Egg		EGG	Egg
			Rectangular		RECT	Rectangular
Pipe Internal Treatment - Rehabilitation	Lining Material	MAX_PIPETREATMENT	CIPP	CIPP	CIPP	Cured In Place Polymer
			Epoxy Lining	EPOXYLINE	EPOX	Epoxy
			Expandable PVC	EXPANDABLELINE	FPVC	Fold and Form PVC
			Polymeric Lining	POLYMERICLINE	PYMR	Polymer
			Spiral Wound Lining	SPIRALWOUNDLINE	RIBL	Ribloc Spiral PVC
What is the pipe installation method	InstallationMethod	MAX_PIPEINSTALL	Bored		BORE	Bore
			PipeBurst		BURT	Pipe Bursting

Table 2.c.(iv). NZ Beta Standard Core Pipe Reticulation Code list mapped to Healthy Waters standard Pipe Code list with Picklist and Description.

NZ BETA STANDARD ATTRIBUTES	MAPPED HW ATTRIBUTE	NZ BETA STANDARD VALUE PICKLIST	NZ BETA VALID VALUE DESCRIPTION	NZ BETA VALID VALUE CODE	HW MAPPED VALID VALUE CODE	HW VALID VALUE DESCRIPTION
Specific type of Access Chamber	Manhole type	MAX_CHAMBERCLASS	Inspection Chamber	INSPECT	INSP	Inspection Chamber
			Manhole Chamber	MANHOLE	STND	Standard Manhole
Chamber Construction Material	Manhole Material	MAX_MATERIALCIVIL				
Lid Material	MH Lid/Cover Material	MAX_MATERIALCIVIL	Steel		STEL	Steel
					GAIR	Galvanised Iron or Steel
Type of security on access	MH Lid/Cover Locking Mechanism Type	MAX_SECURITY	Unsecured		NONE	None
			Bolted		STRP	Strapped
					OTHR	Other
Chamber Lid Style	MH Lid/Cover Type	MAX_LIDSTYLE	Standard grated		GRAT	Grated
			Wavy grated			

Furthermore, gap analysis for the population statistics of each asset layer present in Auckland Council's dataset was done and the results are provided in supplementary material [SA_6]. The values extracted for population statistics identifies with the Table 2.b-2.

2.c.1. (a) - Alignment for Quake Centre Outcomes:

Since the feature attributes nomenclature differ vastly in the Auckland Council's data as compared to Healthy Waters, the assessment of alignment of Pipe, and Manhole and Chamber classes to NZ Beta Standard was done separately following the same methodology. The results of AC data mapping to NZ standard mapping are as given in tables – 2.c.1, 2.c.(v), and 2.c.(vi). The final spreadsheets are also provided in the supplementary material [SA_5].

Table.2.c-(v). Mapped Core Pipe Attributes against Auckland Council Data.

NZ Beta -Standard	**AC**
Pipe category	SAPGEO_MASTER_swPipeTable_SW_ASSET_TYPE
Asset Spatial Representation	geometry_name
Spatial Model Function	
At Asset- Downstream asset ID	
At Asset- Downstream asset Type	
At Asset- Downstream RL at base of structure in meters (m).	SAPGEO_MASTER_swPipeTable_SW_DEPTH_DOWNSTREAM_M
To Asset - Upstream asset ID	
To Asset- Upstream asset Type	
To Asset- Upstream RL at base of pipe in meters (m).	
Pipe Label (Local Nomenclature)	
Pipe Network Purpose	
Material of pipe	SAPGEO_MASTER_swPipeTable_SW_MATERIAL
Nominal diameter in mm	SAPGEO_MASTER_swPipeTable_SW_DIAMETER_MM
Length of the pipe	SAPGEO_MASTER_swPipeTable_SW_LENGTH_GIS_M, Shape_length
Pipe Grade (1: Value) Drop by length	
Pressure class (PN)	
Stiffness rating (SN)	
Concrete Load Class	
Pipe Shape	
Average burial depth to invert of pipe	SAPGEO_MASTER_swPipeTable_SW_INVERT_LEVEL_DOWNSTREAM_M
Pipe external wrapping or sleeving	
Pipe Internal Treatment - Rehabilitation	SAPGEO_MASTER_swPipeTable_SW_MATERIAL_LINING
Internal Diameter Post Treatment	
Pipe Treatment Installation Date	SAPGEO_MASTER_swPipeTable_SW_LINING_DATE
What is the pipe installation method	SAPGEO_MASTER_swPipeTable_SW_INSTALLATION_TYPE

Table.2.c-(vi). Mapped Core Access Chamber Attributes against AC Data.

NZ Beta Standard	AC
Specific type of Access Chamber	SAPGEO_MASTER_swManholeAndChamberTable _SW_ASSET_TYPE
Asset Spatial Representation	
Spatial Model Function	
RL at access of structure in meters (m).	SAPGEO_MASTER_swManholeAndChamberTable _SW_COVER_LEVEL_M
RL at base of structure in meters (m).	SAPGEO_MASTER_swManholeAndChamberTable _SW_INVERT_LEVEL_M
Structure Foundation Type	
Structure Treatment Type	
Chamber Construction Material	SAPGEO_MASTER_swManholeAndChamberTable _SW_MATERIAL_LINING
Lids fitted - YES/No	SAPGEO_MASTER_swManholeAndChamberTable _SW_HAS_STEPS_OR_LADDER
Lid Material	SAPGEO_MASTER_swManholeAndChamberTable _SW_MATERIAL_COVER
Step irons fitted - YES/no	
Ladder fitted - YES/no	
External load capacity - tonnes	
Chamber Width (m)	
Chamber Length (m)	
Chamber Height (m)	SAPGEO_MASTER_swManholeAndChamberTable _SW_DEPTH_TO_INVERT_M
Chamber Diameter - (m)	
Type of security on access	SAPGEO_MASTER_swManholeAndChamberTable _SW_LOCKING_MECHANISM
Chamber max working height -m	
Chamber Lid Shape	
Chamber Lid Style	SAPGEO_MASTER_swManholeAndChamberTable _SW_COVER_TYPE
Network Flushing Point	

Due to the absence of a more current pick list for the valid values of attributes in the AC dataset, there is a probability that the current AC data supports more valid values for attributes than mapped results derived from the HW standard.

2.c.2. Tauranga City Council

To assess the alignment of TCC’s data and NZ Beta Standard, the shortened results of mapping main reticulation was showed below. The metadata Standard was a shortened version of NZ Beta Standard. The full version of results can be found in supplementary material “Mapping results”.

2.c.2.1 Assessment of alignment to NZ pipe -Standard

Core attributes mapping for pipes of three water.

Table 2.c.i) Core pipe attribute mapping (Shortened version)

Metadata Standard	Stormwater	Wastewater	Watersupply
Pipe category	Type	Type	Type
Asset Spatial Representation			
Spatial Model Function			
At Asset- Downstream asset ID	ToID	ToID	ToID
At Asset- Downstream asset Type	ToType	ToType	ToType
At Asset- Downstream RL at base of structure in meters (m).			
To Asset - Upstream asset ID	FromID	FromID	FromID
To Asset- Upstream asset Type	FromType	FromType	FromType
To Asset- Upstream RL at base of pipe in meters (m).			
Pipe Label (Local Nomenclature)			
Pipe Network Purpose			
Material of pipe	Material	Material	Material
Nominal diameter in mm	Diameter	Diameter	Diameter
Length of the pipe	Length	Length	Length
Pipe Grade (1: Value) Drop by length			
Pressure class (PN)	Class?	Class?	Class?
Stiffness rating (SN)			
Concrete Load Class			
Pipe Shape	MainShape	MainShape	MainShape
Average burial depth to invert of pipe			
Pipe external wrapping or sleeving			
Pipe Internal Treatment - Rehabilitation			
Internal Diameter Post Treatment			
Pipe Treatment Installation Date			
What is the pipe installation method			

2.c.2.2 Assessment of alignment to NZ Access Chamber -Standard

Core attributes mapping for manholes of stormwater and wastewater.

Table 2.c.(ii) Core manhole attribute mapping (Shortened version)

Metadata Standard	Stormwater	Wastewater
Specific type of Access Chamber		
Asset Spatial Representation		
Spatial Model Function		
RL at access of structure in meters (m).	LidLevel	LidLevel
RL at base of structure in meters (m).	InvertLevel	InvertLevel
Structure Foundation Type		
Structure Treatment Type		
Chamber Construction Material		
Lids fitted - YES/No		
Lid Material		
Step irons fitted - YES/no		
Ladder fitted - YES/no		
External load capacity - tonnes		
Chamber Width (m)		
Chamber Length (m)		
Chamber Height (m)	Depth	Depth
Chamber Diameter - (m)	Diameter	Diameter
Type of security on access		
Chamber max working height -m		
Chamber Lid Shape		
Chamber Lid Style		
Network Flushing Point		

2.c.2.3 Assessment of alignment to NZ Beta Standard pipe material picklists

Valid values mapping for material of pipes for three water.

Table 2.c.(iii) Mapping results of material picklists

Metadata Standard	Stormwater	Wasterwater	Watersupply
ABS		ABS	
AC	AC	AC	AC
ALK			Alkathene
API-STEEL			
BRICKB			
BB			
B-STEEL			
CI	Cl	Cl	Cl
CIPP		CIPP	
CL-DI		CLDI	CLDI
Cor-PE			
DI		DI	DI
EARTH			
GALV-STEEL			
Glass Reinforced Polymer(code not exists)			
HDPE	HDPE	HDPE	HDPE
LDPE	LDPE		
MDPE100	MDPE?	MDPE?	MDPE?
MDPE80	MDPE?	MDPE?	MDPE?
MLDI			
M-PVC	mPVC	mPVC	mPVC
NOVA	Novaflo?		
O-PVC			
PE	PE80?	PE	
PE100	PE100	PE100	PE100
PVC		PVC(Red)?	
RCRR			
Rib-PVC			
SW-ALUM			
STEEL	Steel		Steel
U-PVC	uPVC	uPVC	uPVC
NOTCAP	Unknown	Unknown	Unknown
CU		Copper	
TIMB	Timber		

2.c.3. Christchurch City Council

In order to evaluate the alignment of the pipe and access chamber asset data captured by the Christchurch City Council to the NZ Beta Standard - Simple Network v1.0, we mapped the asset class, feature attribute name, picklist name and picklist values to NZ Beta Standard the asset class, feature attribute name, picklist name and picklist values if such a match exists. For the attribute name, we first aligned the Christchurch City Council's attributes of each network asset. And then the network attributes were aligned with the metadata standard to avoid mapping the same attribute multiple times for each network type, and the results are provided in the same spreadsheet as the attribute grouping in [SC3].

2.c.3.1 Assessment of alignment to NZ pipe Beta Standard

The attributes mapping for the core pipe attributes is presented in Table C1. The picklist mapping for the NZ Beta Standard pipe picklist is presented in Table C2. The pipe material picklist values mapping to the NZ Beta Standard pipe picklist values is presented in Table C3. The core pipe attributes and picklists mapping are provided in the supplementary material [SC1]. This shows which core pipe attributes and picklists specified by the metadata standard could not be mapped to attributes and picklists in the geospatial data, and also shows the attribute name and picklist name consistency across stormwater, wastewater and water supply.

Table C1. Core pipe attribute mapping for Christchurch City Council.

Metadata Standard	Stormwater	Wastewater	Watersupply
Pipe category	SwPipeType	WwPipeType	WsPipeType
Asset Spatial Representation			
Spatial Model Function			
At Asset- Downstream asset ID	DownstreamFeatureID	DownstreamFeatureID	
At Asset- Downstream asset Type	DownstreamFeature	DownstreamFeature	
At Asset- Downstream RL at base of structure in meters (m).	DownstreamInvert	DownstreamInvert	EndInvert
To Asset - Upstream asset ID	UpstreamFeatureID	UpstreamFeatureID	
To Asset- Upstream asset Type	UpstreamFeature	UpstreamFeature	
To Asset- Upstream RL at base of pipe in meters (m).	UpstreamInvert	UpstreamInvert	StartInvert
Pipe Label (Local Nomenclature)	PipeName	PipeName	
Pipe Network Purpose			
Material of pipe	Construction	Construction	Construction
Nominal diameter in mm	NominalDiameter	NominalDiameter	NominalDiameter
Length of the pipe	drvLength	drvLength	drvLength
Pipe Grade (1: Value) Drop by length	Grade	Grade	
Pressure class (PN)	PressureClass	PressureClass	PressureClass
Stiffness rating (SN)	StiffnessClass	StiffnessClass	StiffnessClass
Concrete Load Class	LoadClass	LoadClass	LoadClass
Pipe Shape	Shape	Shape	
Average burial depth to invert of pipe	Depth	Depth	Depth
Pipe external wrapping or sleeving			
Pipe Internal Treatment - Rehabilitation	Treatment	Treatment	Treatment
Internal Diameter Post Treatment	TreatmentDiameter	TreatmentDiameter	TreatmentDiameter
Pipe Treatment Installation Date	TreatmentDate	TreatmentDate	TreatmentDate
What is the pipe installation method	InstallationMethod	InstallationMethod	InstallationMethod

Table C2. Core pipe picklist mapping for Christchurch City Council.

Metadata Standard	Stormwater	Wastewater	Watersupply
MAX_PIPECLASS	domSwPipeType	domWwPipeType	domWsPipeType
MAX_SPATIALTYPE			
MAX_SPATIALMODEL			
MAX_CONNECTINGASSET	domSwPipeUpstreamFeature		
MAX_CONNECTINGASSET	domSwPipeDownstreamFeature		
MAX_MAX_PIPELABEL			
MAX_MAX_PIPEPURPOSE			
MAX_MATERIALPIPE	domSwPipeConstruction	domWwPipeConstruction	domWsPipeConstruction
MAX_NOMINALPRESSURE	domSwPipePressureClass	domWwPipePressureClass	domWsPipePressureClass
MAX_PRESSSTIFFNESS	domSwPipeStiffnessClass	domWwPipeStiffnessClass	domWsPipeStiffnessClass
MAX_PIPELOADTYPE	domSwPipeLoadClass	domWwPipeLoadClass	domWsPipeLoadClass
MAX_PIPESHAPE	domSwPipeShape	domWwPipeShape	
MAX_PIPESLEEVE			
MAX_PIPETREATMENT	domSwPipeTreatment		domWsPipeTreatment
MAX_PIPEINSTALL	domSwPipeInstallationMethod	domWwPipeInstallationMethod	domWsPipeInstallationMethod

Table C3. Pipe material mapping for Christchurch City Council.

Metadata Standard	Stormwater	Wastewater	Watersupply
Acrylonitrile Butadiene Styr	ABS	ABS	ABS
Asbestos Cement	AC	AC	AC
Alkathene	AL	AL	AL
Amer. Petroleum Inst. Steel Pipe	API	API	API
Brick Barrel	BB	BB	BB
Black Brute	BLBRUTE	BLBRUTE	BLBRUTE
Cast Iron	CI	CI	CI
Concrete Lined Ductile Iron	CLDI	CLDI	CLDI
Corrugated Polyethelene	CORRPE	CORRPE	CORRPE
Ductile Iron	DI	DI	DI
Earthenware	EW	EW	EW
Galvanised Steel	GALV	GALV	GALV
Glass Reinforced Polymer	GRP	GRP	GRP
High Density Polyethelene	HDPE	HDPE	HDPE
Low Density Polyethylene	LDPE	LDPE	LDPE
Medium Density Polyethylene 80	MDPE80	MDPE80	MDPE80
Mortar Lined Ductile Iron	MLDI	MLDI	MLDI
NovaFlow	NovaFlow	NovaFlow	NovaFlow
Oriented Polyvinyl Chloride	OPVC	OPVC	OPVC
Polyethelene	PE	PE	PE
Polyethelene 100	PE100	PE100	PE100
Polyvinyl Chloride	PVC	PVC	PVC
Reinf. Concrete Rubber Ringed	RCRR	RCRR	RCRR
Spirally Wound Aluminium	SPWAL	SPWAL	SPWAL
Steel	STEEL	STEEL	STEEL
Unknown	Unknown	Unknown	Unknown
Unplasticised Polyvinyl Chloride	UPVC	UPVC	UPVC
Bitumen Lined Steel			
Cured In Place Plastic			
Medium Density Polyethylene 100			
Modified Polyvinyl Chloride			
Ribloc- Ribbed PVC Pipe Liner			

Since the development of this project was based only on geodatabases, there were some attributes that cannot be aligned to the NZ Beta Standard, which needed further improvement. For stormwater and wastewater pipe, the Christchurch City Council asset attributes and NZ Beta Standard can be aligned to a high degree, and the water supply attributes were less consistent with stormwater and wastewater.

The alignment of the pipe picklists to the pipe picklists of NZ Beta Standard was also not fully implemented. For water supply, the Christchurch City Council did not have the attributes and picklists to store data about the pipe upstream asset and downstream asset.

2.c.3.2 Assessment of alignment to NZ Access Chamber Beta Standard

The attribute mapping for the core access chamber attributes is presented in Table C4. The picklist mapping for the NZ Beta Standard access chamber picklist is presented in Table C5. The core access chamber attributes and picklists mapping are provided in the supplementary material [SC2]. This shows which core access chamber attributes and picklists specified by the metadata standard could not be mapped to attributes and picklists in the geospatial data, and also shows the attribute name and picklist name consistency across stormwater and wastewater.

Table C4. Core access chamber attributes mapping for Christchurch City Council.

Metadata Standard	Stormwater	Wastewater
Specific type of Access Chamber	SwAccessType	WwAccessType
Asset Spatial Representation		
Spatial Model Function		
RL at access of structure in meters (m).	LidLevel	LidLevel
RL at base of structure in meters (m).	BaseLevel	BaseLevel
Structure Foundation Type		
Structure Treatment Type	SwAccessTreatmentType	WwAccessTreatmentType
Chamber Construction Material	SwAccessConstruction	WwAccessConstruction
Lids fitted - YES/No		
Lid Material		
Step irons fitted - YES/no		
Ladder fitted - YES/no		
External load capacity - tonnes		
Chamber Width (m)	PitWidth	PitWidth
Chamber Length (m)	PitLength	PitLength
Chamber Height (m)	PitDepth	PitDepth
Chamber Diameter - (m)		
Type of security on access	SwAccessSecurity	WwAccessSecurity
Chamber max working height -m		
Chamber Lid Shape	SwAccessLidShape	WwAccessLidShape
Chamber Lid Style	SwAccessLidStyle	WwAccessLidStyle
Network Flushing Point		

Table C5. Core access chamber picklist mapping for Christchurch City Council

Metadata Standard	Stormwater	Wastewater
MAX_CHAMBERCLASS	domSwAccessType	domWwAccessType
MAX_SPATIALTYPE		
MAX_SPATIALMODEL		
MAX_FOUNDATION		
MAX_STRUCTURETREAT	domSwAccessTreatmentType	domWwAccessTreatmentType
MAX_MATERIALCIVIL	domSwAccessConstruction	domWwAccessConstruction
MAX_YESNO		
MAX_SECURITY	domSwAccessSecurity	domWwAccessSecurity
MAX_LIDSHAPE	domSwAccessLidShape	domWwAccessLidShape
MAX_LIDSTYLE	domSwAccessLidStyle	domWwAccessLidStyle

3. Methodology and Discussion:

3.a. Auckland Council

Every Council's data ideally should have the following aspects of the network system present in it:

(i) Physical/ As - built attributes:
(ii) Operational and Maintenance attributes
(iii) Financial description attributes
(iv) Risk attributes

While the Healthy Waters Stormwater data standard does not cover all of the above aspects, it was interesting to see the presence of attributes related to above-mentioned information in the AC's data. Noticeably, the NZ Beta standard currently only contains physical/ as – built attributes. A brief count of attributes in AC data and HW standard mapped to the beta national standard is as follows (Tables 3.a-1 and 3.a.-2):

Table 3.a-(i). Attributes Count - Pipe stormwater network.

Attributes Count	Beta Standard-Pipe	HW standard - Pipe	AC data - Pipe
Total	25	27	106
Matched	-	11	10
Unused	-	15	96

Table 3.a-(ii). Attributes Count - Access Chambers stormwater network.

Attributes Count	Beta Standard-Access Chamber	HW standard - Manhole	AC data - Manhole_And_Chamber
Total	23	23	98
Matched	-	12	10
Unused	-	15	88

3.a.1. General Approach and Problem Solving

- Data Extraction:

The initial process involved loading the .gdb files for Auckland Council and reading it in the FME Workbench and FME Data Inspector. FME Workbench is a visual workflow editor used for developing data transformation tools. When the Data Interperability extension is enabled, you can use it to create spatial ETL tools and custom formats. Using FME, Geodatabase files were converted to .csv format as well as the asset layer 'Pipe' was converted to .kml format. However, the drawback of using FME for data conversion was that the extracted data did not include the geometry type and geographical coordinates present in the database.

Hence, python script given in supplementary material [SA_14] to load the 11 asset layers data in .csv format - with each layer containing geometry type, geometry name, and columns for spatial coordinates (NZTM2000). These layers are provided in the supplementary material [SA_7].

- Gap Analysis:

The next step was to perform gap analysis to report on the population statistics of the data layers. This was again achieved by python script given in supplementary material [SA_1]. The resulting statistics for each layer is provided in [SA_6]. The results show many important attributes in every layer that are not populated i.e. are empty columns. Statistics like Negative, Positive, maximum and minimum values, data format etc. were also calculated.

- Mapping Procedure:

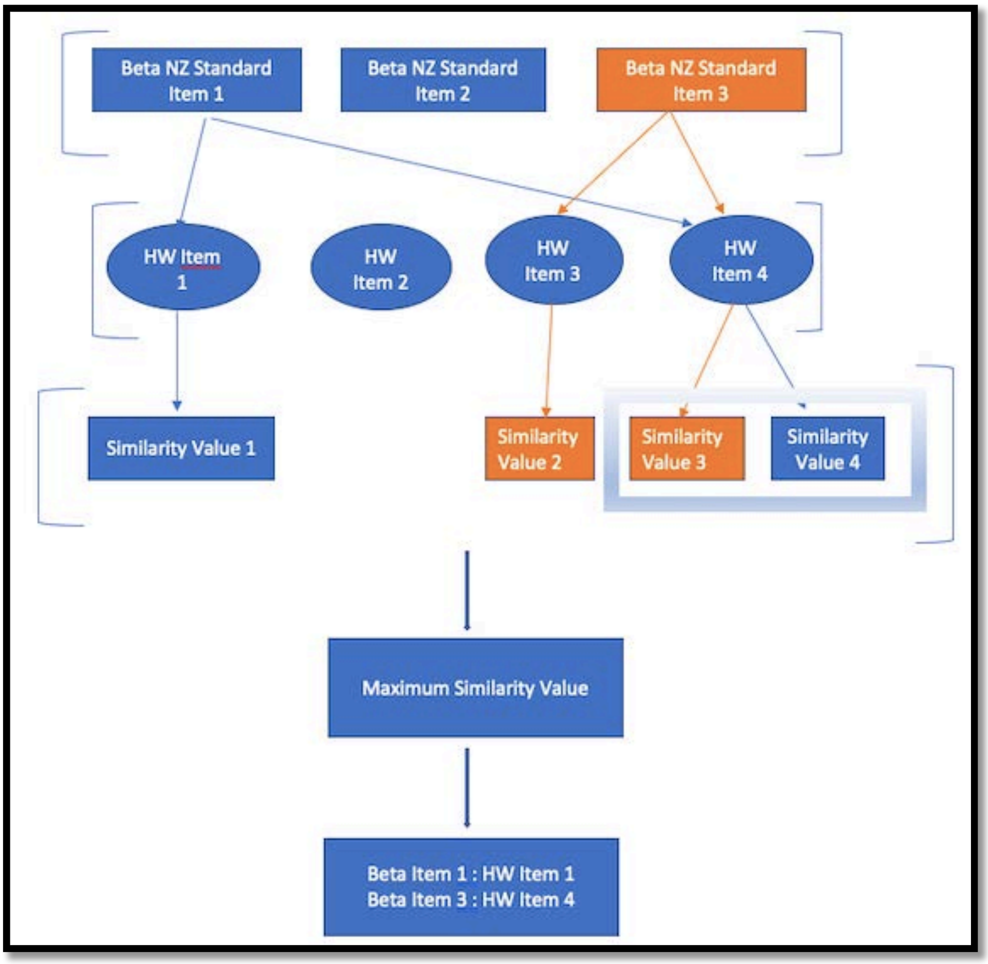
Mapping the nomenclature followed for similar attributes in Healthy Waters Data Standard to the beta NZ standard was automated using various string similarity algorithms in Python. The three algorithms used for this are as follows:

1. Gestalt pattern matching

The module library Difflib provides classes and functions for comparing sequences; and the class difflib.SequenceMatcher his is a flexible class for comparing pairs of sequences of any type, so long as the sequence elements are hashable. The basic algorithm predates, and is a little fancier than, an algorithm published in the late 1980's by Ratcliff and Obershelp under the hyperbolic name "gestalt pattern matching." The idea is to find the longest contiguous matching subsequence that contains no "junk" elements. The same idea is then applied recursively to the pieces of the sequences to the left and to the right of the matching subsequence. The python script developed using this library is given in supplementary material [SA_8] and the resulting csv file is in [SA_11].

A flow chart depicting the general process followed for automatic mapping is as follows: (Graph 3.a.(i))

Graph 3.a.(i) A flowchart of the String Similarity Algorithm



2. Lavenshtein Distance

Fuzzywuzyy is another library that is used for string matching. Fuzzy string matching is a process of finding the strings that match a certain pattern. This library calculates Levenshtein Distance to calculate the differences between sequences, which is a string metric for measuring difference between two sequences. I used the class process.extract() and process.extractOne(). The later method calculates Lavenshtein Distance and gives one attribute corresponding to the attribute in question. The results came out to be better than the method 1. However, by choosing the former class, I can set the limit to the number of results: say limit = 3 gave three attributes corresponding to the target attribute. I can choose the most apt result from the three results. This increased the number of correctly mapped attributes.
The related python script is given in supplementary material [SA_9].

3. TF-IDF cosine algorithm (Package scikit-learn (sklearn))

To apply this algorithm the documents (or the grouped attributes in our case) are first converted to a TF-Idf matrix. Then, the cosine angle between these matrices is calculated, where 1.0 is the Cosine Similarity between the first document with itself. The python script for this algorithm and the csv output is given in supplementary material [SA_10] and [SA_12] respectively. The results are poorest out of the above two methods. As most of the results are matches within the same document.

Lastly, the attribute names were manually checked for best results and were corrected in case of an attribute absent/mismatch. A final mapping spreadsheet for bothasset class is created and given in supplementary material [SA_3].

On comparing the final mapped results to the automatically mapped results, it is found that the methods 1 and 2 performed to the same level of accuracy for the two asset classes; however, method 2. required more manual checking. Method 3 performed the poorest amongst all.

The same approach was used for both Access Chambers as well as Pipes asset class. Additionally, the code lists were mapped manually. The same algorithms can be applied to automatically map codes for attribute values in HW to the National beta standard as well. It is more sensible to map the descriptions of valid values in the two standards rather than the value codes since not every value has a defined valid code. For example, the attribute ‘Pipe_shape’ has valid values as ‘Circle’, ‘Egg’ and ‘Rectangular’ in Beta standard; whereas the mapped values in HW standard are ‘CIRC’, ‘EGG’ and ‘RECT’. Also, there are cases where entirely different codes are used for the same value – say, ‘ABCM’ and ‘AC’ for pipe material described as - ‘Asbestos Cement’. It was also seen that some attributes don’t have any values that can be mapped to the beta standard. For instance, the material list provided for ‘Chamber Construction Material’ do not map to any of the material given in HW standard’s ‘Manhole material’ valid values list.

3.a.2 AC Data Visualisation in Bruce/Nextspace:

Following steps were involved in this process:

- <u>Data Conversion</u>
  The data visualisation in the prototype national infrastructure portal for pipe asset

layer- Bruce requires data in kml or csv format. Another requirement is that the geographical coordinates should be in WGS84 format. Firstly, Hence, the standard NZTM coordinates in the AC data were converted to WGS84 system via the online LINZ coordinate conversion tool. The longitudinal and latitudinal columns for low and hi points of a pipe were calculated in this way, and a new column called 'geometry' for the class – pipe consisting of a string for the combined WGS84 coordinates was added to the file. The same can be obtained by importing the gdb files as .kml datasets in QGIS software. This formatted data now included the following information the asset layer pipe: (as shown in Table.3.a-3)

Table 3.a-3 Pipe Layer columns and descriptions for Bruce Dataset

Columns	Description
Layer	Layer Name
ID	Index starting at 1
Geometry Type	Geometry type, which is defined by numbers
Geometry Name	Geometry name
Longlo	Downstream Longitude
Latlo	Downstream Latitude
Longhi	Upstream Longitude
Lathi	Upstream Latitude
geometry	Combined string of geometry coordinates in WGS84 system
Attributes extracted	Feature attributes

The data extracted in the above format is named as 'Bruce_ACPipe'and provided in the supplementary material [SA_13].

- Getting AC Pipe dataset mapping results for BRUCE

  In order to link the attributes of Pipe layer dataset from Auckland Council to the National Beta Standard, it was necessary to know the mapped values. Since the attribute names mentioned in AC dataset vary moderately from the HW data standard mapped attributes, I manually checked and updated the mapping results for AC dataset alignment to the NZ standard. The dataset - 'Bruce_ACPipe' along with the mapping results were now ready to be imported into BRUCE.

- Visualisation
  The data was imported in a portal common to all the three councils - 'NZQC Bruce account', where it was further visualised and used to run queries.

The link to the common Bruce account is: (credentials required)
https://nzqc.nextspace.host/Bruce~Entities?tab=dashboard

3.a.3 Suggestions and further scope for Quake Centre:

We were successful in dynamically mapping and visualising the data of the three Councils in Bruce. Which means we can map, visualise and analyse dataset from multiple councils as a single dataset, without changing their records. Also, we were able to run multiple single query across all three councils. The data dictionaries built and taxonomy references for attributes and their value list will effectively help automate the schema mapping in the future. The same process done to achieve the stormwater pipe asset layer can be applied on the rest of the layers as well as other network systems – wastewater and potable water systems if collected by other councils. However, this is just a starting step and more councils need to follow the similar of mapping with national metadata standard and incorporating their data in Bruce to help better learn the extent of data alignment within the councils and nationwide. Bruce software is still a work in progress and needs some bug fixing and enhancements to be done. For instance, we need to come up with a User interface in Bruce that can help to view different city councils in one view; and the ability to delete a link of schema mapping, or to update the mapping once it is completed. These bugs and enhancement requirements were discussed with Bruce which they are currently working on.

3.a.4 Suggestions and further scope for Auckland Council

In my viewpoint, the alignment results of HW standard, as well as the Auckland Council data standard to the national standard, reveals many interesting points that can be dug deep to better both the council's standard as well as the beta version national standard. The mapping depicts that there are attributes that can be crucial for the asset information to the council but are not mentioned in the national standard. It should be noted that currently the NZ beta standard does not contain Operational and Maintenance attributes, Financial or Risk attributes.
The population statistics show that there are some attributes in AC data which are relevant to the national standard, but they are not populated (empty columns). The code list of valid value names in HW data standard needs to be updated w.r.t the standard. I noticed, that there are more valid values present in the AC data that map to the beta version national standard; however, such a new updated picklist relating to the current AC data was absent. Hence, the results were limited and might not align with the attribute values in AC's current data.
Additionally, the current dataset shows that the average annual depreciation value cost for stormwater pipes (13558.49 NZD) to the average length of pipe is approximately 416.71 NZD. This can be further grouped annually according to the material of the pipe and its diameter. This can be used to create a regression model for one Council and derive the same for the other councils, or to analyse the factors that play a role towards the depreciation finances.

3.b. Tauranga Council

3.b.1 Processing

3.b.1.i) Preliminary treatment

The gdb files of stormwater, wastewater and potable water were converted to three Excel spreadsheets (xlsx) through FME workbench separately at first, each sheet insides a spreadsheet represents a different asset of a certain water. In order to make all the assets more easily accessible and readable, VBA code was used to extract and each sheet was saved as a separate Excel file within three different folders.

3.b.1.ii) Manual mapping

a) Manual mapping attributes name

Manual mapping is the first step to explore the data, it helps to understand the structure and attributes of assets of each water. The assets of the main reticulation can be easily accessed after the preliminary treatment. However, the terminology and the logic of attributes' name have their own professionalism, which makes it harder to understand what they represent and map to attributes in the NZ Beta Standard.

Frequent and clear communication is a significant important work here. Although I didn't get the expected descriptive documents, the Engineering 101 which Bryan showed me and the process of went over the data with Lynda made me gain a deeper understanding, which enables me to finish the mapping of pipes for stormwater, wastewater and potable water and manholes for stormwater and wastewater.

In terms of fittings, one of three main reticulations, I can only find an asset called nodes in the data. I checked with Greg to see whether it can be mapped with fittings or I'm able to help to improve the data of the node if council would add fittings later. Firstly, we checked the attributes name of nodes, then the GIS visualization. We couldn't find any patterns for understanding and also found some of the assets didn't connect to anything, which didn't make sense in an engineering perspective. I grabbed some ideas of some tools to help detect this situation visually. Then we email to the council and got some information and finally confirmed that nodes are not as same as fittings, in fact, they are even not a physical asset.

b) Manual mapping picklists

I finished the manual attributes name mapping for main reticulations, then I was focused on valid values, in other words: picklists for each attribute. I used R code and extracted the unique values for the attributes in TCC's data which are able to map with the standard. I also asked for full picklists in case the unique values I extracted are incomplete and not covered whole picklists. But some of the attributes still didn't have a provided picklist. So I tried to use my extracted results and accidentally found some errors in data.

In pipes of stormwater data, the unique valid values of "ToType" showed some unnecessary repetitive.

NA
StormWater Manhole
StormWater Structure
StormWater Node
StormWater Sump
StormWater PumpStation
Storm Node
Storm Manhole
StormWater TreatmentDevice
StormWater Miscellaneous
StormWater Soakhole
StormWater manhole

Table 3.b.i)-1 Unique values of “ToType” of stormwater pipe.

We can see that “StormWater Manhole”, “StormWater manhole” and “Storm Manhole” should be the same thing, while for “Storm Node” and “StormWater Node” either. These repeated values would lead to inaccuracy count of each categories' number even some errors while querying for data.

Similar problems also happened in “FromType” of wastewater pipe data and shows in the appendix of manual mapping results.

3.b.1.iii) Gap analysis and statistical analysis

The gap analysis was a summary to manually mapping results, the number of total attributes in NZ Beta Standard and TCC’s data separately, as well as mapped attributes was counted and summarised in a spreadsheet.

The results of the statistical analysis were automatically generated by python code and saved in Excel spreadsheets used suffix as "statistic", which contains the total count, empty/not empty count, unique values count etc. for each attribute of each asset. Due to the limit documentations of python code deal with gdb data, I didn't have too many useful attempts at first, most were failed as it couldn't read the data. With the help of James and small changes were made into the code to adapt to my data, the gdb files were successfully read and achieved statistical analysis.

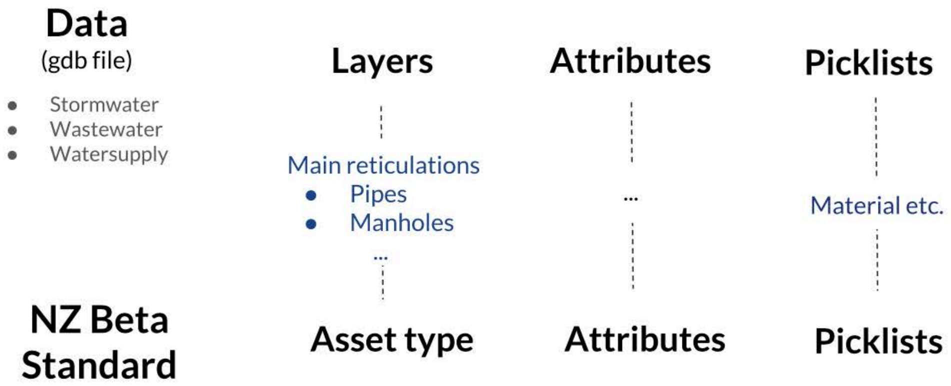
3.b.1.iv) Automatic process for mapping

I discussed with Jing and Manveer about an automatic process for mapping and achieved it successful with the general idea.

To achieve automatic mapping, there are three different classes need to deal with. The layers in the data correspond to asset type in the standard, the attributes and picklists are similar. Firstly, mapping asset type between the data and NZ beta standard, then the attributes within it, finally the picklists of attributes.

I was only focused on the attributes because there is significant difference between the data asset type structure and the standard, which are totally different classification. The algorithms talked later would not help too much of it. As for picklists, in fact, only material picklists need to be mapped while other attributes' picklists were incomplete. So, it was a less repetitive work while code this process would take much longer time.

Graph 3.b.1-1 Mapping class diagram



Through iterate the list of attributes name, the value of similarity was calculated between attributes name of data and standard, then choose the pair which has the maximum similarity and beyond the threshold as the mapping results. This saved a lot of time for manual mapping. It can easily pick up the most similar attributes name, if possible, it probably can help with the attributes name that are relatively different with others.

I used google and found about four different algorithms to calculate the distance between words or strings. I tried two of them, which are Cosine and Levenshtein. The cosine algorithm doesn't consider the order of characters, it works well with the almost same attributes name mapping. However, this kind of situation seldom appears in TCC's data. The Levenshtein distance is sensitive to string length, it calculated the minimum number of single-character edits required to change one word into the other. As the results, it performs much poorer than Cosine algorithm.

3.b.2 Evaluation

- Gap analysis

From the gap analysis results we can see that although NZ Beta Standard has the almost same total amount of attributes for each asset, only a small number of attributes are able to map with the standard. TCC's data didn't have too much overlap with the standard.

Table 3.b-1 Gap analysis results for pipes of three water.

Fields count	NZ Beta Standard-Pipe	SWE-Pipe	WWE-Pipe	WSE-Pipe
Total	60	71	66	62
Matched	-	18	18	17
Unused	-	53	48	45

Table 3.b-2 Gap analysis results for manholes of stormwater and wastewater.

Fields count	NZ Beta Standard-Manhole	SWE-Manholes	WWE-Manholes
Total	53	59	61
Matched	-	13	14
Unused	-	46	47

The consistency between three councils data's mapping results and NZ Beta Standard showed how well the standard can be applied to the different council on a larger scale which discussed in section 5.b.

For Tauranga City Council, the classification of assets is quite different with NZ Beta Standard while the latter is more concerned about physical assets. For TCC both physical assets and non-physical assets exist.

The results was attached in supplementary material “Mapping results”, sheet “Summary”.

- Quality of the data

According to the results of statistical analysis, apart from some existed empty values, there are still many fields unused and have empty values in the data. Some valid values did not consistent with each other.

The results was attached in supplementary material folder <Automatic process results>.

- Automatic process for mapping

From the results, we can see that the automatic process does not have enough correctness compared to the manual results. The basic theory of Cosine algorithm is to compare the character similarity between words (phrases), in other words, it counts how many same characters rather than their true meaning behind it. Therefore, for those phrases who were combined with more words, they are more likely to be automatically mapped with irrelevant attributes.

On the other hand, considering the manual mapping results, only a small number of attributes in TCC’s data are able to map with the NZ Beta Standard. For example, showed in table 3.b-1, there are total 60 different attributes in NZ Beta Standard for pipes, only 18 attributes of the data are able to map with them, which is also a part of the reason that algorithms have pretty low accuracy.

The code was attached in supplementary material folder <code>.

- Visualization

Bruce is a software able to handle the data contains GIS information and visualize on variety of map include earth map. It based on the web browser, therefore it is easy to run and can be access from anywhere. The schema of data can be saved inside Bruce and applied into new data, which let it has ability to handle massive data with complex structure and querying them.

Visualization of TCC's data was achieved within Bruce, which included main reticulations for three waters:
https://tdc.ui.nextspace.host/View_threeWaters

Graph3.b-1 Visualization of main reticulations for three water

- Further investigation

As I mentioned in section 5.c.2.ii), to detect the unusual pattern from geographic view, assets that are not connected to any other assets which are unreasonable from an engineering perspective. For example, some stormwater sumps didn't connect to anything. I checked the standard drawing documents on TCC's IDC website, it seems like at least they should have a sump lead (to connect with pipes). I suggest using convolutional neural networks to detect this pattern: train convolutional neural networks with different feature screening layers to detect the assets have certain features in pictures.

3.c. Christchurch City Council

Defining the differences between the Christchurch City Council metadata standards and the NZ metadata standards visualising the gaps is a problem that was tackled in this project.

3.c.1. Statistical analysis for mapping results of NZ Beta Standard

The statistical analysis for pipe and access chamber assets aligned to the beta standard are presented in Table 3.c.1 and Table 3.c.2.

Table 3.c.1. Statistics computed for pipe mapping result

Attribute count	Beta Standard - pipe	SwPipe	WwPipe	WsPipe
Total	26	52	52	44
Matched	-	21	21	15
Unused	-	31	31	29

Table 3.c.2. Statistics computed for access chamber mapping result

Attribute count	Beta Standard - access chamber	SwAccess	WwAccess
Total	22	39	39
Matched	-	11	11
Unused	-	28	28

For pipe, the Christchurch City Council metadata standard had a high degree of match with the NZ Beta Standard. NZ Beta Standard had a total of 26 attributes for pipe that can align 21 attributes of Christchurch City Council stormwater and wastewater metadata standard. Some NZ Beta Standard attributes which were unable to align to the Christchurch City Council could get by further explored. For instance, the corresponding "Asset Spatial Representation" attribute which could not be directly found in the Christchurch City Council geodatabases, but we can further extract the information programmatically from the geodatabases.

For the access chamber, only half of the attributes that can be mapped to Christchurch City Council metadata standard, some of which related to the lid did not exist in the Christchurch City Council metadata standard.

By integrating the attribute alignment of the three councils (Auckland, Tauranga, Christchurch), the NZ Beta Standard defined the core attributes of the pipe and access chamber, and each council can also know how to improve the management of its respective asset data.

3.c.2. Problems solving

3.c.2.1. Data extraction

Firstly, I extracted the Christchurch City Council asset data from the geodatabases. Before I extracted the data, I used the FME Workbench and FME Data Inspector [6] to view the attributes and data of each layer of each network geospacial database. I used the Geospatial Data Abstraction Library (GDAL) [1] provided by the Open Source Geospatial Foundation (OSGeo) for data extraction. The feature level attributes were extracted for all layers in the geospatial databases, which saved in separated csv files with the layer name. The csv files are provided in supplementary material [SC9] and the python script used to implement the functionality is included in the supplementary material [SC7].

3.c.2.2 Data Quality Assessment

After extracting the attribute values of each asset in the geodatabase provided by the Christchurch City Council, the data quality of the 3-waters asset in Christchurch City Council can be checked first. By statistical analysis of the attribute values stored in the Christchurch City Council geospatial databases, it can be concluded that attribute values such as ID, asset type, service status, ownership, response, location certainty were almost completely captured but other asset attribute values were missing to some extent. But it was impossible to store each attribute value, since even the same asset, there were different asset classifications, such as fittings containing junctions and end cups. The statistics computed for each feature attribute contained in supplementary material [SC10] and the python script for extracting the information is provided in supplementary material [SC7].

3.c.2.3 Assessment of alignment to LINZ metadata standard

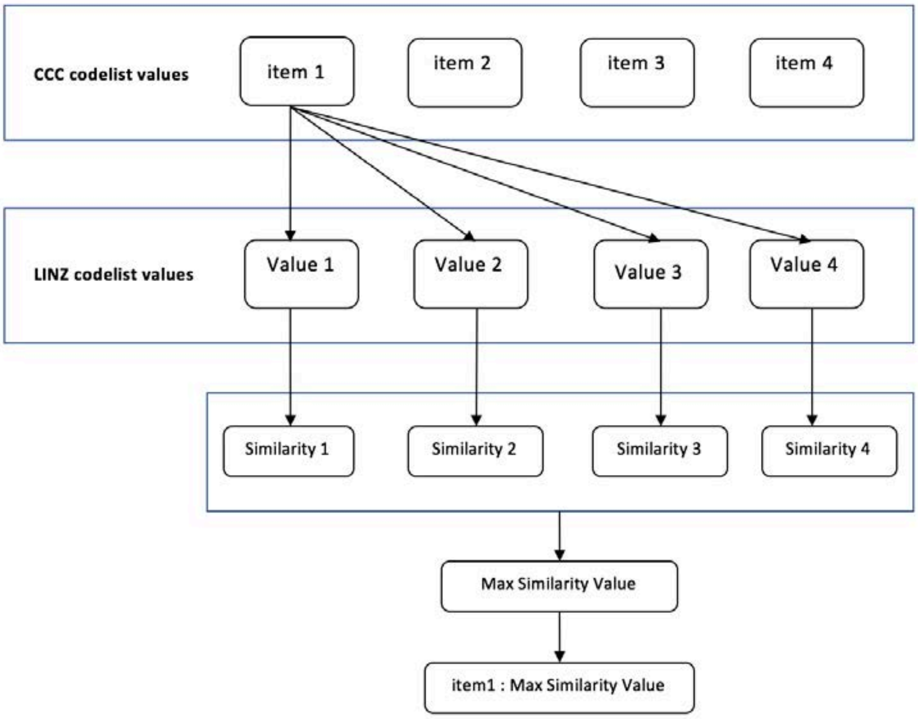
Details on the alignment to LINZ metadata standard is in the Section 6.b of this report.

The gap analysis between Christchurch City Council metadata standardand LINZmetadata standard were performed at four levels, which were asset level, attribute level, codelist level and codelist value level.

The gap analysis of each asset had the same process and the same structure of output, which could be achieved by using python scripts to automatically map to generate the same structure excel file. There were 4 sheets in each asset spreadsheet which were the gap analysis at asset level, attribute level, codelist level and codelist value level. Since these 4 levels were following hierarchy structure, I manually mapped the asset type and attributes and used the python Pandas package to automatically match the codelists of attributes. Then I used my own function (based on difflib library) to automate the alignment of the codelist values.

I used the python difflib library to calculate the similarity between two items and compare the codelist value in Christchurch City Council metadata standard with the codelist values in LINZ metadata standard to get the most similar combination for alignment. The automatic alignment of each codelist value was based on the fact that the codelist was aligned, which reduced the error rate. The flow chart of automatic mapping is provided in Graphic 3.c.1.

Graphic 3.c.1. Automatic mapping flow chart.



The project had to achieve alignment of asset class, attributes, codelists and codelist values, which had a lot of manual work. Using python script to achieve maximum similarity by calculating similarity for automated mapping can greatly improve efficiency. Since the automated mapping was not guaranteed to be 100% correct, manual check was still required after the python script run. But even then, the mapping efficiency can still be improved.

For the analysis of the codelist values, I added a column called statiastic to count the asset codelist value in the corresponding layer. The statistics column was contained in the codelist value level gap analysis which is identified in Table C9 in the Section 6.b of this report. The frequency of each codelist value in the corresponding layer was counted to show the influence of the LINZ metadata standard on the Christchurch City Council assets data. Since the asset data of the Christchurch City Council changed every minute, this statistic can only give the Christchurch City Council a rough degree of influence. The python script used to implement the functionality is included in the supplementary material [SC8]. The automatically mapping results are provided in the supplementary material [SC13] and the final results are provided in the supplementary material [SC6].

3.c.2.4 Assessment of alignment to NZ Beta Standard

In order to analyse the gap between the metadata standards of Christchurch City Council and NZ Beta Standard, Christchurch City Council pipe and access chamber metadata standard were aligned with the NZ Beta Standard, resulting in a missing council Metadata standard asset attributes and Council asset attributes not included in the NZ Beta Standard.

Aligning the Christchurch City Council metadata to the NZ Beta Standard was done after the alignment of the Christchurch City Council metadata standard to the LINZ metadata standard. Since there was no directly loadable spreadsheet to analyse with python's Pandas package, and I only needed to align the pipe and access chamber assets to the NZ Beta Standard, I chose to manually map the asset schema, which was more efficient. The mapping for pipe and access chamber to NZ Beta Standard are provided in the supplementary material [SC4].

3.c.2.5 Visualization

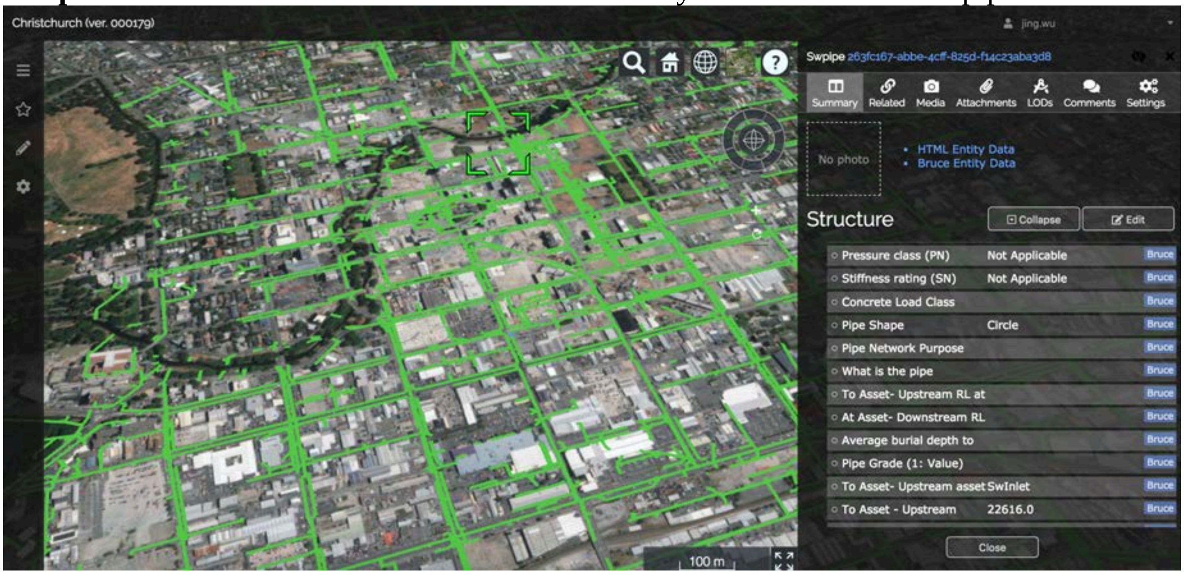
Another goal of this project is to achieve national pipe visualization. Since the layers of geospacial databases provided by the Christchurch City Council did not directly contain coordinate data, I use the GDAL package [1] in Python to extract coordinate data of asset data of different spatial types. Since the project scope only involved the pipe (spatial type was line) and access chamber (spatial type was point), I only extracted the latitude and longitude data of line and point assets. The relevant coordinate data is included in the supplementary material [SC9].

We could get the geometry information by using the feature.GetGeometryRef() function in GDAL library [1]. For the point asset, we could use feature.GetGeometryRef().GetX() and feature.GetGeometryRef().GetY() to get the point latitude and longitude. As for the coordinates of the spatial type of line, the geometry information got by feature.GetGeometryRef().Boundary() contained the upstream and downstream coordinate data. We could use python re library [7] to set a pattern to capture the 4 numbers in the geometry information, which are the latitudes and longitudes of upstream and downstream.

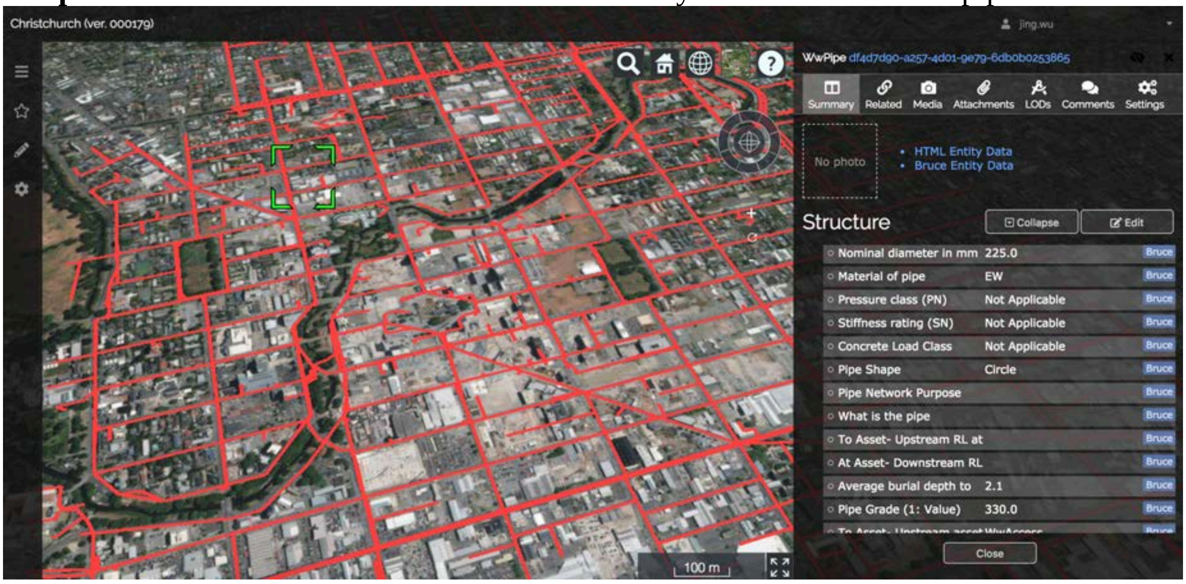
The extracted coordinate data from Christchurch City Council was in NZTM format, but the software (Bruce) that implements visualization only supported WGS84. So, format conversion according to the coordinates supported by Bruce was a must and I created a function implementation transformation in python. The python script used to implement the functionality is included in the supplementary material [SC7]. After the conversion of the coordinate system data was completed, a new column was created in the extracted layer file to store the geometry information of the pipe and access chamber in accordance with the coordinate format required by Bruce. The csv files uploaded to Bruce are provided in [SC11].

The screenshots of the Christchurch City Council stormwater pipe and wastewater pipe are provided in Graphic 3.c.1 and Graphic 3.c.2.

Graphic 3.c.1. Visualization of the Christchurch City Council stormwater pipe



Graphic 3.c.2. Visualization of the Christchurch City Council wastewater pipe



3.c.2. Improvement

3.c.2.1 Alignment evaluation

The current development of this project was based solely on geospatial databases and the Christchurch City Council had other asset management systems which were not considered in this project. The assets attributes that cannot be aligned need to be further improved.

Even with the analysis of geospatial databases, there was still a lot of data that needed to be further explored. For example, "Asset Spatial Representation", the NZ Beta Standard defines the spatial presentation shape of an asset. This attribute was not directly in the Christchurch City Council metadata standard, but the relevant information can be extracted from the geodatabase by using some code processing.

The alignment of codelist values in the codelist also needs to be further refined. In the method of automatically aligning the Christchurch City Council attribute values to the NZ Beta Standard attribute values, the difflib library in the Python standard package was currently used to calculate the similarity of the attribute values, and more similarity algorithms can be tried to select a higher correct rate one.

The NZ Beta Standard and LINZ metadata standard contain all 3-water assets standard, but the project scope only included part of the asset classes. Other asset types are also required the same gap analysis to complete national asset data standardization.

3.c.2.2 Visualization

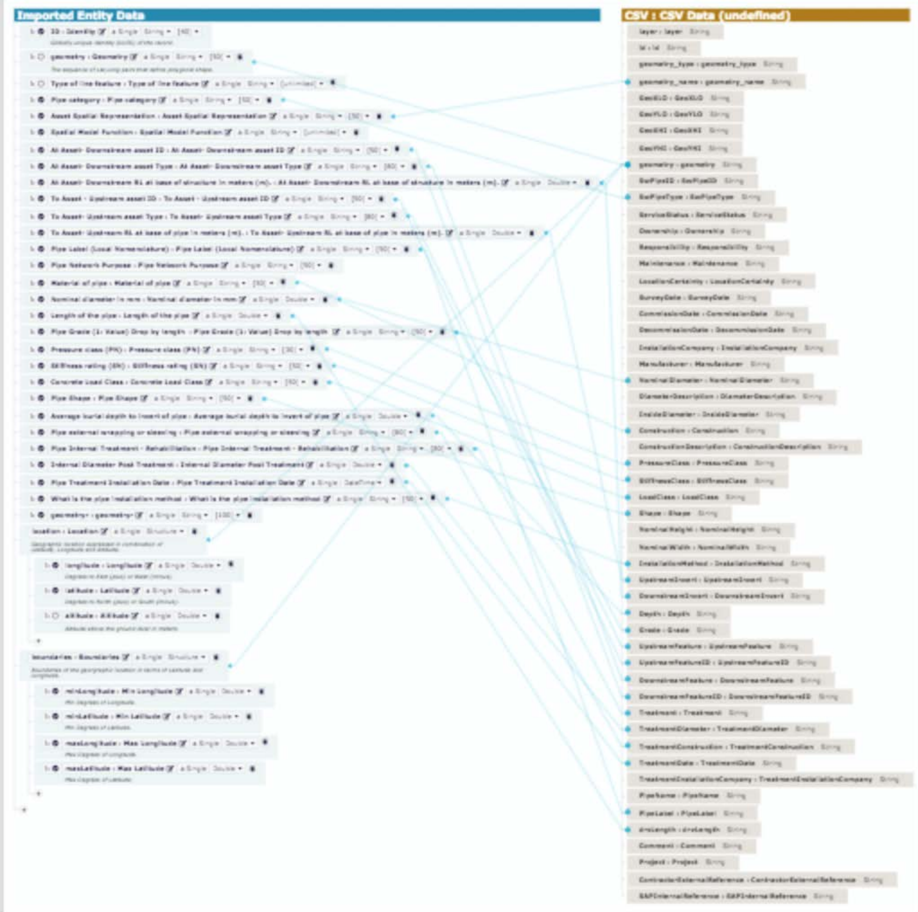
The extraction of geometry information in the geodatabase also needs to be improved. Since the visualization was only based on pipe and access chamber, only the geometry information extraction of point and multilinestring was considered. The geometric information extraction for other spatial types has not been studied in depth.

3.d Higher Level Goals for Quake Centre

3.d.1 Federation (Dynamic Mapping in Bruce)

Firstly, a sample dataset derived from the National Beta standard Asset Pipe was imported to Bruce and the Pipe datasets from the three Councils were then mapped to the NZ standard using the previous mapping information for each council. An example of the mapping links achieved in Bruce are as follows. (Graph 3.d.(i))

Graph 3.d.(i). Snapshot depicting the mapping process in BRUCE for Pipe dataset.



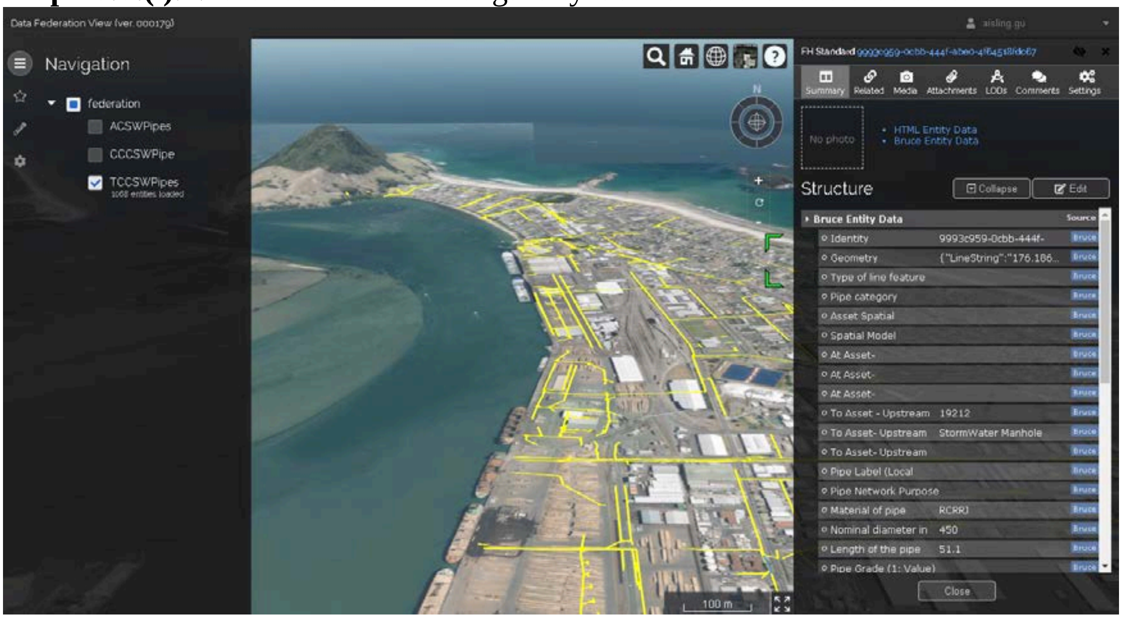
As required by Quake Centre, federation of councils' stormwater pipe data was achieved within Bruce software. In order to integrate data with different structures, we defined a sample NZ Beta - Standard file for pipes and uploaded into Bruce as a "standard" entity. Meanwhile, we also set together about the character limits and data type and whether the attributes were available for indexing, which is an essential part for querying federated data. Then all three councils' stormwater pipe data was uploaded and manually mapped with NZ Beta Standard entity through Bruce schema UI.

3.d.2 Visualization and Query

Through using the methods described above, all of three councils' data was federated as a whole dataset with a consistent schema. The layer of NZ standard now has mapped attributes from all the three councils. UI views and layers were created, and pipes were marked as yellow, also several bookmarks for different areas to help locate where the data will be displayed.

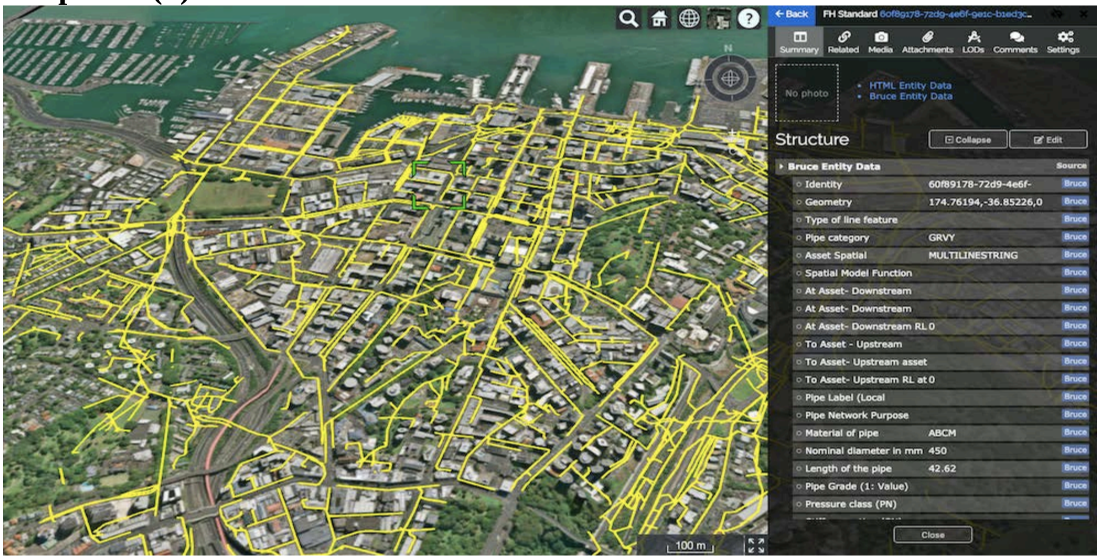
A snapshot of visualisation of NZ Beta standard aligned to the mapped attributes in Tauranga City Council's Stormwater dataset for Pipes is as follows: (Graph 3.d.(i))

Graph 3.d.(i). Visualisation of Tauranga City Council of the federated data



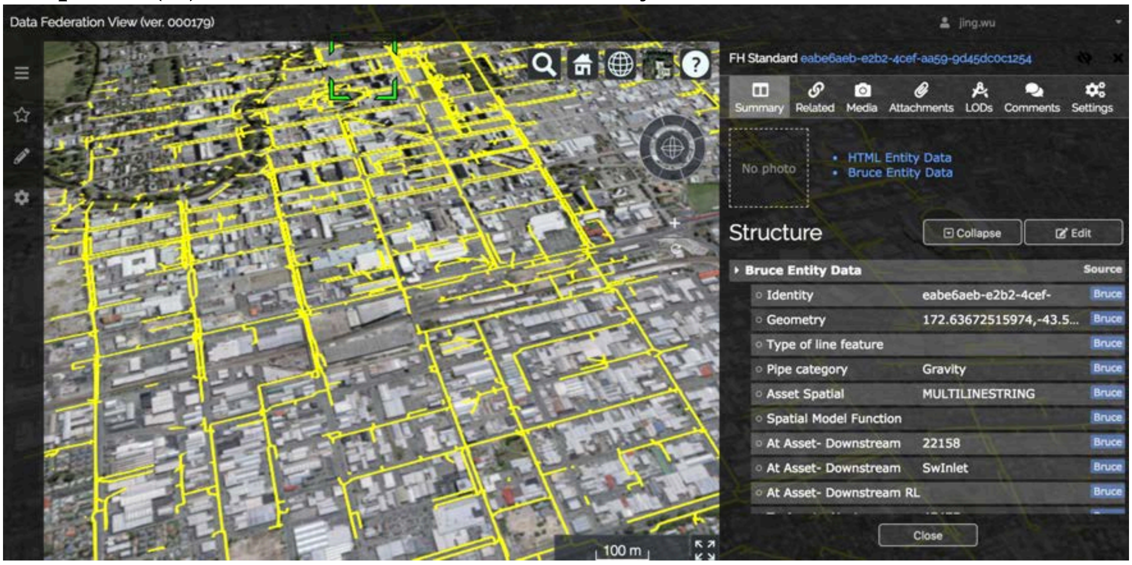
A snapshot of visualisation of NZ Beta standard aligned to the mapped attributes in Auckland Council's Stormwater dataset for Pipes is as follows: (Graph 3.d.(ii))

Graph 3.d.(ii). Visualisation of Auckland Council of the federated data



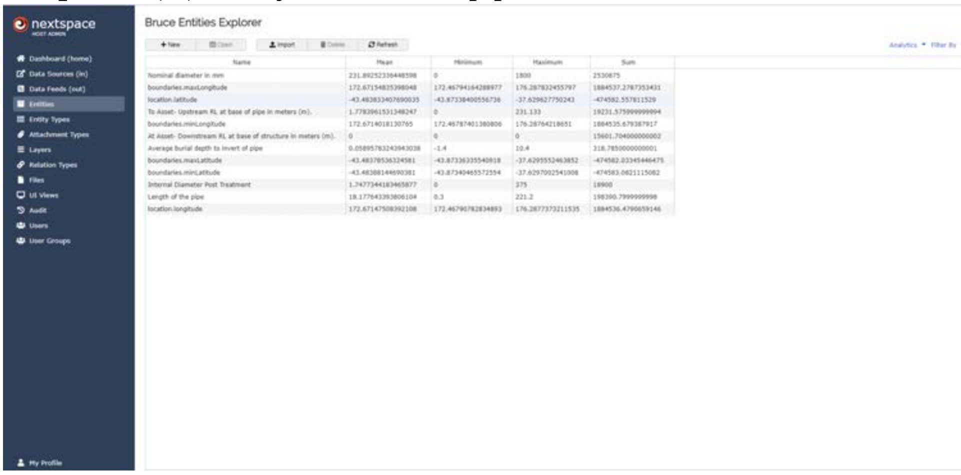
A snapshot of visualisation of NZ Beta standard aligned to the mapped attributes in Christchurch City Council’s Stormwater dataset for Pipes is as follows: (Graph 3.d.(iii))

Graph 3.d.(iii). Visualisation of Christchurch City Council of the federated data

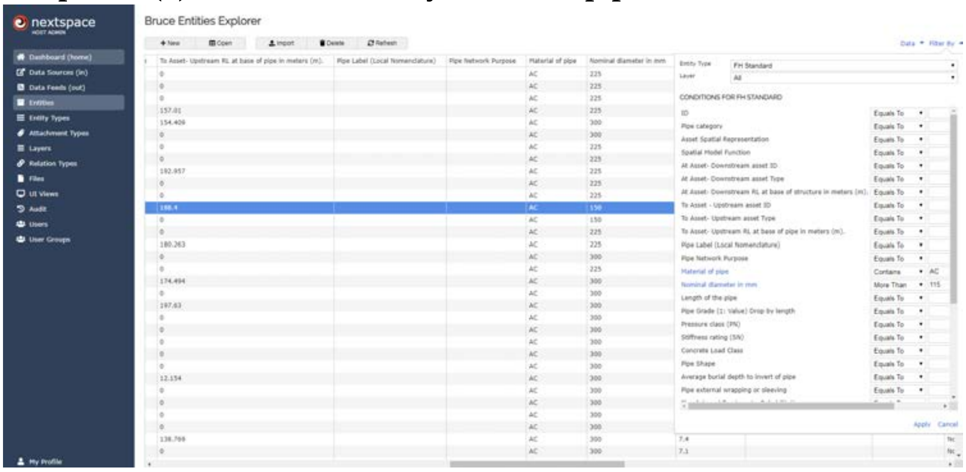


The software provided options of running a single query on the federated data as one entity as well as just one dataset within it. Further basic statistics such as mean/ maximum/ minimum etc. values can be achieved according to the query. An example of this is given in a snapshot 3.d.(iv) and 3.d.(v).

Graph 3.d.(iv). Analytic results for pipe material 'AC' across federated data



Graph 3.d.(v). Combined Query for "AC" pipe material with diameter > 115 meters.



3.d.3 Further Scope of Analysis

We want to analyse the correlation between some attributes of the data that can provide useful information for future. For instance, the annual depreciation cost for the asset layer- pipe might vary according to the attributes- 'Material of pipe', 'Nominal Diameter (mm)' and the 'Length of pipe'. Analysing the relationship between the data values related to these for one council and being able to develop a statistical model such as a regression model can help the council financially and also other councils to check for their compatibility nationally.

4. Supplementary material

4.a. Auckland Council

The directory structure of the supplementary materials of Auckland Council is provided below:

Table SA. Directory structure of the supplementary material of Auckland Council.

<Code>/	
<Auto Mapping Algorithms>/	
Method_1.py	[SA_8]
Method_2.py	[SA_9]
Method_3.py	[SA_10]
<Layers and statistics.py>	[SA_14]
<Documents>/	
<AC data & HW>/	[SA_2]
AC Healthy Water Stormwater Network Asset layers.gdb.zip	
HW Data Standard Tables Definition.xlsx	
<Beta National Standard>/	[SA_1]
FH Asset Meta Data Standard - All Water Assets - Master Data Template.XLSX	
FH Asset Meta Data Standard - Simple - Reticulation v1.0.DOCX	
FH Master Data FH70 Pipe Reticulation.docx	
<Outputs>/	
<AC Stormwater>/	
<Extracted Asset Layers>/	[SA_7](uc T drive)
Catchpit.csv	
…….	
<Final mapped results>/	
Attributes_mapped.xlsx	[SA_3]
Codelists_mapped.xlsx	[SA_4]
Mapped Pipe Data to NZ.xlsx	[SA_5]
<Population Statistics All Layers>/	[SA_6] (uc T drive)
statistics_Catchpit.csv	
…..	
<Automated Mapping Outputs>/	
Method1_difflib.csv	[SA_11]
Method3_cosine.csv	[SA_12]
<Visualisations>/	
AC Pipe data schema.png	
Bruce View 1.png	
Bruce View 2.png	
Bruce View 3.png	
Bruce View 4.png	
Bruce_ACPipe.csv	[SA_13]
Single Query Example.png	
statistics of AC Pipe dataset.png	

Items included in the Auckland Council supplementary material

[SA_8], [SA_9], [SA_10] Auto Mapping Algorithms

These three .py files contains python code for automatically mapping attribute names from Healthy Waters data standard to NZ beta standard. There are 3 .py scripts for the three methods defined in the methodology.

[SA_14] layers and statistics.py
This is the.py script for extracting different asset layers from .gdb format to .csv files, with each asset layer consisting of related geometry type and geometry name.

[SA_2] AC Data & HW
This folder consists of zipped .gdb dataset - ‘AC Healthy Water Stormwater Network Asset layers.gdb’ provided by the Auckland Council. It also contains the spreadsheet for Healthy Waters Data Standard Tables Definition. This spreadsheet contains definitions and standard attributes followed by Auckland Council for all the asset layers.

[SA_1] Beta National Standard
This folder contains all the files rfollowed for this report and for Auckland Council’s outcomes. It consists of the 3 Waters Asset data spreadsheet and the word Document on Simple Reticulation. It also has a separate National standard word document updated just for the Pipe reticulation asset.

[SA_7] Extracted Asset Layers
This folder consists of .csv files of the 11 asset layers extracted using the code in [SA_14] for Auckland Council .gdb dataset.

[SA_3] Attributes_mapped.xlsx
This spreadsheet consists of two spreadsheets for the final mapped attributes results of Pipes and Access Chambers each to the Beta National Standard.

[SA_4] Codelists_mapped.xlsx
This spreadsheet consists of two spreadsheets for the final mapped code list value results of Pipes and Access Chambers each to the Beta National Standard.

[SA_5] Mapped Pipe Data to NZ.xlsx
This spreadsheet consists of the mapping results for Auckland Council dataset layer Pipes to the National Beta Standard.

[SA_6] Population Statistics All Layers
This folder consists of .csv files for the calculated population statistics for each layer in the AC’s .gdb dataset.

[SA_11] Method1_difflib.csv
This .csv file contains mapping results obtained from the python script mentioned in [SA_8]
[SA_12] Method3_cosine.csv
This .csv file contains mapping results obtained from the python script mentioned in[SA_10]

[SA_13] Bruce_ACPipe.csv
This .csv file is the Auckland Council’s Pipe dataset updated according to the requirements of import in BRUCE software. Specifically, geometry for Pipes asset is loaded in WGS64 coordinates and saved as a column named ‘geometry’ for this dataset.

4.b. Tauranga Council

The directory structure of the supplementary materials of Tauranga City Council is provided below:

```
<Code>/
        Automatic process.py
        gis.py
        xlsx.py
        Codelist_extract.ipynb
<Output>/
        Mapping results.xlsx
        Inter-council assessment of consistency.xlsx
        <Automatic process results>/
                SWEPipe-mapping(function name).xlsx
                SWEManhole-mapping(function name).xlsx
                …
        <Tauranga City Councl>/
                <Stormwater>/
                        <gdb>/
                                SDEADMIN_SWMain.xlsx
                                SDEAMIN_SWMain_statistic.xlsx
                                …
                <Wastewater>/
                        <gdb>/
                                SDEADMIN_WWMain.xlsx
                                SDEADMIN_WWMain_stastic.xlsx
                                ...
                <Waterservice>/
                        <gdb>/
                                SDEADMIN_WSMain.xlsx
                                SDEADMIN_WSMain_statistic.xlsx
                                ...
```

4.c. Christchurch City Council

The directory structure of the supplementary materials of Christchurch City Council is provided in Table SC.

Table SC. Directory structure of the supplementary material of Christchurch City Council.

Datasets /	
gdb /	
<network>. gdb.zip	[SC_1]
…	
GISAssetModels.xlsx	[SC_2]
LINZStandards_3Waters.xlsx	[SC_3]
NZ Beta Standard - Simple_-_Reticulation_v1.0.DOCX	[SC_4]
Outputs /	
<network>/	
layers /	[SC9] (uc T drive)
<layer> (layer) .xlsx	
…	
Statistics /	[SC10]
<layer> (statistics) .xlsx	
…	
Bruce /	[SC11]
<layer> (Bruce) .xlsx	
…	
NZ Beta Standard /	
Core Pipe Attribute and Picklist Mapping.xlsx	[SC1]
Core Access Chamber Attribute and Picklist Mapping.xlsx	[SC2]
Pipe Attribute Naming Consistency and Mapping.xlsx	[SC3]
Mapping /	[SC4]
<asset> (NZ Beta Standard) .xlsx	
…	
LINZ /	
Statistics. xlsx	[SC5]
LINZ mapping /	[SC6]
<asset> (gap analysis) .xlsx	
LINZ auto mapping /	[SC13]
<asset> (gap analysis) .xlsx	
…	
Code /	
CCC_gdb_to_geometry.py	[SC7]
CCC_LINZ_auto_mapping.py	[SC8]

Items included in the Christchurch City Council supplementary material:

[SC_1] ArcGIS geodatabase (zip file for each network)

The geospatial data which was grouped by network (stormwater, wastewater and watersupply). There were multiple layers for each network, with a separate layer for each specific asset class and an increasing level of detail.

[SC_2] Christchurch City Council 3-waters metadata standards (spreadsheet)

There were four sheets containing in the spreadsheet. It contains all asset names with definitions, attributes of each asset, domain tables (called codelist in LINZ metadata standards) for each attribute, and domain values (called codelist values in LINZ metadata standards) for each domain table. Using this spreadsheet to get information about each asset from Christchurch City Council.

[SC_3] LINZ metadata standards (spreadsheet)

University of Canterbury provided the LINZ metadata standard they compiled which was formatted as a spreadsheet. There were 2 sheets in the spreadsheet which contained the attributes and codelists of each asset, the codelist values of each codelist.

[SC_4] NZ Beta Standard Asset Metadata Standard (word file)

NZ Beta Standard metadata standard specification document. The metadata standard for the pipe and access chamber is the analysis standard for this project.

[SC1] Core Pipe Attribute and Picklist Mapping (spreadsheet)

The core pipe attribute and picklist mapping contain which core pipe attributes could be aligned to feature attributes and picklist in the CCC metadata standards provided and shows the naming conventions for the CCC metadata standards.

[SC2] Core Access Chamber Attribute and Picklist Mapping (spreadsheet)

The core access chamber attribute and picklist mapping contain which core access chamber attributes could be aligned to feature attributes and picklist in the CCC metadata standards provided and shows the naming conventions for the CCC metadata standards.

[SC3] Pipe Attribute Naming Consistency and Mapping (spreadsheet)

The pipe attribute map displays the same pipe attribute name across the network type, and also maps the attributes of each network to the metadata standard attribute name.

[SC4] Mapping (one spreadsheet for each layer)

The feature level attribute mapping to NZ Beta Standard. This is only provided for the pipe and access chamber asset data, but could be extended to other asset types with some more attribute name.

[SC5] Statistics (spreadsheet)

In the three aspects of attribute, picklist and picklist value, the mapping results are statistically analyzed in four cases of Match, Rename, Not in CCC, Not in LINZ.

[SC6] LINZ mapping (one spreadsheet for each layer)

The feature level attribute mapping to LINZ metadata standard results. This is only provided for the pipe, access chamber, fitting, valve and pump asset data, but could be extended to other asset types with some more attribute name.

[SC7] Python script for extracting data from gdb file (py file)

This python script extracts the data from the gdb file and performs statistical analysis on each layer. Finally, the coordinate data of the pipe and access chamber is converted to WGS84.

[SC8] Python script for mapping CCC standards to LINZ metadata standards (py file)

Assets are mapped at four levels of asset class, attribute, codelist and codelist value and finally formatted into a uniform format for storage in an excel file.

[SC9] Layers (spreadsheet)

The feature level attributes extracted from the geospatial data programmatically.

[SC10] Statistics (one spreadsheet for each layer)

The statistics computed for each feature attribute in each geospatial layer provided.

[SC11] Bruce (one spreadsheet for designated layer)

Asset data for data visualization, and convert its coordinate information into WGS84

[SC13] LINZ auto mapping (one spreadsheet for each layer)

The feature level attributes automatically mapping to LINZ metadata standard. This is only provided for the pipe, access chamber, fitting, valve and pump asset data, but could be extended to other asset types with some more attribute name.

5. References

[1] GDAL/OGR contributors (2018). GDAL/OGR Geospatial Data Abstraction software Library. *Open Source Geospatial Foundation.* http://www.gdal.org/
[2] National Pipe Data Portal (30 July 2018). Alignment of existing pipe asset data with the New Zealand Asset Metadata Standard Vol. 1. Isogonal, Quake Centre.
[3] Python OSGeo Tutorial. MADS projects (2018). Gis_example.py, xlsx_example.py
[4] Geomaps. Online public portal by Auckland Counci. Powered by esri. https://geomapspublic.aucklandcouncil.govt.nz
Softwares Used:
[1] FME workbench and FME data inspector.
[2] Python software
[3] Nextspace/Bruce portal

6. Appendix (Comment out python scripts used)

6.a. Christchurch City Council (LINZ Metadata Standard)

6.a.1. Introduction

New Zealand Land Information (LINZ) divides the water system into three parts, storm water, wastewater and water supply, referred to as 3-waters. The 3-waters Asset Metadata Standard established by LINZ provides asset managers with asset data specifications that support the creation, collection, storage and analysis of data.

Christchurch City Council (CCC) has its own data management structure, and metadata standards for its 3 Waters assets. These have been set up and managed within CCC's asset management system comprising of GIS and SAP. CCC asset metadata standards were set up before LINZ metadata standards and CCC asset metadata standards emerged from asset information requirements from the CCC staff involved with asset management. The CCC metadata standards have progressively been updated based on needs from the staff and expertise & industrial knowledge and expertise within CCC.

With the recent release of LINZ metadata standards, there was a need within CCC to compare its standards with the LINZ standards, in order to help understand and quantify how well CCC metadata standards aligns with LINZ metadata standards. The problem CCC facing was "How to compare these two standards can make CCC understand the difference between its standard and LINZ standard more effectively?"

6.a.2. Purpose

The main goal of this project was to compare the metadata standards for pipe reticulation used in the Christchurch City Council's Asset Management Information System (GIS) with the New Zealand Land Information (LINZ) pipe reticulation metadata standards. By comparing the pipe reticulation metadata standards issued by LINZ with the GIS metadata for pipe reticulation assets managed by CCC, the difference between CCC and LINZ metadata can be obtained from the following four cases:

· Match ------------- values have same name and meaning
· Rename ----------- values have same meaning but different name
· Not in CCC ------- values are not in CCC metadata standard
· Not in LINZ ------ values are not in LINZ metadata standard

6.a.3. Scope

Gap analysis of pipe reticulation asset (pipe, access chamber, fitting and valve) attributes, codelists and codelist values to LINZ metadata standard.

6.a.4. Assessment of alignment to LINZ Metadata Standard

To assess the alignment of the pipe asset data collected by Christchurch City Council to LINZ metadata standard, we mapped the LINZ metadata standard to Christchurch City Council metadata standard, if such a match existed. We also put those metadata standards in LINZ that cannot be mapped to CCC metadata standards and those metadata standards in CCC that cannot be mapped to LINZ metadata standards in the same spreadsheet.

The asset class, attributes, codelists and codelist values of each asset were aligned separately. We manually map asset class and attributes, and automatically align codelists, and codelist values. Eventually all the information is stored in a spreadsheet generated by a python script. The python script provided in the supplementary material [SC8].

Mapping each asset from CCC pipe reticulation to asset from LINZ pipe reticulation generated a spreadsheet of the same structure. This spreadsheet contained four sheets which were the gap analysis at asset level, attribute level, codelist level and codelist value level. The first one is “Asset”, the gap analysis at asset level is identified in Table C6. The second one is “Attributes”, the gap analysis at attribute level is identified in Table C7. The third one is “Codelist name”, the gap analysis at codelist level is identified in Table C8. The last one is “Codelist values”, the gap analysis at codelist value level is identified in Table C9. The results are provided in the supplementary material [SC6].

Table C6. Asset mapping sheet columns and descriptions

Column names	Descriptions
CCC Asset Class	Asset name in CCC
LINZ Asset Class	Asset name in LINZ
Content	Result, whether is matching or not.

Table C7. Attributes mapping sheet columns and descriptions

Column names	Descriptions
CCC Asset Class	Asset name in CCC
CCC Attribute	Attribute of each asset in CCC
Content	Result, whether is matching or not.
LINZ Asset Class	Asset name in LINZ
LINZ Attribute	Attribute of each asset in LINZ
CCC Attribute Data Type	Attribute data type in CCC
LINZ Attribute Data Type	Attribute data type in LINZ

Table C8. Codelist name mapping sheet columns and descriptions

Column names	Descriptions
CCC Asset Class	Asset name in CCC
CCC Attribute	Attribute of each asset in CCC
CCC Codelist	Codelist of each attribute in CCC
Content	Result, whether is matching or not.
LINZ Asset Class	Asset name in LINZ
LINZ Attribute	Attribute of each asset in LINZ
LINZ Codelist	Codelist of each attribute in LINZ

Table C9. Codelist values mapping sheet columns and descriptions

Column names	Descriptions
CCC Asset Class	Asset name in CCC
CCC Attribute	Attribute of each asset in CCC
CCC Codelist	Codelist of each attribute in CCC
CCC Codelist value	Codelist value of each codelist in CCC
Content	Result, whether is matching or not.
LINZ Asset Class	Asset name in LINZ
LINZ Attribute	Attribute of each asset in LINZ
LINZ Codelist	Codelist of each attribute in LINZ
LINZ Codelist value	Codelist value of each codelist in LINZ
Statistics	Statistics on the amount of affected assets

The spreadsheets for pipe reticulation asset mapping showed that the CCC pipe reticulation metadata standards which were matched and not exactly matched to the LINZ pipe reticulation metadata standard. Since LINZ metadata standards are based on pipe reticulation asset data for whole New Zealand, for certain LINZ metadata standards that do not exist in the CCC metadata standard, it is necessary to evaluate according to CCC requirements.

6.a.4. Statistics Summary of each asset mapping

For each asset class, the attribute, codelist and codelist value are counted according to the four cases of Match, Rename, Not in CCC, Not in LINZ. The statistics are identified in Table C10 and the results for each asset are provided in the supplementary material [SC5].

Table C10. Statistics computed for each asset mapping result

Asset Name	Attribute	Codelist	Codelist value
Matched	Count of Matched attributes	Count of Matched codelists	Count of Matched codelist values
Not in CCC	Count of Not in CCC attributes	Count of Not in CCC codelists	Count of Not in CCC codelist values
Not in LINZ	Count of Not in LINZ attributes	Count of Not in LINZ codelists	Count of Not in LINZ codelist values
Rename	Count of Rename attributes	Count of Rename codelists	Count of Rename codelist values

It seems that the attributes, codelists and codelist values that can be matched and renamed account for a small portion by aligning the CCC metadata standard with the LINZ metadata standard. There are many attributes, codelists and codelist values that cannot be aligned for each asset and need to be assessed based on CCC's asset management requirements.

Developing a Proof-of-Concept National Pipe Data Portal

Federation and visualisation of water asset data from distributed asset owners with reference to a common water data standard

Primary Contact Greg Preston
Engineering Core Block
Ground Floor, Rm 138
M +64 27 262 2629
E greg.preston@canterbury.ac.nz

www.quakecentre.co.nz