

## Abstracts

Tuesday, April 29

### Session C2: National Water Quality Portal: Lessons Learned

8:00 – 9:30 am | Room 262

---

#### ***The Water Quality Exchange: A Streamlined Way for Sharing Water Quality Data***

**Charles Kovatch, Kevin Christian, Michael Brennan and Dwane Young**

*US Environmental Protection Agency, Washington, D.C.*

##### **Abstract**

The Water Quality Exchange (WQX) is a framework that makes it easier for States, Tribes, and others to submit and share water quality monitoring data over the Internet. States, Tribes and other organizations can now submit data directly to the publicly-accessible STORET Data Warehouse, making the data also available via the Water Quality Data Portal, using the WQX framework. WQX is designed to allow for the automated sharing of water quality data using a standard format (the WQX Extensible Markup Language (XML) schema) and using agreed upon exchange protocols (the Exchange Network). This standard approach for sharing water quality monitoring data was developed through a coordinated effort with the National Water Quality Monitoring Council (NWQMC), USGS, the States, several Tribes, and a number of volunteer organizations. The WQX Schema has also enabled the development of the Water Quality Data Portal which allows for the integration of USGS monitoring data and data provided to EPA to be made available via a common portal.

During this presentation, EPA will give an introduction to WQX, including a discussion on the types of data that can be shared and the mechanisms by which those data can be shared including an introduction to WQX Web which is an online tool that enables for the easy conversion of a data to the WQX Schema. EPA will also discuss the importance of sharing water quality data, and will discuss some EPA initiatives to increase the amount of data sharing by state, tribal, and volunteer partners.

#### ***Two Years of the Water Quality Portal: Improvements, Lessons, and Plans for the Future***

**James Kreft**

*US Geological Survey, Middleton, Wis.*

##### **Abstract**

The Water Quality Portal (WQP) was launched in April of 2012 as a single point of access for discrete water quality samples stored in the US Geological Survey (USGS) NWIS system and the US Environmental Protection Agency (EPA) STORET system. Since then, thousands of users have visited the Water Quality Portal and used it to download hundreds of millions of results that are pertinent to their interests, and numerous tools have been developed that use WQP web services as a source of data for further analysis. Since the launch of the Portal, the WQP development team at the USGS Center for Data Analytics has worked with USGS and EPA stakeholders as well as a wider user group to add significant new features to the WQP. WQP users can now directly plot sites of interest on a web map based on any of the 14 WQP query parameters, and then download data of interest directly from that map. In addition, the WQP has expanded beyond just serving out USGS and EPA data and has begun providing data from the US Department of Agriculture's Agricultural Research Service STEWARDS system, and is working with others to bring in additional data. Finally, the WQP is now linked to another NWQMC-supported project, the National Environmental Methods Index (NEMI), so WQP users can easily find the method behind the data that they are using. Future work is focused on adding additional biological data and biologically relevant query parameters to broaden the scope of discrete water quality sample types that the WQP can provide. The WQP is

also exploring ways to further integrate with other systems to facilitate the overarching goal of improving access to water quality data for all users.

### ***A National Compilation of Water-Quality Monitoring Data to Support Local, Regional, and National Scale Water Quality Assessments***

**Denise Argue<sup>1</sup>, Jeffrey Deacon<sup>1</sup> and Robert Gilliom<sup>2</sup>**

<sup>1</sup>US Geological Survey, Pembroke, N.H., <sup>2</sup>US Geological Survey, Sacramento, Calif.

#### **Abstract**

A national inventory was initiated in 2011 by the U.S. Geological Survey's (USGS) National Water Quality Assessment (NAWQA) program to compile and assess available water-quality monitoring information in order to make full use of existing data to support local, regional, and national assessments of the quality of the Nation's waters. The compilation includes nutrients, pesticides, and a wide range of additional water-quality data collected by a variety of organizations focused on monitoring the conditions of surface water, groundwater, and macroinvertebrate communities. The data compilation brings together numerous partner data sets into a single consistent format, specifically focusing on historic and current ambient monitoring data that are readily accessible, including from federal, state, and regional government agencies; non-governmental organizations; and data publicly housed in the National Water Information System (NWIS) and the Storage and Retrieval (STORET) system. The initial purpose of the data compilation was to support an assessment by the Northeast-Midwest Institute of the existing water information system in the northeast and Midwestern U.S. in order to address priority environmental issues in the region, such as nutrient enrichment and hydrologic fracturing. As the compilation expanded to national scale, the purposes expanded to include: 1) application of compiled monitoring data to support national scale water-quality status assessments, trends assessments, watershed modeling, and ecological studies; and 2) evaluation of current monitoring activities for the purpose of designing efficient monitoring networks and partnerships. For pesticides, in particular, NAWQA and USEPA Office of Pesticides are working together to enhance the completeness of the compilation so that it can best support assessments of pesticide concentrations in the Nation's streams and groundwater.

### ***Duplicate Water Data – Causes, Implications, Solutions***

**Dorinda Gellenbeck<sup>1</sup> and Jonathon Scott<sup>2</sup>**

<sup>1</sup>US Geological Survey, Denver, Colo., <sup>2</sup>US Geological Survey, Oklahoma City, Okla.

#### **Abstract**

The duplication of water data between databases within an agency and among agencies can lead to incorrect data analyses if not properly handled. Such duplication has been widespread for decades to accomplish data aggregation tasks that serve analyses related to specific areal studies. Resources required to obtain water-quality data are high; therefore, many researchers increased the available data by copying relevant data from other sources. Researchers and agency personnel commonly obtain, reformat, and store all the available data for a particular study area in their separate and respective databases.

Data copies residing in multiple databases sometimes are inconsistent, because of reformatting errors or differences, lack of diligence in copying all associated metadata, and updates made to the original source after the copy was performed. Presently large aggregated database collections are accessible via the internet, which can result in saving time previously expended obtaining, reformatting, and combining data sources. However, these large aggregated data collections also have exposed the duplicate data problem. These duplicate records potentially corrupt analyses with bias due to the inclusion of the same data more than once or with errors that propagate from inconsistent copies.

New duplication of data can be avoided now with the advent of the Water- Quality Portal, a joint effort of the National Water Quality Monitoring Council, U.S. Environmental Protection Agency, and the U.S. Geological Survey

(USGS). Recently the USGS has used a heuristic scoring system to remove duplicate copies of water-quality and other types of data in a new centralized data repository. The task was too time-consuming and tedious to solve with database experts, so an automated process was devised to rank duplicate copies to select the most appropriate version. A similar technique could potentially be used to evaluate duplicated data among agency systems. Additional methods could be used to identify co-located data-collection sites, such as a shared site-identifier system, or a site-identifier alias database.