American Water Resources Association
2010 SUMMER SPECIALTY CONFERENCE
GIS & Water Resources VI
March 29 – 31, 2010
Orlando, FL

Tuesday, March 30
3:30 PM – 5:00 PM
SESSION 21: Hydrologic Information Systems III

HydroPortal and HydroViewer: Registration and Discovery of CUAHSI Web Services - Dean Djokic, ESRI, Redlands, CA (co-authors: Zhumei Qian, Christine Eggers, Zichuan Ye, Jignesh Divecha)

The Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) has developed web services specifications (WaterOneFlow) and infrastructure for query and dissemination of hydrologic (and other) time series data. This paper describes two applications that build on top of CUAHSI services and provide service registration and data discovery capabilities. HydroPortal is a customization of ESRI ArcGIS Server Geoportal extension technology that allows registration of WaterOneFlow services and their metadata in one of many national and international metadata standard formats. HydroPortal allows registration of specific WaterOneFlow parameters in addition to the standard metadata, thus allowing the full metadata functionality of the geoportal to be applied to hydro data. Once registered in the geoportal, the data can be discovered through the default geoportal web application interface or through standard web technologies such as REST or Catalog Service for the Web (CS-W). One HydroPortal can automatically discover services registered in other HydroPortal and Geoportal extension installations. HydroViewer is a web-based map application for discovering data that are published using WaterOneFlow services and registered through one or many HydroPortals. It is based on ESRI ArcGIS Server technology. Using a map interface, users can discover published WaterOneFlow services, query spatial, variable, and temporal space of each service individually or in combination, and combine information from different services into a single "DataCart". The DataCart is an XML structure that defines which service, station, variable, and time periods are of interest to the user and the actual service URLs that serve these data. Once the DataCart is defined, it can be stored for future use on the server, downloaded in XML format, sent to another application that can consume a DataCart, or the actual spatial and temporal data defined by the DataCart can be assembled into an Arc Hydro-compliant geodatabase and delivered to the user. The assimilation of disparate spatial and temporal data into a well established GIS format allows immediate mobilization of downloaded data using COTS GIS technology.

OLAP-Based Analysis, Visualization and Animation of Large Volumes of Hydrologic Observation and Model Data - Matthew Rodriguez, San Diego Supercomputer Center, UCSD, La Jolla, CA (co-authors: Thomas Whitenack, David Valentine, Ilya Zaslavsky)

One of the goals of the CUAHSI Hydrologic Information System project (his.cuahsi.org) is to create a comprehensive portrait of hydrologic observations for the U.S., integrating observational data and metadata from multiple sources, at the national, regional, and local levels. The data are made available via a uniform set of web service interfaces, called CUAHSI Water Data Services (or WaterOneFlow services). Once a source of hydrologic observations is exposed via such set of methods, it is registered in the HISCentral registry (hiscentral.cuahsi.org) and its metadata is harvested into the central metadata catalog. The catalog currently indexes over 8 million time series. In addition, we work with large databases of observations and model-generated data, which may be time-consuming to query using standards relational database tools. We 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 observations data catalogs and the databases with the harvested values, into special representations that support high performance analysis and visualization. The generation of OLAP (Online Analytical Processing) databases, often called data cubes, is an approach to organizing and querying large multi-dimensional data collections. We apply OLAP techniques, using Microsoft SQL Server 2005/08, to the analysis of the catalogs from several agencies. This presentation describes challenges of generating data cube dimensions from hydrologic data catalogs and observations data, and demonstrates an online application designed to map and animate OLAP query results. We also present analytical results primarily related to hydrologic data availability from the observations data catalogs. The results reflect geographic and temporal analysis of available data totals from USGS NWIS, EPA STORET, and USDA

SNOTEL repositories, and spatial and temporal dynamics of available measurements for several key nutrient-related parameters. In this capacity, OLAP data cubes can become a component of HIS Central data discovery system, by providing rapid data aggregations over the HISCentral metadata repository. In addition, we demonstrate how cube dimension can be generated and tuned to answer specific research queries. This approach can be also used for outlier detection in large databases and repositories of hydrologic observations.

Hydrologic Metadata Catalog and Semantic Search Services in CUAHSI HIS - Thomas Whitenack, San Diego Supercomputer Center, UCSD, La Jolla, CA (co-authors: David Valentine, Ilya Zaslavsky, Michael Piasecki, David Tarboton, Jeffery Horsburgh, Timothy Whiteaker, Daniel Ames, David R. Maidment)

The CUAHSI Hydrologic Information System (HIS) relies on a collection of water data services (WDS) that follow a common Water Markup Language (WaterML) specification to exchange hydrologic data and metadata. The services provide uniform access to both government and academic hydrologic data repositories. Federal sources of hydrologic data available within HIS include USGS National Water Information System, EPA STORET, NCDC, USDA, and the Army Corps of Engineers. There are also some state and other data sources. For users wishing to share data on the system a data publication workflow and HIS Server software stack allows anyone to publish their hydrologic observations data. A registry of these services has been created (http://hiscentral.cuahsi.org) and information about each service is harvested into the HIS Central Metadata catalog. The metadata catalog maintains information about observation series available from registered water data services. This includes information about observation sites, measured parameters, periods of record, measurement methods, and other metadata as outlined in the CUAHSI Observations Data Model (ODM). Additional tables store information about data services (the Networks table) and a concept hierarchy used to support keyword based data discovery in the form of an ontology. Maintaining ontology information in the HIS Central metadata catalog is critical due to semantic heterogeneity across different observation networks. Variables measured within each network are associated with a common ontological term. HIS Central provides services for searching by keyword across a federated system of water observation networks. The central service registry, the metadata catalog, and the ontology underlie HISCentral web services that allow both web-based and desktop client applications such as the HydroViewer and HydroDesktop to query hydrologic metadata from multiple observation networks simultaneously. The services allow users to filter their searches by using the ontology keywords, period of record, and the geographical location. Once a client application retrieves this metadata, it can then obtain the data values by connecting directly to the registered service. This paper describes the design of HIS Central registry and metadata catalog, and the HIS Central services that make these data available to distributed applications.

Transition to OGC Standards in CUAHSI Hydrologic Information System - David Valentine, San Diego Supercomputer Center, UCSD, La Jolla, CA (co-authors: Ilya Zaslavsky, David R. Maidment)

The CUAHSI Hydrologic Information System (HIS) is a distributed network of hydrologic data repositories and applications that are integrated using web services. Water data services, the core of the HIS service-oriented architecture, provide uniform access to many sources of observation data, which include hydrologic and related time series from the US Geological Survey (USGS), the Environmental Protection Agency (EPA), the National Climatic Data Center (NCDC), Army Corps of Engineers, US Department of Agriculture, and other federal and state agencies. The HIS also provides a workflow for publishing hydrologic observations collected by academic projects. The services follow Water Markup Language, an XML schema for exchanging hydrologic time series information. It is designed to capture semantics of hydrologic observations discovery and retrieval and express the point observations information model as an XML schema. WaterML 1.x follows the representation of the information model as adopted by the CUASHI Observations Data Model (ODM), as well as specifications and metadata adopted by the federal agency partners of CUAHSI HIS. WaterML and the associated WaterOneflow services have been used successfully for several years, providing a common foundation for exchanging both agency data and data collected in multiple academic projects. In addition, the WaterML specification has been adopted by USGS (the Daily Values and Instantaneous data services from the National Water Information System) and NCDC (the Integrated Surface Daily and Hourly data) in their prototype water data service development. The next step in services development is their harmonization with standards being adopted or developed in neighboring communities, in particular the relevant specifications being explored within the Open Geospatial Consortium (OGC). Development of WaterML 2.0 is one of the foci of the recently established OGC and WMO (the World Meteorological Organization) Hydrology Domain Working Group. This working group has as its mission profiling and tuning OGC standards (GML, O&M, SOS, WCS, WFS) for the water resource domain, thus ensuring WaterML's wider applicability and easier interpretation. In this paper, we review the

ongoing transition of HIS WaterML-based services to a new model that relies on WaterML 2.0 and other OGC specifications.

Interdisciplinary Environmental Time Series Data Management - Stefan Fuest, KISTERS North America, Inc., Citrus Heights, CA (co-authors: Stan Malinky, Michael Natschke)

Environmental Information Systems are often designed for particular data themes. Specific solutions are commercially available, that deliver workflows and analysis tools for groundwater data, meteorological data, dam operation or hydrometric data. Individual proprietary solutions are splitting data into different locations, which makes the process complicated to cross query and analyse interdisciplinary data. The KISTERS TSM (time series management server) is designed to manage environmental data in general. It integrates interdisciplinary time series data as well as descriptive data into one controlled open information system. The foundation is a flexible server infrastructure that has been derived from experiences collected through highly demanding projects over more than 20 years in a global and mature market. KISTERS TSM provides an optimised storage management designed for sparse and dense time series data serving small to large data collection networks. Data from specific disciplines such as hydrometric, meteorological, hydro geological data as well as dam safety can be integrated into one central information system. Automatic data collection processes (telemetry, SCADA, IP based loggers) can directly import into the TSM storage. The quality assurance process is a combination of manual, semi automatic or fully automated validation. The powerful and high sophisticated graphing tool brings all information together and provides optimised tools for interactive data editing. A series of generic and disciplinary calculation methods are available for automatic calculation of derived time series data. These include aggregation to hourly, daily, weekly, monthly and/or annual statistics but also long term data, trends as well as peaks over threshold detection or mean tide calculations. All persistent time series are updated automatically in the background to provide at any point in the time the best data available. Individual reports can be designed to be presented in Business Objects, Microsoft Excel, Microsoft Word or any ASCII output (plain ASCII, XML, HTML). Public accessible data can be disseminated through the KISTERS DataProvider directly to Intra-/Internet Web Solutions but also to spatial software solutions such as provided by ESRI. Open standards for data formats and services (e.g. OGC-WMS/WFS/SOS or CUAHSI-WaterOnflow) are supported as well.