Gap Analysis between LINZ Metadata standards and CCC Metadata standards for 3-Waters Pipe Reticulation Assets

Contributor

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

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

Name	Jing Wu
Organization	University of Canterbury
Phone	+64 21 149 6394
Email	jing.5@hotmail.com

Contents

1. Executive Summary

1.1 Introduction

New Zealand Land Information (LINZ) divides the water system into three parts, stormwater, wastewater and potable water, 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 meta-data standards emerged from asset information requirements from the CCC staff involved with asset management.

With the recent release of LINZ metadata standards, there is 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. Defining the differences between the Christchurch City Council metadata standards and the LINZ metadata standards visualising the gaps is a problem that was tackled in this project.

1.2 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 metadata standards and LINZ metadata standards 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

1.3 Scope

This project focused on gap analysis between NZ Metadata standards and current CCC standards in the Reticulation Structure for 3 waters. The scope involved the reticulation assets of CCC (Pipe, Access Chamber, Fitting, Valve) and the reticulation assets of LINZ metadata standard (Sections 2.4 (stormwater), 2.4 (wastewater), and 2.4 (potable water)). I need to achieve the gap analysis between the Christchurch City Council defined pipe reticulation assets and LINZ metadata standard defined pipe reticulation assets.

1.4 High Level Goal

When the gap analysis was completed, the resulting outputs can help to the next step - National Infrastructure Viewer. The Christchurch City Council pipe and access chamber asset data were mapped to the metadata standards by mapping the CCC and metadata standards. In this way, the pipe and access chamber asset data from Christchurch City Council, Auckland City Council, Tauranga City Council and Wellington City Council can be managed at the national level for visualization.

2. Data and datasets

2.1 Data and dataset provided by Christchurch City Council

Christchurch City Council provided the metadata standard and geospatial data for stormwater, wastewater and watersupply.

The geospatial data provided by Christchurch City Council was grouped by network and formatted as an ArcGIS geodatabase (gdb). There were multiple layers for each network (stormwater, wastewater and watersupply), with a separate layer for each specific asset class and an increasing level of detail. The dataset is provided in the supplementary material [S1].

The metadata standard provided by Christchurch City Council was formatted as spreadsheet (xlsx). There were 4 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. The dataset is provided in the supplementary material [S2].

2.2 Other data and datasets for analysis

University of Canterbury provided the 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. The dataset is provided in the supplementary material [S3].

3. Approach

3.1 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, I 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.

I mapped each asset from CCC pipe reticulation to asset from LINZ pipe reticulation generated a spreadsheet of the same structure. Each 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 1. The second one is “Attributes”, the gap analysis at attribute level is identified in Table 2. The third one is “Codelist name”, the gap analysis at codelist level is identified in Table 3. The last one is “Codelist values”, the gap analysis at codelist value level is identified in Table 4. The results are provided in the supplementary material [SC2].

Table 1. 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 2. 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 3. 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 4. 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 count 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.

The column Statistics in Table 4 is the statistics of the count of assets with the same codelist values. The frequency of each codelist value in the corresponding layer is counted to show the influence of the LINZ standard on the CCC metadata standard. 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.

3.2 Statistics Summary

The gap between CCC and LINZ metadata can be obtained from the following four cases:

- Match
- Rename
- Not in CCC
- Not in LINZ

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 tables are provided in Table 5 – Table 13 and the results for each asset are also provided in the supplementary material [SC1].

Table 5. Stormwater ACCESS CHAMBER statistics

SwAccess	**Attribute count**	**Codelist count**	**Codelist value count**
Matched	0	0	0
Not in CCC	34	7	97
Not in LINZ	26	11	216
Rename	13	4	10

Table 6. Wastewater ACCESS CHAMBER statistics

WwAccess	**Attribute count**	**Codelist count**	**Codelist value count**
Matched	0	0	0
Not in CCC	34	7	96
Not in LINZ	26	11	221
Rename	13	4	8

Table 7. Stormwater PIPE statistics

SwPipe	**Attribute count**	**Codelist count**	**Codelist value count**
Matched	0	0	9
Not in CCC	37	10	255
Not in LINZ	28	11	386
Rename	24	10	36

Table 8. Watersupply PIPE statistics

WsPipe	**Attribute count**	**Codelist count**	**Codelist value count**
Matched	0	0	9
Not in CCC	44	11	261
Not in LINZ	27	10	317
Rename	17	9	33

Table 9. Wastewater PIPE statistics

WwPipe	Attribute count	Codelist count	Codelist value count
Matched	0	0	9
Not in CCC	37	12	256
Not in LINZ	28	9	354
Rename	22	8	35

Table 10. Stormwater FITTING statistics

SwFitting	Attribute count	Codelist count	Codelist value count
Matched	0	0	0
Not in CCC	25	8	111
Not in LINZ	19	6	71
Rename	7	2	8

Table 11. Watersupply FITTING statistics

WsFitting	Attribute count	Codelist count	Codelist value count
Matched	0	0	0
Not in CCC	25	8	111
Not in LINZ	21	7	75
Rename	7	2	8

Table 12. Wastewater FITTING statistics

WwFitting	Attribute count	Codelist count	Codelist value count
Matched	0	0	0
Not in CCC	25	7	107
Not in LINZ	19	6	75
Rename	7	2	8

Table 13. Stormwater VALVE statistics

SwValve	Attribute count	Codelist count	Codelist value count
Matched	0	0	0
Not in CCC	27	7	197
Not in LINZ	21	14	258
Rename	17	3	12

Table 14. Watersupply VALVE statistics

WsValve	Attribute count	Codelist count	Codelist value count
Matched	0	0	0
Not in CCC	28	7	197
Not in LINZ	24	13	268
Rename	16	5	18

Table 15. Wastewater VALVE statistics

WwValve	Attribute count	Codelist count	Codelist value count
Matched	0	0	0
Not in CCC	28	7	197
Not in LINZ	22	12	261
Rename	16	5	20

3.3 Christchurch City Council Data Quality Analysis

Data quality can be analysed by performing a series of statistics on asset data. We generated the quality of the Christchurch City Council Pipe Reticulation Asset Data by performing a statistical analysis of how much specific asset data is stored, how much data is null, how much data is not empty, the maximum and minimum and a range of other statistical analyses for each attribute in each layer. The statistics analysis is identified in Table 16 and the results for each layer in each geospatial dataset file are provided in the supplementary material [SC6].

Table 16. Statistical analysis columns and descriptions of each attribute for each layer

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

By statistical analysis of the attribute values stored in the CCC 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.

3.4 Report

This report has a holistic introduction to the entire project, including the following:
1. Introduce the main analytical methods. (Master spreadsheet)
2. Perform a statistical summary of the mapping results.
3. Data quality assessment of Christchurch City Council asset data.
4. Business decision result record.
5. The realization of national pipeline visualization.

3.5 High Level Goal Approach

In order to achieve the National Infrustration Viewer goal, we needed to get the asset data from the geospatial databases as well as the geometry information.

3.5.1 Layer attributes with latitude and longitude

The attributes of each layer in each network ArcGIS geodatabase for each network (stormwater, wastewater and watersupply) were extracted from the geospatial databases programmatically using the Geospatial Data Abstraction Library (GDAL) provided by the Open Source Geospatial Foundation (OSGeo) [1]. 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 feature level attributes were extracted and saved separately for each layer with added columns storing other extracted data. Layer summary is identified in Table 17 and the extracted layer attributes are provided in the supplementary material [SC5].

Table 17. 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 extacted	Feature attributes

In order to standardize and visualize the stormwater pipe data, the attribute values and geographic information of each feature in stormwater pipe layer were extracted from the geographic database provided by Christchurch City Council.

3.5.2 Federation and visualization

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 [SC3].

After obtaining the coordinates of the pipe, the national pipe visualization can be realized by dynamic mapping to a NZ metadata standard with the coordinates which was supported by software (Bruce). The csv files that uploaded to Bruce are provided in supplementary material [SC7]

4. Discussion

4.1 Problems solving

I extracted the data and data structure of each layer from the spatial database first. This was a general understanding of the asset data from Christchurch City Council. Then I performed a quality analysis on each layer's data. After the data alignment with the LINZ metadata standard was completed, data federation of the three councils was achieved.

4.1.1 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 geospatial 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 [SC5] and the python script used to implement the functionality is included in the supplementary material [SC3].

4.1.2 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 [SC6] and the python script for extracting the information is provided in supplementary material [SC3].

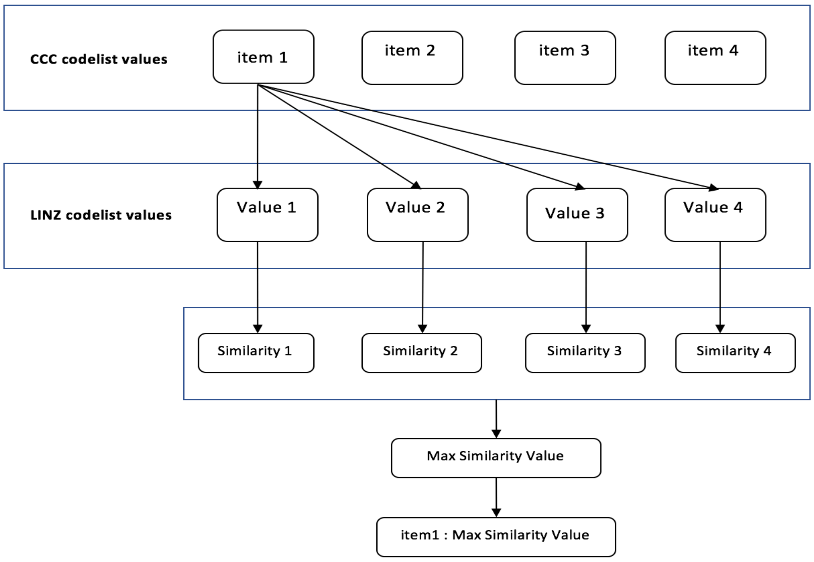
4.1.3 Assessment of alignment to LINZ metadata standard

The gap analysis between Christchurch City Council metadata standard and LINZ metadata 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 [4]) to automate the alignment of the codelist values.

I used the python difflib [4] package to calculate the similarity between two str 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.

Graphic 3. 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 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 [SC4] and the results are provided in the supplementary material [SC2].

4.1.3 Visualization

Another goal of this project is to achieve national pipe visualization. The extracted coordinate information 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 [SC3]. 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 excel files uploaded to Bruce are provided in [SC7].

4.2 Improvement

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.

The alignment of codelist values in the codelist also needs to be further refined. In the method of automatically aligning the CCC attribute value with the LINZ metadata standard codelist value, the difflib package [4] in the Python standard package was currently used to calculate the similarity of the codelist values, and more similarity algorithms can be tried to select a higher correct rate one, such as Fuzzywuzyy library.

The LINZ metadata standard contained all three water assets, but the project scope only included pipes, access chambers, fittings and valves. Other asset types also required the same gap analysis to complete national asset data standardization.

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.

4.3 Difficulties

I have encountered some problems in this project. At first, I didn't know how to define the differentiation between the LINZ metadata standard and the CCC metadata standard. Secondly, the alignment of asset classes and some codelist values was also confusing because of my lack of infrastructure-related knowledge. I completed this project with the help and guidance of CCC finally, which also gave me a better understanding of asset management.

4.3.1 Alignment to LINZ Asset Class

This project involved two asset management data standards for 3-waters, including the LINZ metadata standard and the Christchurch City Council metadata standard. The difference in the classification of assets by two metadata standards directly leaded to the difficulty of mapping their asset types. Only the asset classifications were aligned correctly to continue the alignment of subsequent attributes, codelists and codelist values. For example, SwFitting and SwLateralFitting of CCC were classified as a type of Fitting in LINZ, and the attributes, codelists, codelist values of SwFitting and SwLateralFitting contained in the CCC standards were different from LINZ. The CCC metadata standard classifies asset classes more detailed, perhaps based on CCC's need for asset management. The mapping result of asset classes was supplied by CCC which is provided in supplementary material [SC0].

4.3.2 Alignment to LINZ Codelist Values

For the alignment of codelist values, more difficulty comes from my lack of knowledge about infrastructure engineering. For example, after using automatic alignment of codelist values, due to the lack of information about the pipe material, I was unable to manually check whether the pipe material that was not aligned to the LINZ metadata standard was the same thing with a different name.

4.4 Suggestion

4.4.1 Perfection of LINZ metadata standard

In the LINZ metadata standard, its definition of some asset management was not clear, resulting in the coincidence of values in some codelists. For example, the codelist value of Burst Control is included in both the Valve Purpose and the Valve Type Codelist. the Valve Purpose and the Valve Type Codelist values are provided in Table 18 and Table 19.

Table 18. Valve Purpose Codelist values

Code	Description	Code	Description
AIRIN	Air In	**FLTP**	Flushing Point
AIROUT	Air Out	**IRRIG**	Irrigation
AIRINOUT	Air In & Out	**ISO**	Isolation
PRESBDY	Boundary Press Zone	**LATSUP**	Lateral Supply
BURSTC	Burst Control	**NONE**	No Special Function
BYPASS	Bypass	**NRV**	Non-return/Backflow
CTRLFLOW	Control - Flow	**PRESREDU**	Pressure Reducing
CTRLPRESS	Control - Pressure	**PRESRG**	Pressure Regulation
CTRLFLPR	Control Flow & Press	**PRESRELF**	Pressure Relief
SERV	Customer Service	**PRM**	Pressure Maintaining
DF	Drinking Fountain	**PRV**	Pressure Reducing
DMABDY	DMA Boundary Isolation	**PTR**	Transducer
EMRO	Emergency Only	**SAMPLE**	Sampling Point
EMWR	Emergency Waste Removal	**SCOUR**	Scour
FIREFIGHT	Fire Fighting	**TAP**	Tap
FIRE	Fire Service Connection	**VACSO**	Vacuum Shut Off

Table 19. Valve Type Codelist values

Code	Description	Code	Description
AF	Auto flush	**KN/GATE**	Knife Gate
AIR	Air Release	**L/C**	Level Control
AIRRC	Air-Recycled	**L/GATE**	Lift Gate
ALT	Altitude	**MOTOR**	Motorised
ALT/NRV	Altitude/Non-Return	**NEEDLE**	Needle
ALT/PR	Altitude/Pressure Reducing	**NRV**	Non Return / Reflux / Check
ALT/PS	Altitude/Pressure Sustaining	**P/RLF**	Pressure Relief
ALT/PS/PR	Altitude / Pressure Sustaining / Reducing	**PEN**	Penstock
B/F	Butterfly	**PILOT**	Pilot
BACKFP	Backflow Prevention	**PRV**	Pressure Reducing
BACKFPRPZ	Backflow Prevention RPZ	**PSV**	Pressure Sustaining
BALL	Ball	**PTR**	Transducer
BURSTC	Burst Control	**R/F**	Ring Follower
D/BAR	Drop Bar	**REGR**	Regulator
DAIR	Double Air	**REVS**	Reverse
DIAPH	Diaphragm	**S/C**	Swing Check
FC	Flow Control	**S/COCK**	Stopcock
FERRULE	Ferrule	**SAIR**	Single Air
FGAP	Flap Gap	**SLEEVE**	Sleeve
FLOAT	Float	**SLUICE**	Sluice
FOOT	Foot	**SOLENOID**	Solenoid
GATE	Gate	**STOP**	Stop
HYOFF	Hydrant-Offset	**STOPRC**	Stop-Recycled
ISO	Isolation	**TC**	Test Cock
KEYGATE	Key Gate	**W**	Wheel

Second, some of the contents of the LINZ standard were not clearly stated, and some contents cannot be aligned. For example, the codelist of the Stiffness Class contained three types of SN2500 (Relining, Buried), SN5000 (Minor roads), and SN10000 (Landfill, Well pipes) with the meaning was not clear. CCC classified the stiffness class into 13 categories according to the stiffness, such as SN10 (stiffness 10 kN/m2).

CCC is more detailed about the division of codelist values, which may be more conducive to asset management. For instance, in the pipe material codelist, CCC divided the Polyethylene-based materials into six types: Alkathene, Low Density Polyethylene, Medium Density Polyethylene 80, Polyethylene, Polyethylene 100, and PE-Al-PE Contamination Barrier, while LINZ only had Polyethylene, which referred to all such materials. Also, the LINZ pipe material codelist contained some uncommon material types, such as Acrylonitrile Butadiene Styrene, which may not be necessary for CCC. The mapping result of pipe material was provided by CCC which is provided in supplementary material [SC0].

4.4.2 Perfection of CCC metadata standard

Some content in the CCC metadata standard was missing. But it is necessary to determine whether the content included in the LINZ standard but not included in the CCC is necessary based on the needs of CCC asset management. These are all provided in the supplementary material [SC2].

5. Supplementary Material

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                                   [SC_1]
        …
    GISAssetModels.xlsx                                  [SC_2]
    LINZStandards_3Waters.xlsx                           [SC_3]

LINZ /
    Asset and pipe material mapping.xlsx                 [SC0]
    Statistics. xlsx                                     [SC1]
    LINZ mapping /                                       [SC2]
        <asset> (gap analysis) .xlsx
        …
Python /
    CCC_gdb_to_geometry.py                               [SC3]
    CCC_LINZ_auto_mapping.py                             [SC4]
<network>/
    layers /                                             [SC5]
        <layer> (layer) .xlsx
        …
    Statistics /                                         [SC6]
        <layer> (statistics) .xlsx
        …
    Bruce /                                              [SC7]
        <layer> (Bruce) .xlsx
        …
Visualization /                                          [SC8]
…
```

Items included in the Christchurch City Council supplementary material

[SC_1] ArcGIS geodatabase (gdb)

> 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 for each attribute, and domain values for each domain table.

[SC_3] LINZ metadata standards (spreadsheet)

There were 2 sheets in the spreadsheet which contained the attributes and codelists of each asset, the codelist values of each codelist.

[SC0] Asset and pipe material mapping (spreadsheet)

The Christchurch City Council asset classes, pipe material and LINZ metadata standard mapping results

[SC1] 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.

[SC2] LINZ mapping (one spreadsheet for each layer)

The feature level attribute mapping to LINZ metadata standard. This is only provided for the pipe, access chamber, fitting, valve and pump asset data.

[SC3] 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.

[SC4] Script for mapping CCC metadata 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.

[SC5] Layers (spreadsheet)

The feature level attributes extracted from the geospatial data programmatically.

[SC6] Statistics (one spreadsheet for each layer)

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

[SC7] Bruce (one spreadsheet for designated layer)

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

[SC8] Visualization (png files)

Visualization screenshots on Bruce

6. 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] Python documentation. https://docs.python.org/2/library/difflib.html

Software Used:

[1] FME workbench and FME data inspector.

[2] Python software

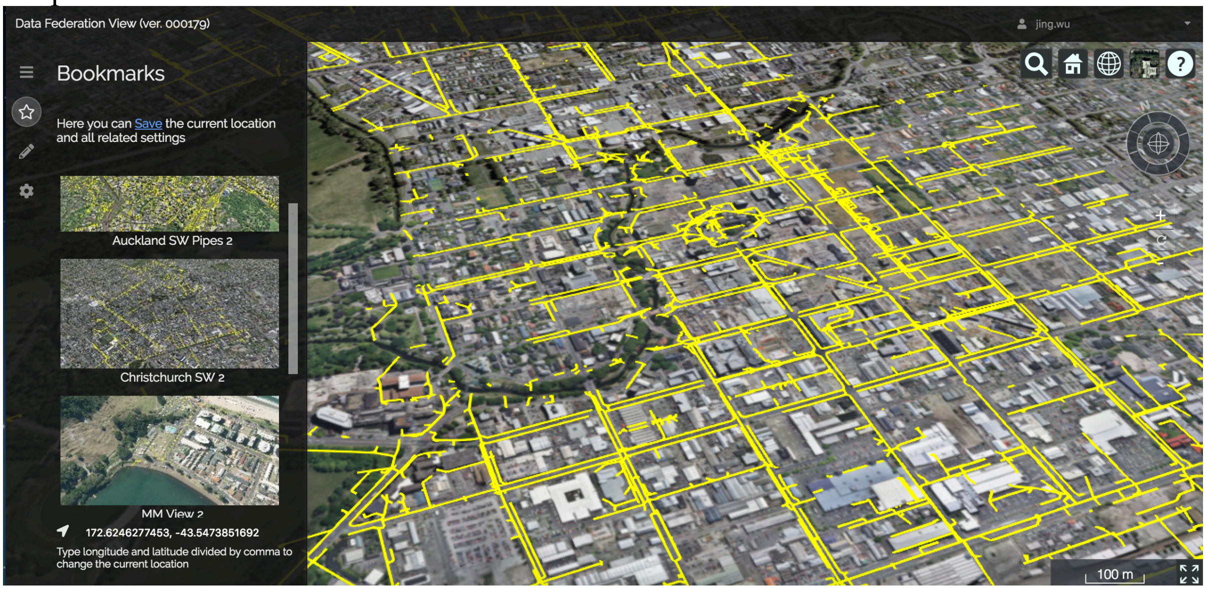
[3] Nextspace/Bruce portal

7. Appendix

Bruce visualisation for Federated data of 3 Councils

Graph 1. Federation data view for Christchurch

Graph 2. Federation data view has different bookmarks

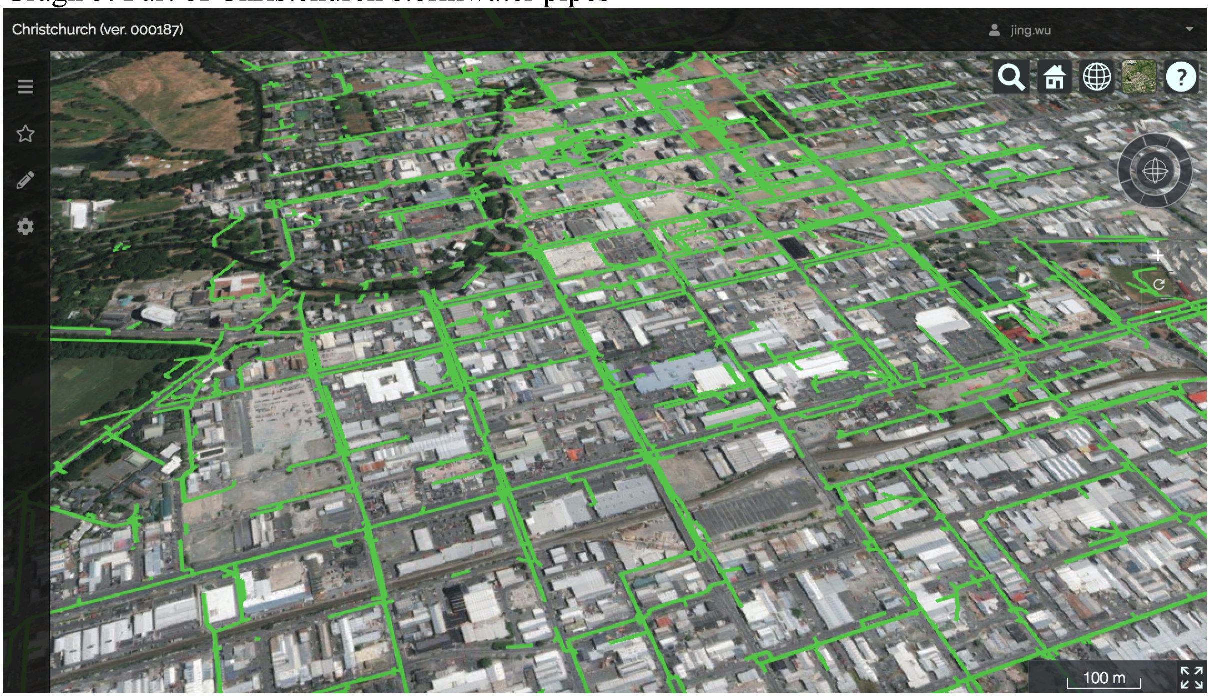


Bruce overall visualisation for Christchurch City Council

Since Bruce is still under development, there are still many features that need to improve. For example, when the height of the visualization is higher than 500m, all assets will disappear on the screen.

Graph 4. Christchurch 3-water pipes visualization

Gragh 5. Part of Christchurch stormwater pipes



Gragh 6. Part of Christchurch wastewater pipes

Gragh 7. Part of Christchurch watersupply pipes