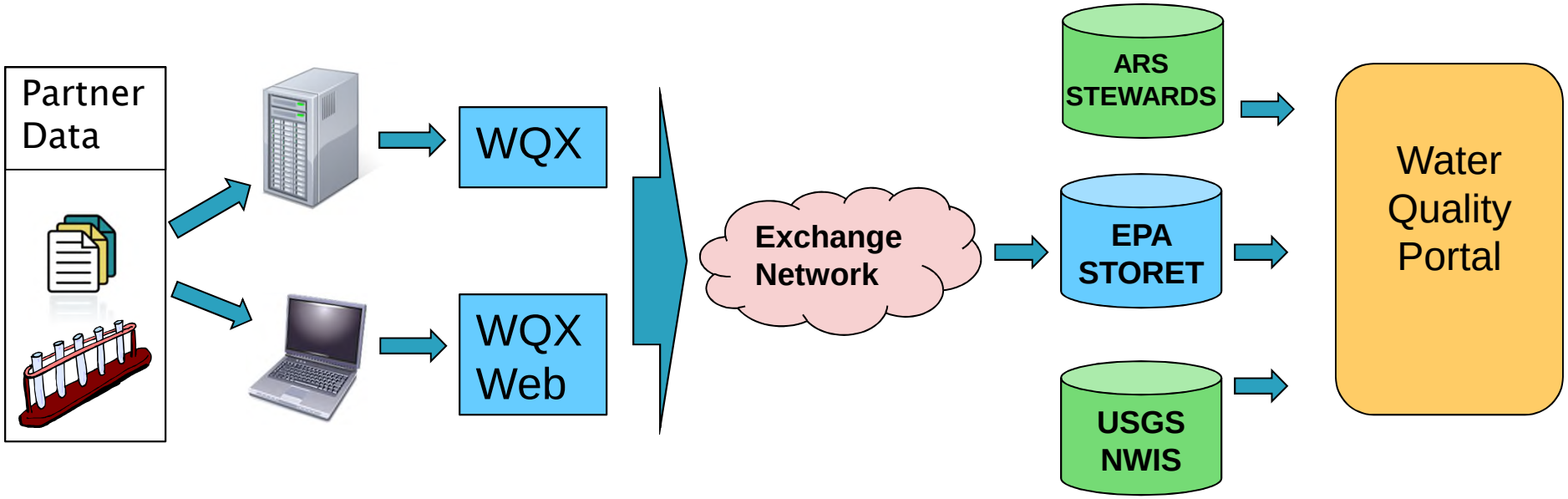
Data Models and Data Sharing Networks: Continuous and Discrete Data

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 role of WQX in Data Sharing



- 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

Metadata Needs May Change Based on Your Data

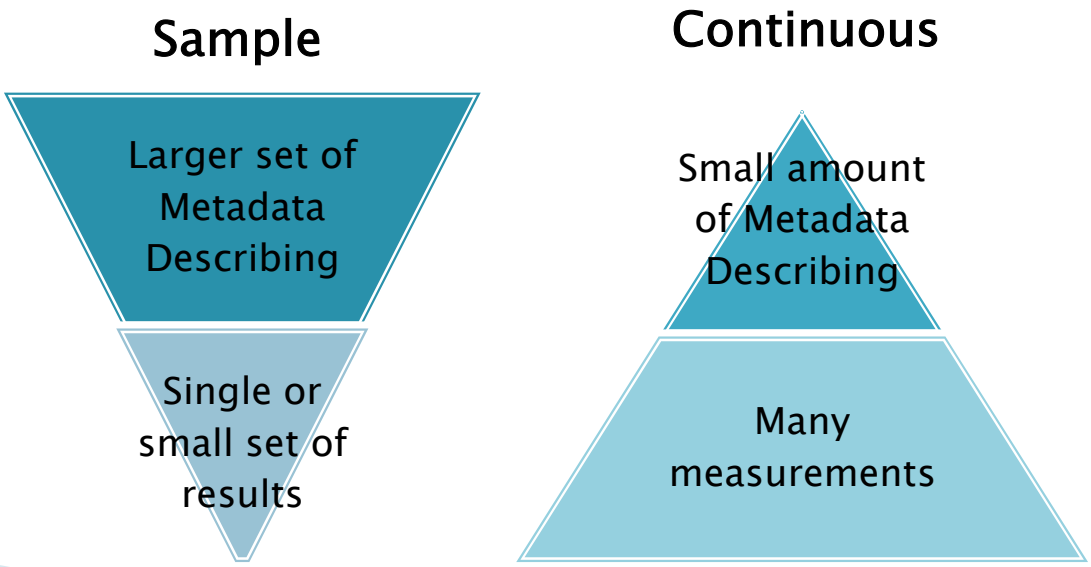
- In the water monitoring world, we can classify monitoring into two types:

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.)

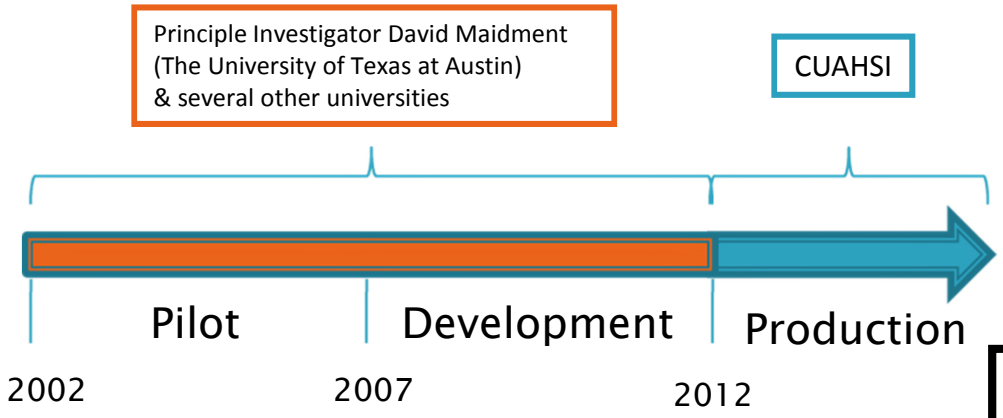
Continuous Monitoring Data

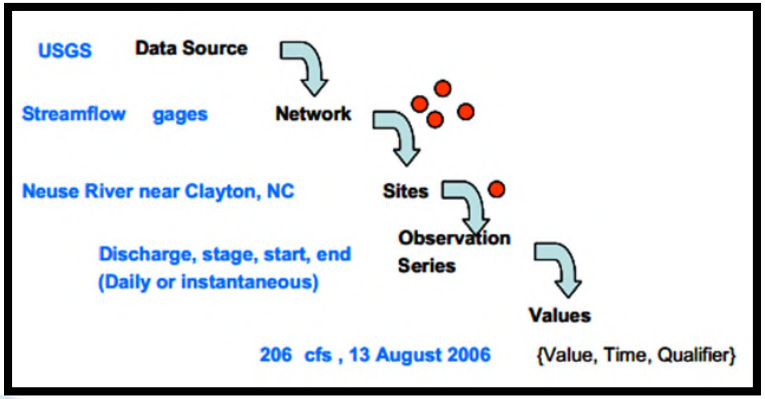
- The data model for continuous data is different from sample data
- EPA recognizes that WQX is probably not the model for this type of data
- EPA is beginning the process of looking at other approaches for continuous data (i.e. WaterML 2)

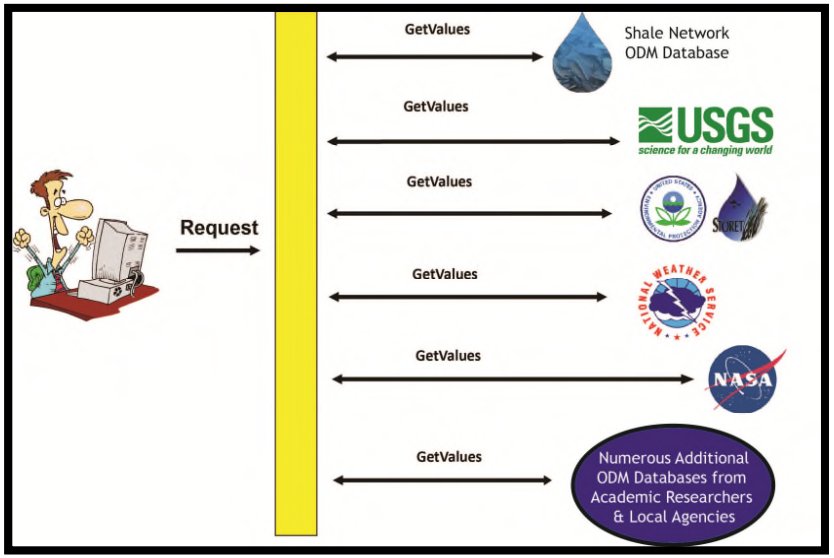


CUAHSI HIS, ODM, and WaterML

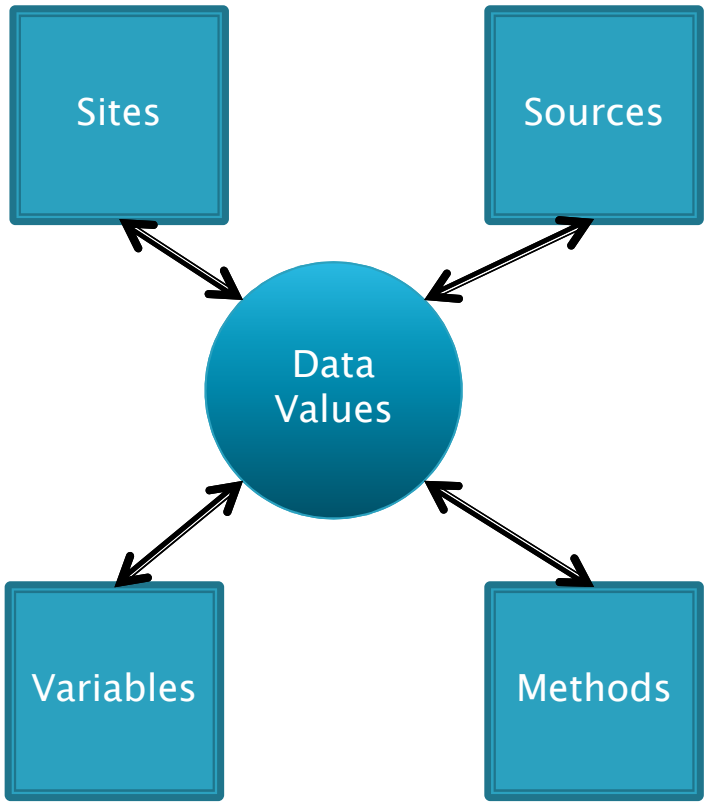
- Hydrologic Information System developed through CUAHSI Community







The Observations Data Model 1 and WaterML1

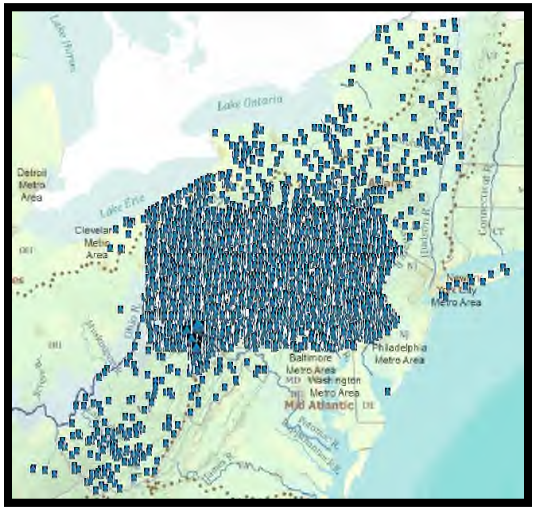


Above: Generalization of ODM 1

```
<timeSeries>
- <sourceInfo xsi:type="SiteInfoType">
    <siteName>Colorado Rv at Austin, TX</siteName>
    <siteCode network="NWIS" siteID="4619631">0815800
  - <geoLocation>
    - <geogLocation xsi:type="LatLonPointType" srs="EPSG
        <latitude>30.24465429</latitude>
        <longitude>-97.694448</longitude>
      </geogLocation>
    </geoLocation>
  </sourceInfo>
- <variable>
    <variableCode vocabulary="NWIS" default="true" variable
    <variableName>Discharge, cubic feet per second</vari
    <units unitsAbbreviation="cfs" unitsCode="35">cubic fee
  </variable>
- <values count="2545">
    <value dateTime="2006-12-31T00:00:00">129</value>
    <value dateTime="2006-12-31T00:15:00">129</value>
    <value dateTime="2006-12-31T00:30:00">129</value>
    <value dateTime="2006-12-31T00:45:00">129</value>
    <value dateTime="2006-12-31T01:00:00">124</value>
    <value dateTime="2006-12-31T01:15:00">129</value>
    <value dateTime="2006-12-31T01:30:00">124</value>
    <value dateTime="2006-12-31T01:45:00">124</value>
```

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)

IPSWICH RIVER WATERSHED ASSOCIATION
The Voice of the River

Above: One of the citizen scientist organizations publishing data with CUAHSI HIS

Suwannee River
WATER MANAGEMENT DISTRICT

NIAGARA PENINSULA CONSERVATION AUTHORITY

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

STORET

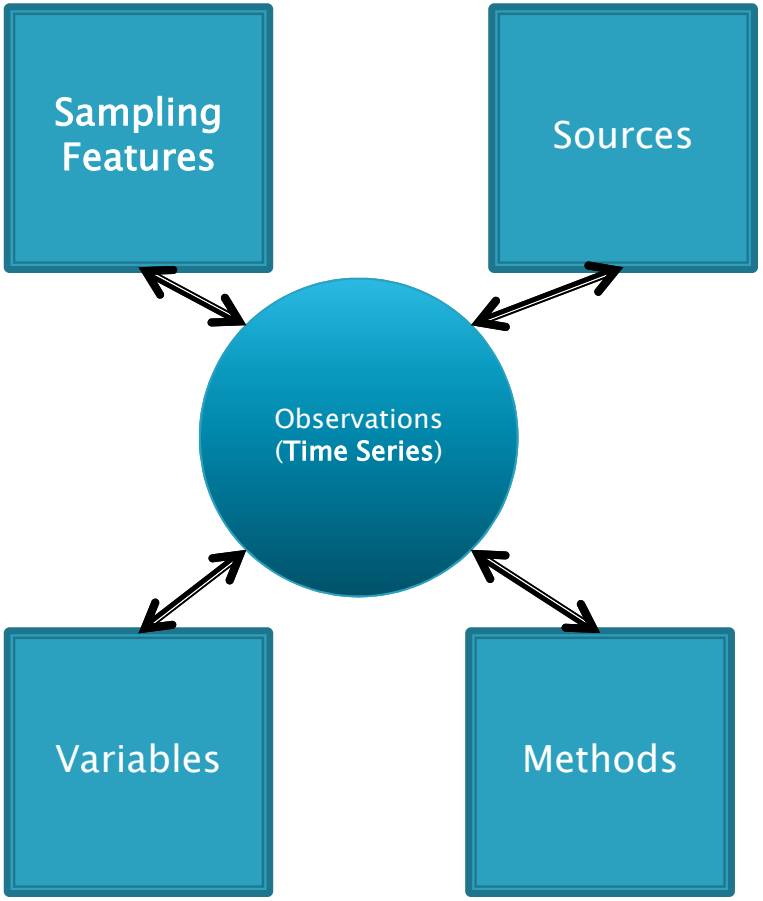
NLDAS

NOAA

USGS
science for a changing world

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.

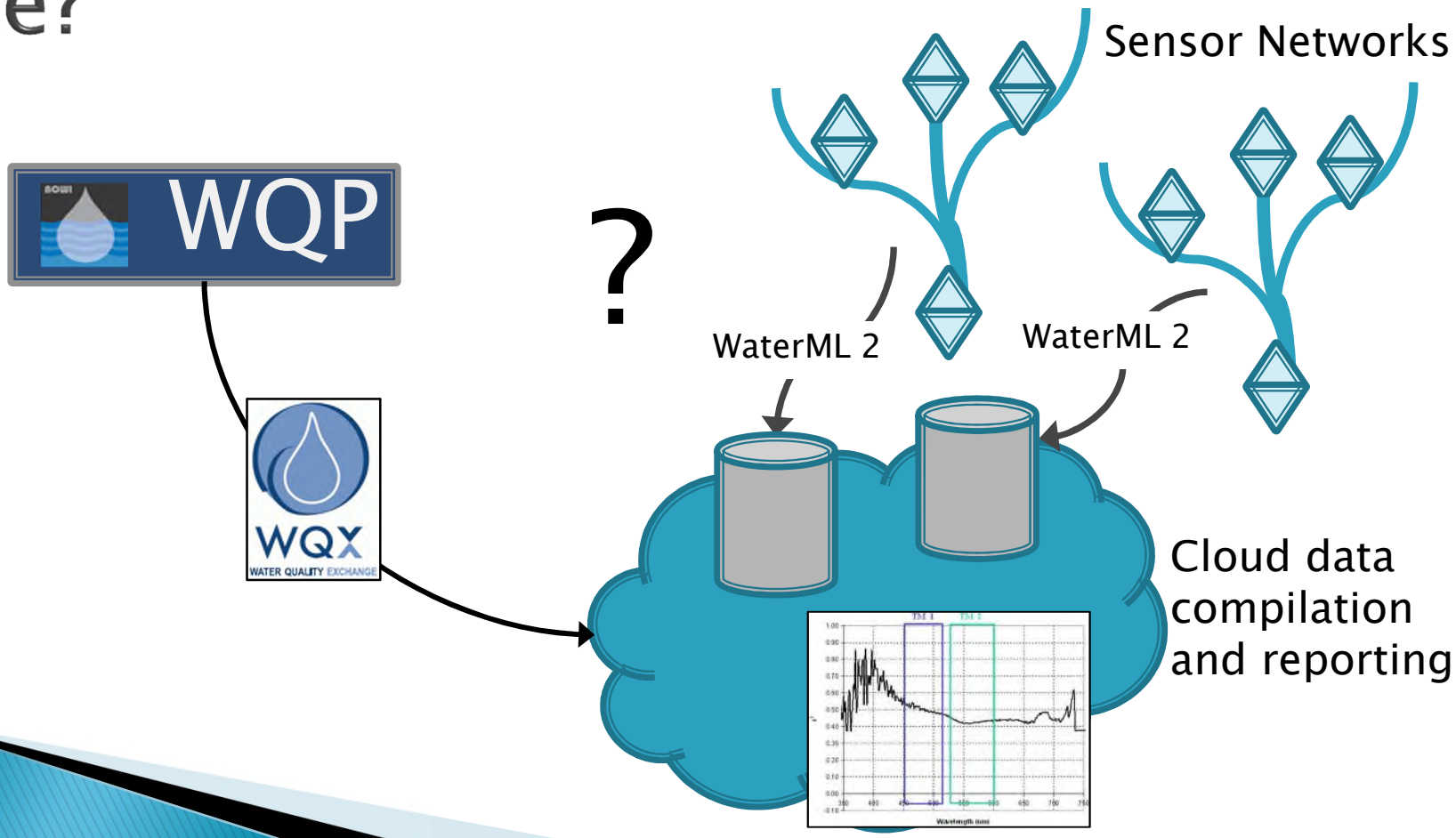
```
28.             <om:observedProperty xlink:href="http://kiwis.kisters.de/parameters/557"
xlink:title="Q"/>
29.             <om:featureOfInterest xlink:href="http://kiwis.kisters.de/stations/1732100"
xlink:title="ATHIEME"/>
30.             <om:result>
31.                 <wml2:MeasurementTimeseries gml:id="Ki.Ts.132042">
32.                     <wml2:temporalExtent>
33.                         <gml:TimePeriod gml:id="Ki.TsTime.1">
34.                             <gml:beginPosition>1990-09-01T00:00:00.000+01:00</gml:beginPosition>
35.                             <gml:endPosition>1990-09-30T00:00:00.000+01:00</gml:endPosition>
36.                         </gml:TimePeriod>
37.                     </wml2:temporalExtent>
38.                     <wml2:defaultPointMetadata>
39.                         <wml2:DefaultTVPMeasurementMetadata>
40.                             <wml2:interpolationType
xlink:href="http://www.opengis.net/def/waterml/2.0/interpolationType/ConstPrec" xlink:title="Constant
in preceding interval"/>
41.                             <wml2:qualifier xlink:href="http://kiwis.kisters.de/statusCodes/40"
xlink:title="40"/>
42.                             <wml2:uom uom="cumec"/>
43.                         </wml2:DefaultTVPMeasurementMetadata>
44.                     </wml2:defaultPointMetadata>
45.                     <wml2:point>
46.                         <wml2:MeasurementTVP>
47.                             <wml2:time>1990-09-01T00:00:00.000+01:00</wml2:time>
48.                             <wml2:value>193.0</wml2:value>
49.                         </wml2:MeasurementTVP>
50.                     </wml2:point>
51.                     <wml2:point>
52.                         <wml2:MeasurementTVP>
53.                             <wml2:time>1990-09-02T00:00:00.000+01:00</wml2:time>
```

Above: Screenshot of WaterML 2 from www.waterml2.org

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?



How do we achieve the goal of developing this network?

How do we get there from here?

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?