Hydrologic Information Systems
Synopsis of Class 20, GIS in Water Resources, Fall 2013

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

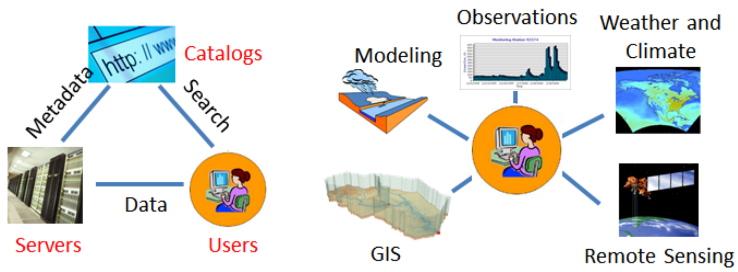
After today's class you should be able to

- 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.

Two concepts, (1) the services oriented architecture; and (2) the desktop hydrologic information system underlie the architecture of the CUAHSI HIS (Figure 1).



a) Services-Oriented Architecture b) Desktop Hydrologic Information System

Figure 1. Hydrologic Information System Overarching Vision

The HIS services-oriented architecture can be viewed as: 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.

The concept of HIS desktop application software is somewhat analogous to Geographic Information System (GIS) desktop software that supports storage and analysis of logically linked data (Tomlinson, 2003). Our 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 HIS services-oriented architecture is comprised of three classes of functionality: 1) data publication (HydroServer), 2) data cataloging (HydroCatalog), and 3) data discovery, access and analysis (HydroDesktop) (Figure 2). 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.

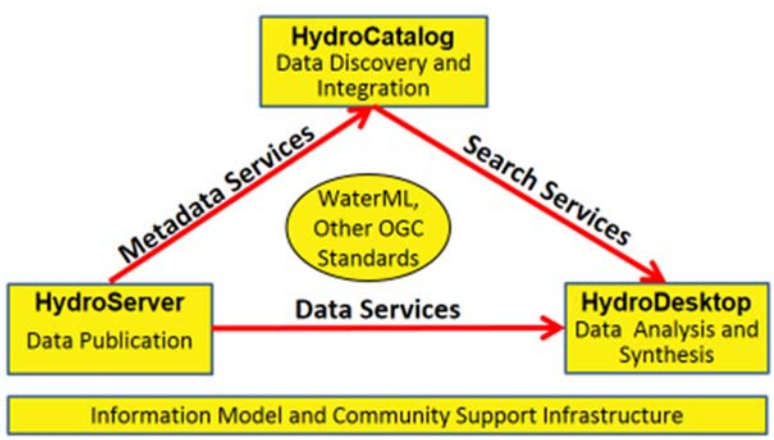


Figure 2. Components of CUAHSI HIS Services Oriented Architecture.

The components shown in Figure 2 either publish or consume information via the following categories of web services:

- Data Services – which convey the actual data.
- Metadata Services – which convey metadata about specific collections or series of data.
- Search Services – which enable search, discovery, and selection of data and convey metadata required for accessing data using data services.

The formats for transmission of information between these systems and the interfaces that enable the communication between them (the connecting arrows in Fig. 2) 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.

At the base of Figure 2 is depicted the information model and community support infrastructure upon which the system is founded. The information model is the conceptual model used to organize and define sufficient metadata about hydrologic observations for them to be unambiguously interpreted and used. Within HydroServer, it is encoded using the Observations Data Model (ODM) (Horsburgh et al., 2008) relational database and the HydroServer Capabilities Database to ensure that data and metadata are stored together. The information model also serves as the conceptual basis for WaterML to ensure that data and associated metadata are transmitted with fidelity when data are downloaded. HydroDesktop implements the information model within its data repository database ensuring that local copies of data retrieved from a server maintain their original context. ODM includes a number of controlled vocabularies for metadata such as units, variable names, sample media etc., where semantic consistency in describing observations is important.

The architecture shown in Figure 2 has evolved as an approach for sharing hydrologic observations data that is general and open to allow broad participation. By relying on standards to define the interactions between components, different organizations can develop components and have them be interoperable. This approach based on standards provides a general foundation and approach for integration and sharing of hydrologic data around the world. This idea extends beyond hydrology and Dangermond and Maidment (2010) have advocated geo-enabling the CUAHSI model as an approach to developing a new web-GIS that will enable the easy integration of large volumes of water data and complex models into simple to use applications that become pervasive. Jack Dangermond is president of ESRI, the company that develops ArcGIS, and the fact that ESRI is advancing this capability makes it accessible to a much broader class of users. The CUAHSI HIS cyberinfrastructure has been developed through federally funded research. One of the challenges with cyberinfrastructure is the business model for its long term sustainability. The fact that ESRI is adopting parts of HIS into its business model for web based geo-services provides one avenue for sustaining and advancing this capability into the future.

HydroShare is a new project supported by the National Science Foundation (NSF) for the development of an online, collaborative environment for sharing hydrologic data and models. The goal of HydroShare 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. Scientists may also publish outcomes (data, results or models) into HydroShare, using the system as a collaboration platform for sharing data, models and analyses. HydroShare is expanding the data sharing capability of the CUAHSI Hydrologic Information System by broadening the classes of data accommodated, creating new capability to share models and model components, and taking advantage of emerging social media functionality to enhance information about and collaboration around hydrologic data and models. One of the fundamental concepts in HydroShare is that of a Resource. All content is represented using a Resource Data Model that separates system and science metadata and has elements common to all resources as well as elements specific to the types of resources HydroShare will support. These will include different data types used in the hydrology community and models and workflows that require metadata on execution functionality. The HydroShare web interface and social media functions are being developed using the Drupal content management system. A geospatial visualization and analysis component enables searching, visualizing, and analyzing geographic datasets. 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. This lecture will describe some of the functionality that is planned and being developed as part of HydroShare and will imagine and reflect on the possibilities that this capability can bring as it becomes part of cyberinfrastructure in the future.

References

Dangermond, J. and D. Maidment, (2010), "Integrating Water Resources Information Using GIS and the web," 2010 AWRA Spring Specialty Conference Geographic Information Systems (GIS) and Water Resources VI, Orlando Florida, American Water Resources Association, Middleburg, Virginia, TPS-10-1, http://his.cuahsi.org/documents/conference-awra2010/Dangermond_Maidment_awrakeynote.pdf.

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, http://dx.doi.org/10.1029/2007WR006392.

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.