CUAHSI

universities allied for water research

CUAHSI HYDROLOGIC INFORMATION SYSTEM: 2009 STATUS REPORT

Edited by David R. Maidment
Center for Research in Water Resources
The University of Texas at Austin

October 23, 2009

Distribution

Acknowledgement

Funding for this research was provided by the National Science Foundation under Grant No. EAR-0413265. In addition, much input and feedback has been received from the CUAHSI Hydrologic Information System development team, from the WATERS testbed network data managers, the CZO Observatory scientists, the ESRI Water Resources Applications team, and from the CUAHSI Program Office. These contributions are gratefully acknowledged.

Table of Contents

Chapter 1. INTRODUCTION

By David Maidment, The University of Texas at Austin

1.1 CUAHSI HYDROLOGIC INFORMATION SYSTEM

The Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI) (http://www.cuahsi.org) is an organization representing 122 US universities, which is supported by the Earth Sciences Division of the National Science Foundation to develop infrastructure and services to advance hydrologic science in the nation's universities. One component of CUAHSI's activity, also funded by the National Science Foundation, is a Hydrologic Information System (HIS) project, which is developing infrastructure and services to improve access to hydrologic data. The overall goals of this project are:

Data Access – providing better access to a large volume of high quality hydrologic data;
Hydrologic Observatories – storing and synthesizing hydrologic data for a region;
Hydrologic Science – supporting science by providing a stronger hydrologic information infrastructure;
Hydrologic Education – bringing more hydrologic data into the classroom.

The purpose of this report is to inform a review in October 2009 of the CUAHSI HIS project by the HIS Standing Committee, a group of water and computer scientists drawn from academia, government and industry, who are charged by CUAHSI's Board of Directors with assessing the progress of the HIS project. This report supplements and updates a comprehensive overview of the HIS project as it stood in July 2008 (HIS, 2008), which described the HIS project as it existed at that time. Three new water data capabilities developed as part of the HIS project were presented in HIS (2008):

Data Storage in an *Observations Data Model* (ODM), which is a standardized relational database structure for storing and describing time series of hydrologic observations data measured at point locations;

Data Access through internet-based *Water Data Services* that enable querying and accessing time series observations data stored at remote locations in ODM databases, and in the water databases of public agencies, and delivery of these data in a consistent data language, called WaterML;

Data Indexing through a *National Water Metadata Catalog,* built as part of HIS Central at the San Diego Supercomputer Center that assembles in a consistent form the metadata that describe the water observation networks of the nation, and enables data searching across these networks. This catalog presently indexes nearly 10 million time series of water observations data collected at nearly 2 million locations in the United States and globally.

The HIS Project is carried out by a multi-university team at five universities:

University of Texas at Austin – David Maidment, Tim Whiteaker, Eric Hersh, James Seppi, Wendy Harrison
San Diego Supercomputer Center – Ilya Zaslavsky, David Valentine, Tom Whitenack, Matt Rodriguez
Utah State University – David Tarboton, Jeff Horsburgh, Kim Schreuders
Idaho State University – Dan Ames, Ted Dunsford, Jiri Kadlec
University of South Carolina – Jon Goodall, Anthony Castronova

The San Diego Supercomputer Center (SDSC) is the cyberinfrastructure partner and houses HIS Central, the hub for providing access to data through web services. The hydrologic science researchers at the other universities focus on developing the hydrologic information technology to serve and process the data. In addition, as part of the recent five-year renewal by NSF of the CUAHSI Program Office, an HIS support person, Yoori Choi, has been added to that office, and Richard Hooper, President and CEO of CUAHSI, is actively assisting the HIS team by developing an ontology of key concepts for indexing hydrologic information.

Defining the nature of a Hydrologic Information System is a challenge. The HIS team is trying to create something that has never existed before, and is operating at the outer limits of current technology. By definition, a system is an array of connected components, and in this instance the components can be defined as software applications that store, access and index hydrologic information. The connection among them is established by *web services* which are automated functions that enable one computer to make appropriate requests of another computer and receive responses through the internet. In this sense, the HIS team and its partners are creating a *services-oriented architecture* for water information. Josuttis (2007) defines a services-oriented architecture as "a concept that applies to large, distributed information systems that have many owners, are complex and heterogeneous, and have considerable legacies from the way their various components have developed in the past." This definition certainly applies to the water resources field which has thousands of agencies and individuals who collect and archive water information in their own way. It also applies to other fields such as insurance, banking, health care and public safety, that also have many institutions and individuals who store and exchange information.

Another way of thinking about a Hydrologic Information System is by analogy with a Geographic Information System (GIS). Tomlinson (2003) states that "a GIS stores spatial data with logically-linked attribute information in a GIS storage database where analytical functions are controlled interactively by a human operator to generate the needed information products." This definition implies that all the information has been harvested and stored in a local database and is then available for analysis and interpretation. However, unlike GIS where the data are static and change little through time, a hydrologic information system is representing phenomena that are inherently dynamic and vary greatly through time.

If we combine these two concepts, the services-oriented architecture, and the desktop application which stores and operates on the information on a local desktop computer, we may define a Hydrologic Information System as shown in Figure 1-1 as comprising three components:

HIS Server is a repository of hydrologic time series data published as WaterOneFlow web services;
HIS Central is a metadata catalog of data accessible through WaterOneFlow web services
Hydro Desktop accesses hydrologic data discovered on HIS Central and acquired from HIS Servers

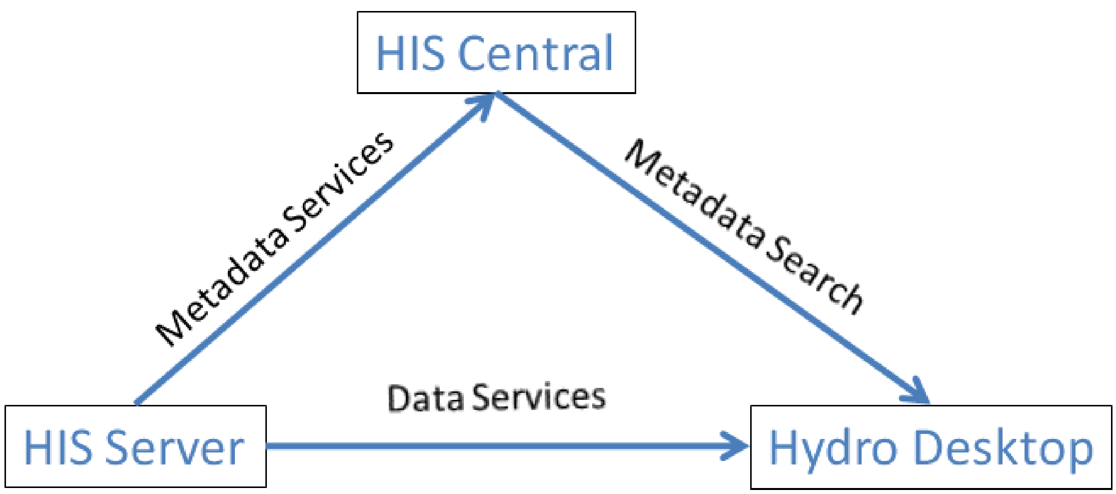


Figure 1-1 The three components of a Hydrologic Information System linked by web services

In developing these definitions it is important to recognize that these three components are not unique to hydrology, and indeed the internet itself is also based on these three elements: information servers that publish text and images in Hypertext Markup Language (HTML), portals such as Google and Yahoo that index published documents and enable searching across them, and applications such as web browsers that access and display the resulting documents and files. Also, it is important to note that, like the internet, there may be multiple instances of each component.

As Figure 1-1 shows, an HIS Server publishes both data and metadata services. For example, for a stream gage, the data consists of the time series of discharge and stage height values recorded at the gage, while the metadata describes the location of the gage, the variables measured there, the period of record, and the organization that operates the gage. By contrast, HIS Central indexes and catalogs only metadata, but it does so across many HIS Servers which contain hydrologic information for a particular geographic region. This supports a user of HydroDesktop who can search the metadata stored in HIS Central and then acquire the data directly through data services provided by HIS Server. Because data and metadata services are public information sources they can be accessed by any appropriate computer application not just Hydro Desktop.

1.2 WHAT'S NEW IN 2009?

Hydro Desktop

A key innovation for CUAHSI HIS in 2009 is the creation of Hydro Desktop. In other words, operating versions of the HIS Server and HIS Central already existed in 2008, and the data and metadata services they provided could then be accessed through programming languages such as Python, or customized applications such as Hydro Excel, an adaptation of Excel that access CUAHSI WaterML time series data services. Hydro Desktop is a new local application that accesses and acquires data from web services, allows for the addition of local files that the user already has available, and provides an open source platform for hydrologic scientists to build hydrologic analysis applications and links to hydrologic models.

As shown in Figure 1-2, Hydro Desktop is intended to access and synthesize water observations data with data from GIS, climate, and remote sensing, and to provide links to hydrologic models. The goal for 2009 is to have an operational Hydro Desktop just for water observations data, so that Hydro Desktop can search through metadata in HIS Central and acquire the corresponding data from a set of geographically distributed HIS Servers, and that goal has been accomplished. Since Hydro Desktop is built using MapWindow as a base, it already has a significant capacity to process and display GIS data. Hydro Desktop includes a local database for time series data and metadata, and this allows hydrologic scientists to store and manipulate information in both space and in time.

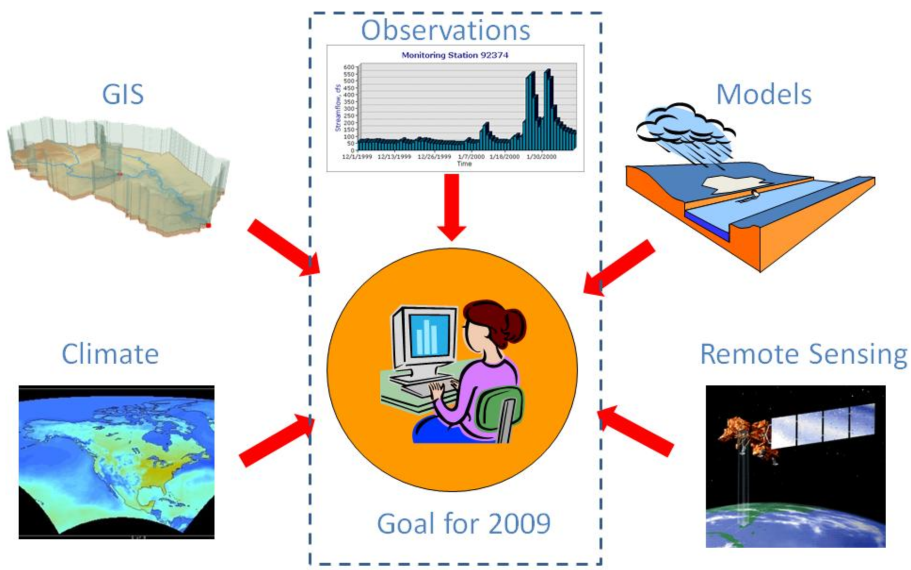


Figure 1-2 The components of Hydro Desktop. For 2009 the focus is on water observations data

Hydro Desktop is managed by a new partner in the HIS team, Dan Ames of Idaho State University, and it builds on the MapWindow open source GIS that he also manages. The source code for Hydro Desktop is maintained in a public subversion repository at http://www.hydrodesktop.org and the first executable version was created in October 2009. Hydro Desktop is undergoing continual development and the October 2009 version is in "Pre-Alpha" form, that is, it performs some functions satisfactorily and is being shared among its software development team, but it is not yet ready for testing by a wider audience.

New CUAHSI Water Data Services

During 2009, a number of new WaterML services have been established and existing ones have been improved and solidified. In September 2009, the USGS began publishing its instantaneous water observations data in WaterML, after having earlier established a WaterML service for their daily values data. Instantaneous data are those measured every 15 minutes or at other intervals that are statistically summarized each 24 hours to provide daily values, such as daily streamflow data. The USGS Instantaneous data service provides access to data at 11188 sites in the United States for a set of 80 variables, a subset of which is measured at each site. The USGS service presently provides instantaneous data up to 60 days before present but the USGS archives of instantaneous data extend much further back in time and these longer data series may be exposed in WaterML at a later time. The US Army Corps of Engineers, Rock Island District, has published a WaterML service for 2210 sites, mainly in the Mississippi River basin, providing data mainly for water levels and water quality. The National Climatic Data Center's Integrated Station Hourly and Daily WaterML services have been indexed at the main CUAHSI HIS Central instance housed at SDSC and can now be searched for information on up to 34 variables measured at 13,628 sites across the globe, with data available up to 36 hours before the present time. Drexel University has established a HydroNexrad web service for gridded precipitation for the Chesapeake Bay Environmental Observatory at 215359 points each located at the center of a cell on a 1km SuperNexrad mesh. HydroNexrad is an application for processing high precision Nexrad data developed by the University of Iowa, so it is interesting that value-added data from academic data sources are now starting to play a major role in HIS Central and CUAHSI Water Data Services.

Usage statistics presented in Chapter 3 show that the amount of data indexed at HIS Central has increased by more than a factor of ten during the past year from 342 million data to 4.3 billion data, largely because of the HydroNexrad service. The usage of CUAHSI Water Data Services to access water observations data has increased by a factor of three since last year from 450 requests per day to 1500 requests per day, and within this, usage of USGS water web services through HIS Central, has increased by a factor of ten from 110 to 1100 requests per day. We anticipate that as the use of HydroDesktop to access data through web services becomes more established, these usage figures will increase further by a significant amount.

A Texas Hydrologic Information System has been established in collaboration with state level water agencies in Texas led by the Texas Water Development Board. An instance of HIS Central at the University of Texas at Austin indexes 10 water data services, summarizing state and local water agency and academic data measured in Texas, together providing more than 1 million water observation time series, where a water observation time series is a set of measurements of one variable, at one site, by one organization, collected over a defined period of time. This is leading to the publication of Texas water observation data "themes," which are datasets compiled from observations data harvested from the core water data services, for quantities such as salinity, evaporation, water temperature and nutrients. The intention with "themes" is to provide access to carefully curated and readily usable data packages of water information, customized for particular subject areas and geographic regions.

The Inland Northwest Research Alliance (INRA) Constellation of Experimental WATERsheds (ICEWATER) project (http://icewater.inra.org) in the Northwest is an operating example of the synthesis of the three components described in Figure 1-1, where the HIS Central instance is located at Utah State University, and the HIS Servers are at Boise State University, Idaho State University, Montana State University, University of Alaska Fairbanks, University of Idaho, University of Montana, Utah State University, and Washington State University. These HIS Servers publish both CUAHSI water observations data services using information stored in the CUAHSI Observations Data Model and geographic datasets stored in ArcGIS Server. Both the observations and metadata about the geographic spatial data services are stored within the SQL/Server relational database, which is the foundation of the HIS Server.

Internationalization of WaterML

The initial version of WaterML, version 1.0, was invented by the CUAHSI HIS team and it presently connects all the HIS components. This is an eXtended Markup Language (XML) designed to convey water observation time series data and metadata through the internet, and to support web service requests for these data.

The most accepted international standards for internet transmission of the geospatial data have been established by the Open Geospatial Consortium (OGC) using GML, or Geographic Markup Language, which is also an XML language like WaterML. Beginning in September 2008, the OGC has joined with the Commission for Hydrology of the World Meteorological Organization to establish an OGC/WMO Hydrology Domain Working Group http://external.opengis.org/twiki_public/bin/view/HydrologyDWG/WebHome. The OGC/WMO partnership at the international level parallels the partnership between cyberinfrastructure experts at SDSC and hydrologic scientists at other universities within the CUAHSI HIS team.

The Hydrology Domain Working Group meets regularly each three months, generally twice in Europe and twice in North America each year, as part of the OGC Technical Committee meetings. In the five meetings already held, a significant amount of progress has been made in learning about technical developments in water web services in the United States, Europe and Australia. The goal of the Hydrology Domain Working Group is to design a new version of WaterML, version 2.0, that will be compliant with OGC and WMO information standards, and to test this with a set of Interoperability Experiments for exchange of water information within countries and across national borders. The development of WaterML through the OGC/WMO working group signals the emergence of international, public domain, open source standards for water data for the first time. If successful and widely adopted, this could have profound regional, national and global impacts on advancing the communication and synthesis of water data to support water science and water management.

Links with Hydrologic Modeling

One of the key goals of hydrologic information access and synthesis is to support hydrologic modeling. Since modeling takes place on many software platforms and in many programming languages, it is not the goal of the HIS project to be doing hydrologic modeling, but rather to be linking to hydrologic models so as to provide information to them and to receive information from them. CUAHSI has a separate initiative, the Community Hydrologic Modeling Platform (CHyMP), whose goal is to construct and share hydrologic modeling components. There are two distinct flavors of hydrologic models, lumped and distributed. Lumped modeling means that a hydrologic system such as a watershed or river basin is represented by a set of spatially discrete entities, such as subwatersheds and river reaches, each of which has representative properties, and the hydrologic model solves a set of ordinary differential equations describing the time variation of hydrologic processes occurring in these entities. Distributed modeling means that the spatial domain is represented by a mesh or grid of computation points, and the hydrologic model solves a set of partial differential equations in space and time.

Several model-data interoperability systems have been developed and the HIS team has tested two of these: the Open Modeling Interface OpenMI, developed with support from the European Commission, which links lumped hydrologic models and data sources, and the Earth System Modeling Framework (ESMF), developed by NCAR and partners, which links distributed models and data sources. An OpenMI plugin for Hydro Desktop has been developed by Jon Goodall and colleagues in the HIS team at the University of South Carolina, a WaterML linkage to ESMF has been developed by the ESMF team. It would be desirable also to understand how Hydro Desktop can be linked to the Community Surface Dynamics Modeling System project.

Adaptation of the Hydrologic Ontology

A critical element of the conceptual model of HIS presented in Figure 1-1 is the capacity to search hydrologic metadata in portals to identify information of particular types. This is complicated by the fact that even though WaterML smoothes out *syntactic heterogeneity*, or variations in data format, it does not eliminate semantic heterogeneity, or differences in the way variables are described. A brilliant piece of research carried out by Bora Beran and Michael Piasecki at Drexel University earlier in the HIS project demonstrated how a centralized metadatabase could be harvested from CUAHSI water data services and how a set of search services for particular variables could be performed over this metadatabase by using an application called HydroSeek. HydroSeek is based on a hierarchy of search terms, or hydrologic ontology, expressed as a connected set of files in the Ontology Web Language (OWL). The HydroSeek application and its underlying metadatabase and search services, were transferred to HIS Central at SDSC and are continuously maintained there. As new CUAHSI water data services are registered at HIS Central and cataloged in the metadatabase, their variables are linked to the concepts in the hydrologic ontology so that they can be included in HydroSeek searches. These search services at HIS Central are now accessible to the Hydro Desktop application operating at any geographic location. Hydro Desktop can select particular classes of information, copy their metadata into its own local database, and then harvest the corresponding observations from their HIS Servers, wherever those are located. This greatly simplifies data searching and acquisition.

As this process evolved, several limitations became apparent. First, it was very difficult to understand and to manipulate the hydrologic ontology as a connected set of fragmented OWL files. However, the main visualization tool for the ontology is in a StarTree viewer, and this viewer is based on a tabular listing of the ontology expressed as a concept paths, where each record in the list specifies the sequence of concepts on the path from the central core concept to a single leaf concept on the outer edge of the concept tree. This simplification suggested that the CUAHSI hydrologic ontology could be constructed in tabular form, and that transformation was accomplished by Kate Marney and David Maidment at the University of Texas. Matt Rodriguez at SDSC has shown that the whole metadatabase and its linkage to concept tables, can be presented as a Data Cube in SQL/Server using its OLAP (OnLine Application Processing) capabilities, which provides rapid access for searching the metadata across a number of search dimensions, such as site location, time period, selected concepts, and media (air, water, soil) within which the observations were made. The result will be a more robust and unified metadatabase to support search services although this system has not been fully implemented yet.

In parallel with the CUAHSI work on hydrologic ontologies, the USGS and EPA have for a number of years been seeking to unify the capacity to search their respective water quality databases by referencing their variable descriptions to a list of substances in the EPA Substance Registry System. For example, the list of substances includes Oxygen, so all forms of Dissolved Oxygen measured in water are linked to the term Oxygen as a substance. The USGS water quality database requires about 1600 substances to index its measured variables. The CUAHSI HIS team has examined this approach to indexing water quality data using the Substance Registry System and has decided to use that approach as a point of departure in its future work on indexing water quality data.

Richard Hooper has converted the hydrologic ontology into a relational database form. In consultation with colleagues in the CUAHSI community, he is the person who will finalize the concept and structure of the CUAHSI hydrologic ontology.

Critical Zone Observatories

One of the key goals of the CUAHSI HIS project is to support the publication of data from hydrologic observatories. These can take several forms, and one that the National Science Foundation Earth Sciences Division has chosen to fund is Critical Zone Observatories, where the Critical Zone is defined to be from the top of the bed rock to the top of the atmospheric boundary layer, and all the phenomena that lie in between those levels, including surface and subsurface hydrologic processes. A number of different types of data are collected at Critical Zone Observatories, but a data inventory compiled of the information being generated at the first three observatories at Penn State, the University of Colorado at Boulder, and University of California, Merced, showed that regularly sensed or irregularly sampled time series of information at point locations produces the main part of the water observations data collected at those observatories. Three additional Critical Zone Observatories are in the process of being established. A supplement to the Critical Zone Observatory grants has been made by NSF to support establishing web sites for CZO data at each observatory, a common CZO Data Portal for observations time series at the San Diego Supercomputer Center, and an adaptation of Hydro Desktop to form a CZO Desktop with similar search capabilities for data indexed in the CZO portal.

Regional HIS Training Courses

Tim Whiteaker at the University of Texas developed a 1.5 day training course in preparing and publishing hydrologic observations data. The course begins with a set of ascii files of time series data measured at a network of observation points in a region, shows how to transform those files and ingest them into the CUAHSI Observations Data Model in SQL/Server, how to provide access to the resulting data through WaterML web services, and how to register this WaterML web service at HIS Central and access its data through HydroSeek. This course has so far been taught at five locations: Melbourne, Australia (November, 2008), Kansas State University (March, 2009), University of Texas at Austin (May, 2009), University of Vermont (June, 2009), and University of Iowa (September, 2009). The Iowa class was taught by Yoori Choi as she transitions into the role of supporting HIS implementation in CUAHSI institutions. A similar course was taught at Boise State University (August, 2009) for participants in the eight universities in the ICEWATER network.

Each of these workshops involves extensive preparation with local partners at the host institution. A dataset of water observations in the local region is selected, and a HIS Server is set up at the host institution to house the dataset once it is compiled into the CUAHSI format. Besides academic participants, representatives of regional water agencies are contacted, and invited to participate, so that a beginning can be made for linkage between academia and government that is needed to establish a regional HIS. Typically, a course involves 20-30 academic and water agency participants. Once the course is over, the pilot regional HIS Server provides a base and application for further development of HIS. In this way, the leadership that CUAHSI water scientists provide from their universities is used as a catalyst for water data synthesis in their state or region.

1.3 WHAT HAVE WE LEARNED?

It may be useful at this point in our project to step back for a moment and assess what lessons have been learned in the five years since the HIS project was initiated, and in particular in the three years since the present five-year grant was awarded in January 2007. Here are some thoughts in that regard:

- A critical contribution of the HIS project is the invention of WaterML and the demonstration that it works to provide a common language for water observations data transmission through the internet. This language now has sufficient adoption by the USGS and other agencies, and progress toward internationalization through the OGC/WMO Hydrology Domain Working Group, that there is a reasonable likelihood that it will evolve into a permanent open source standard for water data publication.
- The amount of water observations data indexed at HIS Central has increased by a factor of ten during the past year from 342 million data to 4.3 billion data, and the number of requests for data from CUAHSI Water Data Services over the same period has increased by a factor of three, from 450 requests per day to 1500 requests per day.
- The Observations Data Model is a comprehensive way of storing and cataloging hydrologic observations time series, and the link between the ODM and its attached WaterML web services is very robust.
- Indexing hydrologic observations metadata across a set of water data services in a centralized metadata catalog is valuable information source for supporting searches for particular data types.
- Hydrologic observations data collected by academic investigators are project oriented – they are collected to support a particular set of research goals funded by a research project of finite duration. The ODM and WaterML web services built at the project site are one way of storing and providing public access to the data during the project's duration, and after it is completed.

- An alternative model is to wait until the project is mature and then capture the data in whatever form it is in, ingest it into the ODM, and store it in a centralized repository at SDSC. This is the path being followed with the Critical Zone Observatories, who are going to internally manage their data in whatever way they choose.
- For hydrologic scientists to comprehensively access the CUAHSI services-oriented architecture for water data, a customized local application is necessary, like the Unidata Local Data Manager. The idea that the data can be captured using web services imbedded in programming languages, or by adaptation of existing software like Excel and ArcGIS, has some validity, but none of these approaches is configured to provide a comprehensive interface to a services-oriented architecture. This is why we have developed Hydro Desktop. We intend that by engaging at first a small group of users of Hydro Desktop in its design and development, we can gradually solidify and expand the capability of the application and the size of the audience within the CUAHSI community that uses it.
- Registering WaterML web services and building and maintaining a National Water Metadata Catalog at HIS Central requires continual investment and effort. Having a single national source of such metadata has some advantages but it does not serve all purposes. Equivalent needs and focus exist within states and regions who may have little interest in information from distant locations in the nation, but a great interest in synthesizing national, state and local information for limited geographic areas. This requires the deployment of a standardized Hydro Portal whose contents can be cataloged, and searched from within neighboring Hydro Portals and by external applications. We think it is possible to accomplish this by using the Hydro Portal adaptation of the ESRI Geoportal Extension programmed for the HIS team by the ESRI Water Resources Applications group[1], but this capability is not fully tested yet and will take time to solidify and systematize.
- The idea of applying CUAHSI HIS technology at the state and regional level has the capacity to capture state and local water observations data that are not included in national datasets compiled by federal agencies. The regional and state-level HIS workshops conducted during the past year have been valuable in establishing this role in some parts of the nation. The leadership with their state and region provided by CUAHSI scientists from their universities is a valuable method of innovation and deepening the nation's repository of indexed water data accessible in WaterML. Because of the neutral role of academia in interaction with government water agencies, CUAHSI scientists have a unique opportunity to provide this leadership. If not us, then who? If not now, then when?
- The HIS Server and HIS Central components of the HIS are founded on mature commercial information technology that is adapted to serve the purposes of HIS. This approach is used so that as the underlying technology evolves with emerging information standards, the HIS server and search capabilities themselves can evolve with minimal need for maintenance effort by the HIS team.
- The HydroDesktop component of HIS is an open source, public domain application with a public subversion repository for its source code to encourage others in the HIS and related communities to join in engaging in its development by producing plug-ins that provide additional functionality. This is the part of HIS that the users directly interact with and wish to enhance so it is important that this component is available free of charge and is eventually cross-platform compatible, especially between the Windows and Macintosh operating systems.
- The elements of the Hydrologic Information System shown in Figure 1-1 are probably complete enough that there is no need to add major additional components. What is needed over the remaining two years of the project is to solidify the technology and engage and train the CUAHSI community in its use.

---

[1] The HIS team wishes to express its appreciate to the Environmental Systems Research Institute and to Dean Djokic and his programming team in the ESRI Water Resources Applications Group for their many contributions to our research, which are provided free of charge to the HIS project.

1.4 Scope of This Report

This report contains eight chapters, beginning with this introduction, Chapter One. Chapter Two summarizes the Hydro Desktop application and its development. Chapter Three describes HIS Central and CUAHSI Water Data Services. Chapter Four describes the architecture of HIS Server and its application in support of the ICEWATER network among universities in the Northwest. Chapter Five describes the Texas Hydrologic Information System as an example of CUAHSI technology applied at the state rather than national level. Chapter Six summarizes the present conception of the Hydrologic Ontology. Chapter Seven describes our work on linking hydrologic data and models. Chapter Eight summarizes our HIS education and communication activities.

1.5 References

HIS (2008), “CUAHSI Hydrologic Information System: Overview of Version 1.1”, Produced by the CUAHSI HIS Project Team and edited by David R. Maidment, http://his.cuahsi.org/documents/hisoverview.pdf

Josuttis, N.M., (2007), “SOA in practice – the art of distributed system design”, O’Reilly Press, Sebastopol, CA, 324p.

Tomlinson, R., (2003), “Thinking about GIS”, ESRI Press, Redlands CA, 283p.

Chapter 2. HydroDesktop

By Dan Ames, Idaho State University

2.1 Initial Objectives

HydroDesktop (formerly referred to as HIS Desktop) is a new component of the HIS project intended to address the problem of how to obtain, organize and manage hydrologic data on a user's computer to support analysis and modeling. As designed and developed thus far, HydroDesktop is focused on facilitating the discovery and access of hydrologic data and, secondarily, providing support for data manipulation and synthesis. It also provides a visualization tool for HIS Server based data and a platform for the integration of HIS data.

HydroDesktop is being developed using the open source MapWindow 6 open source GIS (geographic information system) software development toolkit together with a number of existing open source spatial and temporal data analysis and visualization tools. In this way, the full software package meets the requirement of being completely open source and accessible for third party developers and researchers. The source code for HydroDesktop is maintained at http://www.HydroDesktop.org which is based on a community open source software development portal maintained by Microsoft. Additionally, HydroDesktop's design maintains the use of a plug-in architecture that allows for the extension of the program's core functionality and encourages third party participation in the open development community.

This chapter provides a brief overview of the HydroDesktop program as it exists thus far then address progress regarding data discovery and manipulation and community development.

2.2 Overview

As stated in the HIS Desktop Functional Specifications design document (June 2009):

> "HIS Desktop will be a platform for the integration of hydrologic data, which can be used in analysis applications such as R, MATLAB, and Excel, or in custom code developed by the end user. The HIS Desktop design paradigm includes the use of a plug-in architecture and data abstraction layer that will allow extension of the core functionality. HIS Desktop will provide local access to data obtained from distributed data services that are part of the Internet-based, SOA that the CUAHSI HIS project has developed for the sharing of hydrologic data."
> (See http://hydrodesktop.codeplex.com/Wiki/View.aspx?title=Functional%20Specifications)

To support these goals, we have developed HydroDesktop with a familiar primary interface similar to most desktop GIS programs with the addition of tools and forms specifically related to time series data visualization and analysis.

Figure 2-1 shows the HydroDesktop default interface as it appears on a user's local computer. Included are a *Main Menu*, *Toolbar*, *'View' Tabs*, *Legend*, and a main *Map Display*. The *Map Display* is the main visualization tool while the other portions of the interface provide tools for searching, obtaining and managing data. It is anticipated that the program will be used by a wide variety of users from graduate students to university faculty to scientific researchers. With this wide base of users in mind, HydroDesktop maintains a user friendly interface that can be easily managed regardless of the operator's technical background.

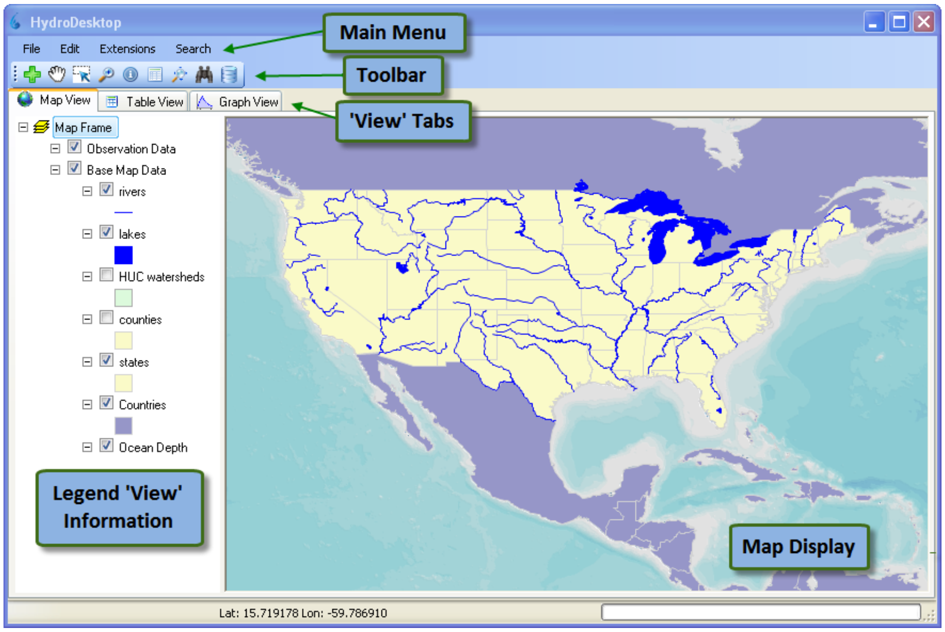


Figure 2-1 HydroDesktop Main Interface

2.3 Discovering and Accessing Data

HydroDesktop uses the methods from the HIS Central metadata catalog Application Programming Interface (API) to provide search capabilities across the catalog to determine relevant data sets for a specific user. A series of data discovery forms has been developed to provide the user with a quick and easy, step by step means of defining the type of data they are looking for and the source from which they wish to find it. These forms can be accessed simply by clicking the *Search and Download Data* button in the *Toolbar*. These forms present the user with the option to input one or more of the subsequent search criteria.

The first form allows the user to establish whether to search the HIS Central catalog or to search a local metadata source.

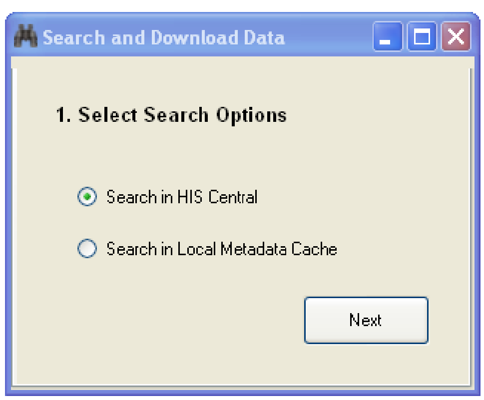


Figure 2-2 First Data Discovery Form

The second form creates a latitude/longitude bounding box to serve as the spatial constraint on the query. This form makes use of the interactive map to set the context for the data discovery. An area of interest is often used as a filter for narrowing a search for data. The box can be input by typing in coordinates, by drawing a rectangle on the HydroDesktop map, or by selecting a feature from one of the layers in the HydroDesktop map (e.g., a watershed boundary, a county or a state – the extent of the feature is converted to a latitude/longitude box). Note that this and subsequent forms can be skipped or that the search can be run at any given point in the series of forms. All that is necessary is that at least one of the criteria be entered.

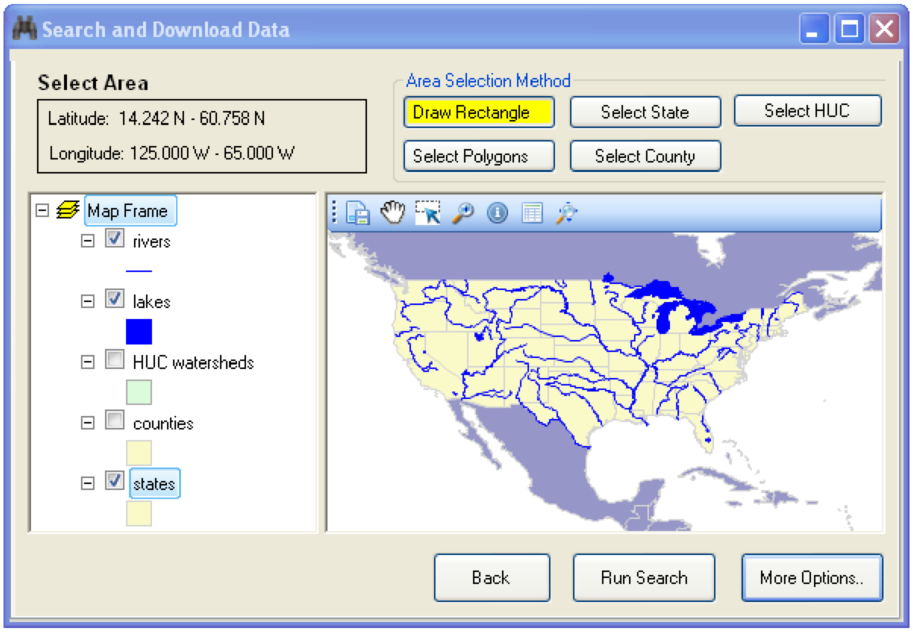


Figure 2-3 Second Data Discovery Form

The third form allows the user to search concepts from the HIS Ontology. The ontology-based search uses keywords that are either input by the user or selected from a list.

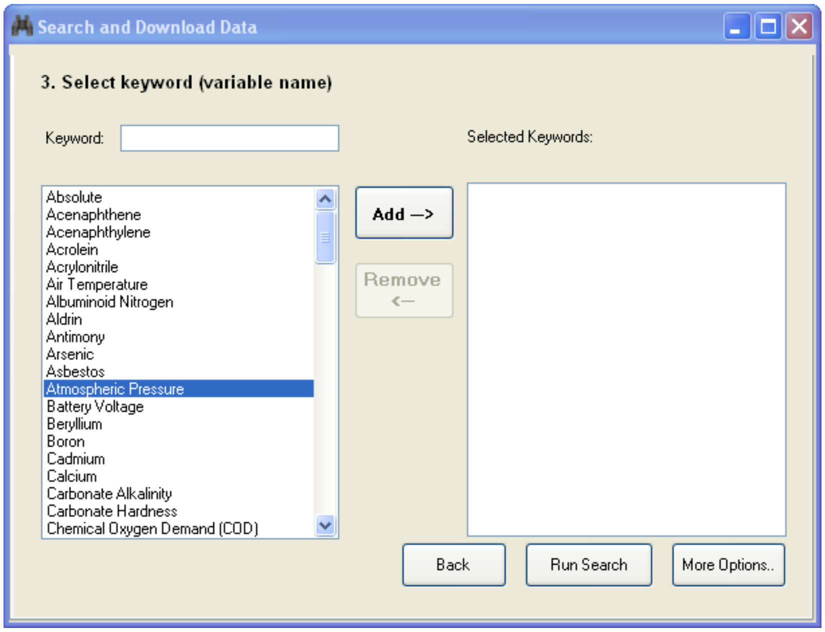


Figure 2-4 Third Data Discovery Form

The fourth form allows the user to enter a start date and an end date to serve as the temporal constraint on the query. This function is especially important to the purposes of the HIS project in that it deals with observational time series data as well as simple geospatial data.

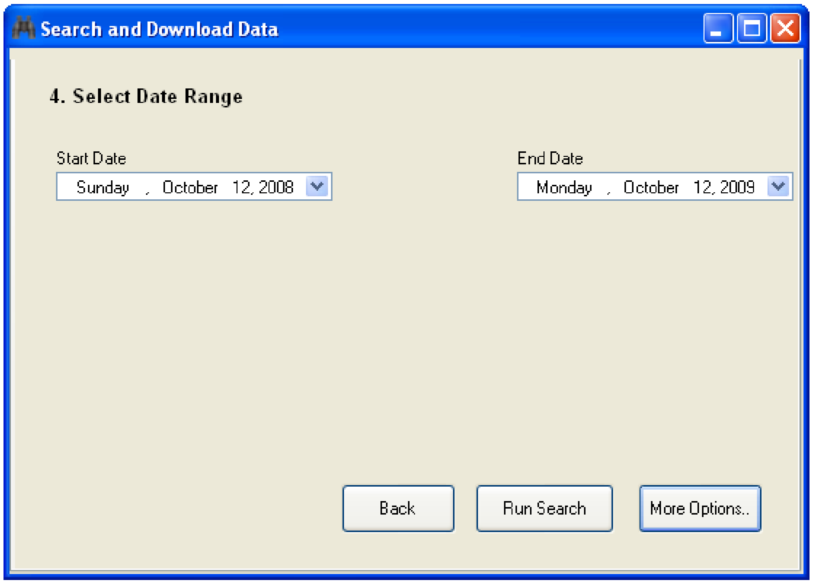


Figure 2-5 Fourth Data Discovery Form

The fifth form provides a list of web services to include in the search. This is a user-specified subset of the web services registered at HIS Central that limits search results to only a selected set of services. The user may also elect to search through the entire catalog of registered HIS Central web services.

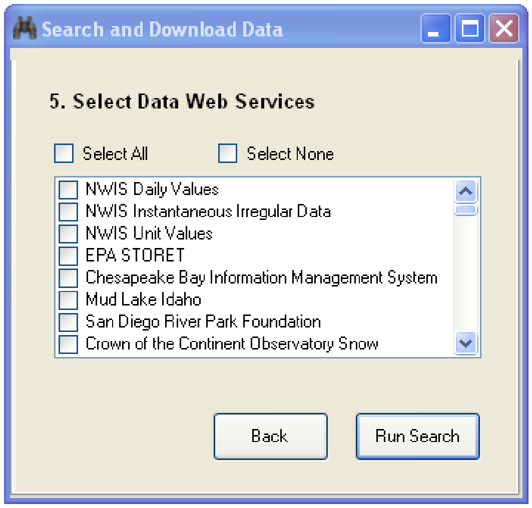


Figure 2-6 Fifth Data Discovery Form

All of the data series cataloged at HIS Central that meet the search criteria are returned in a format that allows the user to easily streamline and store the results of the query. Once the search is finished, the user can view the results plotted on a map or in table format. Figures 2-7 and 2-8 show the results of the data discovery process as it appears in these two different views. This process will be discussed in the next section.

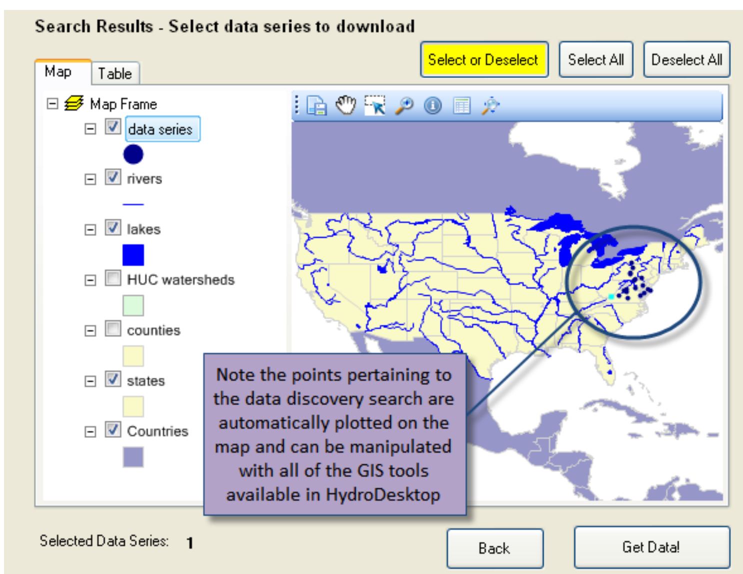


Figure 2-7 Data Discovery Results in Map Form

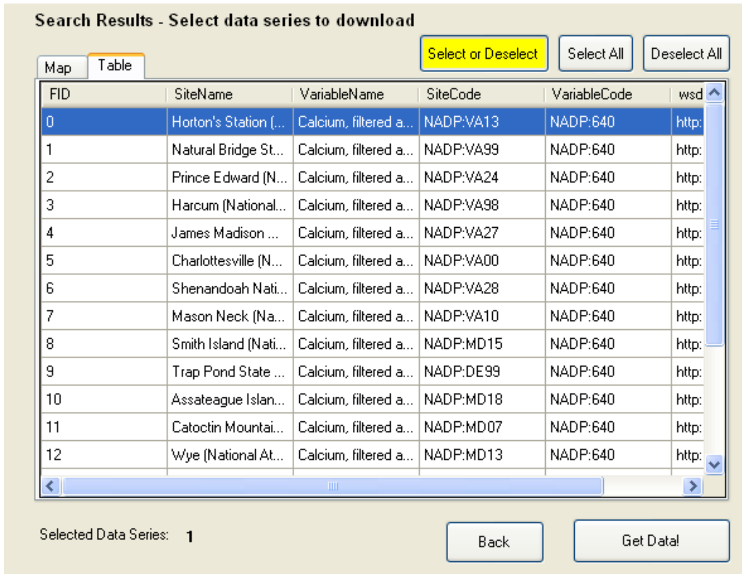


Figure 2-8 Data Discovery Results in Table Form

The data discovery forms allow the user to access data which is supported by HIS Central metadata web services. HydroDesktop also provides the capability to access web services that have not been registered with HIS central and add them the catalog of searchable services. This option makes use of datasets that are stored on individual or regional HIS Servers but not necessarily registered with HIS Central. Again, a simple button click is all that is necessary to initiate this function. The user simply clicks the *Metadata Fetcher* button in the *Toolbar* which opens the 'HIS Metadata Fetcher' window. This window allows the user to add new web service description languages (WSDL's) to the current HIS Central catalog simply by entering the URL for the unregistered web service. The desired service is then added to the list of searchable web services as seen in figure 2-6 and the user can then add the unregistered service to their data discovery search criteria.

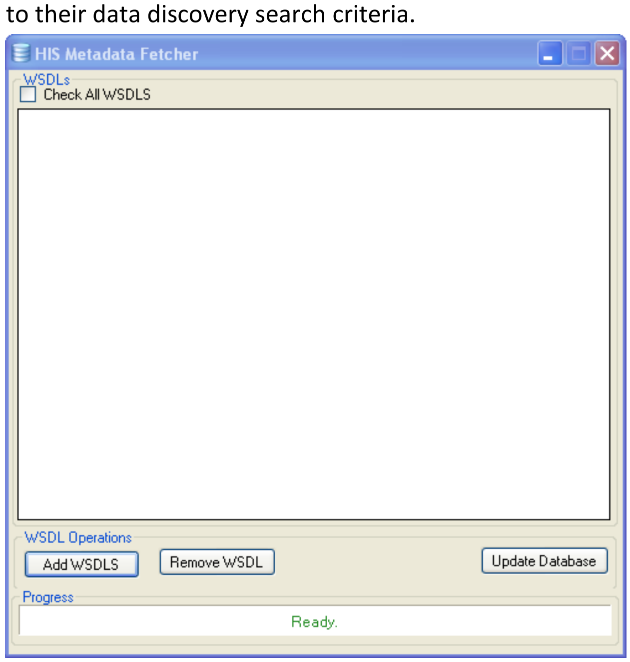


Figure 2-9 HIS Metadata Fetcher Window

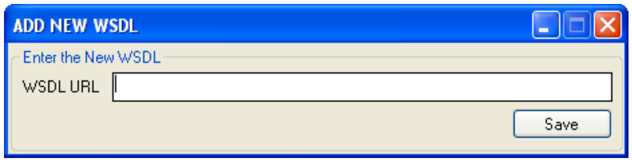


Figure 2-10 Add New WSDL Window

2.4 DATA DOWNLOAD, MANIPULATION AND SYNTHESIS

As an interactive GIS, HydroDesktop focuses on visualization. Once the data discovery process described above is complete the results are plotted as layers on the HydroDesktop map. HydroDesktop uses functions in the underlying MapWindow GIS system to provide users with the ability to visualize and manipulate a variety of vector, raster and image GIS data types. It also includes functionality for navigating the map as well as many other GIS tools and features. All of the functionality provided by MapWindow for visualization and manipulation of GIS datasets is available within HydroDesktop.

The map is presented in another window in the series of 'Search and Download Data' forms. From this new form, the user can operate the various GIS functions to continue to narrow down the selected data before moving on to the download process which is initiated simply by clicking the *Get Data!* button.

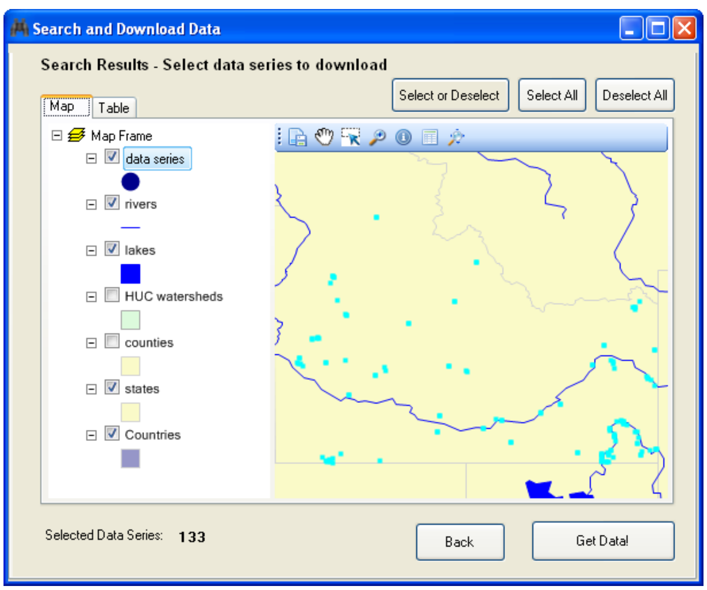


Figure 2-11 'Get Data!' Form

At this point the user is presented another form which allows them to assign the chosen data to a new or pre-existing user defined "theme." Also the user is able to establish how to handle duplicate data where the new data overlaps with previously downloaded data sets.

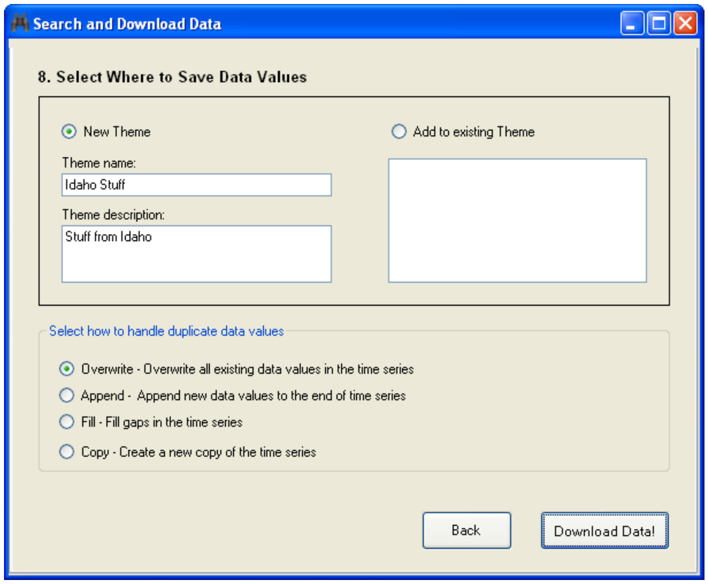


Figure 2-12 'Download Data!' Form

Once observational data have been retrieved and downloaded, HydroDesktop provides users with tools for visualizing and analyzing the data. A GIS data layer showing the locations of the data sites is added to the HydroDesktop *Legend* and displayed in the *Map Display*. The user can then manipulate the layer's symbology, properties and label set up among many other features. Also, the user can switch to 'Table View' using the *'View' Tabs* and quickly access attribute tables pertaining to each data site.

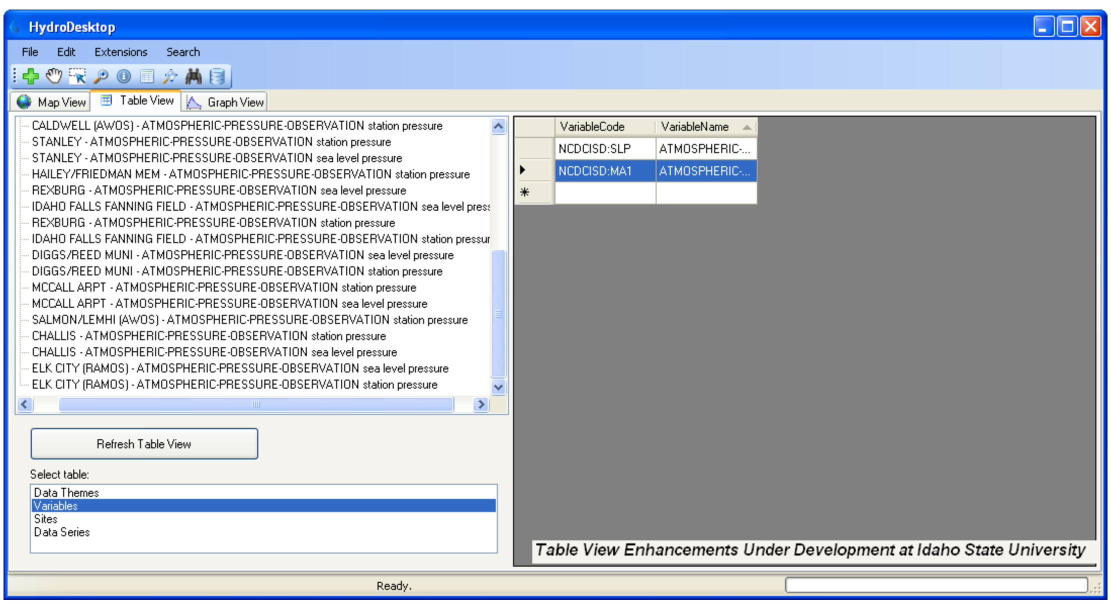


Figure 2-13 HydroDesktop Table View

2.5 ADDITIONAL PLUGINS

The tabbed interface and plug-in architecture presented in HydroDesktop have been developed such that third party programmers and researchers can relatively easily add functionality to the software by programming custom code and compiling it to a separate dynamic link library (DLL) that is "dropped" in an appropriate folder and is recognized by HydroDesktop. This plug-in architecture will allow for a distributed development effort both during the active development phase as well as in the future.

At present, Utah State University is actively engaged in developing the graphing and time series visualization tools that will appear under the "graph" tab. This is being done using the plug-in interface to both test the approach and also to simplify and separate that programming effort from programming of the core system. This has also allowed the USU team to build their plug-in using the VB.NET programming language instead of C# which HydroDesktop primarily uses, hence allowing for significant code reuse from their other existing projects. An alpha version of the Time Series Analyst for HydroDesktop plug-in will be available at the end of 2009.

The OpenMI based HydroModeler plug-in under development at University of South Carolina is another example of this distributed software development approach. Figure 2-14 shows the HydroDesktop interface with the added OpenMI plug-in. For more information on this plug-in refer to chapter 6.

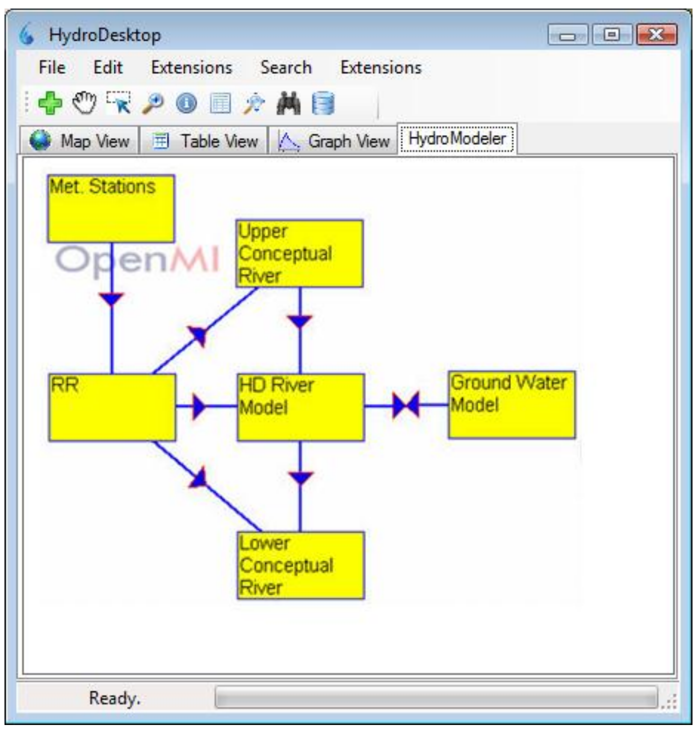


Figure 2-14 HydroDesktop with OpenMI based HydroModler plugin

2.6 OPEN SOURCE COMMUNITY DEVELOPMENT

HydroDesktop is under active development and a community of users/developers is growing - largely from the developer ranks of the CUAHSI project – but with growing interest from third party researchers. The source code for HydroDesktop is maintained at http://www.hydrodesktop.org which is based on a community open source software development portal maintained by Microsoft. Also available at this site is a download for the 0.1.3568 Alpha Installer for the software that has full HIS Central search capabilities. The community web site also includes discussion forums, a bug tracking system and a documentation Wiki. We have endeavored to make have a relatively low bar of entry for participation in the open development community by making the source code and bug tracking system and forums open access for all to see, with various levels of authentication required for posting to the site.

Also, recognizing the value of community based code development as a means of ensuring end-user adoption, this project has adopted an "iterative" or "spiral" software development approach where 1) the general project requirements and hard boundary conditions are specified at the outset (this requirements document has been posted on the community portal); 2) an initial brief functionality requirements list is developed; 3) the initial limited system is produced primarily by the core funded developer team, but with voluntary external programmer support as it becomes available (ongoing development can be seen in the form of repository commit notifications at the web site; 4) testing and bug fixes by the developer team; 5) deployment of an installation package for end-users (a first installer is now available); 6) collection of bug notices and feature requests from end-users; 7) identification of specific bugs and features to be addressed in a new release; 8) addition of these features by the developer team, etc.

This development approach is the most common approach used by open source projects because of its flexible and dynamic nature. This model is well suited to a community project where it is difficult (and often not useful) to fully-specify the functionality set required for a software release. Rather it is desirable to maintain an open structure that can easily be extended through the development of third party plug-ins to support as-yet unknown functions and capabilities, as well as a clear policy on how code is moved into the core system, and how external developers are included in the developer team.

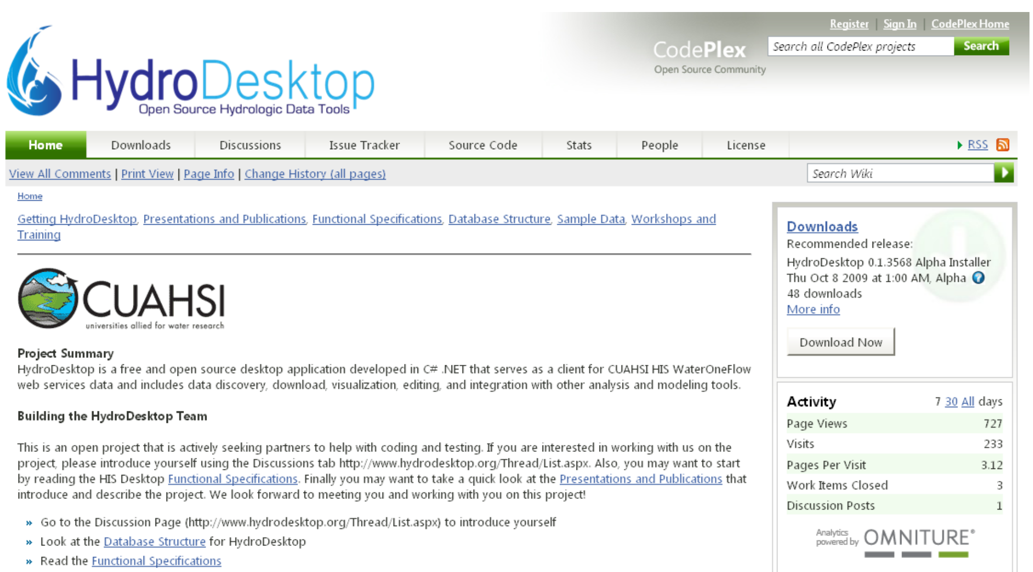


Figure 2-15 HydroDesktop Home Page

The HydroDesktop development team set out to create an HIS tool for obtaining, organizing and managing CUAHSI HIS data on a local user's computer with emphasis placed on modeling and analysis. This is being achieved by tailoring the open source MapWindow GIS toolkit and other existing software to suit the needs of the CUAHSI HIS, creating a series of user friendly tools and forms that focus around data discovery, download and manipulation, and establishing an online open source software development community to update and maintain the software.

Complete information on HydroDesktop is found at http://www.hydrodesktop.org.

Chapter 3. HIS CENTRAL AND CUAHSI WATER DATA SERVICES

By Ilya Zaslavsky, Tom Whitenack, and David Valentine, San Diego Supercomputer Center

The main focus of the SDSC team of CUAHSI HIS continued to be development of robust, governed and maintainable Hydrologic Information Infrastructure, capitalizing on the rapid software development in the first two years of the project. HISCentral and CUAHSI Water Data Services are core components of this infrastructure. This chapter reviews last year's progress in both components, while background information can be found in Chapters 3 and 6 of the 2008 HIS Overview http://his.cuahsi.org/documents/hisoverview.pdf.

3.1 WATER DATA SERVICES

The principles and composition of the CUAHSI HIS service oriented architecture are described in detail in the 2008 Overview report. In brief, CUAHSI WaterOneFlow web services comprise two types of functions: metadata services and data services. Metadata services (GetSites, GetSiteInfo, GetVariableInfo) specify the list of observation sites, the variables measured at those sites, the period of record for each variable and details about the units, method of measurement of the variables, and the source of the data service. The data service (GetValues) provides time series of observations data, one series at a time, where a series is defined by the service that provided it, the site at which the data was measured, the variable the data describes, and the period of record of the information. A National Water Metadata Catalog is maintained at the San Diego Supercomputer Center, which contains a single record for each time series in each of the more than forty data services now registered in that catalog. The catalog now contains more than 8 million such metadata records, and is the most comprehensive inventory of water measurements in the United States.

Data services through the GetValues function are provided from wherever the data are stored. Generally, government agencies such as the USGS, Corps of Engineers and National Climatic Data Center write custom-programmed versions of the GetValues function on top of their existing water observations databases. This does not require any rearrangement of their internal data structure and just allows another way of publishing and accessing the data in parallel to access through their web site. Academic organizations and some state and local water agencies use the CUAHSI Observations Data Model to store their water observations data, and this supports all CUAHSI WaterOneFlow web service functions, both metadata and data.

Over the last year, significant effort was focused on adding new water data services and harvesting respective metadata catalogs, in particular from the National Climatic Data Center, NCDC, and the US Army Corps of Engineers. In September 2009 the USGS published a new WaterML-compliant GetValues service, for instantaneous data in the National Water Information System, to add to the USGS WaterML-compliant service for Daily Values published earlier. Instantaneous data are much more descriptive of hydrologic processes and more valuable for hydrologic research than are daily data. To date, instantaneous data are accessible only for the most recent 60 days but that period may be extended later. Catalogs for these and other CUAHSI Water Data Services have been regularly updated. The updated statistics and service descriptions are presented in section 3.2.1. Over the last year, the total number of hydrologic measurements available through the system increased by more than a factor of 10 -- from 342 mil to 4.3 billion, mostly by provision at Drexel University of an extensive set of information about precipitation in the Chesapeake Bay area provided through the HydroNexrad system from the University of Iowa.

3.1.1 NEWLY ADDED NATIONAL WATER DATA SERVICES:

The following list of water data services have been added or updated in the last year at HIS Central. These are all significant because they required collaboration between the CUAHSI team and the agency or university involved. Each represents a custom implementation of the WaterML/WaterOneFlow web service specification to provide access to important hydrologic data sources.

1) **USGS NWIS Daily Values and Instantaneous Values services:** These services have been coded to take advantage of the GetValues REST services published by USGS, which are WaterML-compliant. Both USGS services were developed in collaboration with the SDSC CUAHSI team. The instantaneous service provides real-time access to 15 minute data for over 11,000 stream gauge sites.
2) **NCDC Integrated Surface Daily and Integrated Surface Hourly services:** have been developed to take advantage of the respective GetValues services published by NCDC. As with USGS, both services are WaterML-compliant, and have been developed with assistance from the SDSC team. The NCDC services provide hourly and daily weather data for 25,000 sites, world-wide.
3) **The Army Corps of Engineers RiverGages service:** this is another WaterML-compliant service recently published by the US Army Corps of Engineers and developed in consultation with the CUAHSI team. The RiverGages service provides river gage data for 2200 nationwide sites, which record data for over 40 different variables.
4) **HydroNEXRAD:** This service is hosted at the Drexel University and offers hourly precipitation data across a dense matrix of 215,359 sensor points from south western Virginia to New Hampshire. The SDSC CUAHSI team worked with Drexel through several attempts to harvest the metadata until we were finally successful. With 3.7 billion values, this service represents the largest number of data values in our system by far and was an excellent test of the scalability or our system.

3.1.2 THE CURRENT STATUS OF WATER DATA SERVICES

In addition, a range of new water data services for academic projects and for state agencies have been developed and added to HISCentral, by different members of the CUAHSI team working in collaboration with state and local data management and research groups. Some of the more significant of them include: services developed for Texas Water Development Board, HydroNEXRAD data service at the University of Iowa, the NOAA National Weather Service Advanced Hydrologic Prediction Service, and the National Atmospheric Deposition Program data. Table 3-1 provides summary statistics for hydrologic data available via Water Data Services as registered in HISCentral (see table 4-3 from the 2008 HIS Overview report for comparison). Services implemented by US Federal Agencies are highlighted at the top. The number of data accessible through CUAHSI water data services has increased by more than a factor of ten, from 342 million to 4.3 billion, due mainly to the HydroNexrad data from Drexel University described above.

Table 3-1 Statistics by organization of data available from web services cataloged in HIS Central. Services highlighted in brown are provided by federal water agencies

Network Name	Site Count	Value Count	Earliest Observation
NWISDV	29894	274762525	1/1/1900

Network Name	Site Count	Value Count	Earliest Observation
EPA	362645	94349967	1/1/1900
NWISUV	11185	77935488	60 DAYS
NCDC ISH	11555	3000000*	1/1/2005
NCDC ISD	24770	18165478	1/1/1892
NWISIID	369148	15501245	1/9/1867
NWISGW	827200	8325633	1/1/1900
RIVERGAGES	2206	263101295	1/1/2000
HydroNEXRAD	215359	3773019600	1/1/2007
LBL	24	6439473	5/23/2006
CIMS	894	5418382	7/2/1949
LittleBear	13	5382506	4/27/2005
MPE	9308	5010626	1/1/2005
ODMDCEW2	65	4631781	12/22/1998
LittleBearRiver	12	4040614	4/27/2005
IIHRNexrad	142	3356312	1/1/2006
SRBHOS	5	2182460	1/1/1997
NADP	19	1713885	7/25/1978
Sevilleta	41	1642555	7/1/2006
SFe-SRGWL	1089	998718	8/29/1905
SFe_GWL_USGS	190	960108	1/21/1933
MudLake	5	940977	5/25/2007
BaltimoreGW	7	333045	3/13/2008
SFe_SWFGWL	118	328854	3/26/1935
IIHRTippingB	3	275017	6/1/2006
SFe_CTDSondes	6	262347	6/8/2007
IIHRTippB	4	253251	6/1/2006
SFe_MICROWAVECITRA	1	219116	3/24/2004
ParadiseCreek	25	213267	10/1/2001
TWDB_Sondes	74	182697	8/3/1987
CCBay	98	105494	10/22/1987
TWDB_Wind	43	86408	8/1/1987
BaltOD	12	85952	11/14/2000
IIHRWQ	3	84631	8/31/2007
TWDB_Tides	105	65184	8/1/1987
Australian_ODM	1662	59088	2/28/2000
BESOD	11	42972	10/15/1998
COTCsnow	2	37734	6/24/2007
BaltPrecip	5	34238	11/14/2000
MAST	22	25480	6/13/2001

Network Name	Site Count	Value Count	Earliest Observation
SFe_ISUS	1	20664	6/15/2007
SantaFe-RainDlySRWMD	11	14371	2/29/2000
SFe_YSI	1	9288	7/12/2007
TxSt_Blanco	10	5683	10/1/2003
LakeChamplain1	5	4775	5/7/1992
SFe_Storet	17	4484	2/14/1990
SBRP	32	3008	7/9/2007
SDRPF	18	2424	10/14/2004
SFe-FLStoret	17	1730	2/14/1990
TIFP_LowerSabine	171	3670	5/16/2006
TIFP_LowerSanAntonio	162	3137	3/27/2006
Sonora	4	22	1/8/2001
Totals (Oct 2009)	1,867,108	4,336,790,286	
Totals (Aug 2008)	1,754,974	342,680,556	

* estimated

3.1.3 Water Data Service Usage statistics

A useful metric to measure the usage of the Water Data Services is to know the number of times a GetValues request has been processed. Table 3-2 below illustrates that for the last 11 months of 2008, 135,363 GetValues requests were processed for all CUAHSI HIS Water Data Services. Of these, 34,808 (25%) of those were to the USGS NWIS services. Already the total number for 2009 has increased by a factor of three to 407,781 total GetValues requests, with the USGS NWIS services processing the majority, increasing nearly ten-fold from their usage in 2008.

GetValues Requests	All Services	NWIS
February to December 2008	135,363	34,808
January to October 2009	407,781	298,093
Total:	543,144	332,901

Table 3-2 GetValues request received for two similar periods of time for 2008 and 2009.

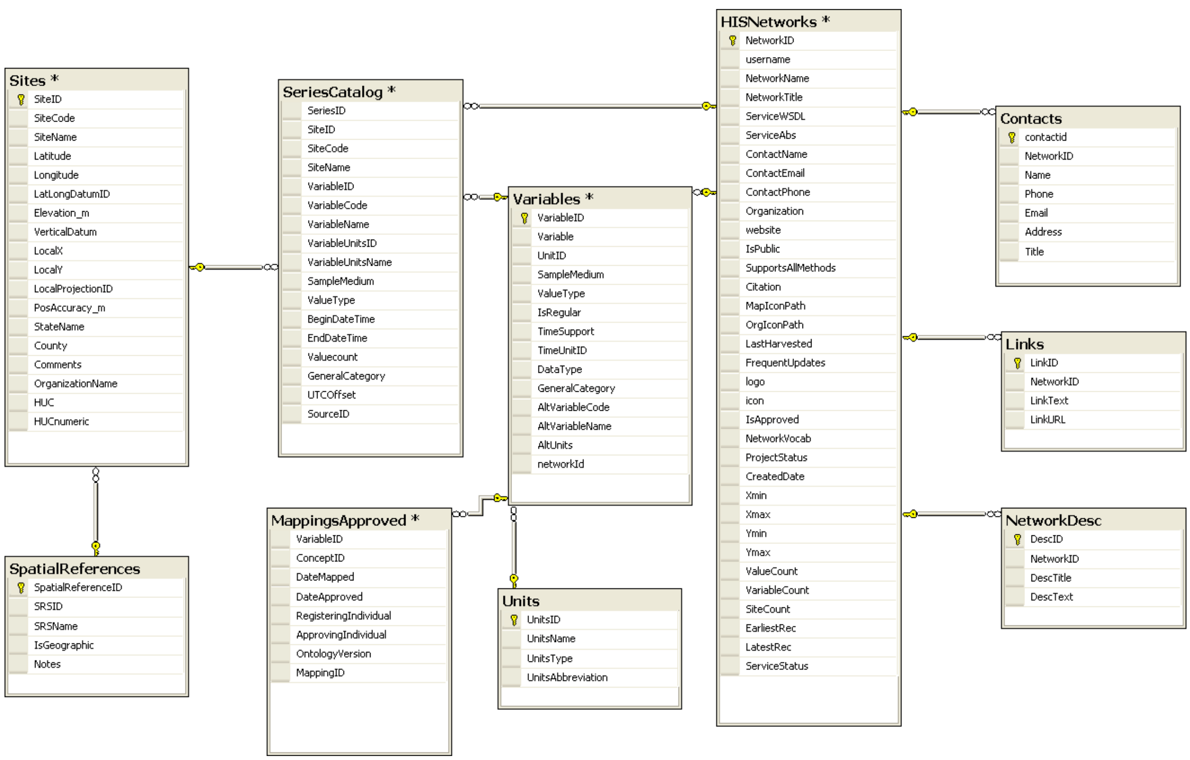
3.2 HIS Central Metadata Catalog

The HIS Central application manages the national registry of water data services and the national metadata catalog, and enables data managers to register their water data services in the central registry. In addition, HISCentral supports data discovery services over the central catalog. The main focus of HISCentral development in the last year has been making the registration of water data services easier and more complete, and making the content of the metadata catalog available via HISCentral web services, so that it can be queried from various HIS applications, in particular HydroDesktop.

3.2.1 National service registry and metadata catalog

The water data service registry and the national water metadata catalog store information about each registered service. The metadata catalog hosts detailed service metadata that support map-based and semantics-based discovery of available time series. During the last year, the schema of the metadata catalog was finalized and documented (Figure 3-1). While sharing common semantics with the Observations Data Model, this schema includes service metadata (the “HISNetworks” table) and ontology mappings (“MappingApproved”), and is centered around SeriesCatalog. This structure demonstrated acceptable performance on ingest and select queries, and enables rapid semantics-aware search across multiple networks.

Figure 3-1 Core table schema of the Central Metadata Catalog



3.3 HIS Central Web Service

A set of HISCentral web services exposing the content of central service registry and metadata catalog, has been added this year, so that client applications have a standard and consistent way of accessing the water data service metadata.   The HIS applications currently being developed with hydrologic data discovery functionality (e.g. HydroDesktop) are taking advantage of this new mechanism. The service is available from **http://water.sdsc.edu/hiscentral/webservices/hiscentral.asmx**, and includes the following methods:

1) Geographic selection of Water Data Services, and Service metadata:
    a. **GetWaterOneFlowServiceInfo** (returns a list of registered services with service metadata as available in the HISNetworks table)
    b. **GetServicesInBox** (returns a list of registered services that overlap a given Lat/Long box)
2) Ontology methods:
    a. **GetSearchableConcepts** (Returns a list of all searchable concepts from the HIS ontology)
    b. **GetOntologyTree** (returns a tree of concepts in XML format, starting with the input concept as the root and including its child nodes).
3) Methods that enable resolution of search concepts to variables across water data services, and search for variables:
    a. **GetMappedVariables** (returns WaterML-formatted list of all variables with their metadata that are either mapped to or fall under the input ontology concept, across specified water data services).
    b. **GetMappedVariablesInBox** (returns WaterML-formatted list of all variables with their metadata found within a bounding box)
4) Methods for selecting sites based on various criteria:
    a. **GetSitesInBox** (returns WaterML-formatted list of all sites that fall within the bounding box, have variables that are mapped to or fall under the Ontology Concept, overlap the date range of interest, have a minimum number of data values, and are within the list of services)
    b. **GetSitesInService** (returns the same information as the previous method, for a specified service)
    c. **GetSitesInHUC** (returns the same information as the previous method, for a specified HUC)
5) Methods for selecting data series based on various criteria:
    a. **GetSeriesCatalogForSite** (returns a WaterML-formatted list of all data series available for the site that are either mapped to or fall under the input ontology concept, overlap the input date range, and have a minimum number of observations.
    b. **GetSeriesCatalogForBox** (returns the same information as the previous method, for all sites within a specified Lat/Lon box)
    c. **GetSeriesCatalogForHUC** (returns the same information as the previous method, for all sites within a specified HUC)
6) Additional metadata method, to enable search by HUCs:
    a. **GetHydrologicUnits** (returns a list of all USGS 8-Digit Hydrologic Unit Codes, including HUC Code and HUC Name)

All these methods are critical components in the interaction between HydroDesktop and other data discovery clients developed within the HIS project, and HISCentral components: the service registry, metadata catalog, and parameter ontology. Using these services, the client applications can implement ontology-based search, or other types of searches over the information stored at HISCentral. The service methods have been designed and implemented based on discussions with HydroDesktop and HydroViewer developers. The methods have been tested and shown to perform to specification.

3.4 Service Standardization Efforts

Further adoption of CUAHSI HIS infrastructure, its scaling to managing a larger number of water data services, and the needs of seamless integration with data from other domains, place additional standardization requirements on the service messaging schema adopted by CUAHSI HIS. Development and adoption of WaterML 1.0 created a prototype web service for water observations data, which led us into collaboration with Open Geospatial Consortium (OGC), the leading international community standards organization for geospatial data. The milestones of this emerging collaboration over the last year include:

- A Hydrology Domain Working Group was organized, as a joint activity between OGC and WMO (World Meteorological Organization) Commission for Hydrology, through the efforts of David Maidment of CUAHSI HIS, working in collaboration with OGC staff, CSIRO, and other interested groups
- A series of preparatory meetings with key OGC and WMO members during the early phases of the collaboration: at OGC Technical Committee meetings in Atlanta and Athens (over phone), followed by a visit to WMO (April'09), and the first two official meetings of the Hydrology DWG held in Boston (June'09) and Darmstadt (Sept'09)
- Ilya Zaslavsky of the CUAHSI HIS project at SDSC was elected co-chair of the Hydrology DWG at the Boston meeting. The other co-chairs are Ulrich Looser (GRDC/WMO) and David Lemon (CSIRO). The responsibility of co-chairs is preparation of meetings, development of the DWG's work plan, and organization of joint activities within the group.
- The focus of the HDWG is on standardization of hydrologic data exchange, building on earlier schemas developed by CUAHSI HIS (WaterML 1.0), CSIRO (WDTF), and other organizations, and on emerging OGC specifications such as Observations and Measurements. The HDWG will focus on harmonizing these different specifications and analyze their respective use cases, to arrive at a common feature model, a common time series (observations) model, and an agreed upon semantic schema.
- A first international water data Interoperability Experiment (IE) is being organized now under the aegis of HDWG, with participation from USGS, Canadian Geological Survey, CSIRO, and CUAHSI HIS. The IE, focused on groundwater interoperability across the US-Canadian border, is now being considered for approval by the OGC Architecture board.

Further development of WaterML 2.0 and alignment of Water Data Services with this emerging specification and with other OGC standards, will take CUAHSI HIS to a new level of interoperability with data available from neighboring disciplines, and with applications and services available from both open source and commercial software projects. This is the first time that open source water data standards for the internet have been developed and it is a very significant development with global impact.

Chapter 4. HIS SERVER: A PLATFORM FOR PUBLISHING SPACE-TIME HYDROLOGIC DATASETS

By Jeffery S. Horsburgh, David G. Tarboton, and Kimberly A. T. Schreuders, Utah State University

HIS Server is a computer server that contains a collection of datasets, web services, tools, and software applications that allow data producers to store, publish, and analyze space-time hydrologic datasets. HIS Server is designed to permit local control of the data, while still being part of a distributed, national/international system allowing universal access to the data. In this chapter, we describe the HIS Server architecture and software stack, including tools for managing and publishing time series data for fixed point monitoring sites as well as spatially distributed, GIS datasets that describe a particular study area, watershed, or region. We then describe a case study implementation of HIS Server for publishing water resources related space-time datasets for a network of research watersheds in the northwestern United States called the Inland Northwest Research Alliance (INRA) Constellation of Experimental Watersheds (ICEWATER). The sharing of data in a common format is one way to stimulate interdisciplinary research and collaboration. It is anticipated that the growing, distributed network of HIS Servers will facilitate cross-site comparisons and large scale studies that synthesize information from diverse settings, making the network as a whole greater than the sum of its parts in advancing hydrologic research. Details of the ICEWATER network can be found at http://icewater.inra.org.

4.1 HIS SERVER DESIGN OVERVIEW

HIS Server was designed to enable producers of hydrologic data to join the growing, distributed network of published hydrologic data services that has been enabled by the CUAHSI HIS by establishing their own server and publishing their own data resources. The HIS Server software tools make the data available via Web services on the Internet and, when registered at HIS Central, the data become discoverable through HIS client applications such as Hydroseek and HydroDesktop. HIS Server includes a suite of software tools for local management of data on the server, as well as a set of Internet-based applications for accessing, visualizing, and downloading hydrologic data. Additionally, HIS Server enables users to publish spatial datasets for a study area or region and link them with observational datasets that have been created using the Observations Data Model (ODM) and the WaterOneFlow web services.

HIS Server introduces the concept of a study area, or “region.” A region is an area within which hydrologic data are collected. A region may be an experimental watershed, field site, or other geographic location. The region concept is important as an organizing principle within HIS Server as it provides a way for spatially distributed datasets describing a particular area, or region, to be associated with point observations data collected within that region. It also enables the capability to host data for several regions on a single HIS Server. Each region can be associated with one or more ODM databases and WaterOneFlow services that contain observational data for that region as well as one or more spatial data services that describe the region (e.g., a watershed boundary, stream locations, digital elevation model, etc.). This advances HIS Server to the point where it can be the platform for a complete Digital Watershed, describing both the spatial and temporal characteristics of a region.

4.2 HIS SERVER FUNCTIONAL COMPONENTS

Figure 4-1 shows an overview of the major HIS Server components. Time series observations made at fixed points (e.g., stream gages, weather stations, etc.) are loaded into one or more ODM databases. Each ODM database is attached to a WaterOneFlow web service that publishes the data contained within the ODM database on the Internet. Spatial datasets (e.g., watershed boundaries, stream lines, digital elevation models, etc.) are published as spatial data services using ArcGIS Server. A HIS Server database contains a listing of all services that have been published (both WaterOneFlow, and spatial services), as well as appropriate metadata to describe each service. A configuration tool has been built that enables HIS Server managers to edit the contents of the HIS Server database.

A suite of Internet applications connects to the HIS Server database and makes the data published on the HIS Server available on the Internet. These include: 1) an Internet map application that can be used to combine both spatial datasets and time series data published using ODM and WaterOneFlow services; 2) the Time Series Analyst, which was designed as a visualization window into an ODM Database, 3) a HIS Server website that presents information about the data services that have been published on the HIS Server; and 4) the HIS Server Capabilities Web service, which publishes the capabilities (i.e., a metadata description of all of the services that the HIS Server contains) of the HIS Server in a machine readable format. The following sections describe each of these major functionalities.

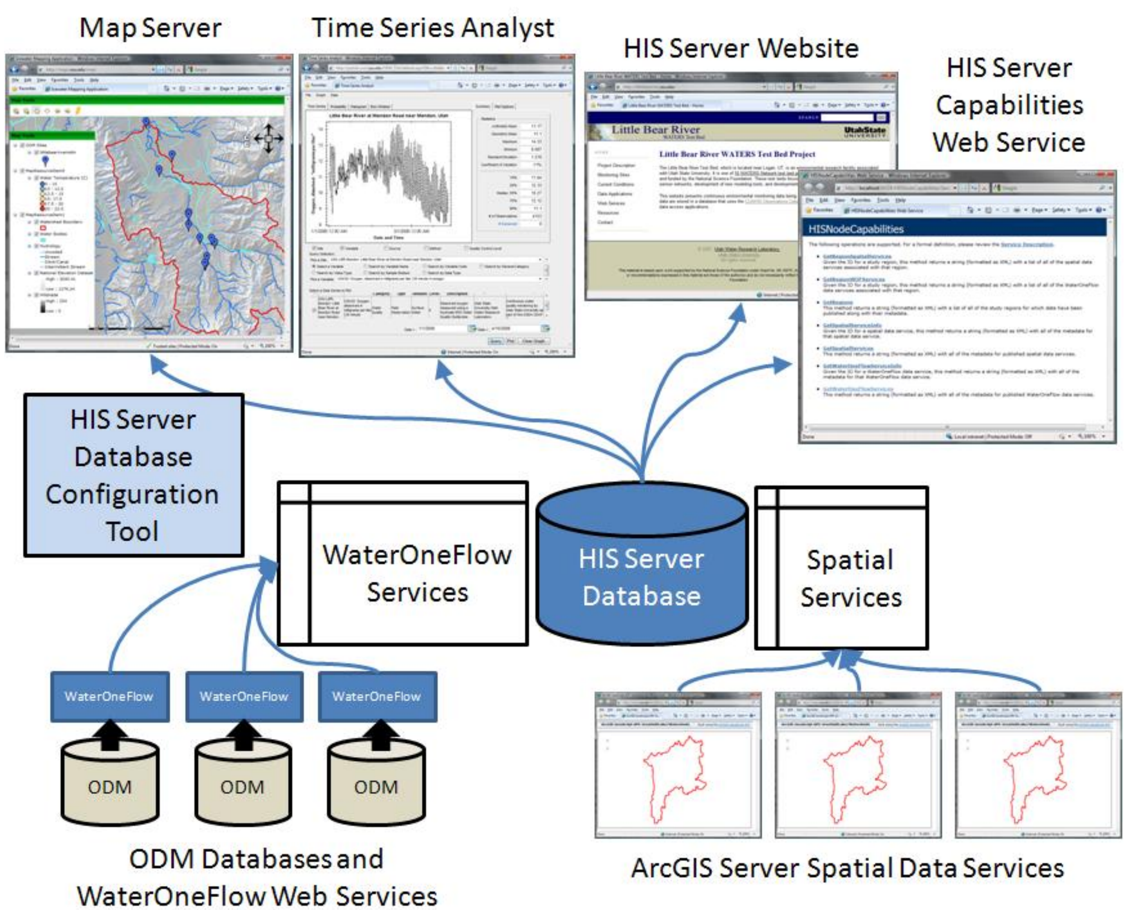


Figure 4-1 General overview of HIS Server components.

4.2.1 Publishing Point Observations with ODM and WaterOneFlow

ODM and the WaterOneFlow Web services serve as the basis for publishing point observations data on an HIS Server. Point observations data are loaded into an ODM database, which is then connected to the WaterOneFlow Web services. The WaterOneFlow Web services transmit the data over the Internet in Water Markup Language (WaterML) format. In addition to ODM and the WaterOneFlow Web services, HIS Server includes several utilities for working with ODM databases. ODM, WaterOneFlow, and the ODM utilities are described briefly below.

The Observations Data Model (ODM) – ODM is a relational data model for storing, managing, and manipulating point observations data. Each HIS Server implements one or more ODM databases, depending on the number of research watersheds/locations hosted by the HIS Server. All of the point observations data for each region will be entered into an ODM database, and a region may have one or more ODM databases. ODM provides the persistent storage mechanism for the data. Because ODM databases are implemented in a relational database management system (i.e., Microsoft SQL Server), they support a variety of applications through the use of SQL queries that can be passed to the database to retrieve and manipulate data. Additionally, once data have been entered into an ODM database, they can be published using the WaterOneFlow web services. Figure 4-2 shows the logical data model for ODM. The latest version of ODM is available at http://his.cuahsi.org/odmdatabases.html.

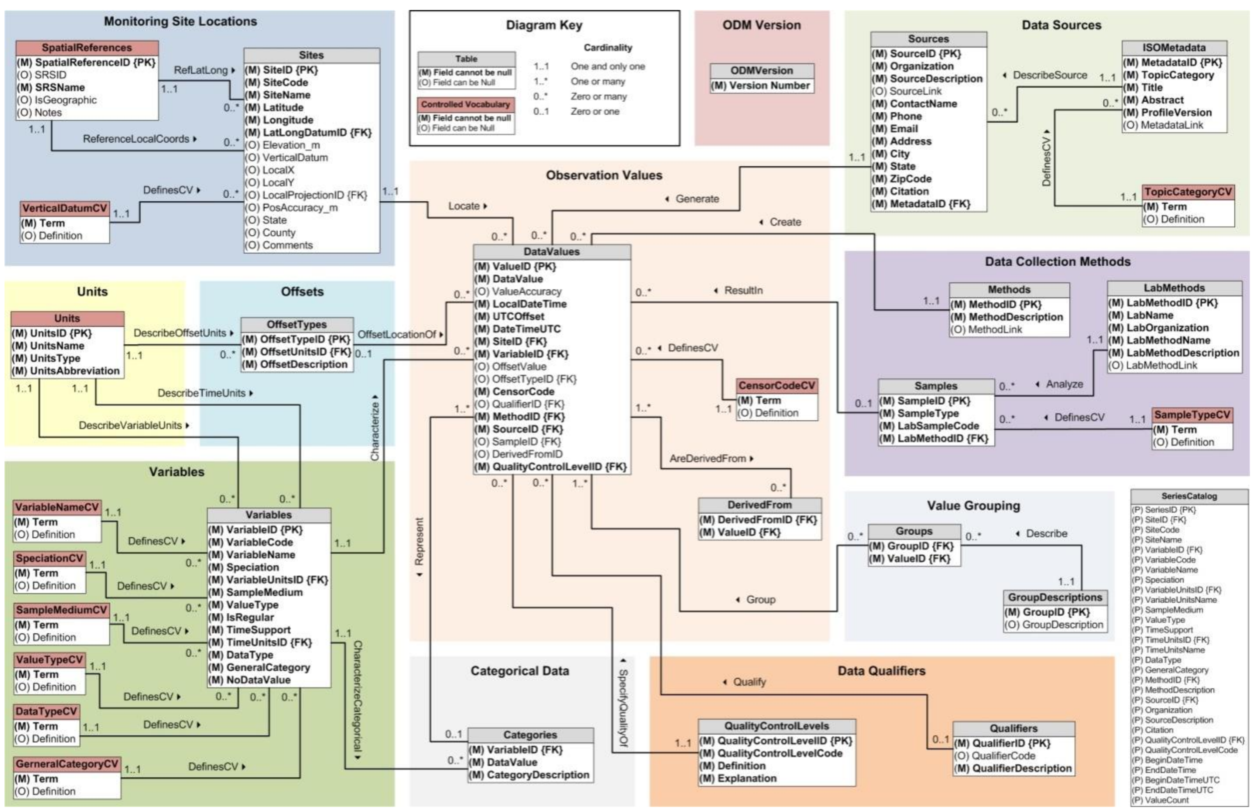


Figure 4-2 ODM logical data model. The primary key field for each table is designated with a {PK} label. Foreign keys are designated with a {FK} label. The lines between tables show relationships with cardinality indicated by numbers and labeled with the name and directionality of the relationship. Required (Mandatory) data fields are indicated with an M and are in bold, while optional data fields are indicated with an O.

The ODM Data Loader (ODMDL) – The ODMDL application was created to allow HIS Server administrators to load data into an instance of ODM. ODMDL accepts as input data in table format (Microsoft Excel or comma separated values .csv) that is sufficient that it can be loaded into ODM without violating any ODM constraints. ODMDL protects the security and consistency of an ODM database because it provides users with a set of tools for validating and loading their data into ODM. This minimizes the potential for human caused errors in loading these data into an ODM database. The ODMDL input file formats are similar to the table structures in ODM, but they do provide users with some flexibility in specifying the required metadata. Users do not need to perform any specialized programming to parse and load the data, and ODMDL ensures that the data are fully qualified with valid metadata when they are loaded. The latest version of ODMDL is available at http://his.cuahsi.org/odmdataloader.html,

The ODM Streaming Data Loader (ODMSDL) – The ODMSDL was designed for streaming continuously measured sensor data generated by a monitoring and telemetry system into an ODM database. Similar to ODMDL, ODMSDL is a file based data loader, but takes as input datalogger files that have a single date column and potentially multiple columns of data. ODMSDL provides simple visual tools for mapping streaming data files to the ODM schema and for specifying all of the required metadata, which means that users do not need to perform any specialized programming to parse and load the data and that the data are fully qualified with valid metadata when they are loaded. ODMSDL can be run manually or automatically as a Windows scheduled task. The latest version of ODMSDL is available at http://his.cuahsi.org/odmsdl.html.

ODM Tools –ODM Tools provides HIS Server administrators with a set of value added tools that they can use to better manage data stored in ODM. These tools are organized into three general areas: 1) query and export; 2) visualize; and 3) edit. The "Query and Export" functionality allows users to find the data that they are interested in and export it to a simple format that can be used with a variety of analysis software. The "Visualize" functionality allows users to quickly plot and summarize data using a variety of plot types and descriptive statistics. The "Edit" capability of ODM Tools was designed to provide users with a simple set of tools that they can use to edit existing data series (e.g., for performing QAQC on continuous data streams) and to create new data series from existing data series.

ODM Tools protects the security and consistency of an ODM database because it provides users with a set of automated tools for performing many of the most common database transactions. ODM Tools also allows users to export data from their ODM database with an accompanying metadata file. This allows users to work with local copies of data series exported from their ODM database while preserving the provenance of the data via the metadata file. ODM Tools also provides a mechanism by which users can interact with the ODM database without having to learn the complexities of its relational structure. Finally, for more advanced users, the source code of the ODM Tools application provides an example of how applications can be built on top of the CUAHIS HIS ODM. Figure 4-3 shows the graphical user interface of the ODM Tools software application. The latest version of ODM Tools is available at http://his.cuahsi.org/odmtools.html.

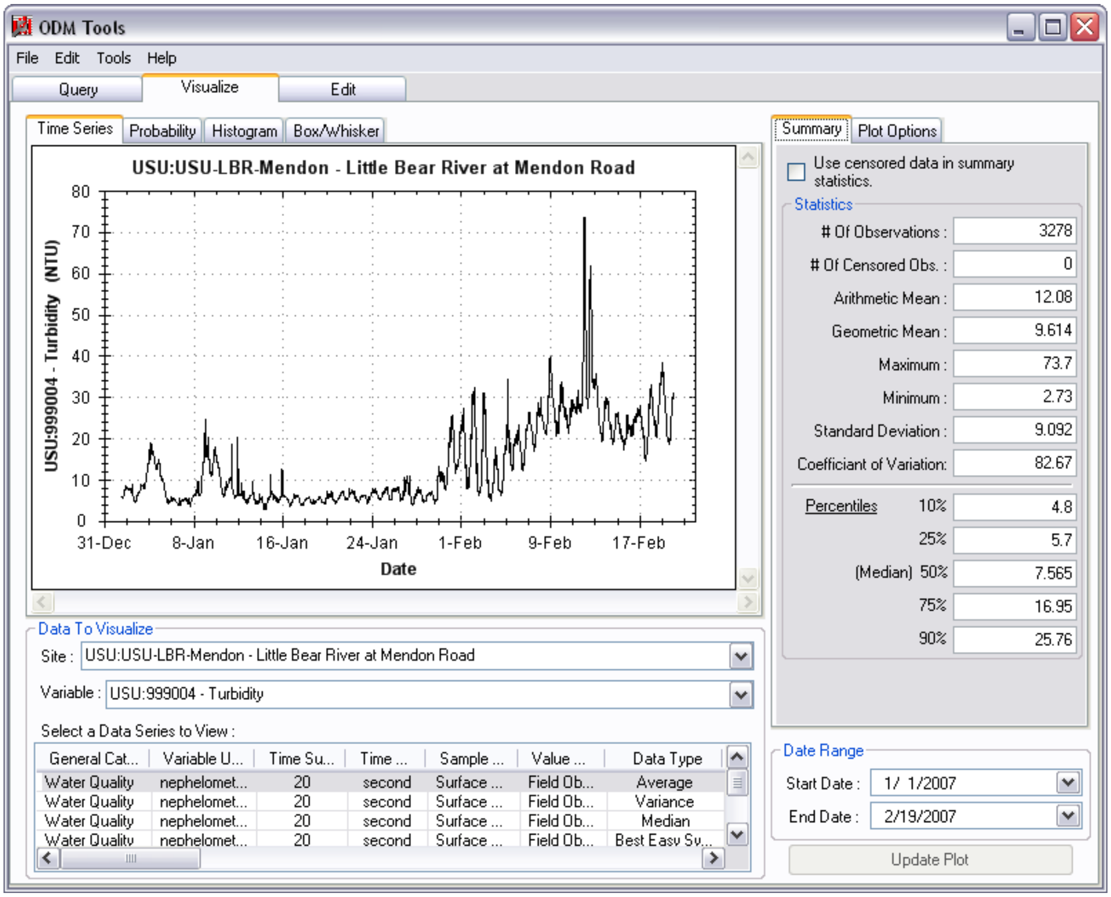


Figure 4-3 ODM Tools Visualize tab.

The WaterOneFlow Web Services – The WaterOneFlow Web services are designed to be implemented on top of an ODM database to publish point observations data on the internet. The WaterOneFlow Web services transmit data in Water Markup Language (WaterML) format and consist of a set of methods (i.e., GetSites, GetSiteInfo, GetVariableInfo, GetValues) that can be called from many different programming languages and software environments for retrieving data from an ODM database over the Internet. These methods have been implemented within a single Web application that is easily installed and configured. Each HIS Server can implement a set of WaterOneFlow Web services for each ODM database that contains data to be published. This ensures that all of the data published on any HIS Server are available on the Internet in a standard, interoperable (i.e., platform and programming language agnostic) format. In addition, WaterOneFlow Web services can be registered with CUAHSI HIS Central (http://hiscentral.cuahsi.org/), enabling the services to be discovered and consumed by central CUAHSI HIS applications like Hydroseek (http://www.hydroseek.org).

4.2.2 Publishing Spatial Datasets Using ArcGISServer

HIS Server relies on the publication of spatial datasets as Web services using Open Geospatial Consortium (OGC) standards. ESRI's ArcGIS Server is used as the engine for publishing these services and is capable of publishing vector data using the Web Feature Service (WFS) standard and raster data using the OGC Web Coverage Service (WCS) standard. Hybrid services that contain both vector and raster data can be published as image services using

OGC's Web Map Service (WMS) standard. The OGC standards are open formats that can be consumed by a variety of different GIS client software.

GIS datasets are inherently tied to a location or geographic extent, and it is for this reason that the concept of a study area or "region" is so important within HIS Server. The region gives context to the geographic extent that the spatial dataset describes. For example, a user might create a watershed boundary for the Little Bear River Experimental Watershed region, a digital elevation model for the T.W. Daniels Experimental Forest region, or a blue line coverage of streams for a region defined as the State of Utah. A region may have many spatial datasets that describe its characteristics.

Spatial data services can contain a single spatial dataset such as a watershed boundary or digital elevation model, or multiple spatial datasets can be combined to create a composite map service. This provides HIS Server administrators with flexibility in defining spatial data services that are appropriate for the regions that they are defining. It also enables them to reuse spatial data services for multiple regions – for example, a single map service that covers the whole United States might be used as the base map for multiple regions for which data are published on the HIS Server. Figure 4-4 shows a map service for the Little Bear River Experimental Watershed in which three layers have been published as a single service using ArcGIS Server.

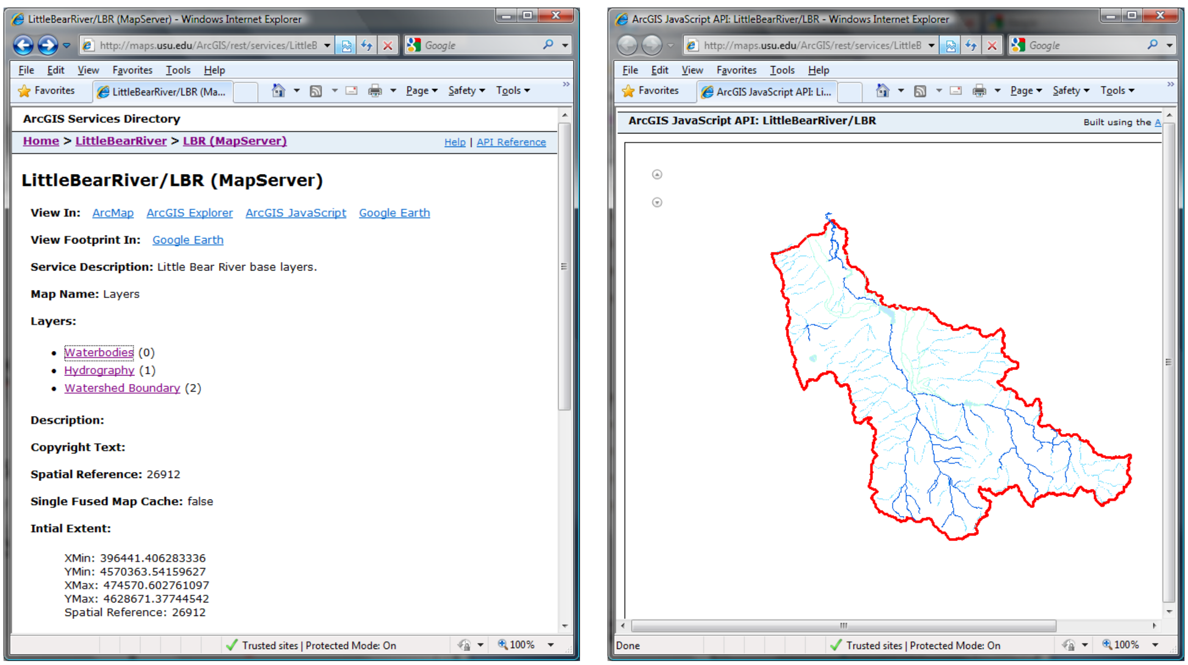


Figure 4-4 Composite map service for the Little Bear River Experimental Watershed with water bodies, streams, and the watershed boundary.

4.2.3 HIS Server Database and Configuration Tool

Each HIS Server has a database that catalogs the list of regions for which data have been published and the list of observational and spatial data services that have been published on the server. The database also contains metadata describing each region and service. The purpose of this database is to define the capabilities of a HIS

Server, and it is populated by the server administrator as new services are brought online. Figure 4-5 shows the database schema for the HIS Server database.

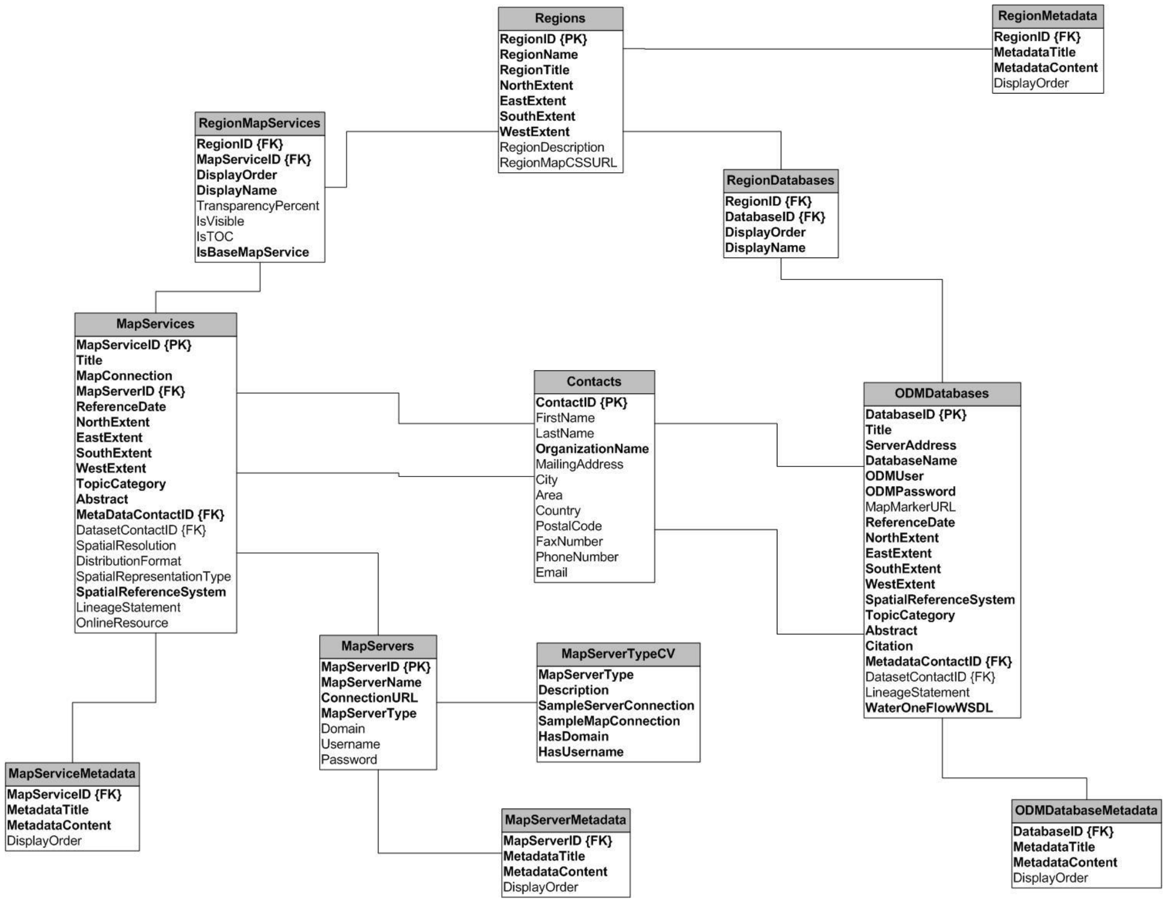


Figure 4-5 Database schema for the HIS Server database.

A configuration tool has been built that enables HIS Server administrators to define the contents of this database. It provides a simple forms based interface for HIS Server administrators to define regions and services and the metadata that describe each. The inputs to the forms are written to the appropriate tables in the database so that HIS Server administrators can edit the database through the application rather than editing each table in the database individually.

4.2.4 HIS Server Capabilities Web Service

Each HIS Server can implement a set of Web services that publish its capabilities as defined within the HIS Server database. The HIS Server Capabilities Web service includes Web methods that return in XML format the list of regions for with data have been published along with region metadata, the published point observations data services (based on ODM and WaterOneFlow) and appropriate metadata, and the list of published spatial data services along with appropriate metadata. By doing so, all of the capabilities of the HIS Server are published in a machine readable format that can be discovered by central registration and cataloging services (i.e., HIS Central). This is somewhat different than the previous model where users manually registered services with a central metadata catalog. With the implementation of the HIS Server Capabilities Web service, the HIS Server becomes self describing (i.e., a programmer or machine can discover all of the available capabilities of the HIS Server simply by calling the Capabilities Web service) and can be registered and harvested automatically.

The HIS Server Capabilities Web service connects directly to the HIS Server database and uses the tables that list the data services that have been published on the HIS Server. As HIS Server administrators define new regions or services within their HIS Server database, they are immediately available to the Capabilities Web service. The service has the following methods:

GetRegions – This method returns a list of all of the study areas or regions for which data have been published on the HIS Server.
GetRegionInfo – This method returns the metadata description for a selected region for which data have been published on the HIS Server.
GetWaterOneFlowServices – This method returns a list of all of the WaterOneFlow Web services that have been published on a HIS Server.
GetWaterOneFlowServiceInfo – This method returns a metadata description for a selected WaterOneFlow Web service published on the HIS Server.
GetMapServices – This method returns a list of all of the spatial data services that have been published on the HIS Server.
GetMapServiceInfo – This method returns a metadata description for a selected spatial data service published on the HIS Server.
GetRegionWaterOneFlowServices – This method returns the list of WaterOneFlow Web services that have been associated with a region.
GetRegionMapServices – This method returns a list of the spatial data services that have been associated with a region.

4.2.5 HIS Server Data Presentation and Visualization Tools

A suite of data presentation and visualization tools have been created for HIS Server. These tools are intended to provide a graphical user interface, or public interface, to the data holdings of the HIS Server. The following sections describe each software tool in more detail.

Internet Map Application – Each HIS Server can implement an Internet map server interface to the observational data services and the GIS data services that are published on that server. The map interface is a Web application that runs in a web browser and allows users to browse data that are published for a particular region. Each region can have its own map. The application is based on ArcGIS Server, and combines both spatial datasets and the locations of point observations in a single map (e.g., the map server plots the locations of monitoring sites and provides site information and links to the data when users click on a site on the map). The map server application dynamically generates its content using information contained within the HIS Server database. The map application is capable of presenting data from multiple published observational and GIS data services for a region, and the list of available services is generated from the HIS Server database, which can be edited by HydroSever administrators. When services are added to these tables, they automatically appear in the Internet Map Server application. Figure 4-6 shows the HIS Server map application.

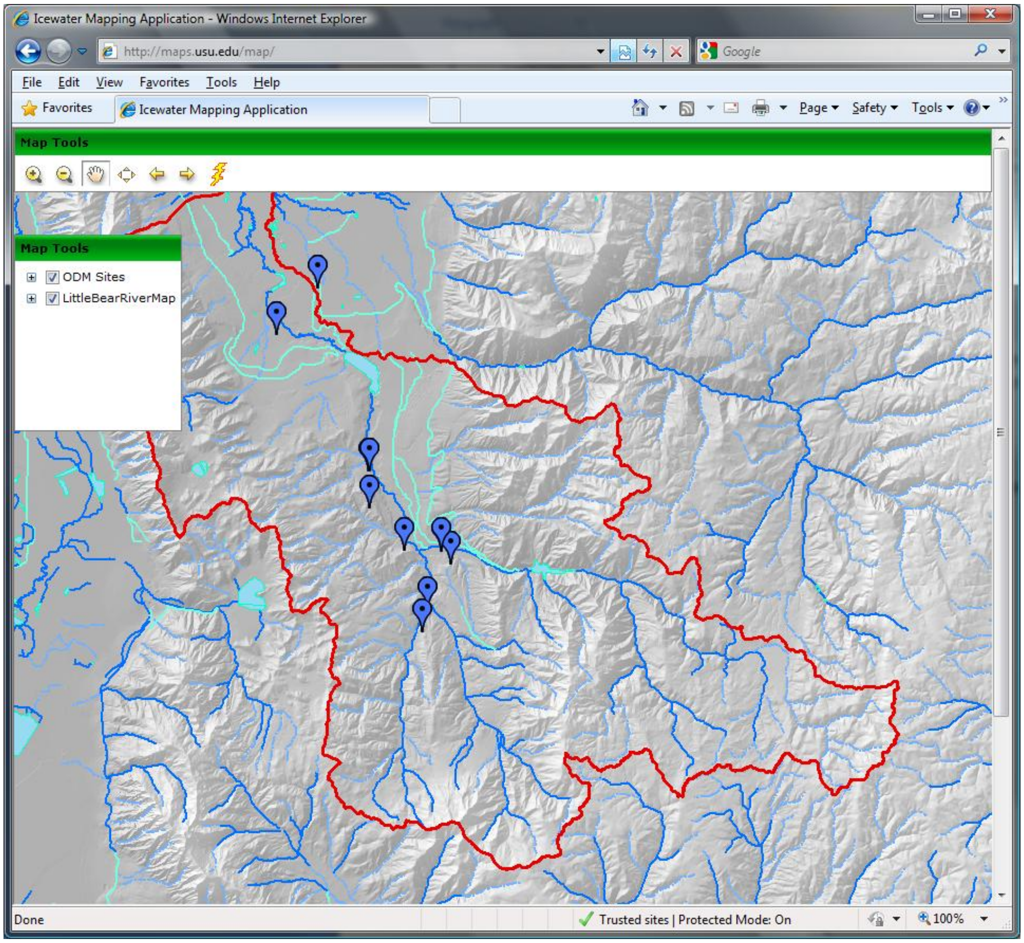


Figure 4-6 HIS Server map application for the Little Bear River Experimental Watershed region. In this instance, a single point observations data service (ODM Sites) has been combined with a single spatial data service (LittleBearRiverMap).

Time Series Analyst – Each HIS Server can host an instance of the Time Series Analyst application that enables users to visualize and generate descriptive statistics for selected point observations datasets. The Time Series Analyst is a Web application that enables users to screen/preview datasets prior to download so that they can make sure that the data are what they want. An example of the Time Series Analyst is available at http://tsa.usu.edu/odm_tsa/. The Time Series Analyst can be linked to the Internet map application so that when users click on a monitoring site on the map they are presented with a link to visualize/summarize the available data at the selected site using the Time Series Analyst. The Time Series Analyst is capable of connecting to any number of ODM databases hosted on a HIS Server, and has a unique calling interface that can be used to launch the application with a selected ODM database, site, variable, and date range. Figure 4-7 shows a screen shot of the Time Series Analyst.

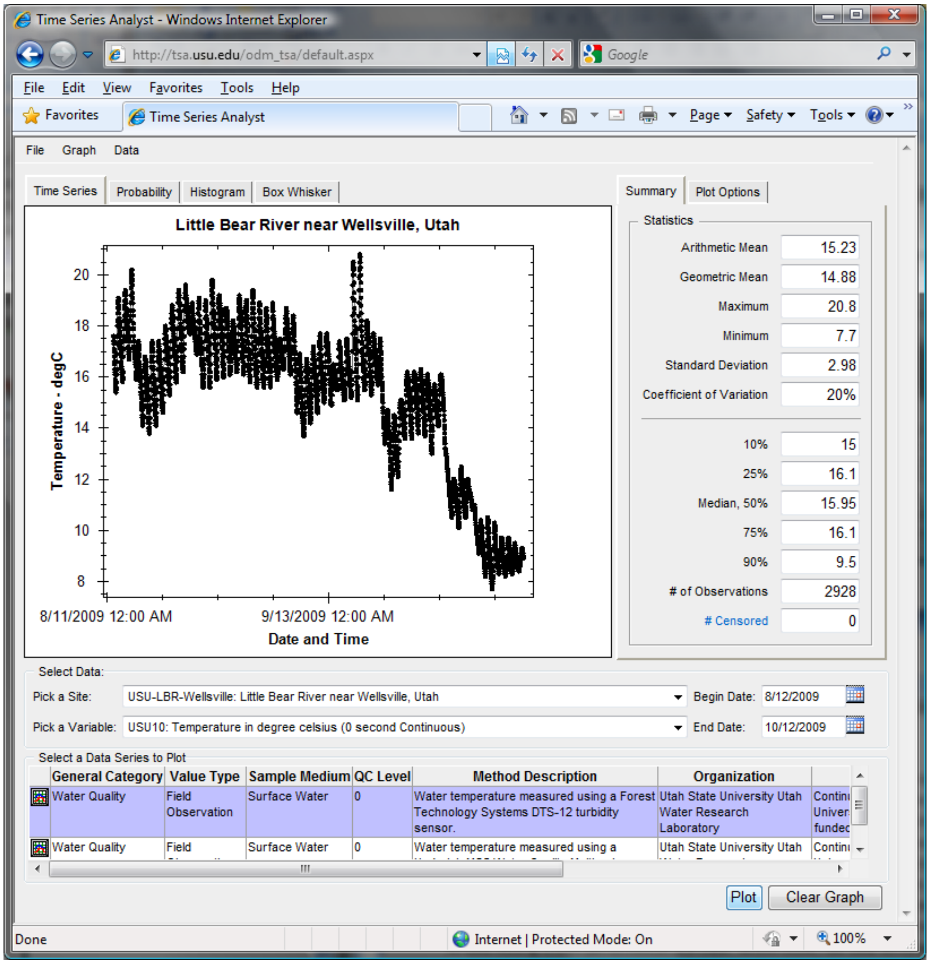


Figure 4-7 HIS Server Time Series Analyst application.

HIS Server Website – Each HIS Server can implement a website that provides information about the HIS Server and details for each published data service that resides on that HIS Server. The majority of the content of this website is dynamically generated from the HIS Server database and the database(s) that hold the observational data. This enables the website to be dynamically updated through adding content to the HIS Server database rather than editing the code of the website. The HIS Server website contains the following components/pages: 1) an overall/opening page that describes the HIS Server; 2) a dynamically generated listing of published point observations data services; 3) a dynamically generated listing of published spatial data services; 4) a page with links to the map application and Time Series Analyst; and 5) a data query and download page that will allow users to more easily query for data from one or more published observational data service and then download the data.

4.3 ICEWATER: Using HIS Server to Build a Regional Network of Published Data

Over the past several years, researchers at universities affiliated with the Inland Northwest Research Alliance (INRA) have been collecting water resources datasets at a number of experimental watersheds in the western

United States. Experimental watersheds in the INRA region span a number of climate, human development, and disturbance gradients, and researchers are investigating several different research themes, including snowmelt responses to climate change, groundwater - surface water interactions, modeling of hydrologic response, land use change, and arctic river processes. It is hoped that integration of data from these watersheds will facilitate cross-site comparisons and larger scale studies that synthesize information from diverse settings, making the network as a whole greater than the sum of its parts.

In this section, we describe the establishment of and support for the INRA Water Resources Consortium Constellation of Experimental Watersheds Information System Network (ICEWATER). ICEWATER is a distributed network of HIS Servers, that is being used to publish and integrate the data holdings from each INRA university. This network is illustrated in Figure 4-8 and comprises HIS Servers at Boise State University, Idaho State University, Montana State University, University of Montana, University of Alaska Fairbanks, University of Idaho, Utah State University, and Washington State University. Utah State University hosts ICEWATER Central that provides centralized support for this regional HIS network. The goals of ICEWATER are: 1) establishment of a common information system for data sharing, analysis and archiving, building upon and extending the CUAHSI HIS; 2) establishment of a common modeling framework to facilitate sharing and model interoperability; and 3) establishment of common base characterization datasets such as digital elevation models (DEMs) from LIDAR, land cover and land use from remote sensing, for research sites within the ICEWATER network that provide detail beyond nationally available information.

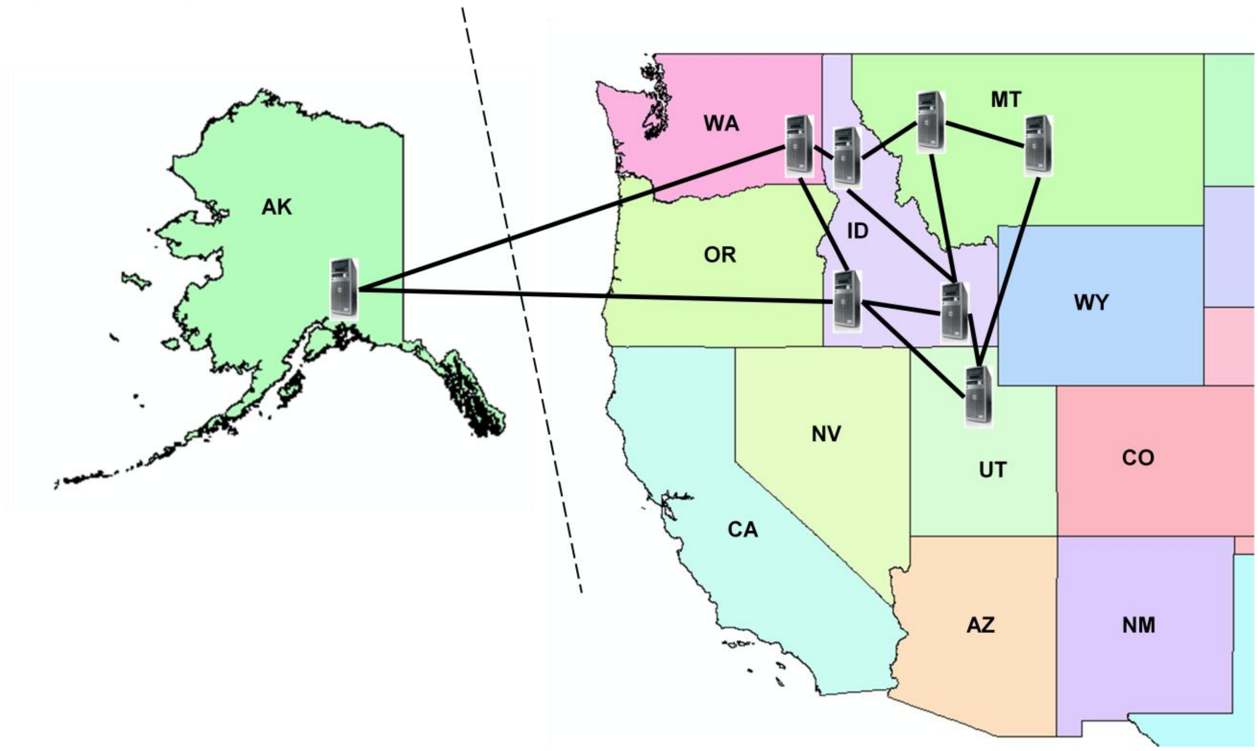


Figure 4-8 INRA Constellation of Experimental WATersheds (ICEWATER) Network.

4.3.1 ICEWATER Implementation

ICEWATER is an ongoing effort that uses the CUAHSI HIS software to build a regional HIS that is also part of the national network of data published using HIS tools. A HIS Server has been established at each INRA university, and on each HIS Server the data holdings of one or more regions (e.g., experimental watersheds) are being published. Each ICEWATER HIS Server is in the process of implementing one or more ODM databases connected to WaterOneFlow web services, spatial data services published using ArcGIS Server, a HIS Server database and Capabilites Web service, and the whole suite of data access, presentation, and visualization tools described above. As WaterOneFlow services are coming online, they are being registered with HIS Central so that they become part of the national HIS data network.

4.3.2 ICEWATER Central: Supporting a Regional HIS

On additional component of ICEWATER is functionality that is referred to as ICEWATER Central (http://icewater.inra.org), where support is provided for HIS Server administrators and data managers at each of the INRA Universities. ICEWATER Central consists of personnel who provide support for HIS Server administrators and data managers via telephone and via a group email listserv, and a website with the following components: 1) general information about INRA and ICEWATER; 2) resources for HIS Server administrators and data managers, including a listing of all hardware and software needed to establish a HIS Server; and 3) a catalog of and presentation of the published data services from each ICEWATER HIS Server.

The ICEWATER Central functionality is not intended to replace functionality of the HIS Central website. Rather, it is intended to give the ICEWATER Network an identity that is separate from the CUAHSI HIS, but in a way that is still entirely compatible with the national network of published data that is enabled by the CUAHSI HIS. Because each ICEWATER data manager is registering their services with HIS Central the entire contents of the ICEWATER regional HIS will be available through HIS Central and can be discovered and accessed using HIS tools like HydroDesktop and Hydroseek.

4.4 Summary

In this chapter, we have described the CUAHSI HIS Server, which is a computer server that integrates collections of point observations and spatially distributed GIS datasets, Web services, and software applications to enable data producers to store, publish, and analyze space-time hydrologic datasets. HIS Server was designed to permit data managers affiliated with universities, research watersheds, experimental sites, etc. to establish their own server and publish their own data resources. This permits local control of the data, and through registration with HIS Central, each HIS Server can become part of a distributed, national/international system that enables universal access to the data.

The HIS Server software stack includes ODM, the WaterOneFlow Web services, ArcGIS Server for publishing spatial datasets, a HIS Server database and Capabilities Web service, and a suite of data presentation and visualization tools that includes a map application, Time Series Analyst, and a HIS Server website. Also included are a set of utilities for managing and working with time series data for fixed point monitoring sites in ODM, including ODM Tools, ODMDL, and ODMSDL.

We have also described how HIS Server can be used to build a regional HIS that has both a regional identity and participates in the national network of published hydrologic data that is enabled by the CUAHSI HIS. Indeed, HIS Server is the basis for the INRA ICEWATER network where a HIS Server has been established at each INRA university and where data for a variety of experimental watersheds within the western United States are now coming online, whereas before they were confined to the files of individual researchers or available on the Internet in inconsistent formats. It is anticipated that the growing, distributed network of HIS Servers, such as the ones being built within the ICEWATER network, will facilitate cross-site comparisons and larger scale studies that synthesize information from diverse settings, making the network as a whole greater than the sum of its parts in advancing hydrologic research.

Chapter 5. TEXAS HIS

By Tim Whiteaker, University of Texas at Austin

The University of Texas at Austin is working with the Texas Water Development Board, other state and local water agencies, and academic investigators, to develop a state hydrologic information system for Texas. Through Texas HIS, a number of state datasets are now being published in a consistent way, and some datasets are now accessible which were not available before the advent of Texas HIS. This chapter describes how Texas HIS was established, what data services are now available through the Texas HIS, and insights and technologies developed through its application.

5.1 BUILDING A STATE HIS

Funding for Texas HIS was provided by the Texas Water Development Board (TWDB), an agency whose mission is "to provide leadership, planning, financial assistance, information, and education for the conservation and responsible development of water for Texas," (Texas Water Development Board, 2009). Inspired by accomplishments of the national HIS effort, TWDB wanted to facilitate the creation of an HIS for Texas. Working with a motivated state agency proved to be critical for the development of a statewide HIS. TWDB provided leadership and organized stakeholders to identify and assemble key datasets for Texas.

Regular outreach efforts keeps interested parties updated and involved in Texas HIS progress. In October, 2008, The University of Texas at Austin led a 1-day Texas HIS workshop which demonstrated how the national HIS worked and how the Texas HIS would function. Afterward, on a periodical basis, TWDB holds conference calls in which Texas HIS stakeholders discussed accomplishments and laid out future plans for Texas HIS. In May, 2009, The University of Texas at Austin led another 1-day Texas HIS workshop focused on how to publish and access data with Texas HIS. As the system had matured significantly by that point, the workshop included hands-on training in which participants loaded a sample dataset into a CUAHSI Observation Data Model (ODM) database and then published the data with the ODM WaterOneFlow web service (a version of WaterOneFlow specifically designed to read data from an ODM database). Then participants used HydroExcel to access data from their web service and to verify that they had published the data correctly. As a result of the workshop, The University of Texas at Austin continues to provide guidance to workshop participants such as John McEnery as The University of Texas at Arlington who is now setting up an HIS Server at that University, primarily to publish in CUAHSI HIS format data from the National Weather Service and US Army Corps of Engineers offices in the Dallas-Fort Worth region.

Because many agencies were still learning how to publish data with HIS, The University of Texas at Austin took the lead role in loading Texas data into Texas HIS. The quickest method of publishing static datasets was for the agency or data provider to deliver the data to The University of Texas at Austin where the HIS team would translate the data into an ODM database using the ODM Data Loader and publish the data with the ODM WaterOneFlow web service (Figure 5-1). This is how most Texas WaterOneFlow services have been published.

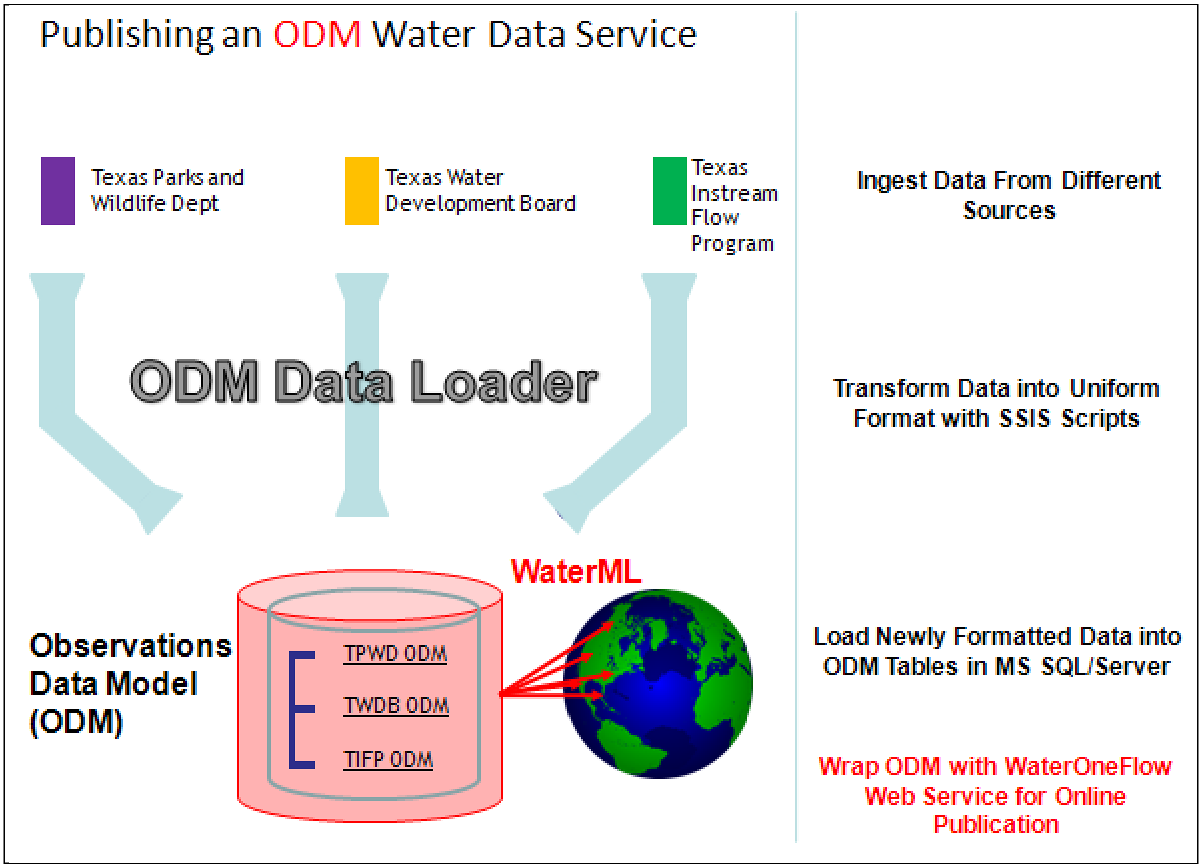


Figure 5-1 Publishing Data with the Observations Data Model and WaterOneFlow

In the case of the Texas Coastal Ocean Observation Network (TCOON), a near-real-time data source, a local copy of the data could not be loaded into a static ODM database at The University of Texas at Austin since the data are regularly updated. However, time series values from TCOON are available from the TCOON website at http://lighthouse.tamucc.edu/pq. The University of Texas at Austin thus created a hybrid WaterOneFlow web service that converts client queries into the appropriate web request to the TCOON website, and then translates the result into WaterML. This handles the GetValues request of WaterOneFlow. To handle GetSites, GetSiteInfo, and GetVariableInfo, The University of Texas at Austin obtained an XML file with site and variable descriptions from TCOON and loaded the information into an ODM database. The hybrid web service looks in this ODM database for the “metadata” about sites and variables, and performs “web scraping” on the TCOON website when time series values are needed (Figure 5-2). Because all results are returned in WaterML format, the client is insulated from the complexities of how the hybrid service is working.

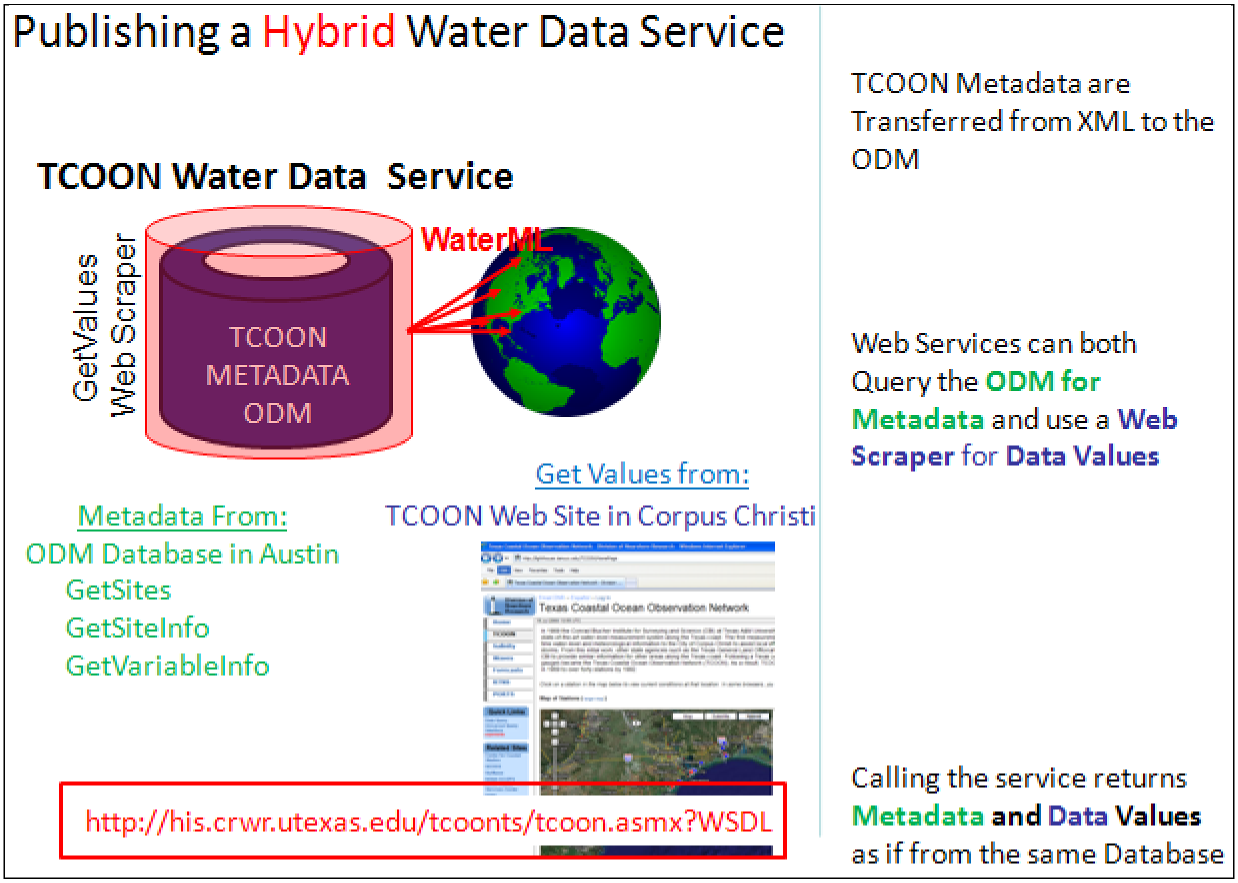


Figure 5-2 Publishing Data with a Hybrid Water Data Service

In the case of the water quality data published by Texas A&M University Corpus Christi, the data are accessed from a WaterOneFlow service housed by Texas A&M University Corpus Christi, as personnel there had the expertise to house their data in an Observations Data Model and to install their own WaterOneFlow web service. The list of WaterOneFlow services in Texas HIS is presented in Table 5-1. These services are currently indexed at the website http://data.crwr.utexas.edu/. TWDB is currently developing a Texas HIS Viewer for these services, which will be accessible at http://waterdatafortexas.org/.

Table 5-1 Texas HIS WaterOneFlow Services

Data Source	Service Address
TWDB - Coastal Water Quality	http://his.crwr.utexas.edu/TWDB_Sondes/cuahsi_1_0.asmx?WSDL
TWDB – Pan Evaporation	http://his.crwr.utexas.edu/TXEvap/cuahsi_1_0.asmx?WSDL
Texas A&M Corpus Christ – Coastal Water Quality	http://ccbay.tamucc.edu/CCBayODWS/cuahsi_1_0.asmx?WSDL
Texas Instream Flow Program – Lower Sabine	http://his.crwr.utexas.edu/SabineBio/cuahsi_1_0.asmx?WSDL

Texas Instream Flow Program – Lower San Antonio	http://his.crwr.utexas.edu/SanAntonioBio/cuahsi_1_0.asmx?WSDL
Texas Natural History Museum – Texas Fish Atlas (Percidae species only)	http://his.crwr.utexas.edu/Percidae/cuahsi_1_0.asmx?WSDL
Texas State University – Aquatic Biology	http://his.crwr.utexas.edu/BioODws2/cuahsi_1_0.asmx?WSDL
TCOON – Coastal Water Levels and Conditions	http://his.crwr.utexas.edu/tcoonts/tcoon.asmx?WSDL
Texas Parks and Wildlife Department – Coastal Water Quality	http://his.crwr.utexas.edu/TPWD/cuahsi_1_0.asmx?WSDL
Texas Commission on Environmental Quality – Water Quality	http://his.crwr.utexas.edu/TRACS/cuahsi_1_0.asmx?WSDL
National Weather Service – Historical Precipitation for Austin, TX	http://his.crwr.utexas.edu/nwsmpe/cuahsi_1_0.asmx?WSDL

A descriptive page for each Texas WaterOneFlow web service is available at http://data.crwr.utexas.edu/wsdl.html. The page highlights site and variable details of the service, and includes an interactive map showing site locations and which uses the ESRI HydroBasemap to provide the spatial context for the sites. Hovering the mouse over a site will pop up brief site information (Figure 5-3).

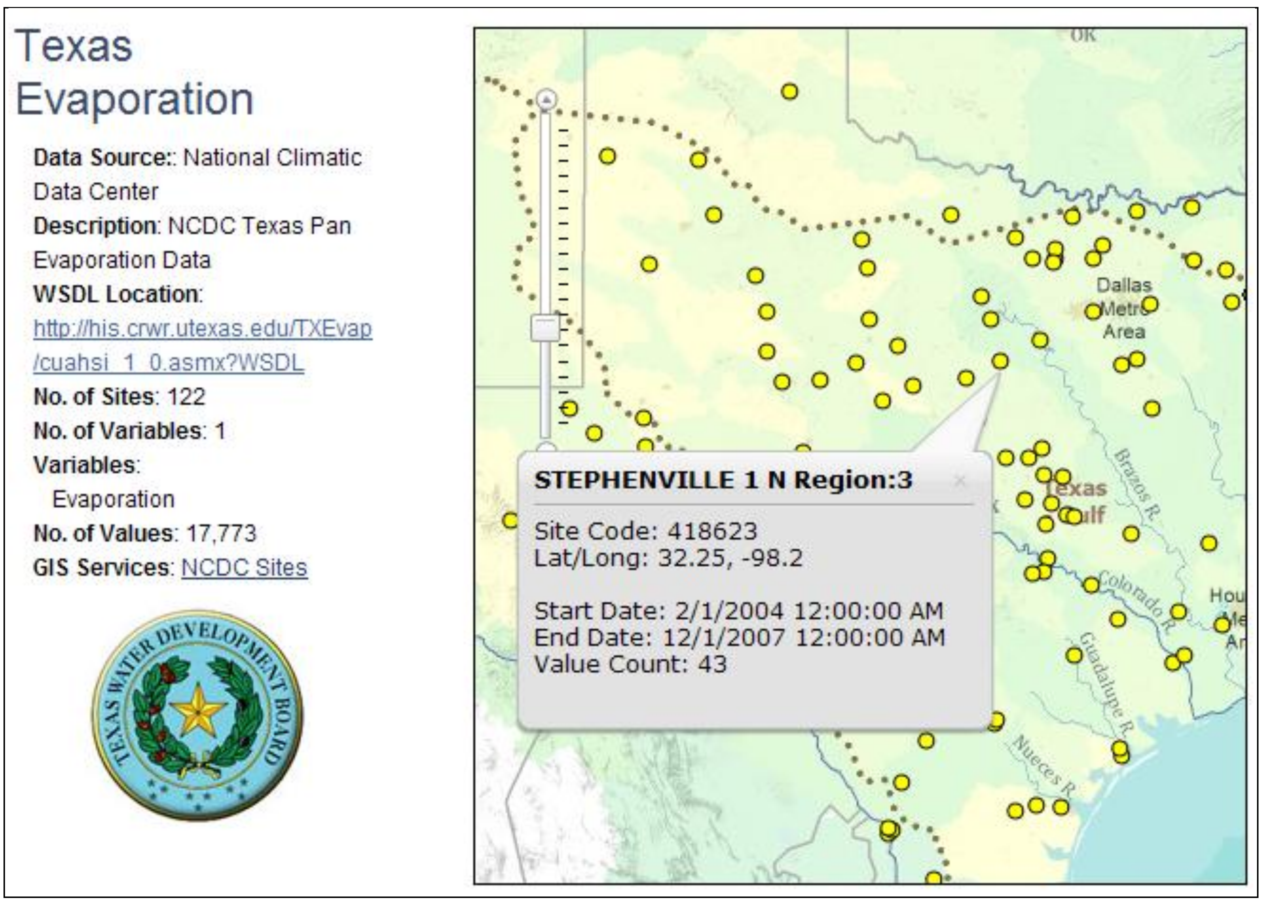


Figure 5-3 Descriptive Page for Texas Evaporation WaterOneFlow Web Service

For services which access ODM databases housed at The University of Texas at Austin, an application called the ODM Time Series Analyst has been hooked up to the interactive maps for those services. The Time Series Analyst plots graphs of time series variables and statistics from an ODM database. By clicking on a site location in the map, the user invokes a new browser window that shows a Time Series Analyst view onto a variable at the clicked site.

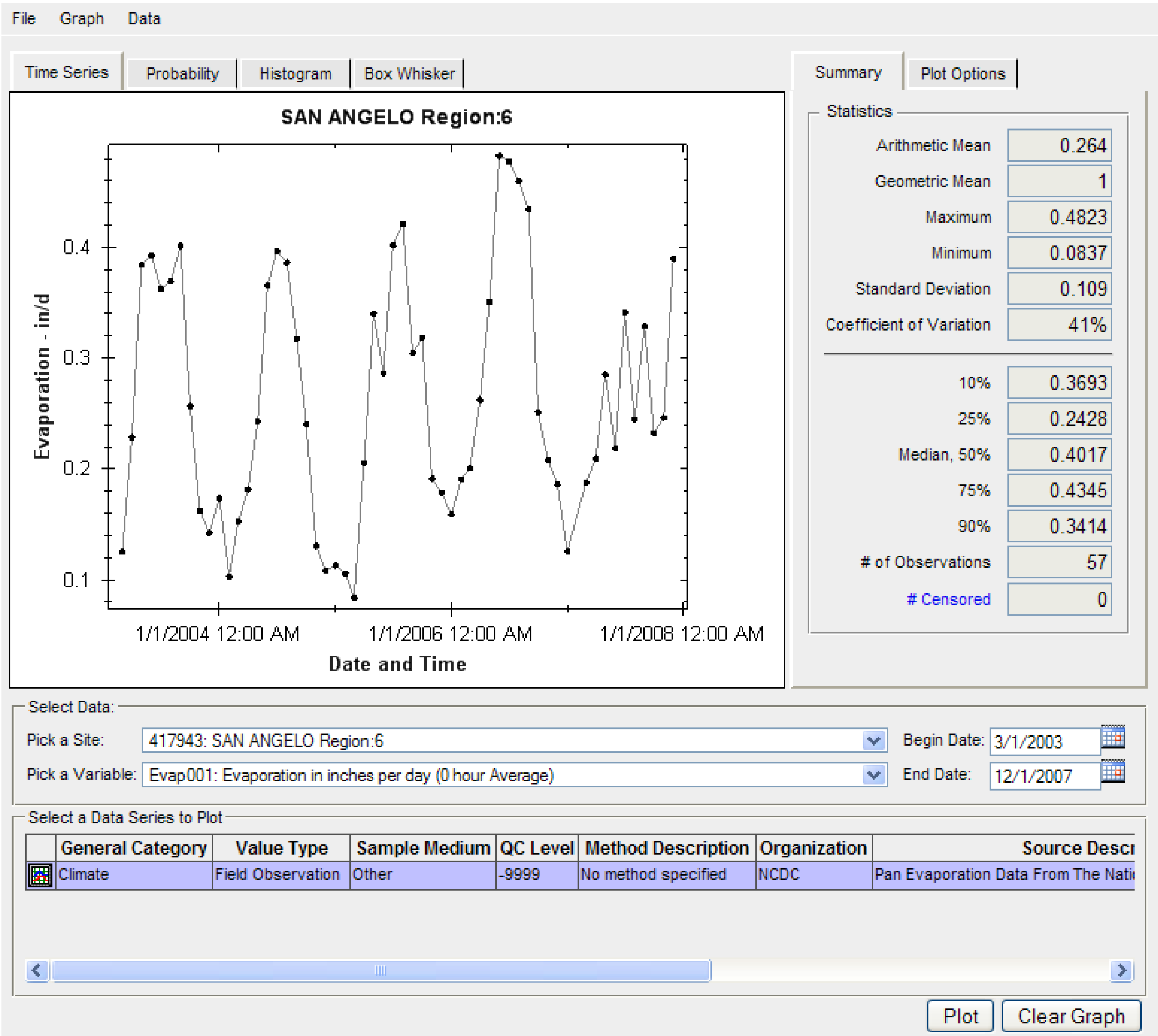


Figure 5-4 Time Series Analyst Plot of Evaporation

By accessing the Data menu within the Time Series Analyst, the data plotted in the graph can be downloaded to the client. Thus, for services supported by ODM databases housed at The University of Texas at Austin, the descriptive page for each service supports an interactive map as well as time series plotting and download. For other services, only the interactive map is provided.

5.2 Thematic Datasets

In addition to gathering data for publication with Texas HIS, The University of Texas at Austin synthesized these data to produce thematic datasets. A thematic dataset is a set of geospatial or time series data grouped together for some purpose, and is one of the new developments that arose from the Texas HIS effort. Information about a thematic dataset is recorded in a thematic dataset catalog table which (a) provides sufficient information that a Web Service call can be issued to get the data, and (b) lists key summary information so that a user can assess what and how much data is contained in each theme. For example, a thematic dataset could be created to summarize salinity time series available across all Texas WaterOneFlow web services. A thematic dataset catalog provides the URI of the service to get each time series, the calling parameters to supply the service, and additional geospatial and attribute information to assist in the selection of time series for download. Thematic dataset catalogs can be published as a Web Feature Service (WFS) for standardized access via a mapping interface.

Both thematic dataset catalogs (description of the time series) and the thematic datasets themselves (the time series data) can be packaged into a database or other format and published online for easy access to the data. As an example, a thematic dataset for salinity in Texas has been prepared in Arc Hydro geodatabase format, and is described at http://data.crwr.utexas.edu/salinity.html.

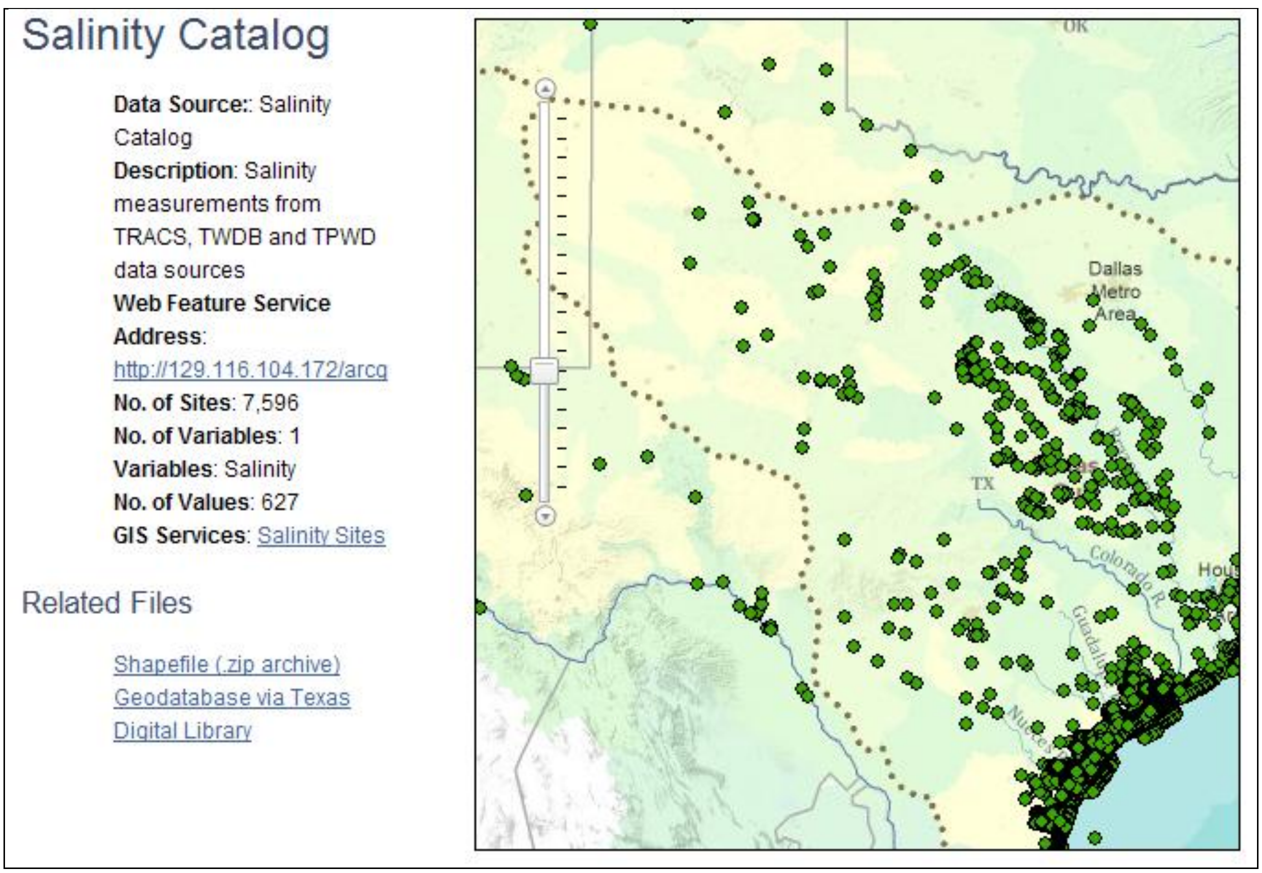


Figure 5-5 Salinity Theme for Texas

A major benefit of maintaining thematic datasets is that users can very quickly access quality controlled data pertinent to commonly utilized themes. Instead of having to make individual web services requests for data across multiple data sources, users can go a single place, download a single file, and have all of the data. The “place” that users go to discover themes is the HydroPortal, described in the next section.

5.3 HydroPortal

Texas HIS WaterOneFlow web services and themes are currently published at http://data.crwr.utexas.edu/. While that website does provide a summary of the services a map showing site locations for each service, it does not permit a search on the metadata for the services, nor does it allow users outside of the HIS team to register new data services. For this reason, The University of Texas at Austin and the CUAHSI HIS team have been working with the ESRI Water Resources Applications group to develop the HydroPortal, a customization of the Geoportal Extension for ArcGIS which enables online resources to be registered and searched via metadata keywords or geospatial coordinates. The Geoportal extension supports common metadata standards such as FGDC and Dublin Core, as well as search standards such as Open Geospatial Consortium’s Catalog Services for the Web. The HydroPortal customization includes metadata entry fields that support registration of WaterOneFlow web services and thematic datasets, along with minor adjustments to the user interface to indicate that the application is a HydroPortal rather than a regular Geoportal.

A prototype HydroPortal is available at http://crwr-idis.crwr.utexas.edu:8080/GPT9/catalog/main/home.page. This portal is purely for testing purposes and will eventually be replaced by a production application as the software matures.

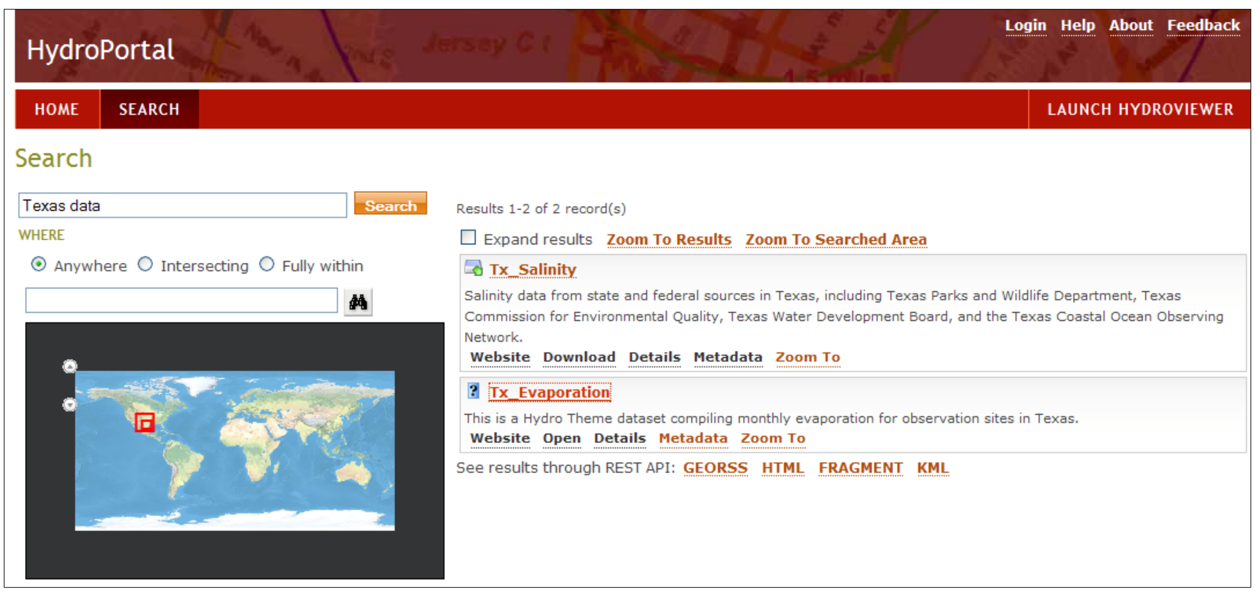


Figure 5-6 HydroPortal supports geospatial and metadata keyword search for online resources

Although an instance of the HydroPortal will be hosted by either The University of Texas at Austin or a Texas agency, the web services supported by HydroPortal will allow it to be called by other applications, which means that the Texas HIS can remain independent from the national implementation of HIS, yet also participate in the national CUAHSI HIS system.

5.4 References

Texas Water Development Board. Web. 12 October 2009. <http://www.twdb.state.tx.us/>.

Chapter 6. An Ontology for Discovery of Hydrologic Data

By Rick Hooper and Yoori Choi[2], Michael Piasecki[3]

6.1 Background

As described in the HIS Status Report (HIS, 2008), an ontology was developed as part of Bora Beran's PhD dissertation, done under the direction of Michael Piasecki, that enabled searching of the HIS Central master metadata catalogue using concepts. A critical functionality provided by this research was not just the ability to search for synonymous terms (e.g., discharge, streamflow) but also to permit searching for a class of measurements at different levels of generality (e.g., from most general to most specific: "nutrients"→ "macronutrients"→"nitrogen" →"nitrate plus nitrite"). Therefore, the expert user could search for a specific chemical ("nitrate plus nitrite"), but a less experienced user could search for "nutrients" and get back a hierarchical table with all the compounds that fell under that designation. Furthermore, because the ontology is not a strict hierarchy, a single leaf concept (e.g., dissolved silica) could appear at multiple places, (e.g., as both a nutrient and a weathering product).

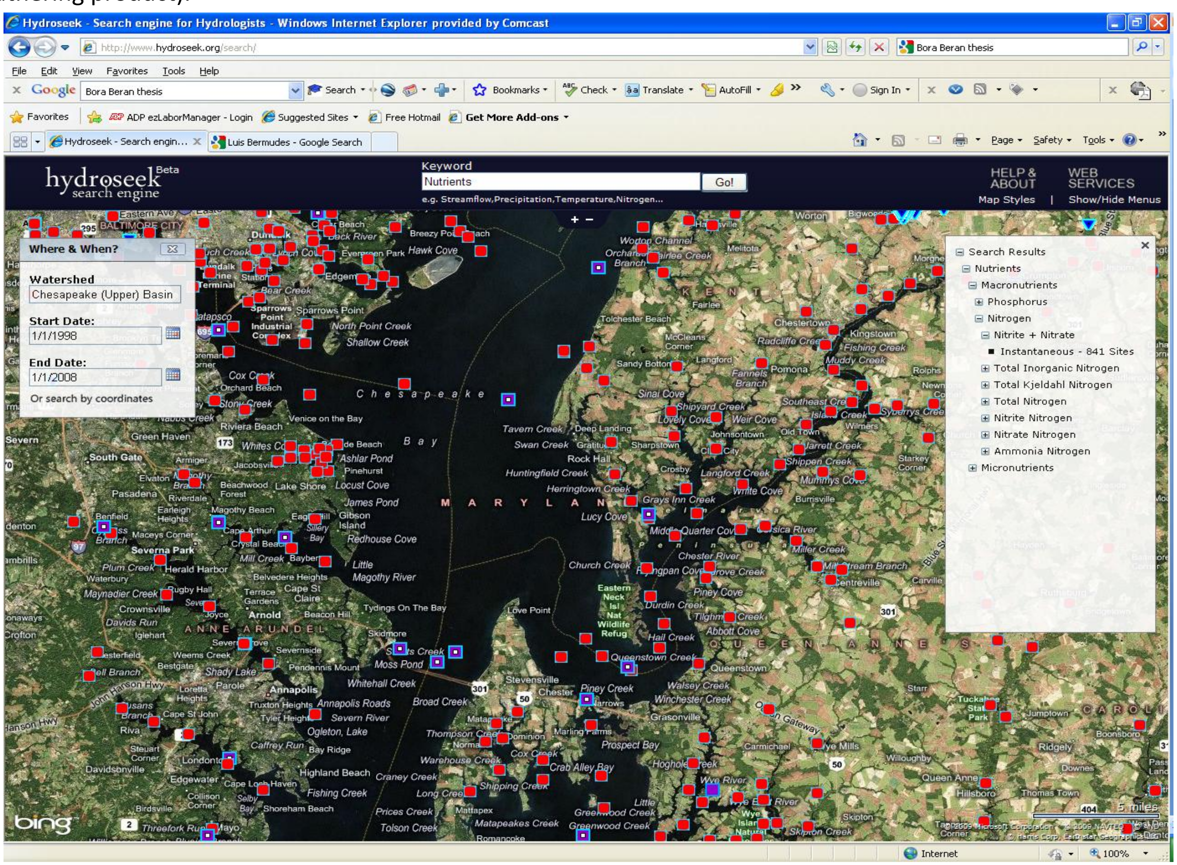


Figure 6-1 Hydroseek user interface

---

[2] Consortium of Universities for the Advancement of Hydrologic Science, Inc., Medford, MA
[3] Drexel University, Philadelphia, PA

The left hand box enables the user to specify space and time coordinates. The right hand return shows a list of measured properties associated with the concept "Nutrients." Different symbols indicate sites of different measurement networks.

Dr. Beran's work (Beran B., 2008) (Beran B., Engineering New Paths to Hydrologic Data, 2009) (Beran B., A Semantic Annotation Tool for the Hydrologic Sciences, 2009) provided a powerful demonstration of what could be achieved through the use of an ontology designed for discovering hydrologic data. The ontology developed by Beran and Piasecki can be viewed at http://hiscentral.cuahsi.org/startree.aspx. His work, however, was limited to a small set (about 150) of the many thousands of parameters found in the data holdings of the US EPA and USGS and focused on chemical compounds found in surface water.

Therefore, the objective of this task is to broaden the ontology to be more comprehensive (i.e., to include more parameters, including biological data) and to reconcile the competing needs for simplicity and precision in the search specification.

6.2 Approach

6.2.1 Scope and Complexity

The purpose of this ontology is to enable *discovery* of hydrologic data, or, more precisely, time-series data collected at a fixed point, known as a site within the information model used in CUAHSI Water Data Services. The basic search interface permits the specification of a space/time "box" for selection of sites. That functionality will not be altered.

However, discovering the property measured is done through a single keyword in the present system. For some properties, such as temperature, the keyword contains the medium in which it was measured (e.g., "Air temperature"). For other properties, the medium is assumed to be water (e.g., "Nitrite"), but it is not explicit. Furthermore, stream sites dominated the data sets considered during HydroSeek development and site type is not explicit in the return. Considering the importance for the user to understand the environment in which the measurement is made and the medium on which it is made, we decided that a minimum of three search parameters must be specified:

- Site Type (e.g., stream, lake, well)
- Medium sampled (e.g., air, water, tissue)
- Property measured (e.g., temperature, nitrate, stage)

The default setting for site type and media may well be "all" in the search interface, but it is important for the user to understand both the setting and the media in which the measurements are made. Although this complicates the search specification and the return values, this was considered to be the least complex approach that provided a return that can be interpreted unambiguously by the user and would help prevent misinterpretation of the data.

The USGS and EPA have site type descriptions (Table 6-1) that can serve as a starting point for the development of this identifier. CUAHSI has a controlled vocabulary within the ODM for media (Table 6-2).

Table 6-1 The coarsest definition of Site Type. These types are further subdivided in the EPA/USGS ontology

Site Type	Definition
Glacier	Body of land ice that consists of recrystallized snow accumulated on the surface of the ground and moving slowly downslope over a period of years or centuries.
Land	A location on the surface of the earth that is not normally saturated with water. (See also: Wetland).
Wetland	Lands where saturation with water is the dominant factor determining the nature of soil development and the types of plant and animal communities living in the soil and on its surface. Wetlands are found from the tundra to the tropics and on every continent except Antarctica. Wetlands are areas that are inundated or saturated by surface or ground water at a frequency and duration sufficient to support, and that under normal circumstances do support, a prevalence of vegetation typically adapted for life in saturated soil conditions. Wetlands generally include swamps, marshes, bogs and similar areas. Wetlands may be forested or unforested, and naturally or artificially created.
Atmosphere	A site established primarily to measure meteorological properties or atmospheric deposition.
Estuary	A coastal inlet of the sea or ocean; especially the mouth of a river, where tide water normally mixes with stream water. Salinity in estuaries typically ranges from 1 to 25 Practical Salinity Units (psu), as compared to oceanic values around 35 psu. See also: tidal stream and coastal.
Ocean	Site in the open ocean, gulf, or sea. (See also: Coastal, estuary, and tidal stream).
Lake	An inland body of standing fresh or saline water that is generally too deep to permit submerged aquatic vegetation to take root across the entire body (cf: wetland). This site type includes an expanded part of a river, a reservoir behind a dam, and a natural or excavated depression containing a water body without surface-water inlet and (or) outlet.
River/Stream	A body of running water moving under gravity flow in a defined channel. The channel may be entirely natural, or altered by engineering practices through straightening, dredging, and (or) lining. An entirely artificial channel should be qualified with the "canal" or "ditch" secondary site type.
Spring	A location at which the water table intersects the land surface, resulting in a natural flow of ground water to the surface. Springs may be perennial, intermittent, or ephemeral.
Well	A hole or shaft constructed in the earth intended to be used to locate, sample, or develop groundwater, oil, gas, or some other subsurface material. The diameter of a well is typically much smaller than the depth. Wells can be used to recharge ground water artificially or to pressurize oil and gas production zones. Additional information about specific kinds of wells should be recorded under the secondary site types or the Use of Site field. Underground waste-disposal wells should be classified as waste-injection wells (FA-WIW).
Subsurface	A location below the land surface, but not a well, soil hole, or excavation.
Facility	A non-ambient location where environmental measurements are expected to be strongly influenced by current or previous activities of humans.
Aggregate water-use establishment	An aggregate class of water-using establishments or individuals that are associated with a specific geographic location and water-use category, such as all the industrial users located within a county or all self-supplied domestic users in a county. The aggregate class of water-using establishments is identified using the national water-use category code and optionally classified using the Standard Industrial Classification System Code (SIC code) or North American Classification System Code (NAICS code). An aggregate water-use establishment site type is used

	when specific information needed to create sites for the individual facilities or users is not available or when it is not desirable to store the site-specific information in the database. Data entry rules that apply to water-use establishments also apply to aggregate water-use establishments.
Aggregate ground-water use	An aggregate of specific sites where ground water is withdrawn or returned and is defined by a geographic area or some other common characteristic. An aggregate ground-water site type is used when it is not possible or practical to describe the specific sites as springs or as any type of well including 'multiple wells', or when water-use information is available only for the aggregate.
Aggregate surface-water-use	An aggregate of specific sites where surface water is diverted or returned and is defined by a geographic area or some other common characteristic. An aggregate surface-water site type is used when it is not possible or practical to describe the specific sites as diversions, outfalls, or land application sites, or when water-use information is only available for the aggregate.

Table 6-2 Media designation from CUAHSI ODM Controlled Vocabulary

Term	**Definition**
Air	Sample taken from the atmosphere
Groundwater	Sample taken from water located below the surface of the ground, such as from a well or spring
Precipitation	Sample taken from solid or liquid precipitation
Sediment	Sample taken from the sediment beneath the water column
Snow	Observation in, of or sample taken from snow
Soil	Sample taken from the soil
Surface Water	Observation or sample of surface water such as a stream, river, lake, pond, reservoir, ocean, etc.
Tissue	Sample taken from the tissue of a biological organism
Unknown	The sample medium is unknown
Not Relevant	Sample medium not relevant in the context of the measurement
Other	Sample medium other than those contained in the CV

Further decisions were made relative to the scope of searches that the discovery ontology would support. We decided to focus the ontology on properties measured and to permit searches to occur a multiple levels of the ontology, as HydroSeek allowed, but that we would *not* support search on the methods used or other metadata elements such a units. Therefore, one could discover “nitrate + nitrite” or “nutrients” but not parameters measured by ion chromatography or by ion-specific electrodes. Such properties are returned with the data as metadata but cannot be discovered using the ontology concepts.

These decisions helped to balance the complexity of the search with the ambiguity of the return. A final design decision remains, however, which is the granularity of the search concepts and the number of properties returned with each call. That issue will be discussed in the following section.

6.2.2 Development of a Master List of Keywords

The USGS and the EPA have been working to reconcile parameters names in the NWIS and STORET data bases for a number of years. EPA maintains the Substance Registry Service (http://iaspub.epa.gov/sor_internet/registry/substreg/home/overview/home.do) as an authoritative list of substances regulated by EPA. The purpose of this service is to provide a common basis for identification of chemicals, biological organisms, and physical properties. We obtained a working list of physical and chemical parameters by the joint USGS/EPA working group, which included their grouping of properties into different classes. In addition, we used a list of biological parameters, compiled by Eric Hersh (Center for Research in Water Resources, University of Texas) from aquatic ecology collections in Texas. These biological parameters included various species that were measured, as well as more aggregated measures. These, too, were grouped into a hierarchy. When combined with the original ontology develop by CUAHSI, the total number of leaf concepts (i.e., the most precise description of the properties measured) is 4165. The resulting list of keywords is *not* viewed as the ultimate list of all keywords, but as a credible starting point that will be built upon by community input.

The groupings provided by USGS/EPA and by TCEQ were used as an initial point for grouping the concepts. Concepts were also reviewed for suitability; some were eliminated because they were deemed to describe methods rather than properties measured; some were redundant. At this stage, 4117 records remain but the review is not yet complete.

The hierarchy that has been developed to date begins with one simple navigation layer (Figure 6-2) that groups the leaf concepts at the most basic level.

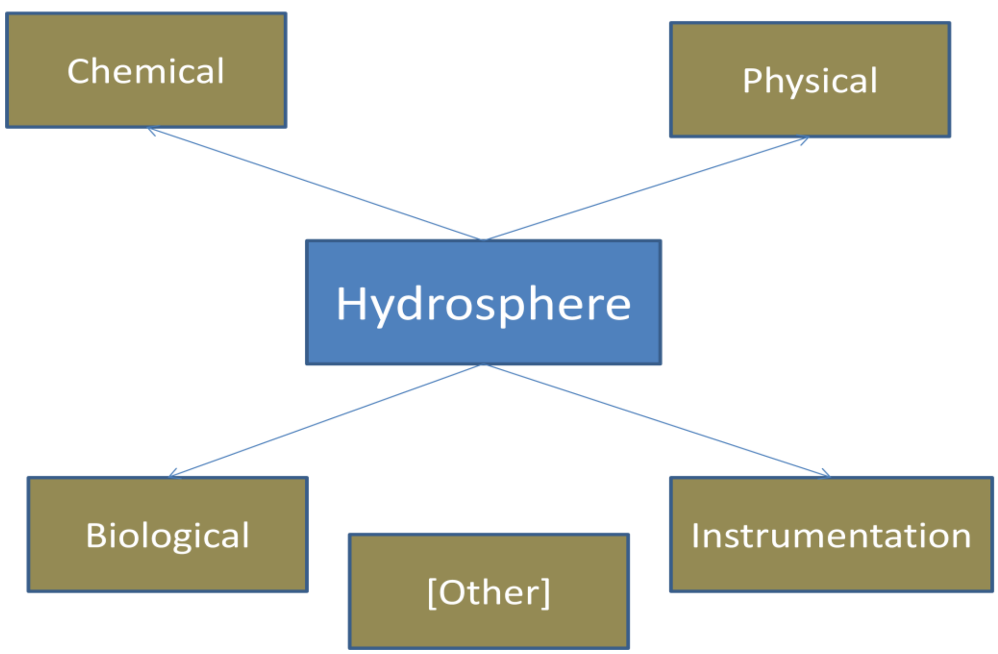


Figure 6-2 Navigation Layer

These properties will not be able to be searched on (they are too general), but rather provide an initial organization of the concepts for navigation purposes. The “Other” group currently has 87 concepts associated with it, but the intent is to distribute or eliminate those concepts to eliminate this group; many of the concepts contained in that group are related to methods (e.g., description of fish seines).

Each of these navigation concepts can be broken down further into subgroups. For the Chemical group, there are three layers of subgroups as shown in Figure 6-3. Most of these terms were defined by the USGS/EPA joint working group, although a few groups have been added for completeness and balance, such as “Bulk Property” and “Composite”. Each of these subgroups are considered “discovery layer.” That is, these are searchable terms and will return all of the measurements of leaf concepts associated with that subgroup, organized by all the subgroups below the one returned by the search.

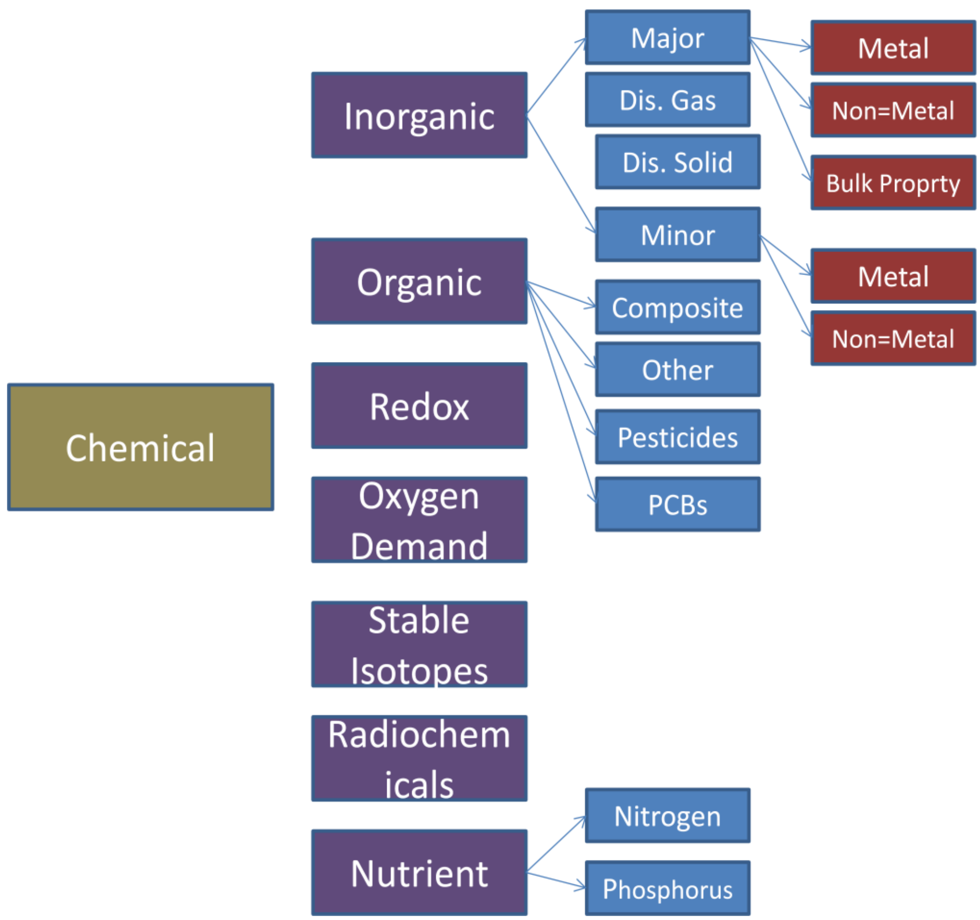


Figure 6-3 Sub-groups contained within Chemical concepts group

The distribution of concept contained within the chemistry layer is highly skewed towards organic chemicals, as seen in Figure 6-4. A decision must be made about how to handle a request for “organic” chemicals. Do we return all the parameters, even if the number is large or request the user to refine the search further?

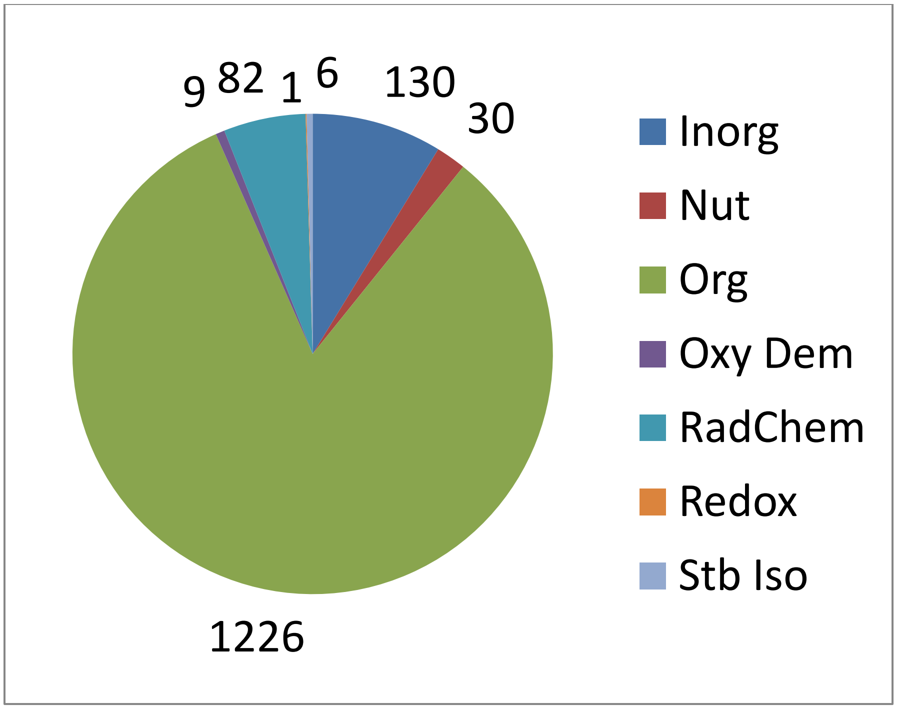


Figure 6-4 The distribution of concept contained within the chemistry layer

When we go to the next level down below this group, the number of parameters in each of the subgroups under inorganic is more feasible as a return to a search inquiry as shown in Figure 6-5.

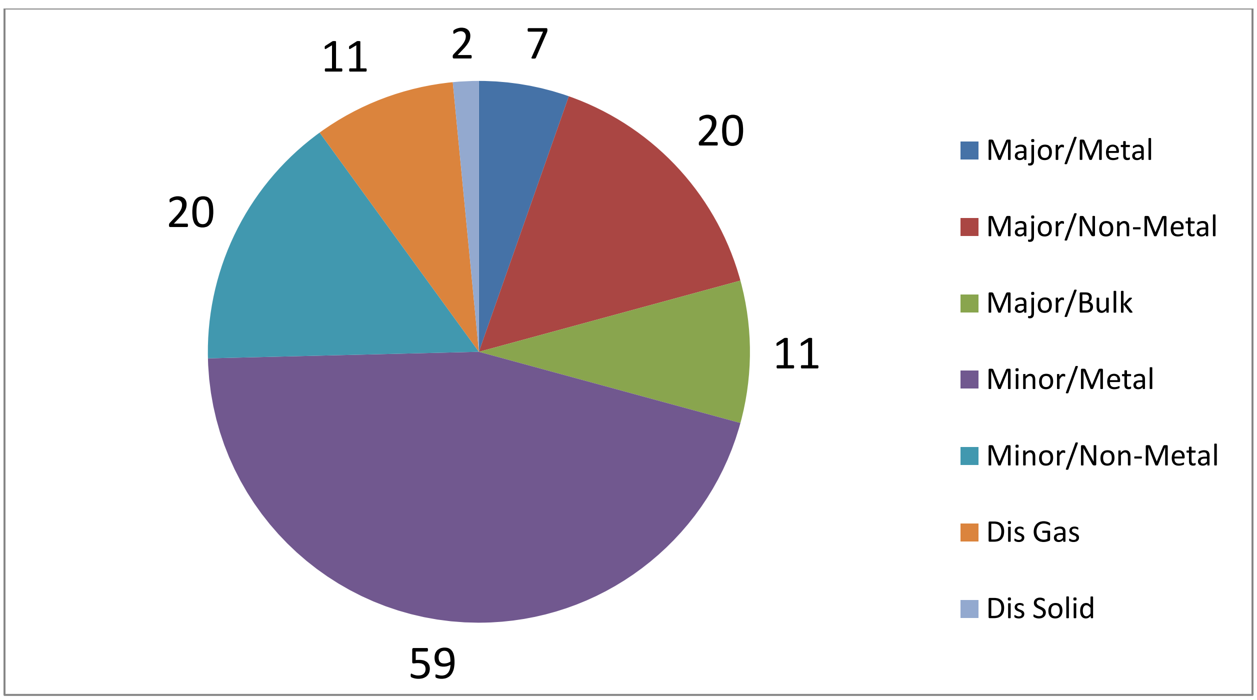


Figure 6-5 Subgroups under Chemical->Inorganic

As can be expected, the organic groupings are far coarser and are likely not very functional for the purposes of discovering data because the returns will contain so many concepts (Figure 6-6). Also troubling is the large number of “Other” organic compounds. Additional groupings could be by chemical property or by use (e.g., herbicide)

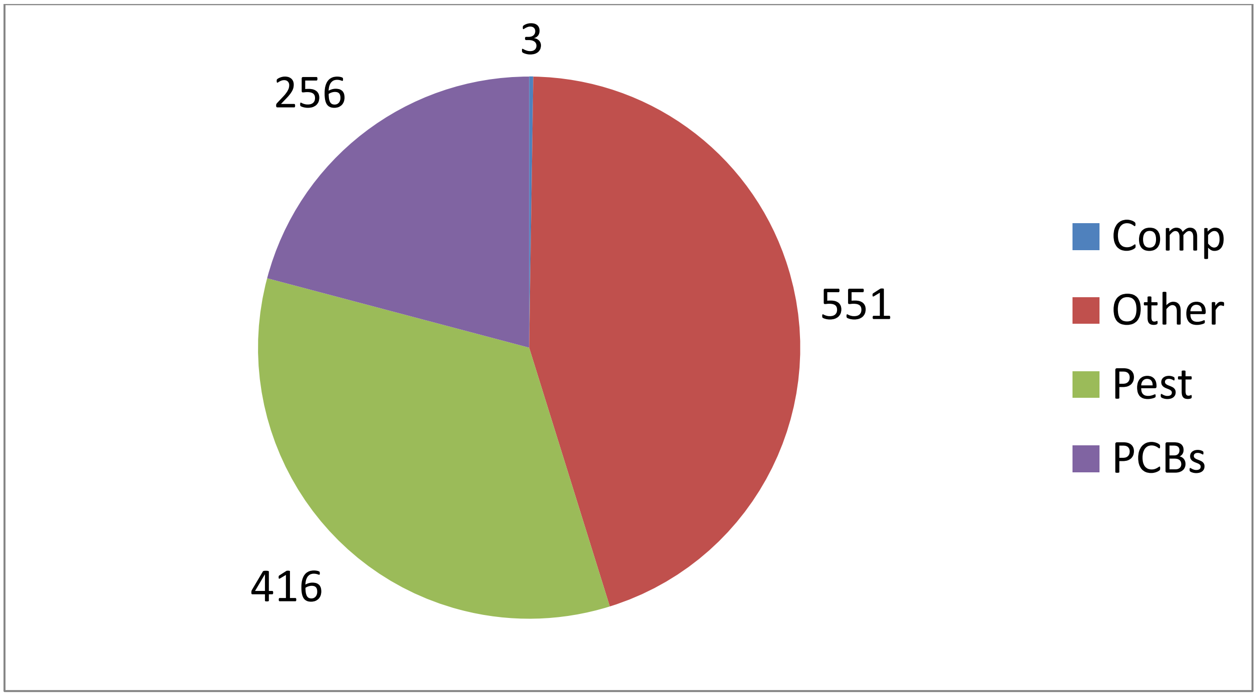


Figure 6-6 Subgroups under Chemical-> Organic

Our team does not contain the expertise to develop these groupings further. We need to get outside assistance with this, but the current hierarchy is adequate to expose to the community for input.

The biological concepts are somewhat more challenging to group. We had to develop more of the hierarchy for the biological concepts than the chemical concepts. The initial hierarchy is shown in Figure 6-7.

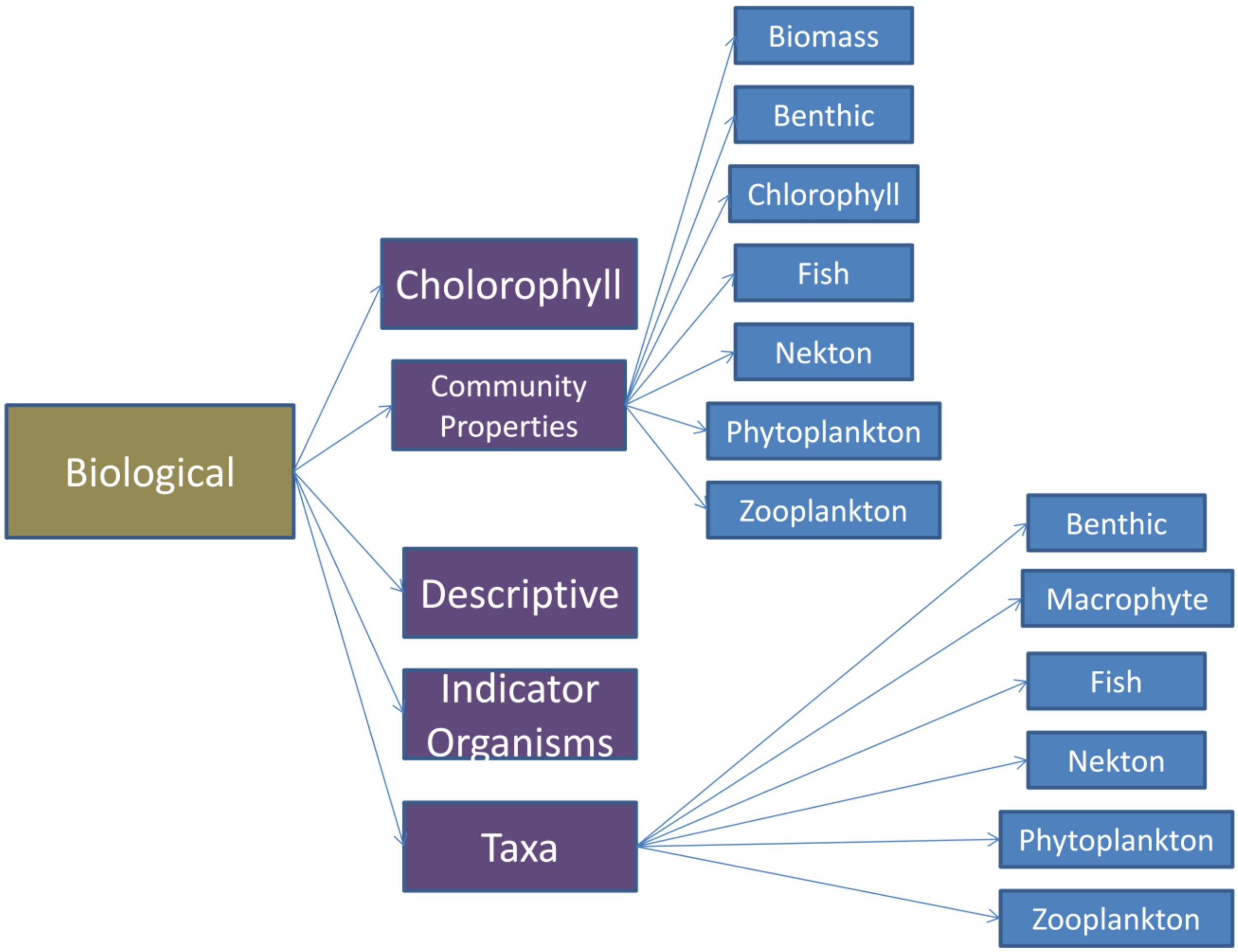


Figure 6-7 Subgroups of biological concepts

The distribution of the leaf concepts under the biological grouping a far more skewed than the chemical briefing, as shown in Figure 6-8. Taxa (which can be at the species, genus, or family level) by far dominate the species concepts. As used by TCEQ, these designations were generally counts of individuals. We decided to allow any measure at the species level (count, density, biomass, etc.) to be associated with this leaf concept. The user would, therefore, be able to discover that some data specific to a taxon was available.

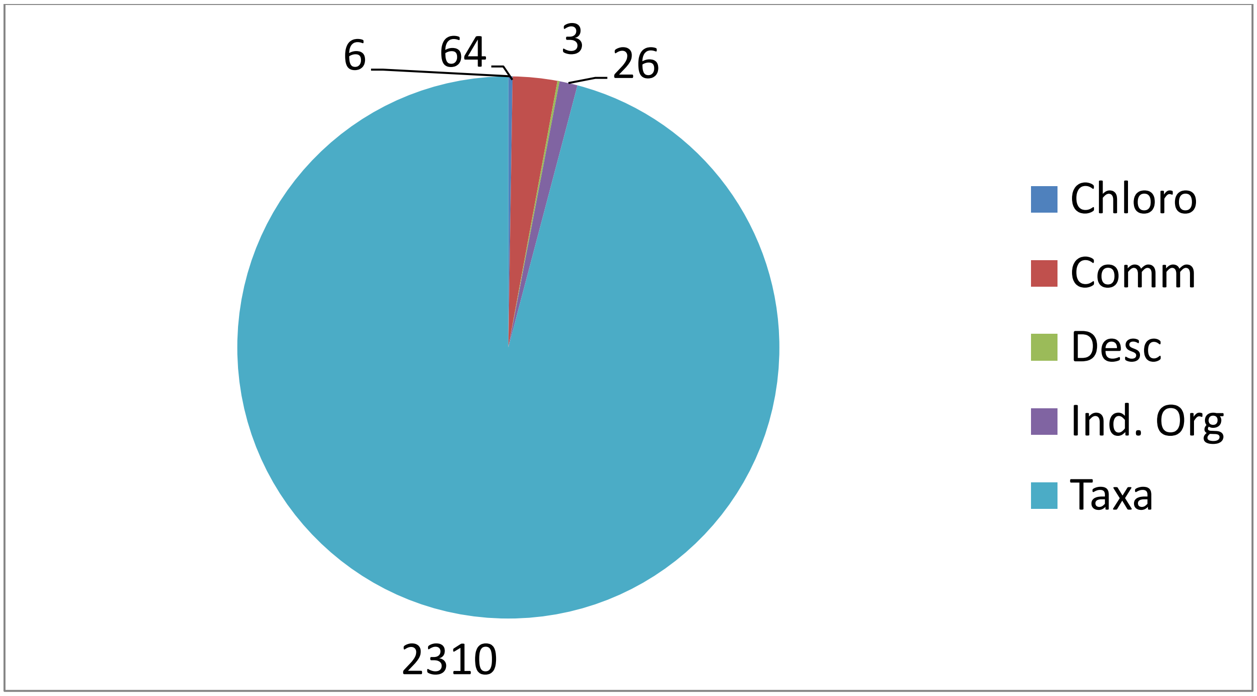


Figure 6-8 Distribution of leaf concepts in Biological group

The taxa are divided into further subgroupings as shown in Figure 6-9. In this particular list, benthic taxa dominate this grouping, but 4 of the 5 groups have large number of leaf concepts associated with them. We are uncertain as the functionality of these groups and subgroups for data discovery of biological parameters, but view this as a sufficiently specific to test these groupings with focus groups of knowledgeable individuals.

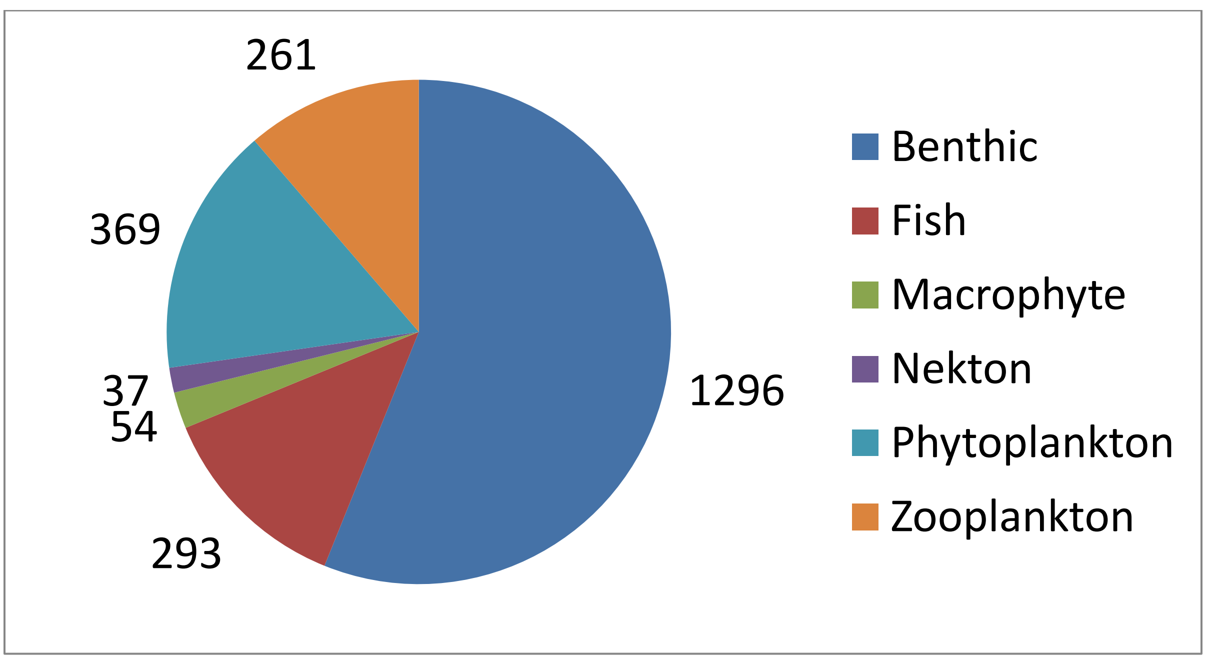


Figure 6-9 Subgroups of taxa

The physical parameter subgroups are not completed as of this writing. In general, there are far fewer physical parameters and defining subgroups is less difficult than for biological or chemical parameters.
The next step after developing an initial master list of concepts with proposed groupings is the conversion of the discovery layers into a list of keywords. This is a non-trivial exercise as some of the leaf concepts are dependent upon their context in the hierarchy. The keywords, however, are a simple flat list of terms. As with the hierarchical ontology, our approach is to make an initial attempt that is factually accurate, expose this discovery interface to focus groups and to modify the keyword list (and ontology hierarchy) as indicated by the user response.

6.3 Conclusions

Building upon the work of Piasecki and Boran, we have begun to build an ontology for discovering hydrologic data, including physical, chemical, and biological measurements. During the initial development of HydroSeek, Piasecki and Boran sought to define discovery layers that returned a 'reasonable' number of concepts—somewhere between 5 and 15 concepts. After reviewing the chemical and biological concepts, it seems unlikely to be able to retain this rule, as many subgroups break into many more than this number of parameters.

Our approach to resolve this issue is to expose this initial discovery ontology to user focus groups to assess its utility and how it must be altered to improve the utility. When a workable version is obtained, this ontology should be hosted in a moderated forum to allow the community to make suggested improvements and updates. A proposal for this work is under consideration at this time in the Geoinformatics competition at NSF.

6.4 References

Piasecki, M., Beran, B., "A Semantic Annotation Tool for the Hydrologic Sciences", *Earth Science Informatics*, Springer, Vol. 2, No. 3, p. 157, August 2009.
http://dx.doi.org/10.1007/s12145-009-0031-x

Beran B., Piasecki M., "Engineering New Paths to Hydrologic Data", *Computers and GeoSciences*, Elsevier, Special Issue, Vol. 35, Issue 4, pp. 753 – 760, April 2009.

Beran B., Piasecki M., "Availability and Coverage of Hydrologic Data in the US Geological Survey National Water Information System (NWIS) and US Environmental Protection Agency Storage and Retrieval System (STORET)", *Earth Science Informatics*, Springer, Vol.1, No.3, November 2008, pp. 119 – 129.

Chapter 7. LINKS TO HYDROLOGIC MODELING

By Jonathan Goodall and Anthony Castronova, University of South Carolina

7.1 OVERVIEW

The role of the University of South Carolina on the CUAHSI Hydrologic Information System (HIS) project is to provide a link between the data synthesis and delivery system developed *within* the HIS project and hydrologic modeling systems developed *external to* the HIS project. There are a large and growing number of modeling systems in development that have relevance to the hydrologic modeling community (OMS, ESMF, CSDMS, OpenMI, etc.). In a broad sense, these modeling systems can be grouped into two classes: (1) those that adopt a lumped view of space (i.e. watersheds, river reaches, etc.), sometimes described as unstructured grids, and (2) those that adopt a distributed view of space (i.e. computational meshes), sometimes described as structured grids. Although the overarching goal of linking the HIS with models is to provide easy access to HIS data within as many of these existing modeling systems as possible, particular focus has been placed on interoperating with both the OpenMI and ESMF modeling systems as case studies of lumped and distributed modeling systems, respectively.

This chapter updates the 2008 summary of the University of South Carolina HIS work as reported in Chapter 8 of the HIS 1.1 documentation (http://his.cuahsi.org/documents/HISOverview.pdf). While many of the tools and techniques presented in this document are still valid today, they have been enhanced over the past year. For example, Chapter 8 of the HIS 1.1 document discusses the FetchWaterML application, which is a command-line tool for batch downloading multiple WaterML time series files. The functionality made available in FetchWaterML is now available in a more user friendly way through the new HydroDesktop application. Likewise, Chapter 8 of the HIS 1.1 document discusses the HydroLink application, which is an OpenMI-compliant component that supplies HIS data to OpenMI modeling components. HydroLink has also been embedded within a new extension to HydroDesktop called HydroModeler that supports modeling using the OpenMI standard.

The following sections overview three outcomes from the University of South Carolina team in 2009. First, the team designed and prototyped a system for more easily and rapidly creating new OpenMI modeling components called the Simple Model Wrapper (SMW). Second, the team collaborated with the Earth System Modeling Framework (ESMF) development team to prototype ways for making the HIS and ESMF interoperable. Lastly, the team began building a modeling plug-in for HydroDesktop that will serve as a front end application for supporting hydrologic modeling and analysis from within HydroDesktop.

7.2 CREATING OPENMI HYDROLOGIC PROCESS COMPONENTS

While OpenMI is at its core a standard for linking hydrologic modeling and data components, the OpenMI Association Technical Committee (OATC) provides a Software Development Kit (SDK) for using the OpenMI as a modeling framework. The primary focus of the OATC SDK is on wrapping legacy models. However, an important part of hydrologic science is creating new models to test scientific hypotheses and understand complex hydrologic systems. For this reason, there is a need to allow hydrologists to more easily incorporate their own models as components within an OpenMI workflow. The Simple Model Wrapper is designed to address this need of the hydrologic science community by simplifying the process of creating new OpenMI-compliant hydrologic modeling components.

The Simple Model Wrapper works by allowing a component to be described by two files: (1) a model engine DLL that specifies how the model should be initialized, advanced in time, and finished and (2) a model configuration XML file that specifies model attributes (e.g. description, time horizon, inputs, outputs, etc.). We believe that this is a more intuitive process for creating modeling components than implementing the interfaces provided in the OATC SDK. The SMW code works a translator between the Model.dll and other components within an OpenMI component workflow as depicted in Figure 7-1.

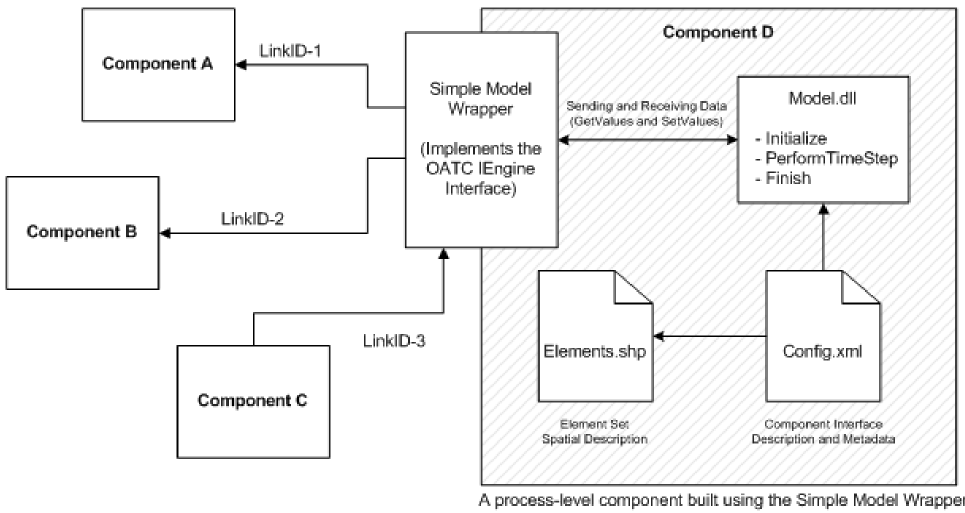


Figure 7-1 General Design of an OpenMI modeling component when using the Simple Model Wrapper approach

We have also begun to build a modest library of OpenMI-compliant hydrologic modeling components using the SMW. Each model component is accompanied with a "unit test" that checks the component with known problems taken from popular hydrology textbooks. The modeling component source code is freely available within the HydroDesktop SVN repository.

7.3 Linking HIS and ESMF through Web Services and WaterML

Working with the ESMF development team, we investigated two approaches for linking the HIS with ESMF[4]. In the first approach we used web services to facilitate data communication between Earth System Modeling Framework (ESMF) model components and the HIS. In the spirit of the overall HIS system architecture, which adopts a service-oriented paradigm, we have investigated if modeling services could be wrapped as OpenMI-compliant components. A motivation for this work is to provide interoperability between the OpenMI and other modeling frameworks, in particular the ESMF. Working with the ESMF developers, we were able to demonstrate that it is possible to wrap an ESMF component exposed as a web service within an OpenMI component. While there is much work to be done on making service oriented modeling more robust, the vision for this work is to eventually find an appropriate API for exposing models as web services. To this end, we believe that Open Geospatial Consortium (OGC) Web Processing Service could serve as such a standard and are now working to investigate ESMF web services following the OGC-WPS specification. Thus, just as the WaterOneFlow API provides access to heterogeneous databases, a standard API for exposing models as services would provide access to heterogeneous

[4] The HIS team would like to acknowledge the ESMF team, specifically Cecelia Deluca, Kathy Saint, and Xinqi Wang, for their collaboration and efforts to achieve interoperability between HIS and ESMF.

models and modeling systems to a wide set of end users. Together these service APIs could provide a powerful environment for hydrologic analysis and modeling using services and workflows as envisioned in Figure 7-2.

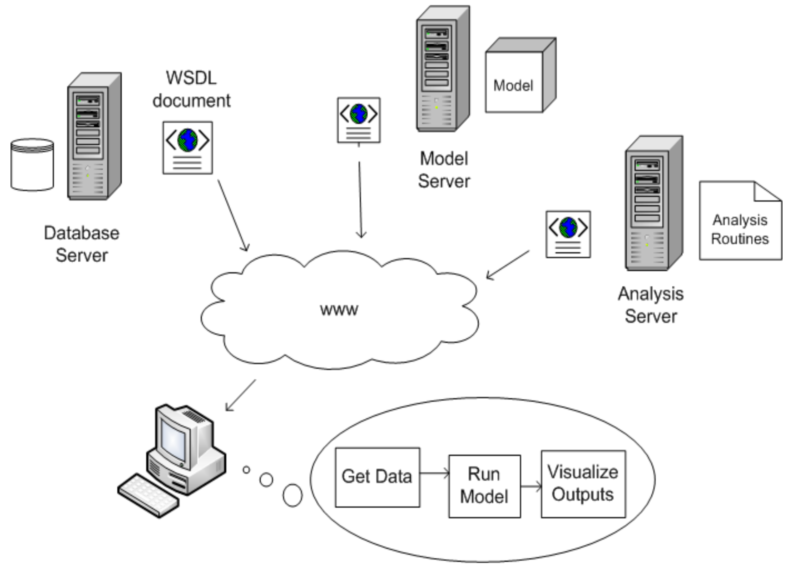


Figure 7-2 Vision for coupling ESMF within OpenMI workflows using a service-oriented architecture

As a second simple demonstration of this concept, and to demonstrate services that could be used to expose custom model components as well as ESMF model components, we creating a modeling component using the Python programming language (which is cross platform and open source) to implement a popular hydrologic modeling routine: The Muskingum Routing method. We then exposed this algorithm using a low-level service protocol called Remote Procedure Call (RPC). A C#.Net OpenMI-compliant component then services as client for the service and allows the Python model to serve as a modeling component within an OpenMI workflow (or configuration).

The second approach for linking the HIS with ESMF was to extend ESMF components to output data in the WaterML format. The ESMF team was able to create a proof-of-concept example where an ESMF component was extended to output time series using the WaterML schema, in addition to other supported output files (e.g. NetCDF). Because WaterML is the standard communication language used within the HIS system, model outputs from ESMF components could be easily consumed as data services by HIS client applications. The disadvantage of this approach over the web service approach is client applications could not control model runs – they would only able to access model output files. For many applications, however, this will be a suitable level of interoperability between the HIS and ESMF.

7.4 HydroModeler: A Modeling Plug-In for HydroDesktop

The third outcome from the University of South Carolina is the HydroModeler plug-in for the HydroDesktop application. HydroDesktop is intended to be the primary client application for interaction with the Water Data Services (WDS). An important design criteria for HydroDesktop is to adopt an open architecture where third party developers can add plug-ins that appear as new toolbars or tabs within HydroDesktop. While the core HydroDesktop application handles data access, management, and visualization, hydrologic modeling is beyond its core design specifications. For this reason, we are building a plug-in for HydroDesktop called HydroModeler that will allow end users to model hydrologic systems using OpenMI-compliant components.

The HydroModeler plug-in in its current state allows users to have essentially the same capabilities as the OpenMI Configuration Editor, but from within the HydroDesktop environment (Figure 7-3). However, by embedding this OpenMI Configuration Editor functionality within HydroDesktop, it opens the door to more tightly integrating the OpenMI Configuration Editor with HydroDesktop core functionality for data management, geoprocessing, and visualization. One example is work underway to remove the native OpenMI Configuration Editor geographic data viewer in favor of HydroDesktop GIS capabilities. Other planned additions to HydroModeler include tools for managing and authoring modeling components, more fluid interaction between the map, graph, and model tabs, and other enhancements focused on providing a use friendly, plug-and-play modeling environment within HydroDesktop.

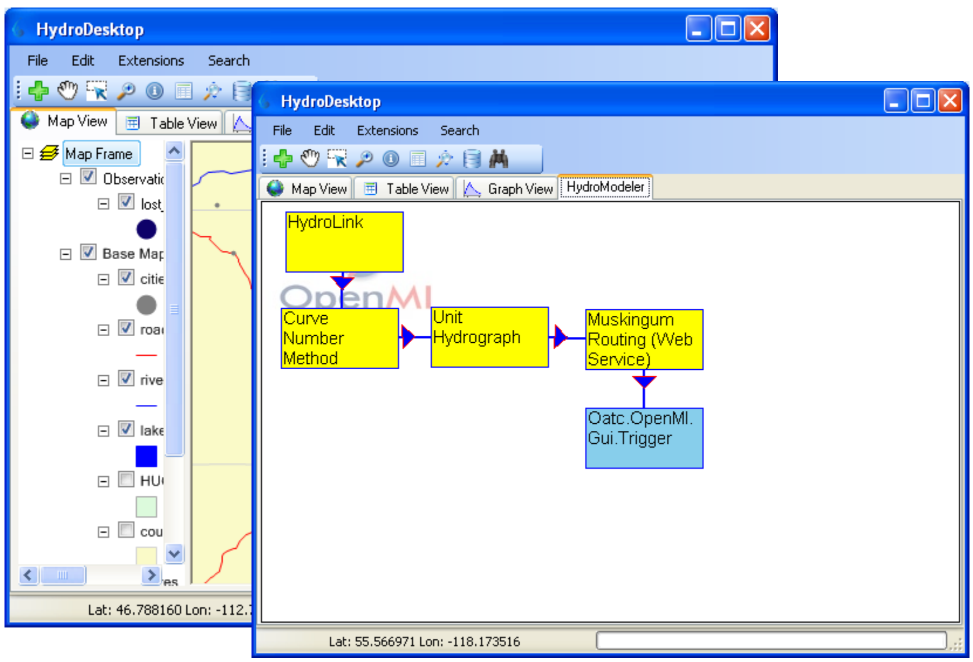


Figure 7-3 Prototype HydroModeler Plug-in to provide modeling capabilities from within the HydroDesktop application

7.5 Goals for 2010

The primary goal and focus of the University of South Carolina team for 2010 will be on the HydroModeler plug-in for HydroDesktop – more specifically we will focus on enhancing and documenting its usability for hydrologic scientists. There is great potential for this application to be a useful tool for hydrologic analysis when properly integrated with other tools within HydroDestkop. Our team will also continue to work with the ESMF developement team, as well as other model framework teams, to enhance interoperability between water-related modeling systems and the HIS.

Chapter 8. HIS EDUCATION AND OUTREACH

By Tim Whiteaker[5], Kimberly A. T. Schreuders[6], Yoori Choi[7]

8.1 HIS WORKSHOPS

In response to demand for training on how to publish data with HIS, the HIS team developed and led seven HIS workshops between October, 2008, and September, 2009. Each workshop was tailored to use data from the region in which the workshop was held, with the intent of showing participants that, "Yes, I can do this for my region." Workshop locations are shown in Table 8-2.

Table 8-2 HIS Workshop Locations and Dates

HIS Workshop Location	Dates
The University of Texas at Austin, Austin, TX	October 27, 2008
Bureau of Meteorology, Melbourne, Victoria, Australia	November 10-12, 2008
Kansas State University, Manhattan, KS	March 24-25, 2009
The University of Texas at Austin, Austin, TX	May 27, 2009
University of Vermont, Burlington, VT	June 4-5, 2009
Boise State University, Boise, ID	August 4-5, 2009
University of Iowa, Iowa City, IA	September 28-29, 2009

At early workshops in Texas and Australia, presentations and demonstrations bracketed discussions about HIS. At later workshops the majority of workshop time involved hands-on training where participants completed the entire data publication process starting with raw data and ending with their data being available for search in national applications such as HydroSeek. While workshop lengths have varied between 1 and 3 days, the most common HIS workshop format is 1.5 days with a schedule similar to the one below:

Day 1 9:30am-5:00pm

Morning	Introduction to CUAHSI HIS
09.30 – 09.40	Welcome
09.40 – 11.00	Introduction to the CUAHSI Hydrologic Information System (HIS)
11.00 – 11.20	Break
11.20 – 12.30	HIS Central: A catalog of the nation's water data

12.30 – 13.30 Lunch

[5] University of Texas at Austin
[6] Utah State University
[7] Consortium of Universities for the Advancement of Hydrologic Science, Inc., Medford, MA

Afternoon	Loading Data into ODM
13.30 – 15.00	How to map your data to the HIS Observations Data Model (ODM)
15.00 – 15.20	Break
15.20 – 17.00	How to load your data into ODM and analyze it

Day 2 8:00am-12:00pm

Morning	Publishing a Data Service
08.00 – 09.30	How to publish and use your data online with HIS
09.30 – 09.50	Break
09.50 – 11.00	How to register your data with HIS Central
11.00 – 12.00	Wrap-up: publication options, discussion and other questions

While much of the workshop material can be reused, a significant portion is developed from regional data at each workshop location. Additionally, the workshop computers are all set up on location, giving the technical staff of the workshop host thorough training in how to set up HIS Servers. The materials prepared for each workshop generally include:

- Data / Files
  - Raw data files of local hydrologic observations data to be used in the hands-on training portion of the workshop
  - Solution files consisting of transformed raw data files for loading into the Observations Data Model database, to be used of the transformation process could not be completed by the workshop participant due to time or other constraints
  - Workshop document giving detailed descriptions of all steps performed in the hands-on training
  - Workshop agenda
  - Survey evaluating workshop and presenter quality, as well as identifying needs of the community regarding HIS functionality
  - Demonstration files and procedures
  - PowerPoint presentations about HIS and data publication
- Hardware (set up by the workshop host)
  - Lab computers for hands-on training, installed with software according to the HIS Lite configuration described at http://his.cuahsi.org/hisserver.html
  - Copy of workshop data and files with unique observation site locations for each computer
- User accounts at a special HIS Central website set up specifically for workshop use

Surveys to evaluate the workshop were handed out to all participants at the end of each workshop. Completed surveys were then compiled and delivered to the CUAHSI program office. In overall, participants are very satisfied about the workshops. It was very valuable to them for the future use in their research work. In general, participants are very happy for workshop content and materials we provide during the workshop and have hands-on experience to complete the entire data publication process. Reception of the HIS workshops has been generally good, and usually sparks further interaction between regional players and the HIS team as new HIS servers are set up by those who attended the workshop.

The first five HIS workshops were led by Dr. Tim Whiteaker of The University of Texas at Austin. Beginning with the Vermont HIS workshop in June, 2009, Dr. Whiteaker began training Yoori Choi, the present CUAHSI User Support Specialist, in how to prepare and present HIS workshops. Under guidance from Tim Whiteaker, Ms. Choi led her first HIS workshop at the University of Iowa in September, 2009. This event marked the final phase in transitioning primary responsibility of giving HIS workshops from The University of Texas at Austin to the CUAHSI User Support Specialist.

The sections below list details for each HIS workshop that was given between October, 2008 and September 2009.

8.1.1 Texas HIS, October 27, 2008, Austin, TX

Workshop Location: The University of Texas at Austin, Austin, TX
Sponsor: Texas Water Development Board
Contact: Jorge Izaguirre <Jorge.Izaguirre@twdb.state.tx.us>
Duration: 1 day
Audience: Texas agency staff, consultants, researchers, and students
Topics:

- Background of national HIS
- Motivation for Texas HIS
- waterdatafortexas.org – one-stop data viewer and data access portal
- Data publication methods with the Observations Data Model (ODM) and WaterOneFlow

Demos:

- Searching for data with HydroSeek
- Accessing data with HydroExcel
- Loading Texas data into an ODM database
- Publishing Texas data with WaterOneFlow

Hands-on Training: There was no hands-on training in this workshop.

The demos for loading and publishing Texas data used water quality data collected by the Texas Water Development Board for Christmas Bay along the Texas Coast. The data were modified for workshop use.

8.1.2 CUAHSI HIS, November 10-12, 2008, Melbourne, Australia

Workshop Location: Bureau of Meteorology, Melbourne, Victoria, Australia
Sponsor: Bureau of Meteorology
Contact: Tony Boston <T.Boston@bom.gov.au>
Duration: 3 days
Audience: 30 staff members from the Bureau of Meteorology and CSIRO
Topics:

- Introduction to CUAHSI and CUAHSI HIS
- HIS Server

    - Observations Data Model
    - WaterOneFlow Web Service
    - WaterML
    - ODM Tools
    - Data Loaders
- HIS Central
    - HIS Central Metadata Catalog
    - Ontology, HydroTagger
    - HydroSeek
- WRAP Hydro tool set and Schematic Processor – A methodology for supporting water rights analysis in geographic information systems

Demos:

- Searching for data with HydroSeek
- Accessing data with HydroExcel
- Loading data into an ODM database
- Publishing data with WaterOneFlow
- Registering a data service with HIS Central

Hands-on Training: There was no hands-on training in this workshop.

The demos for loading and publishing data used water quality data collected by the Texas Water Development Board for Christmas Bay along the Texas Coast. The data were modified for workshop use.

8.1.3 CUAHSI HIS, March 24-25, 2009, Manhattan, KS

Workshop Location: Kansas State University, Manhattan, KS
Sponsors: Provost's Office Targeted Excellence Program at Kansas State University, Kansas National Science Foundation EPSCoR program, Department of Civil Engineering at Kansas State University
Contact: David Steward <steward@ksu.edu>
Duration: 1.5 days
Audience: 20 researchers, students, and local agency staff
Topics:

- Introduction to CUAHSI and CUAHSI HIS
- HIS Server
    - Observations Data Model
    - WaterOneFlow Web Service
    - ODM Tools
    - Data Loaders
- HIS Central
    - HIS Central Metadata Catalog
    - Ontology, HydroTagger
    - HydroSeek

Demos:

- Searching for data with HydroSeek
- Accessing data with HydroExcel

Hands-on Training:

- How to map Kansas data to the HIS Observations Data Model
- How to load Kansas data into an ODM database and analyze it
- How to publish and use Kansas data online with HIS
- How to register Kansas data with HIS Central

The hands-on training used groundwater levels measured at wells in the Ogallala Aquifer. These data were extracted from the Water Information Storage and Retrieval Database (WIZARD) and remain largely untouched, giving the user the experience of what it takes to load real data into HIS. Each well has many water level observations measured sporadically through time, sometimes months apart, sometimes years apart.

8.1.4 Texas HIS, May 27, 2009, Austin, TX

Workshop Location: The University of Texas at Austin, Austin, TX
Sponsor: Texas Water Development Board
Contact: Jorge Izaguirre <Jorge.Izaguirre@twdb.state.tx.us>
Duration: 1 day
Audience: 35 Texas agency staff, consultants, researchers, and students
Topics:

- CUAHSI HIS, and how it led to Texas HIS
- HIS for Data Publication
- Texas HIS Viewer

Demos:

- Accessing data with HydroExcel

Hands-on Training:

- How to load Texas data into the Observations Data Model
- How to publish Texas data online with Water Data Services

The hands-on training used time series of water quality data measured in several Texas bays from 1987 to 1995. Measurements are taken at various depths below the water surface and sporadically in time. The data include measurements of salinity, dissolved oxygen, conductivity, temperature, and pH.

8.1.5 CUAHSI HIS, June 4-5, 2009, Burlington, VT

Workshop Location: University of Vermont, Burlington, VT
Sponsors: Vermont EPSCoR

Contact: Kristen Hallock <khallock@uvm.edu>
Duration: 1.5 days
Audience: 17 researchers, students, and local agency staff
Topics:

- Introduction to CUAHSI and CUAHSI HIS
- HIS Server
  - Observations Data Model
  - WaterOneFlow Web Service
  - ODM Tools
  - Data Loaders
- HIS Central
  - HIS Central Metadata Catalog
  - Ontology, HydroTagger
  - HydroSeek

Demos:

- Searching for data with HydroSeek
- Accessing data with HydroExcel

Hands-on Training:

- How to map Vermont data to the HIS Observations Data Model
- How to load Vermont data into an ODM database and analyze it
- How to publish and use Vermont data online with HIS
- How to register Vermont data with HIS Central

The hands-on training used water quality data measured for the Lake Champlain Long-term Water Quality and Biological Monitoring Project. The data include measurements of nitrogen, phosphorus, temperature, total suspended solids, and chlorophyll a, taken from 1992 to 2007.

8.1.6 ICEWATER HIS Training, August 4-5, 2009, Boise, ID

Workshop Location: Boise State University, Boise, ID
Sponsors: INRA ICEWATER Project
Contact: Pam Aishlin <pamaishlin@boisestate.edu>
Duration: 2 days
Audience: 15 data managers, researchers, students, and local agency staff
Topics:

- Review of HIS and Why HIS
- Overview of ICEWATER CI Architecture Using HIS
- Setting Up and Configuring HIS Server
  - Introduction to the Observations Data Model with Examples
  - ODM Streaming Data Loader and ODM Data Loader

    - Visualizing, Querying and Editing Data Using ODM Tools
    - Publishing Data Using WaterOneFlow Web Services
- Registering WaterOneFlow Web Services with HIS Central and Tagging Variables
- HIS Desktop for ICEWATER
- Phase II HIS Server Components for ICEWATER
    - Publishing Spatial Datasets Using ArcGIS Server
    - Implementing Time Series Analyst
    - Implementing ICEWATER Map Server

Demos:

- HydroSeek
- HydroExcel
- ODM Streaming Data Loader
- ODM Data Loader
- Loading Data Using SSIS
- Visualizing, Querying and Editing Data Using ODM Tools
- HIS Central
- Registering WaterOneFlow Web Services and Tagging Variables
- HIS Desktop for ICEWATER
- Time Series Analyst
- ICEWATER Map Server

Hands-on Training:

- Translating and Loading Data into ODM
- Working with Data in ODM Using ODM Tools
- Publishing an ODM Database with WaterOneFlow Web Services
- Registering WaterOneFlow Web Services with HIS Central and Tagging Variables
- Implementing Phase II ICEWATER HIS Components
- Working with Your Own Data

This workshop was focused on providing ICEWATER data managers with the skills necessary to set-up and load data into an HIS Server with the new geographic extensions. Most of the hands-on training used water quality data measured for the Lake Champlain Long-term Water Quality and Biological Monitoring Project. The data include measurements of nitrogen, phosphorus, temperature, total suspended solids, and chlorophyll a, taken from 1992 to 2007. In addition, each participant was encouraged to bring some of their own data to load into an ODM database in the final exercise. The exercises allowed each participant to completely build their own full HIS Server with geographic extensions by the end of the two days.

8.1.7 CUAHSI HIS, September 28-29, 2009, Iowa City, IA

Workshop Location: University of Iowa, Iowa City, IA
Sponsors: The University of Iowa
Contact: Marian Muste <marian-muste@uiowa.edu>

Duration: 1.5 days
Audience: 29 data managers, researchers, students, and local agency staff
Topics:

- Introduction to CUAHSI and CUAHSI HIS
- HIS Server
  - Observations Data Model
  - WaterOneFlow Web Service
  - ODM Tools
  - Data Loaders
- HIS Central
  - HIS Central Metadata Catalog
  - Ontology, HydroTagger
  - HydroSeek

Demos:

- Searching for data with HydroSeek
- Accessing data with HydroExcel

Hands-on Training:

- How to map Iowa data to the HIS Observations Data Model
- How to load Iowa data into an ODM database and analyze it
- How to publish and use Iowa data online with HIS
- How to register Iowa data with HIS Central

The hands-on training used water quality data measured for the Clear Creek Watershed. The data include measurements of dissolved oxygen, NO3N (nitrogen, nitrate (NO3) as N, unfiltered), pH, temperature, and turbidity, taken from July 2008 to September 2008.

8.2 User Support

Yoori Choi was hired as the User Support specialist by CUAHSI on August 1, 2009. This marks a commitment for continued user support beyond the life of the HIS Project and an assumption of an important role by the CUAHSI office.

The User Support Specialist:

- Supports the user community through synchronous and asynchronous communication
- Communicates user requests and issues to the HIS development team
- Develops training materials
- Interacts with users who publish data and analysts who seek to discover and to access data using HIS services, both in a one-on-one support mode and in formal training settings
- Reviews documentations developed by HIS team for completeness, accuracy and comprehensibility

We have implemented Water Data Services (WDS) User Forum web site.
http://blog.hydrologicscience.org/?page_id=13
This is the community center for HIS users. Our initial user group is participants who attended CUAHSI HIS Workshops. We post any recent news about our HIS work and respond to any issues posted by users either technical problem or any suggestion for improvement for our work. Depending upon the issue, we transmit it to the development team for action. In addition, we raise user comments during HIS weekly conference call to group for consideration. This forum is one direct way we can communicate with users and know what they want to get from HIS work.

We have many efficient tools for Hydrologic community. Now it's time more to know what users think and what users want to see in our future work and what they want to benefit from our work. Thus through forum web site and workshop and conference/meetings, we will interact with the user community.

8.3 HIS Website

The HIS website at http://his.cuahsi.org continues to be improved and updated. Significant changes this year include:

- Continuously updated the What's New section on the home page
- Added an introductory video to home page
- Added a new page for HydroDesktop
- Added a new page for Time Series Analyst
- Posted new versions of the various HIS software components as they have become available
- Added new pages for the recently held HIS Workshops
- Added new publications to the Publications page
- Added several software tutorials and case studies
- Updated the Feedback Form to work with the Jira Bug Tracking system at SDSC
- Added Google Analytics capability to track use of the website

8.3.1 Google Analytics Summary for HIS Website

We continue to have a significant and wide audience for the HIS website. It should be noted that as the quantity and quality of the content on the website has grown over the year, the dwell time has consistently increased each quarter.

Quarter	Visits	Page Views	Dwell Time (minutes)	Unique Visitors	Unique Countries
First Quarter 2009*	3294	10425	3:30	1657	70
Second Quarter 2009	4312	15578	3:53	2266	80
Third Quarter 2009	4265	15004	4:12	2156	82

* Partial quarter. Tracking began on 1/19/09.

The top three web pages visited (other than the home page) were the same every quarter and were as follows:

1. WaterOneFlow Web Services & WaterML—wofws.html
2. ODM Databases—odmdatabases.html
3. HydroExcel—hydroexcel.html

Additional top pages varied slightly over the 3 quarters. Here is the list for the 3rd quarter 2009:

4. Using Data—datausers.html
5. HydroObjects—hydroobjects.html
6. ODM Tools—odmtools.ntml
7. HIS Server—hisserver.html
8. Publishing Data—datapublishers.html
9. Publications—publications.html
10. DASH—dash.html

8.4 Collaborations

Collaborations are another avenue for involving the hydrologic community in the HIS project. The HIS team initiated and/or expanded collaborations with the following individual and/or organizations in the past year:

- The Renaissance Computing Institute (Renci) on the Sensor Data Bus (SDB) project to share ODM ideas and developments, including ODM Tools. See http://www.sensordatabus.org and http://ogc.codeplex.com.
- INRA's ICEWATER Network to develop and test HIS Servers including the new Spatial Data Publication Extension. See http://icewater.inra.org.
- Idaho State University (ISU) and Australia's Commonwealth Scientific and Industrial Research Organization (CSIRO) on HydroDesktop and Hydrologist Workbench. See http://www.csiro.au/science/Hydrologists-workbench.html.
- Nermin Sarlak of the Republic of Turkey on implementing the HIS system at several test sites in that country.

8.5 HIS Development Wiki

The HIS team continues to develop and expand its HIS Development Wiki (http://river.sdsc.edu/wiki/). Major new topics added or expanded this year include:

- HIS Glossary and Acronyms List (contains over 130 terms and definitions)—http://river.sdsc.edu/wiki/HIS%20Glossary.ashx
- WaterML and WaterOneFlow Issues (including draft information on version 1.1—http://river.sdsc.edu/wiki/WaterML.ashx, http://river.sdsc.edu/wiki/CUAHSI%20Web%20Services%20Specification.ashx, http://river.sdsc.edu/wiki/WaterOneFlow.ashx
- New and Improved Custom WaterOneFlow Services—http://river.sdsc.edu/wiki/NCDC%20Webservice%20Development%20Notes.ashx, http://river.sdsc.edu/wiki/USGS%20NWIS%20webservices.ashx
- HIS Central Metadata Web Services—http://river.sdsc.edu/wiki/HISCentralWebServices.ashx
- Ontology Issues—http://river.sdsc.edu/wiki/Ontology%20System%20Requirements.ashx, http://river.sdsc.edu/wiki/Ontology%20API.ashx

8.6 Publications and Presentations

Formal publications and presentations are another very necessary path for communication, especially in the academic community. The HIS team produced the following publications and presentations in the last year.

- Ames, D. P., J. S. Horsburgh, J. Goodall, D. G. Tarboton, T. Whiteaker, and D. R. Maidment (2009), Introducing the open source CUAHSI Hydrologic Information System desktop application (HIS Desktop), in Anderssen, R. S., R. D. Braddock, and L.T.H. Newham (eds) 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, pp. 4353-4359, ISBN: 978-0-9758400-7-8.
- Beran B., and M. Piasecki, (2009), Engineering new paths to hydrologic data, Computers and GeoSciences, 35(4), 753-760, doi:10.1016/j.cageo.2008.02.017.
- Beran B., and M. Piasecki (2008), Availability and coverage of hydrologic data in the US Geological Survey National Water Information System (NWIS) and US Environmental Protection Agency Storage and Retrieval System (STORET), Earth Science Informatics, 1(3-4), 119-129, doi:10.1007/s12145-008-0015-2.
- Horsburgh, J. S., D. G. Tarboton, M. Piasecki, D. R. Maidment, I. Zaslavsky, D. Valentine and T.Whitenack, (2009), "An integrated system for publishing environmental observations data," Environmental Modelling & Software, 24(8): 879-888, http://dx.doi.org/10.1016/j.envsoft.2009.01.002.
- Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky, (2008), "A Relational Model for Environmental and Water Resources Data," Water Resour. Res., 44: W05406, doi:10.1029/2007WR006392.
- Maidment, D. R., R. P. Hooper, D. G. Tarboton and I. Zaslavsky (2009), "Accessing and Sharing Data Using CUAHSI Water Data Services," in Hydroinformatics in Hydrology, Hydrogeology and Water Resources, Edited by I. Cluckie, Y. Chen, V. Babovic, L. Konikow, A. Mynett, S. Demuth and D. A. Savic, Proceedings of Symposium JS4 held in Hyderabad, India, September 2009, IAHS Publ. 331, Hyderabad, India, p.213-223, http://iahs.info/redbooks/331.htm.
- Tarboton, D. G., J. S. Horsburgh, D. R. Maidment, T. Whiteaker, I. Zaslavsky, M. Piasecki, J. Goodall, D. Valentine and T. Whitenack, (2009), "Development of a Community Hydrologic Information System," 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, ed. R. S. Anderssen, R. D. Braddock and L. T. H. Newham, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, July 2009, p.988-994, http://www.mssanz.org.au/modsim09/C4/tarboton_C4.pdf.
- Tarboton, D. G., J. S. Horsburgh, T. L. Whiteaker, D. R. Maidment and I. Zaslavsky, (2008), "Accessing and Sharing Data Using the CUAHSI Hydrologic Information System," Eos Trans. AGU, 89(53): Fall Meet. Suppl., Abstract C21A-0504 INVITED.
- Tarboton, D. G., D.R. Maidment, I. Zaslavsky, M. Piasecki, J. Goodall, (2008), "The CUAHSI Hydrologic Information System", presentation to Western States Water Council, Salt Lake City, Utah, November 19.
- I. Zaslavsky, D. Valentine, and D.R.Maidment (2008) "Integration of Hydrologic Observations from Government and Academic Data Collections with the Consortium of Universities for the Advancement of Hydrologic Sciences (CUAHSI) Hydrologic Information System". In: Brady, S.R., Sinha, A.K., and Gundersen, L.C., editors, 2008, Geoinformatics 2008-Data to Knowledge, Proceedings: U.S. Geological Survey Scientific Investigations Report 2008 -5172, pp. 60-62.
- B. Beran, J. Goodall, D. Valentine, I. Zaslavsky, M. Piasecki (2009) "Standardizing Access to Hydrologic Data Repositories through Web Services," geows, pp.64-67, 2009 International Conference on Advanced

Geographic Information Systems & Web Services. ISBN: 978-1-4244-3363-6. Presented by David Valentine (Best Conference Paper Award)