American Water Resources Association
2015 ANNUAL WATER RESOURCES CONFERENCE

November 16-19, 2015

Denver, CO

Wednesday, Nov. 18
10:30 AM – 12:00 Noon
SESSION 49: Open Water Data 2

Introducing the WaterML R Package for Retrieving Data from CUAHSI WaterOneFlow Web Services - Jiri Kadlec, Brigham Young University, Provo, UT (co-authors: B. St Clair, D. P. Ames, R.Gill)

An increasing number of hydrological observations from different institutions and research teams in the world is becoming published on the internet in the WaterML data exchange format approved by the Open GIS Consortium (OGC). The WaterML format includes not only the raw data, but also important metadata about the site, variable, units, methods, and quality control levels of the time series observations. The R computational environment is used by many hydrologists for statistical data analysis. We present the design and development of a new WaterML R package that provides access to the Consortium of Universities for Advancement of Hydrologic Science (CUAHSI) Hydrologic Information System (HIS) HydroServers as a means for storing and managing data. Using the WaterML R package, the user can retrieve and analyze data from HydroServers of multiple organizations that are listed in the CUAHSI Water Data Center catalog and the Global Earth Observation System of Systems data catalog of the World Meteorological Organization. The new WaterML R package is presented in terms of its functional requirements and design, with the express goal of providing support for four core web methods defined by the HydroServer WaterOneFlow web services specification. The WaterML R package supports reading data from all versions of WaterML (1.0, 1.1 and 2.0) and accessing both versions of the WaterOneFlow web services (SOAP and REST). The WaterML R package is open-source software and it has been published on the Comprehensive R Archive Network (CRAN) on http://cran.r-project.org/web/packages/WaterML . Any R user can install it using the "install packages" option in R. We have tested the WaterML R package in the context of data collected as part of a large ecological manipulation experiment. The resulting system allows research scientists to use a familiar statistical computation environment, R, together with the open source HydroServer software (for data archival and sharing). We also developed a new HydroServer data upload web service to facilitate data upload to a PHP version of HydroServer called HydroServer Lite directly from the WaterML R package presented here.

Automated Approaches for Georeferencing Data to the NHDPlus and Implications for the Open Water Data Initiative - Dwane Young, U.S. EPA, Washington, DC (co-authors: W. Reid, D. Boezio, B. Bergenroth)

EPA has long seen the value of referencing geospatial data to a common geofabric. EPA has invested significantly in associating various program data (i.e. impaired waters, assessed waters, permitted facilities, monitoring locations, etc.) as events with the National Hydrography Dataset Plus (NHDPlus). This has allowed for numerous cross-program analyses, including upstream and downstream analyses and watershed characterization. EPA has also deployed a number of services off this dataset to provide access to this wealth of information via open services that have seen significant use. Although this dataset provides tremendous value, it does require a significant investment to update and maintain it. To help reduce the cost of updating this dataset, EPA pursued a new automated approach for

georeferencing the data to the NHDPlus. This new approach uses the NHDPlus catchments as the units to which the data are georeferenced, and then those catchments can be used to make cross-program associations, as a unit of measure for demonstrating progress under EPA's Strategic Plan, and for other analyses. EPA developed this approach through a Georeferencing Pilot. This pilot was very successful, and clearly demonstrated the cost saving and time saving potential. What used to take weeks to months to complete, now took minutes to hours. Following the success of the pilot, EPA pursued the development of this approach into a production level process. One of the goals that EPA set for this process is that it would be developed using an open-source/cross- platform approach so that anybody could have access to the code and software and run the process without any expensive third-party software. EPA expects to complete this new Catchment Indexing Process (or CIP) in early Fall 2015. The CIP tool is being developed in PostgreSQL and PostGIS (an open source database and geospatial data management system) with a Python interface. Processing times using this new tool have proven to be very promising. For example, the new tool is capable of processing an entire state's impaired waters layer (stream segments) in less than 10 minutes, depending upon the number of impaired waters. Although some follow-up QA is required, the amount of time to develop these datasets is significantly reduced. This tool has the potential to be valuable to the implementation of the OWDI. In order for data to be discoverable, they must first be associated with a common geofabric. This tool provides a simple and fast approach that accomplishes that association. Although this approach may not work for all use cases, for many uses it is a tremendously valuable approach. As part of this presentation, EPA will demonstrate the tool as well as how EPA will use these data to demonstrate progress under EPA's strategic plan.

Sensor Data Sharing Approaches in Support of an OWDI Future - Tad Slawecki, LimnoTech, Ann Arbor, MI (co-authors: D. Young, P. McClellan, B. Perez)

Sensor Data Sharing Approaches in Support of an OWDI Future Water quality sensors are playing an ever-larger role in water quality monitoring efforts. These sensors are also beginning to have the capability to monitor new parameters that were considered impossible to monitor via a sensor only a few years ago. For example, EPA has issued a Nutrient Sensor Challenge (http://nutrients-challenge.org) which has the goal of incentivizing the marketplace to develop lower cost nutrient sensors. With all of this excitement about sensors, the question arises: "What to do with the data?" In concert with the Open Water Data Initiative (OWDI), and as a component of EPA's E-Enterprise Initiative, EPA has undertaken a series of actions to identify solutions for making sensor data easily discoverable. The first step that EPA is taking is to work with the National Water Quality Monitoring Council to develop a data strategy. This strategy (expected to be available in September 2015) will provide a roadmap for how to make use of existing data standards, such as those developed by the Open Geospatial Consortium (OGC) for sharing sensor data (http://www.opengeospatial.org/standards/sos and http://www.opengeospatial.org/standards/waterml) for sharing both water quality and water quantity data collected by sensors from the diverse water monitoring community (Federal, state, tribal, local, etc.). These standards are being successfully implemented by the USGS and NOAA, and EPA is now in a unique position to help expand that data sharing capability to the rest of the water monitoring community and to encourage innovation in the private sector to make the adoption of these standards for that community as simple as possible. EPA has also initiated two demonstration projects to evaluate the approaches that are recommended in the data strategy. These demonstration projects will occur in two watersheds, and will start in the Fall of 2015. These projects will provide an opportunity to test the principles of the OWDI when it comes to sensor data, and demonstrate the value of following an OWDI approach. The OWDI provides a vision for how water data can be made both discoverable and interoperable. The framework envisioned under OWDI meshes nicely with developing an interoperable

sensor network. In order for a sensor network to be successful, the principles of the OWDI would need to be followed, including: 1) having a catalog of available sensor data sets, which would include the metadata about those sensors and the parameters that they measure; 2) providing standard services that supply data in a common format using common ontologies; 3) tying sensor locations to a common geofabric enabling the discovery of sensors through upstream, downstream, or other network analyses; and 4) enabling an open marketplace where the private sector, including academia, can discover, analyze, and communicate the data in innovative ways. In this presentation, we will present the recommended approach from the data strategy, and the work to date in the demonstration watersheds. We will also discuss some of the challenges identified in trying to make this vision a reality.

Using Open Data and Software for Water Resources to Drive Transparency and Innovation - Steve Malers, Open Water Foundation, Fort Collins, CO

Water is perhaps our most precious natural resource. Good stewardship and optimization of this resource for multiple uses through monitoring, protection, and management is an obligation of society and a requirement for sustainable economic use of the resource. Promoting the development and use of open data and software for water resources extends the interest in water as a public resource to data and tools about that public resource. Open platforms can provide a framework for education, collaboration, and government programs that lead to transparency and innovation. Transparent access to data ensures that entities focus on data and science- driven policies and decisions. Access to data also allows researchers and industry to identify problems and consequently focus on solutions. This paper examines recent progress in the use of open data and software for water resources and challenges to progress, such as concerns about privacy and fear that data might be misinterpreted or misused.