Automating Data Management and Sharing within a Large-Scale, Heterogeneous Sensor Network

Jeffery S. Horsburgh

Amber Spackman Jones
Stephanie Reeder
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

Rain, Snow, and Climate

Sapflux

Soil water and chemistry

Water quality

Water quantity

Groundwater

Mobile Platforms
Gradients Along Mountain To Urban Transitions (GAMUT) Network

- Ecohydrologic observatory deployed in 3 watersheds: 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 \textit{in situ} and re-locatable sensors
Gradients Along Mountain to Urban Transitions (GAMUT) Network

Climate/Terrestrial Sites

Aquatic Sites

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Instrument</th>
<th>Variables Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campbell</td>
<td>HC253</td>
<td>Air Temperature and Relative Humidity</td>
</tr>
<tr>
<td>Apogee</td>
<td>ST110</td>
<td>Air Temperature</td>
</tr>
<tr>
<td>Campbell</td>
<td>CS106</td>
<td>Barometric Pressure</td>
</tr>
<tr>
<td>RM Young</td>
<td>5303</td>
<td>Wind Speed/Direction</td>
</tr>
<tr>
<td>Geonor</td>
<td>TB-200</td>
<td>Precipitation</td>
</tr>
<tr>
<td>Judd</td>
<td>Depth Sensor</td>
<td>Snow Depth</td>
</tr>
<tr>
<td>Huskethlux</td>
<td>NR01</td>
<td>Incoming and Outgoing Shortwave and Longwave Radiation</td>
</tr>
<tr>
<td>Apogee</td>
<td>SP-230</td>
<td>Incoming Shortwave Radiation</td>
</tr>
<tr>
<td>Apogee</td>
<td>SQ-110</td>
<td>Incoming and Outgoing Photosynthetically Active Radiation</td>
</tr>
<tr>
<td>Apogee</td>
<td>SI-111</td>
<td>Surface Temperature</td>
</tr>
<tr>
<td>Acclima</td>
<td>ACC-SEN-SDI</td>
<td>Soil Moisture, Temperature, and Conductivity at 5 cm, 10 cm, 20 cm, 50 cm, 100 cm below ground</td>
</tr>
<tr>
<td>Campbell</td>
<td>CS210</td>
<td>Enclosure Humidity</td>
</tr>
<tr>
<td>Campbell</td>
<td>18166</td>
<td>Enclosure open door sensor</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Instrument</th>
<th>Variables Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>YSI</td>
<td>599100-01</td>
<td>Dissolved Oxygen</td>
</tr>
<tr>
<td>YSI</td>
<td>599870-01</td>
<td>Specific Conductivity and Water Temperature</td>
</tr>
<tr>
<td>YSI</td>
<td>599795-02</td>
<td>pH</td>
</tr>
<tr>
<td>YSI</td>
<td>599101-01</td>
<td>Fluorescent Dissolved Organic Matter (fDOM)</td>
</tr>
<tr>
<td>YSI</td>
<td>599102-01</td>
<td>Blue Green Algae and Chlorophyll $a$</td>
</tr>
<tr>
<td>Campbell</td>
<td>CS451</td>
<td>Water Depth and Water Temperature</td>
</tr>
<tr>
<td>FTS</td>
<td>DTS-12</td>
<td>Turbidity and Water Temperature</td>
</tr>
</tbody>
</table>
Gradients Along Mountain to Urban Transitions (GAMUT) Network
Sensor Data Acquisition

Logan River Sites

Red Butte Creek Sites

Provo River Sites

MDF LoggerNet Server

MDF Database Server

MDF Web Server

Sensor Data Acquisition

Base Station

LoggerNet Backup

ODM Streaming Data Loader

Logan ODM Database

Red Butte ODM Database

Provo ODM Database

Map Based Display

Time Series Display (and other plot types)

Visualization and Editing of Sensor Data Streams

ODM Tools Python

File-Based Archival and Publication

WaterOneFlow Web Services

Archival of original .dat files

Archival of ODM Databases

Virtual Machine Backups

Backup Script

Backup Script

Data QC

Data QC

Data QC
Data Loading and Storage
Data Loading and Storage

Data Loading and Storage
Web-Based Data Access

MDF LoggerNet Server
- Base Station
- LoggerNet Backup
- ODM Streaming Data Loader
- .dat DataLogger Text Files
- Backup Script
- Archival of ODM Databases
- Virtual Machine Backups

MDF Database Server
- Logan ODM Database
- Red Butte ODM Database
- Provo ODM Database
- Data QC

MDF Web Server
- Map Based Display
- Time Series Display (and other plot types)
- File-Based Archival and Publication
- WaterOneFlow Web Services

Logan River Sites
- Base Station

Red Butte Creek Sites
- Base Station

Provo River Sites
- Base Station
Web-based Data Access

CUAHSI HIS
Sharing hydrologic data
Enhanced Web-Based Time Series Data Access and Visualization
Development of a QAQC Plan

Quality Assurance:
“...protocols developed and adhered to in a way that minimizes inaccuracies in the data produced ... produces high-quality data while minimizing the need for corrective measures to improve data quality.”
- Site Standardization
- Data Curation (Datalogger Programs and Files, Data Averaging, Database Structure, Equipment Management)
- Replicate Sensors
- Factory Maintenance
- Field Maintenance Schedule and Procedures
- Field Calibration Schedule and Procedures
- Manual Data Monitoring
- Automated Data Monitoring and Alerts
- Recording Events

Quality Control: “occurs after the data are generated and tests whether they meet the necessary requirements for quality outlined by the end users.”
- Data Qualifiers and Flagging
- ODM Tools Python (data management software)
- Quality Control Levels
- Data Processing Steps

QAQC: Automated Alerts

Technicians receive email alerts daily

1. **Power**: battery voltage < 12 volts
2. **Persistence**: value of a variable is unchanging
3. **Updates**: data are not being reported
4. **NaNs**: sensor is reporting “NaN” values

Additional alerts will be implemented as needed (e.g., variable-specific range checks, internal consistency, spatial consistency).
QAQC:
Data Visualization and Management

- Multiple Plot Types
- Plot Display Options
- Dynamic Zooming and Panning
- Date Range Restrictions
- Filters on Data Series
- Build Query
- Export Data Series
- Time Series Selection

Field Names:
- SiteName
- SiteCode
- SiteCode
- SiteName
- VariableName
- VariableCode
- Value
- Value
- Spacing

SELECT * FROM [Attributes] WHERE [SiteName] = 'Little Bear 11'
QAQC: Data Visualization and Management
QAQC: Post Processing

- Data Editing Tools
- Python Code Console
- Dynamic Data Editing Display
- Python Script Editor
Monitoring Equipment Management

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

Who programmed this datalogger?

When was the last time we cleaned this sonde?

What were the field conditions of our discharge measurement?

Has this turbidity sensor been serviced at the factory?

What is this dissolved oxygen sensor’s calibration history?

Which soil moisture sensor is buried at 40 cm?

How long has that battery been deployed?
Monitoring Equipment Management

Data Model developed to relate:
- Equipment
- Field Activities
- Deployments
- Calibrations
- Measured Variables
- Datalogger Programs

Database serves as underlying structure to web interface.
Monitoring Equipment Management

- Record site information, site visits, and field activity details
- Store information on physical equipment
- Track equipment deployments, calibrations, service events
Open Source Code Repositories

- ODM Tools Python – Sensor Data Management
  - https://github.com/UCHIC/ODMToolsPython
- ODM2 Sensor – Sensor equipment management
  - https://github.com/UCHIC/ODM2Sensor
- ODM Streaming Data Loader
  - https://github.com/UCHIC/ODMStreamingDataLoader
- WEBTSA – Time series data visualization
  - https://github.com/UCHIC/WEBTSA
Summary

• Researchers are collecting more data using in situ sensors at multiple sites
• Much of the data management workflow can (and should!) be automated
• The tools described have sped the time from collection to analysis and facilitate sharing and publication of the data
Questions?

Acknowledgements

iUTAH Watershed Technicians:
Chris Cox
Joe Crawford
Dave Eiriksson

UWRL Student Programmers:
Juan Caraballo
Jacob Meline
Mario Matos
Maurier Ramirez