Data Models and Data Sharing Networks: Continuous and Discrete Data

Dwane Young, U.S. EPA Office of Water
Jon Pollak, CUAHSI
4/30/2014
Water Quality Exchange (WQX)

- Allows for the computer-to-computer exchange of water quality data
  - Designed for water quality samples (physical, chemical, microbiological, biological, and habitat)
  - Been available since 2007
- The schema was designed in partnership with the states and tribes
- The data exchange is used by EPA, other federal agencies, states, tribes, citizen groups, and local governments
- More information can be found at: www.epa.gov/storet
The water quality data portal (an EPA/USGS partnership) provides access to over 232 million water quality results. These data are all available via web services which can be incorporated into any other third party application. For more information on the portal see: www.waterqualitydata.us
In the water monitoring world, we can classify monitoring into two types:

**Discrete Monitoring**
- A *sample* is taken and sent to a lab for further analysis.
- Typically a one-time event that can be repeated as needed.

**Continuous Monitoring**
- A *sensor* is used to record a continuous stream of data about 1 particular analyte or a small set of analytes (i.e. flow, dissolved oxygen, pH, etc).
- Values are reported at set intervals (i.e. every 15 minutes, 1 hour, etc.).
Continuous Monitoring Data

- The data model for continuous data is different from sample data
- EPA recognizes that WQX is probably not the model for this type of data
- EPA is beginning the process of looking at other approaches for continuous data (i.e. WaterML 2)
CUAHSI HIS, ODM, and WaterML

- Hydrologic Information System developed through CUAHSI Community

Principle Investigator David Maidment (The University of Texas at Austin) & several other universities

Pilot Development Production

2002 2007 2012

USGS Data Source
Streamflow gages
Neuse River near Clayton, NC
Discharge, stage, start, end (Daily or instantaneous)

206 cfs, 13 August 2006 (Value, Time, Qualifier)

GetValues
GetValues
GetValues
GetValues

Shale Network ODM Database
NASA
Numerous Additional ODM Databases from Academic Researchers & Local Agencies
The Observations Data Model 1 and WaterML1

Above: Generalization of ODM 1

Above: Screenshot of WaterML 1
Success Stories

Above: Shale Network is an RCN studying impacts of fracking

Above: Little Bear Experimental Watershed (Utah State University)

Above: Examples of local/regional government agencies who have published data in CUAHSI HIS

Above: Examples of federal agencies whose data has been published with CUAHSI HIS
ODM2 and WaterML2

Above: Generalization of ODM 2. This version is extensible and more flexible to accommodate ex-situ data.

Remaining Challenge: ODM2 has not been mapped to WaterML2.

New Standard: TimeSeriesML (?)
What would a national monitoring network that included both discrete and continuous data look like?

Sensor Networks

WaterML 2

WaterML 2

Cloud data compilation and reporting
How do we achieve the goal of developing this network?
Should we be thinking bigger than just water quality data?

What’s the big picture?

What would we like this network to be able to do?

What kinds of questions do we want to ask?