

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

Jeffery S. Horsburgh

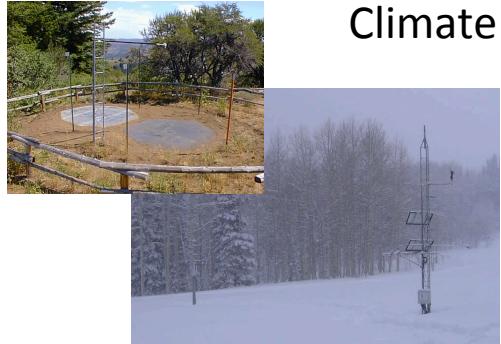
Amber Spackman Jones  
Stephanie Reeder



# 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



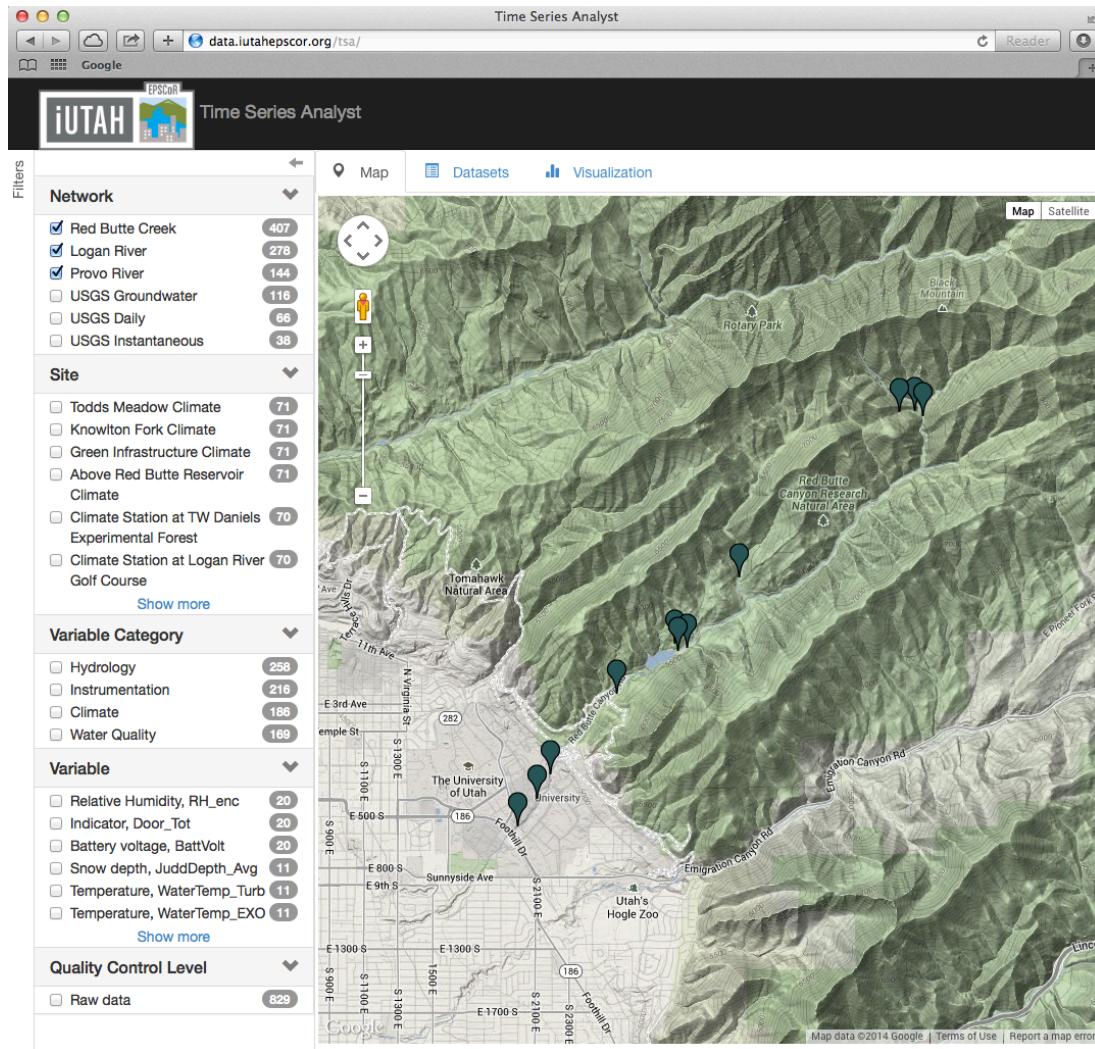
Groundwater



Mobile Platforms



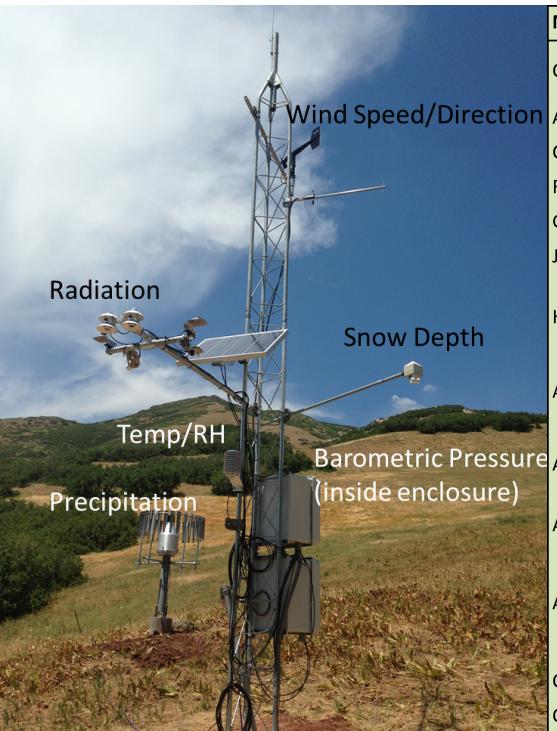
# Gradients Along Mountain To Urban Transitions (GAMUT) Network



- 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 *in situ* and re-locatable sensors

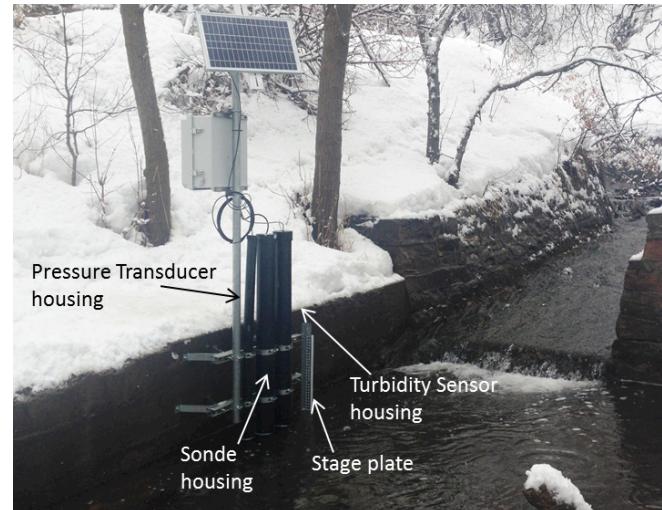
# Gradients Along Mountain to Urban Transitions (GAMUT) Network

## Climate/Terrestrial Sites



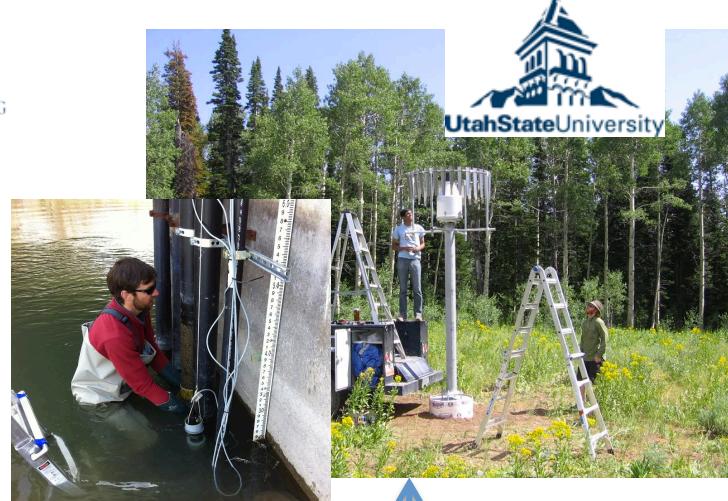
Manufacturer	Instrument	Variables Measured
Campbell	HC2S3	Air Temperature and Relative Humidity
Apogee	ST110	Air Temperature
Campbell	CS106	Barometric Pressure
RM Young	5303	Wind Speed/Direction
Geonor	TB-200	Precipitation
Judd	Depth Sensor	Snow Depth
Hukseflux	NR01	Incoming and Outgoing Shortwave and 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

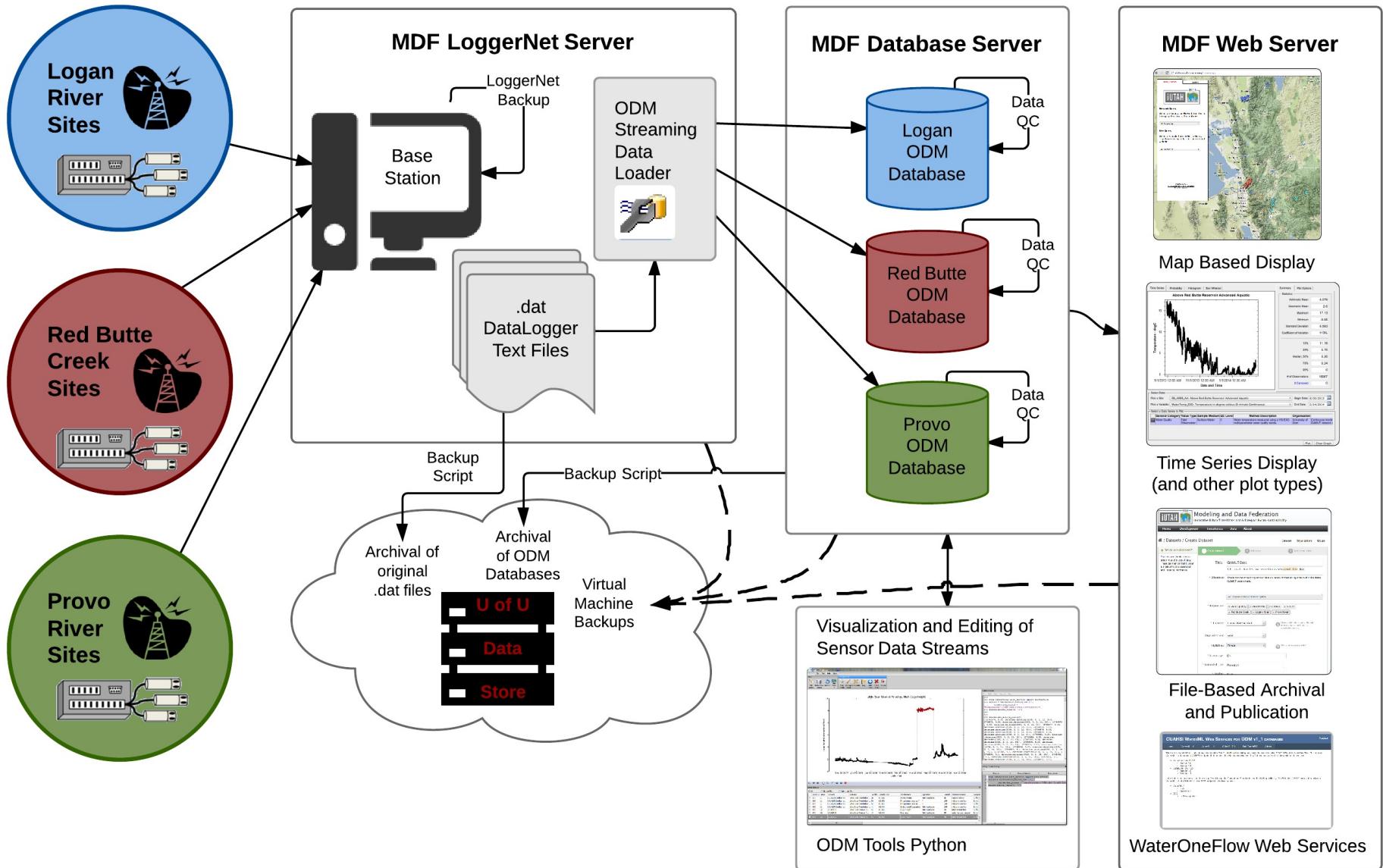


Manufacturer	Instrument	Variables Measured
YSI	599100-01	Dissolved Oxygen
YSI	599870-01	Specific Conductivity and Water Temperature
YSI	599795-02	pH
YSI	599101-01	Fluorescent Dissolved Organic Matter (fDOM)
YSI	599102-01	Blue Green Algae and Chlorophyll <i>a</i>
Campbell	CS451	Water Depth and Water Temperature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

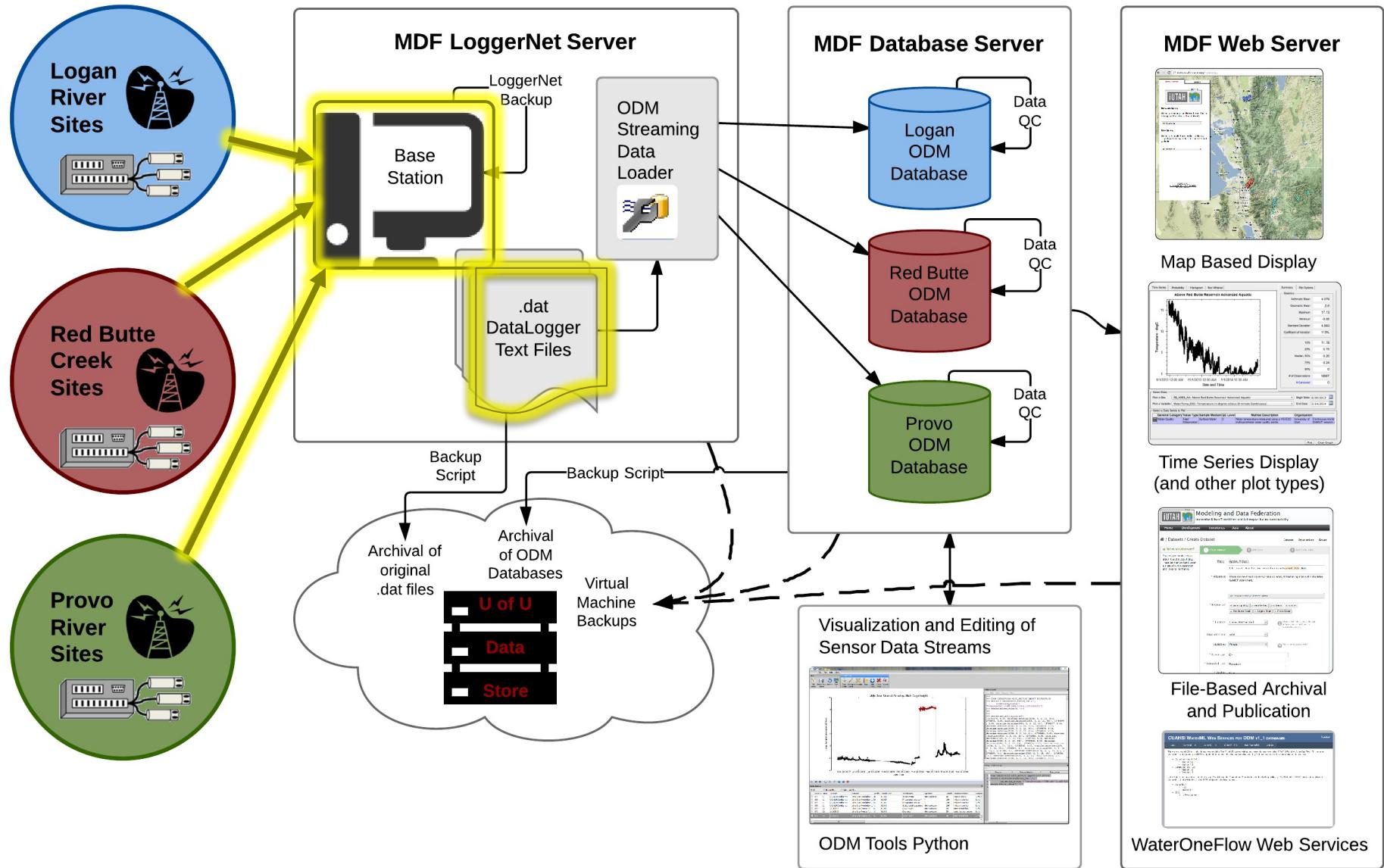
# Gradients Along Mountain to Urban Transitions (GAMUT) Network



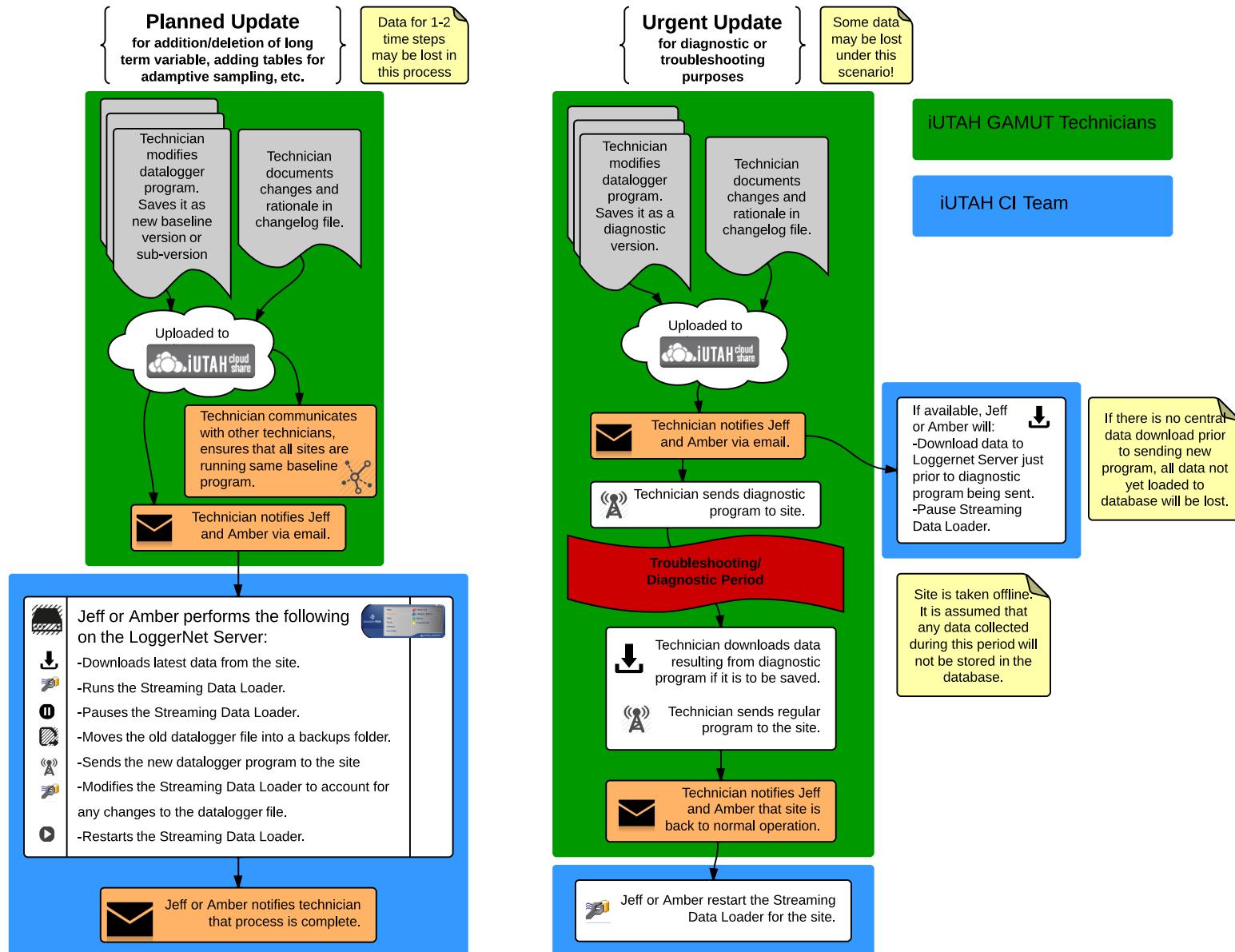
# GAMUT Data Workflow



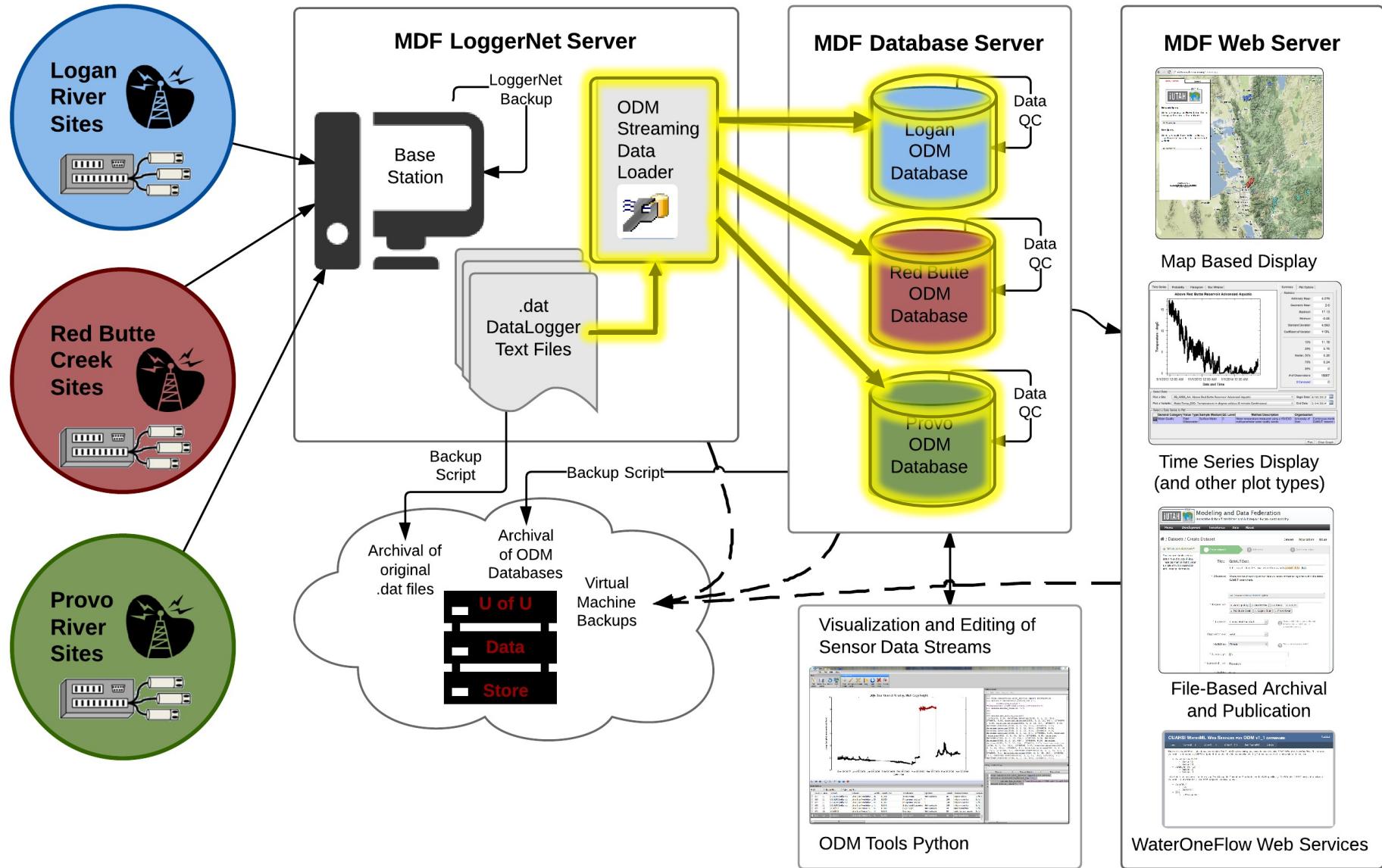
# Sensor Data Acquisition



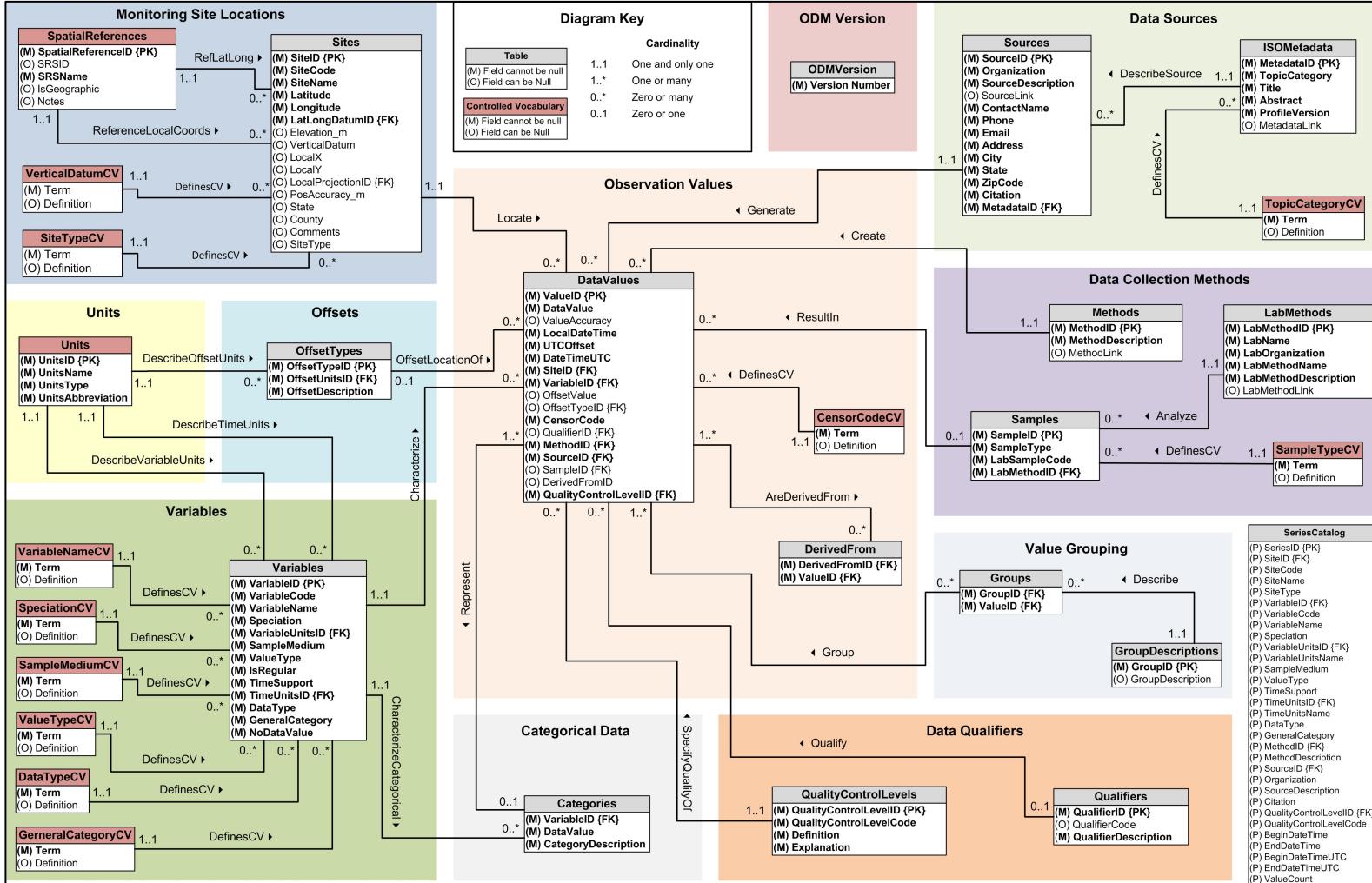
# Datalogger Program Updates



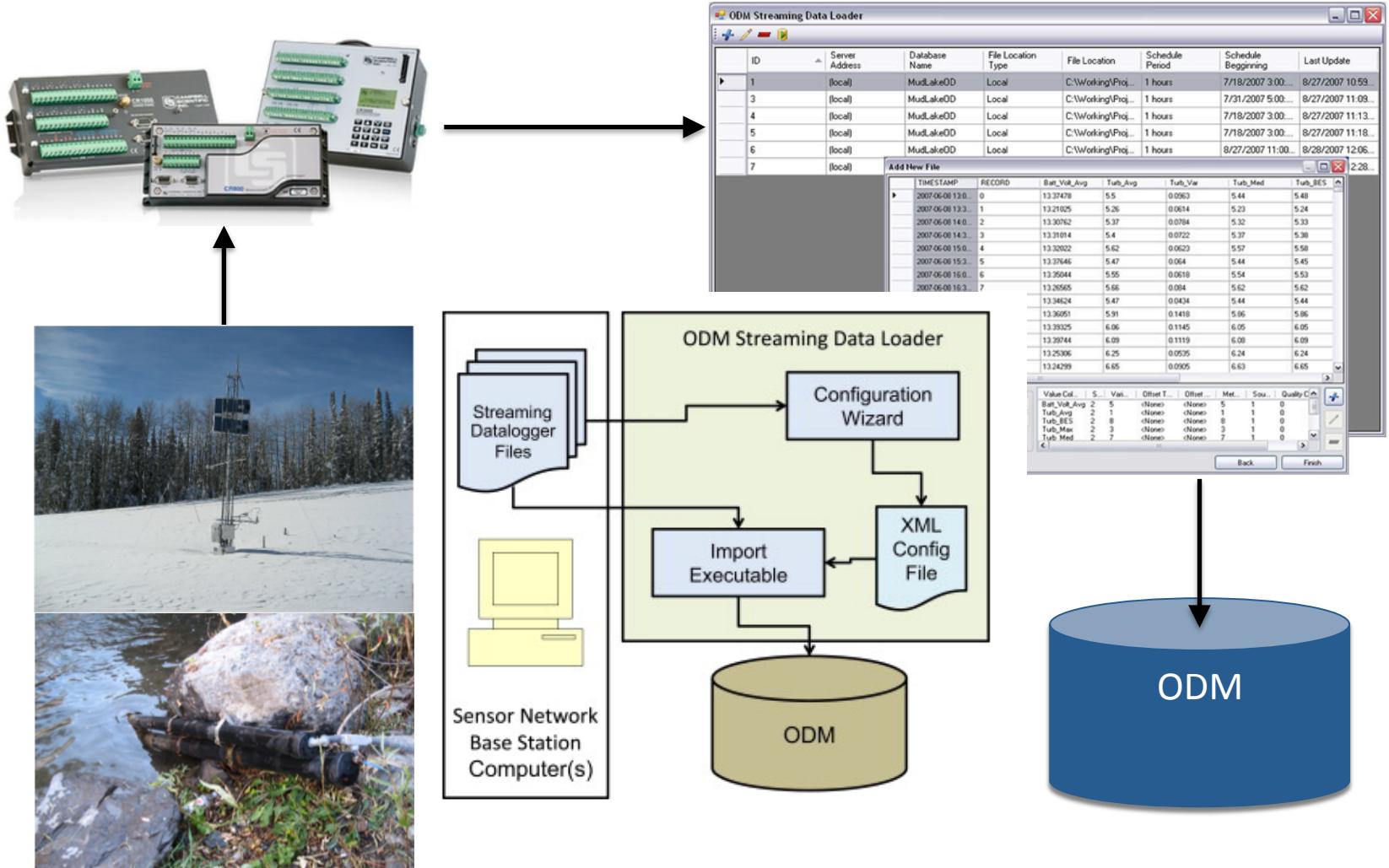
# Data Loading and Storage



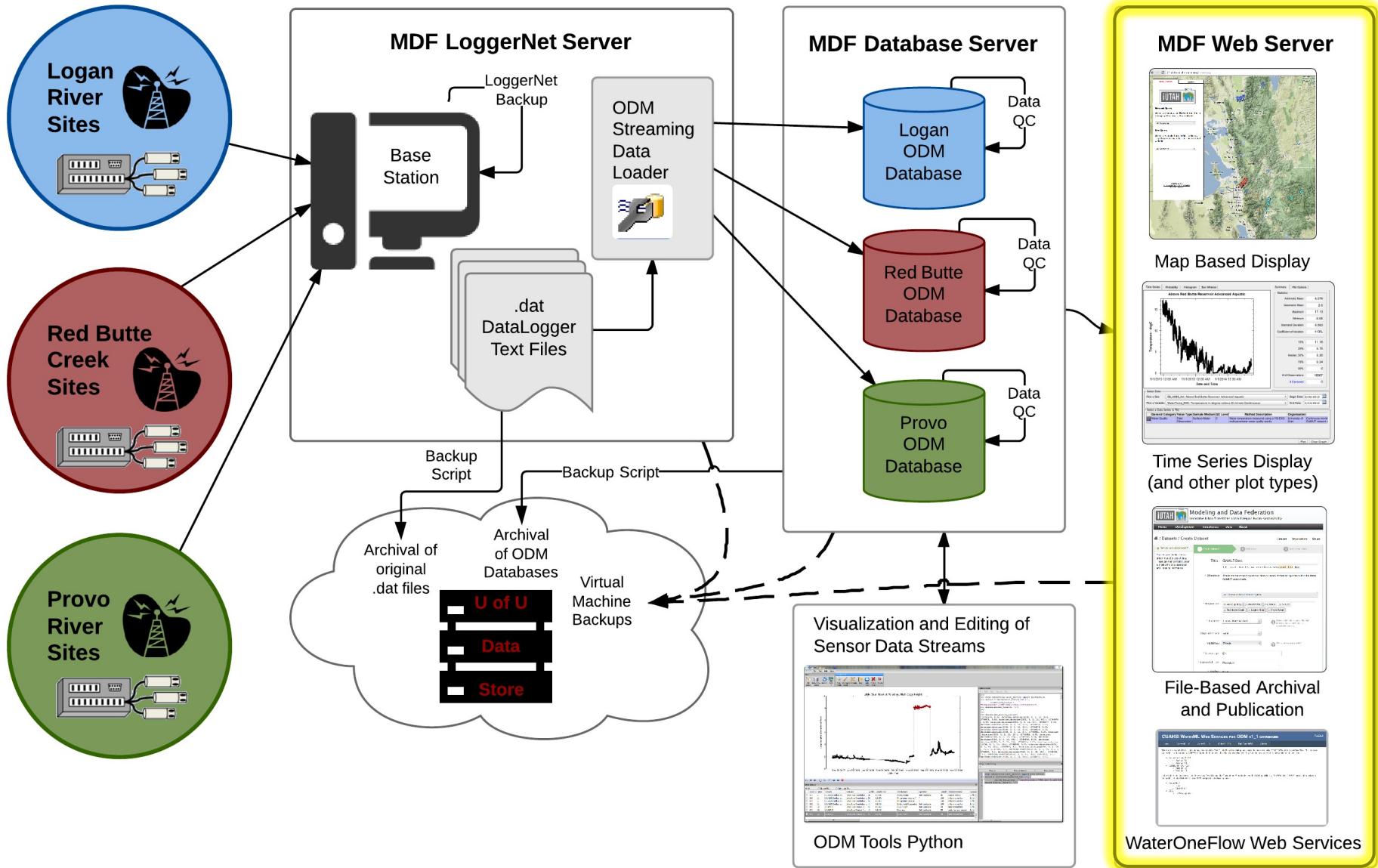
# Data Loading and Storage



# Data Loading and Storage

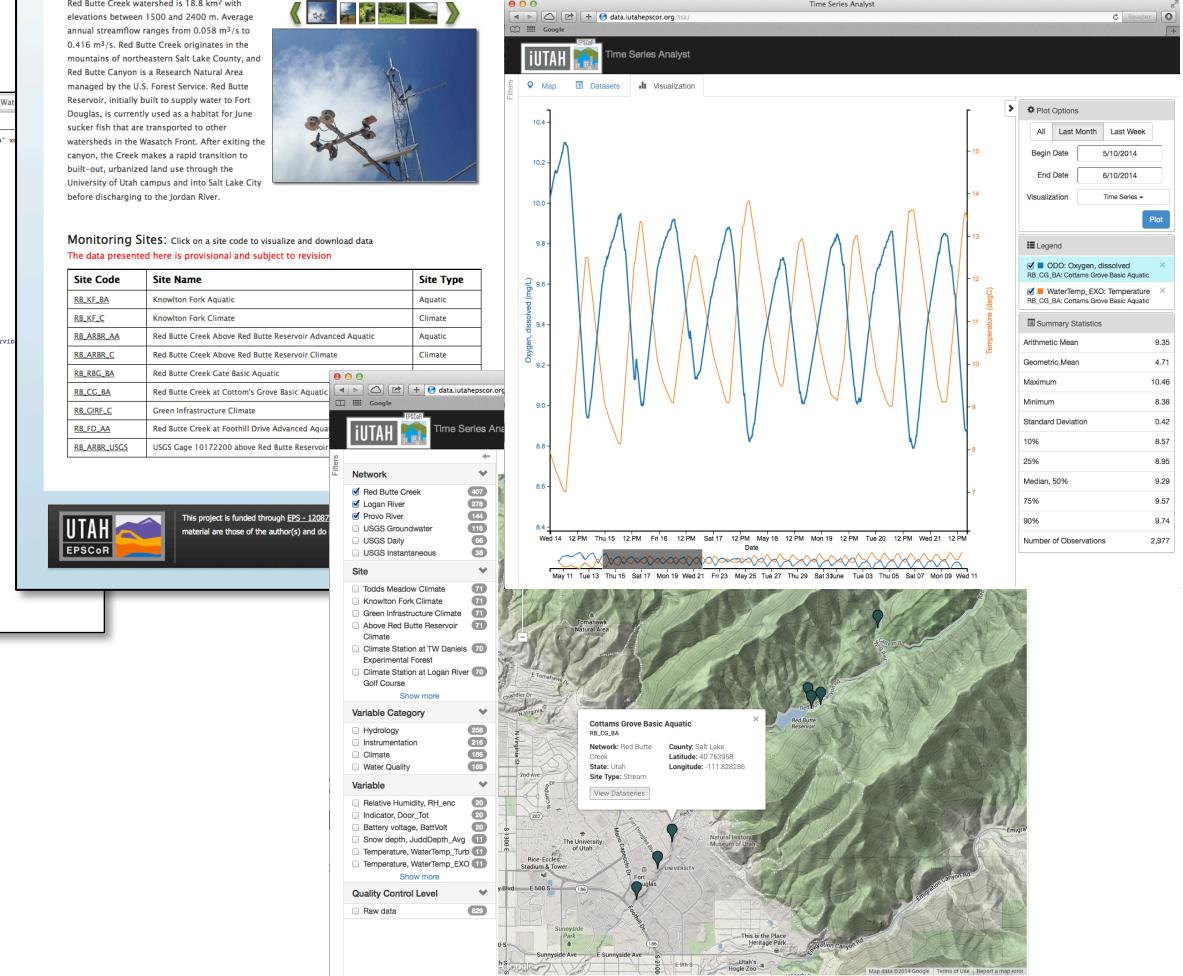


# Web-Based Data Access



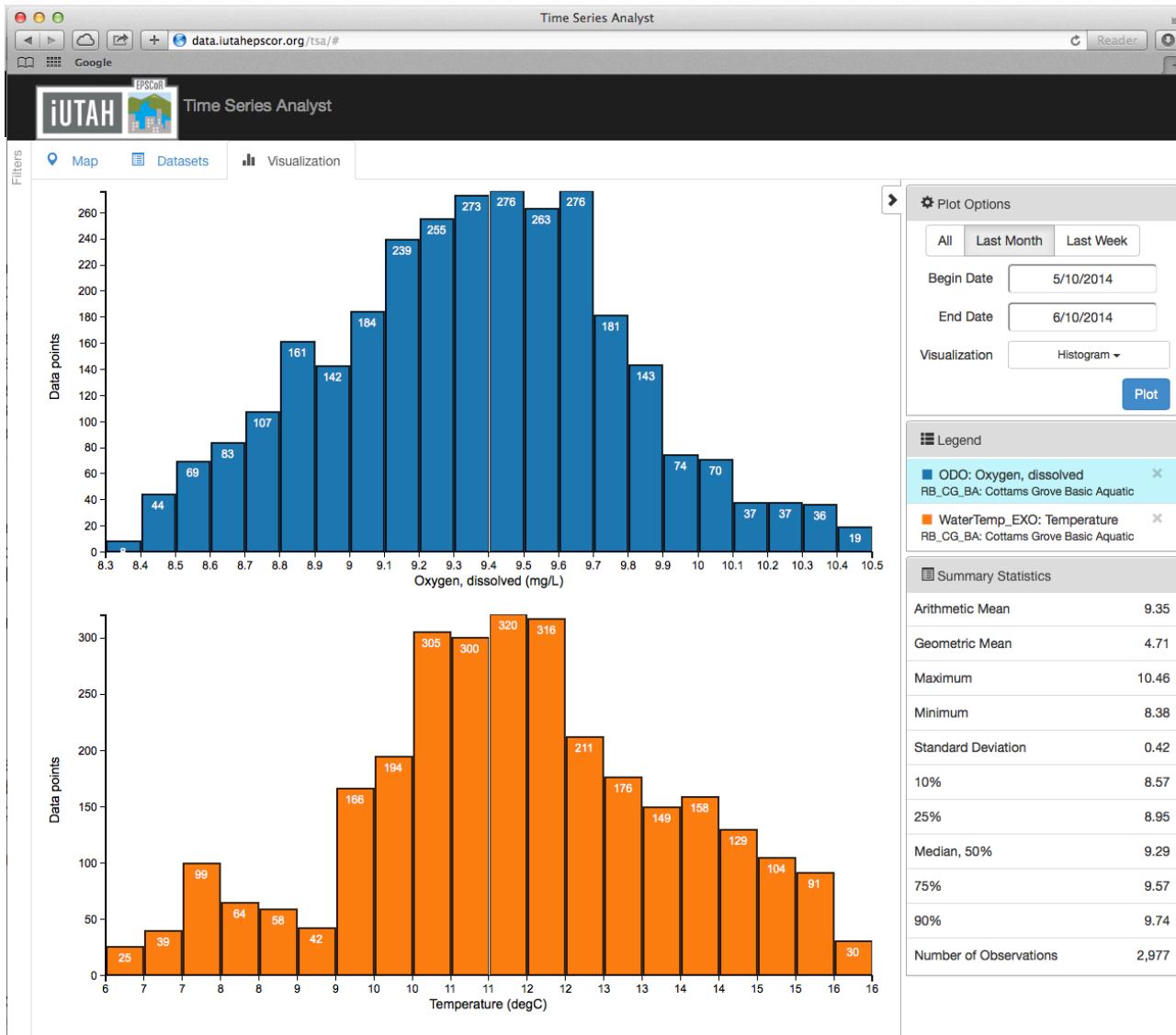
# Web-based Data Access

 EPSCoR  
**iUTAH**  Modeling and Data Federation  
innovative Urban Transitions and Aridregion Hydro-sustainability

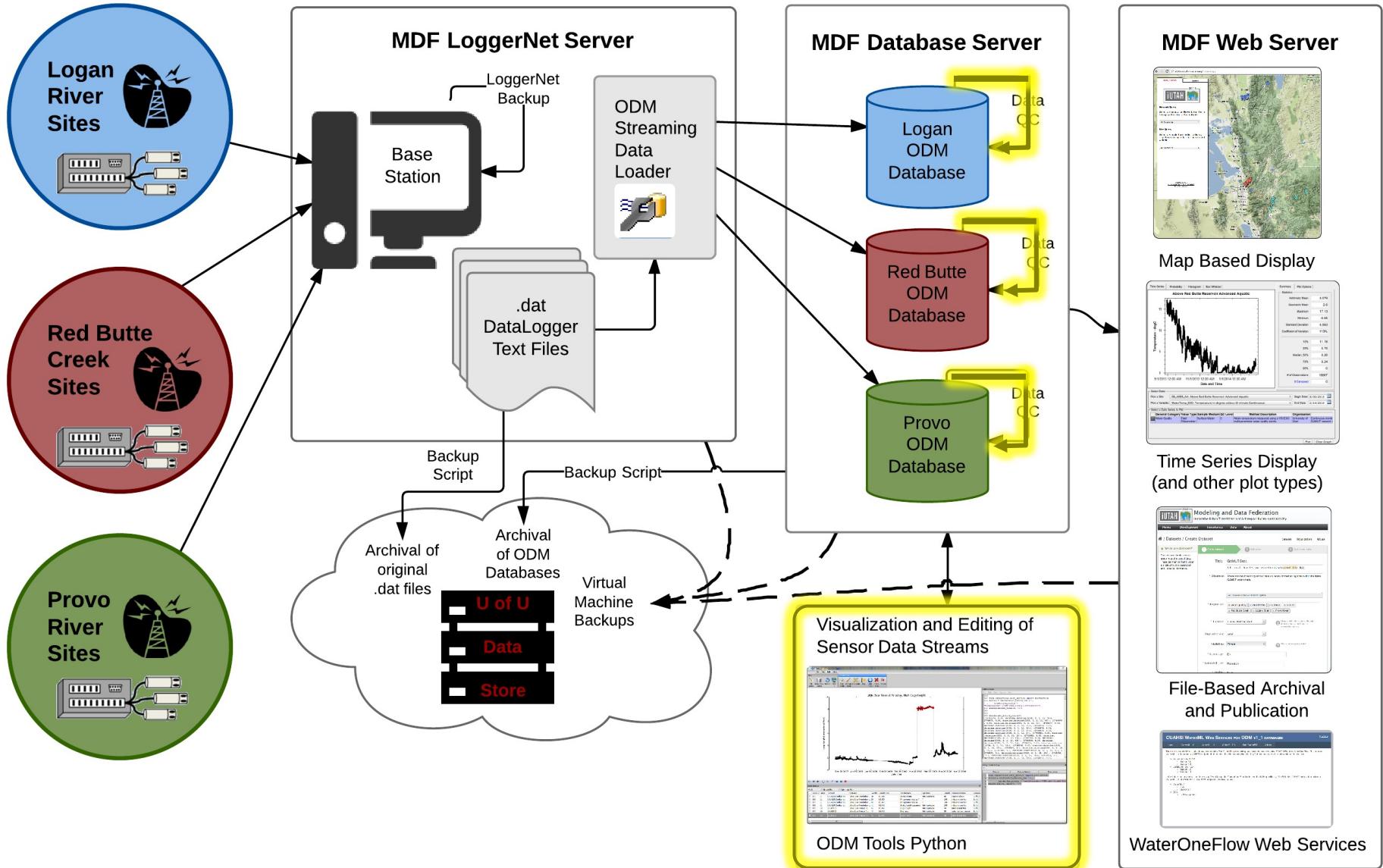


- CUAHSI  
HIS  
*Sharing hydrologic data*

# Enhanced Web-Based Time Series Data Access and Visualization



# Sensor Data QAQC



# Development of a QAQC Plan

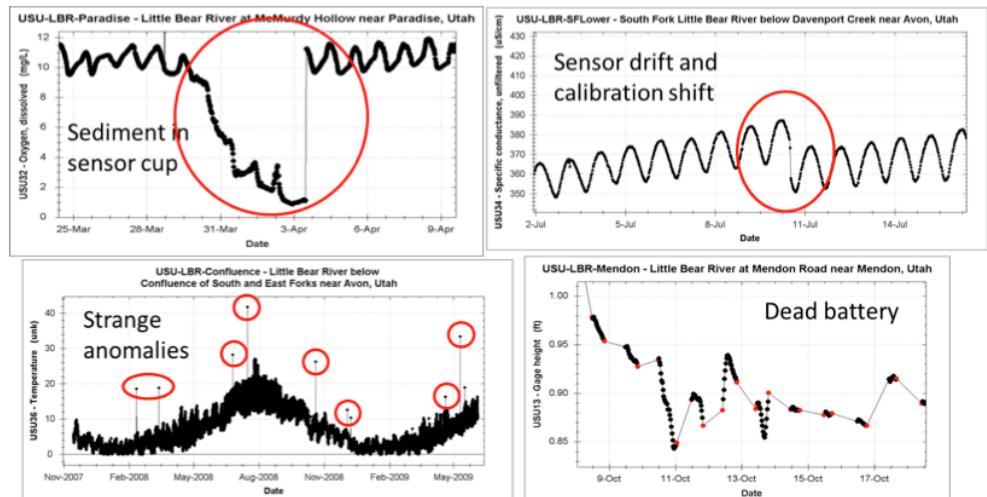
## Quality Assurance:

“...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

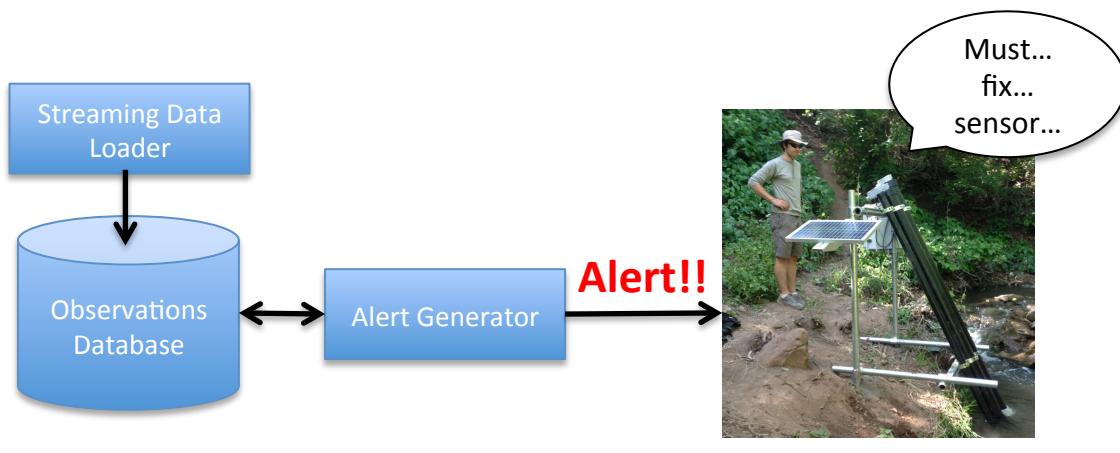


# QAQC: Automated Alerts

Technicians receive email alerts daily

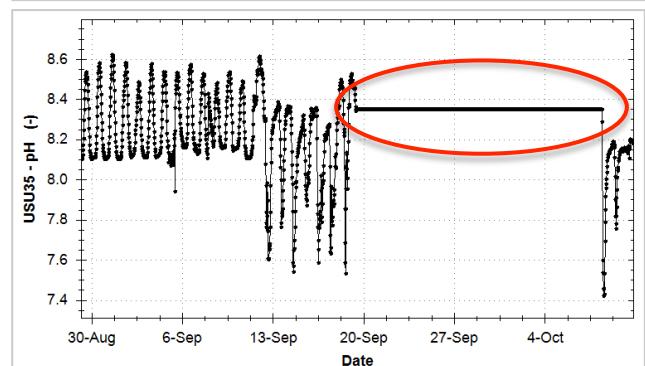
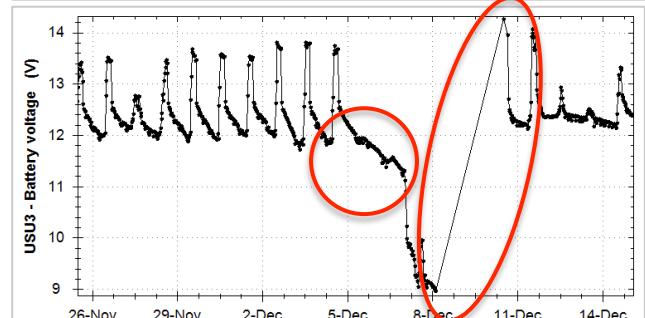
1. **Power:** batter voltage < 12 volts
2. **Persistence:** value of a variable is unchanging
3. **Updates:** data are not being reported
4. **NANs:** sensor is reporting “NaN” values

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 have repeated values in the past day.						
SiteID	SiteCode	VariableID	VariableCode	DataValue	Start	End
3	LR_MainStreet	61	ODD_Sat	84.90	2014-02-17 21:15	2014-02-18 01:15
4	LR_TWDEF_C	9	Precip_Tot	9.74	2014-02-17 18:00	2014-02-18 00:15
4	LR_TWDEF_C	13	SWIn_NR01_Avg	-4.68	2014-02-17 20:30	2014-02-18 00:30
4	LR_TWDEF_C	14	SWOUT_NR01_Avg	.00	2014-02-17 18:30	2014-02-18 04:00
4	LR_TWDEF_C	24	PARIn_NR01_Avg	.00	2014-02-17 18:30	2014-02-18 04:00
4	LR_TWDEF_C	25	PAROut_Avg	.00	2014-02-17 18:45	2014-02-18 04:00
4	LR_TWDEF_C	41	SoilCond_20cm_Avg	.00	2014-02-17 06:15	2014-02-18 04:00
4	LR_TWDEF_C	45	SoilCond_50cm_Avg	.00	2014-02-17 06:15	2014-02-18 04:00
4	LR_TWDEF_C	48	SoilTemp_100cm_Avg	-2.05	2014-02-17 09:15	2014-02-17 13:15
4	LR_TWDEF_C	49	SoilCond_100cm_Avg	.00	2014-02-17 06:15	2014-02-18 04:00
4	LR_TWDEF_C	91	Precip_HrDiff	.00	2014-02-17 19:00	2014-02-18 01:00
5	LR_GC_C	5	BP_Avg	86.00	2014-02-17 21:45	2014-02-18 01:45
5	LR_GC_C	9	Precip_Tot	-.9999.00	2014-02-17 06:15	2014-02-18 04:45
5	LR_GC_C	45	SoilCond_50cm_Avg	.00	2014-02-17 06:15	2014-02-18 04:45
5	LR_GC_C	91	Precip_HrDiff	.00	2014-02-17 06:15	2014-02-18 04:45

(15 rows affected)



# QAQC:

# Data Visualization and Management

Multiple Plot Types

Plot Display Options

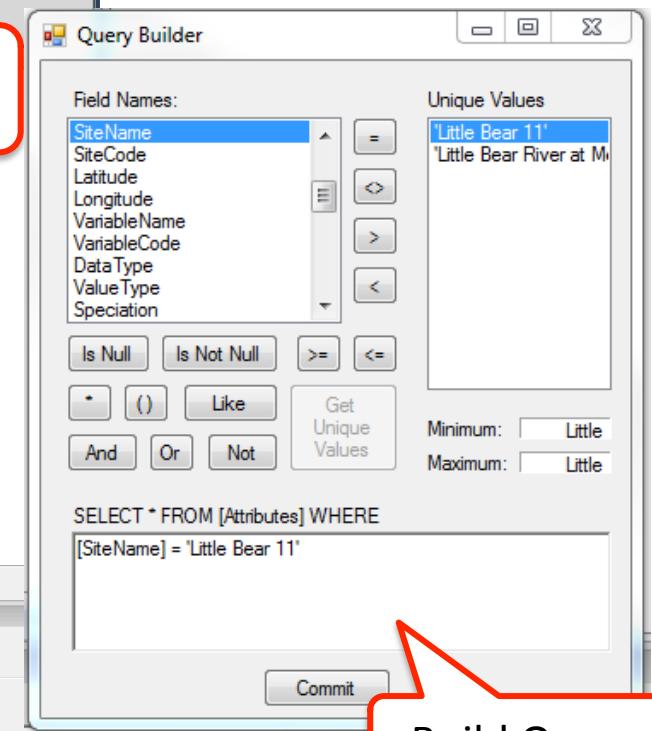
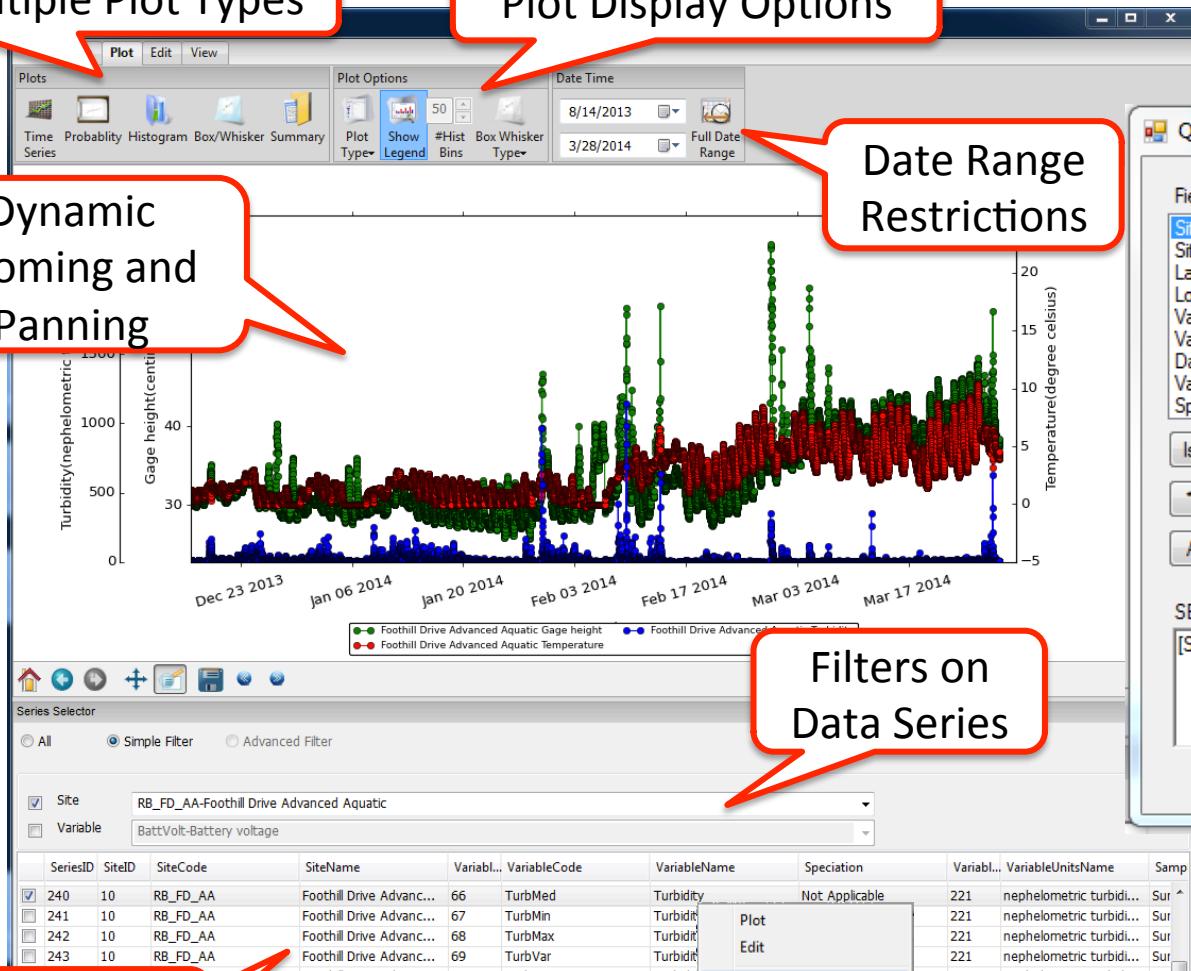
Dynamic  
Zooming and  
Panning

Date Range  
Restrictions

Filters on  
Data Series

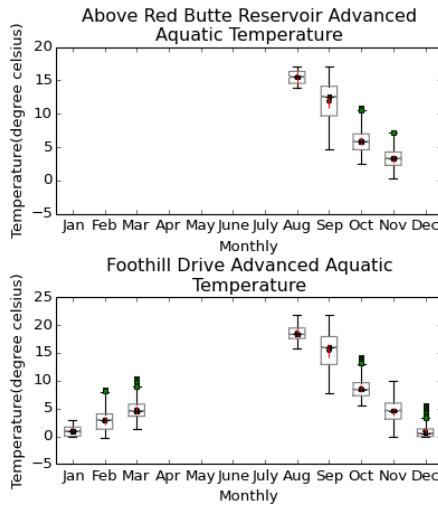
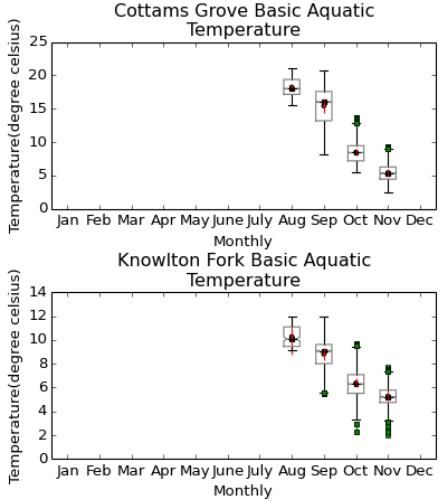
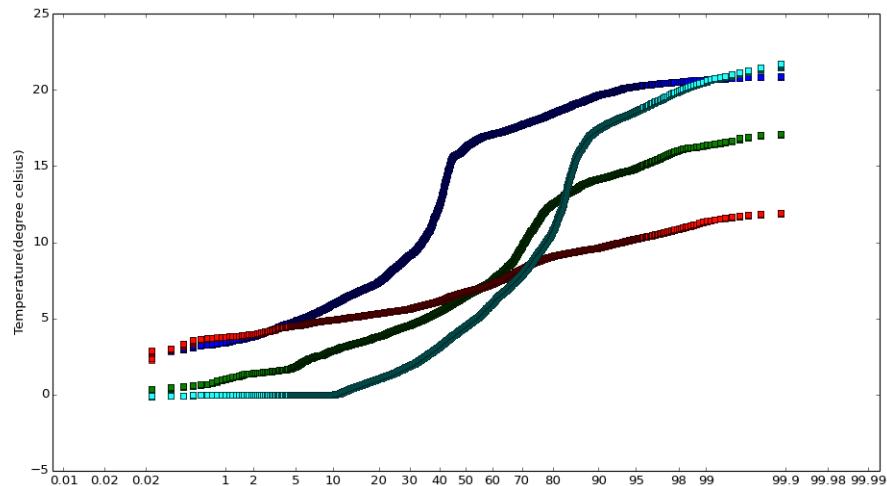
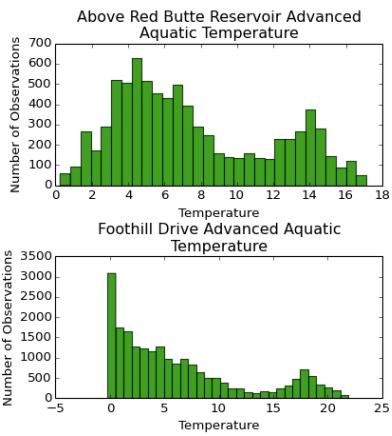
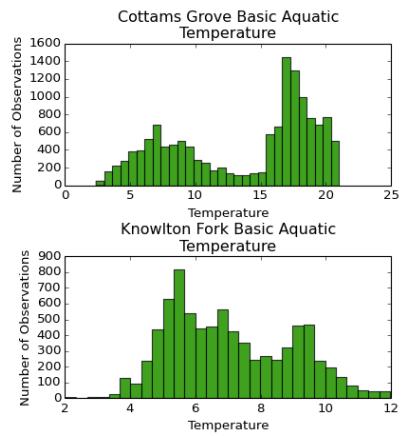
Time Series  
Selection

Export Data Series



# QAQC:

# Data Visualization and Management



# QAQC: Post Processing

Dynamic Data Editing Display

Data Editing Tools

Python Code Console

Python Script Editor

Little Bear River at Paradise, Utah Gage height

Gage height(international foot)

Date Time

Main

File Plot Edit View

Edit Functions

Series

Derive New Series

Restore Save

Filter Points Change Interpolate Flag Add Point Delete Point Record

Python Console

File Edit View Options Help

```
>>> from odmservices.edit_service import EditService
>>> series = EditService(series_id= 171,
...     connection_string =
"mssql+pyodbc://ODM:odm@(local)/LittleBear11")
>>> series.filter_value(8, '>')
>>>
>>> series.get_active_points()
[(2766974, 9.04, datetime.datetime(2008, 3, 2, 11, 30)), (2766975, 9.04, datetime.datetime(2008, 3, 2, 11, 45)), (2766976, 9.04, datetime.datetime(2008, 3, 2, 12, 0)), (2766977, 9.04, datetime.datetime(2008, 3, 2, 12, 15)), (2766978, 9.04, datetime.datetime(2008, 3, 2, 12, 30)), (2766979, 9.04, datetime.datetime(2008, 3, 2, 12, 45)), (2766980, 9.04, datetime.datetime(2008, 3, 2, 13, 0)), (2766981, 9.04, datetime.datetime(2008, 3, 2, 13, 15)), (2766982, 9.05, datetime.datetime(2008, 3, 2, 13, 30)), (2766983, 9.06, datetime.datetime(2008, 3, 2, 13, 45)), (2766984, 9.06, datetime.datetime(2008, 3, 2, 14, 0)), (2766985, 9.08, datetime.datetime(2008, 3, 2, 14, 15)), (2766986, 9.09, datetime.datetime(2008, 3, 2, 14, 30)), (2766987, 9.1, datetime.datetime(2008, 3, 2, 14, 45)), (2766988, 9.1, datetime.datetime(2008, 3, 2, 15, 0)), (2766989, 9.1, datetime.datetime(2008, 3, 2, 15, 15)), (2766990, 9.1, datetime.datetime(2008, 3, 2, 15, 30)), (2766991, 9.1, datetime.datetime(2008, 3, 2, 15, 45)), (2766992, 9.09,
```

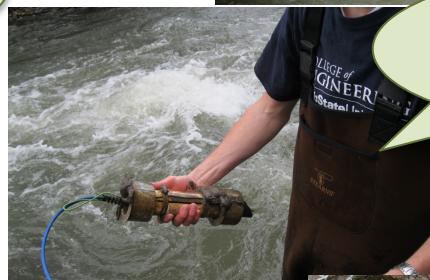
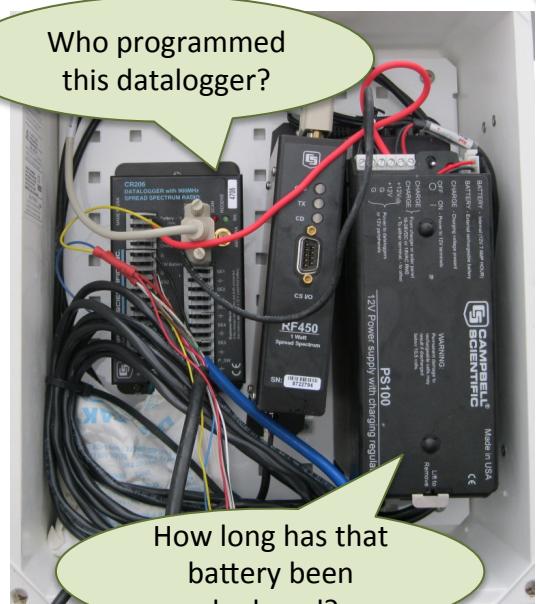
Editing: Script Test.py

Execute Execute Selection Execute Line

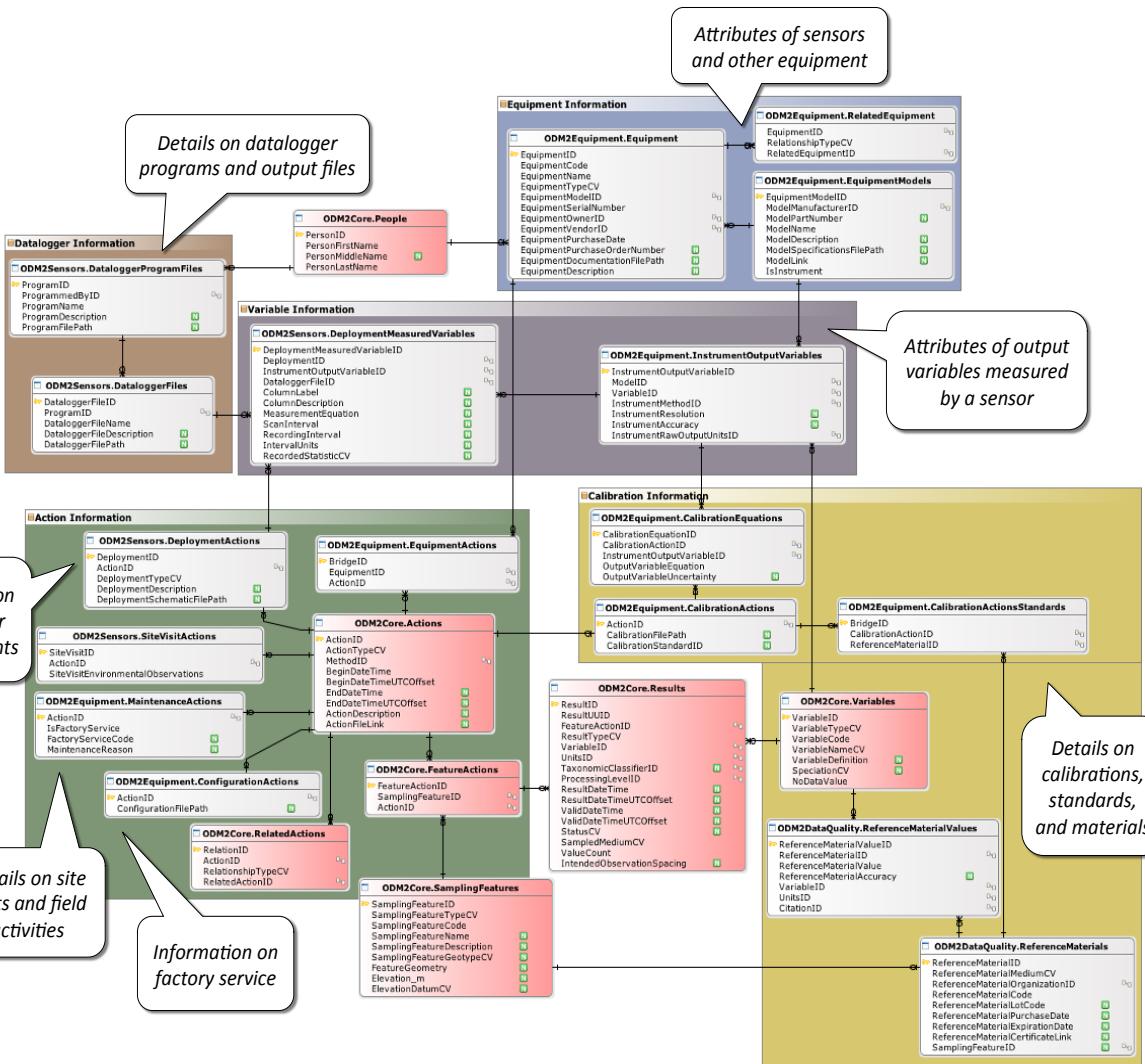
```
1 from odmservices.edit_service import EditService
2 series = EditService(series_id= 171,
3     connection_string =
"mssql+pyodbc://ODM:odm@(local)/LittleBear11")
4 series.filter_value(8, '>')
```

SeriesID	SiteID	SiteCode	SiteName	Variable...	VariableCode	VariableName	Speciation	Variable...	VariableUnitsName	SampleM
165	11	USU-LBR-Confluence	Little Bear River below...	36	USU36	Temperature	Not Applicable	96	degree celsius	Surfa
166	11	USU-LBR-Confluence	Little Bear River below...	39	USU39	Phosphorus, total as P	P	199	milligrams per liter	Surfa
167	11	USU-LBR-Confluence	Little Bear River below...	40	USU40	Phosphorus, total as ...	P	199	milligrams per liter	Surfa
168	11	USU-LBR-Confluence	Little Bear River below...	41	USU41	Solids, total Suspended	Not Applicable	199	milligrams per liter	Surfa
169	12	10105900	Little Bear River at Pa...	42	USU42	Gage height	Not Applicable	48	international foot	Surfa
170	12	10105900	Little Bear River at Pa...	43	USU43	Discharge	Not Applicable	35	cubic feet per second	Surfa
171	12	10105900	Little Bear River at Pa...	42	USU42	Gage height	Not Applicable	48	international foot	Surfa

# Monitoring Equipment Management



# 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

**Sites**

**Site Visits**

**Equipment**

**Vocabularies**

**Amber Logout**

**iUTAH EPSOil**

**Sites** Manage Sites   **Site Visits** Manage Site Visits   **Equipment** Manage Equipment   **Vocabularies** Manage Vocabularies

**New Site**

**Keyword**

Site Code	Site Name	Site Group	Site Type
RB_KF_C	Knowlton Fork Climate	Red Butte Creek	Climate
RB_KF_R	Knowlton Fork Repeater	Red Butte Creek	Repeater
RB_KF_BA	Knowlton Fork Basic Aquatic	Red Butte Creek	Basic Aquatic
RB_KF_S	Knowlton Fork Sapflux	Red Butte Creek	Sapflux

**Site Details**

**Site Name:** Main Street Bridge Basic Aquatic

**Latitude:** 41.721091   **Longitude:** -111.50749   **Lat/Long Datum:** WGS84

**Elevation:** 1377

**Deployment Measured Variables**

Variable Name	Recorded Status
Oxygen, dissolved	
pH	
Specific Conductance	
Temperature	
Oxygen, dissolved, transducer signal	

**New Deployment**

**View Deployed Equipment**   **View Site Visits**   **New**

This material is based upon work supported by the National Science Foundation under Grant No. 0953237. Any opinions, findings, 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.

**Site Visits**

**Amber Logout**

**iUTAH EPSOil**

**Sites** Manage Sites   **Site Visits** Manage Site Visits   **Equipment** Manage Equipment   **Vocabularies** Manage Vocabularies

**New Site Visit**

**Keyword**

Date	Site Name	Site Group	Crew
2014/02/11 10:30 AM	TWDEF Climate Station	Logan River	Chris Cox, Jobie Carlise, Joel Martin
2014/02/05 12:00 PM	TWDEF Climate Station	Logan River	Chris Cox, Allison Chan
2014/02/03 12:50 PM	Golf Course Climate	Logan River	Chris Cox
2014/02/03 12:00 PM	Miller	Miller	
2014/01/21 03:35 PM	Miller	Miller	
2014/01/24 03:00 PM	Gruber	Gruber	
2014/01/24 01:00 PM	Miller	Miller	

**Site Visit Details**

**Site Visit Location**

**Site Visit Description**

**Field Activities Performed**

**Deployment**

**Equipment Details**

**Deployment Details**

**Deployment History**

**Factory Service History**

**Calibration History**

**Manufacturer Information**

**Vendor Information**

**Deployment Description**

**Measured Variables**

**New Deployment**

**New Deployment Measured Variable**

**View Equipment's Deployment History**

**Go to Site Visit**

# Monitoring Equipment Management

## Equipment

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



## Site Visits



## Field Activities



- Activity type
- Description
- Date

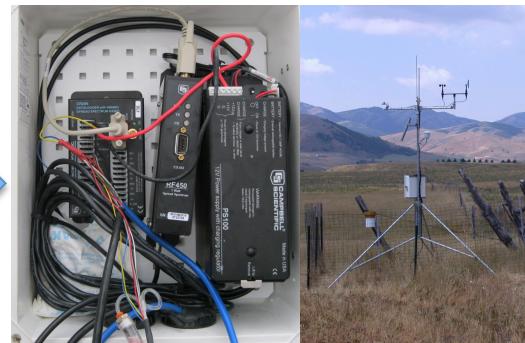
- Location
- Date
- People
- Conditions

## Calibrations



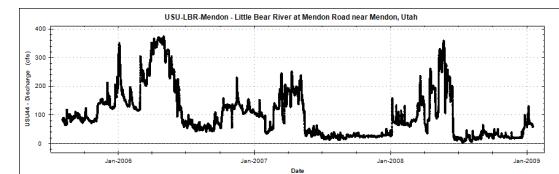
- Method
- Standard

## Deployments



- Deployment type
- Description
- Dates
- Offsets

## Time Series Observations



# 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



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.

## Acknowledgements

iUTAH Watershed Technicians:

Chris Cox

Joe Crawford

Dave Eiriksson

UWRL Student Programmers:

Juan Caraballo

Jacob Meline

Mario Matos

Maurier Ramirez

# Questions?

