New Approaches for Sharing Continuous Monitoring Data

Data Standards in Action
Outline

• Introduction to the Interoperable Watersheds Network
• Core Principles of the IWN
  • Data Standards
  • Metadata
  • Architecture
• Pilot Successes
• Data Appliance: Now Available
• What’s Next?
What is the Interoperable Watersheds Network?

• The Interoperable Watersheds Network was a demonstration project that completed in 2016 that focused on evaluating approaches to improve sensor data sharing

• It was based on knowledge gained from a recommendations report that EPA developed in 2014

• The project focused on addressing three major areas:
  • Data Standards
  • Metadata
  • System Architecture
Principle 1: Common Data Standards

• We needed a common way to represent and communicate the data

• Standards for sensor data already exist, there was no need to create new standards
  • OGC Sensor Observation Service
  • OGC Water ML 2 and Sensor ML

• The Open Geospatial Consortium is an open-source, international standards setting body
Principle 2: Discoverable Metadata

• Needed a standard way to answer the following questions:
  • What data are available and for what parameters?
  • What data can I use?
  • What’s the quality of the data?

• IWN had to develop standard ways to do this (no existing standard existed)

• Further work needs to be done in this area

 zap1

  • nitrate* (11/10/2016 - 02/13/2017)
  • oil* (11/10/2016 - 02/13/2017)
  • total_suspended_solids* (11/10/2016 - 02/13/2017)
  • e_coli* (11/10/2016 - 02/13/2017)
Principle 3: Open Architecture

How do you solve the problem of multiple data providers with large amounts of data that have the potential to change every 3-15 minutes?

• Used a central catalog/index that references every data owner’s assets with the corresponding metadata for each sensor
• Allowed for quick searching and discovery of available data
• This approach is similar to how Google allows you to search the internet
• Actual data comes from the partners systems in real-time
Project Successes

- It contains data from:
  - 8 Partners:
    - EPA Region 1: 2 Sensors
    - EPA Region 7: 18 Sensors
    - EPA Region 10: 1 Sensor
    - EPA ORD: 3 Sensors
    - NJ DEP: 106 Sensors
    - NJ Meadowlands: 3 Sensors
    - Clermont County, OH: 4 Sensors
    - USGS: 15,541 Sensors (nationwide coverage)
Data Appliance is Now Available

• EPA is happy to announce the availability of an open-source tool that enables you to ingest data and have that data published using the common standards.
  • Leverages the Open Source software ‘52 North’
  • Packaged in a ‘Docker’ container to allow for easy deployment
  • Specifically designed to be deployed in the Cloud
  • From start-to-finish, you can be publishing your data in 15 minutes
• Available at [www.github.com/usepa](http://www.github.com/usepa)
  • Search for Interoperable Watersheds Network
Search Index / Catalog is under development by CUAHSI

- CUAHSI has begun development of the search index
- They plan on having something available by the end of this year
- Will incorporate data from any ‘Data Appliance’, USGS, and NOAA.
### Sensors

<table>
<thead>
<tr>
<th>Sensor Id</th>
<th>Sensor Name</th>
<th>Ingest Frequency</th>
<th>Last Ingested</th>
</tr>
</thead>
<tbody>
<tr>
<td>213204</td>
<td>Nooksack@Ferndale</td>
<td>15 min</td>
<td>2018-04-13 04:00:41.975126</td>
</tr>
<tr>
<td>213205</td>
<td>Nooksack@Lynden</td>
<td>15 min</td>
<td>2018-04-13 04:00:41.975126</td>
</tr>
<tr>
<td>213206</td>
<td>Fishtrap@Lynden</td>
<td>15 min</td>
<td>2018-04-13 04:00:41.975126</td>
</tr>
</tbody>
</table>

Showing 1 to 3 of 3 rows
Add Sensor

**Sensor Information**

<table>
<thead>
<tr>
<th>Sensor ID</th>
<th>Short name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enter sensor ID</td>
<td>Enter sensor’s short name</td>
</tr>
</tbody>
</table>

**Long name**

| Enter sensor’s long name |

**Latitude (WGS84)**

| Enter sensor’s latitude |

**Longitude (WGS84)**

| Enter sensor’s longitude |

**Altitude (optional)**

| Enter sensor’s altitude |

**Timezone**

| Select sensor’s timezone |

**Ingest Frequency (minutes)**

| Enter sensor’s ingest frequency |

*Fields marked with * are required.*
Add Sensor

Sensor Information

Data Location (URL)
- Enter data location (URL)

Data Quality
- Select data quality

Timestamp Data Column
- Enter timestamp data column
- Apply QC to Parameter Data

Parameter Data Columns

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Data Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colored dissolved organic matter (CDOM)</td>
<td>C</td>
</tr>
<tr>
<td>Dissolved Oxygen</td>
<td>F</td>
</tr>
</tbody>
</table>

Fields marked with **are required.**

Save changes
Close
## Quality Control

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Operand</th>
<th>Threshold</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Escherichia coli</td>
<td>&gt;</td>
<td>5000</td>
<td>Discard</td>
</tr>
</tbody>
</table>

Showing 1 to 1 of 1 rows
What’s Next?

• The demonstration catalog is still available, however it is not a good long-term solution

• EPA is beginning discussions with CUAHSI for them to take ownership of the catalog

• EPA completing our ‘Authority to Operate’ in Amazon Web Services for the data appliance, which means that it will become available for any EPA office to use this tool to publish their data

• There is still room for some improvement in how the data appliance works

• Explore integration with the Water Quality Portal

• A Demonstration application is currently available at: http://54.210.62.171
QUESTIONS?

Dwane Young
Young.dwane@epa.gov
202-566-1214

Pilot Lessons Learned:
https://www.epa.gov/sites/production/files/2017-
01/documents/iwn_lessonslearned_final_201612.pdf

Data Appliance: https://github.com/USEPA/Interoperable-
Watersheds-Network-Data-Appliance