



Support:  
EPS 1208732

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



Amber Spackman Jones  
Jeffery S. Horsburgh  
Stephanie Reeder

Mountain Observatories-Reno Nevada- July 17, 2014

Utah State University



# 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



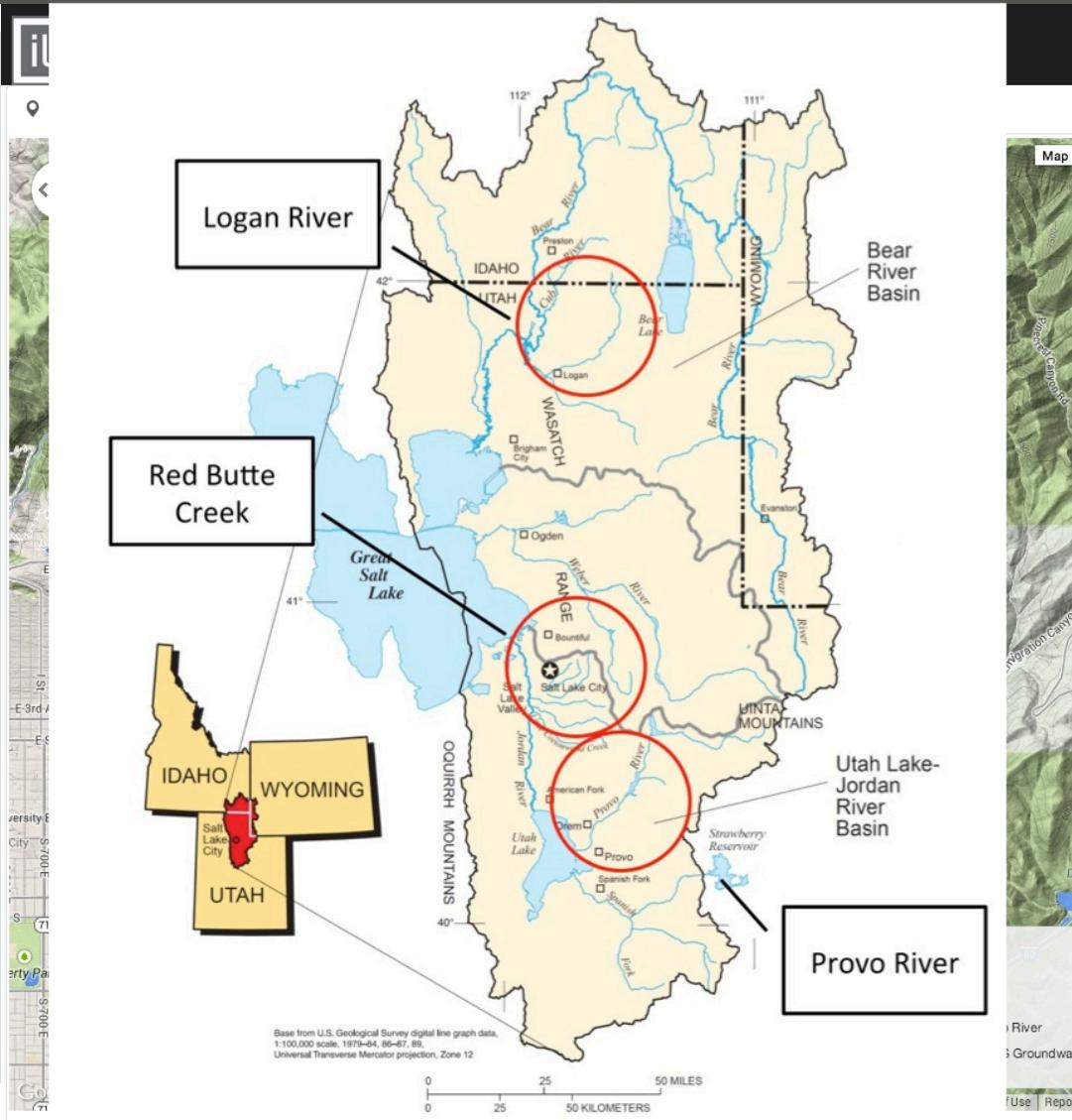
Groundwater



Mobile Platforms



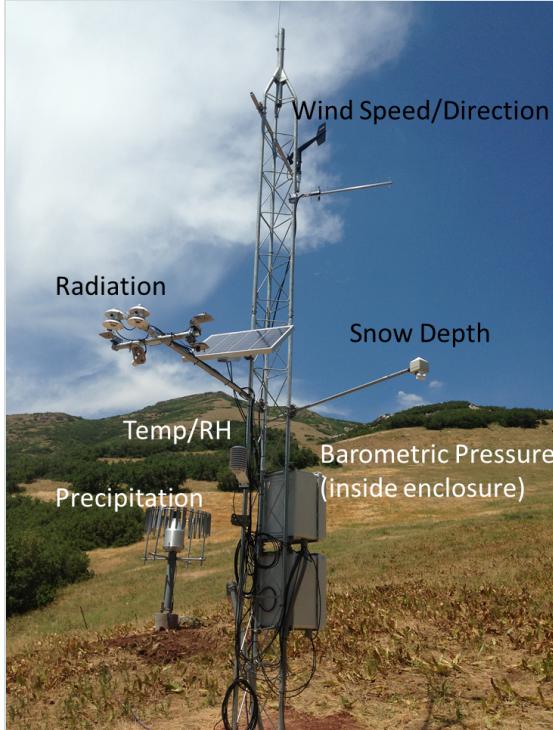
# Gradients Along Mountain To Urban Transitions (GAMUT) Network



- 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 relocatable sensors

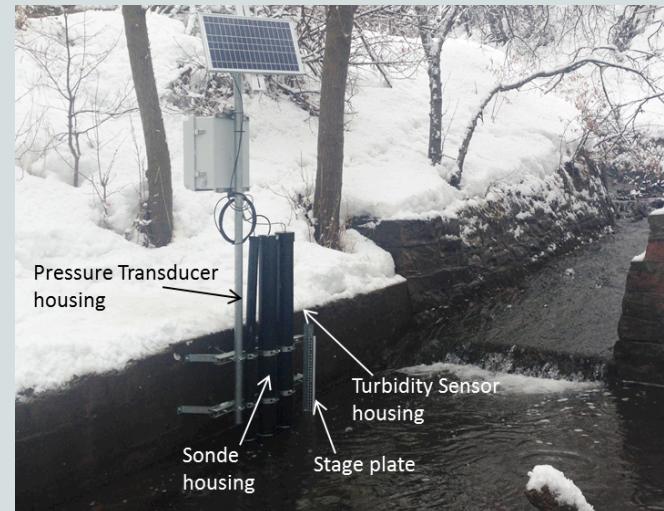
# Gradients Along Mountain to Urban Transitions (GAMUT) Network

## 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

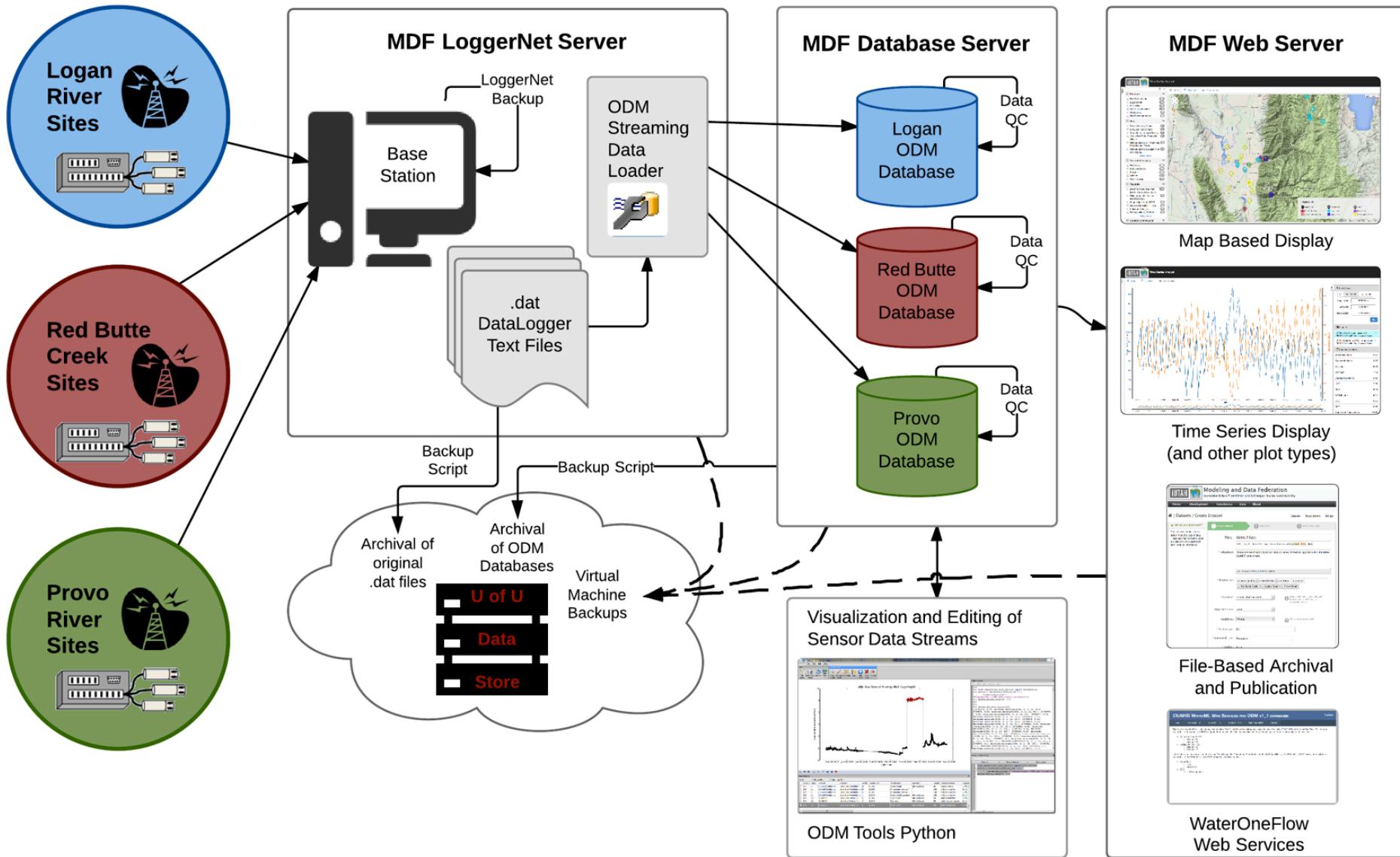


Manufacturer	Instrument	Variables Measured
YSI	599100-01	Dissolved Oxygen
YSI	599870-01	Specific Conductivity and Water Temperature
YSI	599795-02	pH
YSI	599101-01	Fluorescent Dissolved Organic Matter (fDOM)
YSI	599102-01	Blue Green Algae and Chlorophyll <i>a</i>
Campbell	CS451	Water Depth and Water Temperature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

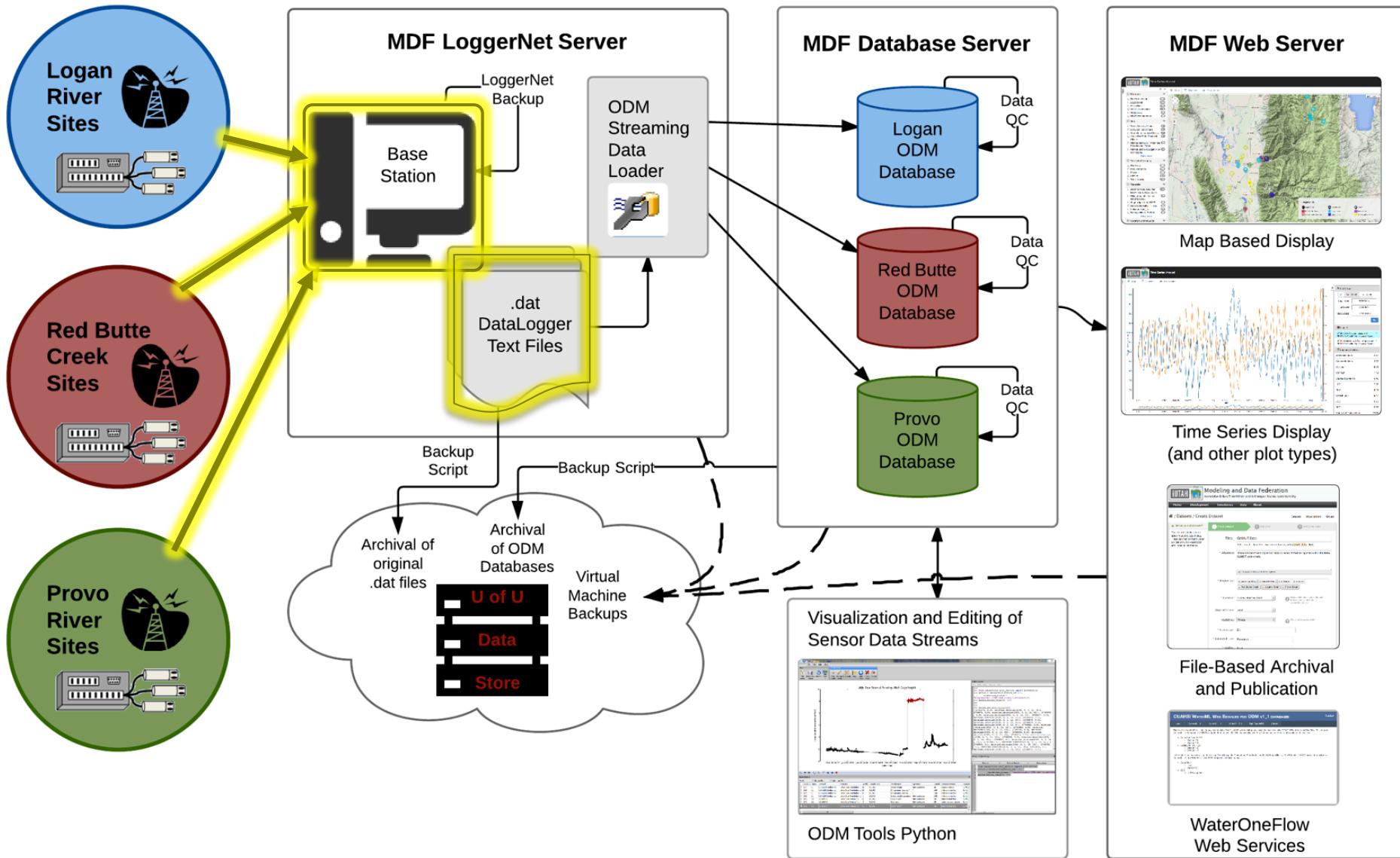
# Gradients Along Mountain to Urban Transitions (GAMUT) Network



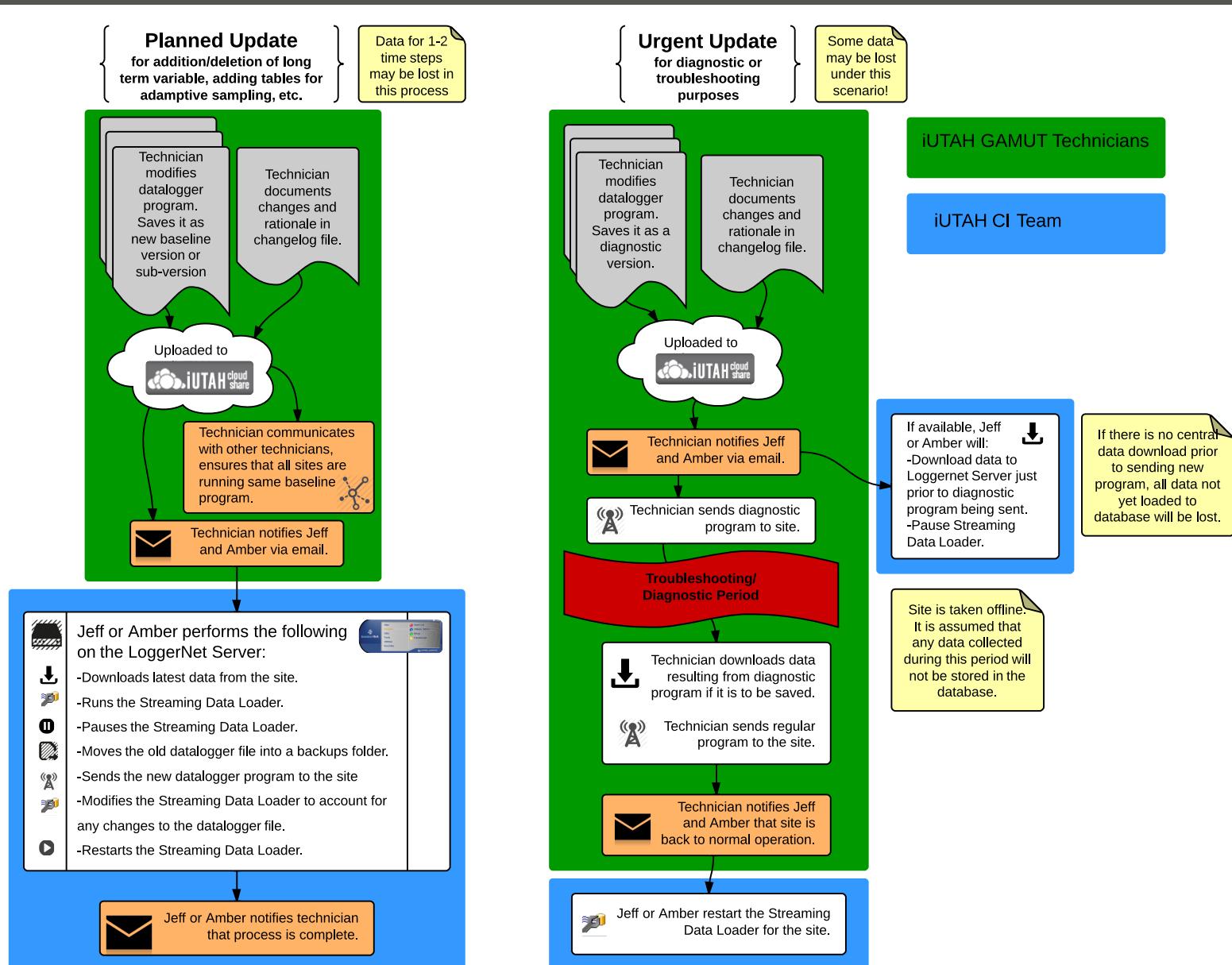
# GAMUT Data Workflow



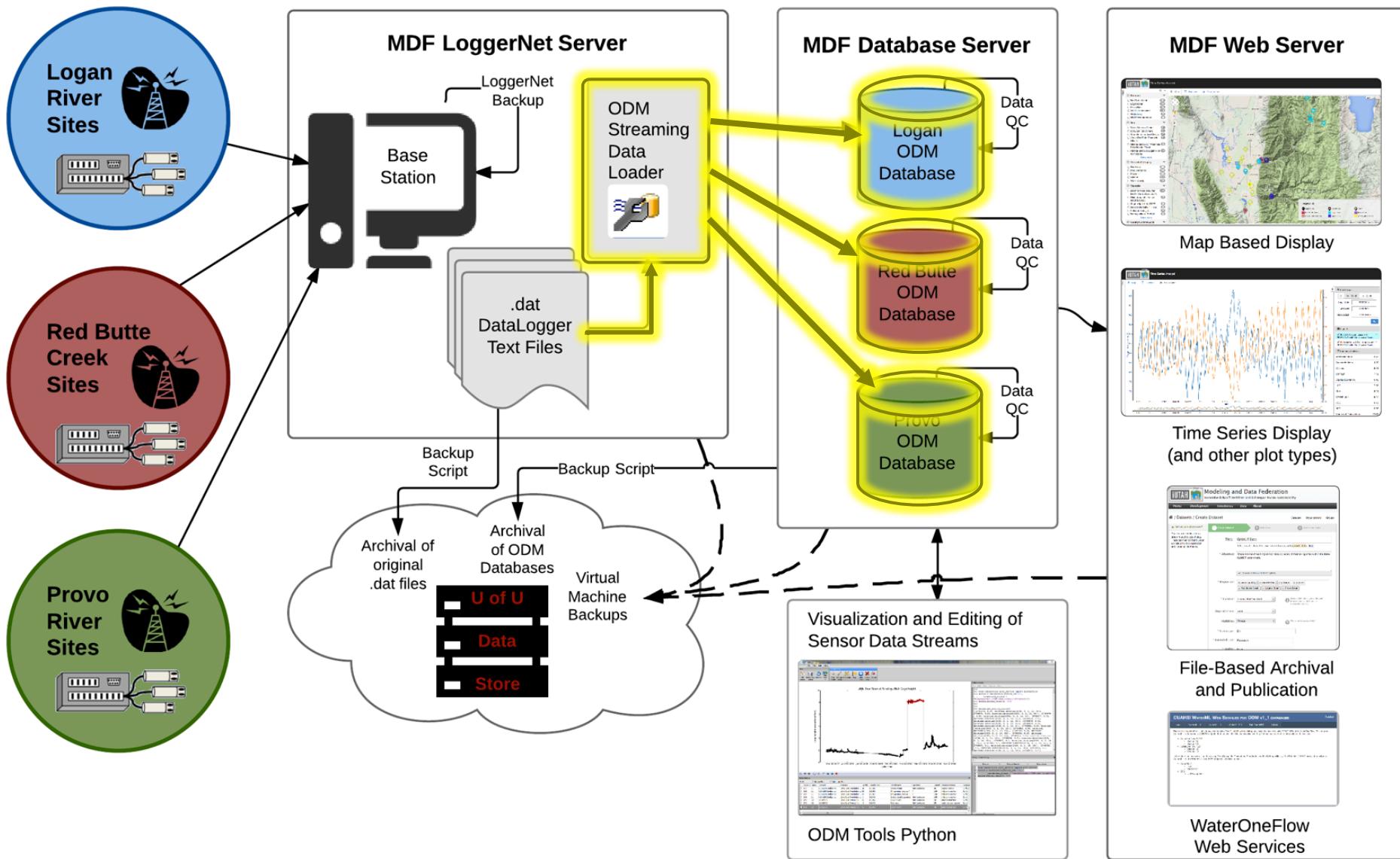
# Sensor Data Acquisition



# Datalogger Program Updates



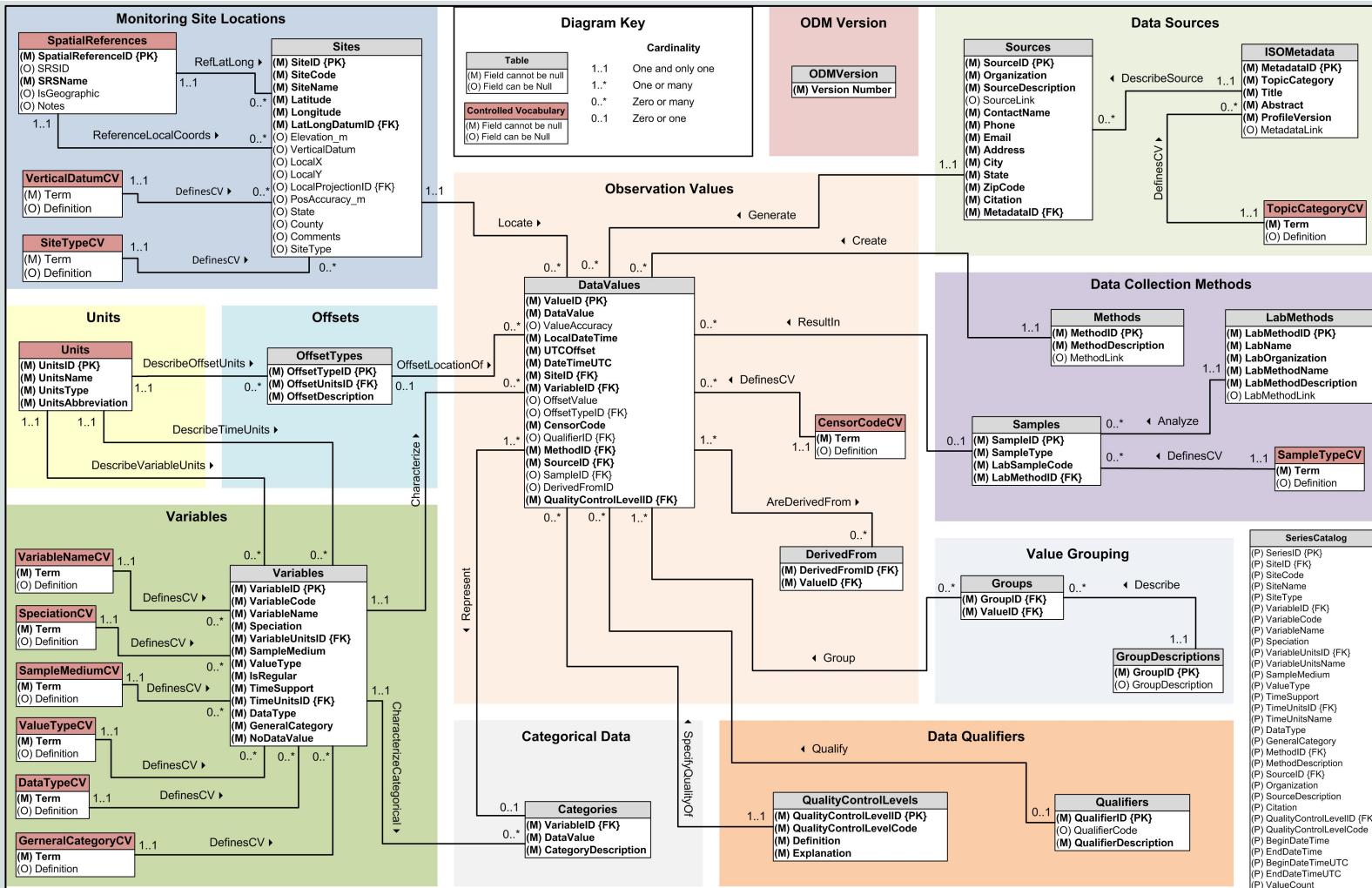
# Data Loading and Storage



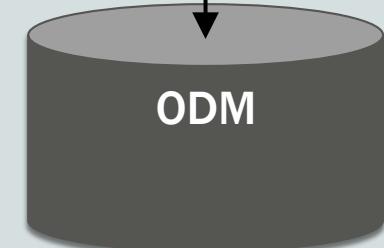
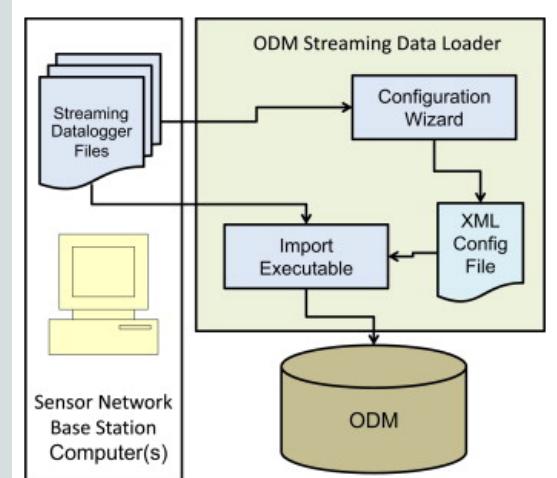
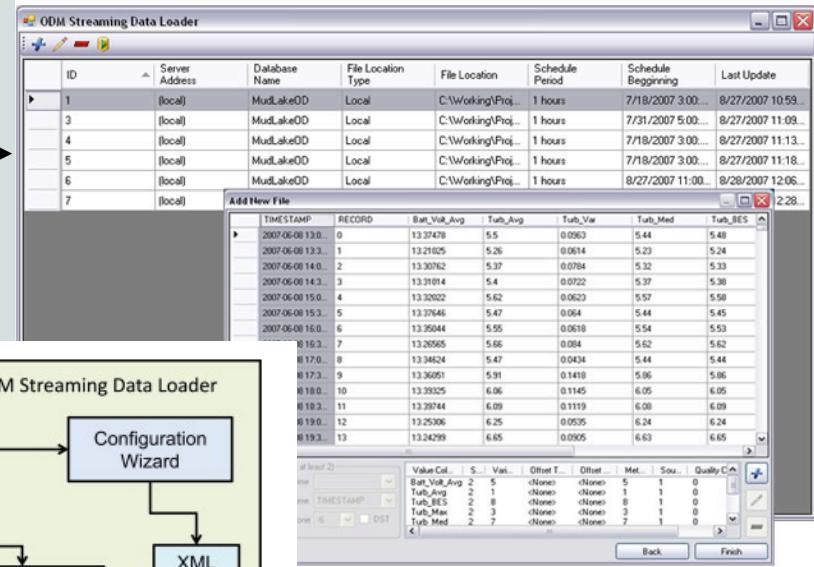
# Data Loading and Storage



CUAHSI  
HIS  
*Sharing hydrologic data*

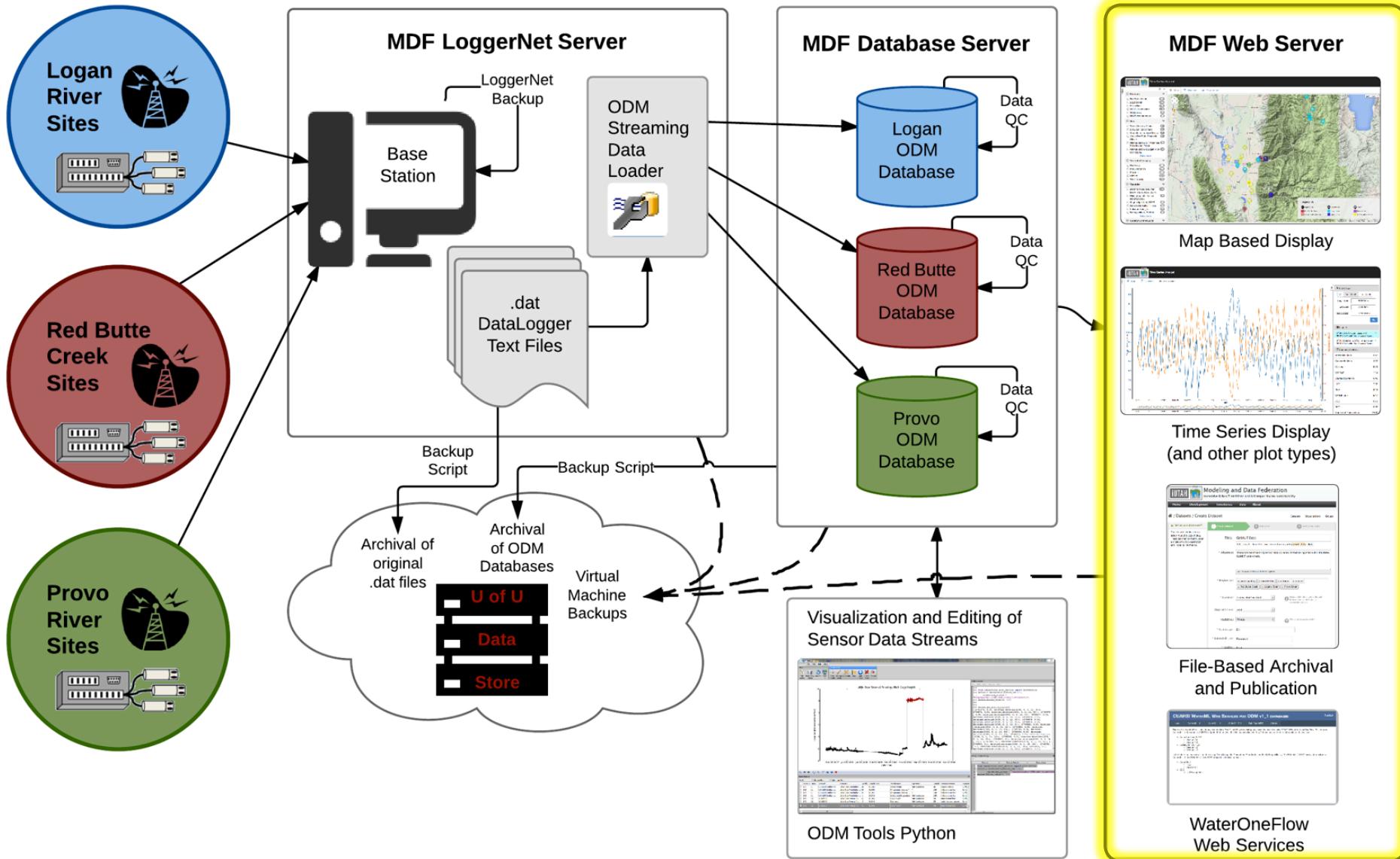


# Data Loading and Storage



ODM

# Web-Based Data Access



# Web-based Data Access

[data.iutahpsc.org/LoganRiverWOF/WaterML\\_1\\_0.aspx](http://data.iutahpsc.org/LoganRiverWOF/WaterML_1_0.aspx)

## CUAHSI WATERML WEB SERVICES FOR ODM v1\_1 DATABASES

[Log In]

Home WaterML 1.0 WaterML 1.1 WaterML 2.0 DataCart WFS Admin

WaterML is delivered over a SOAP API, called "WaterOneFlow". In this version, the API is updated to include a REST endpoint.

There are two methods in these services:

- REST
- SOAP API

Test the Rest Interface  
Site:

Variable:

Begin Date:

End Date:

Generate

VariableURL Site + Series URL Values URL All Sites URL

Generated URL:

[Run the Generated URL](#) [Reset](#)

WaterML is delivered over a SOAP API, called "WaterOneFlow". In this version, the API is updated to include a REST endpoint.

There are two methods in these services:

**SITES**

[utah:LR\_WaterLab\_AA] Logan River at the Utah Water Research Laboratory west bridge  
[utah:LR\_Menden\_AA] Logan River at Menden Road (600 South)  
[utah:LR\_Bridge\_AA] Logan River at 1st Street (Highway 89) Bridge  
[utah:LR\_TWOF] Logan River at the TWOF (Utah Water One Flow) gauging station  
[utah:LR\_GC\_Q] Climate Station at Logan River Gof Course

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<?xml version="1.0" encoding="UTF-8"?>
<Envelope xmlns="http://www.w3.org/2003/05/soap-envelope">
  <Header>
    <To>http://data.iutahpsc.org/LoganRiverWOF/REST/waternal_1_1.svc/siteinfo?location=utah:LR_WaterLab_AA</To>
    <Action>http://www.cuahsi.org/waternal_1_1/GetSiteInfo</Action>
  </Header>
  <Body>
    <GetSiteInfo>
      <@version>1.1</@version>
      <@methodId>1</@methodId>
      <@siteId>utah:LR_WaterLab_AA</@siteId>
    </GetSiteInfo>
  </Body>
</Envelope>
```

**VARIABLES**

[utah:LR\_AirTemp\_Avg] Temperature  
[utah:RH\_Beaufit\_Humidity]  
[utah:DeWP\_Avg] Temperature, dew point  
[utah:VaporPress\_Avg] Vapor pressure  
[utah:WindDir\_Avg] Wind direction  
[utah:WindSp\_Avg] Wind speed  
[utah:WindDir\_Avg] Wind direction  
[utah:WindSp\_Avg] Wind speed  
[utah:Precip\_Tot] Precipitation  
[utah:Precip\_Freq] Precipitation  
[utah:JuddDepth\_Avg] Snow depth  
[utah:Swin\_NR01\_Avg] Radiation, incoming shortwave  
[utah:Swin\_NR01\_Avg] Radiation, outgoing shortwave  
[utah:WVOL\_Corr\_NR01\_Avg] Radiation, incoming longwave  
[utah:WVOL\_Corr\_NR01\_Avg] Radiation, outgoing longwave

**VALUES/TIMESERIES**

[utah:LR\_WaterLab\_AA] Logan River BattVol [Logan River BattVol]  
[utah:LR\_WaterLab\_AA] Logan River RH\_end [Logan River RH\_end]  
[utah:LR\_WaterLab\_AA] Logan River RH\_start [Logan River RH\_start]  
[utah:LR\_WaterLab\_AA] Logan River Temp [Logan River Temp]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_Batt [Logan River Temp\_Batt]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_BattVol [Logan River Temp\_BattVol]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_BattVol\_end [Logan River Temp\_BattVol\_end]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_BattVol\_start [Logan River Temp\_BattVol\_start]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_BattVol\_U [Logan River Temp\_BattVol\_U]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_BattVol\_U\_end [Logan River Temp\_BattVol\_U\_end]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_BattVol\_U\_start [Logan River Temp\_BattVol\_U\_start]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_U [Logan River Temp\_U]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_U\_end [Logan River Temp\_U\_end]  
[utah:LR\_WaterLab\_AA] Logan River Temp\_U\_start [Logan River Temp\_U\_start]  
[utah:LR\_WaterLab\_AA] Logan River U [Logan River U]  
[utah:LR\_WaterLab\_AA] Logan River U\_end [Logan River U\_end]  
[utah:LR\_WaterLab\_AA] Logan River U\_start [Logan River U\_start]  
[utah:LR\_WaterLab\_AA] Logan River V [Logan River V]  
[utah:LR\_WaterLab\_AA] Logan River V\_end [Logan River V\_end]  
[utah:LR\_WaterLab\_AA] Logan River V\_start [Logan River V\_start]

Battery voltage measured by a Campbell Scientific CR3000 datalogger.

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

Home Development Data About

## Logan River

The Logan River watershed is located in the heart of the Bear River range with headwaters near the Utah-Idaho border. The river flows southeast through Logan Canyon—a landscape dominated by formerly glaciated peaks, limestone cliffs, and the occasional sinkhole. The underlying bedrock has numerous caves that create natural springs that contribute to the river's year-round discharge. Near its confluence with the Bear, it is dammed in three locations (First, Second, and Third dams) for hydroelectric generation. After exiting the mountains, the river flows west through Cache Valley and is impacted by a mixture of agricultural and urban environments. The Logan River converges with the Little Bear River in central Cache Valley before flowing north to the main stem of the Bear River and Cutler Reservoir.

**Monitoring Sites:** Click on a site code to visualize and download data  
**The data presented here are provisional and subject to revision**

Site Code	Site Name	Site Type
LR_Mendon_AA	Logan River at Mendon Road (600 South) Aquatic	Aquatic
LR_WaterLab_AA	Logan River at the USU Water Lab Aquatic	Aquatic
LR_MainStreet_BA	Logan River at the Main Street Bridge Aquatic	Aquatic
LR_GCC_C	Golf Course Climate	
LR_TWDEF_C	Experimental Forest	
LR_PB_C	Franklin Basin Climate	
LR_TG_C	Tony Grove Climate	
LR_TG_BA	Logan River near T. G. Wilkins Repeater	
LR_Wilkins_R	Wilkins Repeater	

This project is funded by the U.S. Geological Survey. All data and material are those of the agency.

**Logan River at the Utah Water Laboratory**

Site Code LR\_WaterLab\_AA

Latitude 41.739034 Longitude -111.795274 Lat/long WGS84

Location Projected UTM Zone 13N

Datum	Location		
Elevation	1414.0	Comments	
Local X	None	Watershed	Logan
Local Y	None	Site Type	Stream



Multiple instruments are used to collect data.

**Most Recent Instantaneous Measurements**

Data update time: 2014-06-25 11:45:00, past 24 hours shown.

Temperature WaterTemp (°C) 10.620 degC	Specific Conductance 18.300 µS/cm	pH pH 8.460 pH
Oxygen dissolved ooo 9.830 mg/L	Oxygen dissolved saturation 88.500 % Sat	Turbidity Turbidity 1.490 NTU
Blue-green algae (cyanobacteria), Dyphyco 1 -0.030 RFU	Chlorophyll Fluorescence -0.090 RFU	Colored Dissolved Organic Matter CDOM 0.940 QSU
Gage height slope 55.280 cm		

Time Series Analyst

Network

- Red Butte Creek
- Logan River
- Plover River
- USGS Groundwater
- USGS Daily
- USGS Instantaneous

Site

- Todd Meadow Climate
- Yellow Fork Climate
- Green Infrastructure Climate
- Above Red Butte Reservoir
- Climate Station at TW Daniels Experimental Forest
- Climate Station at Logan River
- Get Coordinates

Show more

Variable Category

- Hydrology
- Instrumentation
- Other
- Water Quality

Variable

- Relative Humidity, RH, enc
- Indicator, Door, Tr
- Temperature, Air, Temp
- Snow Depth, JustDepth, Avg
- Temperature, WaterTemp, Turb
- Temperature, WaterTemp, DWD
- Temperature, WaterTemp, EWD

Quality Control Level

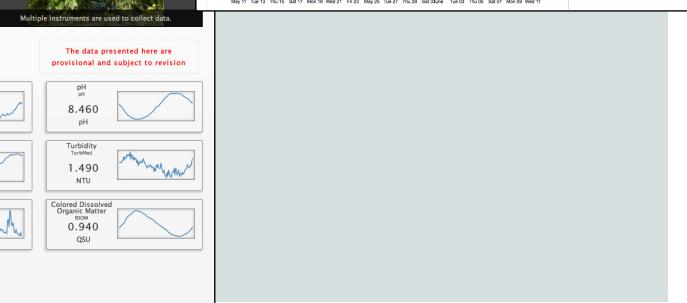
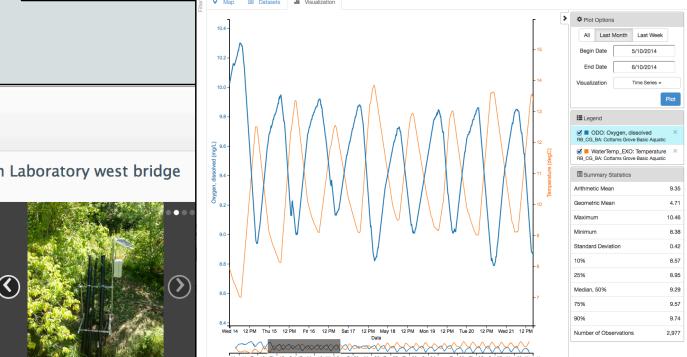
- Raw data

Cottonwood Grove Basic Aquatic  
RBC\_014

Network: Red Butte  
Country: San Luis  
County: Sanpete  
State: Utah  
Elevation: 4070.000000  
Longitude: -111.823286  
Site Type: Stream

View Datasets

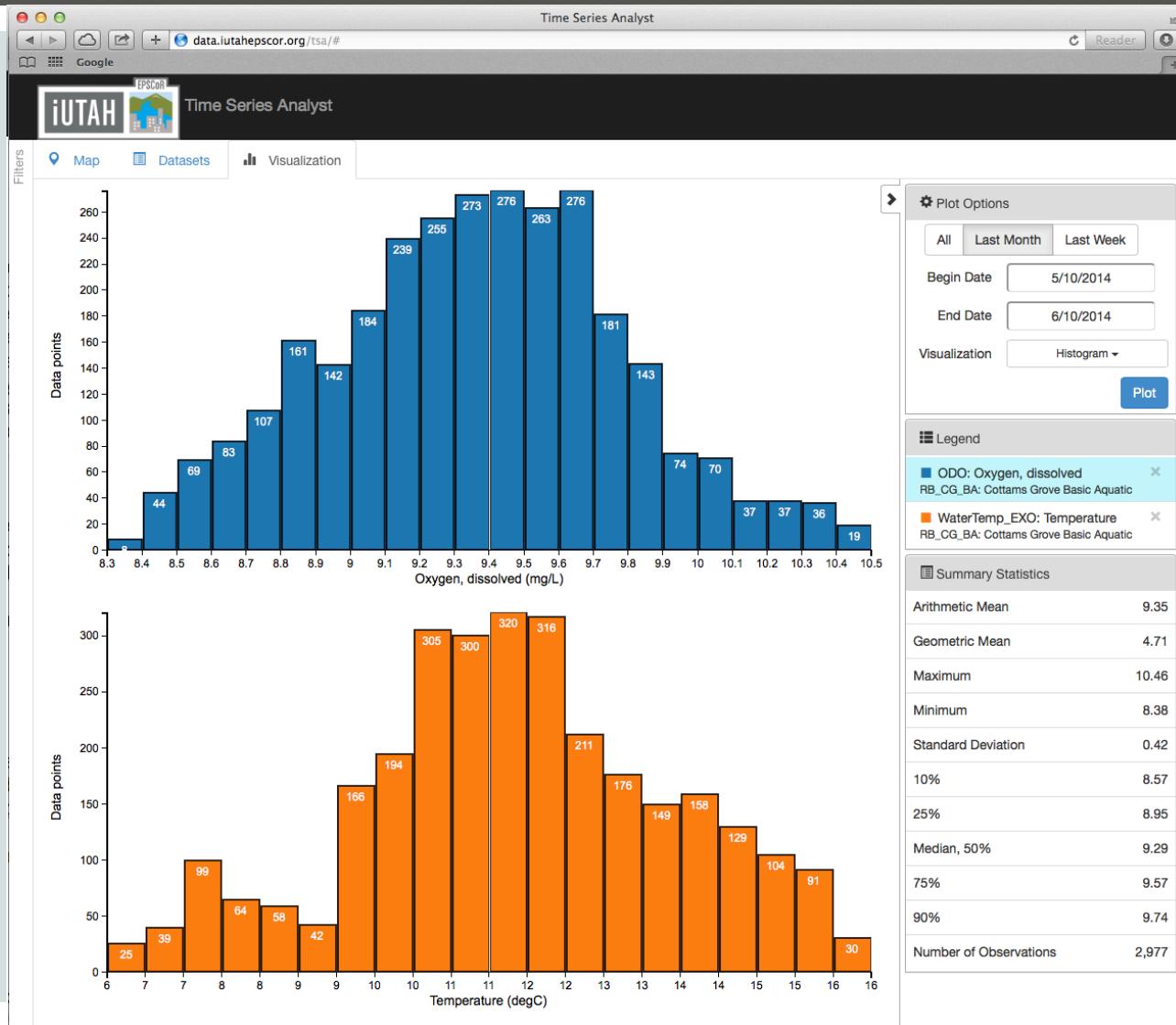
Time Series Analyst



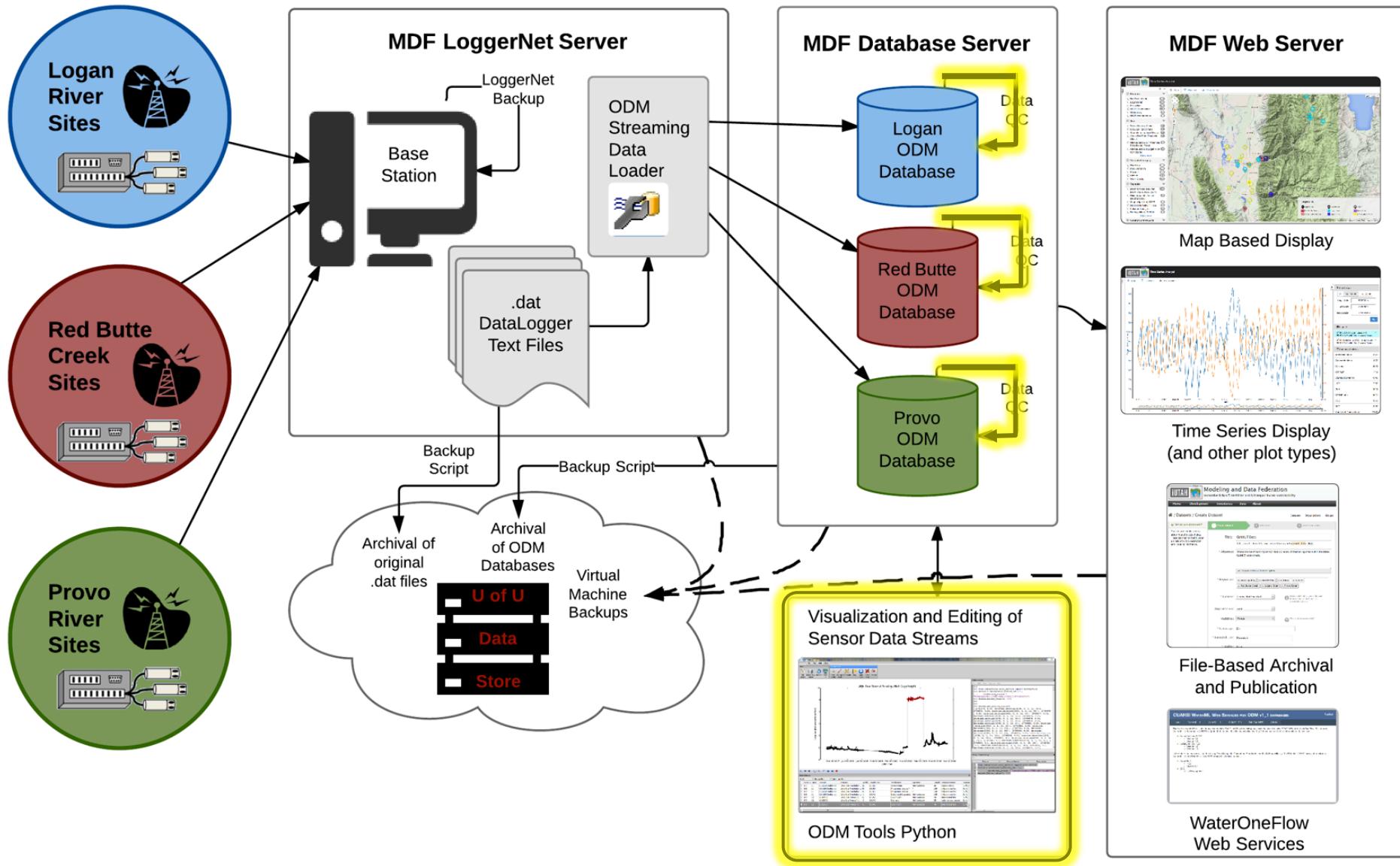
CUAHSI  
HIS

### *Sharing hydrologic data*

# Enhanced Web-Based Time Series Data Access and Visualization



# Sensor Data QAQC



# Development of a QAQC Plan

## Quality Assurance:

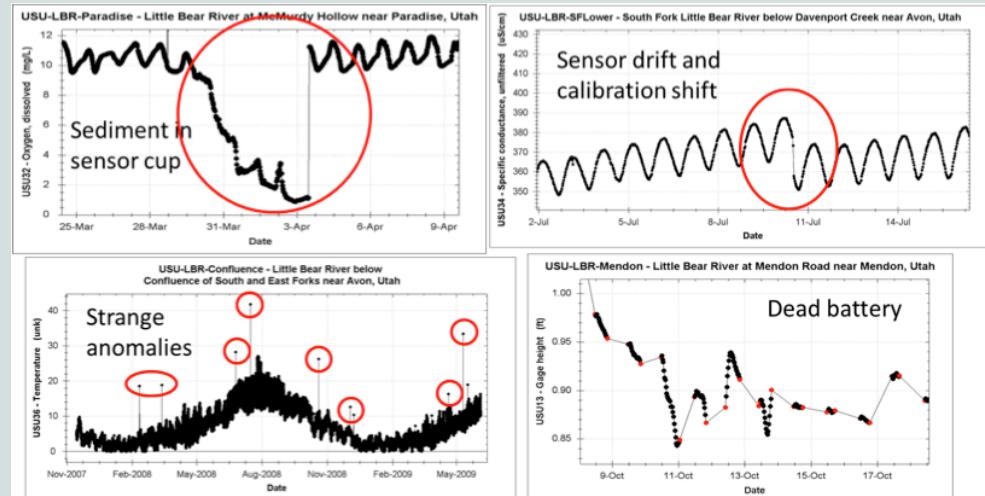
“...protocols developed and adhered to in a way that minimizes inaccuracies in the data produced ... produces high-quality data while minimizing the need for corrective measures to improve data quality.”

- 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

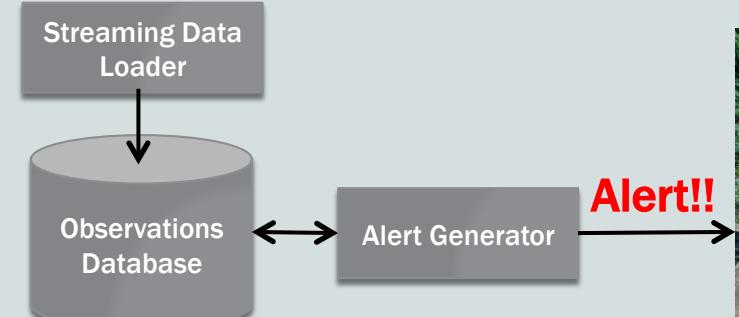


# QAQC: Automated Alerts

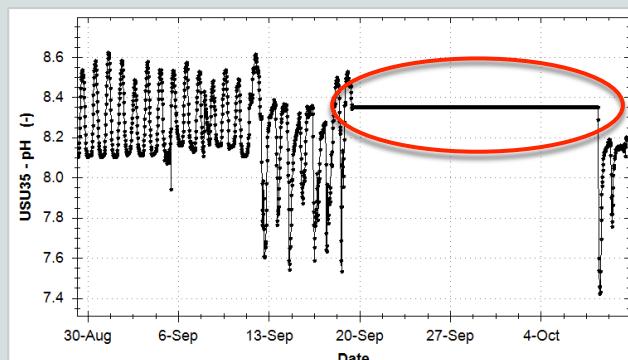
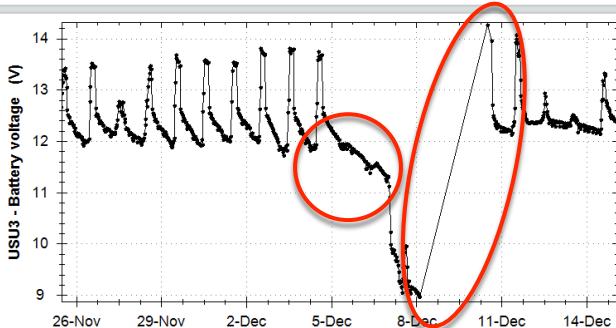
Technicians receive email alerts daily

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 are 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_Avg	.00	2014-02-17 18:45	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 06:15	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	-0.9999	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



# QAQC: Data Visualization and Management

Multiple Plot Types

Plot Display Options

Dynamic  
Zooming and  
Panning

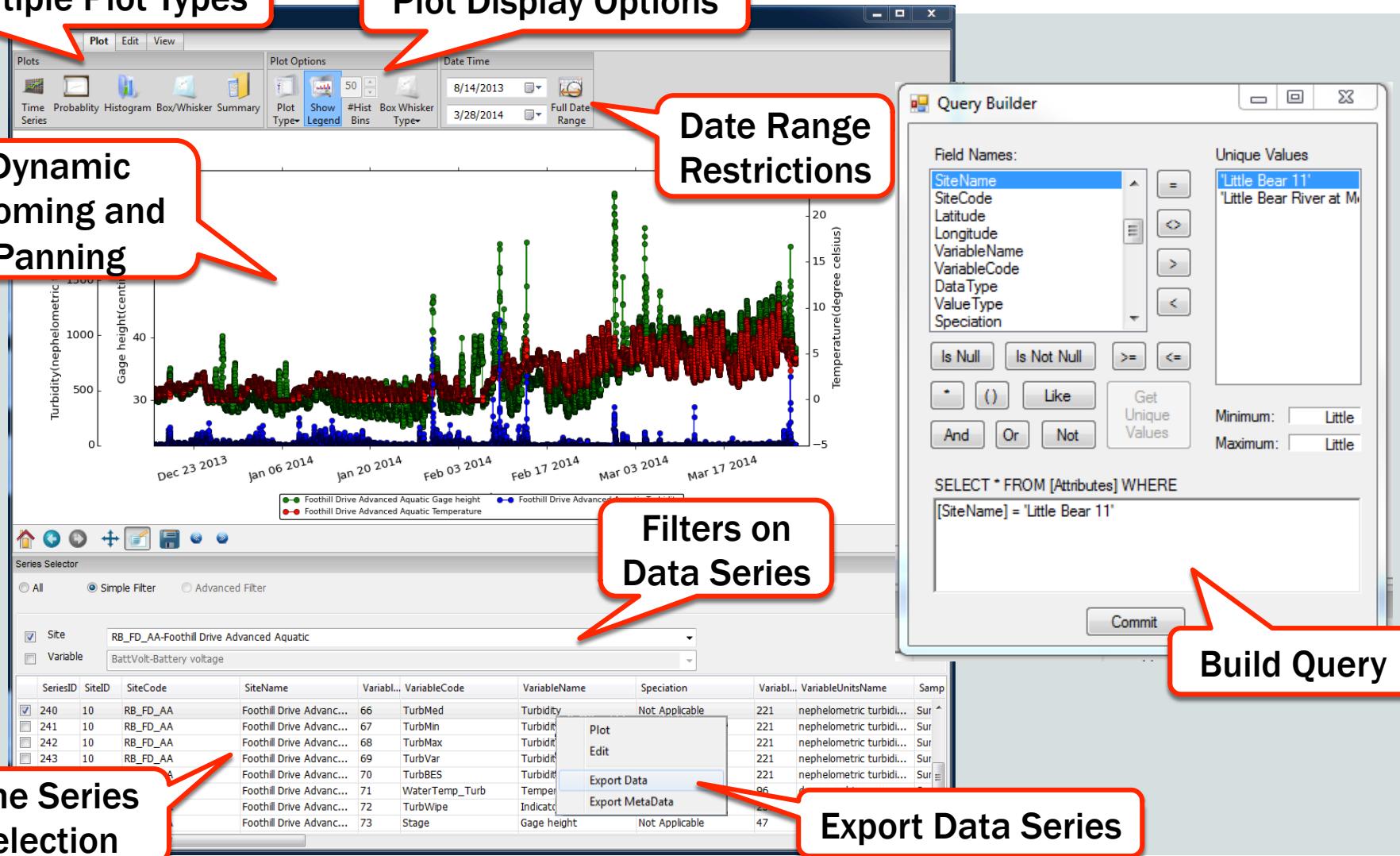
Date Range  
Restrictions

Filters on  
Data Series

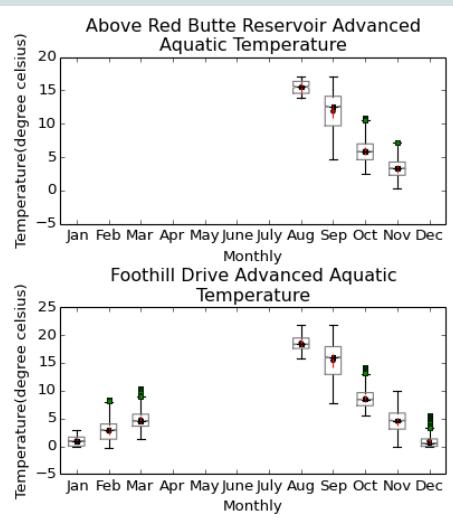
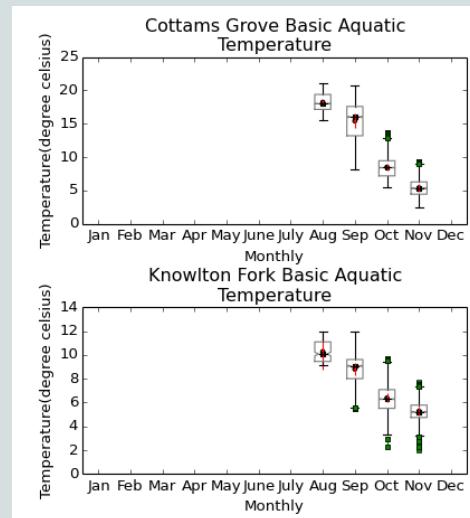
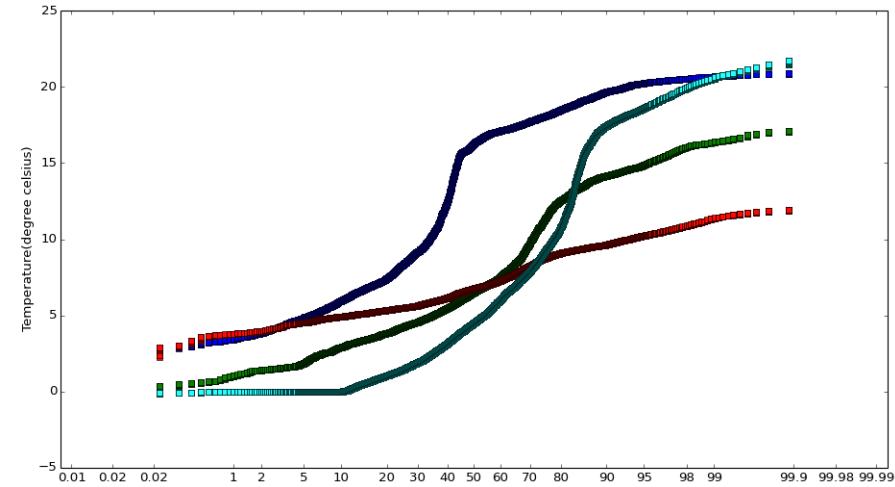
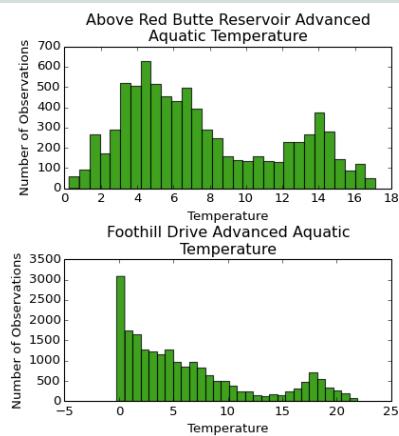
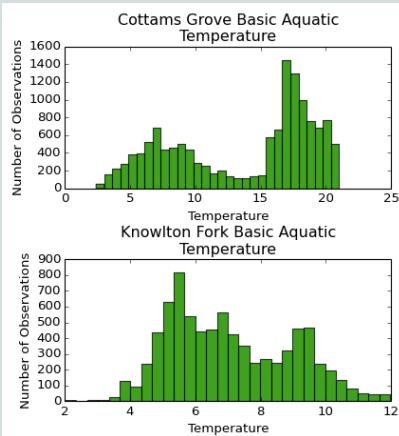
Time Series  
Selection

Export Data Series

Build Query



# QAQC: Data Visualization and Management



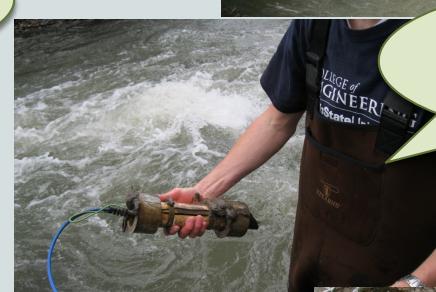
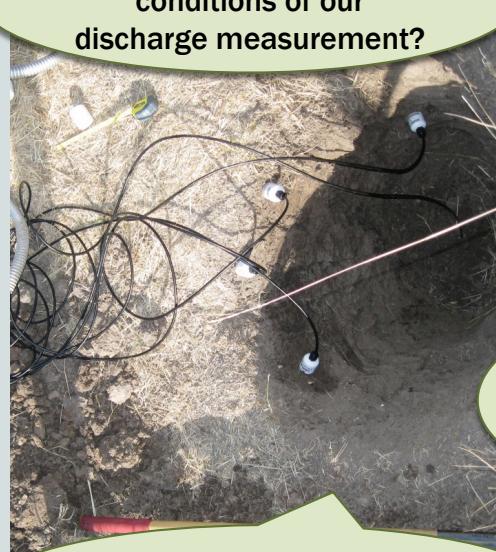
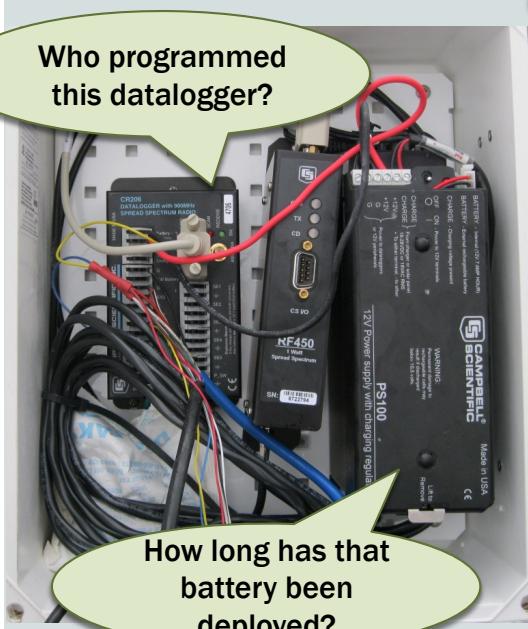
# QAQC: Post Processing

The screenshot displays the ODM Tools software interface, specifically the QAQC Post Processing module. The interface includes:

- Data editing tools:** A toolbar at the top with various icons for editing functions.
- Dynamic data editing display:** A plot titled "Foothill Drive Advance" showing Specific Conductance (microsiemens per centimeter) over time (Date from Jan 03 to Feb 07, 2014). The plot shows several sharp peaks in conductance.
- Data Filter dialog:** A modal window titled "Data Filter" with three tabs: Value Threshold, Data Gaps, and Date. The Date tab is selected, showing fields for "Before:" (8/4/2014) and "After:" (9/14/2013).
- Data selection filters:** A callout pointing to the "Data Filter" dialog.
- Tabular data view and selection:** A table titled "Table View" showing a list of data records with columns: ValueID, DataValue, ValueAccuracy, and LocalDate. The table lists numerous entries from January 2014.
- Python script editor:** A code editor at the bottom containing Python script code for editing a service.

```
3 edit_service = EditService(series_id=261, connection_string='mssql+pyodbc://Amber:xxxxxxxxxxxx@(local)/iUTAH_RedButte_OD')
4 series_service = SeriesService(connection_string='mssql+pyodbc://Amber:xxxxxxxxxxxx@(local)/iUTAH_RedButte_OD')
5 ## To run commands from the python console uncomment and run the following commands ##
6 #edit_service = Tools
7 #series_service = Tools.get_series_service()
8 edit_service.filter_date(datetime.datetime(2014, 1, 31, 0, 0), datetime.datetime(2014, 1, 15, 0, 0))
9 edit_service.filter_date(datetime.datetime(2014, 1, 15, 0, 0), datetime.datetime(2014, 1, 1, 0, 0))
10 edit_service.drift_correction(0.5)
11 edit_service.flag(5)
12
```

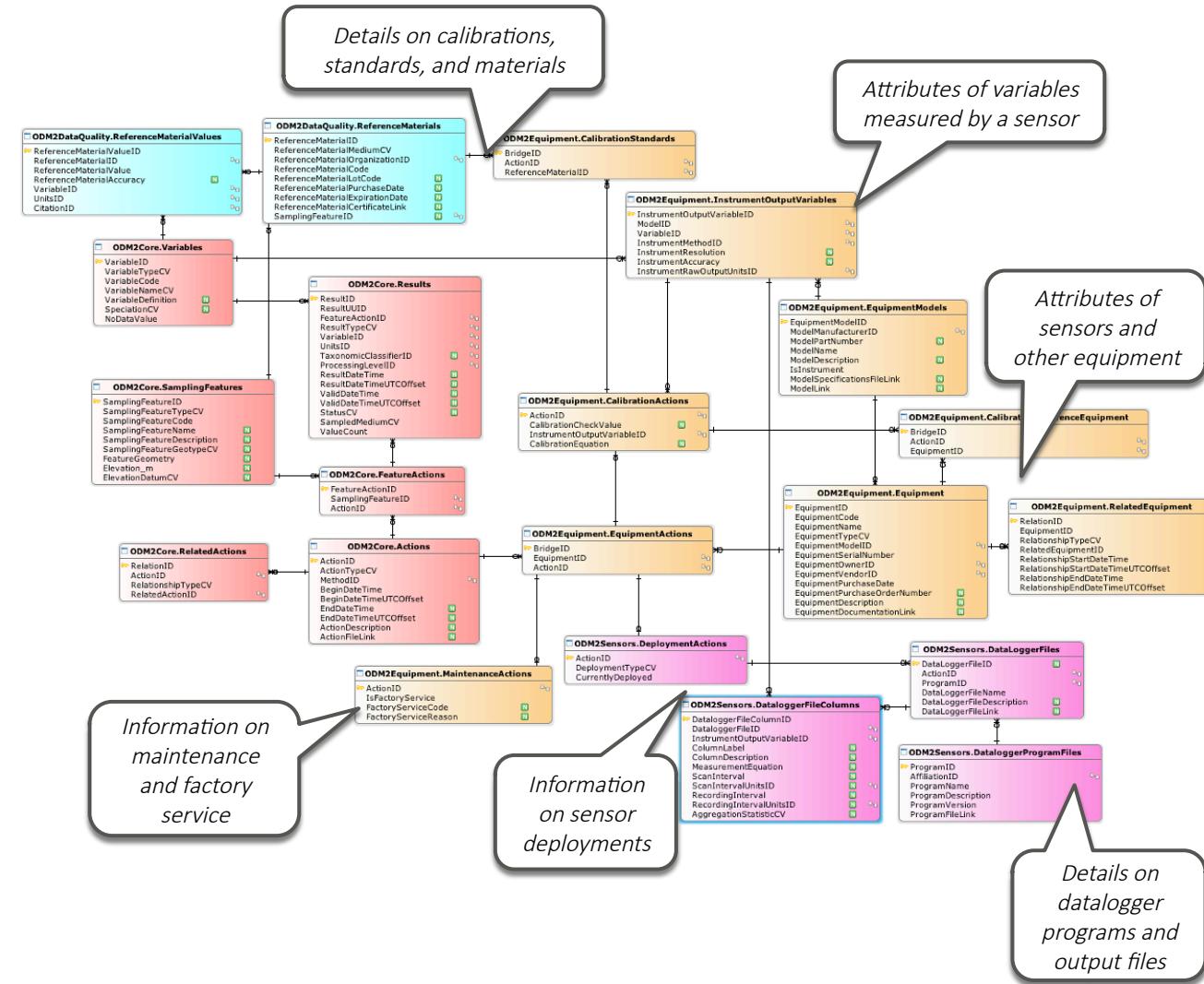
# Monitoring Equipment Management



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.**

# Monitoring Equipment Management

**Sites**

**Site Details**

**Deployment Measured Variables**

**Deployment History of Main Street Bridge Basic Aquatic**

**Site Visits**

**Site Visit Details**

**Site Visit Location**

**Field Activities Performed**

**Equipment**

**Equipment Details**

**Deployment Details**

**Manufacturer Information**

**Vendor Information**

**Deployment Description**

**Measured Variables**

- 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



## Site Visits



## Field Activities



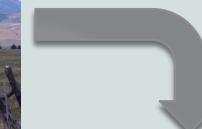
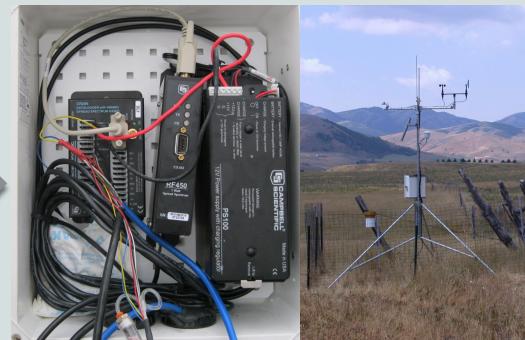
- Activity type
- Description
- Dates
- Date

## Calibrations

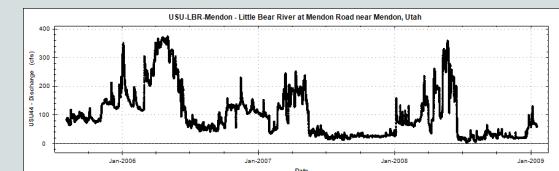


- Method
- Standard

## Deployments



## 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

EPSCOR

iUTAH



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 Eriksson

## 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>