ADDING VALUE TO GIS NETWORK DATA THROUGH MODEL VALIDATION AND VERIFICATION.

TRISTAN JAMIESON - WATERSHED CONSULTATION LTD, BASSAM HALABI –GHD LTD, ONNO URSEM METROWATER LTD

ABSTRACT

Model development, at its most basic level, involves elevating existing information (network data, current and future population data, water usage, rainfall etc) from discrete data sets into a unified and internally consistent database. During this process, the information undergoes rigorous scrutiny. As a result of this analysis, a number of changes to the data often occur. Assuming best practice, these are based either on primary data sets (such as as-built drawings, operator confidence, field monitoring etc) or good engineering judgment and assumption.

Further to the data improvement achieved, a significant review (such as network topology and operation) is undertaken during the validation process. This can identify (amongst other things): unexpected weir settings at CSOs, pumping station operation and capacity, pipe roughness and condition, valve status and PRV settings.

Assuming the changes are correct, the updated network information offers an opportunity to add value to the existing data set. This value can be in the form of confidence in the connectivity, verification of zone/catchment boundaries, identification of erroneous pipe diameters, inverts, conditions and materials, wrongly labelled service connections, nodes located on continuous pipes and duplicated network - to name but a few.

This paper discusses a pilot study undertaken in conjunction with Metrowater Ltd. The study sought to determine the viability of extracting the data improvements from their existing calibrated water distribution models - and installing it into their existing GIS. The pilot study was extremely successful, allowing Metrowater Ltd to add significant value to their existing GIS data, for a fraction of the cost of retrieving the information from the field.

KEYWORDS

GIS, Model Validation, Data Improvement

1 INTRODUCTION

During 2004-2005, Metrowater Ltd (Metrowater) commissioned the development of fully calibrated models for their water supply zones. This involved using the existing Metrowater GIS data as a basis for the models, which was refined, updated and corrected as new information surfaced during the development and calibration process. The level of data scrutiny associated with model development and calibration means significant value can be added to the data during the process. This value can be in the form of confidence in the connectivity, verification of zone boundaries, identification of erroneous pipe diameters and materials, wrongly labelled service connections, nodes located on continuous pipes and duplicated network - to name but a few.

Further to the 'desktop' scrutiny associated with model development and calibration, pressure and flow logging and hydrant testing enabled field inspections of critical sites within the systems. Incorporating any information derived from these inspections into the model data also adds significant value.

In 2008, Metrowater engaged GHD Ltd and Watershed Consultation Ltd to review the data contained in the models, and in the associated documentation. Following this review the companies also worked to provide a pilot study which would determine the value of data improvement held in the modelling studies, define a methodology with which to identify and extract the data, and to upload the data into the Metrowater GIS.

2 PIOLT STUDY OBJECTIVES

The Model of the Konini Water Supply Zone was selected as the basis of the pilot study. The data within the Konin model was integrated based on the following three project objectives:

1. An assessment of the available data. Other than the model itself, the model build process develops a significant amount of documentation which can be used to verify or validate model data. The data and documentation collected for this project included Model GIS exports, current Metrowater GIS and documentation accompanying the original model build projects, and build and calibration reports.
2. Development of a robust methodology for identifying and uplifting the value added data from the models to the current GIS. While it is not onerous to identify differences between the data held in the model and the data in the GIS, the need to ensure that only well verified and validated data would be uplifted to the GIS data base was identified early in the project. This would be achieved through using a robust and auditable methodology.
3. To test and refine the above methodology on the Konini Supply Zone.

3 ASSESSMENT OF AVAILABLE DATA

A data manual was maintained throughout the project, providing a record of - the data received, when it was received and from whom. Each data set received was reviewed and commented on as part of the manual. The manual was then appended to the final report. Aside from maintaining good data handling practice, the purpose of this data manual was to develop a picture of what data was available for the Konini Water Supply Zone. In turn, the viability of undertaking similar projects on other models could be assessed - based on their available data.

Metrowater currently uses five GIS layers to store and retrieve information surrounding the Water Distribution Network. These are:

1. Water mains
2. Valves
3. Hydrants
4. Nodes
5. Supply Zone Extents

The Water Mains layer lists 53 attributes against each water main, the Hydrants layer - 42 against each hydrant, the Valves layer - 29 and the Nodes layer - 30. Many of these attributes (such as length and diameter) can be verified and/or revised as part of the model build procedure. However, there are some exceptions, such as joint type and depth. These have not been considered further and were removed from the study. There are also a number of attributes which pertain to the operation of the network (such as performance and criticality). While it is appreciated that these could be reviewed in light of the modelling results, it was considered to be outside the scope of the project.

The existing model for Konini was built in MWH Soft's H2O MAP water distribution modelling software. This operates a Microsoft Access Database, which is capable of importing and exporting data in common GIS formats, such as ESRI shape files and Mapinfo MID/MIF files. The database is set up so that all data is exported (shape files, dwg, or mid/mif) under the following data sets

1. Junction
2. Pipe
3. Tank
4. Valve

In the case of the data exported from the model, the Water Mains table listed 23 attributes against each water main, Hydrants – 29, Valves -16 and Nodes -29. As discussed above, many of these attributes (such as length and diameter) can be used to verify and/or revised the current GIS data. However, there are some which can't (such as Demand and Average Daily Profile) and these were not considered further.

Some of this information was imported directly from the Metrowater GIS database and remains unchanged, such as the Asset ID and the Year of Installation. However, some information was verified and refined during the model build process, such as pipe length and valve status etc. Additionally, other data has been included for model completeness, such as the representation of WSL bulk meters as tanks and, in some cases, PRV's to represent pressure fluctuations at zone boundaries. Clearly discretion and engineering judgement was required to determine what was appropriate to uplift to the GIS, and what was not.

4 COMPARISON OF MODEL AND GIS DATA

In attempting to use the model data to verify the GIS data, it was important to appreciate which data was discarded and which data was introduced as part of the model build process. It was also crucial to understand the relationship between the data sets, for example valves are represented very differently in the GIS and the Model, and a direct comparison is not immediately achievable.

While considerable value has been added to the original GIS used to develop the Model, there has also been a significant amount of development undertaken directly on the GIS - since the models were developed. Almost half the assets in the Konini Zone were tagged as having been edited since April 2005 (the closure of the Konini model development phase). While this covers the majority of the network, it is important to appreciate that, in all likelihood; only one or two of the attributes for each asset have been changed. The update methodology described later makes measures to avoid the update of any attributes that have been edited post model development.

For historical reasons, the databases for the H2O Map models are referenced in the New Zealand Map Grid coordinate system. However the current GIS is in Transverse Mercator. Because of this difference, the data has undergone a number of translations during importation into the modelling package. Further manipulation again is required to make the model-validated data directly spatially comparable with the current GIS data. These processes are automated, however a minor spatial discrepancy was observed in the H2O MAP data (in the order of a 5mm offset). This was attributed to the transfer process into H2O MAP, associated with a decrease in the accuracy of the asset coordinates.

In order to understand the differences between the two data sets, they have been overlaid in GIS. This process displays the differences in the visual aspects of the data, such as spatial location, and inclusion/exclusion of certain data items etc.

The following figures were produced to visually inspect differences between the two data sets:

- Model Extent: this figure was used to illustrate the number of pipes that were excluded from the GIS during the model development. The majority of these pipes are excluded based on Facility Status. However, a significant number have been excluded for reasons such as, duplicated network, private network and incorrectly labelled facility status.

- Pipe Diameters: this figure was issued to illustrate the differences in pipe diameters between the 2008 GIS and the Model Data set.

- Material Population: Displays the missing Material data in the current GIS which was available from the model database.

- Point Data Issues: This illustrates some of the differences between the 2004 GIS data used to develop the Model and the current GIS data. While some of these issues (in this case orphaned or duplicated nodes/hydrants and valves) were resolved in the current GIS, some new ones have arisen.

The deliverable for the data comparison section of this project was a table that compared the assets (and their associated attributes) for each data set. Each comparable attribute filed was identified as either directly comparable or indirectly comparable. Directly comparable attributes can be mapped using automated functions, while indirectly comparable attributes require some manual assistance, and more importantly engineering judgment. Table 1 summarises this deliverable.

Table 1 Summary of Data Comparison Analysis.

Data Set	Konini GIS Data			Konini Model Data			Comparable Fields	
	Entries in data set	Attri-butes per Data Item	Relevant Attributes	Entries in data set	Attri-butes per Data Item	Relevant Attributes	Directly Compar-able	Indirectly Compar-able
Pipes	6018	53	9	5712	23	9	6	3
Hydrants	2066	42	4	1696	29	4	4	0
Valves	2182	29	6	16	19	6	4	2
Zone Boundaries	1	0	-	1	0	-		1

5 DATA SELECTION METHODOLOGY

Integrating improved asset data back into a GIS data base post modelling study is, in general, a relatively simple process. However, in this specific case, both data sets (model and GIS) have undergone additional improvement/development since the completion of the Konini Model Build and Calibration project. Hence this project, which effectively "reunited" the data sets, and required a methodology that removed the chance of replacing data that has a high level of confidence, with that of a lower level.

Figure 1 below outlines the three main data streams involved in this project. Reuniting divergent data sets, which have both undergone change and development during separation, requires a number of protocols to be maintained to ensure the best possible outcome. These include:

- Ensuring that any data reintroduced into the main database is of better quality than the item it is replacing. This involves understanding the limitations of both the superseded and superseding data items.
- Audit Trails, so that any revision to the existing GIS data set can easily be verified.
- Where spatial deterioration has occurred, only the data attributes are to be updated.

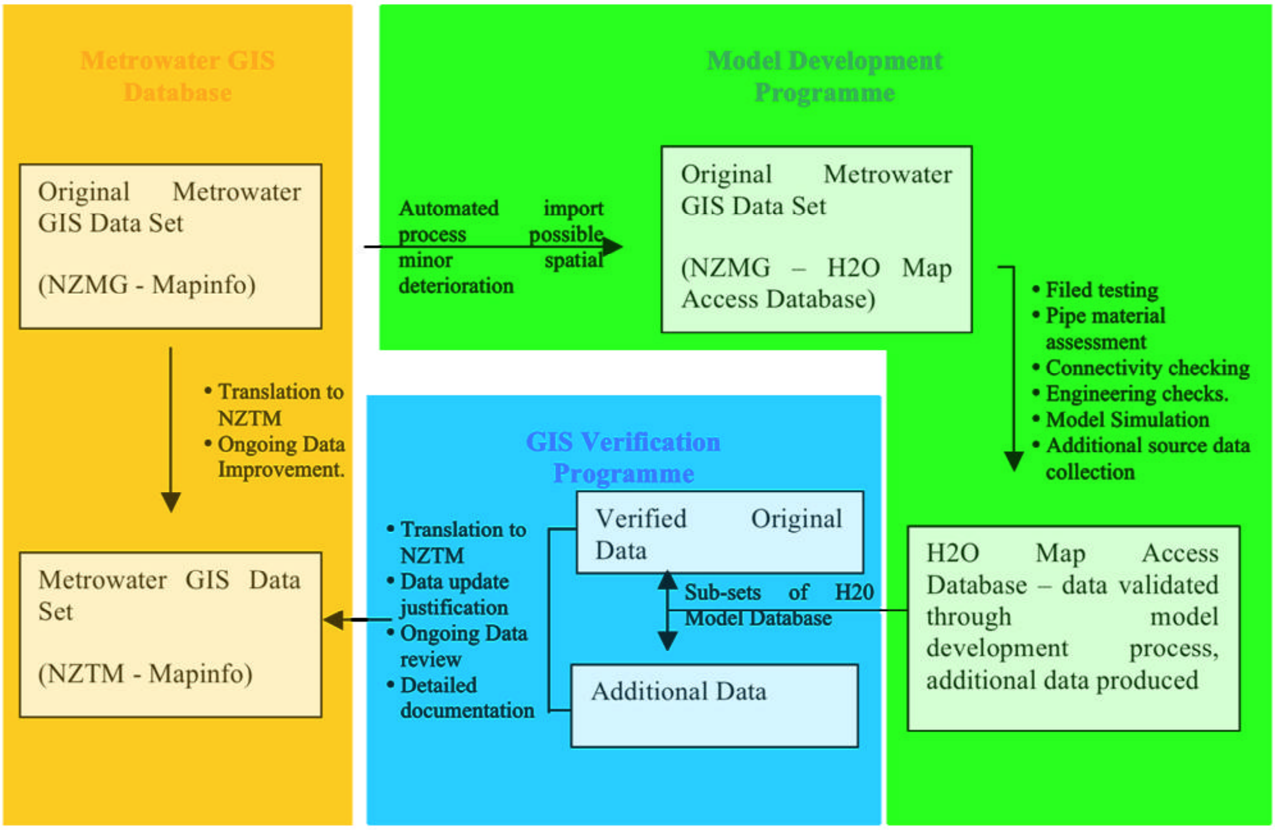

Figure 1 Data routes and processes associated with its development

Figure 1 has split the development of the Water Distribution Asset Data into three distinct, but dependent, processes. Each of these processes is discussed in the following sections.

5.1 METROWATER GIS DATABASE

The data used to develop the H2O Map models was taken from the Metrowater GIS database in 2004. Because each supply zone was modelled separately, and sequentially, the actual date-stamp for the GIS used differs for each model. According to the project notes, any additions/improvements to the GIS data made by Metrowater during the model development period were included in models. Model development was completed in April 2005. Hence any changes or additions to the GIS data since April 2005 should not be subjected to change unless justifiably so.

It is, however, important to note that each item (pipe, valve, hydrant and node) has a number of data fields associated with it. While a data field for a specific item may have been updated in the GIS database post April 2005, this is not necessarily in conflict with updates proposed in the modelling exercise relating to other data fields on that item.

5.2 MODEL DEVELOPMENT PROGRAM

There were a number of data improvement steps undertaken as part of a model development program. These can be conveniently lumped into three distinct categories for the purpose of data improvement:

- Data cleanup using GIS techniques – this involves data sense checks such as removing duplication, removal of discarded assets, connectivity etc, and;
- Simulation added value – a number of data anomalies can be identified during the model calibration process. These include, but are not limited to, pipe diameter, roughness and connectivity. These

attributes will have been manually edited during the model development stage, and so there is generally documentation regarding the assumptions and levels of confidence, and,

- Additional data sources – the model development process is very data intensive and, inevitably, the quest for data uncovers useful data sources that haven't been incorporated into the GIS. In this case, a pipe material database was identified, which was used in the development process for the model, and has not been incorporated into the GIS.

5.3 GIS VERIFICATION PROGRAM

For the purposes of this project, data from the model database was split into two main streams:

- Original unchanged data which, though unaltered, has been verified to some degree by the model development process, and
- Additions and/or revisions to the existing data set.

Obviously, it is extremely difficult to quantify the extent of verification provided to data that remains unchanged by the modelling process. However, it should be recognised that a certain level of data validation has taken place. This project focused only on the data which was altered by the modelling process.

All additions and/or revisions to the April 2005 GIS, which were made during model development, needed to be checked against the current GIS. This process ensured improvements to the data identified during the model development process can be incorporated into the current GIS.

As discussed previously, it was crucial that data revisions recommended as part of the Model Development project did not overwrite those which have taken place since 2005 without some justification. In order to identify when this issue might arise, a two-step process was developed:

Step 1 - takes the 2008 GIS data set through the original model build data cleanup process, and compares the findings to the original model development data clean up results. This comparison determines whether or not the issues identified during the model development have been resolved. If not, it is then justifiable to revise the 2008 GIS data. As an added benefit, the process can identify anomalies which have entered the data set since model development.

Step 2 - uses the GIS edit log created by the original model builders as part of the model development. This log retains the original item attribute, as well as the revised item attribute and the reason for the revision. If the edit log identifies an item in the 2008 GIS, and the current attribute in question is the same as the original attribute, then it can safely be assumed that the attribute hasn't been updated. Hence, it will be recommended for update. Where there is a difference, a selection set will be created so that Metrowater can confirm or reject the revision. In cases where differences between the model data and current GIS are identified, but not documented in the edit log, only null value attributes will be updated.

5.4 DELIVERABLES

The deliverables for the data selection methodology section of this project were three tables which described in detail the process for: a) identifying the differences between the data sets, and b) to ensure that the confidence level of the superseding data is greater than that of the superseded. An example from each of the table is listed below in tables 2, 3 and 4.

Table 2 Model Build Data Clean-Up Techniques – and the Impact on the Existing 2008 GIS (Example 1 of 8 techniques)

Category	Asset Identification	GIS Update Approach	Implications on the Current GIS Database	GIS Update Route
1 Duplicated nodes/ hydrant / valves / main	**<u>Assets with Same Asset_Name</u>** Use MapInfo queries to check assets with duplication are the same for all data fields Export to Excel and check selection process **<u>Assets with Different Asset Name</u>** Used Mapbasic script to locate nodes in close proximity and parallel pipes Checked assets location to make sure that they are of the same type. e.g. 4 hydrants sitting on top of each other.	Visually check through selection to identify assets which fall into this category Remove from selection using Judgement Overlay with H2O Map model Data and query based on Asset ID to ensure that none of these assets were maintained in the model. Tag assets in this category as being excluded from the updated set. Repeat for assets with different names	May identify duplicated assists or identify any assets incorrectly flagged as Facility Status "OOS", "RMVD", "RERR", "RSS"	Create Mapinfo Table containing only data items created as part of duplication identification process that are to be removed/replaced. Items will be labelled appropriately.

Table 3 – Direct and Indirect Comparisons between the Data Set Developed as Part of the Model Build Process, and the current 2008 GIS (Example 1 of 6 techniques)

Category	Asset Identification	GIS Update Approach	Implications on the Current GIS Database	GIS Update Route
Compare pipe set attributes	Use Mapinfo Queries to link model pipe set to current GIS pipe set based on Asset ID. Identify a group of assets in both data sets which don't match, and develop a spatial query to determine a link. Visually inspect each special link to ensure that pairs are matched.	Create new pipe table, which contains only those pipes in both data sets For each comparable attrbute identifed in Table 2, update from both sets. Included original 2004 attribute where possible – this will enable determination of whether current item in 2008 GIS has been updated since 2004	Could identify updates for: • Description • ID • Length • Diameter • Roughness • From Node • To Node • Owner Material	Create Mapinfo Tables containing only data items which are recommended for update. An individual table will be created for each of the attributes identified in the adjacent column.

Table 4 – Using the model data to resolve previously identified issues in the 2008 GIS data

Category	Asset Identification	Compare Model with 2008 GIS Data Set	GIS Update Route
GIS Identified Issues	Use Mapinfo queries to identify current GIS items which have connectivity comments attached to them - pipe table query: *Notes="check connectivity to adjacent watermain" or Notes="check connectivity" or Notes="check connectivity to adjacent*	Save these items as separate table Overlay table with model data and each item is visually inspected in order to understand how the issues are resolved in the model	Create Mapinfo Table containing only data items which are recommended for update with revised notes entry

	main" or Notes="is this right connectivity?"		

6 KONINI WATER SUPPLY ZONE PIOLT STUDY

The methodologies set out in Tables 2, 3 and 4 (examples only) were used to assess the viability and value of extracting data from the Konini Water Supply Zone Model. Each of the 15 techniques was applied to the model data base and/or the GIS. The tables were then updated with the actual findings and presented as deliverables. These updated tables formed the backbone of the audit trail, describing in detail the analysis undertaken on the data and any assumptions used. The tables contain the names of the files produced for the GIS update, so any person involved in the update process can refer to the tables to understand the confidence in each data set.

The methodology above was overwhelmingly successful in identifying where value has been added to the Konini Supply Zone data set through the modelling process. It then goes further to identify anomalous new entries in the current GIS data set, which would have been revised or flagged had they been part of the data set used to develop the model.

In summary, the Pilot Study recommended 10,731 individual edits to the Konini GIS data set. These include attributes such as Pipe Diameter, Network Connectivity, Pipe Node Hydrant and Valve Facility Set Status, Pipe Material, Hydrant Elevation and Co-ordinates, and Duplicated Assets. The edits were prepared in a GIS format, consistent with Metrowater Ltd's protocols for the update of the GIS database.

Entering new data into the GIS database or revising existing data needs to be a manual process. This ensures that each change to the data set has been scrutinised by a person familiar with asset data. Furthermore, Metrowater operate a number of data entry scripts which ensure that proper procedures are maintained during data entry. While there are obvious benefits to these requirements, they necessitate a time intensive data update process. With this in mind, the value of each type of edit was discussed with Metrowater in detail. Edits were then prioritised based on the benefit to the data set, and also on their levels of confidence. A subset of high priority edits was developed and overall 576 changes to the GIS database were made.

7 SUMMARY

The Water Utility Data Verification Project was initiated to take full advantage of the GIS data improvement gained through the development and calibration of Water Distribution Models. An initial data assessment was undertaken to define a methodology by which the improved data could be 'uploaded' to the current GIS. The Konini Pilot Study was undertaken to determine the viability of the proposed methodology, and to quantify the value added to the current GIS.

The methodology developed during this study was overwhelmingly successful in identifying where value has been added to the Konini Supply Zone data set through the modelling process. Furthermore the study identified any anomalous new entries in the current GIS data set, which would have been revised or flagged had they been part of the data set used to develop the model.

In summary, the Pilot Study recommended 10,731 individual edits to the Konini GIS data set, which include attributes such as Pipe Diameter, Network Connectivity, Pipe Node Hydrant and Valve Facility Set Status, Pipe Material, Hydrant Elevation and Co-ordinates, and Duplicated Assets. The edits have been prepared in a GIS format, consistent with Metrowater Ltd's protocols for the update of the GIS database.

In conjunction with the methodology developed, a template was produced for the detailed documentation of the process by which the proposed changes are identified. This template could then be used to undertake the same process on other supply zones. The template contains justifications for the changes, and any associated assumptions that were used during the model build process, thus providing a transparent audit trail.