Hydrologic Information Systems and HydroShare
Synopsis of Class 20, GIS in Water Resources, Fall 2014

The class has to date focused mostly on Geographic Information Systems (GIS) where the emphasis is on spatial geographic data in the context of ArcGIS. Today we shift our view to Hydrologic Information Systems (HIS). GIS has developed to support the storage and analysis of logically linked geographic data (Tomlinson, 2003). Similarly HIS are developing to support the analysis of logically linked hydrologic data. There is much in common, and much of the impetus for HIS has come from GIS. This lecture will describe the HIS functionality that has been developed as part of the CUAHSI HIS project and the ongoing work on HydroShare to enhance capability for sharing and collaborating using hydrologic data and models. A key motivation for HIS is the idea that comprehensive understanding of hydrology requires integration of information from multiple sources. We do not get a complete picture of the hydrologic processes in a watershed from a single sensor or single image. Rather we need to consider together data from multiple sources, collected by different investigators and organizations. Understanding the whole hydrologic system needs to come from combination of information from multiple parts.

Learning objectives

After today's class you should be able to

- Describe the components of a Hydrologic Information System and some of the key functionality of each component
- Reflect on the basic attributes needed to quantify data in an unambiguous way and how to organize this information to enhance analysis
- Discuss the placement of information in a relational data model such as ODM and the relationships required to associate attributes with data values
- Discuss the functionality and models for development of HIS components
- Reflect on and imagine the possibilities that future capability will bring as HIS evolve and become part of our everyday cyberinfrastructure.

The CUAHSI HIS

The CUAHSI HIS is an internet based system to support the sharing of hydrologic data (see http://his.cuahsi.org for details). It is comprised of hydrologic databases and servers connected through web services as well as software for data publication, discovery and access. The system that has been developed provides new opportunities for the water research community to approach the management, publication, and analysis of their data systematically. The system's flexibility in storing and enabling public access to similarly formatted data and metadata has created a community data resource from public and academic data that might otherwise have been confined to the private files of agencies or individual investigators. HIS provides an analysis environment for the integration of data from multiple sources and serves as a prototype for the infrastructure to support a network of large scale environmental observatories or research watersheds. Some of this prototype functionality is now being adopted by agencies such as the USGS and used around the world.

The CUAHSI HIS is built around the concept of a services-oriented architecture comprising: 1) a way of publishing hydrologic data in a uniform way; 2) a way of discovering and accessing remote water information archives in a uniform way; and 3) a way of displaying, synthesizing and analyzing water information and exporting it to other analysis and modeling systems. The connections among components are established by web services (Figure 1). This functionality follows the general paradigm of the Internet. HydroServer publishes data similar to the way Internet web servers publish content. HydroDesktop consumes data published from HydroServer, similar to the way web browsers consume Internet content. HydroCatalog supports data discovery based on indexed metadata similar to the way search engines support the discovery of Internet content. Syntactic (file types and formats) and semantic consistency has been a focus of HIS with an ontology and community controlled vocabulary used to harmonize the terminology used and support thematic key word based data discovery.

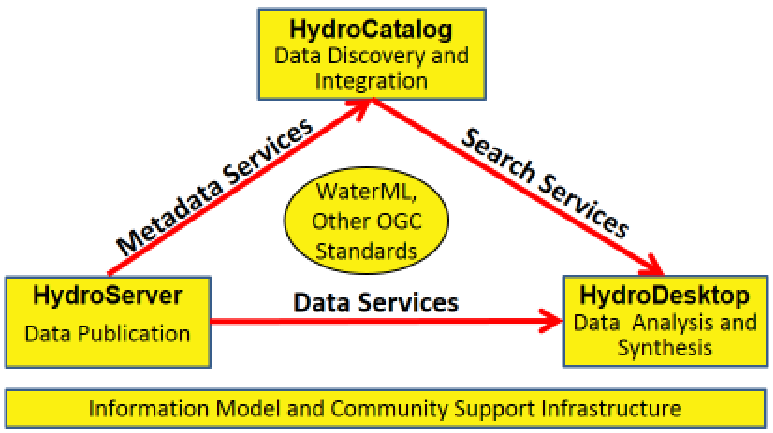

Figure 1. Components of CUAHSI HIS Services Oriented Architecture.

In the CUAHSI HIS implementation, HydroDesktop provides an analysis environment within which data from multiple sources can be discovered, accessed and integrated. HydroDesktop is available from http://hydrodesktop.codeplex.com/ and will be demonstrated during this class.

The formats for transmission of information between these systems and the interfaces that enable the communication between them (the connecting arrows in Fig. 1) are critical to the functioning of the system. CUAHSI HIS has developed WaterML, an XML based language for transmitting observation data via web services (Zaslavsky et al., 2007). WaterML has recently been accepted as an Open Geospatial Consortium (OGC) standard (http://www.opengeospatial.org/standards/waterml). This is an important step as it helps towards the adoption by large agencies and organizations around the world. CUAHSI HIS also relies on other established standards such as World Wide Web Consortium Simple Object Access Protocol (SOAP) and OGC Geographic Markup Language (GML) for transmission of information between the three primary components.

The CUAHSI HIS was designed for sharing time series data, but hydrologists often require other classes of data for their analysis. Indeed a strength of HIS is the ability to generate knowledge by integrating information from multiple sources. For this potential to be realized a broader class of data than addressed in the CUAHSI HIS needs to be accommodated and the collaborative nature of the production of hydrologic research products needs to be supported. HydroShare is a research project being pursued to address these needs. HydroShare is an online collaborative system being developed for open sharing of hydrologic data and models. The goal is to enable scientists to easily discover and access hydrologic data and models, retrieve them to their desktop, or perform analyses in a distributed computing environment that may include grid, cloud, or high performance computing model instances as necessary. It is a collaboration of 10 institutions supported by NSF sustainable software program.

The vision is to enable more rapid advances in hydrologic understanding through collaborative data sharing, analysis and modeling. HydroShare will be a community collaboration website that enables users to easily discover and access data and models, retrieve them to a desktop computer or perform analyses in a distributed computing environment that includes grid, cloud, or high performance computing model instances as necessary. Understanding will be advanced through the ability to integrate information from multiple sources. Products (data, results, models) can then be published as new resources that can be shared with collaborators.

HydroShare uses the web to deliver its functionality in line with software-as-a service and data-as-a service trends in cyberinfrastructure. Sharing and publication of data using the CUAHSI HIS was hard because of the requirement for the publisher to establish or gain access to a HydroServer and for the recipient to use the HydroDesktop client software, which required software installation and had platform dependencies. By being web based HydroShare strives to overcome these limitations.

HydroShare development is being driven by a collaborative data analysis and publication use case that extends the existing CUAHSI HIS data sharing functionality into a dynamic collaborative environment leading to the eventual archival publication of data (Tarboton et al., 2014b) (Figure 2). At (1) data are observed and then loaded (2). In CUAHSI HIS data is loaded into an ODM relational database on a HydroServer that publishes it using web services. Metadata is harvested by the HIS Central catalog, and supports geographic and context based data discovery. A HydroDesktop client user (3) discovers, downloads and analyzes the data, or uses it in a model. Steps 1 to 3 are supported by the existing CUAHSI HIS. HydroShare picks up from here allowing the user to next post the results (data and model) to HydroShare as resources, retaining provenance information on the original data source (4). This will be done through sharing features being added to HydroDesktop. HydroShare will also support direct entry of new resources. Upon ingestion, background actions parse metadata and enable analysis based on rules and policies. The user shares posted resources with colleagues (5), designating who has permission to access the resources. A group collaborates on refining the analysis, model or result. HydroShare tracks provenance supporting reproducibility and transparency. After iteration, the result is finalized and submitted for publication (6). At this point the resources produced (data, model, workflow, paper) are made immutable, access is opened and permanent persistent identifiers (e.g., DOIs) are assigned. The data may be moved to a permanent repository under the auspices of the CUAHSI Water Data Center (7).

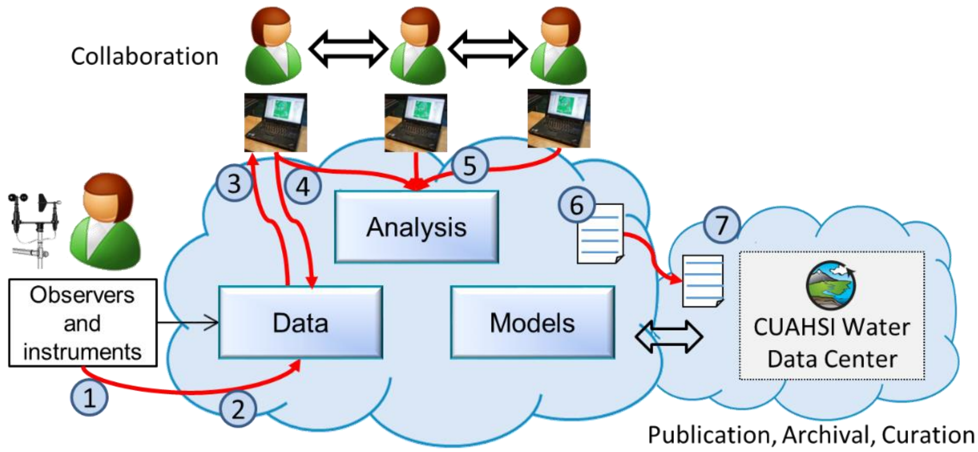

Figure 2. HydroShare collaborative data analysis use case

In HydroShare data is interpreted in a broad sense to mean any product or resource that is an object to be shared in HydroShare. This includes data in a number of formats (time series, geographic features, geographic grids, space time arrays, model programs, model instances, scripts and collections of resources). A structured approach is used for the representation of each resource type using a formal resource data model (Tarboton et al., 2014a). This approach discriminates HydroShare from some other resource sharing systems where objects with any described format are accepted. The structure of HydroShare resources enables their use in analysis and modeling functions so that HydroShare can include a collection of best of practice analysis tools that operate on these resources. These best of practice tools are the incentive to attract users to the system. They provide hydro-value-added functionality over the functionality available in general purpose digital object sharing systems.

The HydroShare architecture is comprised of a website and REST service interface to expose as much functionality as possible via services accessible by clients such as HydroDesktop. The HydroShare web interface and social media functions are being developed using the Django web application framework (Heard et al., 2014). The integrated Rule-Oriented Data System (iRODS) is being used to manage federated data content and perform rule-based background actions on data and model resources, including parsing to generate metadata catalog information and the execution of models and workflows.

HydroShare thus provides a new, web-based system for advancing model and data sharing with sharing features added to HydroDesktop. It provides access to a wider range of types of hydrologic data using standards compliant data formats and interfaces, including models. The standard representation of models facilitates model sharing and discovery functionality as well as transmittal and execution of models in high performance computing systems.

References

Heard, J., D. Tarboton, R. Idaszak, J. Horsburgh, D. Ames, A. Bedig, A. M. Castronova, A. Couch, P. Dash, C. Frisby, T. Gan, J. Goodall, S. Jackson, S. Livingston, D. Maidment, N. Martin, B. Miles, S. Mills, J. Sadler, D. Valentine and L. Zhao, (2014), "An Architectural Overview of Hydroshare, A Next-

Generation Hydrologic Information System," 11th International Conference on Hydroinformatics, HIC 2014, New York City, USA, http://hic2014.org/.

Tarboton, D. G., J. S. Horsburgh, R. Idaszak, J. Heard, D. Valentine, A. Couch, D. Ames, J. L. Goodall, L. Band, V. Merwade, J. Arrigo, R. Hooper and D. Maidment, (2014a), "A resource centric approach for advancing collaboration through hydrologic data and model sharing " 11th International Conference on Hydroinformatics HIC 2014, New York City, USA, http://hic2014.org/.

Tarboton, D. G., R. Idaszak, J. S. Horsburgh, J. Heard, D. Ames, J. L. Goodall, L. Band, V. Merwade, A. Couch, J. Arrigo, R. Hooper, D. Valentine and D. Maidment, (2014b), "HydroShare: Advancing Collaboration through Hydrologic Data and Model Sharing," in D. P. Ames, N. W. T. Quinn and A. E. Rizzoli (eds), Proceedings of the 7th International Congress on Environmental Modelling and Software, San Diego, California, USA, International Environmental Modelling and Software Society (iEMSs), ISBN: 978-88-9035-744-2, http://www.iemss.org/society/index.php/iemss-2014-proceedings.

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

Zaslavsky, I., D. Valentine and T. Whiteaker, (2007), "CUAHSI WaterML," OGC 07-041r1, Open Geospatial Consortium Discussion Paper, http://portal.opengeospatial.org/files/?artifact_id=21743.