HYDROLOGIC DATA: PUBLICATION, RETRIEVAL, AND ANALYSIS USING CUAHSI HYDROLOGIC INFORMATION SYSTEM

COLIN RICKS[1] & ALEX VAZ[2]

Abstract

Hydrologic information has been collected for many years by various organizations, each with its own methodology for publishing and retrieving information. The variety of protocols required to access and retrieve the data is confusing and time consuming. In addition, there is no central repository where researchers can access information. To solve this problem the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) implemented the Hydrologic Information System (HIS) to develop a common standard for storage and retrieval that enhances the accessibility and analysis of hydrologic data. Local HIS servers can be operated by independent organizations, including universities and federal agencies such as the USGS and EPA, that desire to publish hydrologic data. By connecting to these servers any user will be able to search for, identify, retrieve, and analyze hydrologic information over the internet with a standard method. HydroDesktop is the primary software tool used for data retrieval and analysis in the CUAHSI-HIS project. The software provides a standard scheme to retrieve hydrological information from a variety of web sources. Students in the Deer Creek Reservoir research (DCRR) team at Brigham Young University (BYU) have implemented a web-served HIS database and have created tools that assist in uploading data into the local HIS server. This is the first step in adopting the HIS database and HydroDesktop as their primary sources for data storage and retrieval. This study serves as a model for others wishing to store information with

[1] Undergraduate Research Assistant, Civil and Environmental Engineering Department, Brigham Young University, Provo, Utah (colin.ricks@gmail.com)
[2] Undergraduate Research Assistant, Civil and Environmental Engineering Department, Brigham Young University, Provo, Utah (alexjvaz@gmail.com)

the HIS database and provides a description and demonstration of the tools we built to facilitate the retrieval process.

**Keywords:** Hydrologic Information System, HydroDesktop, and HIS Servers.

Introduction

Hydrologic information has been collected for many years and across many organizations throughout the country. Such data may include stream flow, water quality, precipitation, snow, climate conditions, and a variety of other parameters. Among the largest collectors of hydrologic information is the United States Geological Service National Water Information System (USGS-NWIS), which monitors millions of sites for stream flow, water quality, and groundwater. Other federal agencies including the Environmental Protection Agency (EPA) and the United States Department of Agriculture (USDA) also collect and publish hydrologic data. However, each entity has established different methods to publish and retrieve data (Maidment 2008).

Considerable effort is required to learn how to operate these various sites, and the difficulty associated with accessing the hydrologic data can conceal available data from the interested user. In addition, local agencies and researchers that collect hydrologic data often do not publish it, resulting in duplication of efforts and wasted time and resources. If these data are published, researchers may not be aware of their existence and access methods. To provide a solution to these problems, the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) created the Hydrologic Information System (HIS) which provides a standard method to identify, locate, retrieve, visualize, analyze, and publish hydrologic data over the internet (Maidment 2008).

CUAHSI is a group of universities and water resource professionals that focus their efforts in the advancement of hydrologic science. The HIS project is intended to facilitate the accessibility of hydrologic information. Several tools have been created, and others are in development, to accomplish the project's objective. These tools include web services software to discover and download datasets, servers for delivering data, an observation data model (ODM) that provides a framework for storing data, and other related components. These tools work together to form a complete package that is designed to simplify the process of publishing, retrieving, and analyzing hydrologic information.

The HIS central service manages a distributed network of hydrologic data servers. Local HIS servers are operated by individual organizations that publish hydrologic data. Existing government servers, including the USGS NWIS database and the EPA STORET database, are integrated into the system. These data are available through the internet and anyone can retrieve and analyze hydrologic data from many shared sources in a standardized manner without charge as shown in Figure 1 (CUAHSI 2009).
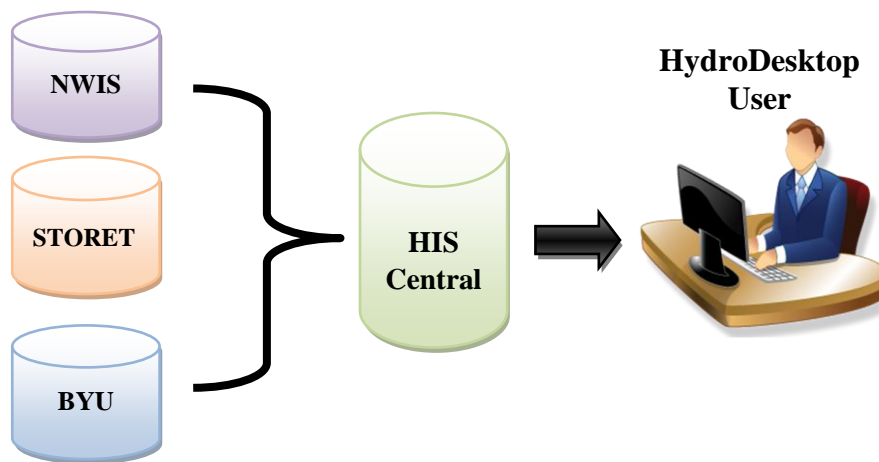
Figure 1. HIS Overview

The primary interface tool of the HIS project is the HydroDesktop application. HydroDesktop is a software tool that enables data query, map-based visualization, data download, editing, graphing, data export to model-specific data formats, and links to modeling systems (Ames, et al. 2009). While this software is still in the early development stage, it is an excellent prospect tool for retrieving and analyzing hydrologic datasets. Initial tools have been developed, and additional capabilities are being added.

Research Goals

The Deer Creek Reservoir Research team (DCRR) at Brigham Young University has collected water samples from Deer Creek Reservoir for the past three years. These include water samples and accompanying analysis, secchi measurements, vertical profiles, and other data. The DCRR intends to publish these datasets using a local HIS server. This will provide local DCRR users better access, archiving, record keeping, and allow others to access these data to benefit the research community and eliminate duplicate data collection.

This paper briefly summarizes the CUAHSI-HIS system and provides a short case study to help researchers understand the process of storing, retrieving, and analyzing hydrologic information within the HydroDesktop system. As part of this case study, we introduce several tools we have developed to aid in uploading data to the system. Specifically vertical profile data which are becoming more common as tools such as electronic water quality sondes gain wider use. As stakeholders and end users gain a greater understanding of this system, more will adopt the CUAHSI network, which will improve hydrologic science as more data becomes available.

HIS Overview and Example

The following subsections will summarize the main procedures for publishing and retrieving data, and provide examples of how the system improves hydrologic data analysis. The purpose of these subsections is to acquaint readers with the HIS system and its uses.

Server

A local database requires a local server which must be set up to manage the database. There are two types of servers: HIS Server Standard and HIS Server Lite. The lite server is sufficient for most organizations and minimizes commercial software requirements but does not include some of the advanced features of the standard server. Typically, a server uses a server operating system (OS) (e.g. Windows Server 2003), rather than a desktop OS with a server database. The HIS software requires the Microsoft Structured Query Language (SQL) be installed on the server to store hydrologic data. CUAHSI has created the ODM, which provides a standard database structure. The blank ODM schema is provided on the CUAHSI website along with instructions on implementing the schema on a Microsoft SQL 2005 server.

A graphical representation of the schema is shown in Figure 2. This schema was designed for water quality data and maintains common data relationships. To make the data useful for researchers requires significant metadata which describe the stored values. These metadata not only provide descriptions of the values, but also provide information on data quality to support research and management objectives. Typical information includes how the data were collected, who collected it, and contact information of the publisher. Other more specific metadata describe the quality and accuracy of the data.
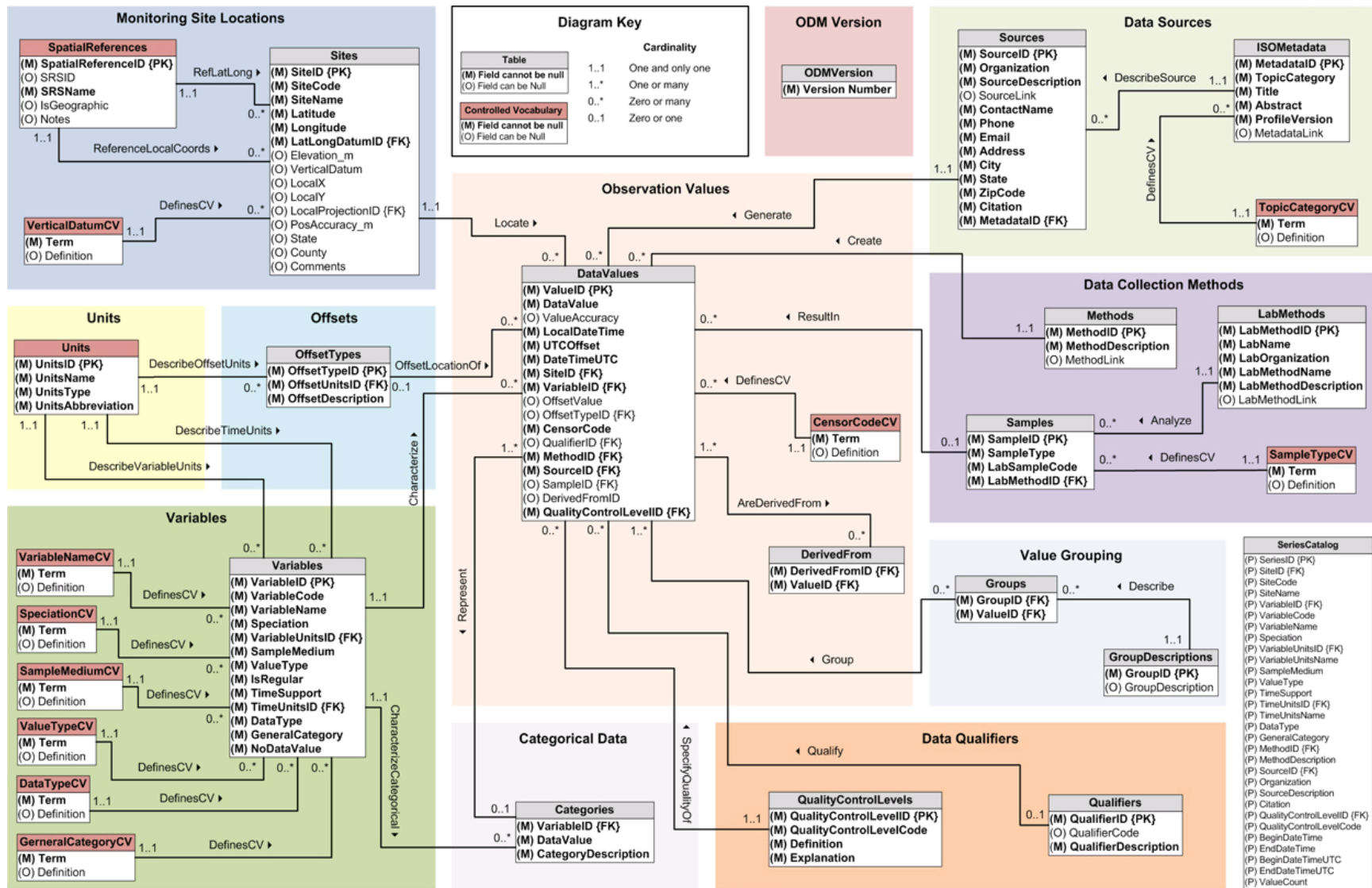
Figure 2. ODM Schema (CUAHSI 2009)

Publishing Data

Data is published through sequential steps: capture and preserve raw data, perform initial quality control, fix erroneous information, store sufficient metadata so data can be unambiguously interpreted and used, and load raw data into the DataValues table (CUAHSI 2008).

The data are loaded into the SQL database using the ODM Data Loader tool (ODMDL) developed and provided by the HIS team. The user can also enter the data manually, but this may be a long and tedious process. The ODMDL will accept files in different formats, such as comma-separated values (csv), text, and Microsoft Excel (Excel) files. We found that the most efficient method was to save data as an Excel file. This allows the data to be edited and formatted easily. Before using the ODMDL tool, the data must be formatted as outlined in the ODMDL software manual (D. Tarboton, et al. 2009).
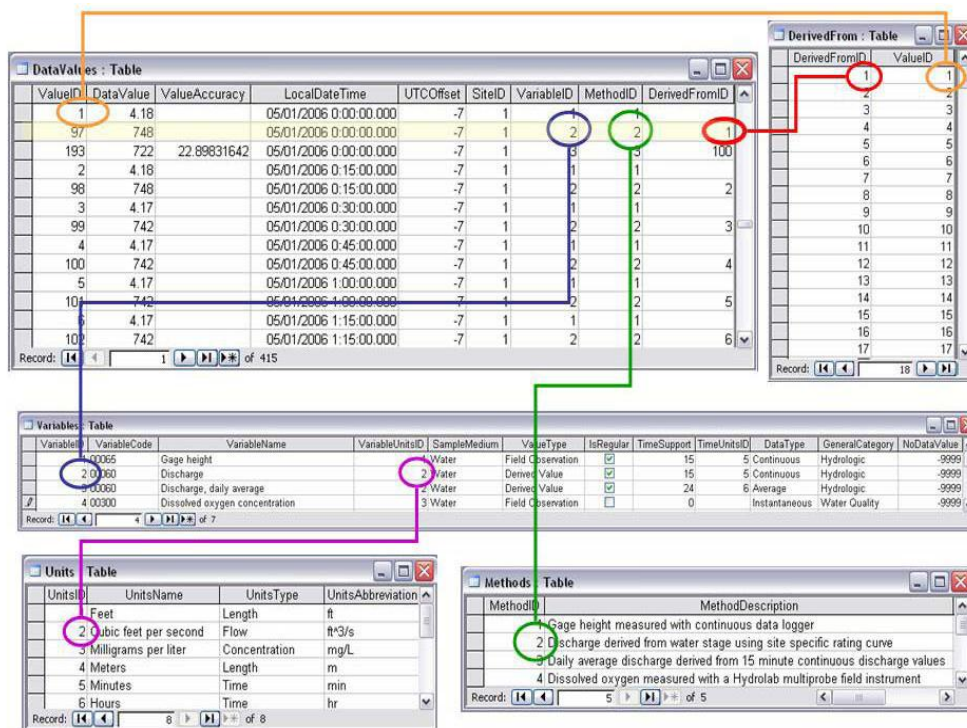

Figure 3. Example of the relationships between tables (CUAHSI 2009)

Figure 3 shows an example of the relationships between the DataValues table and the other tables in the ODM schema. Before the DataValues table can be uploaded to the database, all the other tables must be filled and loaded into the SQL database. This allows the parameters in the DataValues table to be properly linked. Once a table is formatted correctly, the ODMDL tool can be used to upload the data into the SQL database. This process, while straight forward, is prone to errors and can make uploading data time consuming. To facilitate this process, the DCRR team is developing a tool to format the raw data automatically before it can be loaded into the SQL database.

Data Preparation Tool

Because the process of formatting the data for publication can be difficult, we are creating a program using Visual Basic for Applications (VBA) in Excel to assist in preparing large amounts of collected data for publication. The purpose of the program is to automate the formatting process required by the ODMDL. The user enters the raw data and required metadata. The program uses this information and creates a table in the format specified by the ODMDL. The program's interface is shown in Figure 4.

Data Retrieval and Analysis

HydroDesktop is the CUAHSI tool to find and analyze hydrologic data in the HIS network. The software uses a GIS based visualization and data query to search for and identify the desired hydrologic information. The HydroDesktop interface is shown in Figure 5. With the GIS based visualization, users can select a geographic location by selecting a box around the desired region. HydroDesktop then zooms to the specific location where the user can view the available data. An advanced search option is also available when the user is looking for specific information.
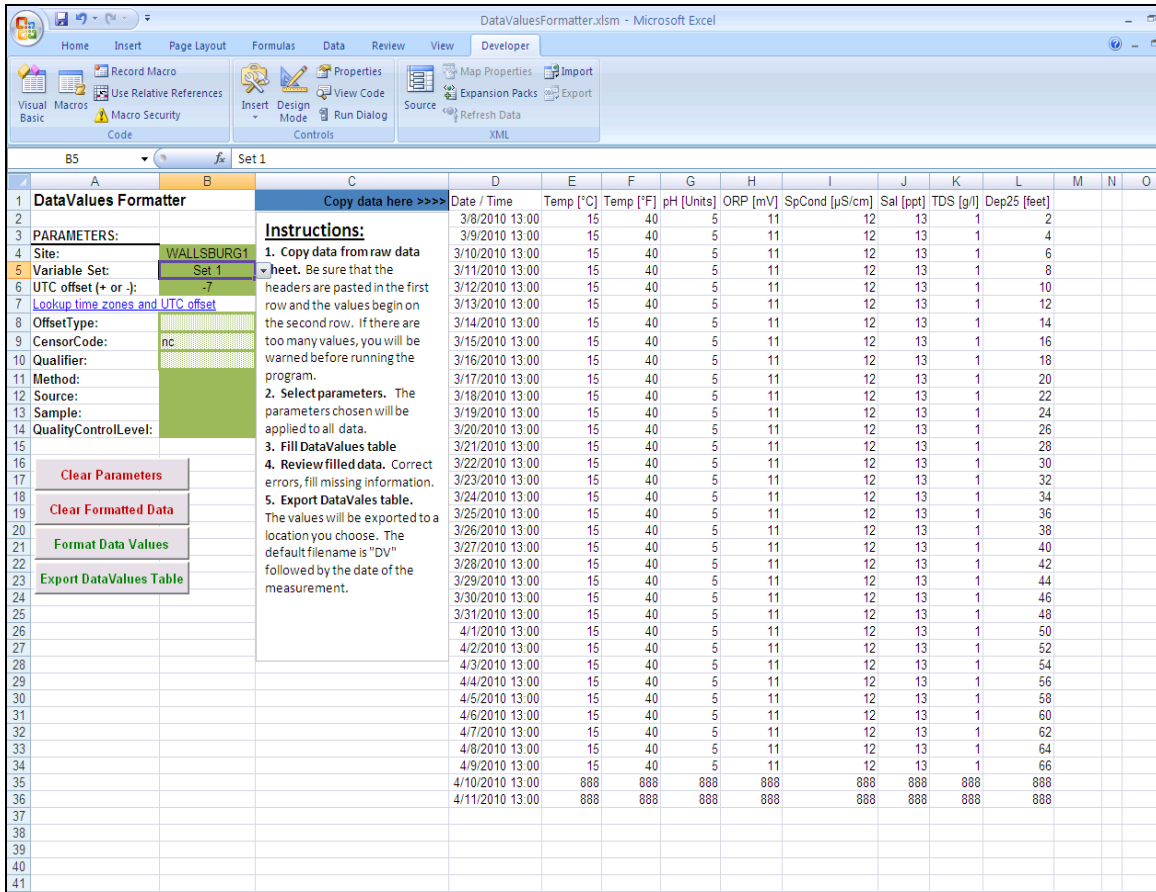
Figure 4. VBA tool created to automate raw data formatting

When data are available for a specific location, the user can view the source of the data before downloading it. Once the data are downloaded, the user can view the metadata that includes, but is not limited to, the source, accuracy, method of measurement, contact information of publisher, and spatial and temporal information. HydroDesktop provides analysis tools to view and analyze the data and also export the data for use in other tools. Data can be viewed in a table or in a graph within HydroDesktop. Data manipulation tools, such as unit conversion, data separation (e.g. monthly or daily), and data formatting are available. Data can be viewed as a table format or exported as a csv file, a common format used by other tools for analysis and modeling such as Excel (Ames, et al. 2009).
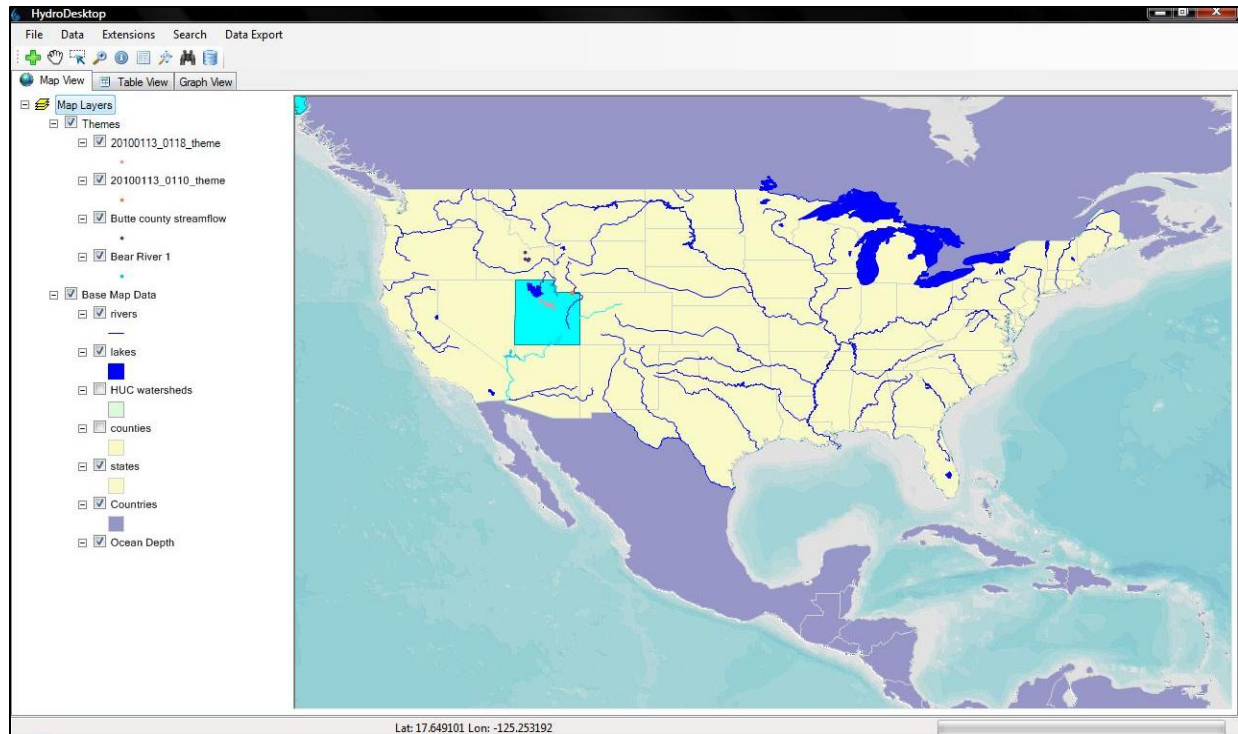
Figure 5. HydroDesktop Interface

Deer Creek Reservoir Case Study

Hydrologic and water quality models require large amounts of data that currently must be obtained from different sources. It is difficult to assemble and verify these data. To demonstrate how the HIS data tools support these efforts, the DCRR team has implemented an HIS database for Deer Creek Reservoir field data. This allows the team to access the data collected by DCCR and integrate it with data from the USGS and EPA on one platform.

The DCCR team is creating a model of Deer Creek Reservoir using a hydrodynamic model known as CE-QUAL-W2 to study nutrient and algal processes. This model demands a large amount of data, including bathymetric data, air temperature, dew point temperature, wind speed, wind direction, cloud cover, solar radiation, inflow and outflow volumes, inflow temperatures, precipitation, evaporation, water quality constituent concentrations, and hydraulic and kinetic

parameters (Williams 2007). Locating and acquiring the supporting data is difficult and time consuming because the data are collected and published by different organizations.

To support this modeling effort, the DCCR, a team of graduate and undergraduate students, have collected water samples and other data at Deer Creek reservoir during the 2008 and 2009 summer field seasons.

Our experience has been that the time required to generate and calibrate our current reservoir model is relatively small compared to the time required to acquire and format the needed information. To evaluate how the HIS database can address this issue we have implemented an HIS database and started to populate it with the data we have collected. This approach also allows us to access the USGS stream flow data, various meteorological data, and EPA water quality values. This effort is just starting, and some results from the model are shown in Figure 6.
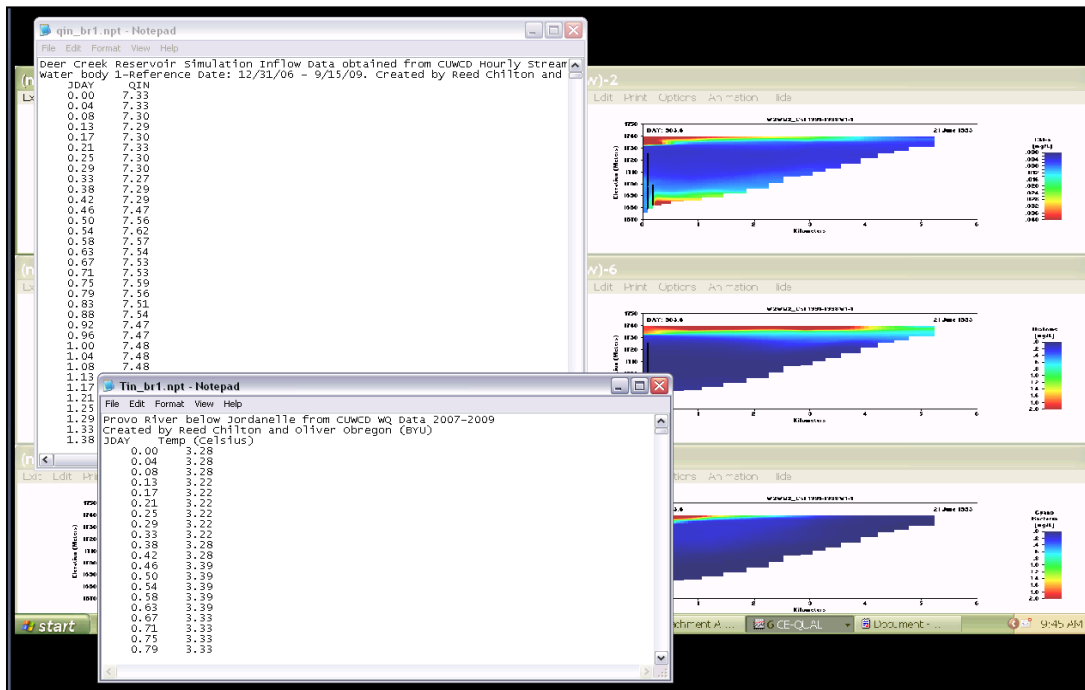

Figure 6. CE-QUAL-W2 model input files and results for Deer Creek Reservoir

Our initial experience is that while the HIS approach has promise, there are issues with formatting and describing the data for uploading. We are still evaluating our procedures and methods and do not have specific conclusions at this point. However, our initial experience prompted us to start the development of the data formatting tool described above.

Mexico Case Study and Monitoring Plan

In addition to the issues associated with assembling and storing hydrologic data, in many cases, data analysis and collection efforts are done by teams that are separated geographically. BYU has several joint water quality projects with Mexico that involve data collection, modeling and analysis. Our experience has shown that maintaining data consistency and providing data access for the separate teams is difficult.

We are using a joint project between BYU and the University of Veracruz (UV) as a case study for the HIS database approach. This project is developing a water quality monitoring plan for the Tuxpan Watershed near Veracruz. BYU is the lead for designing the collection plan, with UV as the lead for field data collection. We have implemented a HIS database and populated it with preliminary data. The ultimate goal is to provide data from the research for study and analysis to enable nearby communities to preserve, protect, and manage important water resources in the area.

CUAHSI-HIS presents an ideal solution for storing water quality data. The UV researchers will be able to store their data in a consistent manner with well-developed metadata. HydroDesktop and Qual2K can be used to analyze and model how conditions change with time.

UV researchers have not decided if they want to share the data they collect and may not want to

commit the resources to set up and maintain a web server at this time. However, these issues do

not prevent them from using HIS for a couple of reasons: data can be stored and shared locally

instead of on the web; using the ODM on a local machine is still a good way to store data

because it maintains quality control and stores metadata; and it can be easily integrated into the

CUAHSI network in the future if desired.

This project has just started, with the BYU student teams traveling to Mexico in the end of

March for initial meetings. We intend to evaluate the use of HIS for this type of collaboration

and hope to have more specific findings in the future. Figure 7 shows the initial database setup

for the Veracruz data. Notice that HydroDesktop provides different plots and basic statistical
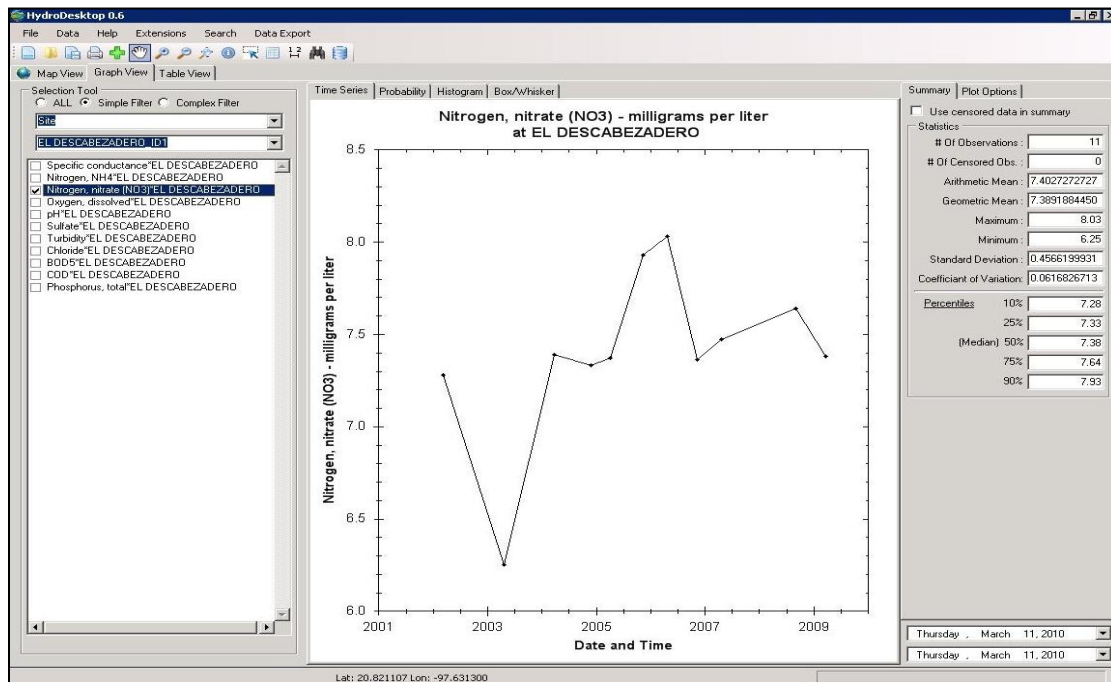
information.

Figure 7. Water quality analysis in HydroDesktop for Vera Cruz Data

Conclusion

The CUAHSI-HIS system presents many benefits for publishers and users of hydrologic data. Data can be stored in a standard format that allows easy comparison of data from diverse groups. Data can also be shared easily and quickly over the internet. Duplication of research and loss of data can be reduced, especially expensive data collection.

The HIS system is not without challenges. It requires knowledge and resources to set up a web server and maintain it. Many of the analysis tools are still being developed and are not intuitive for the average user. Organizations are still joining the CUAHSI network and making contribution to improve the system.

Hydrologic data producers and users should review the CUAHSI documentation and consider the needs of their organization before joining the HIS network. However, any user can access the vast resources that are available through HydroDesktop and other CUAHSI tools.

Our preliminary findings indicate that data formatting and uploading are difficult. We have started to develop tools to assist in this effort. We have also found that in some cases, universal access to data, especially data that have not been validated, may be problematic. This requires either a private server, which loses many of the benefits of the HIS approach, or some sort of tiered access which has not been implemented. We are continuing to study these issues and provide tools to address the problems we identify.

References

Ames, D.P., J. Horsburgh, J. Goodall, T. Whiteaker, D. Tarboton, and D. Maidment. "CUAHSI Hydrologic Information System." *Introducing the Open Source CUAHSI Hydrologic Information System Desktop Application (HIS Desktop).* July 13-17, 2009. http://his.cuahsi.org/publications.html (accessed March 1, 2010).

Beran, Bora, and Michael Piasecki. "Engineering New Paths to Water Data." *Elsevier*, 2008: 753-760.

CUAHSI. "HIS Desktop Functional Specifications." *CUAHSI-HIS Hydrodesktop.* June 2009. http://hydrodesktop.codeplex.com/wikipage?title=Functional%20Specifications&referringTitle=Home (accessed March 1, 2010).

—. *HIS Servers-Lite Version.* 11 19, 2008. http://river.sdsc.edu/wiki/HIS%20light.ashx (accessed 3 1, 2010).

Goodall, Jonathan. "A Geotemporal Framework for Hydrologic Analysis." *CUAHSI Hydrologic Information System.* August 2005. http://his.cuahsi.org/publications.html (accessed March 1, 2010).

Horsburgh, Jeffery, David Tarboton, Michael Piasecki, David Maidment, and Ilya Zaslavsky. "An integrated system for publishing environmental observations data." *Elsevier*, 2009: 897-888.

Maidment, D. R. "Bringing Water Data Together." *CUAHSI Hydrologic Information System.* April 2008. http://his.cuahsi.org/publications.html (accessed March 1, 2010).

Tarboton, D, J Horsburgh, D Maidment, T Whiteaker, I Zaslavsky, and M Piasecki. "Development of a Community Hydrologic Information." *CUAHSI Hydrologic Information System.* July 13-17, 2009. http://his.cuahsi.org/publications.html (accessed March 1, 2010).

Williams, Nicholas. *MODELING DISSOLVED OXYGEN IN LAKE POWELL USING CE-QUAL-W2.* Provo: Brigham Young Univeristy, 2007.

Additional Information

For a complete summary of the Hydrologic Information System the reader is encouraged to visit CUAHSI's website at http://his.cuahsi.org.