Geophysical Research Abstracts,
Vol. 11, EGU2009-6800-2, 2009
EGU General Assembly 2009

A Services-Oriented Architecture for Water Observations Data

D.R. Maidment (1), I. Zaslavsky (2), D. Valentine (2), D.G. Tarboton (3), T. Whitenack (2), T. Whiteaker (1), R. Hooper (4), and D. Kirschtel (4)

(1) Center for Research in Water Resources, University of Texas, Austin, (2) San Diego Supercomputer Center, University of California, San Diego, (3) Utah Water Research Laboratory, Utah State University, Logan, (4) Consortium of Universities for the Advancement of Hydrologic Science, Inc, Washington DC

Water observations data are time series of measurements made at point locations of water level, flow, and quality and corresponding data for climatic observations at point locations such as gaged precipitation and weather variables. A services-oriented architecture has been built for such information for the United States that has three components: hydrologic information servers, hydrologic information clients, and a centralized metadata cataloging system. These are connected using web services for observations data and metadata defined by an XML-based language called WaterML. A Hydrologic Information Server can be built by storing observations data in a relational database schema in the CUAHSI Observations Data Model, in which case, web services access to the data and metadata is automatically provided by query functions for WaterML that are wrapped around the relational database within a web server. A Hydrologic Information Server can also be constructed by custom-programming an interface to an existing water agency web site so that responds to the same queries by producing data in WaterML as do the CUAHSI Observations Data Model based servers. A Hydrologic Information Client is one which can interpret and ingest WaterML metadata and data. We have two client applications for Excel and ArcGIS and have shown how WaterML web services can be ingested into programming environments such as Matlab and Visual Basic. HIS Central, maintained at the San Diego Supercomputer Center is a repository of observational metadata for WaterML web services which presently indexes 342 million data measured at 1.75 million locations. This is the largest catalog water observational data for the United States presently in existence.

As more observation networks join what we term "CUAHSI Water Data Federation", and the system accommodates a growing number of sites, measured parameters, applications, and users, rapid and reliable access to large heterogeneous hydrologic data repositories becomes critical. The CUAHSI HIS solution to the scalability and heterogeneity challenges has several components. Structural differences across the data repositories are addressed by building a standard services foundation for the exchange of hydrologic data, as derived from a common information model for observational data measured at stationary points and its implementation as a relational schema (ODM) and an XML schema (WaterML). Semantic heterogeneity is managed by mapping water quantity, water quality, and other parameters collected by government agencies and academic projects to a common ontology. The WaterML-compliant web services are indexed in a community services registry called HIS Central (hiscentral.cuahsi.org). Once a web service is registered in HIS Central, its metadata (site and variable characteristics, period of record for each variable at each site, etc.) is harvested and appended to the central catalog. The catalog is further updated as the service publisher associates the variables in the published service with ontology concepts. After this, the newly published service becomes available for spatial and semantics-based queries from online and desktop client applications developed by the project. Hydrologic system server software is now deployed at more than a dozen locations in the United States and Australia.

To provide rapid access to data summaries, in particular for several nation-wide data repositories including EPA STORET, USGS NWIS, and USDA SNOTEL, we convert the observation data catalogs and databases with harvested data values into special representations that support high-performance analysis and visualization. The construction of OLAP (Online Analytical Processing) cubes, often called data cubes, is an approach to organizing and querying large multi-dimensional data collections. We have applied the OLAP techniques, as implemented in Microsoft SQL Server 2005/2008, to the analysis of the catalogs from several agencies. OLAP

analysis results reflect geography and history of observation data availability from USGS NWIS, EPA STORET, and USDA SNOTEL repositories, and spatial and temporal dynamics of the available measurements for several key nutrient-related parameters.

Our experience developing the CUAHSI HIS cyberinfrastructure demonstrated that efficient integration of hydrologic observations from multiple government and academic sources requires a range of technical approaches focused on managing different components of data heterogeneity and system scalability. While this submission addresses technical aspects of developing a national-scale information system for hydrologic observations, the challenges of explicating shared semantics of hydrologic observations and building a community of HIS users and developers remain critical in constructing a nation-wide federation of water data services.