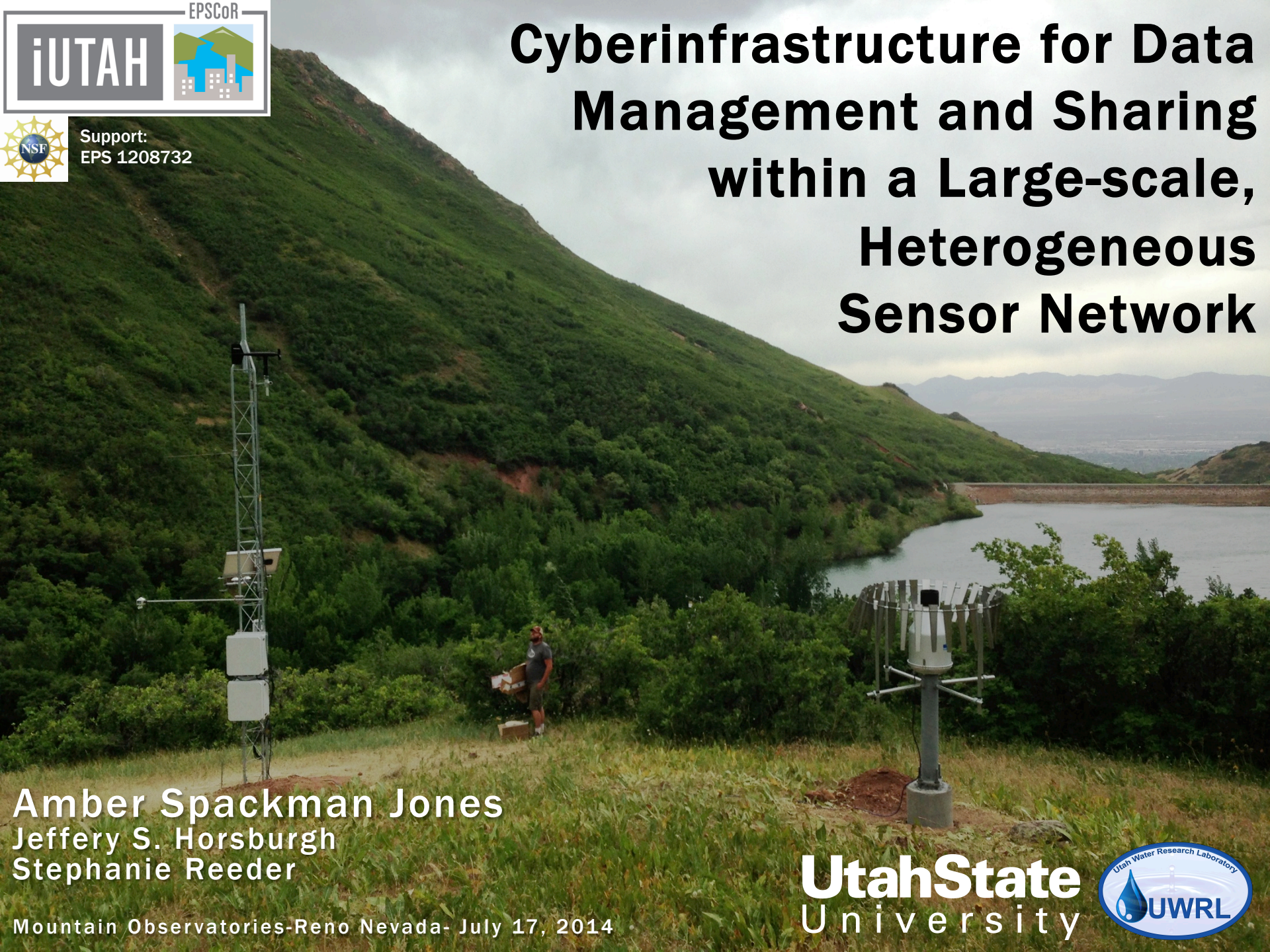


Cyberinfrastructure for Data Management and Sharing within a Large-scale, Heterogeneous Sensor Network

Support:
EPS 1208732

Amber Spackman Jones
Jeffery S. Horsburgh
Stephanie Reeder

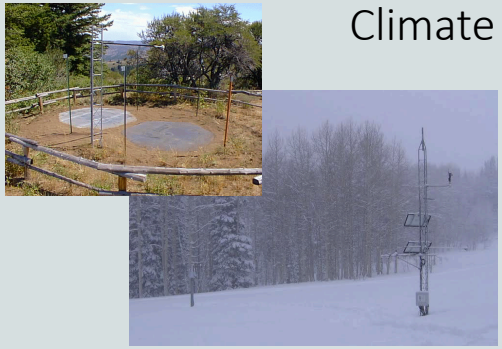
Utah State University
Utah Water Research Laboratory
UWRL

Mountain Observatories-Reno Nevada- July 17, 2014

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

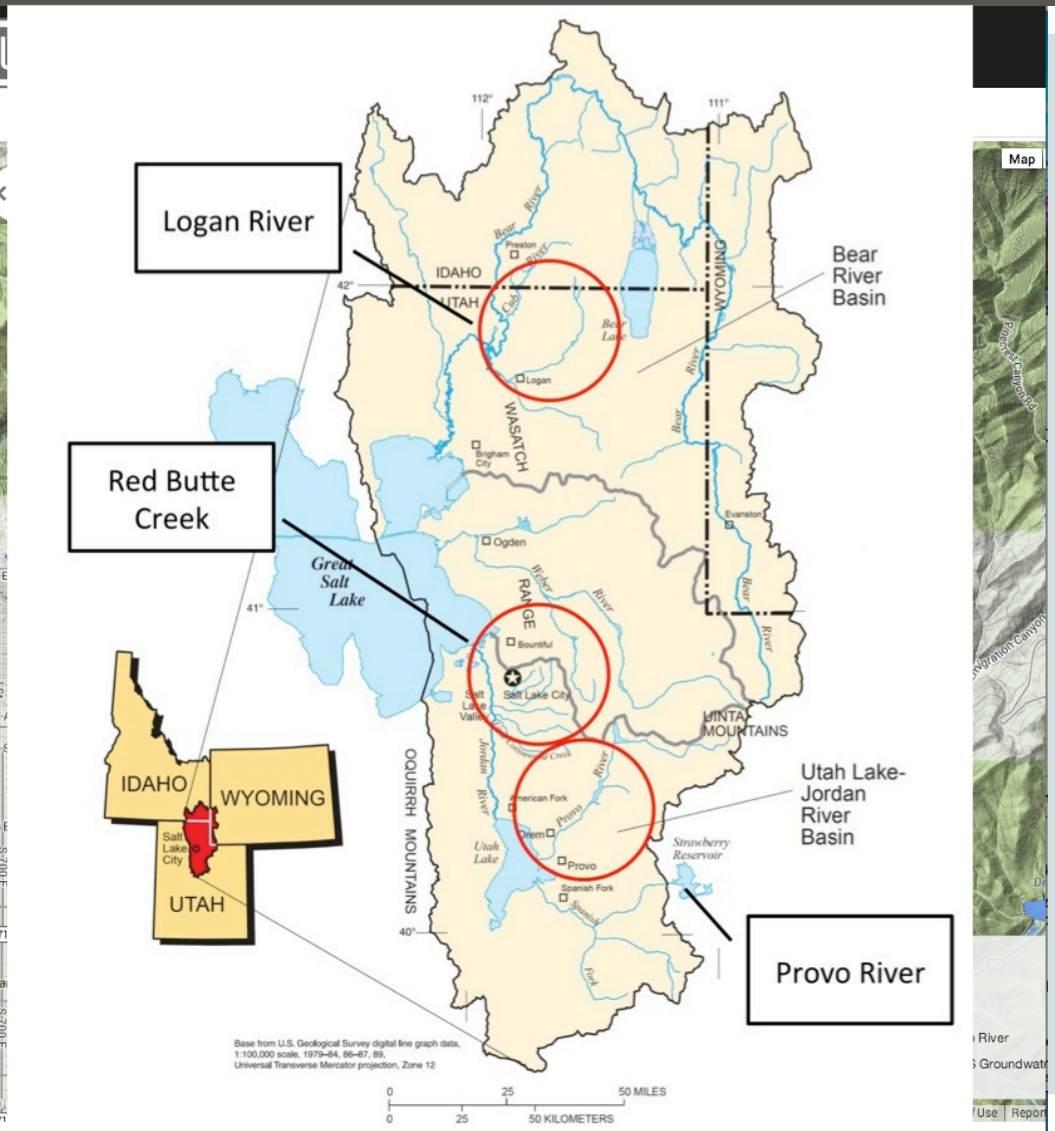
Water quality

Water quantity

Groundwater

Mobile Platforms

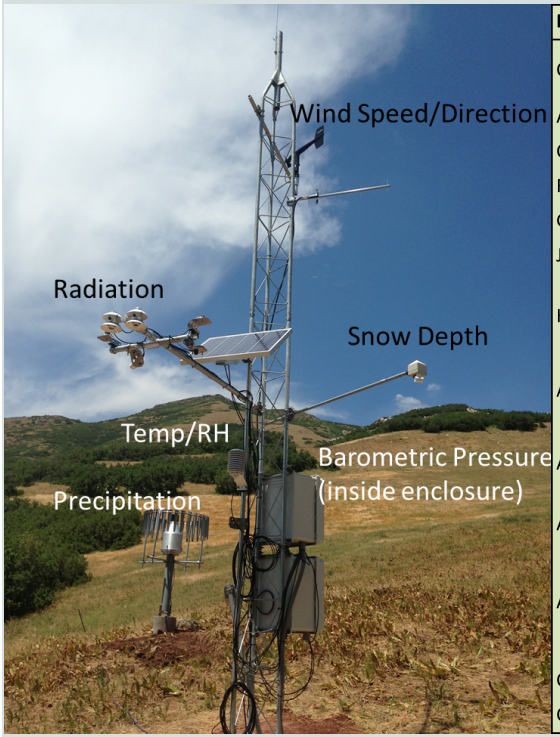
Gradients Along Mountain To Urban Transitions (GAMUT) Network



- Ecohydrologic observatory deployed in 3 watersheds 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

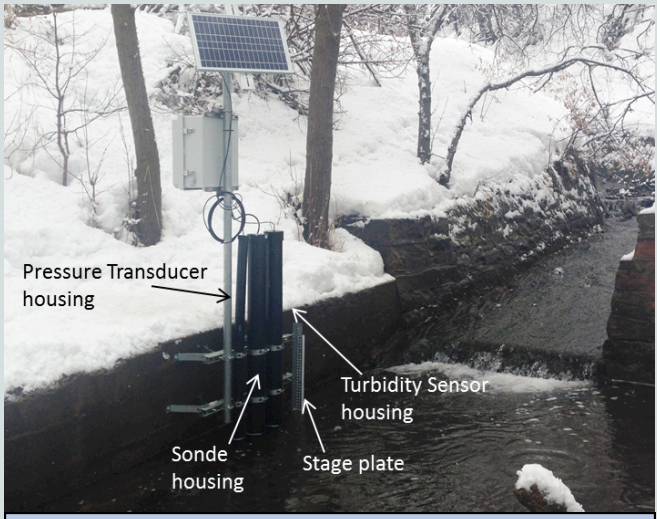
Gradients Along Mountain to Urban Transitions (GAMUT) Network

Climate/Terrestrial Sites



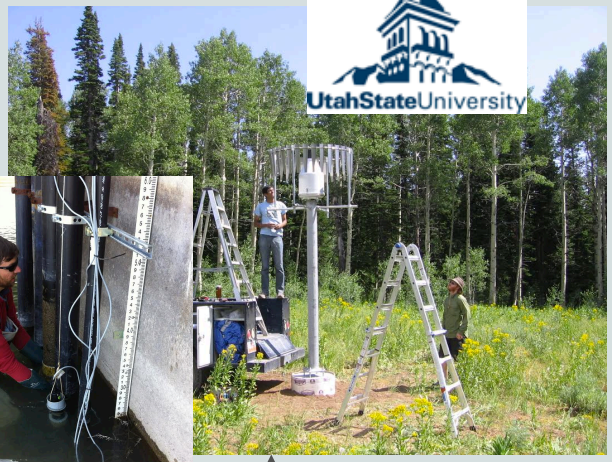
Manufacturer	Instrument	Variables Measured
Campbell	HC2S3	Air Temperature and Relative Humidity
Apogee	ST110	Air Temperature
Campbell	CS106	Barometric Pressure
RM Young	5303	Wind Speed/Direction
Geonor	TB-200	Precipitation
Judd	Depth Sensor	Snow Depth
Hukseflux	NR01	Incoming and Outgoing Shortwave and Longwave Radiation
Apogee	SP-230	Incoming Shortwave Radiation
Apogee	SQ-110	Incoming and Outgoing Photosynthetically Active Radiation
Apogee	SI-111	Surface Temperature
Acclima	ACC-SEN-SDI	Soil Moisture, Temperature, and Conductivity at 5 cm, 10 cm, 20 cm, 50 cm, 100 cm below ground
Campbell	CS210	Enclosure Humidity
Campbell	18166	Enclosure open door sensor

Aquatic Sites



Manufacturer	Instrument	Variables Measured
YSI	599100-01	Dissolved Oxygen
YSI	599870-01	Specific Conductivity and Water Temperature
YSI	599795-02	pH
YSI	599101-01	Fluorescent Dissolved Organic Matter (fDOM)
YSI	599102-01	Blue Green Algae and Chlorophyll *a*
Campbell	CS451	Water Depth and Water Temperature
FTS	DTS-12	Turbidity and Water Temperature

Gradients Along Mountain to Urban Transitions (GAMUT) Network



BYU BRIGHAM YOUNG UNIVERSITY

UtahStateUniversity

THE UNIVERSITY OF UTAH

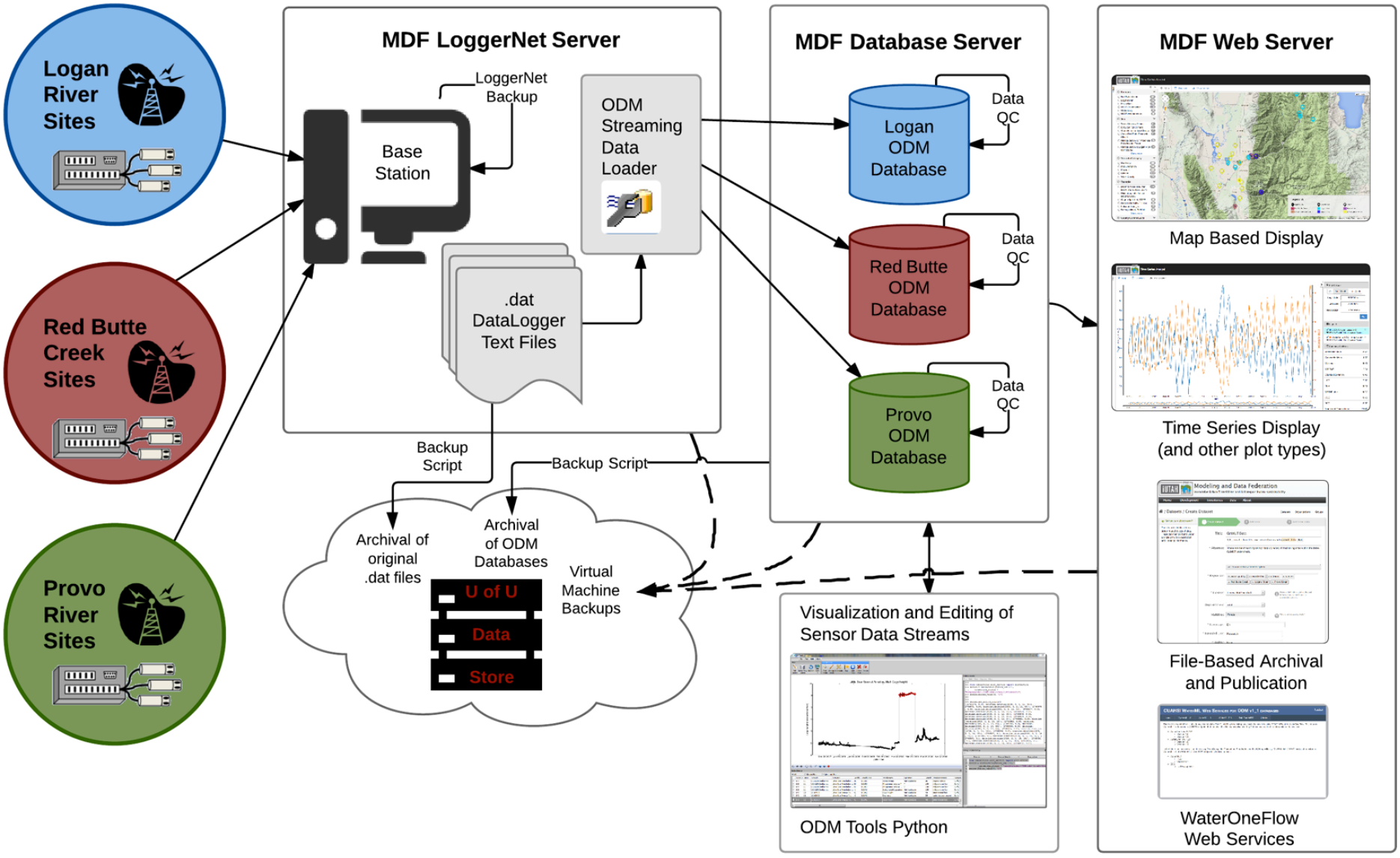
iUTAH EPSCoR

Modeling and Data Federation

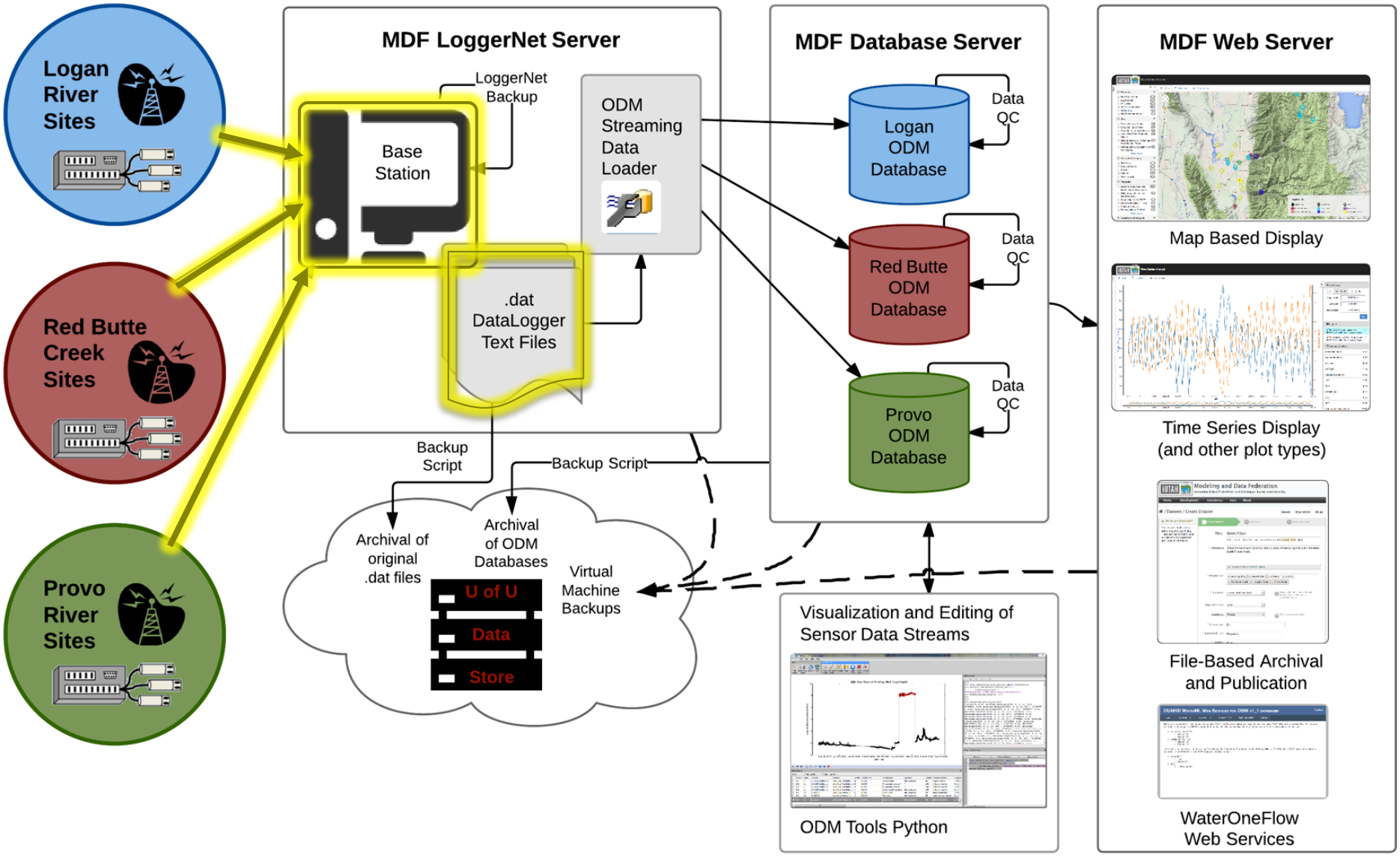
innovative Urban Transitions and Aridregion Hydro-sustainability

Utah Water Research Laboratory UWRL

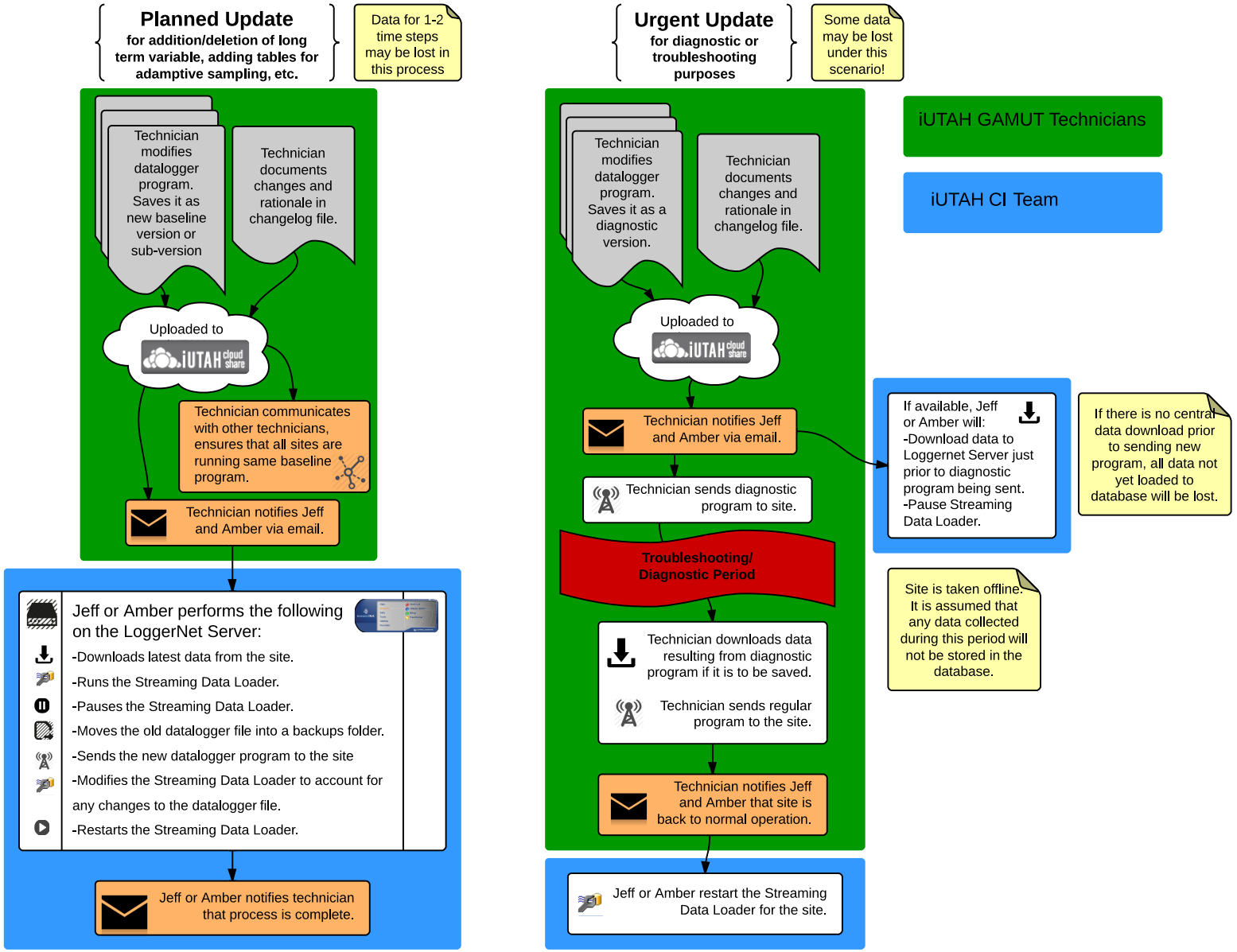
GAMUT Data Workflow



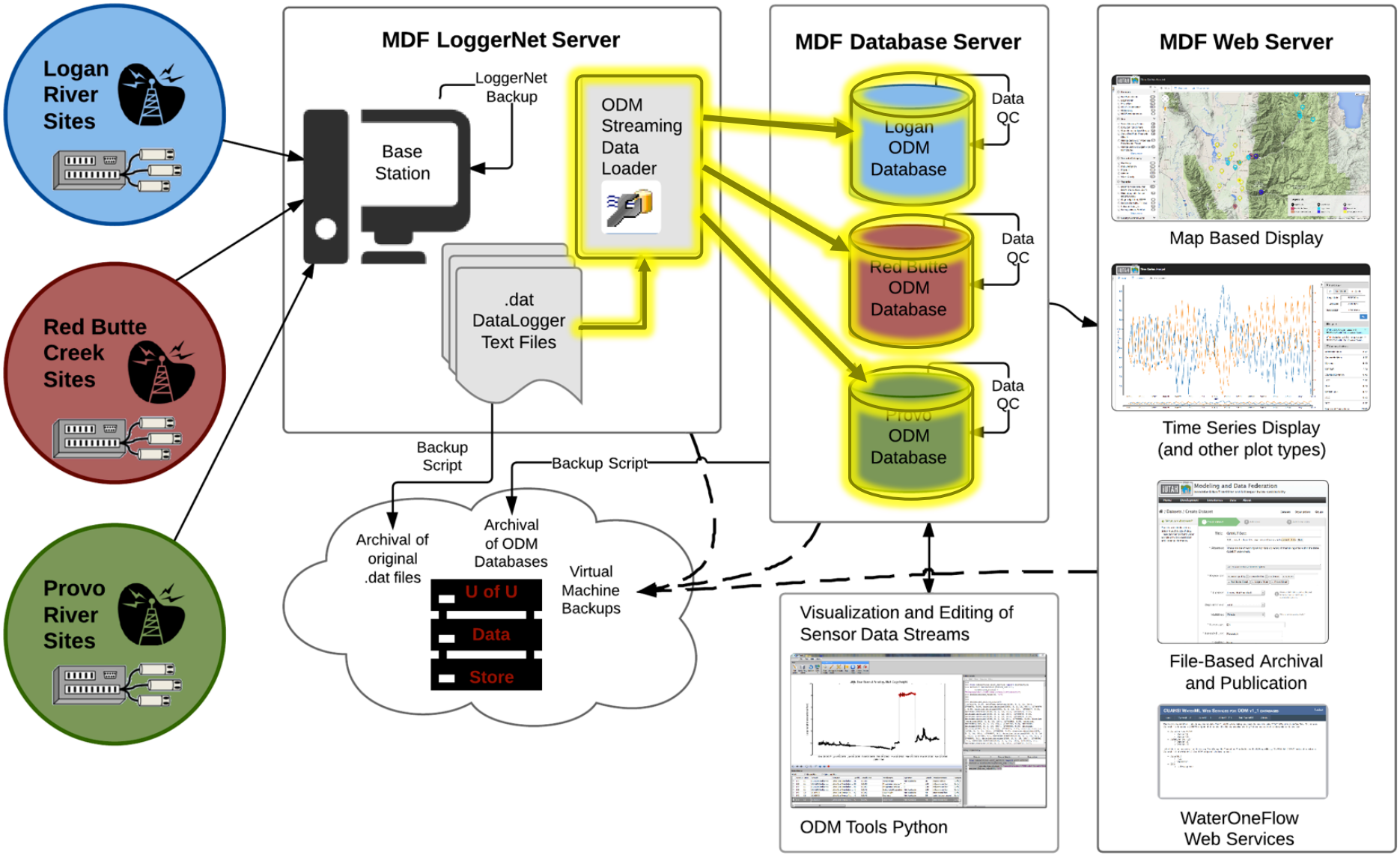
Sensor Data Acquisition



Datalogger Program Updates

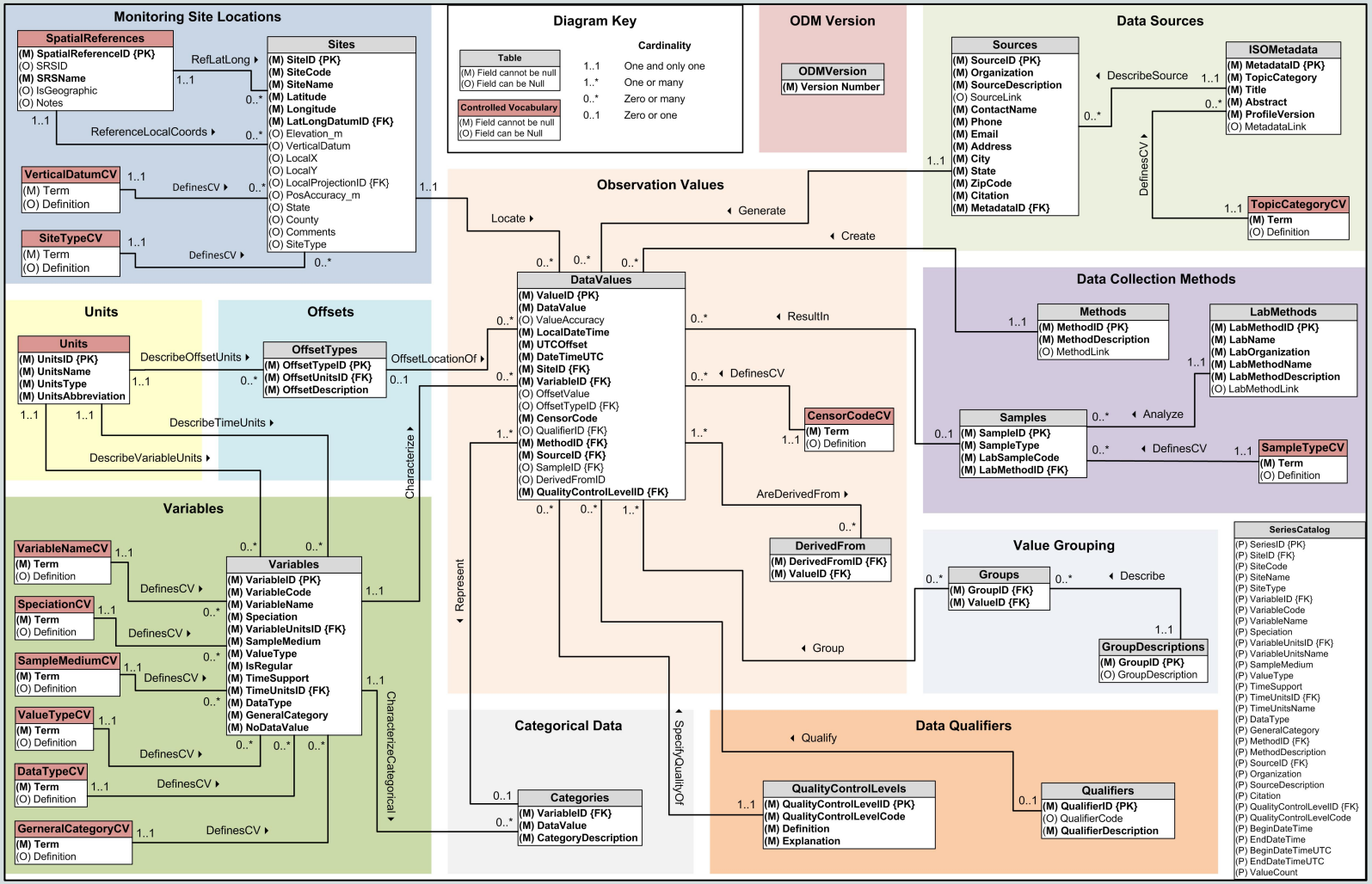


Data Loading and Storage



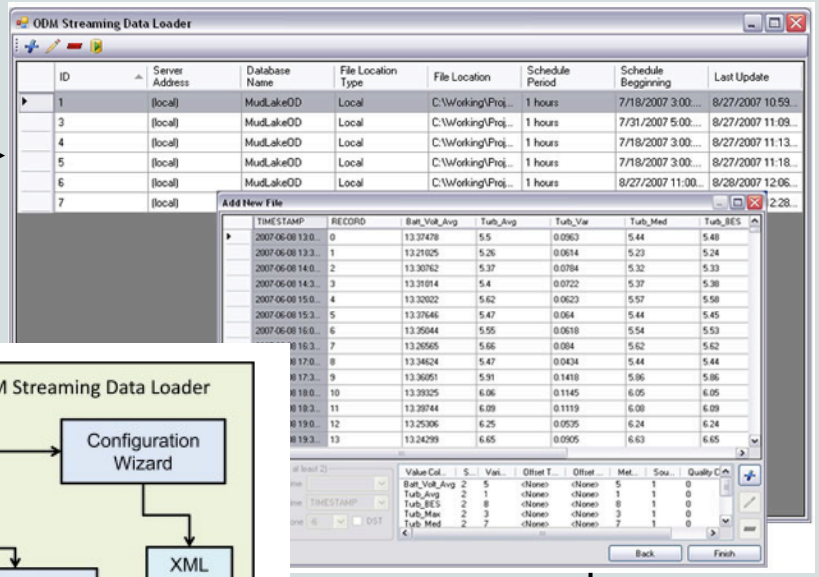
Data Loading and Storage

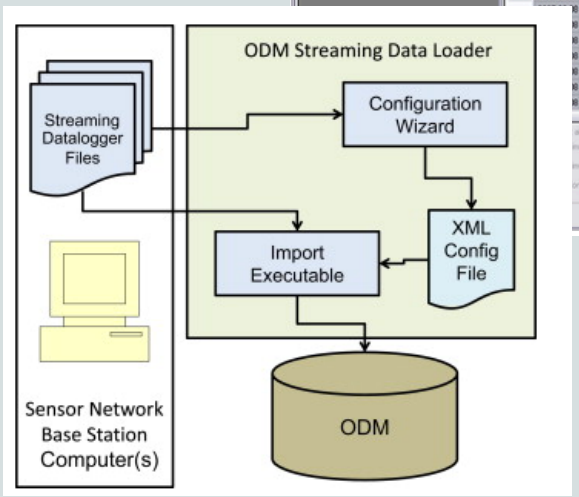
CUAHSI HIS
Sharing hydrologic data

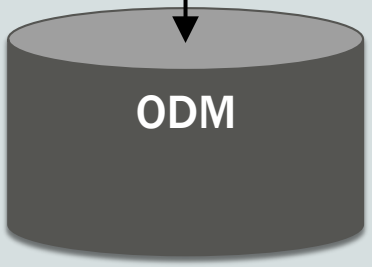


Horsburgh, J. S., D. G. Tarboton, D. R. Maidment, and I. Zaslavsky (2008), A relational model for environmental and water resources data, *Water Resources Research*, 44, W05406, doi:10.1029/2007WR006392.

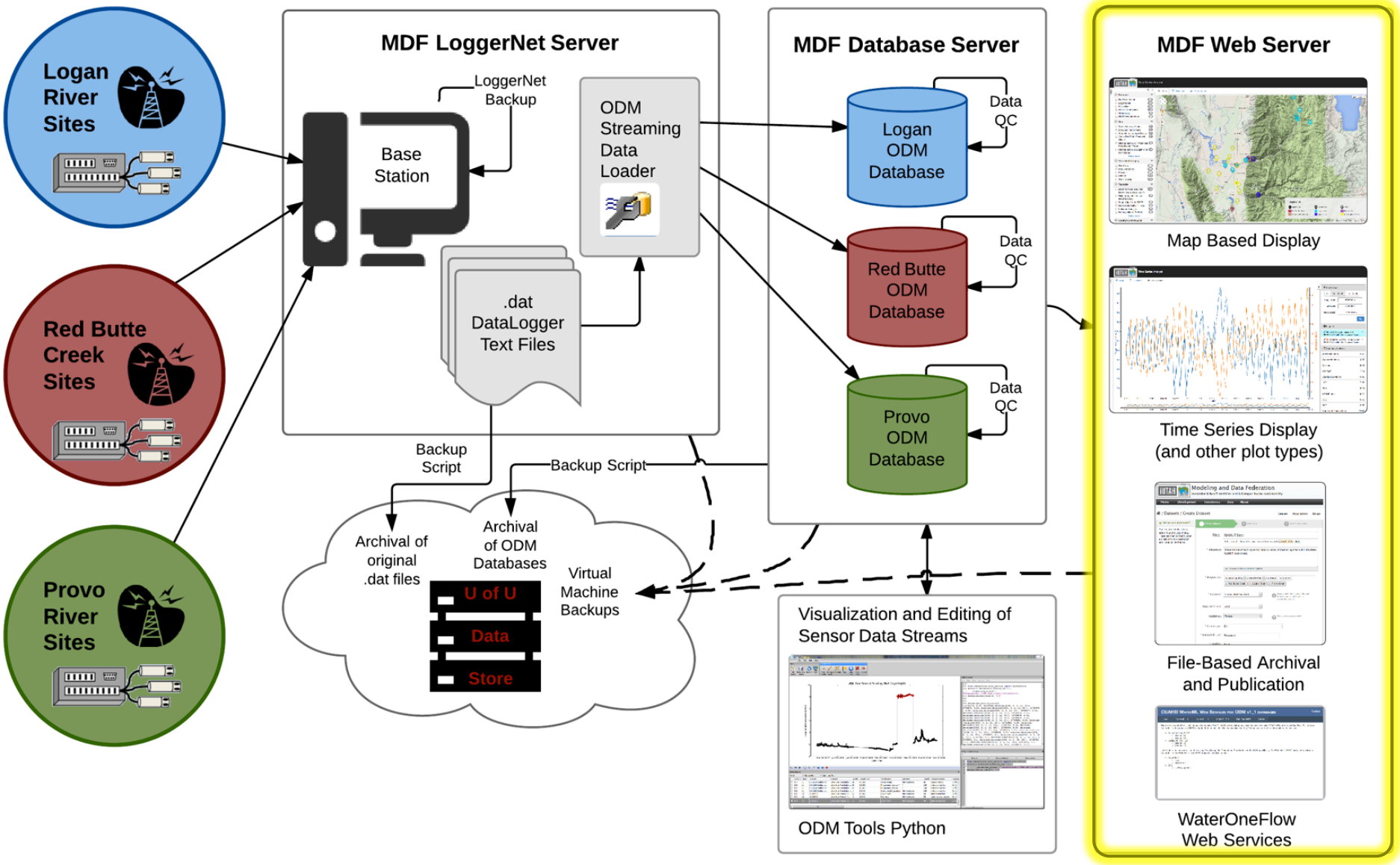
Data Loading and Storage



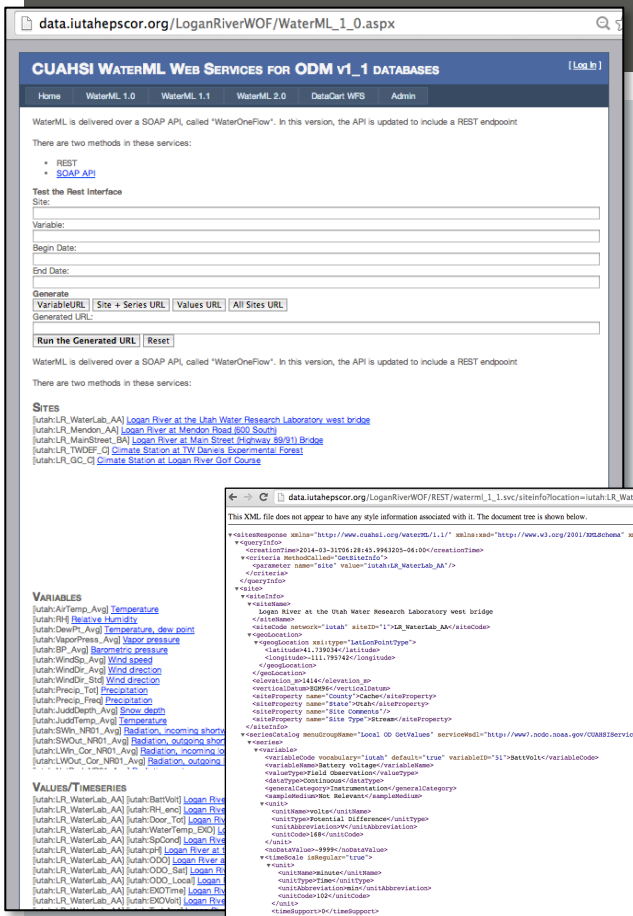


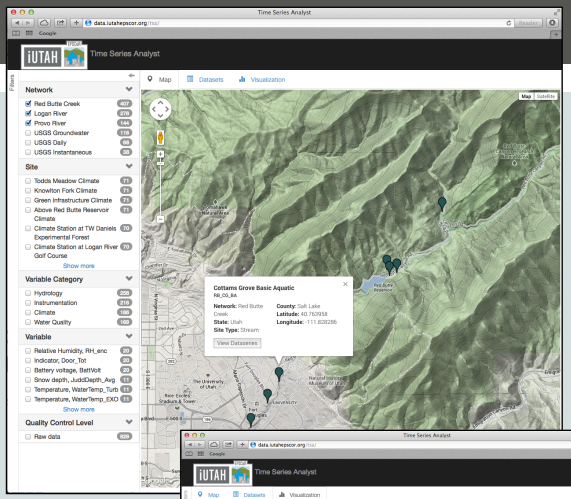


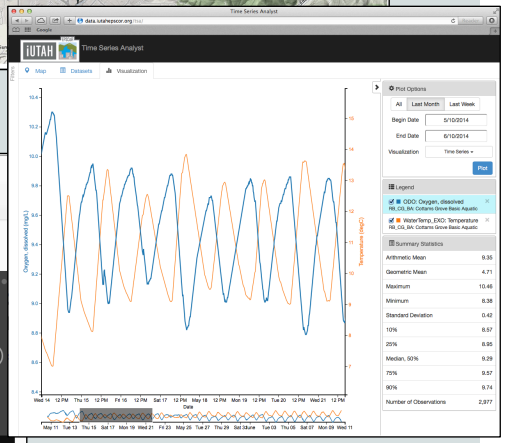
Web-Based Data Access

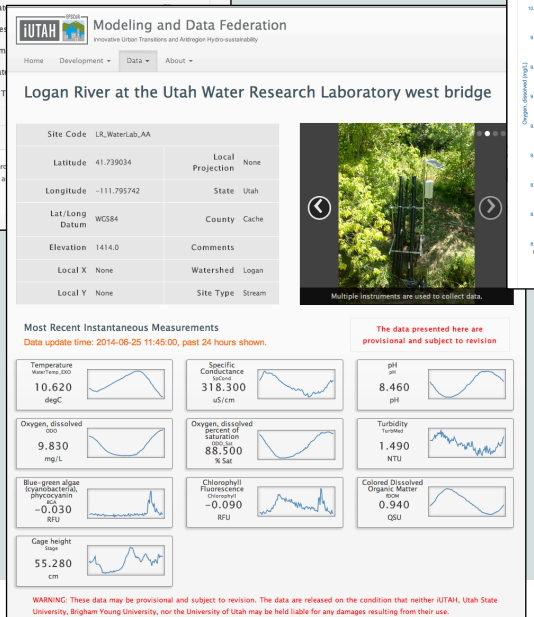


Web-based Data Access



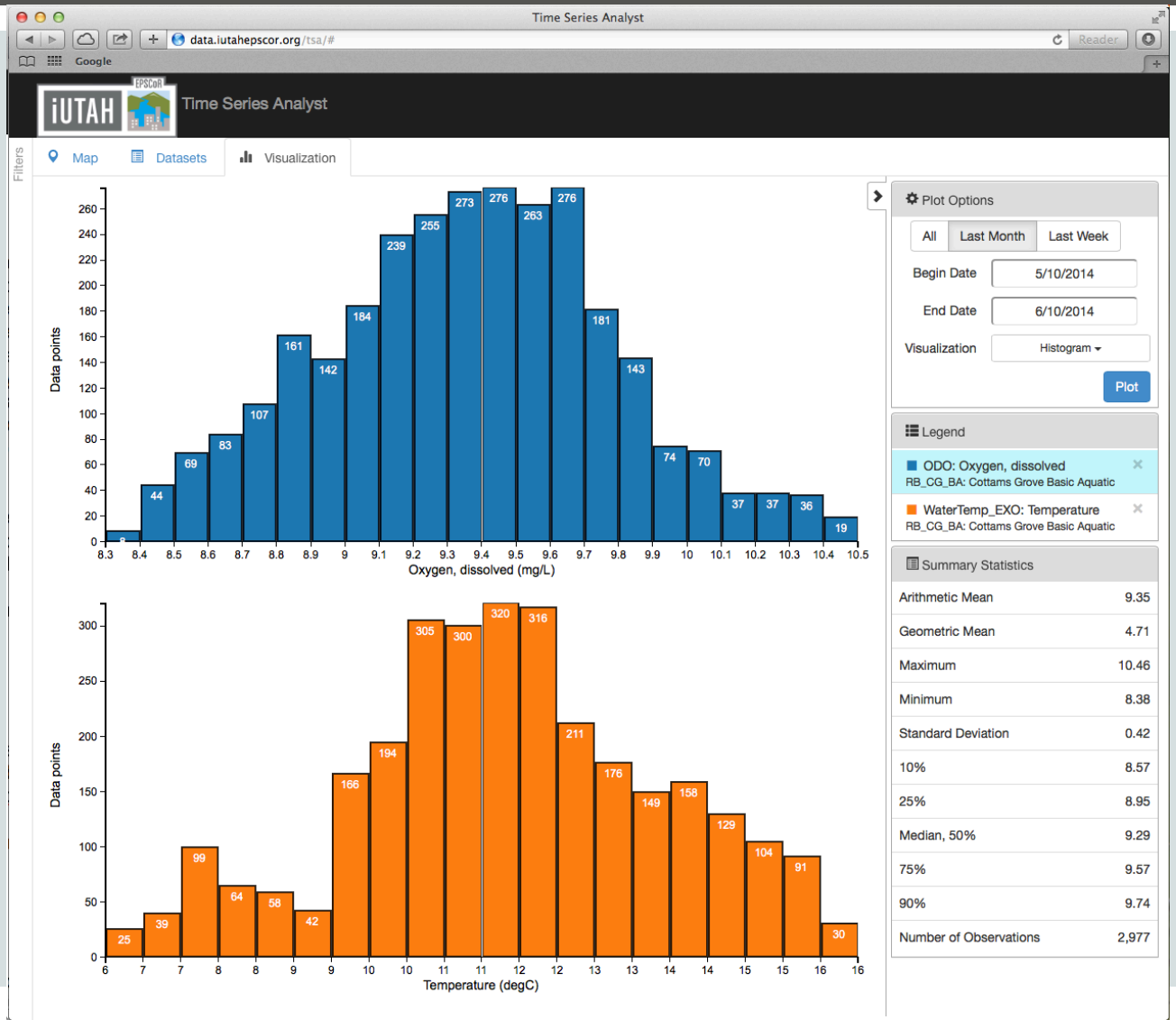




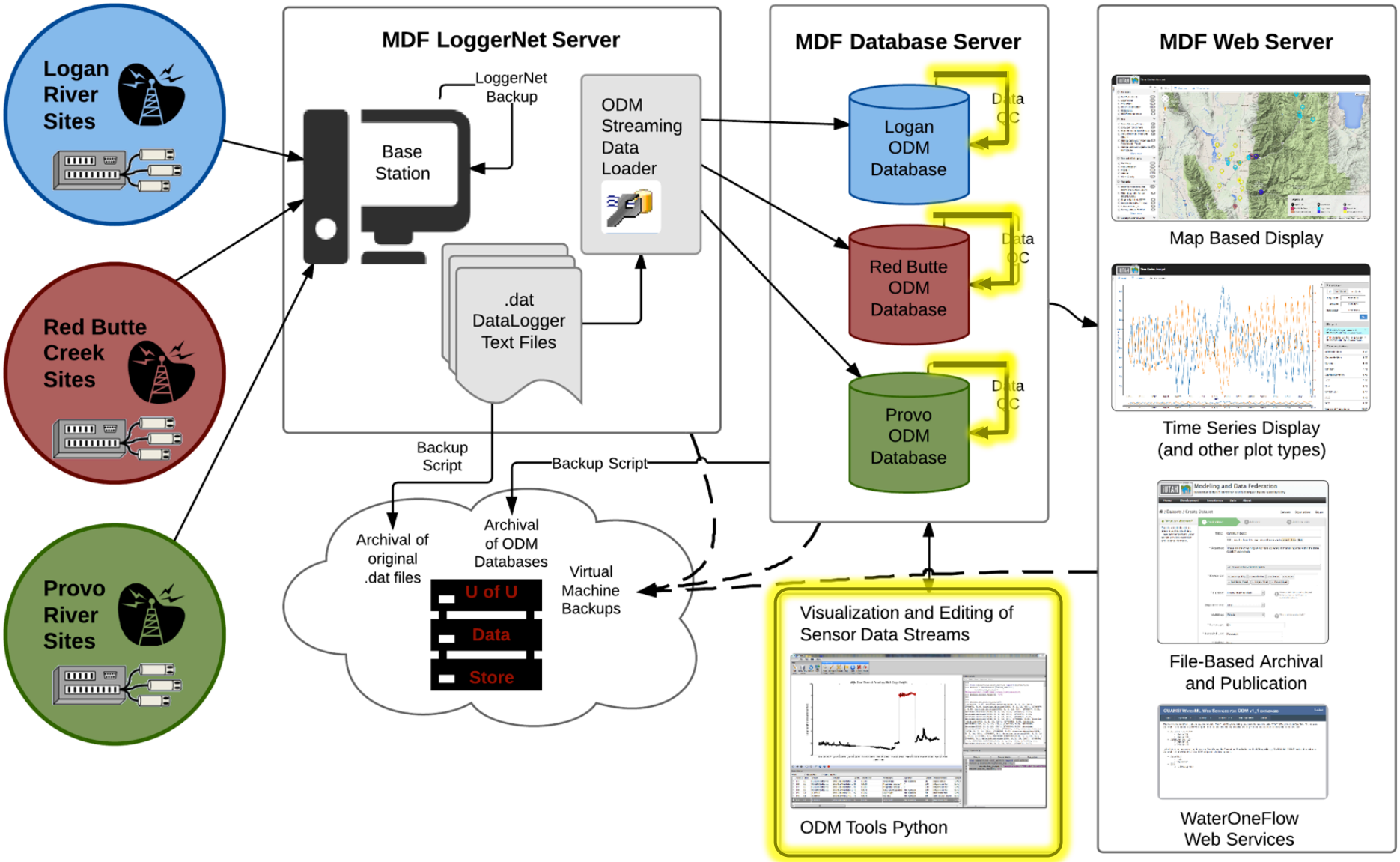


CUAHSI HIS
Sharing hydrologic data

Enhanced Web-Based Time Series Data Access and Visualization



Sensor Data QAQC



Development of a QAQC Plan

Quality Assurance:

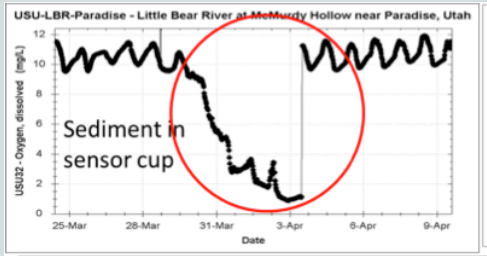
"…protocols developed and adhered to in a way that minimizes inaccuracies in the data produced … produces high-quality data while minimizing the need for corrective measures to improve data quality."

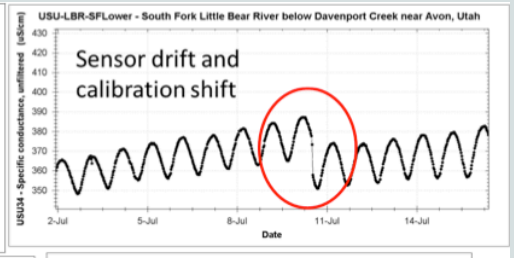
- Site Standardization
- Data Curation (Datalogger Programs and Files, Data Averaging, Database Structure, Equipment Management)
- Replicate Sensors
- Factory Maintenance
- Field Maintenance Schedule and Procedures
- Field Calibration Schedule and Procedures
- Manual Data Monitoring
- Automated Data Monitoring and Alerts
- Recording Events

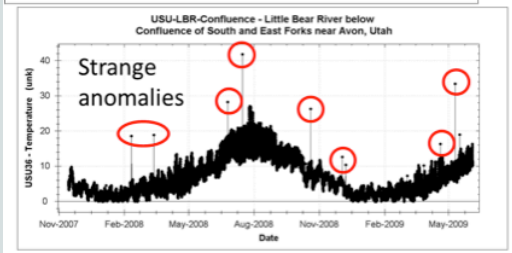
Quality Control:

"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







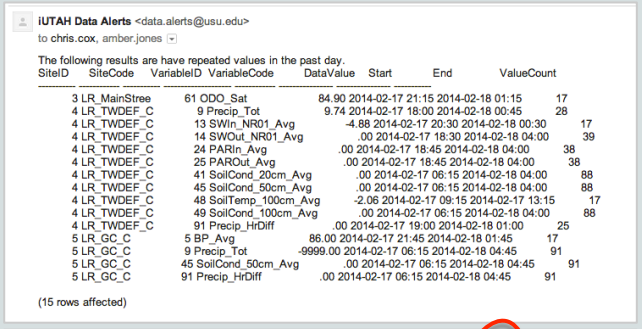
Campbell, J. L., Rustad, L. E., Porter, J. H., Taylor, J. R., Ethan, W., Shanley, J. B., … (2013). Quantity is Nothing without Quality. *BioScience*, *63*(7), 574–585. doi:10.1525/bio.2013.63.7.10

QAQC: Automated Alerts

Technicians receive email alerts daily

1. Power: batter 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).

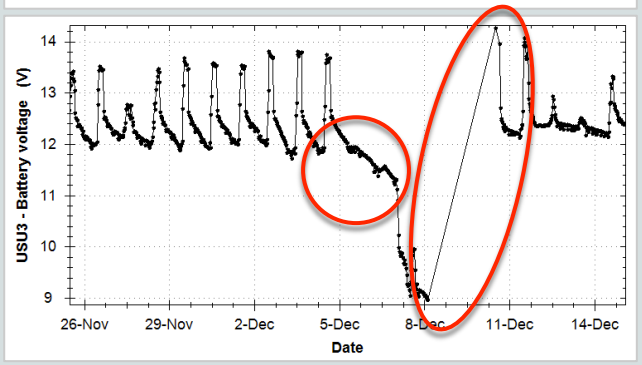


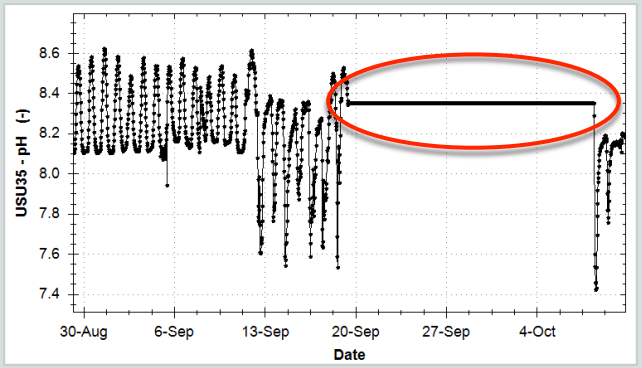
iUTAH Data Alerts <data.alerts@usu.edu>
to chris.cox, amber.jones

The following results are have repeated values in the past day.

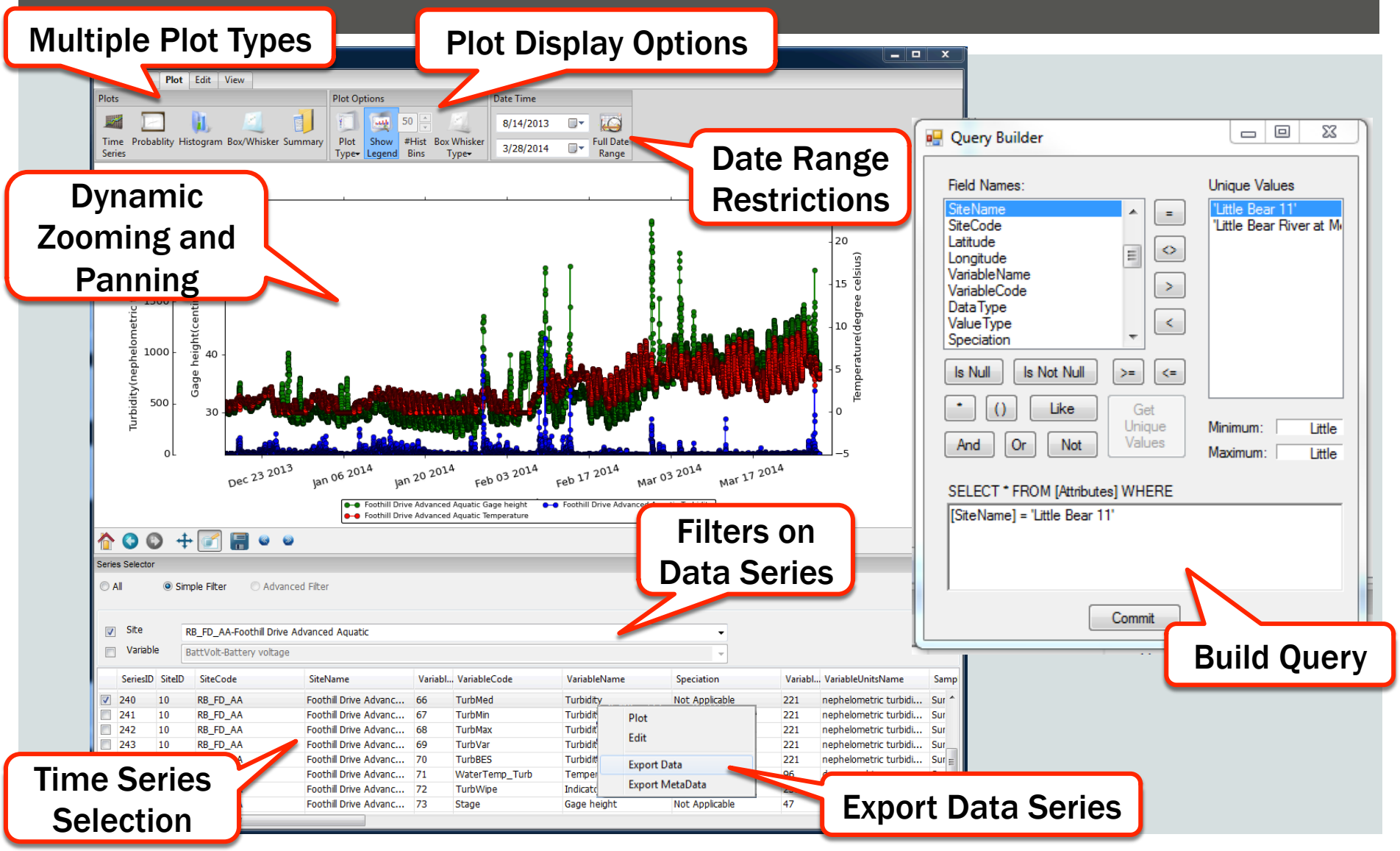
SiteID	SiteCode	VariableID	VariableCode	DataValue	Start	End	ValueCount
3	LR_MainStree	61	ODO_Sat	84.90	2014-02-17 21:15	2014-02-18 01:15	17
4	LR_TWDEF_C	9	Precip_Tot	9.74	2014-02-17 18:00	2014-02-18 00:45	28
4	LR_TWDEF_C	13	SWIn_NR01_Avg	-4.88	2014-02-17 20:30	2014-02-18 00:30	17
4	LR_TWDEF_C	14	SWOut_NR01_Avg	.00	2014-02-17 18:30	2014-02-18 04:00	39
4	LR_TWDEF_C	24	PARIn_Avg	.00	2014-02-17 18:45	2014-02-18 04:00	38
4	LR_TWDEF_C	25	PAROut_Avg	.00	2014-02-17 18:45	2014-02-18 04:00	38
4	LR_TWDEF_C	41	SoilCond_20cm_Avg	.00	2014-02-17 06:15	2014-02-18 04:00	88
4	LR_TWDEF_C	45	SoilCond_50cm_Avg	.00	2014-02-17 06:15	2014-02-18 04:00	88
4	LR_TWDEF_C	48	SoilTemp_100cm_Avg	-2.06	2014-02-17 09:15	2014-02-17 13:15	17
4	LR_TWDEF_C	49	SoilCond_100cm_Avg	.00	2014-02-17 06:15	2014-02-18 04:00	88
4	LR_TWDEF_C	91	Precip_HrDiff	.00	2014-02-17 19:00	2014-02-18 01:00	25
5	LR_GC_C	5	BP_Avg	86.00	2014-02-17 21:45	2014-02-18 01:45	17
5	LR_GC_C	9	Precip_Tot	-9999.00	2014-02-17 06:15	2014-02-18 04:45	91
5	LR_GC_C	45	SoilCond_50cm_Avg	.00	2014-02-17 06:15	2014-02-18 04:45	91
5	LR_GC_C	91	Precip_HrDiff	.00	2014-02-17 06:15	2014-02-18 04:45	91

(15 rows affected)

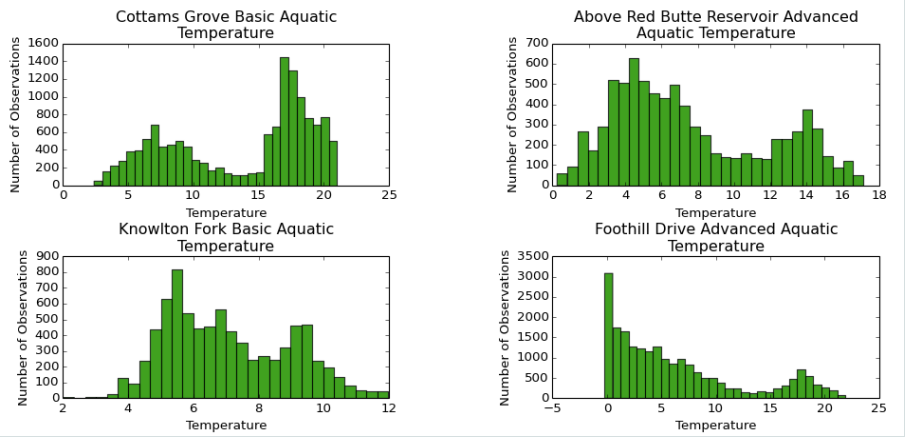


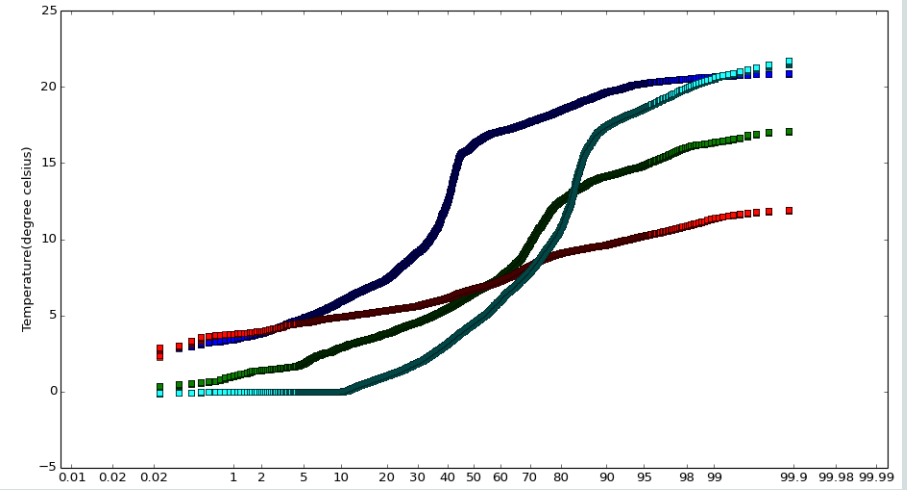


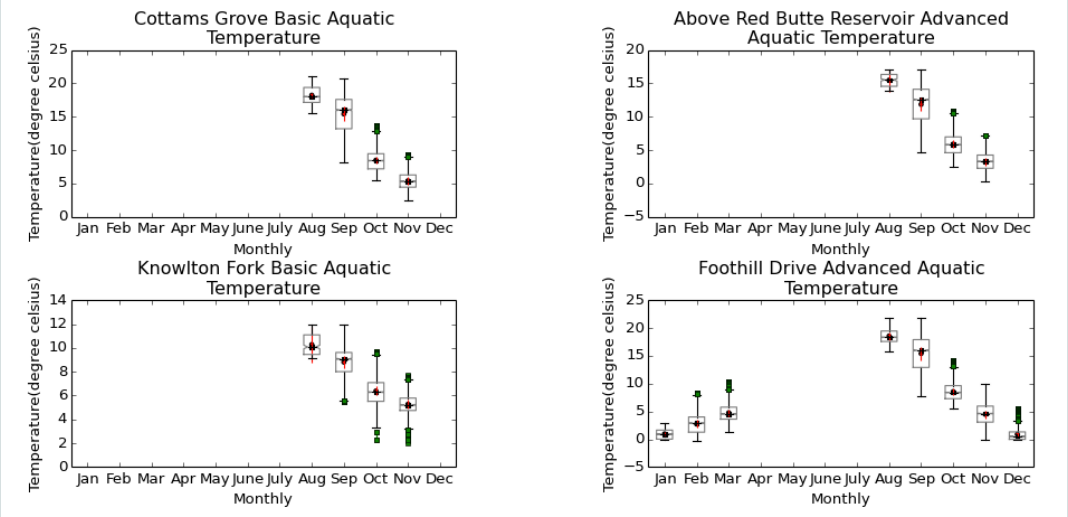
QAQC:
Data Visualization and Management



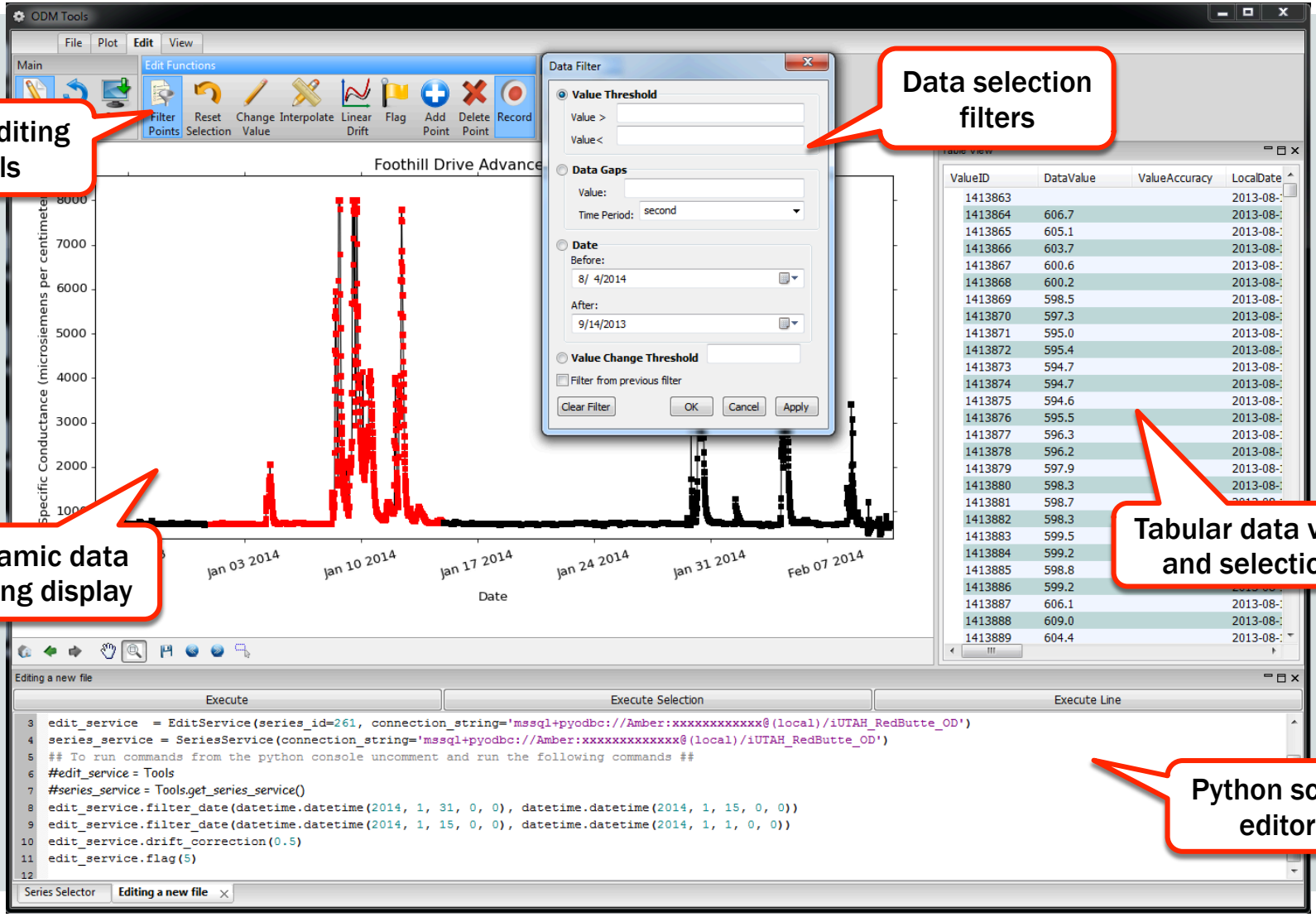
QAQC:
Data Visualization and Management







QAQC: Post Processing



Data editing tools

Data selection filters

Dynamic data editing display

Tabular data view and selection

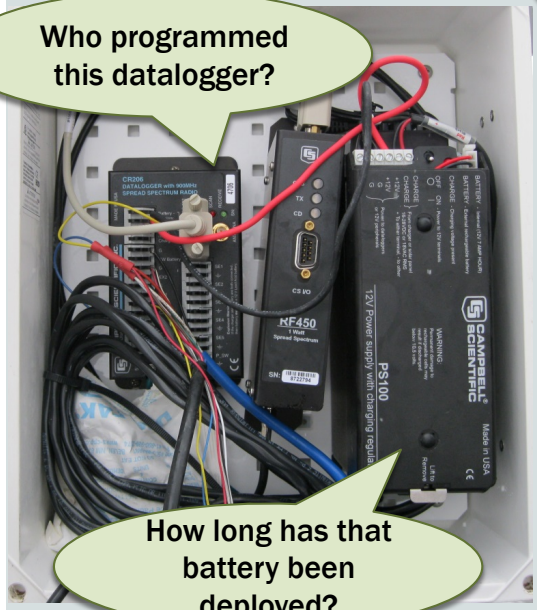
Python script editor

Monitoring Equipment Management

When was the last time we cleaned this sonde?

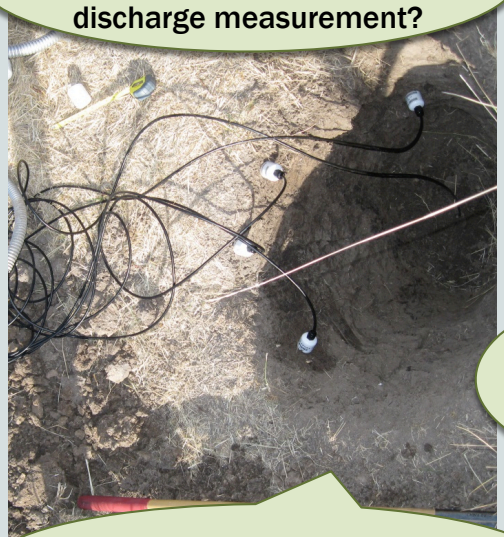
What were the field conditions of our discharge measurement?

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



Who programmed this datalogger?

How long has that battery been deployed?

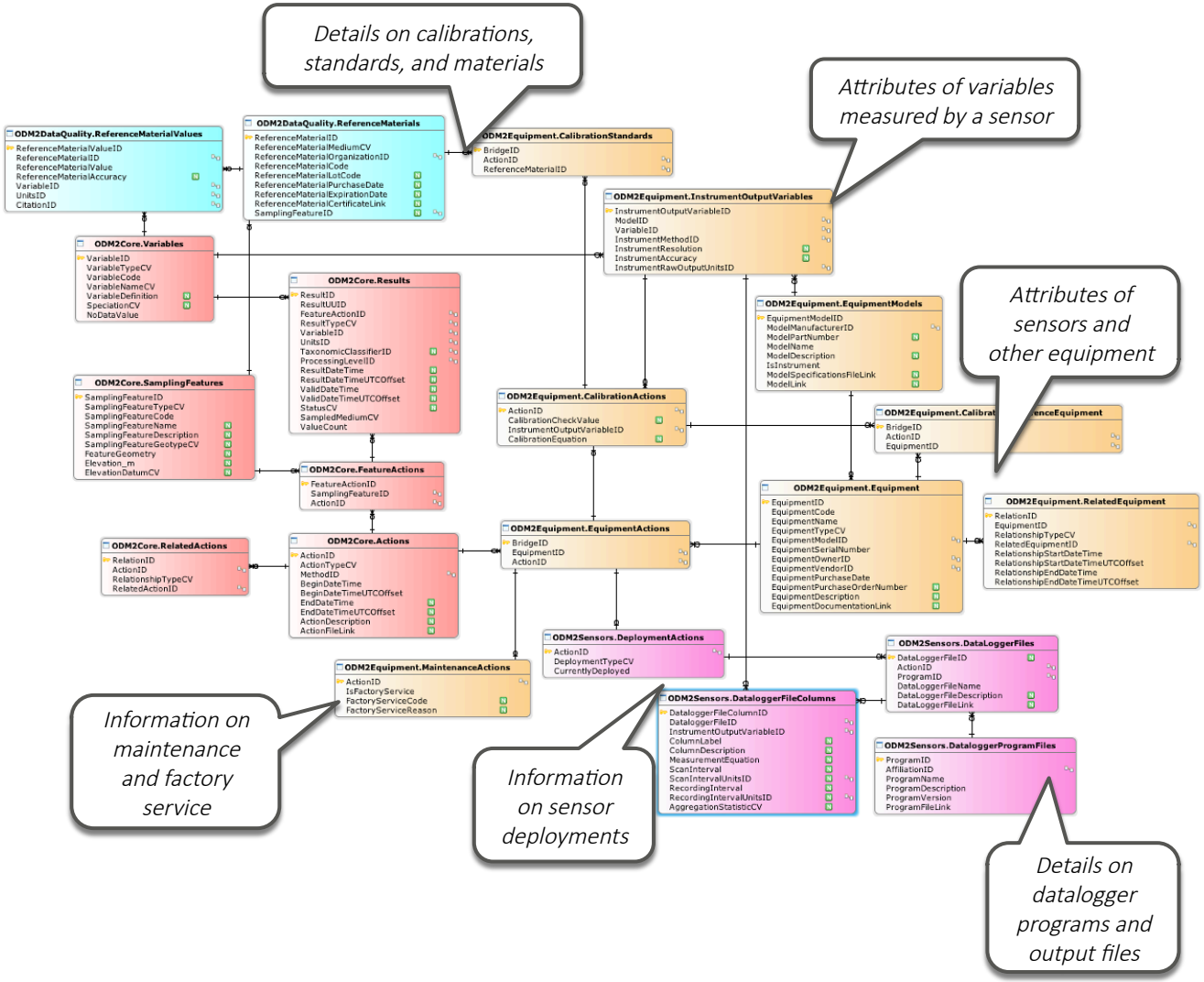


Which soil moisture sensor is buried at 40 cm?

Has this turbidity sensor been serviced at the factory?

What is this dissolved oxygen sensor's calibration history?

Monitoring Equipment Management



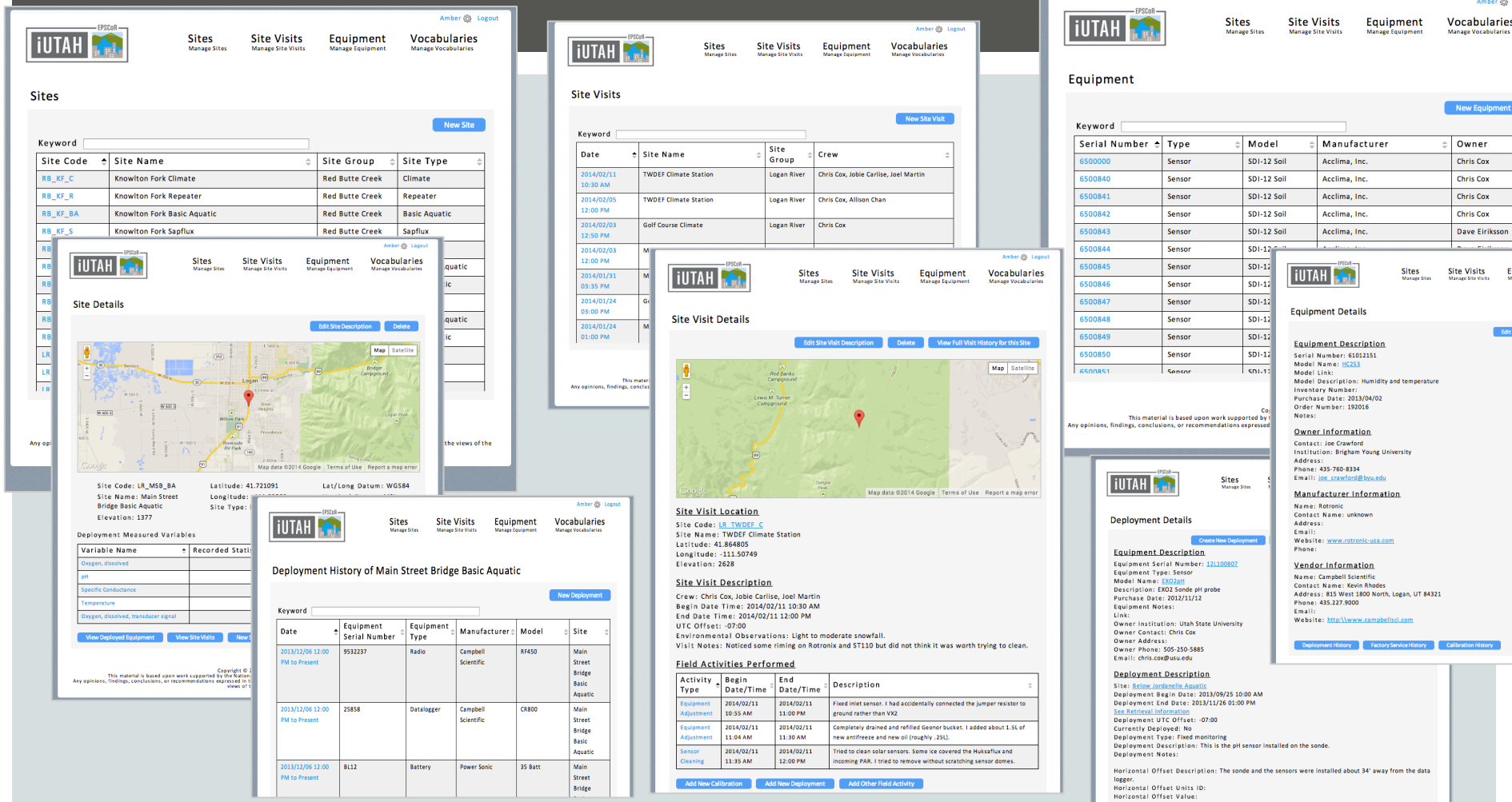
Data Model developed to relate:

- Equipment
- Field Activities
- Deployments
- Calibrations
- Measured Variables
- Datalogger Programs

Database serves as underlying structure to web interface.

EPSCoR iUTAH

Monitoring Equipment Management



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

Monitoring Equipment Management

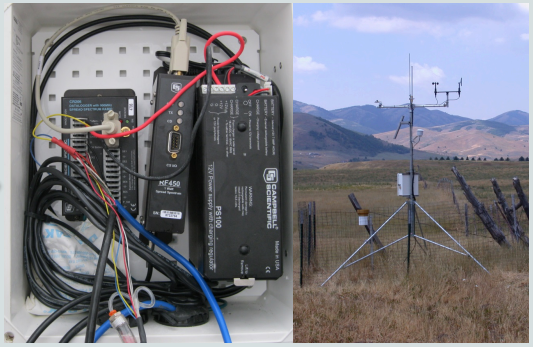
Equipment

- Model
- Serial number
- Owner
- Vendor
- Manufacturer
- Service history

Calibrations

- Method
- Standard

Deployments



- Deployment type
- Description
- Dates
- Offsets

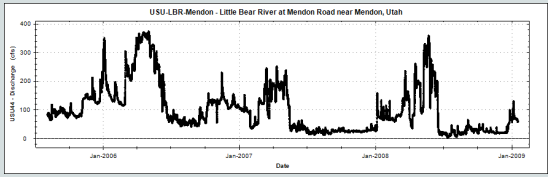
Site Visits

- Location
- Date
- People
- Conditions

Field Activities

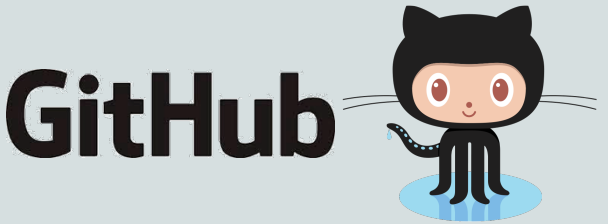
- Activity type
- Description
- Date

Time Series Observations



Open Source Code Repositories

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



Summary

- Researchers are using *in situ* sensors at multiple sites more commonly resulting in large datasets
- Data enhanced by metadata standards and descriptions
- 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
- 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

Acknowledgements

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.

iUTAH Watershed Technicians:

Chris Cox
Joe Crawford
Dave Eiriksson

UWRL Student Programmers:

Juan Caraballo
Jacob Meline
Mario Matos
Maurier Ramirez

Utah State University

Amber Jones
amber.jones@usu.edu

Jeff Horsburgh
jeff.horsburgh@usu.edu

Stephanie Reeder
stephanie.reeder@usu.edu

http://data.iutahepscor.org