Streaming Sensor Data: Tools for Acquisition, Management, and Visualization

Amber Spackman Jones
Jeffery S. Horsburgh • Stephanie L. Reeder
Juan Caraballo • Denver Smith
Zach Yoshikawa • Mario Matos

National Water Quality Monitoring Conference • Tampa, FL • May 5, 2016
In Situ Environmental Monitoring

Winning the water war starts with winning the battle on data

3 March 2016

The water meter buried in your front yard isn’t exactly the most cutting-edge piece of technology. While smart meters are read only once per month, resulting in rough usage data—often rounded to the nearest 1,000 gallons. With the limited data, water utility managers can distinguish individual users, such as sprinklers versus toilets, or determine usage by time of day. This limits their ability to spot costly leaks or see opportunities for water conservation. And it gives water users no useful information about how and when they use water.

But for those in the market for $300,000 per instrument, viable sensors are already on the way.

“People are trying to save water, and the only way to do that is to have an instrument that tells you how much water you’re using,” said Dr. Jeff Horstburg, Assistant Professor of Civil and Environmental Engineering at Utah State University. “But it’s not going to work until we have the right kind of instruments.”

The water meter buried in your front yard isn’t exactly the most cutting-edge piece of technology. While they are accurate, most residential water meters are read only once per month, resulting in rough usage data—often rounded to the nearest 1,000 gallons. With the limited data, water utility managers can distinguish individual users, such as sprinklers versus toilets, or determine usage by time of day. This limits their ability to spot costly leaks or see opportunities for water conservation. And it gives water users no useful information about how and when they use water.

Horsburgh was selected by the National Science Foundation to receive the prestigious CAREER award grant and will receive $507,000 to help fund this ongoing research. He and his team have developed an inexpensive technology to make existing analog meters smarter. The program will create new tools that collect, store, and transform high-resolution data into useful information products for water managers.

ầu data management in science and geoscience

The water meter buried in your front yard isn’t exactly the most cutting-edge piece of technology. While they are accurate, most residential water meters are read only once per month, resulting in rough usage data—often rounded to the nearest 1,000 gallons. With the limited data, water utility managers can distinguish individual users, such as sprinklers versus toilets, or determine usage by time of day. This limits their ability to spot costly leaks or see opportunities for water conservation. And it gives water users no useful information about how and when they use water.

Horsburgh was selected by the National Science Foundation to receive the prestigious CAREER award grant and will receive $507,000 to help fund this ongoing research. He and his team have developed an inexpensive technology to make existing analog meters smarter. The program will create new tools that collect, store, and transform high-resolution data into useful information products for water managers.

However, with surges in community movement and communication, we can design a shared cyberinfrastructure to support most of the requirements of field-based research. This cyberinfrastructure would preserve all field data and metadata, and make it publicly available while enhancing time-consumed and resource-intensive post-processing and data analysis.
Gradients Along Mountain To Urban Transitions (GAMUT) Network

- Ecohydrologic observatory deployed in 3 watersheds in northern Utah, USA: Logan River, Red Butte Creek, Provo River
- Watersheds with similar water source (high elevation snow) but different land use transitions
- Measures aspects of water inputs and outputs and water quality over mountain-to-urban gradient
- Mix of aquatic and terrestrial *in situ* and re-locatable sensors
One day = 96 observations
One week = 672 observations
One month = 2880 observations
One year = 35,040 observations
So far (~3 years) = 105,000+ observations

The Data Deluge

Times 14 Aquatic Sites with ~26 Variables
Times 14 Climate Sites with ~74 Variables
Plus different versions of the data (raw versus checked) = over 200 million observations

You need some infrastructure to manage and share the data.
Challenges to Managing Sensor Data

- Volume of data
- Data heterogeneity
- Multiple watersheds
- Multiple institutions
- Multiple personnel
- Scale
- Data quality assurance and quality control
- Standardize data editing
- Synchronize timing, data access, equipment tracking
Cyberinfrastructure

ODM2: Observation Data Model v. 2
GAMUT Data Workflow

- LoggerNet Server
  - Base Station
  - ODM Streaming Data Loader
  - .dat DataLogger Text Files
  - Backup Script
- Database Server
  - Logan ODM Database
  - Red Butte ODM Database
  - Provo ODM Database
- Web Server
  - Map Based Display
  - Time Series Display
  - File-Based Archival and Publication
  - WaterOneFlow Web Services

- Logan River Sites
- Red Butte Creek Sites
- Provo River Sites
Data Loading and Storage

Logan River Sites

Red Butte Creek Sites

Provo River Sites

LoggerNet Server

Base Station

ODM Streaming Data Loader

.dat DataLogger Text Files

Backup Script

Archival of original.dat files

Archival of ODM Databases

Virtual Machine Backups

Logan ODM Database

Red Butte ODM Database

Provo ODM Database

Database Server

Data QC

Data QC

Data QC

Web Server

Map Based Display

Time Series Display
(and other plot types)

File-Based Archival and Publication

WaterOneFlow Web Services

Visualization and Editing of Sensor Data Streams

ODM Tools Python
Data Loading and Storage
Data Loading and Storage

Sensor Network Base Station

Text-based Data Files

YAML Configuration File

Streaming Data Loader Configuration Wizard

Streaming Data Loader Executable

ODM2 Database

Database Server

ODM2
Web-Based Time Series Data Access and Visualization
Web-Based Time Series Data Access and Visualization

The image displays a web-based system for accessing and visualizing time series data. The interface includes filters for different networks, such as the Logan River, and various quality control levels. The system allows users to select data and export it as a .zip file. The table shows specific data points with columns for Plot, Series, Network, Site Code, Variable Code, Variable Name, Quality Control Level, and Num. The selected data point is highlighted, showing details like oxygen concentration and other environmental variables.
Web-Based Time Series Data Access and Visualization
Web-Based Time Series Data Access and Visualization

![Graph showing time series data with oxygen and temperature values over time.](image)
Web-Based Time Series Data Access and Visualization
Web-Based Time Series Data Access and Visualization
Sensor Data QAQC

- Sediment in sensor cup
- Sensor drift and calibration shift
- Strange anomalies
- Dead battery
Sensor Data QAQC

- **Series Selection Filters**
- **Plot Display Options**
- **Date Range Restrictions**
- **Multiple Plot Types**
- **Dynamic Zooming and Panning**
- **Time Series Selection**

The diagram shows a data visualization tool with various features such as plot display options, date range restrictions, multiple plot types, dynamic zooming and panning, series selection, and filters.
QAQC: Post Processing

- Common Editing Tools
- Filters on Data Points
- Commit Edits to the Database
- Dynamic Data Editing Display
- Python Code Console
- Python Script Editor
Managing Physical Infrastructure

What were the field conditions of our discharge measurement?

Has this turbidity sensor been serviced at the factory?

When was the last time we cleaned this sonde?

Who programmed this datalogger?

Which standards did we use for that calibration?

Which sensors are deployed at this site? Who installed them?

How long has that battery been deployed?

When was the last time we changed desiccant?

What is the calibration history of this fDOM sensor?
Managing Physical Infrastructure

Flowchart:
- Data Quality
- Annotations
- Generic Extension
- Results
- Provenance
- Sampling Features
- Laboratory Analyses
- Extension Properties
- External Identifiers
- Controlled Vocabularies
- Equipment

ODM2
Managing Physical Infrastructure
Managing Physical Infrastructure
Managing Physical Infrastructure

### Equipment

<table>
<thead>
<tr>
<th>Equipment Serial Number</th>
<th>Equipment Type</th>
<th>Equipment Code</th>
<th>Equipment Model</th>
<th>Manufacturer Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>53253</td>
<td>Sensor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53252</td>
<td>Sensor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>53251</td>
<td>Sensor</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Person Details

- **Name:** Chris Cox
- **Organization:** Utah State University
- **Address:** chris.cox@usu.edu
- **Phone:** 505-250-5885
- **Email:** chris.cox@usu.edu

### Equipments owned by this person

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>Model</th>
<th>Type</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>0651301-G-00243</td>
<td>BSP65-12</td>
<td>Solar panel</td>
<td>Northern Arizona Wind and Sun</td>
</tr>
<tr>
<td>0651301-G-00244</td>
<td>BSP65-12</td>
<td>Solar panel</td>
<td>Northern Arizona Wind and Sun</td>
</tr>
<tr>
<td>1028</td>
<td>ST-110</td>
<td>Sensor</td>
<td>Apogee Instruments</td>
</tr>
<tr>
<td>120519010198</td>
<td>SPM20P</td>
<td>Solar panel</td>
<td>Northern Arizona Wind and Sun</td>
</tr>
<tr>
<td>120519010205</td>
<td>SPM20P</td>
<td>Solar panel</td>
<td>Northern Arizona Wind and Sun</td>
</tr>
<tr>
<td>120519010206</td>
<td>SPM20P</td>
<td>Solar panel</td>
<td>Northern Arizona Wind and Sun</td>
</tr>
<tr>
<td>120519010208</td>
<td>SPM20P</td>
<td>Solar panel</td>
<td>Northern Arizona Wind and Sun</td>
</tr>
<tr>
<td>120519010210</td>
<td>SPM20P</td>
<td>Solar panel</td>
<td>Northern Arizona Wind and Sun</td>
</tr>
<tr>
<td>120519010212</td>
<td>SPM20P</td>
<td>Solar panel</td>
<td>Northern Arizona Wind and Sun</td>
</tr>
<tr>
<td>120519010213</td>
<td>SPM20P</td>
<td>Solar panel</td>
<td>Northern Arizona Wind and Sun</td>
</tr>
</tbody>
</table>
Managing Physical Infrastructure
Open Source Code

• WEBTSA – Time series data visualization
  – https://github.com/UCHIC/WEBTSA

• ODM Streaming Data Loader
  – https://github.com/ODM2/ODM2StreamingDataLoader

• ODM Tools Python – Sensor Data Management
  – https://github.com/ODM2/ODMToolsPython

• ODM2 Sensor – Sensor equipment management
  – https://github.com/UCHIC/ODM2Sensor
Conclusions

- Researchers are using *in situ* sensors at multiple sites more commonly resulting in large datasets.
- Much of the data management workflow can (and should!) be automated.
- Visualization access can be made broadly available.
- Reproducible QC removes the requirement that only one person perform post-processing and increases confidence in the process.
- The tools described have sped the time from collection to analysis and facilitate sharing and publication of the data.
- Sustainability is enhanced by using community-developed standards and open-source tools.
- Data reusability facilitated by ability to access data and metadata programmatically.
- Data Available at [http://data.iutahepscor.org](http://data.iutahepscor.org)
Questions?

This project is funded by National Science Foundation grant EPS-1208732.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

http://data.iutahepscor.org