

## A new data management system for tracking licenced wastewater discharges in Queensland

C. Clech-Goods <sup>a</sup>, I. Ramsay <sup>a</sup>, N. Christiansen <sup>a</sup>, S (J). Shen <sup>a</sup> and C. Garnett <sup>a</sup>

<sup>a</sup>*Department of Environment and Resource Management (DERM), Queensland  
Email: [Celine.Clechgoods@derm.gov.qld.au](mailto:Celine.Clechgoods@derm.gov.qld.au)*

**Abstract:** In Queensland, the regulation and monitoring of environmentally relevant activities that release wastewater to the environment is strongly underpinned by data collected by the licensee rather than by the regulator. This is the case for metalliferous, coal and coal seam gas mines, sewage treatment plants, refineries and other industries throughout Queensland, many of which discharge to waterways that ultimately affect key areas such as the Great Barrier Reef, the Great Sandy Straits, Moreton Bay and the Murray Darling Basin. In addition to assessing the wastewater quality and quantity, many licensees are required to collect information on the environment into which they discharge. This may be a stream, an area of land or groundwater. In remote parts of Queensland, monitoring data collected by a licensee may be the only source of information available for that environment. Historically, this “third party” data was rarely available in an electronic format and generally not up-to-date, with limited information about how the monitoring was undertaken.

The importance of this data has led to the development of the Wastewater Tracking and Electronic Reporting System (WaTERS) - a third generation application and database developed by the Queensland Department of Environment and Resource Management (DERM). The system has been designed in a generic manner to facilitate the implementation of the system to a diverse set of clients. The system focuses primarily on capture, storage and processing of data. Clients submit data via the Internet using specific Microsoft Excel<sup>®</sup> templates that are created by WaTERS. A monitoring report is automatically generated on submission and compares the data against licence limits and regulatory requirements. This report is directly provided to both the client and the regulator. Raw monitoring data is available to internal officers and clients via separate applications and processes.

The system has been designed to collect monitoring data from any location within an activity site where wastewater is stored, treated, or discharged (end of pipe) and from the receiving environment. Data capturing is not only limited to water quality and quantity information but can also include more complex biological monitoring data such as aquatic invertebrate surveys. The reliability of the data is better known because WaTERS records specific provenance “metadata” information about the company and persons carrying out the monitoring, the sample collection methods, the instruments, and the associated laboratory including their analysis methods and accreditations. Furthermore, the system can collect additional data about the site infrastructure and operational processes. This provides a better understanding of the activity and water management across the whole site. The key challenge is to ensure that sufficient information is collected by the system to enable full and reliable use of the data while balancing the effort required by the client to collate information and prepare data submissions.

This paper discusses the design of WaTERS with specific focus on the system data and data management requirements.

**Keywords:** *data management, wastewater, water quality monitoring, metadata, third party data.*

## 1. INTRODUCTION

Water quality and aquatic ecosystem health monitoring is time consuming and expensive particularly in remote and regional areas. While the importance of long-term monitoring is recognised and required for condition and trend assessment, it is not feasible to fund monitoring programs for all rivers and streams in Australia. In Queensland, the regulation of Environmentally Relevant Activities under the Environmental Protection Regulation 2008 and monitoring of wastewater releases to the environment relies on data collected by the licensee rather than by the regulator. This is the case for metalliferous, coal and coal seam gas mines, sewage treatment plants, refineries and other industries throughout Queensland, many of which discharge to waterways linked to key conservation areas such as the Great Barrier Reef, the Great Sandy Straits, Moreton Bay and the Murray Darling Basin. In addition to assessing the wastewater quality and quantity, many licensees are required to monitor and collect information on the environment into which they discharge. This may be coastal waters, a stream, an area of land, or groundwater. Often this may be the only environmental information available for remote locations. The importance of these datasets has led to the development of a new data management system – a third generation application and database developed by the Queensland Department of Environment and Resource Management (DERM).

Prior to the development of the DERM data management systems, the licensees generally provided annual summary reports to DERM. Raw monitoring data was difficult to obtain in a suitable format or in a timely manner. To improve the accessibility of data and information on discharge monitoring, the first data management system called “Point Source Database” (PSD) was developed and implemented in 2006. PSD focused on collecting and storing raw monitoring data in an electronic format to support compliance and government decision making. Participating licensees, or clients, entered monitoring results into pre-prepared Microsoft Excel<sup>®</sup> templates that were uploaded to the database system via email. Upon receipt of a data submission, a summary report comparing results to licence limits was generated automatically by the system and emailed to the data submitter, database administrator and compliance officer. This system was innovative in that it allowed data to be uploaded from ‘third parties’ into the database, created standardised reports and instantly linked the data to graphing and load calculation software which are available to DERM officers. This system required a significant administrative effort to work with clients to diagnose problems with data upload and to audit the data received for quality assurance purposes. To help streamline the process, the system was improved in 2010 to allow clients to upload their data via a secure web interface. This system provided instant feedback on formatting errors and performed cursory quality assurance checks prior to allowing uploads. The web application had a number of other data management improvements such as allowing clients to automatically generate templates, view their data, create reports on demand and update their user details. However, this system was only designed to store discharge monitoring data and not monitoring data about the receiving environment.

The third generation data management system has been designed, developed and tested by DERM. This system is known as the “Wastewater Tracking and Electronic Reporting System” (WaTERS). It includes the features of the previous systems but also accommodates the capture and storage of monitoring data from any location within an activity site where wastewater is stored, treated or discharged. Another key enhancement is the inclusion of metadata regarding the provenance of the monitoring data including details of collecting organisations (e.g. environmental consulting company) and sample collecting and analysis methods. Finally, the system can store data and information about the site infrastructure and operational processes. This provides comprehensive information about the activity including the details of water management across the whole site. The key challenge is to ensure that sufficient information is collected by the system to enable full and reliable use of the data while balancing the effort required by the client to collate information and prepare data submissions.

This paper discusses the design of WaTERS with specific focus on the system data and data management requirements.

## 2. SYSTEM DATA

The data in WaTERS is a combination of regulatory information, monitoring results and metadata. The regulatory data, which includes information from the licence, is compared to the monitoring results to produce automated monitoring reports to support compliance interpretation. The monitoring results are typically discrete time series, and allow for the identification of changes, trends and potential exceedances of licence limits. The metadata at the dataset level focuses on the provenance of the data to provide the end users with information about the quality of the data and the source of further information if required. Both the regulatory and monitoring data were collected by the previous systems, while the new system provides a greater capacity to capture metadata both at a time series level and at a dataset level.

Reference information about the provenance of the monitoring data such as the collecting organisations, sample collection methods and laboratory details are provided by the client but uploaded and managed by the system administrators/power users (see Section 3.2 for user definition). The client has to establish the connections between the data and the reference information when configuring the site before submitting the data. The design is aimed at decreasing the data input effort by clients and administrators while ensuring adequate quality controls are in place for any data collected.

The following sections discuss the system data in more detail.

### **2.1. Regulatory Data**

Regulatory data is derived from official documents such as an Environmental Authority or other legal approval documents, but can also be derived from other sources such as monitoring plans or information provided by the licensee. This information describes the monitoring data context and compliance framework required by DERM. It includes:

- Client details: name, contact information, primary contact person,
- Compliance details: approval reference number, approval limits (end of pipe limits and investigation triggers),
- Site information: location (spatial coordinates, address, lots/plans),
- Monitoring details: location (spatial coordinates and streams), type of sample (e.g. water, sediment, etc), collection location (e.g. environmental water, discharge to water, stored water), indicator measured (e.g. pH, temperature).

Regulatory information is largely entered and updated by the system administrator – given this information is often of a legislative nature. Client user details, on the other hand, can be modified by the clients. There are processes in place to ensure the information stored is accurate and up-to-date (see Section 3.3).

### **2.2. Time Series Monitoring Data**

The time series monitoring data is collected and uploaded to the system by the clients. The majority of the time series data in the system is ongoing but the system can handle single data points such as a weather observation or presence/absence of an odour. The assessment of water quality and aquatic ecosystem health involves monitoring many types of indicators. This includes physico-chemical indicators in water and sediment, biological indicators (e.g. aquatic macro-invertebrates or fish surveys) and biota analysis. For certain data such as physico-chemical results, only basic metadata is required (time and monitoring location) to understand the result. However, for more complex data, additional metadata is necessary for a correct interpretation. For example, in the case of aquatic macroinvertebrates survey, it is necessary to record the individual hydraulic unit (riffle, pool or run) where the samples are collected; for heavy metal analyses in an organism, it is essential to record the type of tissue analysed. Capturing and storing a group of metadata for each result was a challenge to the design of the data management system. A hybrid relational and object database was identified as the most practical solution to store such results (see Section 3.1).

### **2.3. Data Set Information**

Some of the data stored by WaTERS is associated with a group of data or a data submission (or dataset). This information generally covers metadata about the time series data that has been collected: its provenance and the methodology used to collect data. As part of their licence conditions, clients are generally required to use sampling methodology consistent with the Department's Sampling Manual (DERM, 2009). Licences generally also include requirements, i.e. the sampling must be conducted by suitably qualified staff and the samples must be analysed by an accredited laboratory (e.g. NATA). Analysis methods should be accurate enough to detect the compound of interest to enable comparison with relevant guidelines, triggers or limits. This helps to assess the origins of the data which can significantly influence the quality of the results. The system also has the capacity to collect and store quality assurance and quality control information such as duplicates, blanks and calibrations data. WaTERS has been designed to collect this information to provide a basis for the data users to determine an uncertainty level of the data as well as to help check whether monitoring meets the requirements of the conditions. DERM is planning to develop a quality rating of the data based on this information.

## 2.4. Other Data & Information

The system has also been designed to improve the notification process whereby a client informs the regulation branch of DERM of water discharge events and of non-compliance events. Notification involves electronically submitting a completed form/document and associated monitoring data. Usually this is carried out at the initiation of a release or non-compliance event and at a latter time when the release has ceased or more monitoring data has been collected. These notifications were previously sent to DERM offices in a hard copy format and were manually tracked. WaTERS automatically coordinates this process by providing reminders and allowing searches of submitted information.

The system can also store data and information about a site's infrastructure including storage dams, and treatment processes such as the disinfection process used by the site. This data typically includes design specifications about the size, the location, the number of treatment units and even design diagrams. This information is useful for a better understanding of the level of water management/treatment at the site level and allows for benchmarking of performance between treatment types or between different sites.

## 3. DATA MANAGEMENT

### 3.1. Data capture and storage

The data management system was designed as a convenient way for a client to provide raw monitoring data in an electronic format to DERM. DERM deals with a variety of clients with varying data management skills and systems. Therefore, the system was designed with a user friendly web interface and the data capture process was simplified as much as possible. The system also includes email features to remind each client when data submission is due or overdue. Figure 1 describes the system workflow.

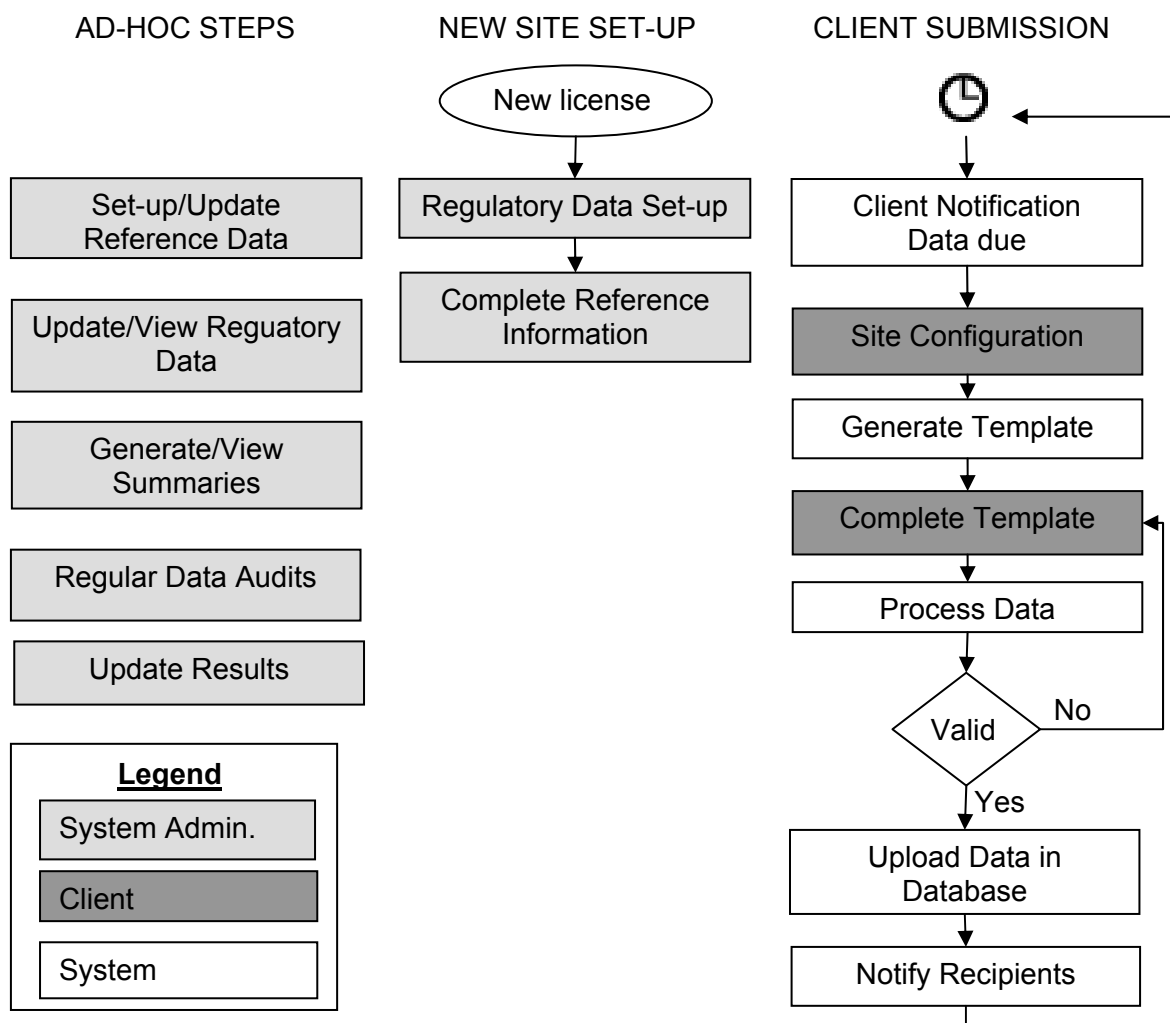


Figure 1. System workflow.

The client accesses the data management system via a secure web interface. Once the clients have completed their site configuration details, they download a template generated by the system specifically for their site. This template is a multi-spreadsheet Microsoft Excel<sup>®</sup> workbook that contains some prefilled regulatory information to allow the capture of time series monitoring data with associated provenance information. Each spreadsheet of the template contains a specific data type such as in-situ surface water quality, laboratory water quality, sediment quality, qualitative observations, aquatic macro-invertebrates survey, fish survey and coral and seagrass survey. This format, where different types of data are segregated into separate spreadsheets in a single template, has been selected because the data sources and formats in each spreadsheet are usually similar. Therefore, population and completion of the template should be easier. Many clients have their own data management system from which they could export the data. While others may cut and paste data from existing documents or adopt data processing software such as Water Quality Analyser (see Tennakoon *et al.*, 2011) to process their data before populating their templates. Once the dataset is prepared, the client submits the completed templates to WaTERS via the web interface. Data submissions are then tracked and audited by the system and the system administrator (see Section 3.3).

The data is stored in a backend database called “Integrated Water Quality Database” (IWQD) which operates on an Oracle 11g Release 2 Enterprise installation on a UNIX<sup>®</sup> server. The database is replicated into a backup server using Oracle “Streams” features which “hot mirrors” all data in the system (Oracle 2009). This will significantly reduce service downtime and increase data security.

IWQD has been designed and modelled to be flexible, generic and uses object-relational features of Oracle<sup>®</sup>. It can store and process a much greater range of information and new data types compared to the previous database. The database is also flexible in that it allows input data to be automatically uploaded in various units, i.e. is unit aware. Most calculations and other operations are directly developed in the database as functions and procedures to maximise the efficiency of numerical processing. Future modification and enhancement to the new database can be easily achieved by using object relational features such as type evolution (modifying existing object fields/methods) and inheritance (extending existing object hierarchy) (Oracle, 2010). Geographical information can also be stored and processed in IWQD with Oracle Locator (Oracle, 2011).

### 3.2. Data access and reporting

The system has four levels of users: system administrator, power user, DERM user, and data submitter (or client). Each of these levels has different authorities which govern the privileges to deal with data and functions. The system allows administrators and power users to access data for all sites in the system. Power users assist the system administrator and are the first point of contact for the clients but have read-only access to the regulatory information. Clients can view and search information only on their site configuration, their data submissions and their event notifications; they can also generate summaries. DERM officers can generate summaries and view data for all sites. (See Figure 2).

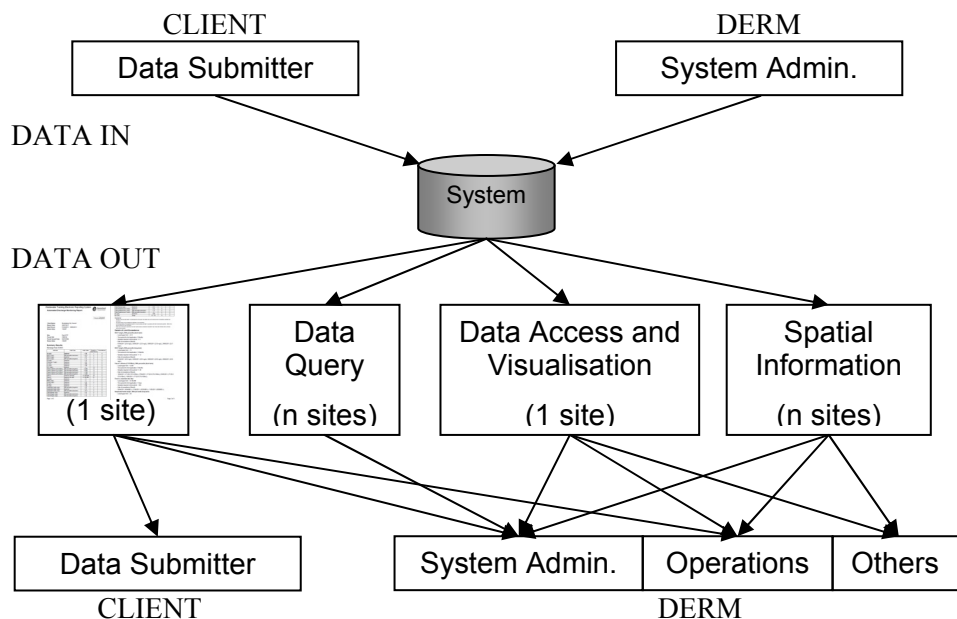


Figure 2. Data capture, access and reporting.

An important feature of the system is the ability to generate summaries. The system produces a site configuration summary automatically before the generation of a template, which lists for the site: the sampling locations, the indicators measured, the instruments, the sample collection methods, the collecting organisations, the laboratories and the analysis methods. The monitoring report provides a summary of what data has been submitted and compares the monitoring data to applicable limits or triggers stored in the system for that site. Any potential exceedances are highlighted and can guide further investigation. The report also indicates when there is insufficient data to undertake a limit check calculation such as a percentile or load value which requires a minimum number of samples in a specified period. Note that the frequency of monitoring data submitted is not checked automatically although information is provided on the number of data points submitted. Both summaries can be generated on demand by the users for a specific period or site.

WaTERS is part of a broader system including several tools to access and view the data. The interconnectivity of the system with other databases, tools and other softwares (e.g. Microsoft Excel<sup>®</sup>, Adobe Acrobat Reader<sup>®</sup>) makes the presentation and interrogation of the data very easy and quick.

In addition, DERM staff can access data stored in the database via two internal applications which provide easy access to licence limits, raw and interpreted monitoring data, as well as geographic coordinates and spatial maps. One is a .Net application that allows the user to: access raw data, which can be saved in various formats; visualise and compare the data to limits; and estimate pollutant loads. The other is mapping software where spatial layers containing regulatory and metadata from the system are presented and can be mapped against other departmental spatial layers such as waterways and catchments.

The data in the system is also used for a broader range of organisational reporting requirements within the Department. For example, the data is used to assist with State of Environment Reporting and Moreton Bay Healthy Waterways Report Cards. Further, the data available in WaTERS is likely to be used in the future for reporting purposes in the Great Barrier Reef Catchments. External organisations do not have direct access to information in the system about other organisations. However, this information is generally provided electronically via email on request. Data can be provided on a site, region or stream/basin basis.

### **3.3. Data maintenance and audits**

The regulatory information in the system is used to generate reports that could lead to compliance investigations. Therefore, it is important to keep this information up-to-date and accurate. Regulatory data is entered by a system administrator and audited by a second administrator. Project managers from the Operations area of DERM are responsible for notifying a system administrator of any new/modified licences involving water and sediment quality monitoring and of any amendments to existing licence limits. Some organisational processes are in place to ensure the system administrator is notified when a new or modified approval is signed. In addition, clients are encouraged to contact the system administrator if any information in the system is not correct. In such a case, the administrator checks the information with the Operations area of DERM and gets notification from them before updating the data.

A number of procedures and strategies have been included in the system to ensure that the monitoring data and metadata provided by the licensee is correct. The submitted data is checked automatically for formatting, obvious typographical errors and against the expected ranges of each indicator on submission. The metadata is checked against their site configuration. If the data complies, it is uploaded into the database. The system administrator checks the compliance summary report for obvious errors or missing data, and clients are encouraged to scrutinize their reports for potential errors as well. Finally, the system administrator graphs the data and checks for any visual abnormalities. If an abnormality is observed, the client is contacted and if an error has been verified then the data may need to be corrected by the data administrator or resubmitted by the client (the original submission is deleted). Finally, the administrator conducts regular audits on the data, for issues such as missing data, as part of database maintenance.

### **3.4. Data custodianship and licensing**

Most of the data collected by WaTERS is associated with approvals under the Queensland Environmental Protection Act 1994 and Environmental Protection Regulation 2008. When this is the case, the administering authority (i.e. DERM) is required to keep a public register and make the data publically available. As mentioned above, DERM supplies data to external parties on request. The requestor has to accept a licence agreement before the data is supplied. Note that a disclaimer is included in all data requests and reports from the system stating “that DERM cannot guarantee the quality of the data as it is collected by “third parties”. Ensuring the quality of the data is the sole responsibility of these parties”.

The system has the capability to accommodate the Government Information Licensing Framework (QGCI, 2010) and related licenses. The majority of the data in the system is likely to be classified as Creative Commons licences (type BY).

#### **4. DISCUSSION AND CONCLUSIONS**

WaTERS is a third generation software application that has been designed to expand the capacity of the organisation to collect information and data related to licences for wastewater releases to the environment. The system has the ability to capture a wide range of indicators, monitoring locations and related metadata at both the time series level and at a dataset level. The system is primarily designed for DERM officers to check the results of monitoring data for potential non-compliances and also check that the monitoring results are obtained as prescribed in the licence conditions. The system includes some design features and capabilities that will support the regulatory functions of the organisation (such as event reporting and information on infrastructure) and allow for ease of expansion and modifications in the future.

A key challenge in the design of WaTERS has been to ensure that adequate information can be collected by the system for an end user to check the data provenance, but also to ensure that the process of data submission by clients is not overly time-consuming. Achieving this balance generally meant that metadata has to be gathered at a dataset or data submission level rather than at a data point level. The success of the system will be tested in the future as it is implemented to clients.

The experience of DERM to-date is that the ability to store and manage data varies significantly between clients. Generally, DERM has had success persuading clients to submit their data electronically instead of providing hard copy reports. This is partly due to the perceived benefits of reducing the analysis and reporting effort, assisting with decision making and generally making better use of the data. With time, the system is likely to encourage clients to establish a good data management culture in their organisations given the larger amount of data that may need to be submitted to WaTERS.

Gathering information on monitoring metadata is an important first step to help define the quality of the data. It constitutes a good basis for the development of a quality rating system in the future. This could include requirements for training and auditing in regards to water quality monitoring that might be undertaken by DERM or other representatives.

Despite a focus on internal delivery of the data, the external delivery of data is possible even if not automated at this stage. Because WaTERS is part of a broader system, automated delivery on behalf of the client to organisations such as the Bureau of Meteorology is envisaged for the future. The incorporation of data access and graphing for clients would be an ideal area for future development.

#### **REFERENCES**

- Department of Environment and Resource Management. (2009) Monitoring and Sampling Manual 2009, Version 2. ISBN 978-0-9806986-1-9.
- Oracle. (2009). Oracle Streams Concepts and Administration, 11g Release 2 (11.2).
- Oracle. (2010). Oracle Spatial Developer's Guide, 11g Release 2 (11.2).
- Oracle. (2011). Oracle Database Object-Relational Developer's Guide 11g Release 2 (11.2).
- Queensland Government Chief Information Office. (2010). Government Information Licensing Framework v1.0.0. Queensland Government Enterprise Architecture Guideline – Shaping government ICT to support business outcomes.
- Tennakoon, S., I. Ramsay, S. Shen, and N. Christiansen. (2011). WQA: an integrated DSS and statistical package for water quality data management, processing and analysis. MODSIM Congress, Perth, WA, 12-19 December 2011, Conference Proceedings.