Observation Data Model (ODM) For Rincon Bayou, Nueces Delta

GRADUATE PROJECT

Submitted to the Faculty of
the Department of Computing Sciences
Texas A&M University-Corpus Christi
Corpus Christi, TX

in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Computer Science

by

Ramya Kurmana
Fall 2010

Committee Members

Dr. Long-zhuang Li
Committee Chairperson

Dr. Paul Montagna
Committee Member

Dr. David Thomas
Committee Member
Acknowledgement

Apart from the efforts of me, the success of any project depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the successful completion of this project.

I would like to show my greatest appreciation to Dr. Paul Montagna and Mr. Kevin Nelson. I extend my sincere thanks for their tremendous support and help. I feel motivated and encouraged every time I attend their meetings. Without the encouragement and guidance from them, this project would not have materialized.

The guidance and support received from the members Dr. Long-zhuang Li, Committee Chairperson and Dr. David Thomas, Committee Member, was vital for the success of the project. I am grateful for their constant support and help.
ABSTRACT

The restoration and enhancement project of Rincon Bayou and marsh in Nueces Delta is being carried out by The Coastal Bend Bays & Estuaries Program (CBBEP). A wide range of research and investigation is being carried out that includes an operation to determine the influence of diversions in restoring the marsh structure and functionality and further to determine better operational methods and explore environmental benefits. The Consortium of Universities for Advancement of Hydrologic Sciences (CUAHSI) is a consortium of 126 universities, colleges, and research institutions from the U.S. and around the world. The CUAHSI has developed a central repository called Observation Data Model (ODM), which is generic and robust enough to encapsulate observations from outside the hydrological paradigm, such as the biological data from Rincon Bayou. In spite of the robustness of the ODM, some limitations have been reported regarding its ability to accurately store the data. In this graduate project, some of the challenges of the ODM have been highlighted. In order to address these challenges, we develop an application that overcomes the limitations being highlighted as part of this project.
# TABLE OF CONTENTS

Acknowledgements .............................................................................................................. ii
Abstract ................................................................................................................................ iii
Table of Contents ....................................................................................................................... iv
List of Figures ........................................................................................................................... v

1. BACKGROUND .................................................................................................................. 1
   1.1. Rincon Bayou Project in Nueces Delta ........................................................................ 1
   1.2. Purpose ....................................................................................................................... 1
   1.3. CUAHSI ..................................................................................................................... 2
   1.4. Observations Data Model ......................................................................................... 3
   1.5. Data Integration ......................................................................................................... 6
       1.5.1. Web Services for Data Integration .................................................................. 6
   1.6. HydroExcel ............................................................................................................... 7
   1.7. Previous Works ....................................................................................................... 10

2. PROJECT RATIONALE .................................................................................................... 12
   2.1. Potential Users .......................................................................................................... 14

3. SYSTEM OVERVIEW ...................................................................................................... 15
   3.1. Software Required .................................................................................................... 15
   3.2. Design Overview ...................................................................................................... 16
   3.3. Functional Workflow ............................................................................................... 24

4. IMPLEMENTATION GUIDELINES ............................................................................... 32

5. TEST CASES & RESULTS ............................................................................................ 45

6. CONCLUSION AND FUTURE WORK ........................................................................... 66
BIBLIOGRAPHY AND REFERENCES .......................................................................................... 67

LIST OF FIGURES

Figure 1.1: Key components of CUAHSI-HIS ........................................................................ 2
Figure 1.2: ODM Schema ........................................................................................................ 4
Figure 1.3: Data Integration ...................................................................................................... 6
Figure 1.4: Retrieve data for the data service ......................................................................... 7
Figure 1.5: Web services ......................................................................................................... 8
Figure 1.6: Analyzing series catalog ...................................................................................... 8
Figure 1.7: Site summary ....................................................................................................... 9
Figure 1.8: The time series .................................................................................................... 10
Figure 3.1: Application home page ....................................................................................... 16
Figure 3.2: Water module form ............................................................................................ 17
Figure 3.3: Browse file window ............................................................................................. 18
Figure 3.4: Attributes to be imported ................................................................................... 18
Figure 3.5: Variables to be imported .................................................................................... 19
Figure 3.6: Add new site form ............................................................................................. 20
Figure 3.7: Add new variable form ....................................................................................... 21
Figure 3.8: Add new source form ........................................................................................ 22
Figure 3.9: Delete record form ............................................................................................. 23
Figure 3.10: Import Form workflow ...................................................................................... 26
Figure 3.11: Add New Variable Form workflow ..................................................................... 27
Figure 3.12: Add New Site Form workflow .......................................................................... 28
Figure 3.13: Add New Source Form workflow ..................................................................... 29
Figure 5.10: Add New Site Evaluation Test .................................................................52
Figure 5.11: Add New Site Validation Test .................................................................53
Figure 5.12: Add New Site Test Case Result ...............................................................54
Figure 5.13: Add New Site Test Case Result ...............................................................54
Figure 5.14: Add New Source Evaluation Test ..........................................................55
Figure 5.15: Add New Source Validation Test ...........................................................56
Figure 5.16: Add New Source Test Case Result .........................................................56
Figure 17: Retrieve Record Query Test & Result .......................................................57
Figure 5.18: Delete Record Data Option Evaluation Test .........................................58
Figure 5.19: Delete Record Data Option Validation Test .........................................59
Figure 5.20: Delete Record Data Option Evaluation Test .........................................59
Figure 5.21: Delete Record Data Option Test Case Result .......................................60
Figure 5.22: Delete Record SINCode Option Evaluation Test ...............................60
Figure 5.23: Delete Record SINCode Option Validation Test ...............................61
Figure 5.24: Delete Record SINCode Option Evaluation Test ...............................61
Figure 5.25: Delete Record SINCode Option Test Case Result ..............................62
Figure 5.26: Delete Record Data Filename Option Evaluation Test ......................62
Figure 5.27: Delete Record Data Filename Option Validation Test ......................63
Figure 5.28: Delete Record Data Filename Option Evaluation Test ......................63
Figure 5.29: Delete Record Data Filename Option Test Case Result ....................64
Figure 5.30: Delete Record Data Filename Option Test Case Result ....................64
Figure 5.31: Delete Record Test Case Result In HydroExcel .................................65
1. BACKGROUND

1.1. Rincon Bayou Project in Nueces Delta

In 1994, a restoration and enhancement project was started to divert fresh water into the Rincon Bayou and marsh in Nueces Delta. This project began with the construction of a channel in order to divert fresh water into the marsh. An investigation of the project was carried out in the Rincon Bayou area, which was initially funded by the U.S. Bureau of Reclamation and later by the City of Corpus Christi. The main idea behind these diversions was to restore and enhance marsh function and largely to implement and manage environmental flows from the Nueces River water system. After study and management for 14 years, the Nueces Delta experience presents guidelines for adaptive management of freshwater inflows in other parts of the Texas and also the United States in general [Montagna et al., 2009].

1.2. Purpose

To determine the influence of diversions in restoring the marsh structure and functionality and further to determine better water system operational methods, various types of data have been collected from different data sources located at four different institutions: the Center for Coastal Studies, Texas A&M University-Corpus Christi (TAMUCC); Conrad Blucher Institute, TAMUCC; Harte Research Institute for Gulf of Mexico Studies, TAMUCC; and The University of Texas Marine Science Institute [Montagna et al., 2009]. With a goal to integrate these diverse data sets into one system, the researchers in the weather forecasting community and the Consortium of Universities
for the Advancement of Hydrologic Sciences (CUAHSI) have developed the cyber-infrastructure to coordinate data from various sources and provide data-driven tools. This would allow for complex integrated analysis.

1.3. CUAHSI

CUAHSI stands for the Consortium of Universities for the Advancement of Hydrologic Sciences. CUAHSI’s Hydrologic Information System (CUAHSI-HIS) is a project carried out to provide access to the water information universally. It also aims to provide access to the data sources, models, and tools that facilitate creation, viewing and analysis of the activities of hydrologic systems. This project consists of three modules (Figure 1.1):

- a national cyber-information system for sharing hydrologic data,
- research into hydrologic information science, and
- support for the hydrologic information community [CUAHSI-HIS Website]

Figure 1.1: Key Components of CUAHSI-HIS

[CUAHSI-HIS Website]
1.4. Observation Data Model (ODM)

The Observations Data Model (ODM) is intended to provide a standard format for effective sharing of information between investigators and to allow analysis of information from disparate sources both within a single study area or hydrologic observatory and across hydrologic observatories and regions. The ODM is designed to store hydrologic observations with sufficient ancillary information (metadata) about the data values to provide traceable heritage from raw measurements to usable information allowing them to be unambiguously interpreted and used. Although designed specifically with hydrologic observation data in mind, this data model has a simple and general structure that will also accommodate a wide range of other data, such as from other environmental observatories or observing networks (Figure 1.2).

The data is currently stored in various research institutes each of them using different systems to store data and in different formats. It becomes difficult and time consuming to access data in different formats, locations and formats to analyze, forecast or generate reports for any decision making. The data available in the ODM is exposed by the research institutes by invoking Web services and transforming the data into a common format there by allowing easy access to perform any analysis and forecasting [Montagna 2009]. Web services are Application Programming Interfaces (API) which can be accessed through Hyper Text Transfer Protocol (HTTP). Web services are used to integrate different web based applications over an Internet protocol backbone using open standards like Extensible Markup Language (XML), Simple Object Access Protocol
(SOAP), Web Service Definition Language (WSDL) and Universal Description Discovery and Integration (UDDI).

Figure 1.2: ODM Schema [Tarboton et al. 2008]
The data sources for this project focused on Rincon Bayou in the Nueces River Delta and Nueces Estuary system in Corpus Christi, Texas. Researchers from the Center for Coastal Studies at Texas A&M University - Corpus Christi (CCS), Harte Research Institute for Gulf of Mexico Studies at Texas A&M University - Corpus Christi (HRI), and the University of Texas Marine Science Institute (UT) have collected environmental observations in the region since 1994. Each observation by every institution was made at a specific location. The participating groups, which provided data in MS Access databases, MS Excel spreadsheets, and SAS-based data files, employed a variety of data management methods.

In a report submitted to the Coastal Bend Bays & Estuaries Program (CBBEP), Montagna et al. have determined the robustness of the ODM to encapsulate observations from outside the hydrological paradigm. For this purpose, a self-imposed restriction on ad hoc augmentation of the schema was used, meaning the ODM was not modified. To load the data into the ODM, Extraction, Transformation and Loading (ETL) process was used. The CUAHSI Web services were used to publish the data. The report concludes that the ODM is reasonable robust, but there are some considerable limitations in the ability to store laboratory method metadata and replicate sample numbers. Other limitations are lack of a foreign key to the data systems outside the ODM and the inability to store the hierarchy of biological names. Currently, there is a need to address these limitations to make the ODM robust enough to encapsulate observations from outside the hydrological paradigm.
1.5. Data Integration

Data integration is the process of merging data present in different data sources to provide a combined view of the data to the users (Figure 1.3). Data integration is incorporated in scenarios where there is a need to retrieve data from distributed data sources, in order to show combined final data.

![Figure 1.3: Overview of Data Integration](http://www.euthenicsit.com/services/data_integration.php)

1.5.1. Web Service for Data Integration

Web services can be used as the prototypes for retrieving and combining relative data from diverse data sources and to give an instant access to the user/client. These Web services act as interfaces to access data from diverse data sources and provide a unified view of the data in reply to a query or a request made by a user. In order to find, retrieve and analyze the data in the ODM, we use a tool called HydroExcel which is an Excel spreadsheet that allows the user to directly access the WaterOneFlow Web services.
1.6. HydroExcel

The HydroExcel spreadsheet provides the user with the data from both the national data providers and the university sources. Here we use HydroExcel version 1.1.3 for Microsoft office 2007. The functionality of communicating with and retrieving the data from the Web service i.e. WaterOneFlow is done using VBA macros and an object library called HydroObjects. The HydroExcel spreadsheet consists of several worksheets which allow a user to work with different aspects of the Web service [Whiteaker 2010].

The primary task is to find and retrieve the information for a data service. For this, the ‘Data Source’ worksheet has to be activated (Figure 1.4). Here we instruct HydroExcel to connect with a Web service we would like to work with. This would activate the selected Web service. Once the selected web service is activated, all the worksheets of HydroExcel are loaded with the information of that particular web service. The Data Source worksheet (Figure 1.5) lists all the services associated with the selected web service and the also tells the user which worksheets in HydroExcel should work with the selected service [Whiteaker 2010].
After a Web service has been activated we need to build up Series Catalogs, which is a worksheet in the HydroExcel spreadsheet that provides information of a list of websites available in the Web service selected and activated earlier (Figure 1.6). The Series Catalog worksheet is used to retrieve information for every Website available in the Web service being currently used.
The Series Catalog downloaded previously needs to be analyzed. The contents of the Series Catalog can be analyzed using a pre-loaded pivot table in the HydroExcel spreadsheet. This pivot table analyzes the contents of the Series Catalog and produces a new table that shows the number of measurements of a particular variable at each of the sites in the series catalog. All this can be done in the Site Summary worksheet of the HydroExcel spreadsheet (Figure 1.7).

Figure 1.7: Site Summary

Until now, we have been using the HydroExcel spreadsheet to explore the information available in the Web service. The Time Series worksheet (Figure 1.8) allows the user to download a time series of a particular variable and provides information on real situations at the intended site [Whiteaker 2010].
1.7. Previous works

In 2007, Goodall et al. developed a machine accessible interface for the National Water Information System (NWIS). It is an online repository of historical and real-time stream flow, water quality and ground water level observations maintained by the United States geological Survey (USGS). This system serves a middle-layer abstraction between the NWIS database and hydrologic analysis systems, which allows substitution of NWIS server by such analysis systems for on-demand data access [Goodall et al. 2008].

An integrated system for publishing environmental observations data was developed by Horsburgh et al. in 2009. The authors have proposed a new method for publishing
research data that consists of point observations. This method uses a standard observations model that makes use of controlled vocabularies for environmental and water resources data along with Web services for transmitting data to consumers [Horsburgh et al. 2009].

CUAHSI has developed ODM tools such as ODDataLoader and Streaming Data Loader. These tools can be used for importing the observations data into the ODM. The ODDataLoader alleviates simple data loading into the ODM. The ODDataLoader does not have as steep of a learning curve, and is useful when the data do not have quirks that could only be handled with more sophisticated data loading means. On the other hand, the ODM Streaming Data Loader (SDL) application was created to allow administrators of local instances of the ODM to automate the process of loading their streaming sensor data from text files generated by their monitoring and telemetry system into an instance of the ODM [Horsburgh et al. 2007] [Whiteaker et al. 2007]

Even though these tools are useful in loading observations data to the ODM, they are not efficient enough to tackle the challenges of the ODM (Chapter 2). We need an application which provides an efficient platform for importing different hydrological observations into ODM and also handle some of the problems in the ODM. In order to accomplish this need, this project aims at developing an application which acts as an interface between the ODM and the user to perform different data transactions on the ODM. This application which provides the user with different features will be efficient enough tackle the challenges of the ODM.
Observation Data Model (ODM) is generic and robust enough to encapsulate observations from outside the hydrological paradigm, such as the water quality or biological data from Rincon Bayou. In spite of the robustness of the ODM, some limitations have been reported regarding its ability to maintain the data more precisely. As part of this project, we have highlighted some of the challenges of the ODM:

- Inability to handle replicate data samples.
- Lack of a foreign key to the data systems outside the ODM.

Inability to handle replicate data samples:

Typically, when hydrographic measurements are taken, values for water temperature, salinity, pH, and others are taken at the same time and depth. Logically, and practically, they are represented in a single record for each site, date, and depth. Indeed, all of the native data sources used in this synthesis employed this structure. Therefore, two records, each containing values taken at the same site, date, and depth can be considered replicates. After transformation into the ODM, the original record can be reconstructed by grouping the values by date, site, and depth. However, when there are replicate measurements this reconstruction is not possible without something denoting which replicate a data value belongs to. Facilities for easily denoting replicate associations are lacking [Montagna 2009].
Lack of a foreign key to data systems outside ODM:

Integrity of the data sources with the database schema is very a key aspect. In order to maintain this integrity, there should be some referential object specified to each of the data sources with respect to the database schema. This kind of referential object is lacking in ODM. In other words the ODM lacks a foreign key that refers to data systems outside ODM.

This project aims to address the challenges mentioned above, in order to increase the efficiency of ODM to maintain and provide integrated data for analysis on diversions restoring marsh structure and function. This is accomplished by developing an application which acts as an interface between the ODM and the user. This application will be able to handle the replicate data samples efficiently while importing the data files. While importing the data files to the ODM through this application, every data value being imported will be accommodated with a unique identifier code (See Section 3.3). This code is incorporated to help the users identify a data value and its replicates. This will address the issue of handling replicate samples in the ODM. This also helps in easily identifying replicate associations. Now that we have a unique identifier for every data value in the ODM and a reference to the data values (i.e., data file name), identifying the replicate associations will be an easy task. This application also addresses the issue of foreign key. For every data value imported into the ODM, its corresponding data file name is also imported (See Section 3.3). Using the relations between their corresponding tables, the data values can be referred back to their original sources. Here, the data file
name will act as the reference to the data values. This addresses the lack of foreign key in the ODM.

2.1. Potential Users

This application can be used by investigators and participating organisations involved in determining the influence of diversions in restoring the marsh structure and functionality. All the users who are authorized to involve in performing transactions on ODM can be considered as the intended users of this application. This application provides a graphical interface that is easy to use in importing the observation data into the ODM. The features of this application allow the intended users to perform different data transactions on the ODM. This application provides a more efficient interface between the users and the ODM, as it addresses some of the issues in the ODM.
3. SYSTEM OVERVIEW

3.1. Software Required

The application for the Observation Data Model (ODM) for Rincon Bayou, Nueces Delta is developed in an environment where different kinds of platforms were used for different purposes. This section lists out the software used in the development of the application.

- **MS SQL Server 2008**: Relational Database Management System for our project. Serves an efficient and feature-rich database system for the development and deployment of server-based applications.

- **MS Visual C# Express 2010**: Developing environment for our application. Provides an efficient and integrated platform for developing applications, testing and debugging.

- **HydroExcel Sheet – Version 1.1.3**: In order to find, retrieve and analyze the data in the ODM, we use a tool called HydroExcel which is an Excel spreadsheet that provides the user with the data from both the national data providers and the university sources.

- **WaterOneFlow Web Service**: provide programmatic access to a growing collection of national, state, and individual investigator hydrologic observation repositories.
3.2. Design Overview

The goal of this project is to develop an application to support the maintenance of data transactions on ODM for Rincon Bayou, Nueces Delta. This application was developed in Microsoft Visual C# Express 2010. The application is designed in such a way that it provides an easy and user friendly medium between the ODM and the user. This section gives an overview of the application design.

- The GUI of the application provides the options for different modules of ODM which include Atmosphere, Biota, Water Nutrients, Water and Sediment (Figure 3.1).
• The connections between the application and the ODM are established using SQL Server. Here, we incorporated Microsoft SQL Server 2008 to act as the data server between the application and ODM.

• Although the different modules related to the ODM are displayed on the main menu, this application only allows the use of Water module. So, observational data related to the Water module can only be imported to the ODM (Figure 3.2).

Figure 3.2: Water module form
• Users can use the application to import the data files into the ODM directly. We provide the user with different options while importing the data files. The user may browse and select the file to be imported (Figure 3.3).

![Figure 3.3: Browse file window](image)

• User may use the drop down menus and text boxes to specify the data entries that has to be imported along with data file (Figure 3.4, Figure 3.5).

![Figure 3.4: Attributes to be imported](image)
• The attributes to be specified by the user while importing a while include, *SiteID* - a unique identifier which refers to a particular sampling location; *SourceID* - a unique identifier which refers to a particular data source; *Time Difference* – the local offset time difference in hours from UTC; *Date Column* - ; *Time Column* - ; *Bay* – the bay area from where the observations have been taken; *Station* – the station in the bay area specified, where the observations have been taken [Tarboton 2008].

![Figure 3.5: Variables to be imported](image)

- *Date* – Date the values were recorded; *Time* – Time the values were recorded; *Temp* – the temperature variable to be imported; *Depth* - Water Depth; *DO Conc* – Dissolved Oxygen Concentration; *DO Charge* – Dissolved Oxygen Charge; *DO%* - Dissolved Oxygen saturation percentage; *SpCond* – Specific Conductance; *Salinity* – Dissolved Salt levels in water; *pH* - measure of the activity of dissolved hydrogen ions.
• Users may need to add any new entities to ODM, for example, adding a new variable, a new source or a new site. We provide the user with the options to add these new entities. The Figures 3.6, 3.7 and 3.8 represent the forms provided for the users to add a New Site, a New Column, and a New Source respectively.

![Add New Site Form](image)

Figure 3.6: Add new site form

• In the Add New Site form, the user has to specify the values including: Site Code – a unique code used by the organization collecting data, to represent a particular site; Site Name – name of the sampling site from where the organization collects the data; Latitude – latitude of the sampling site in decimal degrees; Longitude – longitude of the sampling site in decimal degrees (east +ve/ west –ve); LatLongDatumID - Identifier that references the Spatial Reference System of the latitude and longitude coordinates; Elevation_m – sampling location’s elevation in metres; VerticalDatum – vertical datum of the elevation; LocalX – X coordinate of local projection; LocalY – Y coordinate of local projection; LocalProjectionID - Identifier that references the Spatial Reference System of the local coordinates;
*PosAccuracy_m* – Positional Accuracy in metres; *State* – State the site is situated in; *County* – County the site belongs to; *Comments* – any comments regarding the new site that is being added [Tarboton 2008].

Figure 3.7: Add new column form

- The *Add New variable* form needs the user to specify the following values: *Variable Code* – a unique code to the variable, used by the organization collecting data; *Variable Name* – name of the new variable; *Speciation* – to identify how the data value is expressed; *VariableUnitsID* – a unique ID that represents the Units of the data value; *SampleMedium* – the medium of the sample used to collect the data; *ValueType* – describes the type of data value been added; *IsRegular* – indicates whether the data values for the new variable are from regularly sampled time series; *TimeSupport* – indicates the time support (or temporal footprint) of the data value; *TimeUnitsID* – a unique ID that represents the units in which the time support is measured; *DataType* – defines the type of data value been recorded;
GeneralCategory – indicates the category under which the variable is classified [Tarboton 2008].

![Add Source Form](image)

Figure 3.8: Add new source form

- When the user selects the option to *Add New Source*, a new form pops up (Figure 3.8). Here the user will have to provide information regarding the new source which include, *Organization* – name of the source of the data values; *SourceDescription* – describes the data source; *SourceLink* – link to the data source; *ContactName* – name of the person to be contacted at this source; *Phone* – contact telephone number; *Email* – electronic mail address; *Address* – postal address of the contact person; *City* – city of the contact person; *State* – state in which the contact person lives; *ZipCode* – zip code; *Citation* – reference to the data collected at the source; *MetadataID* – a unique ID that represents the record of this source in ISOMetadata table.
Figure 3.9: Delete record form

- This application also provides the user with an option where he/she can delete the data from the database just by a simple click (Figure 3.9). We provide the user with three deletion options. 1) Deleting an entire file from the database by providing the Data Filename; 2) Deleting a particular row of entries from the database by providing the SINCode, which is a unique code for that particular row; 3) Deleting the entries from a particular date, file name and containing a particular variable. The user has to provide the Data in the following format $Date_{[VariableID]}VariableName_Filename$, where the Date has to be provided in DD/MM/YYYY format; the VariableID would be the unique identifier of the variable entries to be deleted and its corresponding VariableName; Filename would be the name of the file, where the entries to be deleted belong to.
3.3. Functional Workflow

The main functionality of the application is to support the maintenance of observations from outside the hydrological paradigm in the ODM. This application serves as an efficient interface for data transactions with the ODM. As mentioned earlier in Chapter 1, this application provides data transactions related to Water module only. The data file types to be used to import the data would be restricted to Microsoft Excel files (*.xls, *.xlsx, *.csv, *.cdf) only.

Activities or data transactions that are allowed on the ODM include: Importing a data file to the ODM; Adding a new variable into the ODM; Adding a new site into the ODM; Adding a new source into the ODM; Deleting records from the ODM. Different forms have been generated to carry the activities mentioned above. Of these, the main form where all the forms would be generated from is the import form, which is used to import the MS Excel data files to the ODM. The data transferred using this form would directly affect the DataValues table in the ODM. All the data from the Excel file selected and the data provided on the form will be inserted into the DataValues table, corresponding to the format the ODM is designed.

The selected variables would be imported from the Excel sheet into the ODM, where these variable values are inserted one below the other in the same columns; and of course the other data values are updated simultaneously. The name of the data file being imported would be stored in the GroupDescriptions table. The GroupDescriptions table and the DataValues table have their primary keys GroupID and ValueID respectively in
the Groups table. This would keep a relation of the Filename with its data in the 
DataValues table. Each data value row updated in the DataValues table will be given a 
unique ValueID which is auto-incremental. Here, the data Filename inserted into the 
ODM acts as a reference to the entries in the DataValues table. For example, consider a 
set of entries corresponding to a given Variable. A user likes to know the original file 
from where these entries have been imported. In that case, a user can trace back the 
entries in the DataValues to its original Data Filename by using the relation between the 
DataValues and GroupDescriptions tables. This would help the user to find the source of 
the set of entries considered.

At the same time, another unique code representing every single row will be recorded in 
the Samples table under the LabSampleCode column while importing the data file. This 
unique code is called as SINCode (Sample Identification Name Code), would be recorded 
in the format:

\[
WQ\_YYYYMMDD\_HH\:MM\_Bay\_Station\_VariableID\_RepNum
\]

Where \(WQ\) is a standard identifier; \(YYYYMMDD\_HH\:MM\) is date and time when the data 
value was recorded; \(Bay\) at which the variable is recorded; \(Station\) where the 
observations were made; \(VariableID\) is a unique identifier for the corresponding 
variable; \(RepNum\) is a counter for the number of replications for the corresponding 
variable.

During the import of data from files to the DB, we have integrated a SQL query which 
would check for the replications of each data set being imported (Figure 3.10). The data
sets would be compared with the existing entries with reference to the Site, Date, Depth and a Unique ID (which would be the file name being imported). Consider a data value X that already exists in the ODM, and it has been allocated a SINCcode as follows:

\[ \text{WQ}_20020202 \ 20:20\_\text{Rep}1 \]

Now consider that while importing a new data file to the ODM, a data value Y is found to be a replicate of data value X. This replicate sample is given a SINCcode as follows:

\[ \text{WQ}_20020202 \ 20:20\_\text{Rep}2 \]

Users can now identify the Y as a replicate sample of X. The user can also identify the number of replicate samples present for a data value. Thus, the problem of handling the replications of datasets in the Database is solved with the help of this application.

![Diagram](image)

Figure 3.10: Import Form workflow

While importing the data files to the ODM, if the users need to add some new entity members which are not available in the menu options provided, the applications allows addition of new entity members which would be reflected in the ODM. When the users
add entity members to the ODM, for example, adding a new variable or a new site; we give them the options as: ‘Add New Column’ – to add new variable desired; for the Site ID, from the drop down menu: ‘Add New Site’ – to add a new site desired; for the Source ID, from the drop down menu: ‘Add New Source’ – to add a new source desired

In order to add a new column that has to be imported to the ODM; the user has to click on the Add New Column button on the import form (Figure 3.11). This would launch a pop-up form, where the user has to provide the details regarding the new variable that has to be added in the form and imported to the ODM. The Variables table will be updated with the information regarding the new variable.

When the user has to add a new site to the ODM, he/she has to select the Add New Site option from the drop down menu provided for the Site ID (Figure 3.12). Then user has to key in the details of the new site and submit the form. When the data file is imported to the ODM, the values would be inserted into the DataValues table, where the SiteID
column would be filled with a unique ID generated for the new site added. This unique ID is generated in the Site table which is first updated immediately after submitting the Add New Site form.

Every file and the data in it imported to the ODM have been provided by a data source. While importing a data file to the ODM through the application, users have to select the data source from the drop down list provided. If they could find their source in the list, they would be able to add that source there and then proceed with importing the data file. When the user selects the Add new Source option, a new form pops-up and the user will have to provide the details of the new source he/she desires to add (Figure 3.13). Once the source form is submitted, the details of the new source would be reflected immediately in the Source table in the ODM. Later when the data file is imported with the new source, the SourceID column in the DataValues table will filled with a unique and dynamically incremented value that represents the source from where the data has
been collected. This unique value of the source is the SourceID from the Sources table, which is updated immediately after the submission of add new source form.

Figure 3.13: Add New Source Form workflow

There may be some data in the ODM database which the user thinks has been entered incorrectly; or the user may think it is not needed in the database. In such conditions, the user may like to delete that particular data which is incorrect or unwanted (Figure 3.14). The user is provided with such an option to delete the data securely from the database, without affecting the rest of the data. The user would have different conditions on deleting the data from the database. Considering such conditions, we provide the user with three different options for deleting the data. The first option would allow the deletion of an entire data file from the database based on the data file name. The second option allows the user to delete a particular row from the database table based on a unique code provided for that row i.e., SINCode. The third option allows deletion of a particular set of data from the database based on the date, variable Id, and data file name,
which should be given in the format: DD/MM/YYYY_VariableID_Filename. The deletion activities directly affect the DataValues table and also any other tables directly associated with it and the deletion activity.

Figure 3.14: Delete Record Form workflow

Along with features of the application, we provide a query which is designed to retrieve records from the ODM when a File name is given. This feature is not a part of the application we have designed, but an additional feature just for the convenience of the users. It is a SQL Group Query, which accepts a data file name and displays the data values and its corresponding attributes with respect to the given data file name. This Query joins different tables with the given data file name as a reference to give a retrieve and give a combined display of its data values and their corresponding data values. The SQL Query designed for the data retrieval goes as follows:
SELECT
    a.SiteID, c.SiteName, b.VariableName, Bay, Station, 
    DataValue, LocalDateTime, a.VariableID, UTCOffset, 
    DateTimeUTC, CensorCode, MethodID, SourceID, 
    QualityControlLevelID
FROM dbo.DataValues a
JOIN Variables b ON a.VariableID = b.VariableID
JOIN sites c ON c.SiteID = a.SiteID
INNER JOIN Groups G ON g.valueid = a.valueid
INNER JOIN GroupDescriptions gd ON g.GroupID = gd.GroupID
AND
    gd.GroupDescription = 'FILENAME' -- GIVE THE FILE NAME HERE
GROUP BY
    a.SiteID, c.SiteName, b.VariableName, Bay, Station, 
    DataValue, LocalDateTime, a.VariableID, UTCOffset, 
    DateTimeUTC, CensorCode, MethodID, SourceID, 
    QualityControlLevelID
4. IMPLEMENTATION GUIDELINES

This section guides a user through step by step implementation of different features of the application.

Launching the ODM Application

Step by step procedure for launching the application.

**Step 1:** The user has to launch the Visual Studio 2010 and opens the application named ODM.

**Step 2:** Open the project and the application loads, build the application project.

**Step 3:** Once the building is successful, then debug the application project.

**Step 4:** The application launches a new window with a GUI (Figure 4.1).

![Application home page](image_url)

Figure 4.1: Application home page
Importing Water data to the ODM

Step by step guidelines to import Water data to the ODM.

**Step 1:** Select the Water module (Figure 4.2)

![Figure 4.2: Selecting the water module](image)
Step 2: In the Water module screen, the user has to browse and select the file to be imported to the ODM (Figure 4.3).

Figure 4.3: Browsing the data file
Step 3: User enters the work sheet number to be imported from the excel data file (Figure 4.4).

![Figure 4.4: Enter the sheet number](image)

Step 4: Once the user selects a file to be imported, he/she now has to select the SiteID and SourceID using the drop down menu provided (Figure 4.5).

Step 5: User has to enter the details of Bay and Station into the text boxes provided (Figure 4.5).
Step 6: Now the user has to select the variable data values to be imported into the ODM from the file selected. User can use the drop down menus provided in the form (Figure 4.5).

![Figure 4.5: Provide values for all the fields](image)

Step 7: Once the variables are selected, finally the user would click on ‘Import Data’ to submit the form (Figure 4.5).
Adding a New Site to the ODM

Step by step implementation to add a new site into the ODM.

**Step 1:** In the Water Module import form, the user has to set the SiteID. If the user did not find the desired option in the given drop down menu, he/she has to select the option *Add New Site* in the menu (Figure 4.6).
**Step 2:** A new form pops-up and the user have to key in the details of the new site to be added. Click on *Save* to submit the new site (Figure 4.7).

![Image of new form with details to enter]

Figure 4.7: Enter the details of new site
Adding a New Variable Column to the ODM

Guidelines for adding a new variable column

**Step 1:** While importing a data file to the ODM, the user may add new variable entity into it. In order to add a new variable the user has to click *Add New Column* in the form (Figure 4.8).

![Figure 4.8: Click on Add New Variable Column](image-url)
**Step 2:** On the click of *Add New Variable* button, a new form pops-up. Here the user have to key in the details of the new variable he/she wants to add, and click on *Save* to submit the new variable (Figure 4.9).

![Figure 4.9: Enter the details of the new variable](image-url)
Adding a New Source to the ODM

Step by step procedure to add a new source to the ODM.

**Step 1:** While importing a data file to the ODM, the user may add a new source entity into it. In order to add a new source the user has to select the *Add New Source* option from the drop down menu for SourceID in the form (Figure 4.10).

Figure 4.10: Select Add New Source
Step 2: A new form pops-up and the user have to key in the details of the new source to be added. Click on Save to submit the new source (Figure 4.11).

![Image of new form]

Figure 4.11: Enter the details of the new source
Deleting a Record

Guidelines on how to use the Delete Record feature in this application.

**Step 1:** This provides the user with options to delete the data from the ODM database.

The user has to select the *Delete Record* option from the home menu of the application (Figure 4.12).

![Select Delete Record](image)

**Figure 4.12: Select Delete Record**

**Step 2:** Selecting the *Delete Record* option opens a new form with different options for the user to delete the unwanted or incorrect data. The user may choose any of the three
options (Figure 4.13).

- Provide a *Data Filename* to delete an entire file from the database (or)
- Provide a *SINCode* to delete a particular row of entries from the database (or)
- Provide the *Data* in the given format to delete entries from a particular date, file name and containing a particular variable.

![Figure 4.13: Choose the deleting option](image)

**Step 3:** Click the *Delete* button corresponding to the option selected to delete the data.
5. TEST CASES AND RESULTS

Reliability of the applications has always been a major factor in Software Engineering. When a new software or application is developed, it is very important to evaluate whether it actually accomplishes the purpose it has been built for. In order to evaluate an application, its implementation and performance has to be tested. Testing an application also involves performing validation tests on it. For example, if the user misses entering any important information while performing data transactions, an error should be posed cautioning the user. This application is tested, evaluated and validated based on some test cases which represent the implementation and performance of the application.

Test Case 1: Evaluation and Validation of Import Data Form

The test case is to check if our application import a data file into the ODM successfully.

- Run the application to import a data file (Figure 5.1).

![Figure 5.1: Import Data Evaluation Test](image-url)
• If the user misses to enter any of the required fields and clicks *Import Data*, an error message pops-up asking the user to fill in the missing fields. The form not submitted unless all the fields are filled with valid entries. For example if the user fills in all the fields, but forgets to select the Site ID and tries to submit the form. Then, an error message ‘Select SiteID’ is displayed (Figure 5.2).

![Figure 5.2: Import Data Validation Test](image-url)
When all the fields are filled in with their specific entries, the submission of the import form would be accepted and the data file would be imported to the ODM (Figure 5.3).

Result for Test Case 1:

- To check the result, the user has to manually run a SQL Query (Figure 5.4) and look for the imported values in the result pane (provided the user has access SQL).
Figure 5.4: Import Data Test Case Result

- Alternatively, the result of the above test case can be viewed in HydroExcel.
- Launch the HydroExcel sheet and check if the data values have been uploaded.
- The HydroExcel sheet shows the data values from the file that has been uploaded (highlighting is done manually); hence the data file has been imported to ODM successfully (Figure 5.5).

Figure 5.5: Import Data Test Case Result
If a user feels the entries imported into the ODM are not correct, one can delete those entries using the *Delete Record* option provided in the application.

**Test Case 2: Evaluation and Validation Test for Add New Column Form**

This test case is to check the functionality Add New Column feature in the application.

- While importing a data file to the ODM, user wants to add a new variable column entity to the ODM. Clicks on the *Add New Column* button and a new form pops-up.
- User enters the details of the new variable column entity to be added into the ODM and clicks *Save* (Figure 5.6).
- User has to follow the naming standards according to the Control Vocabulary table in the ODM while giving a unique code to the variable.

![Figure 5.6: Add New Column Evaluation Test](image)
• If the user misses to enter any of the required fields and clicks Save, an error message pops-up asking the user to fill in the missing fields. The form is not submitted unless all the fields are filled with valid entries. For example if the user fills in all the fields, but forgets to select the Data Type and tries to submit the form. Then, an error message ‘Select a Data Type’ is displayed (Figure 5.7).

![Add New Variable Validation Test](image)

Figure 5.7: Add New Variable Validation Test

• User has to fill in the required fields and then submit the form. Only then, the new variable entity is added into the ODM.

**Result for Test Case 2:**

• To check the result, a SQL Query is run manually (Figure 5.8). The result of this test case is manually highlighted for visibility.
Figure 5.8: Add New Variable Test Case Result

- Also, launch the HydroExcel sheet to view the new variable been added in the Variables worksheet. Result of the test case is highlighted (Figure 5.9).

Figure 5.9: Add New Variable Test Case Result

Once a new variable entity is added into the ODM, it cannot be removed immediately using the same application. The data values corresponding to a given variable can be deleted using the *Delete Record* option, but the variable entity itself cannot be removed using this application. If the user wishes to do so, one has to contact the Database Administrator for assistance.
Test Case 3: Evaluation and Validation test for Add New Site Form

This test case is to check the functionality Add New Site feature in the application.

- While importing a data file to the ODM, user wants to add a new site into the ODM. Selects the *Add New Site* option from the drop down menu for SiteID, and a new form pops-up (Figure 5.10).

- Users should make sure they give a unique SiteCode and SiteName to avoid confusion due to repeated Site Names. No check constraints are applied for these fields, as to check for the uniqueness.

![Figure 5.10: Add New Site Evaluation Test](image)

52
• User enters the details of the new site to be added into the ODM and clicks Save.

• If the user enters partial information and clicks Save, an error message pops-up asking the user to fill in the missing fields. The form is not submitted unless all the fields are filled with valid entries. For example if the user fills in a few fields, but neglects the rest of the fields and tries to submit the form. Then, an error message ‘Enter all Information’ is displayed (Figure 5.11).

![Add New Site Validation Test](image)

Figure 5.11: Add New Site Validation Test

**Result for Test Case 3:**

• To check the result, an SQL query is run manually. The result of the test case is highlighted (Figure 5.12).
• Launch the HydroExcel sheet to check the Sites worksheet for the new site added.

The result of the test case is highlighted (Figure 5.13).

Once a new site entity is added into the ODM, it cannot be removed directly using this application. The user has to contact the Database Administrator for assistance.
Test Case 4: Evaluation and Validation Test for Add New Source Form

This test case is to check the functionality of the Add New Source feature in the application.

- Select the *Add New Source* option from the drop down menu for the Source ID. A new form pops-up, then key in the details of the new source and click *Save* (Figure 5.14).

![Add New Source Form](image)

Figure 5.14: Add New Source Evaluation Test

- When submitting the form, if the user enters all the information but forgets to fill in a field, an error message pops-up asking the user to fill in the missing fields. The form is not submitted unless all the fields are filled with valid entries. For example if
the user fills in all the fields, but forgets the Metadata ID field and tries to submit the form. Then, an error message ‘Select a Metadata ID’ is displayed (Figure 5.15).

![Add New Source Validation Test](image)

**Figure 5.15: Add New Source Validation Test**

**Result for Test Case 4:**

- To check whether the new source has been added, check the Source table in the ODM by an SQL query manually. The result of the test case is highlighted (Figure 5.16).

```sql
SELECT * FROM Sources
```

![Add New Source Test Case Result](image)

**Figure 5.16: Add New Source Test Case Result**
Once a new source entity is added into the ODM, it cannot be removed directly using this application. The user has to contact the Database Administrator for assistance.

**Test Case 5: Retrieve Record with File Name**

Run the SQL Query designed for retrieval of records by giving a valid File Name (Figure 5.17). And the result of the query is displayed in the Results pane (highlighted for visibility).

**Result for Test Case 5:**

![Query and Result](image)

*Figure 5.17: Retrieve Record Query Test & Result*

If the Filename given does not exist, no records are displayed in the results pane.
Test Case 6: Evaluation and Validation Test for Delete Record Form

This test case to check the functionality of the three different deletion features available in the application.

Test Case 6.1: Deleting records by giving Data in the following format (Figure 5.18):

\[ \text{DD/MM/YYYY}_{[\text{VariableID}]}\text{VariableName}_\text{Filename} \]

![Figure 5.18: Delete Record Data Option Evaluation Test](image)

- If the user gives any invalid entry, for example if there is no record corresponding to the given entry. Then a error message pops-up saying ‘No Records are Found’ (Figure 5.19).
• If the details entered by the user is valid, then the records would be deleted and a message pops-up saying ‘Records Deleted Successfully’ (Figure 5.20).
Result for Test Case 6.1:

- To check the result, run a SQL Query (Figure 5.21). The results pane shows no data as all the data related to the query have been deleted by the Data option.

![Query executed successfully.](image)

Figure 5.21: Delete Record Data Option Test Case Result

Test Case 6.2: Deleting records using SINCode which has to be given in the following format (Figure 5.22):

$$WQ_{YYYYMMDD} \text{ HH:MM} \_ \text{Bay}_\text{Station}_\text{VariableID}_\text{RepNum}$$

![Delete Record SINCode Option Evaluation Test](image)
• If the user gives any invalid entry, for example if there is no record corresponding to the given entry. Then a error message pops-up saying ‘No Records are Found’ (Figure 5.3).

![Figure 5.23: Delete Record SINCode Option Validation Test](image)

• If the details entered by the user is valid, then the records would be deleted and a message pops-up saying ‘Records Deleted Successfully’ (Figure 5.24).

![Figure 5.24: Delete Record SINCode Option Evaluation Test](image)
Result for Test Case 6.2:

- To check the result, run a SQL Query (Figure 5.25). The results pane is empty as all the data related to the query has been deleted by using $SINCode$ option.

![Figure 5.25: Delete Record $SINCode$ Option Test Case Result](image)

Test Case 6.3: Deleting an entire file from the ODM. Give the $Data Filename$ to be deleted (Figure 5.26).

![Figure 5.26: Delete Record $Data Filename$ Option Evaluation Test](image)
• If the user gives any invalid entry, for example if there is no record corresponding to the given entry. Then a error message pops-up saying ‘No Records are Found’ (Figure 5.27).

Figure 5.27: Delete Record Data Filename Option Validation Test

• If the details entered by the user is valid, then the records would be deleted and a message pops-up saying ‘Records Deleted Successfully’ (Figure 5.28).

Figure 5.28: Delete Record Data Filename Option Evaluation Test
Result for Test Case 6.3:

- To check the result, run the SQL Query for data retrieval using the *Data Filename*.

  The results pane is empty as all the data file has been deleted (Figure 2.9).

![Figure 5.29: Delete Record Data Filename Option Test Case Result](image)

- Also check the GroupDescriptions table, where the *Data Filename* is not displayed as it is deleted (Figure 5.30).

![Figure 5.30: Delete Record Data Filename Option Test Case Result](image)
In all the delete cases, the user can alternatively check whether the records have been deleted or not. This is by checking the Value Count column in the Series catalog worksheet of the HydroExcel sheet (Figure 5.31). If the Value Count (highlighted) has decreased to the previous, then the records are deleted successfully.

![Figure 5.31: Delete Record Test Case Result In HydroExcel](image)

In all delete cases, users have to note that records once deleted using the application cannot be retrieved back using the same. There is no feature available in the application, to retrieve the deleted records. If the user wants to retrieve the deleted records, one has to immediately contact the Database Administrator for assistance.
6. CONCLUSION AND FUTURE WORK

This application serves as an efficient interface between the user and the ODM and can be used to perform different data transactions. Using this application, importing Water data will be much easier and is done more efficiently. The functional design of this application addresses some of the challenges in ODM, by providing an efficient platform through which the data transactions can be handled more precisely. The implementation and performance of this application and all its features are tested, validated and evaluated. The evaluation results show the efficiency of the application in all the data transactions performed on the ODM.

As discussed in Section 3.3, this application is designed to allow transactions related to Water data only. Even though the other modules are shown on the GUI of the application, these modules are actually disabled. As a scope for future work, this application can be extended to handle data types related to other categories including: Atmosphere, Biota and Sediment. This will allow users to perform the data transactions related to Atmosphere data, Biological data and Sediment data.
BIBLIOGRAPHY AND REFERENCES


Ernest S.C. and T. Whiteaker; Loading Observations Data with the ODDataLoader (version 1.0); CUAHSI-HIS; July 2007.

Horsburgh J. S. and J. Berger; ODM Streaming Data Loader Version 1.0; CUAHSI-HIS; October 2007.


Montagna P.A. and K. Nelson; Observation Data Model (ODM) For Rincon Bayou, Nueces Delta; Harte Research Institute for Gulf of Mexico Studies Texas A&M University-Corpus Christi; January 2009.
Montagna P.A., Hill E.M. and B. Moulton; Role of science-based and adaptive management in allocating environmental flows to the Nueces Estuary, Texas, USA; Ecosystem and Sustainable Development VII pp. 559-570; 2009.


Whiteaker T.; HYDROEXCEL 1.1.3 SOFTWARE MANUAL (2010); Centre for Research in Water Resources at the University of Texas at Austin; September 2010.