Geophysical Research Abstracts,
Vol. 11, EGU2009-6824-3, 2009
EGU General Assembly 2009

The evolution of the CUAHSI Water Markup Language (WaterML)

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

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

The CUAHSI Hydrologic Information System (HIS, his.cuahsi.org) uses web services as the core data exchange mechanism which provides programmatic connection between many heterogeneous sources of hydrologic data and a variety of online and desktop client applications. The service message schema follows the CUAHSI Water Markup Language (WaterML) 1.x specification (see OGC Discussion Paper 07-041r1). Data sources that can be queried via WaterML-compliant water data services include national and international repositories such as USGS NWIS (National Water Information System), USEPA STORET (Storage & Retrieval), USDA SNOTEL (Snowpack Telemetry), NCDC ISH and ISD(Integrated Surface Hourly and Daily Data), MODIS (Moderate Resolution Imaging Spectroradiometer), and DAYMET (Daily Surface Weather Data and Climatological Summaries). Besides government data sources, CUAHSI HIS provides access to a growing number of academic hydrologic observation networks. These networks are registered by researchers associated with 11 hydrologic observatory testbeds around the US, and other research, government and commercial groups wishing to join the emerging CUAHSI Water Data Federation. The Hydrologic Information Server (HIS Server) software stack deployed at NSF-supported hydrologic observatory sites and other universities around the country, supports a hydrologic data publication workflow which includes the following steps: (1) observational data are loaded from static files or streamed from sensors into a local instance of an Observations Data Model (ODM) database; (2) a generic web service template is configured for the new ODM instance to expose the data as a WaterML-compliant water data service, and (3) the new water data service is registered at the HISCentral registry (hiscentral.cuahsi.org), its metadata are harvested and semantically tagged using concepts from a hydrologic ontology. As a result, the new service is indexed in the CUAHSI central metadata catalog, and becomes available for spatial and semantics-based queries.

The main component of interoperability across hydrologic data repositories in CUAHSI HIS is mapping different repository schemas and semantics to a shared community information model for observations made at stationary points. This information model has been implemented as both a relational schema (ODM) and an XML schema (WaterML). Its main design drivers have been data storage and data interchange needs of hydrology researchers, a series of community reviews of the ODM, and the practices of hydrologic data modeling and presentation adopted by federal agencies as observed in agency online data access applications, such as NWISWeb and USEPA STORET.

The goal of the first version of WaterML was to encode the semantics of hydrologic observations discovery and retrieval and implement water data services in a way that is generic across different data providers. In particular, this implied maintaining a single common representation for the key constructs returned to web service calls, related to observations, features of interest, observation procedures, observation series, etc. Another WaterML design consideration was to create (in version 1 of CUAHSI HIS in particular) a fairly rigid, compact, and simple XML schema which was easy to generate and parse, thus creating the least barrier for adoption by hydrologists. Each of the three main request methods in the water data web services - GetSiteInfo, GetVariableInfo, and GetValues – has a corresponding response element in WaterML: SiteResponse, VariableResponse, and TimeSeriesResponse. The strictness and compactness of the first version of WaterML supported its community adoption. Over the last two years, several ODM and WaterML implementations for various platforms have emerged, and several Water Data Services client applications have been created by outside groups in both industry

and academia. In a significant development, the WaterML specification has been adopted by federal agencies. The experimental USGS NWIS Daily Values web service returns WaterML-compliant TimeSeriesResponse. NCDC is also prototyping WaterML for data delivery, and has developed a REST-based service that generates WaterML-compliant output for its integrated station network. These agency-supported web services provide a much more efficient way to deliver agency data compared to the web site scraper services that the CUAHSI HIS project developed initially. Adoption of WaterML by the US Geological Survey is particularly significant because the USGS maintains by far the largest water data repository in the United States.

For version 1.1, WaterML has evolved to reflect the deployment experience at hydrologic observatory testbeds, as well as feedback from hydrologic data repository managers at federal and state agencies. Further development of WaterML and enhancement of the underlying information model is the focus of the recently established OGC Hydrology Domain Working Group, whose mission is to profile OGC standards (GML, O&M, SOS, WCS, WFS) for the water resources domain and thus ensure WaterML

ś wider applicability and easier implementation. WaterML 2.0 is envisioned as an OGC-compliant application schema that supports OGC features, can express different types of observations and various groupings of observations, and allows researchers to define custom metadata elements.

This presentation will discuss the information model underlying WaterML and describe the rationale, design drivers and evolution of WaterML and the water data services, illustrating their recent application in the context of CUAHSI HIS and the hydrologic observatory testbeds.