Implementation of Cyberinfrastructure and Data Management Workflow for a Large-Scale Sensor Network

Amber Spackman Jones, (435) 797-7147, amber.jones@usu.edu
Jeffery S. Horsburgh, (435) 797-2946, jeff.horsburgh@usu.edu

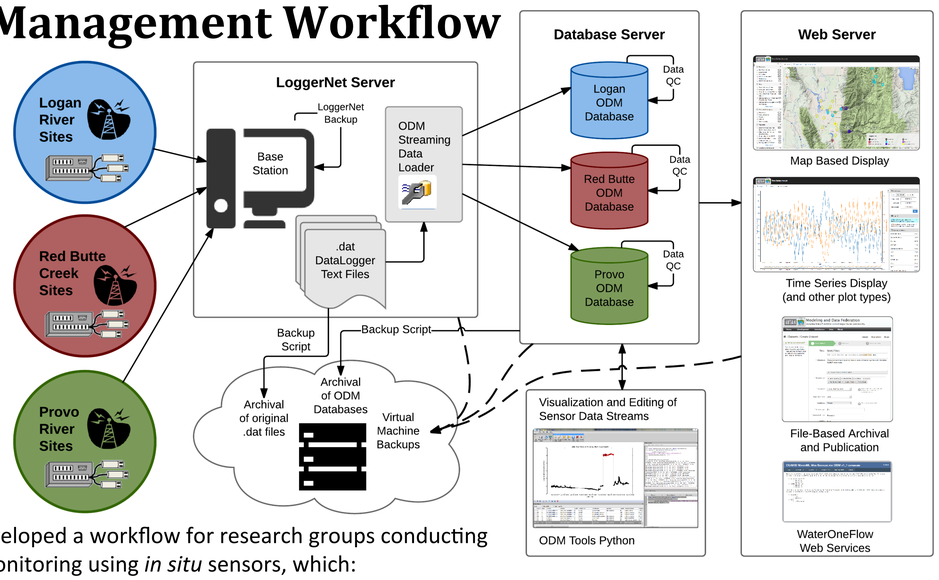
IN41A-3645

This project is funded by National Science Foundation grants 1208732 and 1224638. 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.

Background

Monitoring with *in situ* environmental sensors and other forms of field-based observation presents challenges for data management, particularly for large-scale networks consisting of multiple sites, sensors, and personnel. The availability and utility of these data in addressing scientific questions relies on effective cyberinfrastructure that facilitates transformation of raw sensor data into functional data products. It also depends on the ability of researchers to share and access the data in useable formats. In addition to addressing the challenges presented by the quantity of data, monitoring networks need practices to ensure high data quality, including procedures and tools for post processing.

Data Management Workflow

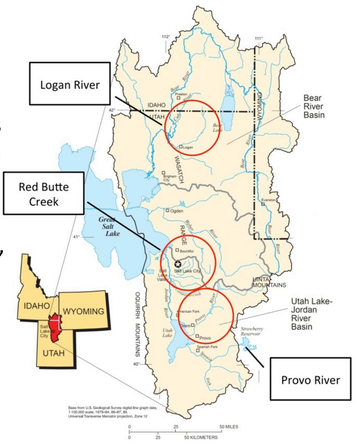


We have developed a workflow for research groups conducting long term monitoring using *in situ* sensors, which:

- **Automate**s much of the data flow, reducing time from collection to analysis.
- Makes data available in near **real time** on the Internet for a broad set of users.
- Incorporates **metadata standards** and **open-source tools**, enhancing data sustainability and reusability.
- Implements **standard procedures** and tools for versioning and **QA/QC**.
- Links observations to **information on equipment**, deployments, calibrations, and other maintenance events.
- **Archives** data with redundant backups.

Application

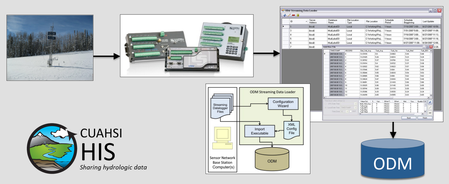
This workflow and associated tools have been implemented for the iUTAH (innovative Urban Transitions and Aridregion Hydro-sustainability) network of aquatic and terrestrial sensors. The variety of environmental sensors and the multi-watershed, multi-institutional nature of the network necessitate a well-planned and efficient workflow for acquiring, managing, and sharing sensor data, which should be useful for similar large-scale and long-term networks. See http://data.iutahepscor.org.



Workflow Features

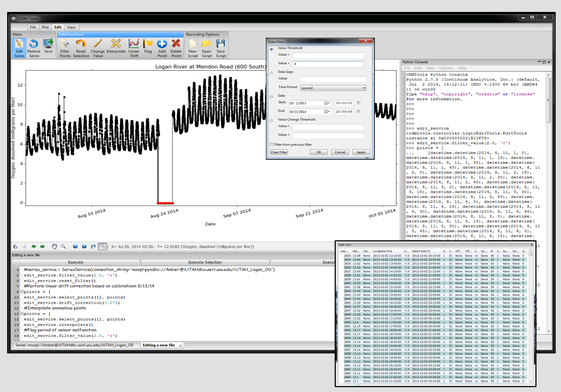
1. Field Data to Database Transfer

- Sensor data are acquired using telemetry and regularly imported to a Campbell Scientific, Inc. LoggerNet Server.
- The Streaming Data Loader application is used to map and automatically transfer data from text files to operational databases.
- An instance of the ODM relational database is deployed on a SQL Server to store data for each watershed in the network.
- ODM is a component of the CUAHSI HIS that facilitates clear metadata, traceable data heritage, interoperability, and semantic and syntactic consistency.



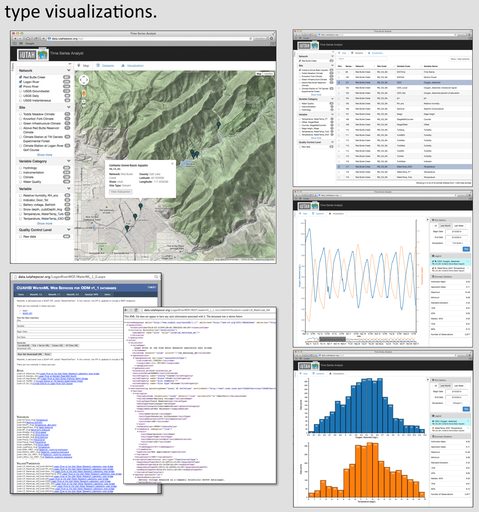
2. Data Quality Assurance/Quality Control

- A QA/QC Plan is being followed by all personnel for consistency in practices to ensure high quality data.
- Automated alerts run on the SQL Server that regularly check incoming data. If pre-defined criteria are met, technicians are notified and can address issues.
- We've developed ODM Tools Python, a desktop application for visualizing, managing, and post-processing sensor data.
- ODM Tools Python features a Python code console and Python script editor that record all data edits.
- Edits can be reviewed and are reproducible.



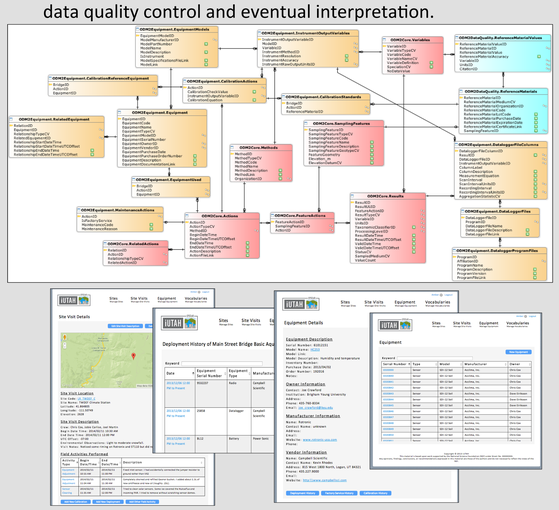
3. Web-Based Data Publication

- Data are published via CUAHSI HIS WaterOneFlow web services.
- Data are published via a file-based repository system for each site on an annual basis.
- Data are discoverable and accessible via an online visualization tool.
- The web-based application includes faceted searching, a map-based display, time series selection and export, and various plot type visualizations.



4. Physical Infrastructure Tracking

- We developed a data model to relate equipment, field activities, deployments, calibrations, measured variables, and datalogger programs.
- The database serves as the underlying structure for a web interface, which permits recording and accessing this information.
- A direct linkage to the data being collected will facilitate data quality control and eventual interpretation.



Open Source Code Repositories on GitHUB:

https://github.com/UCHIC/ODMStreamingDataLoader
https://github.com/UCHIC/ODMToolsPython
https://github.com/UCHIC/WEBTSA
https://github.com/UCHIC/ODM2Sensor