

DEVELOPMENT OF AN INFORMATION SYSTEM FOR THE HYDROLOGIC COMMUNITY

MICHAEL PIASECKI

*Department of Civil, Architectural, & Environmental Engrg, Drexel University
3141 Chestnut Street, Philadelphia, PA 19104, USA*

DAN AMES, Idaho State University, *921 South 8th Ave., Pocatello, ID 83209*

JON GOODALL, University of South Carolina, *Columbia, SC 29208*

RICK HOOPER, CUAHSI, *2000 Florida Ave, N.W., Washington, D.C. 20009*

JEFF HORSBURGH, Utah State University, *Logan, UT 84322-1400*

DAVID MAIDMENT, *University of Texas, 1 University Station, Austin TX 78712*

DAVID TARBOTON, Utah State University, *Logan, UT 84322-1400*

ILYA ZASLAVSKY, San Diego Supercomputer Center, *San Diego, CA 92037*

The Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI) Hydrologic Information System (HIS) project is developing CyberInfrastructure (CI), and services to support the advance of hydrologic science in the United States. CUAHSI HIS developed CI supports storage of water observations data in a relational database, publication via web-services on the internet, federation with water observations data published by water agencies such as the U.S. Geological survey, and search across the various data holdings in the network. These advanced data access and analysis capability are provided through the use of web services using a format known as WaterML. An Observations Data Model (ODM) has been developed to provide a standard relational database schema for storing observatory and individual investigator data and also to serve as a central water metadata catalogue. We report on the latest developments of the Hydrologic Information Science (HIS) such as HydroServer and an open-source based GIS desktop application (HydroDesktop) for accessing web-services supported water information. We provide an overview of the system, the current state, and future plans.

INTRODUCTION

The Consortium of Universities for the Advancement of Hydrologic Sciences (CUAHSI) (<http://www.cuahsi.org>) is an international organization of universities and hydrologic scientists focused on fostering advancements in the hydrologic sciences that is supported by the U.S. National Science Foundation.. The main goal of this effort is to develop, prioritize and disseminate a broad-based research and education agenda for the hydrologic sciences derived from a continuous process that engages both research and application professionals. The effort has self-organized itself into several sub interests such as instrumentation, information science, synthesis, and education each of which

pursues specific goals and objectives to enhance the main CUAHSI goal. These groups pursue funding on individual tracks, however are overseen by CUAHSI's Board of Directors and also engage community based advisory committees that oversee and provide feedback into the development process.

The Hydrologic Information System (HIS) project aims at developing CyberInfrastructure, CI, and services to support the advance of hydrologic science in the United States (Maidment [1]; Horsburgh *et al.* [2]; Tarboton *et al.* [3]). Hydrologic information science involves the description of hydrologic environments in a consistent way, using data models for information integration and software tools that are either customized or based on commercial products to directly interact with the deployed data storage and communication components.

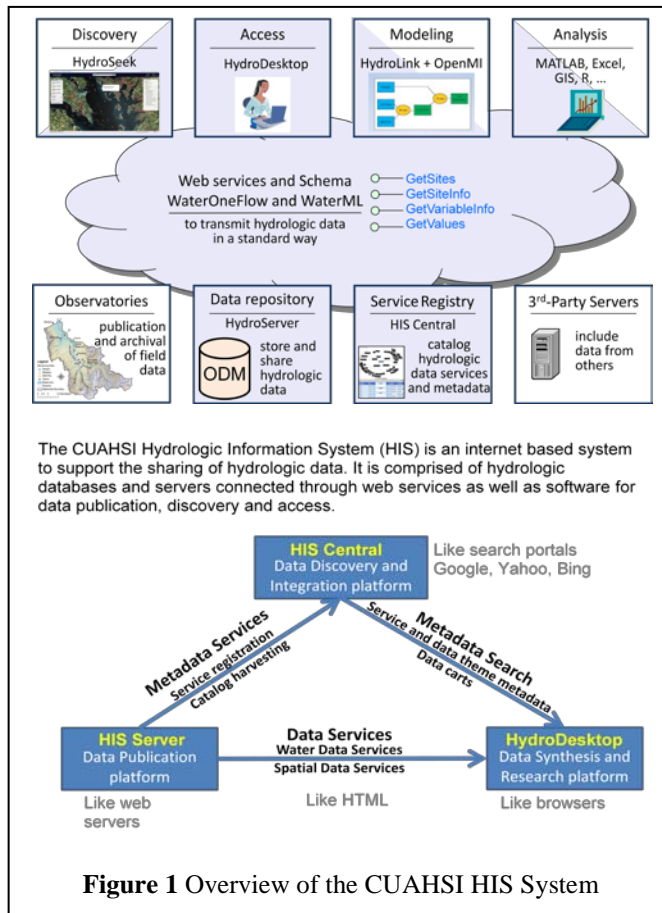
CUAHSI HIS developed CI supports storage of water observations data in a relational database (Horsburgh *et al.* [2]), publication via web-services (Zaslavsky *et al.* [4]) on the internet, federation with water observations data published by water agencies such as the U.S. Geological survey, and search across the various data holdings in the network. These advanced data access and analysis capability are provided through the use of web services using a format known as WaterML (Zaslavsky *et al.* [4]). An Observations Data Model (ODM) has been developed (Horsburgh *et al.* [5]) to provide a standard relational database schema for storing observatory and individual investigator data and to serve as a foundation for the hydrologic information system tools and web services.

This paper will i) give a general overview of the HIS system components, and ii) report on the latest developments of the Hydrologic Information Science (HIS) group. In addition to having developed applications that can ingest the HIS web services such as HydroExcel (EXCEL interface), HydroViewer (ArcGIS based), HydroSeek (ontology based search engine with map interface, Beran and Piasecki [6]), recent efforts have focused in the development of a open-source based GIS application (HydroDesktop) and also a server set up for a regional scale system called HydroServer for accessing web-services supported water information. We will provide an overview of the system, the current state, and future plans.

SYSTEM OVERVIEW

The HIS Environment

CUAHSI's Hydrologic Information Systems (HIS) program has been developing to provide infrastructure to support the interdisciplinary study of hydrologic and related environmental systems, across spatial and temporal scales. The goal of the HIS program is to develop tools that integrate the storage and distribution of data and that facilitate analysis, visualization, and modeling of data. Within the context of the CUAHSI HIS project, several data storage and delivery tools and standards have been developed as depicted in Fig.1 (upper image). These include WaterOneFlow web services (a defined



set of web-services for discovery and download of observation data); WaterML (an XML based language for transmitting observation datasets via web services); HydroServer (a hardware/software system for delivering WaterML from a web server); Observations Data Model (ODM an extensive database schema used to store all data elements associated with point observation data); ODM Tools (a toolkit for managing data in an ODM SQL Server database); HIS Central (a set of web-based tools and central data repository at the San Diego Super Computer Center for hosting and forwarding

data from national databases in the WaterML format); HydroExcel (a Microsoft Excel spreadsheet with built-in macros for accessing and retrieving data from a HydroServer into a local file); HydroGet (an extension for ArcGIS that can access and ingest HIS data via a map interface); HydroSeek (a website that aggregates data from multiple HIS Servers and provides ontology based data search capabilities); HydroObjects (programming components used to parse WaterML into standard objects and classes used in custom software projects); and other related tools, services, and data schemas (Goodall et al. 2008, Maidment et al. 2006). The lower image in Fig.1 shows the relationship between the two new developments and the central metadata catalogue, in which HydroServer is a local database installation (either individual researcher or for a region) and HydroDesktop the user download application that either communicates with a local (or regional) DB installation) or the central metadata catalogue.

Each of these HIS project components fills a specific capability niche that has either been deemed a fundamental building block by the HIS team or has been identified by users as a need. Together, the tools components present a relatively complete software

“stack” to support the consistent storage and delivery of hydrologic and other environmental observation data. Another important arena of activity is the increased participation of the CUAHSI HIS team in standardization efforts, i.e. to move the WaterML signature into the OGC process of developing international standards. In addition, team members are also involved in the WMO and other international agencies to further develop metadata standards for hydrologic data, i.e. via the domain working group for hydrology (DWG_Hydro). Also, the HIS team is playing an increasingly pivotal role in bringing standardization efforts to US governmental agencies who seek to benefit from the HIS developments in trying to unify and overcome system and semantic heterogeneities. As a result, the HIS has developed ties to all major data agencies in the US both on national and state levels.

While these developments have resulted in wide spread recognition for the work done and also build considerable interest in the cyberinfrastructure developed, it has also become clear that there is no one system that would fulfill all user’s needs. Consequently, the system developed is in fact a conglomeration of various components (see above in Figure 1) that have various use trajectories addressing differing user data discovery and retrieval interests. The HIS team is thus continuously developing new application that are in response to what the user community expressed as missing items.

HydroServer

HydroServer is intended to enable producers of hydrologic data to join a growing, distributed network of published water data services by establishing their own server and publishing their own data. HydroServer has a number of features that aid in achieving this goal: it publishes data via Web services, includes software for management of data on the server, it enables publication of spatial datasets for a study area or region and the linkage of spatial datasets with point observational datasets that are hosted on the same server, and, when services are registered with a central cataloging service called HIS

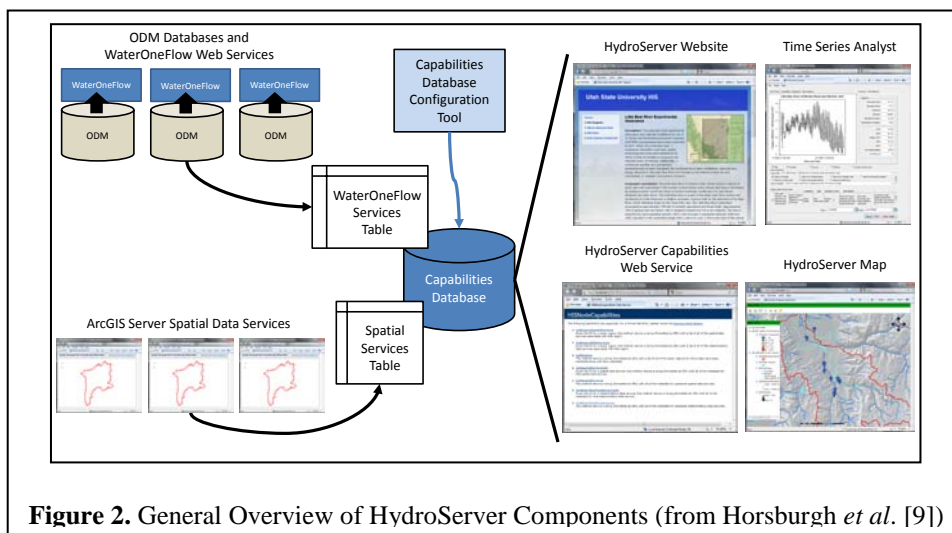


Figure 2. General Overview of HydroServer Components (from Horsburgh *et al.* [9])

Central, the data become discoverable through HIS client applications.

HydroServer uses the concept of a study “region” that can have any extent for example a watershed but also a region with a political boundary. It encompasses point time series data sets that is collected in this region but also permits geospatial data sets describing this region to be associated with the point data sets. Thus, it is a concept for a digital watershed that describes both the spatial and temporal characteristics of a region. Originally conceived in its first version (Whitenack et al., [7]; [8]) in 2007 this version, based on MS SQL Server 2008 and ArcGIS Server 9.3.1 represents a number of important updates such as a new mapping interface, capabilities database and services, Time Series Analyst, and a configurable HydroServer website. Figure 2 shows an overview of the major components of HydroServer.

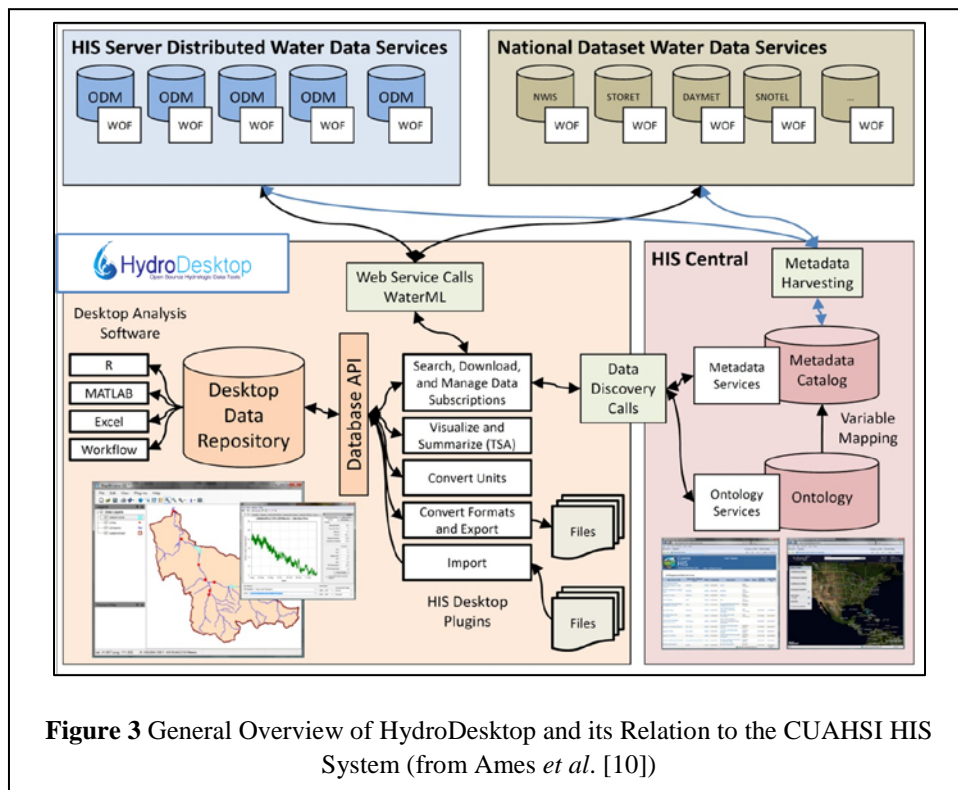
The system consists of one (or several) instances of the Observations Data Model (ODM) database into which observations (time series data in the broadest sense) made at fixed points are loaded. Each ODM database publishes its contents using WaterOneFlow Web services that publish the data on the Internet in Water Markup Language (WaterML). New in this configuration is the presence of a geospatial database which publishes its contents as services using ESRI’s ArcGIS Server. Also new is a so-called Capabilities Database that 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 enables HydroServer administrators to edit the contents of the Capabilities database. A suite of Web applications connects to the Capabilities database and makes the data published on the HydroServer available via a Web browser thus making the HydroServer capabilities self describing. These include: 1) an Internet map that combines GIS datasets and time series published using ODM and WaterOneFlow services for a region; 2) the Time Series Analyst for visualizing observational data, 3) a HydroServer website that describes the regions and services published on the HydroServer; and 4) the Capabilities Web service, which publishes the capabilities (i.e., a metadata description of the services that the HydroServer contains) of the HydroServer in extensible markup language (XML) format. The services available include the following methods: 1) GetRegions; 2) GetRegionInfo; 3) GetWaterOneFlowServices; 4) GetWaterOneFlowServiceInfo; 5) GetMapServices; 6) GetMapServiceInfo; 7) GetRegionWaterOneFlowServices; and 8) GetRegionMapServices. The latest version of HydroServer including documentation is available at <http://his.cuahsi.org>.

HydroDesktop

One of the new developments of the HIS team is a client based workbench that is a highly tuned and customized “hydrology” version of a public domain GIS environment, called HydroDesktop. HydroDesktop is intended to solve the problem of how to obtain, organize, and manage hydrologic data on a user’s computer to support analysis and modeling. This software fills a gap in the HIS project providing a visualization tool for

HIS Server based data. The software is intended to be a platform for the integration of HIS 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 includes the use of a plug-in architecture and data abstraction layer that will allow extension of the core functionality. Following this approach, the system provides local access to data obtained from distributed data services that are part of the internet-based, service oriented architecture (SOA) that the CUAHSI HIS project has developed for the sharing of HIS data.

HydroDesktop is based on the open source GIS code MapWindow (<http://www.mapwindow.org>) which is a client-side (desktop) software tool that ultimately will run on multiple operating systems and will provide a highly usable level of access to HIS services. The software is envisioned to provide many key capabilities of existing HIS tools (data query, map-based visualization, data download, local data maintenance, editing, graphing, etc.) as well as new capabilities not currently included in any of the existing HIS components (data export to some model-specific data formats, linkage with integrated modeling systems such as OpenMI, and data upload to the HIS server from the local desktop software). The role of HydroDesktop within the HIS framework is depicted in Figure 3.



The HydroDesktop features several functionalities that are new to the HIS application suite. While it uses the same web service calls that are at the core of HIS CI, it embeds a download strategy that allows the user to preserve the data locally once downloaded. The sequence is to save a copy of the download file(s) in XML format on the user's hard disk with the purpose of saving them as retrieved from the web service calls. These files are then parsed and the content is subsequently stored in a local relational database (which is part of the HydroDesktop distribution) using the available data series metadata from the web service calls. The relational schema of the data repository database is semantically similar to the CUAHSI ODM database design (Horsburgh *et al.* [2]) with similar naming conventions and data types, but has been modified and extended to facilitate management of the data series that have been downloaded and storage of provenance information.

SUMMARY AND FUTURE WORK

This paper provides a short overview of the CUAHSI HIS activities and what has been developed to date. It also points out the ongoing activities related to standardization efforts, which are a necessary step to bring the technical developments to bear at a larger scale and thus substantially increase significance of the work carried out. The involvement in the WMO and OGC workgroups are activities spanning the globe and will work for the HIS efforts.

The latest developments, i.e. HydroServer and HydroDesktop, are both appliances that are designed to fill two gaps hitherto not covered; the former now expands the concept of a digital watershed by adding a geospatial database side-by-side the time series data base thus providing geospatial context to the point oriented time series observations. This installation facilitated the creation of a "capabilities" database whose contents are exposed via a specifically designed set of web services. This makes the HydroServer installation completely self describing.

HydroDesktop is a downloadable application that has been derived from the public domain GIS environment called MapWindow. It has been customized to access WaterOneFlow web services thus permitting search and retrieval capabilities that at the core of the HIS efforts thus making it a GIS based hydrologic workbench. Because of its plug-in framework HydroDesktop can easily be extended to include modules for analysis, visualization, curation and formation of local data subsets that can be retrieved and conglomerated for themes and regions of interest.

The next steps of the HIS team are the continued drive for standardization and hardening of the software components currently under development. The HIS team has been able to obtain additional funding to work on emerging complex aspects of the work, most recently the further development of a hydrologic keyword ontology and an associated community based development system for it. Another attempt is currently being made to seek funding for exploring the use of the "cloud computing" capabilities for the HIS components.

ACKNOWLEDGEMENTS

Components of the work were in part or in full supported by National Science Foundation grants EAR 0413265, and EAR 0622374, and by the Inland Northwest Research Alliance (<http://www.inra.org>). Special acknowledgements go to the CUAHSI HIS development team for providing continued feedback and leadership while developing the HIS cyberinfrastructure.

REFERENCES

- [1] Maidment, D. R., ed.. CUAHSI Hydrologic Information System: 2009 Status Report, Consortium of Universities for the Advancement of Hydrologic Science, Inc, 79 p., http://his.cuahsi.org/documents/HISOverview_2009.pdf (2009)
- [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 Modeling and Software*, 24, 879-888, doi:10.1016/j.envsoft.2009.01.002 (2008)
- [3] Tarboton, D. G., J. S. Horsburgh, D. R. Maidment, T. Whiteaker, I. Zaslavsky, M. Piasecki, J. Goodall, D. Valentine, and T. Whitenack. Development of a community Hydrologic Information System, in Anderssen, R. S., R. D. Braddock, and L.T.H. Newham (eds) 18th World IMACS Congress and MODSIM09 (2009) 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. 988-994, ISBN: 978-0-9758400-7-8.
- [4] Zaslavsky, I., Valentine, D., Whiteaker, T. (Eds.),. CUAHSI WaterML. OGC Discussion Paper OGC 07-041r1. Version 0.3.0. <http://www.opengeospatial.org/standards/dp> (2007)
- [5] Horsburgh, J. S., D. G. Tarboton, D. R. Maidment, and I. Zaslavsky. A relational model for environmental and water resources data, *Water Resources Research*, 44, W05406, doi:10.1029/2007WR006392 (2009)
- [6] Beran B., Piasecki M., "Engineering New Paths to Hydrologic Data", *Computers and GeoSciences*, Elsevier, Special Issue, Vol. 35, 4, pp. 753 – 760, April (2009)
- [7] Whitenack, T., D. Valentine, I. Zaslavsky, D. Djokic. Hydrologic information system server: the software stack and the initial deployment experience. *Geoinformatics 2007*, http://gsa.confex.com/gsa/2007GE/finalprogram/abstract_122332.htm (2007a)
- [8] Whitenack, T., Zaslavsky, I., Valentine, D., Djokic, D. Data Access System for Hydrology, *Eos Trans. AGU*, 88(52), Fall Meet. Suppl., Abst. H13H-1685 (2007b)
- [9] Horsburgh, J., Tarboton, D., Schreuders, K., Maidment, D., Zaslavsky, I., Valentine, D., HydroServer: A Platform for Publishing Space-Time Hydrologic Data, WARA 2010 Spring Specialty Conference, Orlando, FL, March 29-31, (2010)
- [10] 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, 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, , July 2009, pp. 4353-4359, ISBN: 978-0-9758400-7-8.