18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Water Quality Visualisation and Tracking – Generic Decision Support Tool

I. Ramsay, S. Shen and S. Tennakoon

eWater Cooperative Research Centre, Department of Environment and Resource Management (DERM)
No 80, Meiers Road, Botany Building, Indooroopilly, QLD 4068.
Email: ian.ramsay@epa.qld.gov.au

Abstract: There is increasing pressure on organisations to better utilise and improve access to their environmental data. For large organisations, environmental data is often collected and contained in databases (e.g. Oracle-based), and where resources are available, front-end software (e.g. written in Java or .Net) is developed to provide data access, processing or visualisation. One of the challenges is that developing and maintaining this software is expensive. This is compounded by the often changing needs of the organisations, in terms of what data has to be collected and how it is reported. The databases can usually accommodate additional datasets with limited effort, however, the front-end software is typically hard coded and more difficult to modify. Many organisations do not have the IT expertise or resources to make such changes. This is particularly the case for organisations dealing with water monitoring and management, whether in the area of water supply, water treatment, wastewater treatment or environment monitoring. Within the e-Water CRC, many partners have highlighted the need to improve their ability for data access and visualisation within their organisation.

This paper presents an outline of a new and novel data management software module that is being developed as part of the e-Water CRC in collaboration with the Queensland Department of Environment and Resource Management (formerly Environmental Protection Agency). The module will allow access to an organisation's corporate database with relative ease. A prototype module has been developed with universal data access capability that requires minimal configuration for data retrieval from almost any type of relational database. The module is generic in that it can be used for any time series data set (such as flow rate, water quality etc) and for multiple sites (within a stream or at a specific geographical point). It also features the ability to store information on objectives or guidelines to allow performance assessment. The module will include basic data management functions such as time series management, spatial selection and representation, data availability assessment and data series comparison (visual and statistical). The module is developed under Microsoft .NET Framework 3.5 and runs on most Microsoft Windows variations.

The prototype of the module will be presented in the paper including demonstration of how the module can be used.

Keyword: *Water quality, Data visualisation, Data Management, Decision support systems*

1. INTRODUCTION

In Australia, many organisations invest a substantial amount of time and money in environmental monitoring. Despite information technology advancements, monitoring data is often not readily accessible within or between organisations. Environmental monitoring data is typically stored electronically in one of a number of ways, from simple Microsoft Excel spreadsheets through to sophisticated relational databases. Regardless, manual processing of data and detailed reporting is time consuming and resource intensive and as a result occurs infrequently. For automated data processing, front-end software in combination with databases can be used to allow organisations to better access, process or visualise data. One of the challenges is that the data collection and reporting needs of organisations can be diverse and often changes. Databases can usually accommodate additional datasets with relatively limited effort, however, the front-end software is typically hard coded and expensive to modify. Many organisations do not have the resources to automate much of their data processing and reporting needs or have sufficient IT expertise to make changes to this type of software when the need arises.

An example of this was shown in a recent article by Loeffler (2008) on data warehousing for Sydney Water Corporation business information. It found that a variety of databases, systems architectures, data formats and reporting tools were used for the organisation as shown in Figure 1. It identified that information access, extraction, analysis and reporting capabilities were limited to a small group of specialist power users in the organisation. The reliance on these individuals was found to limit opportunities for ad-hoc reporting, detailed/in-depth analysis and access to transparent real-time information.

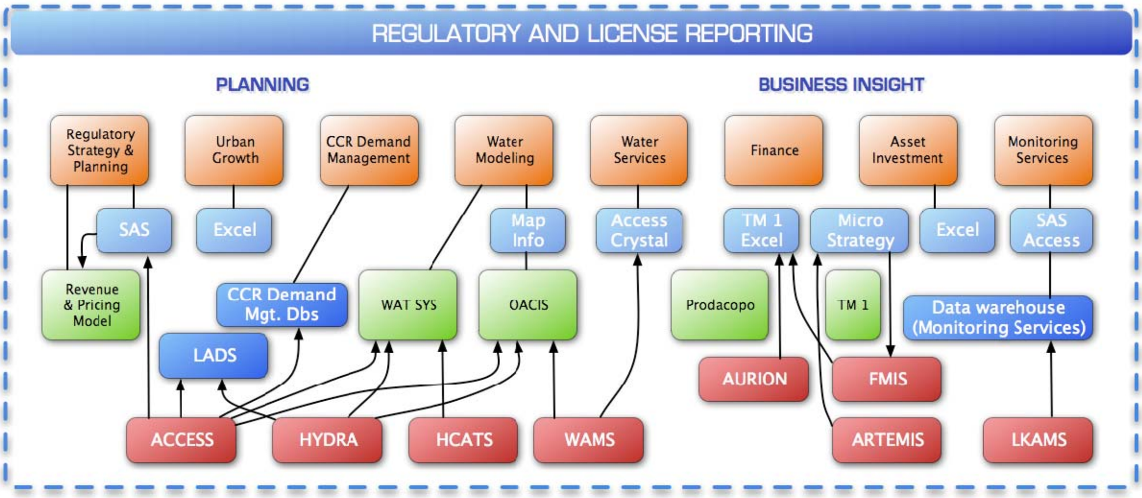

Figure 1. Sydney Water's reporting structure Competency Centre Project in 2005 before implementing Business Intelligence Solution. (reproduced from Loeffler, 2008)

Within the e-Water Cooperative Research Centre (eWater CRC), many partners have highlighted the need to improve data access, visualisation and reporting within their organisation. These partners are organisations dealing with monitoring data for areas such as water storage and regulation, water supply and treatment and waterway protection. A key focus of the CRC has been in development of software tools to assist these organisations with monitoring and assessment of water quality issues. The software package "Water Quality Analyser" (Tennakoon, 2008) has recently been developed for water quality data processing and assessment. This tool currently relies on manual importing of data from text files in a specific format which could limit its uptake by potential users.

One solution to enable software tools to interact with multiple data systems is to develop a "plug-in" module that requires minimal configuration for accessing data from a particular type of data source. The module could be built in a generic way so that any set of time series data can be uploaded and analysed (such as flow rate, water quality etc) and for multiple sites (within a stream or at specific geographical points) where common requirements for the data processing are involved. The module needs to incorporate a universal data access technique to be able to adapt to the diverse nature of data storage systems within organisations. Once this is achieved, simple assessment, visualisation and reporting functions will be incorporated into the software module. The module will then be incorporated in the Water Quality Analyser (and potentially other software) to allow for more specific data processing needs such as guideline setting, trend analysis or load assessment.

This paper describes the elements of such a “plug-in” data management module. The elements focus primarily on data importing and generic data storage structure. Potential data visualisation and assessment functions will also be presented. The development process will be briefly discussed.

2. DATA MANAGEMENT MODULE

2.1 Importing Data

The data management module needs to be flexible to adapt to user and organisational needs. The current options for data importation will include:

1. User entered information
2. Importing from any data source with user-defined SQL queries
3. Uploading from saved data

User entered information is the most time consuming and inefficient way that information can be entered into the data management module. However, this option needs to be available and may be used in some circumstances for small data sets.

Importing data from different data sources can usually result in completely different software approaches. In this tool, the OLE DB (Object Linking and Embedding for DataBases) technology is used for data access. OLE DB is designed to build on the success of ODBC (Open Database Connectivity Overview). ODBC provides an interface for accessing data in a mixed environment of relational and non- relational database management systems. ODBC provides an open, vendor- neutral way of accessing stored data and is a universal data access interface. OLE DB, in many respects, can be considered as a superset to ODBC. OLE DB includes all the functionality of ODBC with improvements on ease of deployments, efficiency, and capability of accessing multidimensional databases.

Data links to corporate databases are typically hard coded in software and only modified by expert programmers. The data management module will be designed in a generic way with generic variable structures based on common features of water quality data (see Data Storage & Formats). As a result no changes to the module structure will be required when implementing the system on a new database. The change will occur for the modules “plug” where the generic variables have to be defined in terms of the structure of the new database. Data will then be sourced from the database using saved Structured Query Language (SQL) queries. The importation is controlled by the user but an expert (an “electrician”) with some knowledge of the database will be required to assist with initially setting up or modifying these SQL queries. The effort involved will be extremely minimal compared to setting up new front-end database applications.

2.2 Data Storage & Format

The key to the data management module is the data structure and file format for storage. Table 1 describes the four types of data and how this will be stored by the module. This includes project information, defined SQL queries, series data and meta-data. Project information, series data and meta-data will be stored as separate XML files. Series data is where key numerical time series information is stored and XML files will include information on visual styles/annotations/exclusions/detection limits for use in Dundas Charts which is a charting control for .NET framework (Dundas Data Visualization, Inc, 2009). Defined SQL queries that are needed for retrieval of data from data sources will be stored on the application level and can be modified as required.

Series data and meta-data can be imported and exported separately in the software. Series data can also be imported and stored separately for different indicators or properties. Series data can include any water related property or indicator that is defined in the meta-data table and may also be separated based on data sources (e.g. in-situ monitoring, laboratory results, modelled data). The main generic meta-data schema can be grouped into:

1. *Series, Site and Water System tables.* These tables involve the key information about the time series, e.g. indicators (Table 2a), the system (Table 2b) and the data site to which the sites belong (table 2c). In addition to multiple indicators, multiple sites and systems can be included. All tables include unique

database identifiers, a name and additional meta-data information. Meta-data for a site mainly relate to spatial information such as latitude and longitudes. An additional column called NextSiteID is adopted to assist with riverine systems where upstream and downstream association is typically important.

2. *Water-Type related Objective tables.* These tables involve information related to water quality objective, water quality guidelines or targets that are used to assess data (Table 3b). In the example shown, the water–type relates to an ecosystem classification of a waterway (Table 3a), however, it could easily relate to a specific use such as human or industrial use. The water-type is associated with the site table and needs to be defined for each site if an objective or guideline assessment is to be used.
3. *Series Group tables.* These include the group name (Table 4a) and group series (Table 4b) tables and are particularly relevant where a large number of series are used and grouping of series is needed for presentation, mapping or reporting.

Table 1. Data Storage and Related Information

	Information Type	Storage Location and Description
1	Defined SQL queries	This is stored on application level for retrieval of data from data sources.
2	Project File	An XML file which stores list of series in <SeriesID, Filename> pairs as well as filename of the meta-data file.
3	Series Data	An XML file which stores serialised Dundas chart series with information on visual styles/annotations/exclusions/detection limits.
4	Meta-Data	An XML file which stores information on water systems/indicators/objectives.

Tables 2a, b and c. Sample of Series, Water System and Site tables (* indicates primary key)

a. Series Table		
ID*	Name	Unit
15001w	*Total Phosphorus*	*mg/L*
07081w	*Total Nitrogen*	*mg/L*
05000w	*Flow*	*m^3/day*
…		

b. Water System Table		
ID*	Name	Type
BNS1	*Sample River*	*River*
…		

c. Site Table						
ID*	Name	Latitude	Longitude	WaterSystemID	NextSiteID	WaterTypeID
xxx1	*X1*	*27.xxx*	*159.xxx*	*BNS1*	*xxx2*	*LC*
xxx2	*X2*	*27.xxx*	*159.xxx*	*BNS1*	*xxx3*	*LC*
xxx3	*X3*	*27.xxx*	*159.xxx*	*BNS1*		*UC*
…						

Tables 3a and b. Sample of Water-Type and Objective tables (* indicates primary key).

a. Water Type Table	
ID*	Name
UC	*Upper Catchment*
LC	*Lower Catchment*
…	

b. Objective Table					
SeriesID*	WaterType*	Start Date*	End Dtae*	Lower	Upper
15001w	*UC*	*01/01/2000*	*31/12/2003*		*12*
15001w	*LC*	*01/01/2000*	*31/12/2003*		*11*
07081w	*UC*	*01/01/2000*	*31/12/2003*		*35*
07081w	*LC*	*01/01/2004*			*38*
…					

Tables 4a and b. Sample of Group Name and Group Series tables (* indicates primary key)

a. Group Name Table	
ID*	Name
NUTR1	*Xxx nutrients*
NUTR2	*Yyyy nutrients*
…	

b. Group Series Table	
GroupID	SeriesID
NUTR1	*15001w*
NUTR1	*07081w*
NUTR2	*07081w*
…	

2.3 Data Validation and Data Export

Data validation functions includes features for checking and cleaning the time series data, such as outlier removal, detection limits handling and data points inclusion/exclusion with explanatory notes. The module will also have the capability of examining a time series and performing temporal aggregation, gap detection and gap filling.

All data and meta-data will be able to be exported in a number of formats. Formats will include csv files, Microsoft Excel spreadsheets or XML files.

3. DATA VISUALISATION AND ASSESSMENT TECHNIQUES

Data visualisation in current eWater Tools are discussed by Tennakoon (2008, 2009) and include trend assessment, guideline derivation and load predictions. In addition to these more complex functions, a number of basic data visualisation and assessment techniques can be incorporated as additional features to the data management module. Some of these techniques are relevant to any data assessment while others are specific to water quality assessment. The proposed techniques are discussed in more detail below.

3.1 Spatial Selection and Representation

A key requirement by many users assessing spatially distributed data is to have a map-based system to select, assess and potentially present water quality data. Various options exist to provide map-based navigation including Dundas Map, ThinkGeo Map Suite and many other third-party controls. The current module prototype utilises Dundas Map control to provide this feature and allows both quantitative and qualitative data to be visualised on a geographical map (see Figure 1 for examples). User interactivities on this map is also planned to simplify operations of the software.

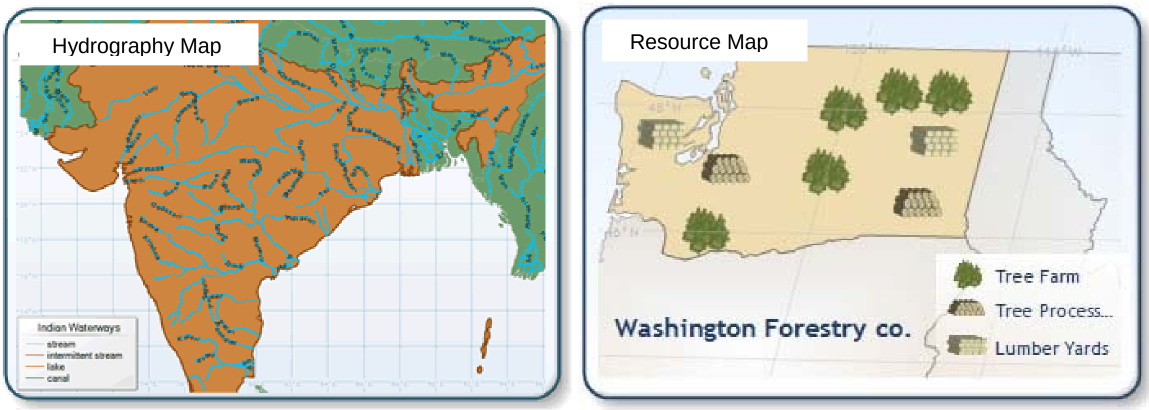

Figure 1. Two examples of Dundas Map control visualisation (Dundas Data Visualization, Inc, 2009)

3.2 Data Availability Assessment

The data management module can be used in combination with a data availability assessment as shown in Table 5 to determine the temporal and spatial availability of data for different series data (i.e. indicators). This assessment is useful for checking the status of newly collected monitoring data, for responding to data request or for new projects where data availability is not known. The availability assessment can be done for selected time periods, water systems, or series groups.

Table 5. Example of Data Availability Assessment for Sample River (Sites X1, X2 and X3) for six years. Value are number of data points available.

	Year					
Indicator	2002	2003	2004	2005	2006	2007
Total Phosphorus	0	0	36	36	36	0
Total Nitrogen	0	0	36	36	36	0
Electrical Conductivity	36	33	36	36	36	36
pH	36	33	36	36	36	36
Dissolved Oxygen	36	33	36	36	36	36
Flow	350	345	350	356	356	356

3.3 Condition Assessment

For series data that have associated guidelines or objectives, a condition assessment can be undertaken over the time period for which data is available. Different types of objectives could be included in the database including medians or percentiles assessed over defined time periods, most commonly twelve months. Table 2 shows objectives based on maximums (and minimums). A number of options for condition assessment will be available including time series comparison against objectives, summary tabulated statistics of compliance for sites or water systems and spatial presentation of compliance for selected indicators.

Various visual aids are available for statistical assessment. The following are on the development plan:

- Percentile control lines
- Smoothing curves (moving average, LOWESS algorithm, etc.)
- Frequency plot
- Whisker box plot

3.4 Series Data Comparison

The function of comparing multiple series data, both visually and statistically is extremely useful for many applications. One good example is for validating of computer model predictions against monitored data. Numerous software tools developed by the eWater CRC assist with modelling of particular water issue (see CRC Toolkit, http://www.toolkit.net.au). In addition to visual aids discussed above, statistical comparison using techniques such as least squares regression will be provided.

3.5 Automated Reporting

Automated reporting and exporting of information will be important to many users. Automated reporting will include user-selected charts, meta-data information and selected statistical results. Reports will be exported to Microsoft Excel or PDF format. Charts can be saved separately as pictures.

4. MODULE DEVELOPMENT

The data management module has been initially developed with a single time series import. Basic features tested in this stage include:

1. OLE DB data importing from Microsoft Excel, CSV, Oracle Database, Microsoft SQL Server as well as FoxPro (Hydstra's data store),
2. Basic data exporting,
3. Dundas charting with interactive data checking/cleaning,
4. Data storage structure efficiency and portability - Project file/Series Data/Meta-data XML files loading/parsing/saving, etc.

The data management module will then be extended to multiple time series. Functionality will be extended to include time series management (series aggregation, gap detection, gap filling), spatial representation, data availability assessment, condition assessment and data series comparison (both numerical and visual).

Once complete, the prototype will be tested on two or more different corporate databases for adaptability and performance testing. The module will then be incorporated into the Water Quality Analyser and tested in this software.

5. CONCLUSIONS

A data management software module is presented in this paper that is designed with universal data access capability to allow access to virtually any corporate relational database. This will be underpinned by the Object Linking and Embedding for Data Bases (OLE DB) technology and stored SQL queries that can be modified as required. The module is also being developed in a uniquely generic way in that it can be used for any set of time series data, designed in this case to assist water monitoring and management. Data and meta-data will be imported separately and stored in XML files. The data management module will assist with fundamental data management needs such as time series management, spatial selection and representation, data availability assessment, condition assessment and data series comparison. The module will be designed as a “plug-in” for the e-Water CRC’s Water Quality Analyser but also has potential broader applications with other data processing software.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support from the Queensland EPA and e-Water CRC.

REFERENCES

Dundas Data Visualization, Inc (2009), Figures sourced from Dundas Map online gallery: http://www.dundas.com/Gallery/Flash/Map/

Loeffler, U. (2008), Implementing a BI Competency Centre – a Sydney Water case study. IDC case study. Cited by “IDC - The life of your enterprise: how Sydney Water Corporation leveraged BI. http://www.strategicpath.asia/page/Business_Data_Management/Business_Intelligence/Data_Integration/Data_Integration/Definitions_Benefits__ROI/Independent_Insights/IDC_-_The_life_of_your_enterprise_how_Sydney_Water_Corporation_leveraged_BI/

Tennakoon, S. B. (2008), *Water Quality Analyser, Software User Guide*. eWater CRC, Innovation Centre, University of Canberra, Canberra.

Tennakoon, S. B., D. Robinson and S. Shen (2009), Decision Support System for Temporal Trend Assessment of Water Quality Data. 18th IMACS World Congress - MODSIM09 International Congress on Modelling and Simulation, 13 – 17 December 2009, ISBN: 978-0-9758400-7-8.