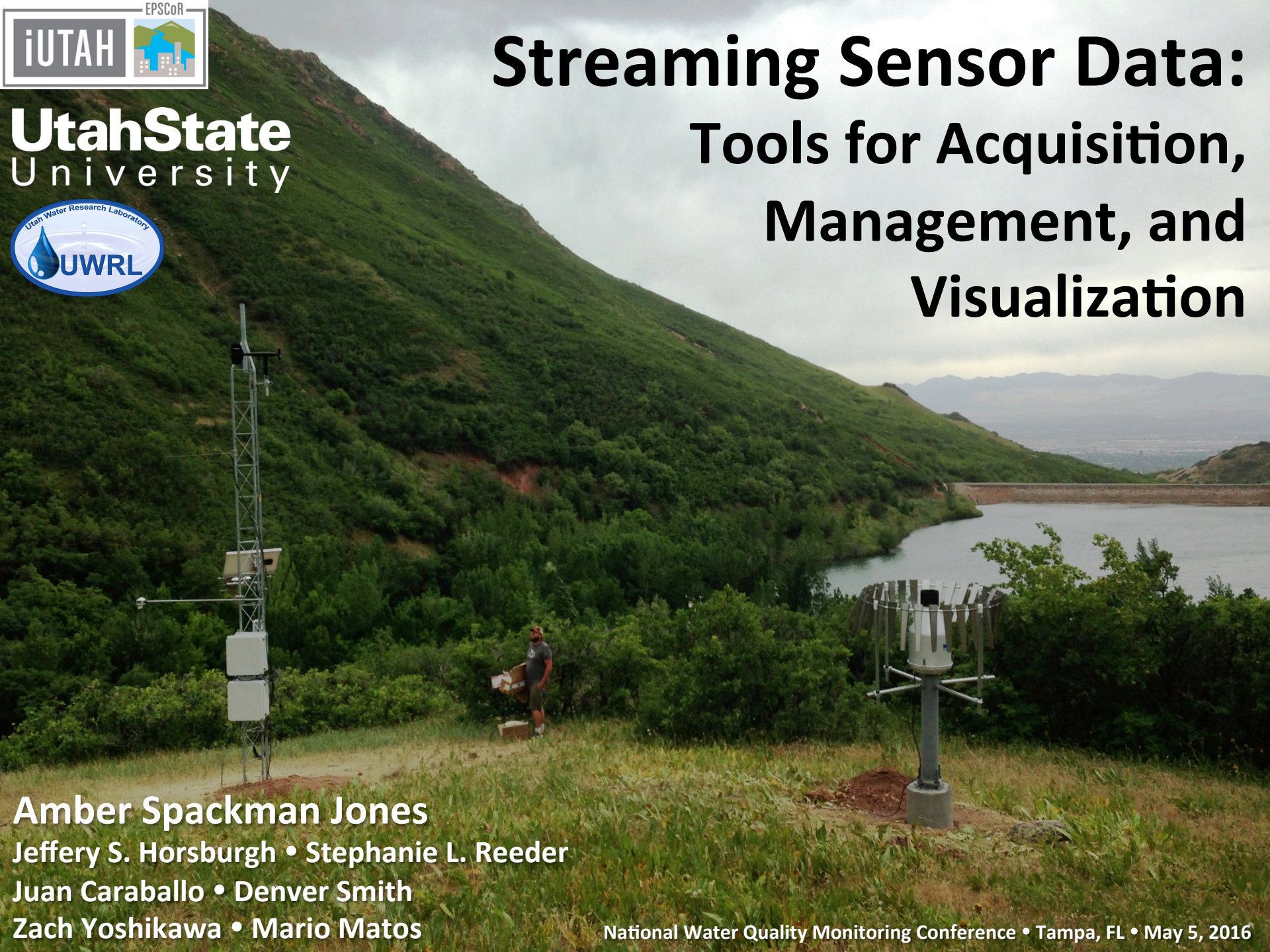
Streaming Sensor Data: Tools for Acquisition, Management, and Visualization



Amber Spackman Jones

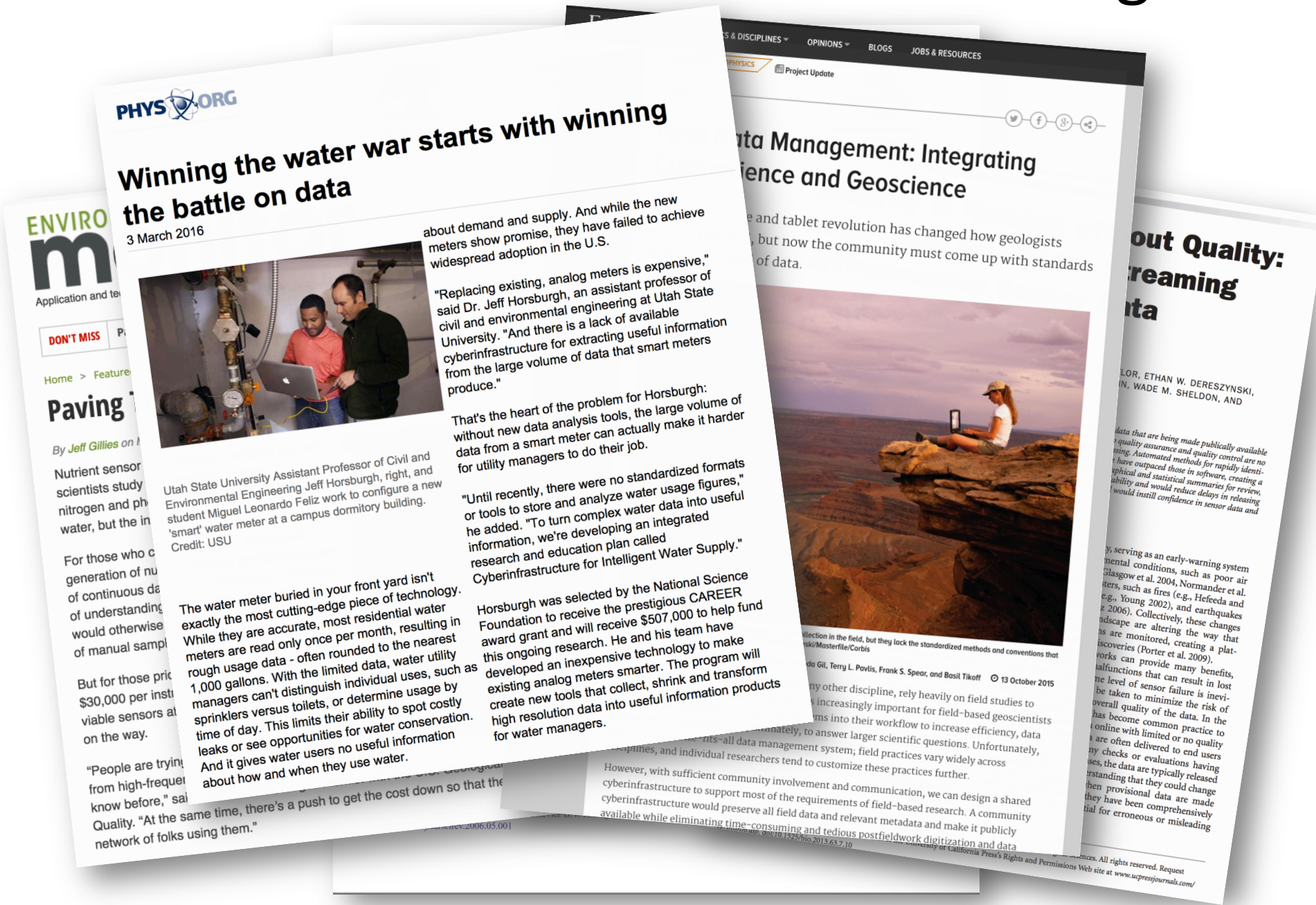
Jeffery S. Horsburgh • Stephanie L. Reeder

Juan Caraballo • Denver Smith

Zach Yoshikawa • Mario Matos

National Water Quality Monitoring Conference • Tampa, FL • May 5, 2016

In Situ Environmental Monitoring



PHYS.ORG

Winning the water war starts with winning the battle on data

3 March 2016

Utah State University Assistant Professor of Civil and Environmental Engineering Jeff Horsburgh, right, and student Miguel Leonardo Feliz work to configure a new 'smart' water meter at a campus dormitory building. Credit: USU

The water meter buried in your front yard isn't exactly the most cutting-edge piece of technology. While they are accurate, most residential water meters are read only once per month, resulting in rough usage data - often rounded to the nearest 1,000 gallons. With the limited data, water utility managers can't distinguish individual uses, such as sprinklers versus toilets, or determine usage by time of day. This limits their ability to spot costly leaks or see opportunities for water conservation. And it gives water users no useful information about how and when they use water.

about demand and supply. And while the new meters show promise, they have failed to achieve widespread adoption in the U.S.

"Replacing existing, analog meters is expensive," said Dr. Jeff Horsburgh, an assistant professor of civil and environmental engineering at Utah State University. "And there is a lack of available cyberinfrastructure for extracting useful information from the large volume of data that smart meters produce."

That's the heart of the problem for Horsburgh: without new data analysis tools, the large volume of data from a smart meter can actually make it harder for utility managers to do their job.

"Until recently, there were no standardized formats or tools to store and analyze water usage figures," he added. "To turn complex water data into useful information, we're developing an integrated research and education plan called Cyberinfrastructure for Intelligent Water Supply."

Horsburgh was selected by the National Science Foundation to receive the prestigious CAREER award grant and will receive $507,000 to help fund this ongoing research. He and his team have developed an inexpensive technology to make existing analog meters smarter. The program will create new tools that collect, shrink and transform high resolution data into useful information products for water managers.

ENVIRO m

Application and te

DON'T MISS

Home > Feature

Paving

By Jeff Gillies on

Nutrient sensor scientists study nitrogen and ph water, but the in

For those who c generation of nu of continuous da of understanding would otherwise of manual sampl

But for those pric $30,000 per inst viable sensors a on the way.

"People are trying from high-frequen know before," sai Quality. "At the same time, there's a push to get the cost down so that the network of folks using them."

S & DISCIPLINES OPINIONS BLOGS JOBS & RESOURCES

PHYSICS Project Update

ata Management: Integrating ience and Geoscience

e and tablet revolution has changed how geologists , but now the community must come up with standards of data.

llection in the field, but they lack the standardized methods and conventions that ski/Masterfile/Corbis

da Gil, Terry L. Pavlis, Frank S. Spear, and Basil Tikoff 13 October 2015

ny other discipline, rely heavily on field studies to s increasingly important for field-based geoscientists ms into their workflow to increase efficiency, data nately, to answer larger scientific questions. Unfortunately, nts-all data management system; field practices vary widely across plines, and individual researchers tend to customize these practices further.

However, with sufficient community involvement and communication, we can design a shared cyberinfrastructure to support most of the requirements of field-based research. A community cyberinfrastructure would preserve all field data and relevant metadata and make it publicly available while eliminating time-consuming and tedious postfieldwork digitization and data

doi:10.1525/bio.2013.63.7.10

ences. All rights reserved. Request Permissions Web site at www.ucpressjournals.com/

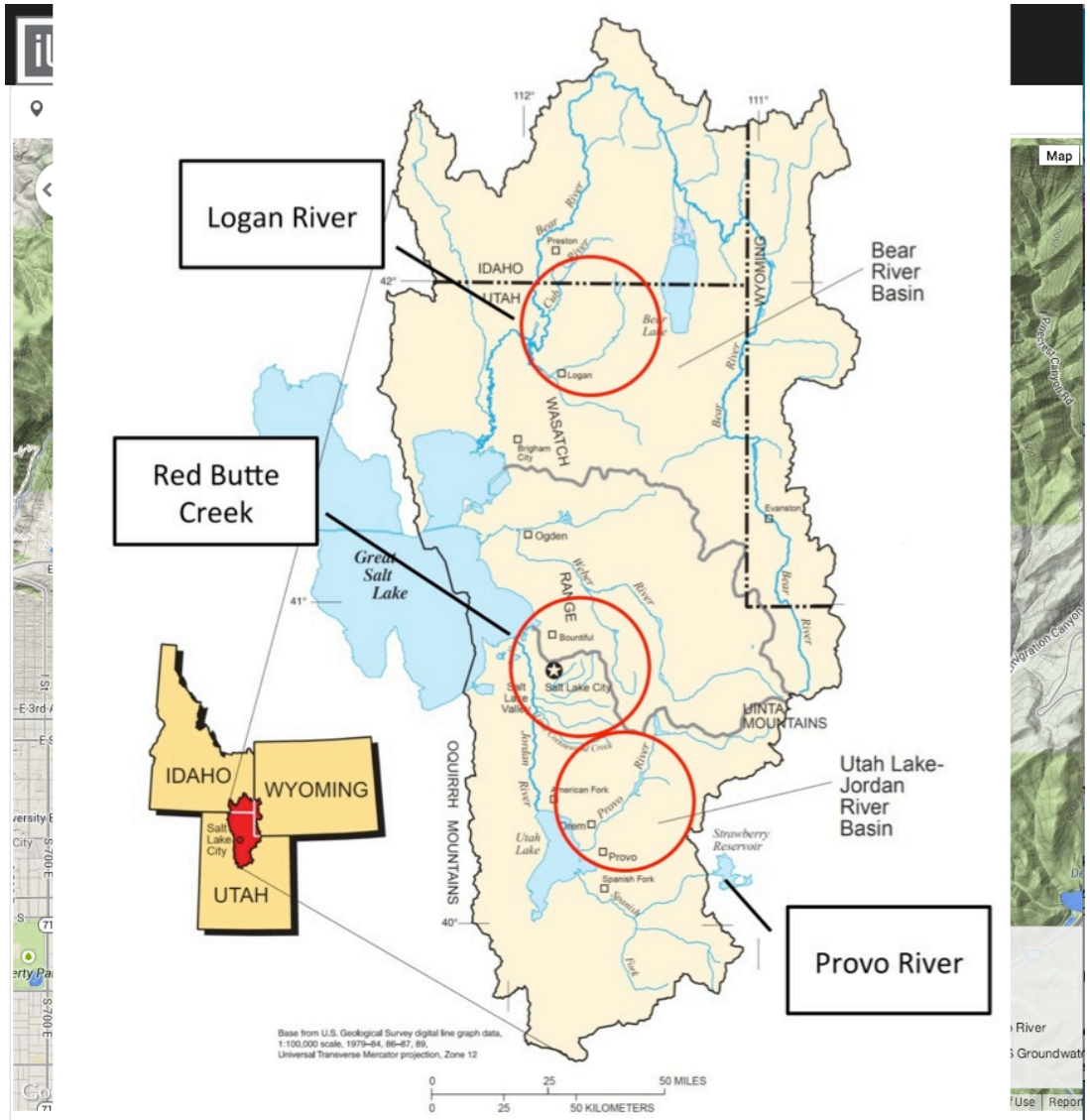
out Quality: reaming ta

LOR, ETHAN W. DERESZYNSKI, N, WADE M. SHELDON, AND

data that are being made publically available quality assurance and quality control are no sing. Automated methods for rapidly identi- have outpaced those in software, creating a phical and statistical summaries for review, ability and would reduce delays in releasing would instill confidence in sensor data and

y, serving as an early-warning system nental conditions, such as poor air Glasgow et al. 2004, Normander et al. ters, such as fires (e.g., Hefeeda and e.g., Young 2002), and earthquakes z 2006). Collectively, these changes ndscape are altering the way that ns are monitored, creating a plat- iscoveries (Porter et al. 2009). orks can provide many benefits, alfunctions that can result in lost me level of sensor failure is inevi- be taken to minimize the risk of overall quality of the data. In the has become common practice to online with limited or no quality s are often delivered to end users ny checks or evaluations having ses, the data are typically released rstanding that they could change hen provisional data are made hey have been comprehensively tial for erroneous or misleading

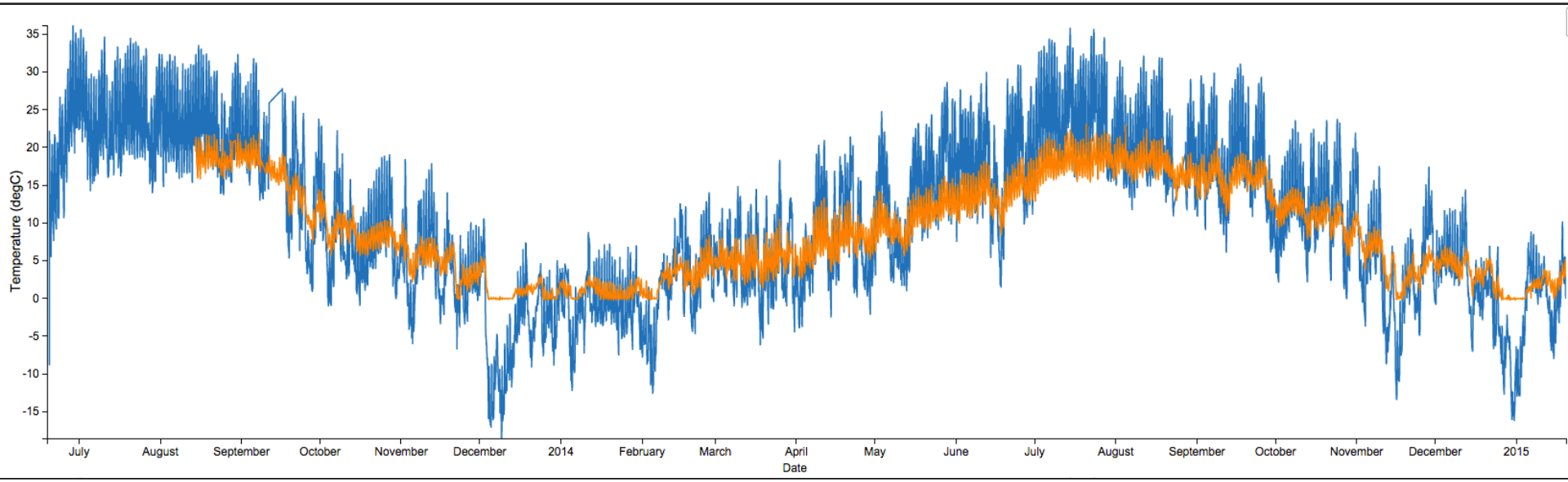
Gradients Along Mountain To Urban Transitions (GAMUT) Network



- 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

EPSCoR iUTAH

The Data Deluge



One year = 35,040 observations
So far (~3 years) = 105,000+ observations

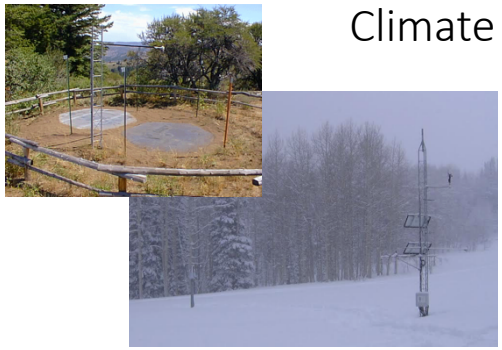
Times 14 Aquatic Sites with ~26 Variables
Times 14 Climate Sites with ~74 Variables
Plus different versions of the data (raw versus checked) = **over 200 million** observations

You need some infrastructure to manage and share the data.

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

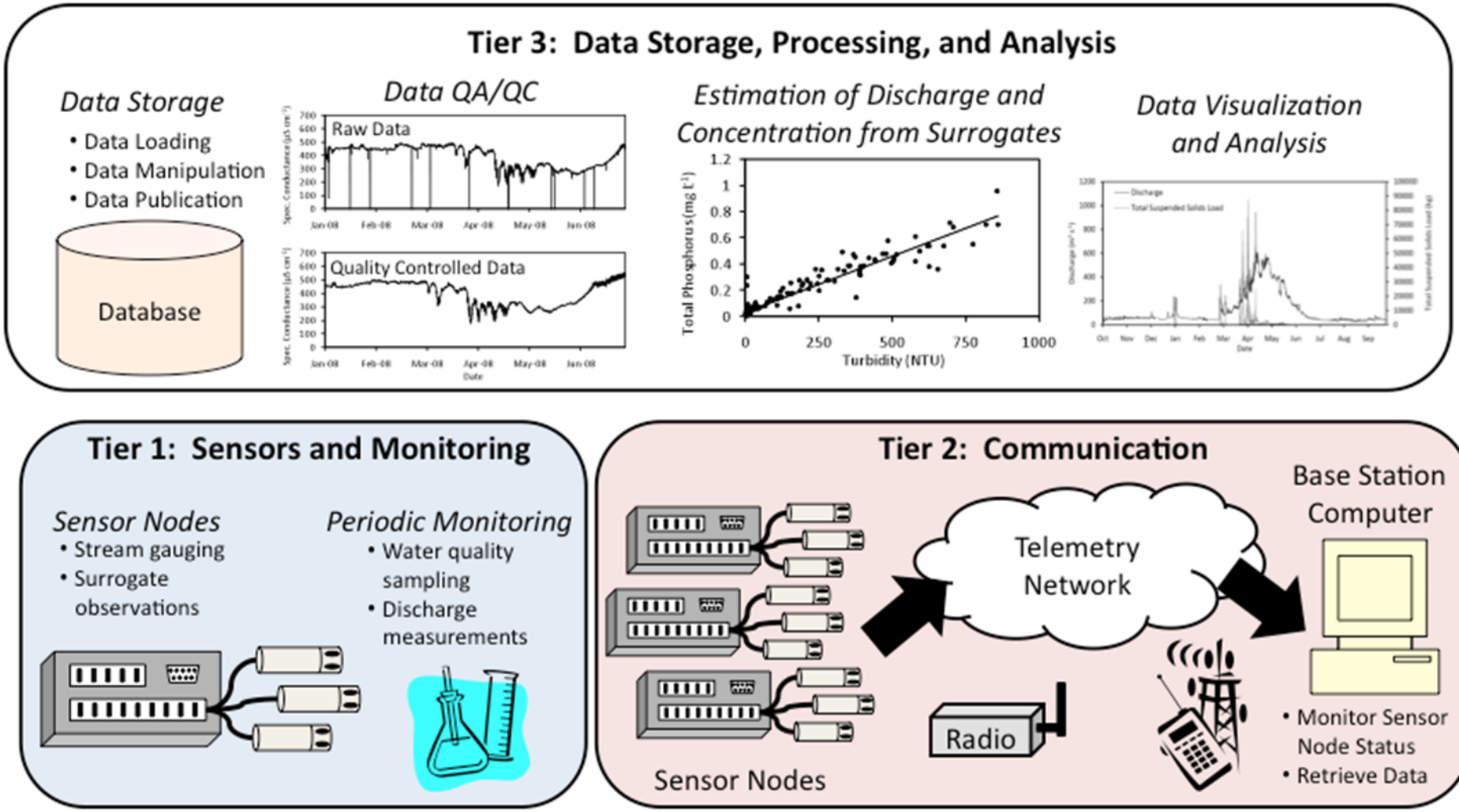
Water quality

Water quantity

Groundwater

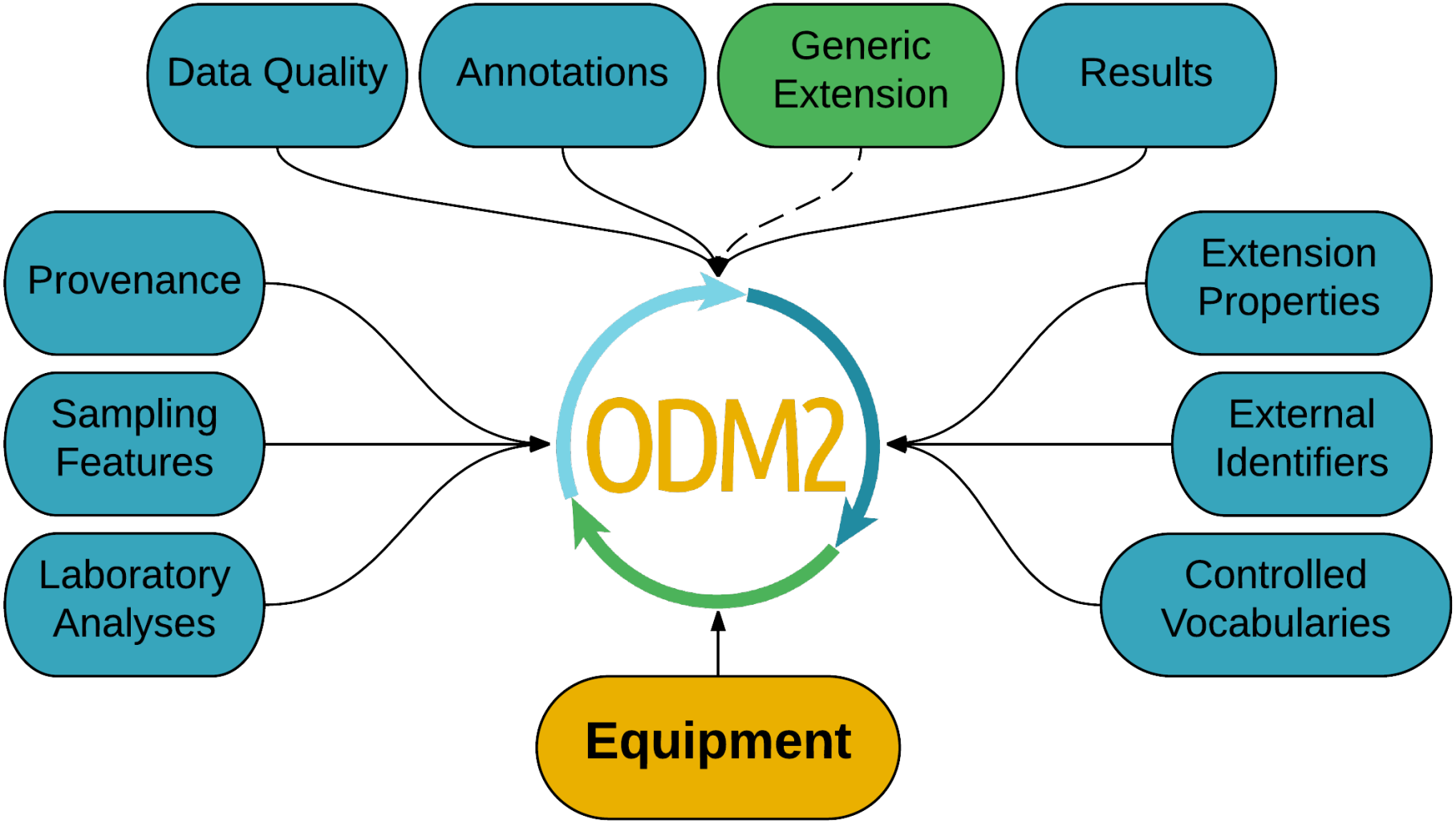
Mobile Platforms

Cyberinfrastructure

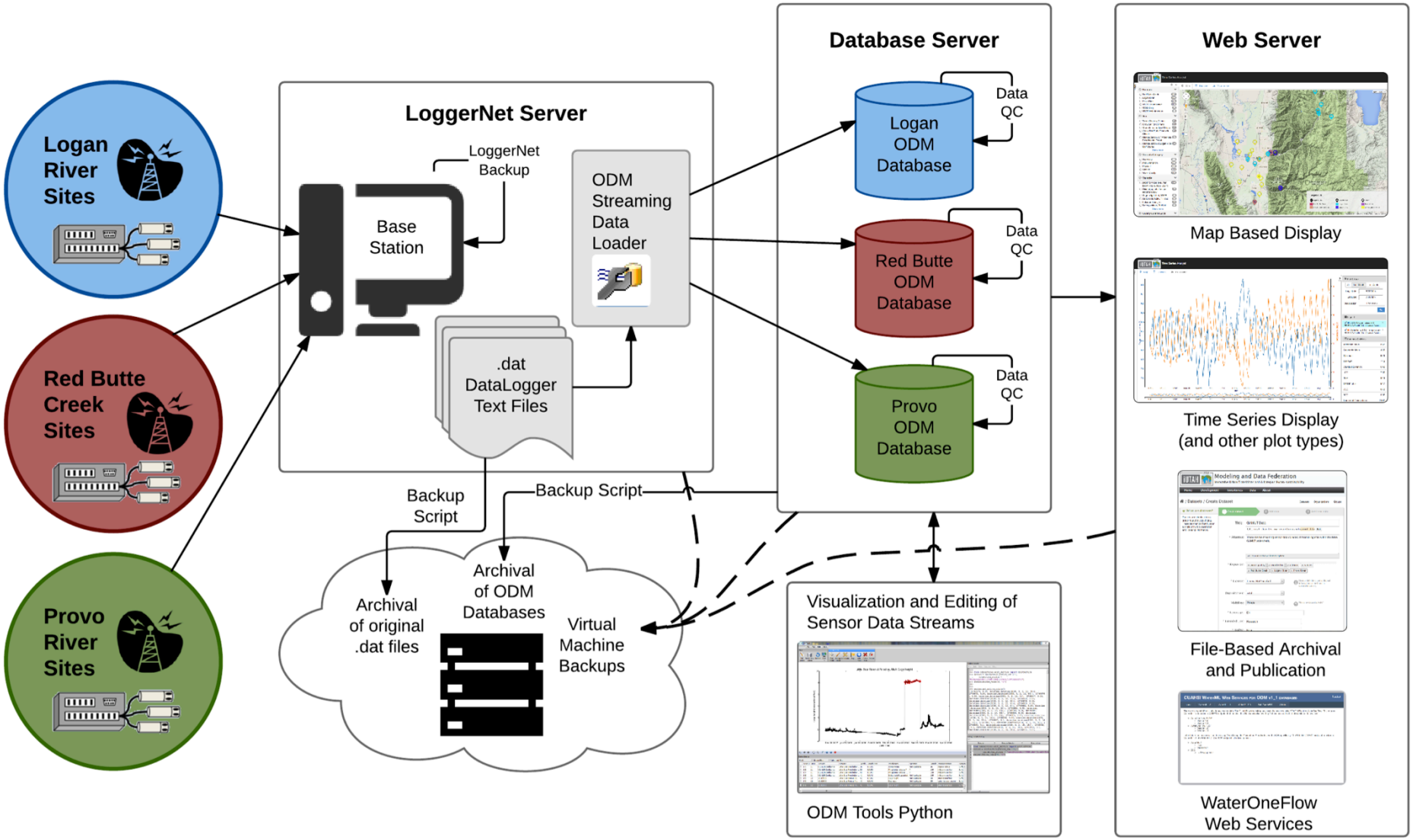


Horsburgh, J. S., A. Spackman Jones, D. G. Tarboton, D. K. Stevens, and N. O. Mesner (2010), A sensor network for high frequency estimation of water quality constituent fluxes using surrogates, *Environmental Modelling & Software*, 25, 1031-1044, doi:10.1016/j.envsoft.2009.10.012.

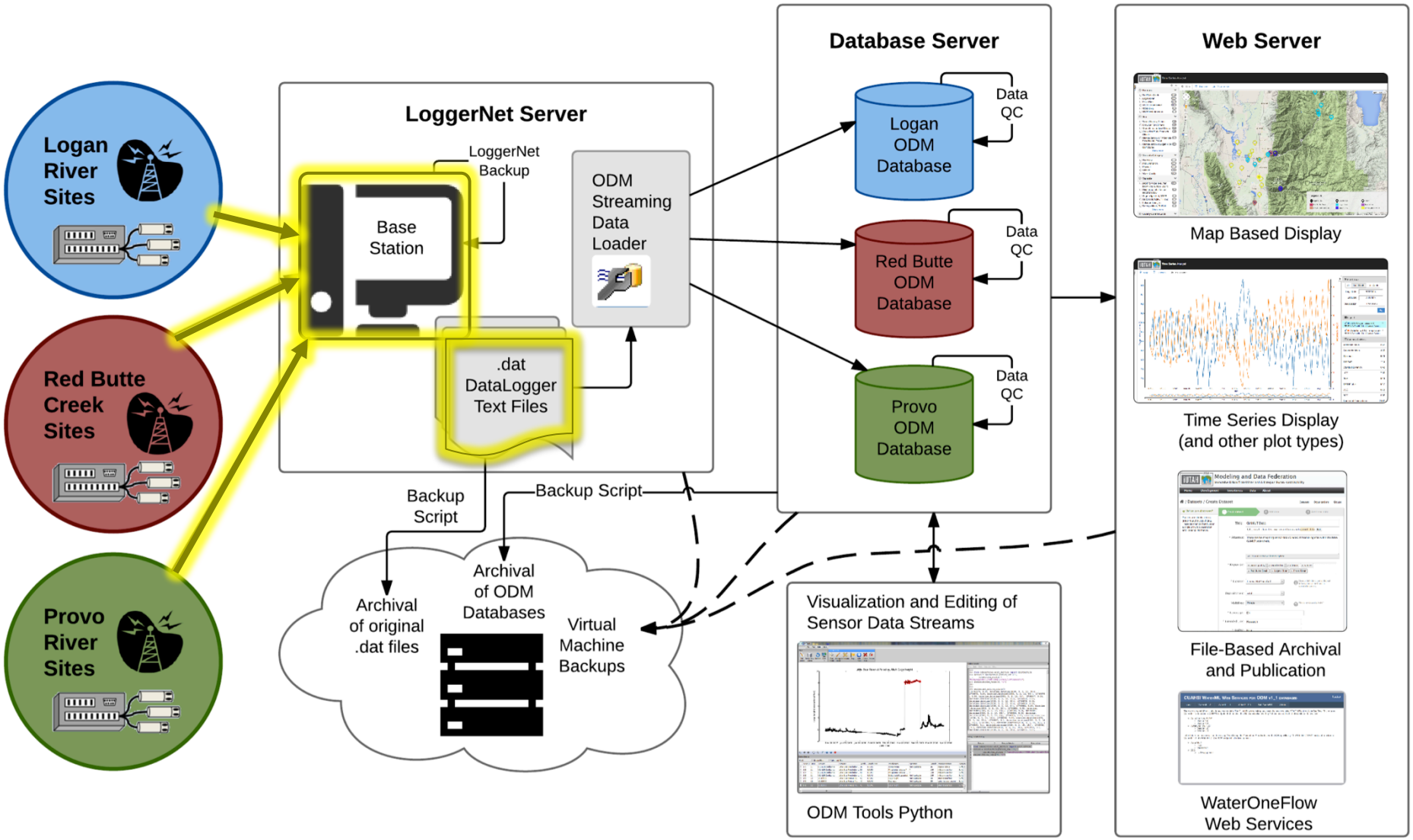
ODM2: Observation Data Model v. 2



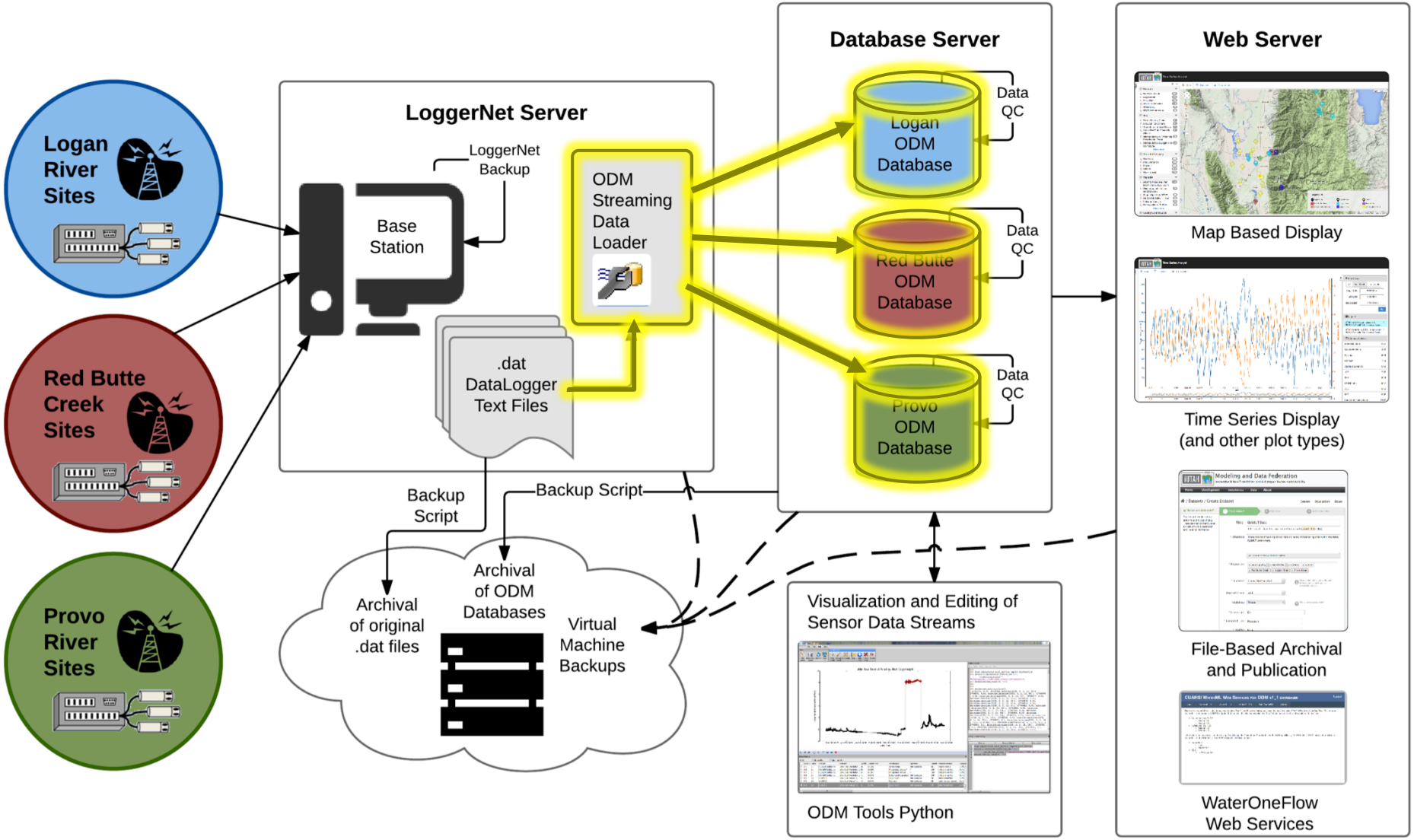
GAMUT Data Workflow



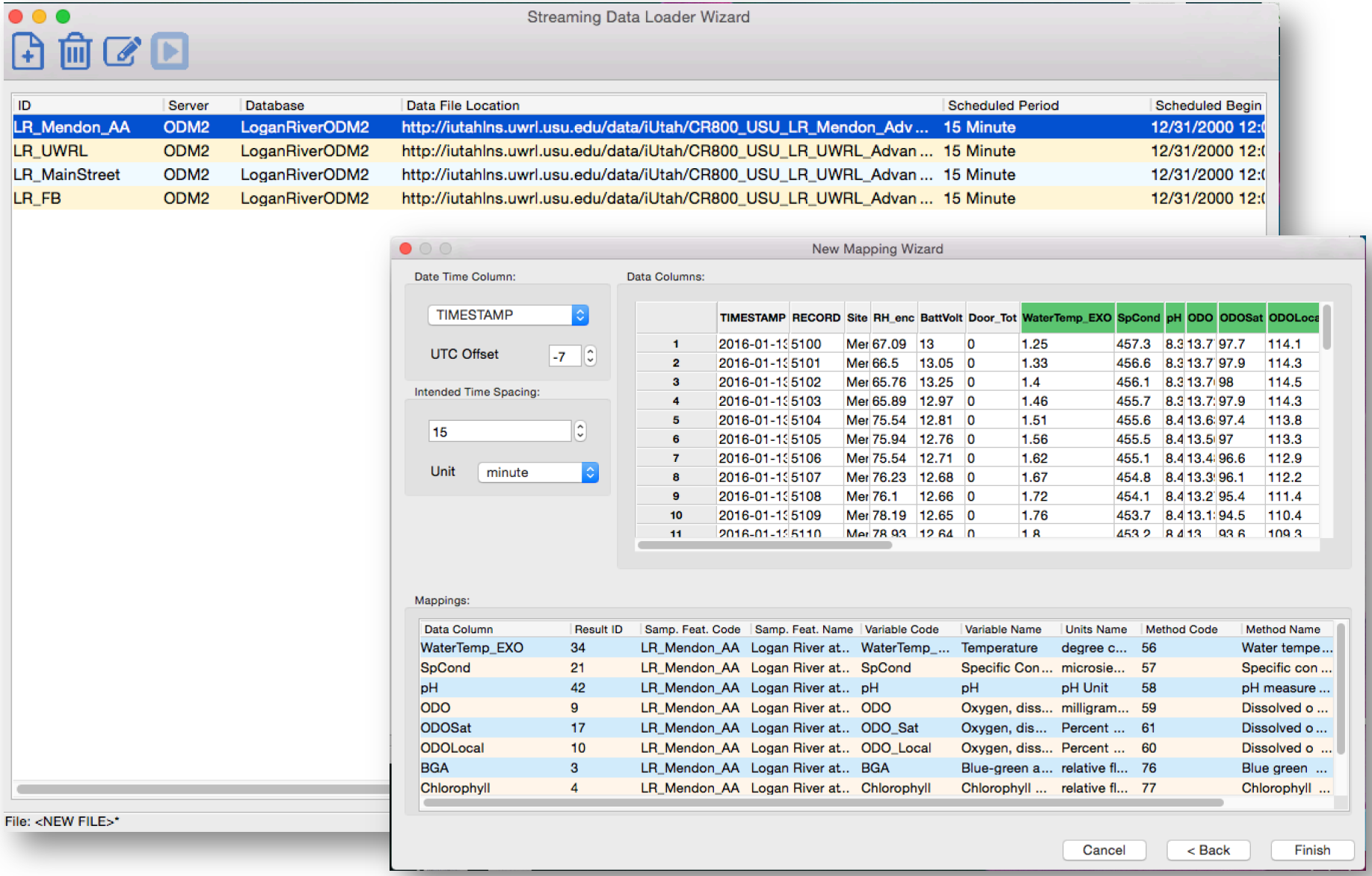
Sensor Data Acquisition



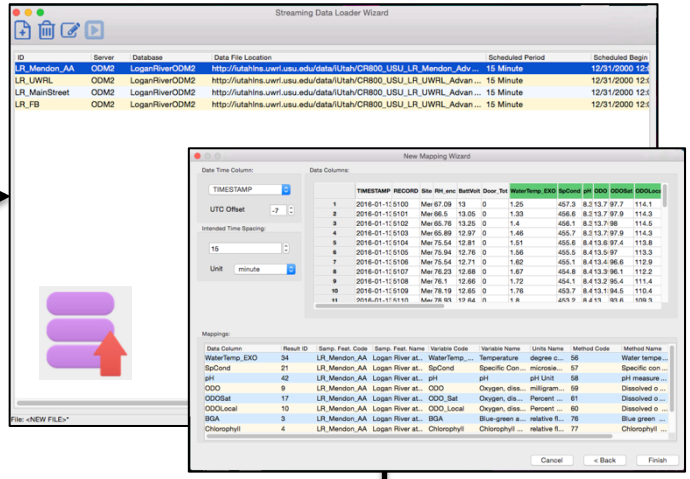
Data Loading and Storage

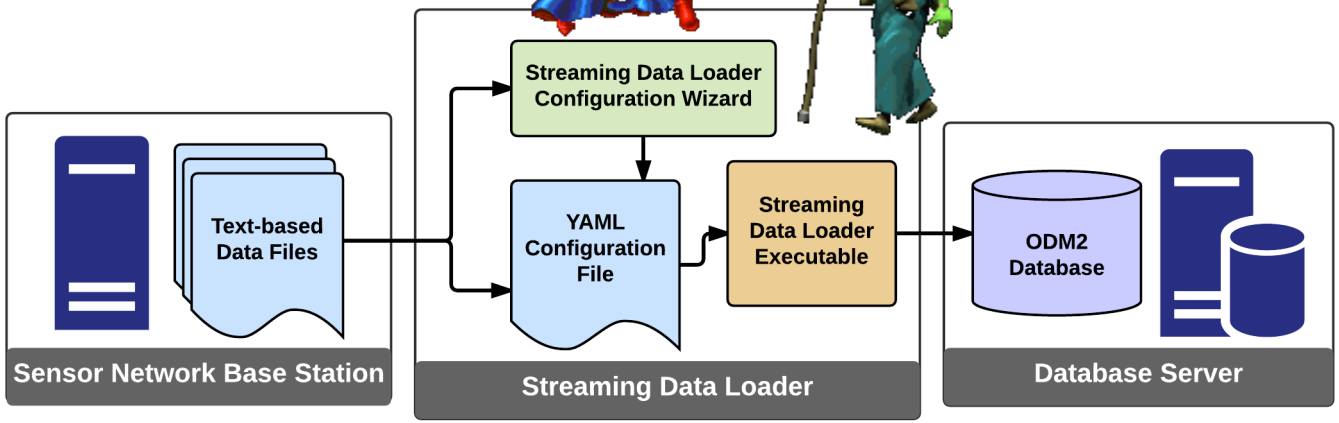


Data Loading and Storage

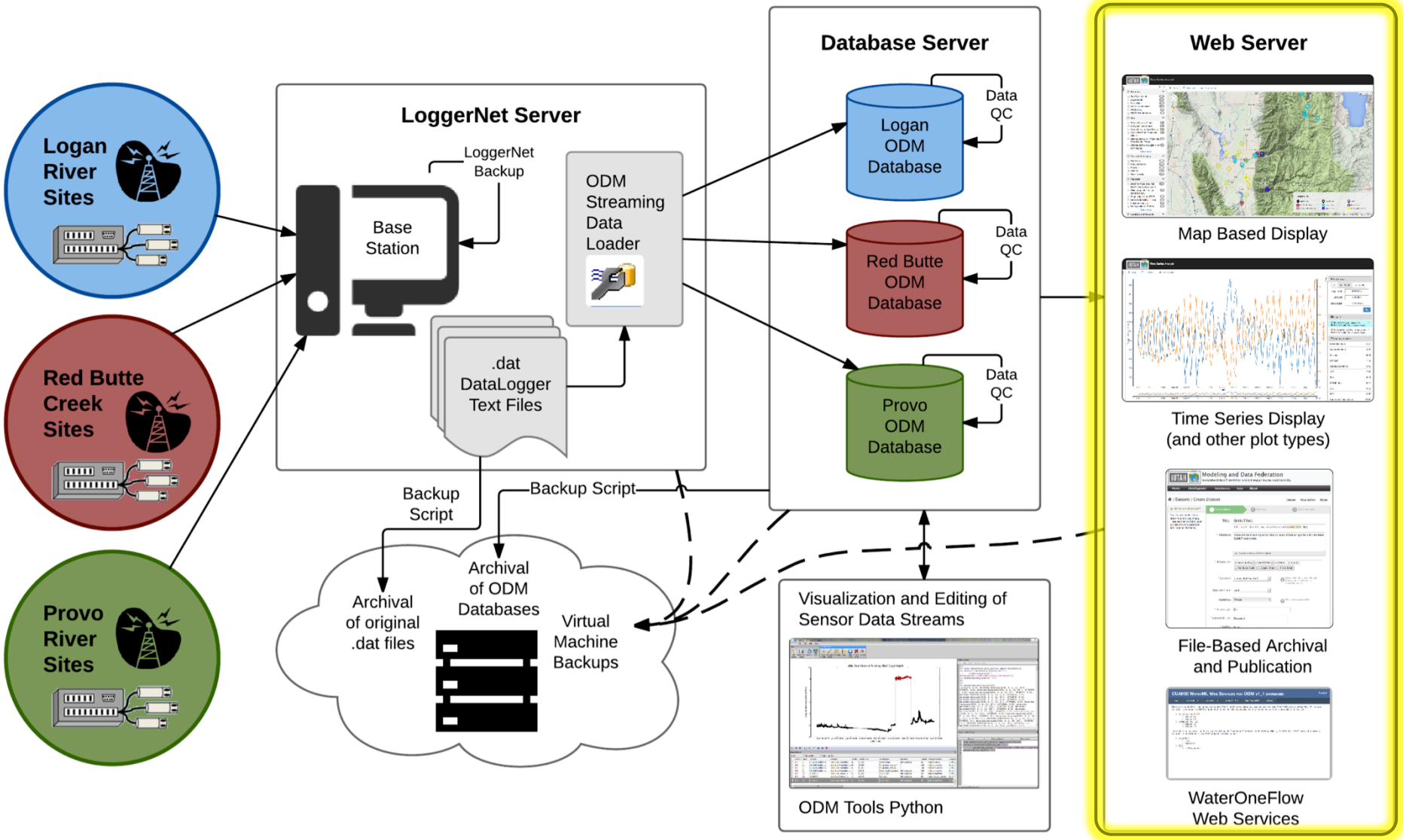


Data Loading and Storage

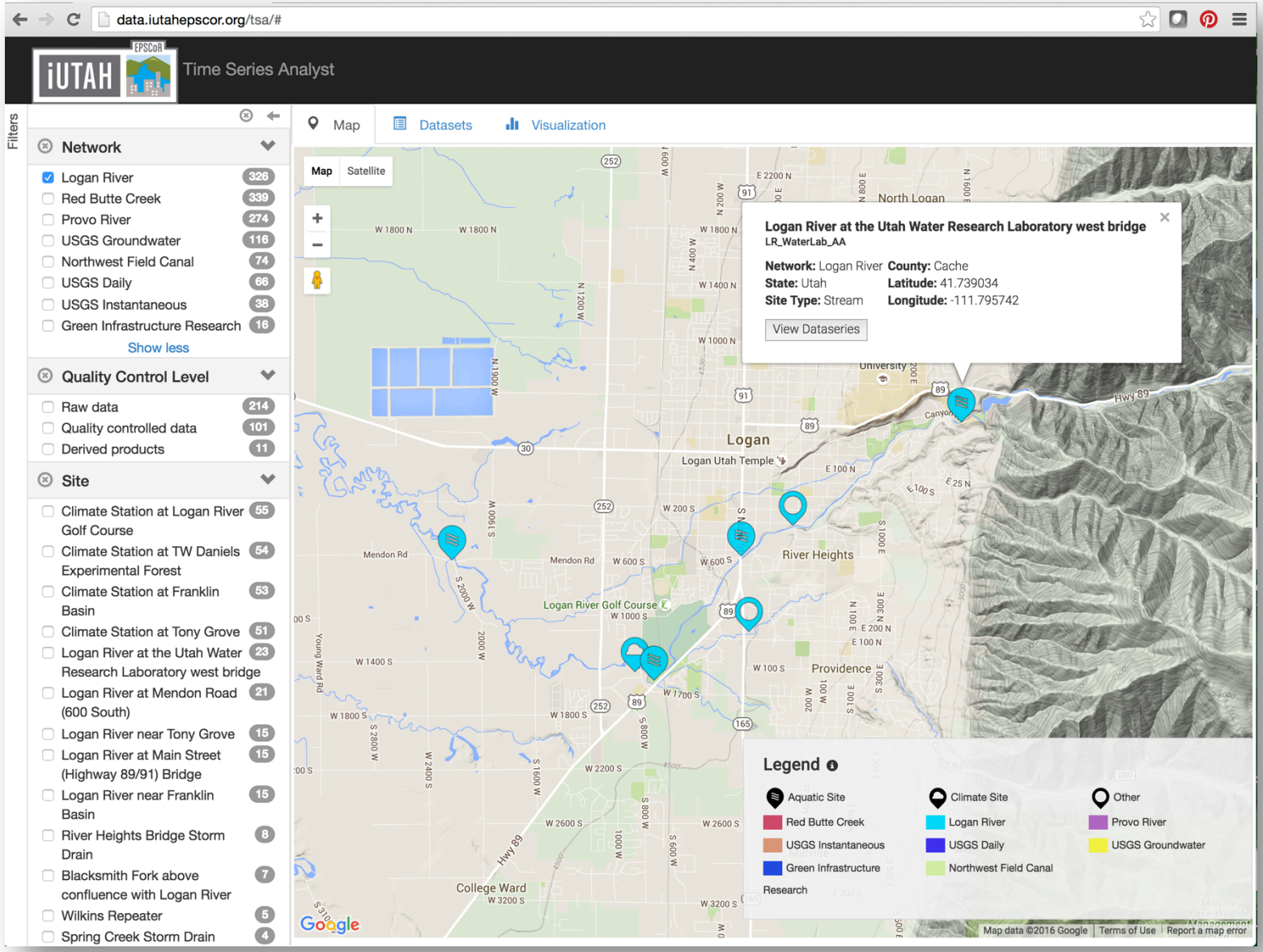




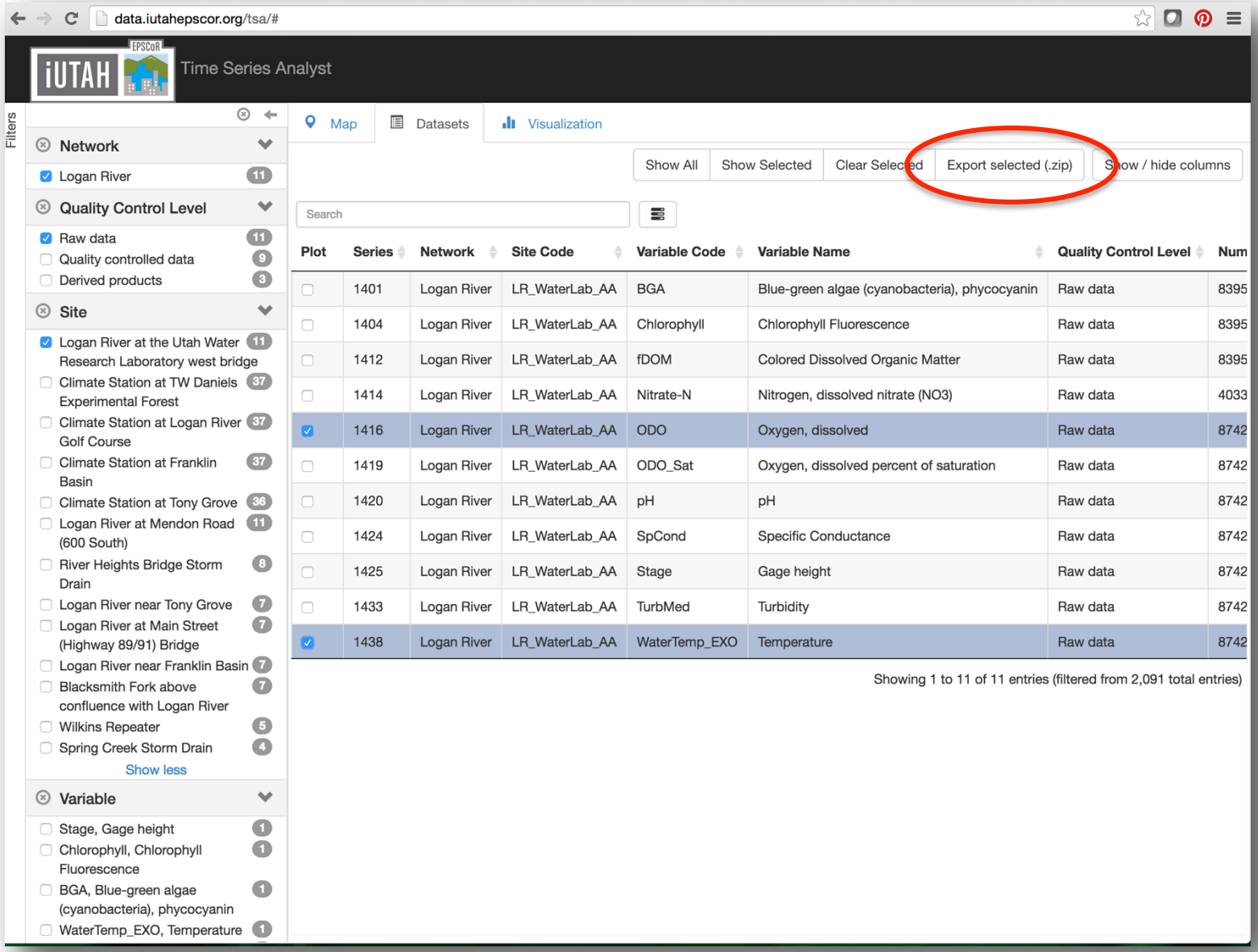
Web-Based Data Access



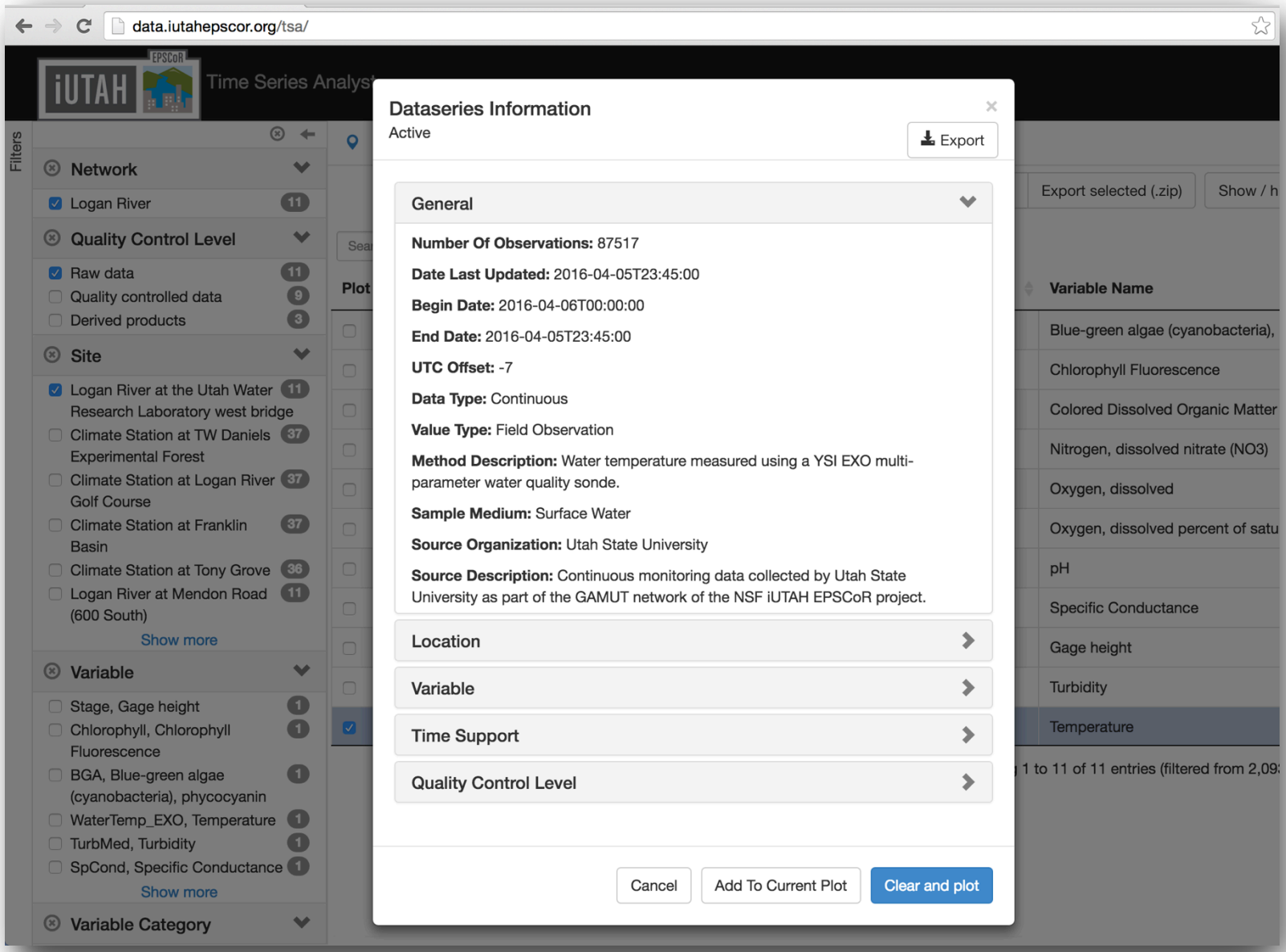
Web-Based Time Series Data Access and Visualization



Web-Based Time Series Data Access and Visualization

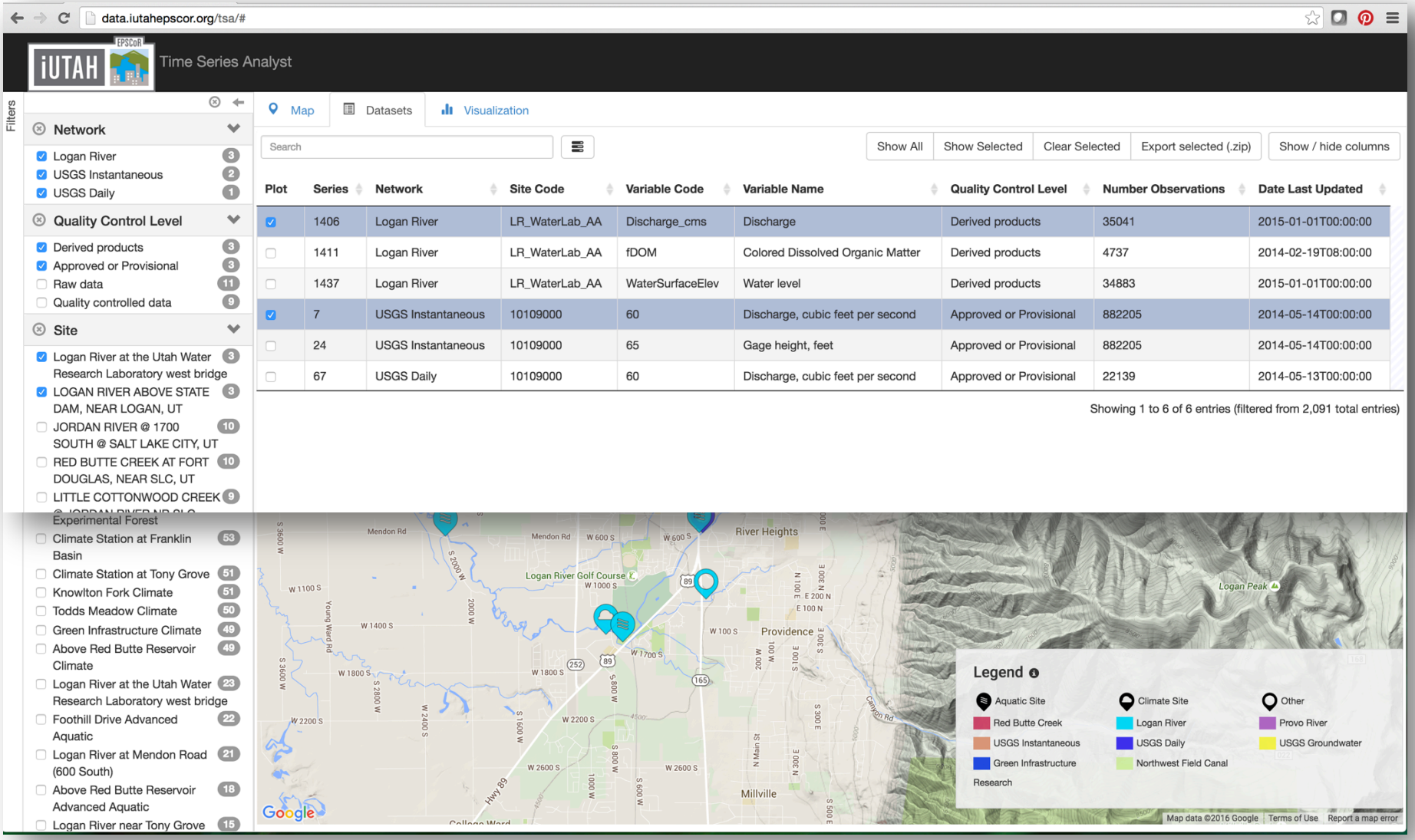


Web-Based Time Series Data Access and Visualization



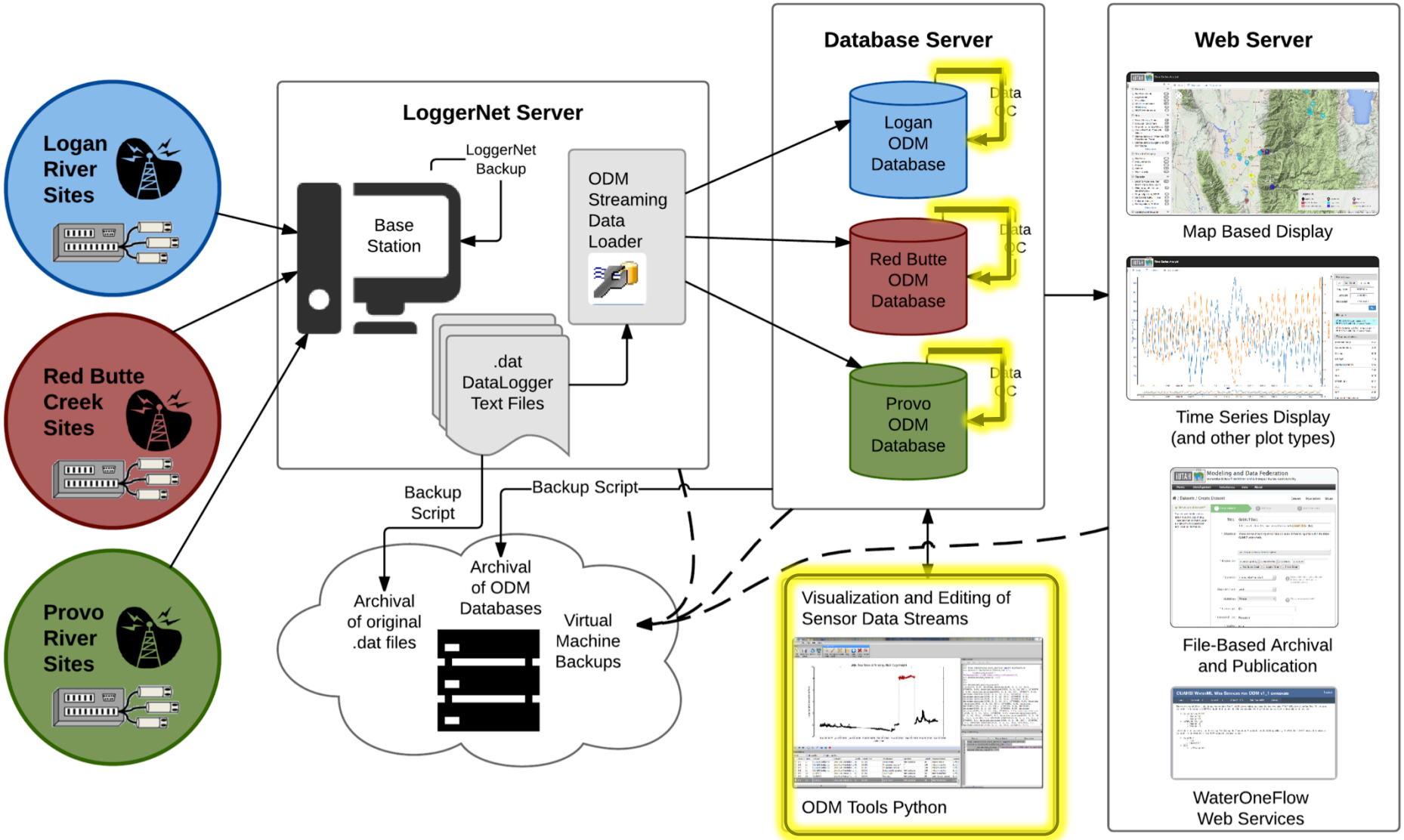
Web-Based Time Series Data Access and Visualization

Web-Based Time Series Data Access and Visualization

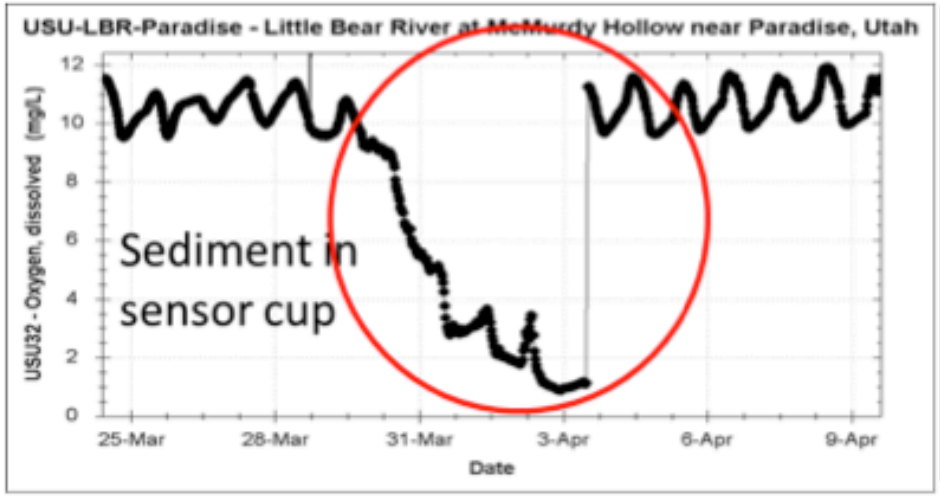


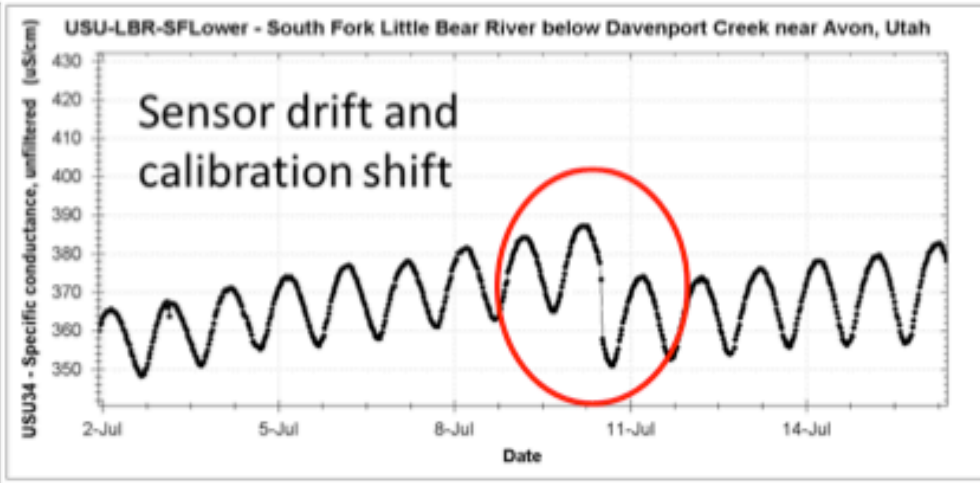
Web-Based Time Series Data Access and Visualization

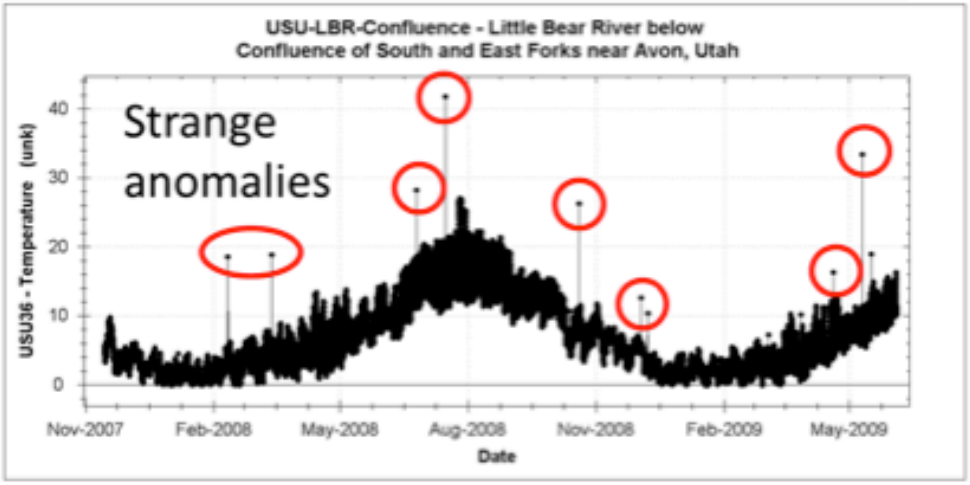
Sensor Data QAQC



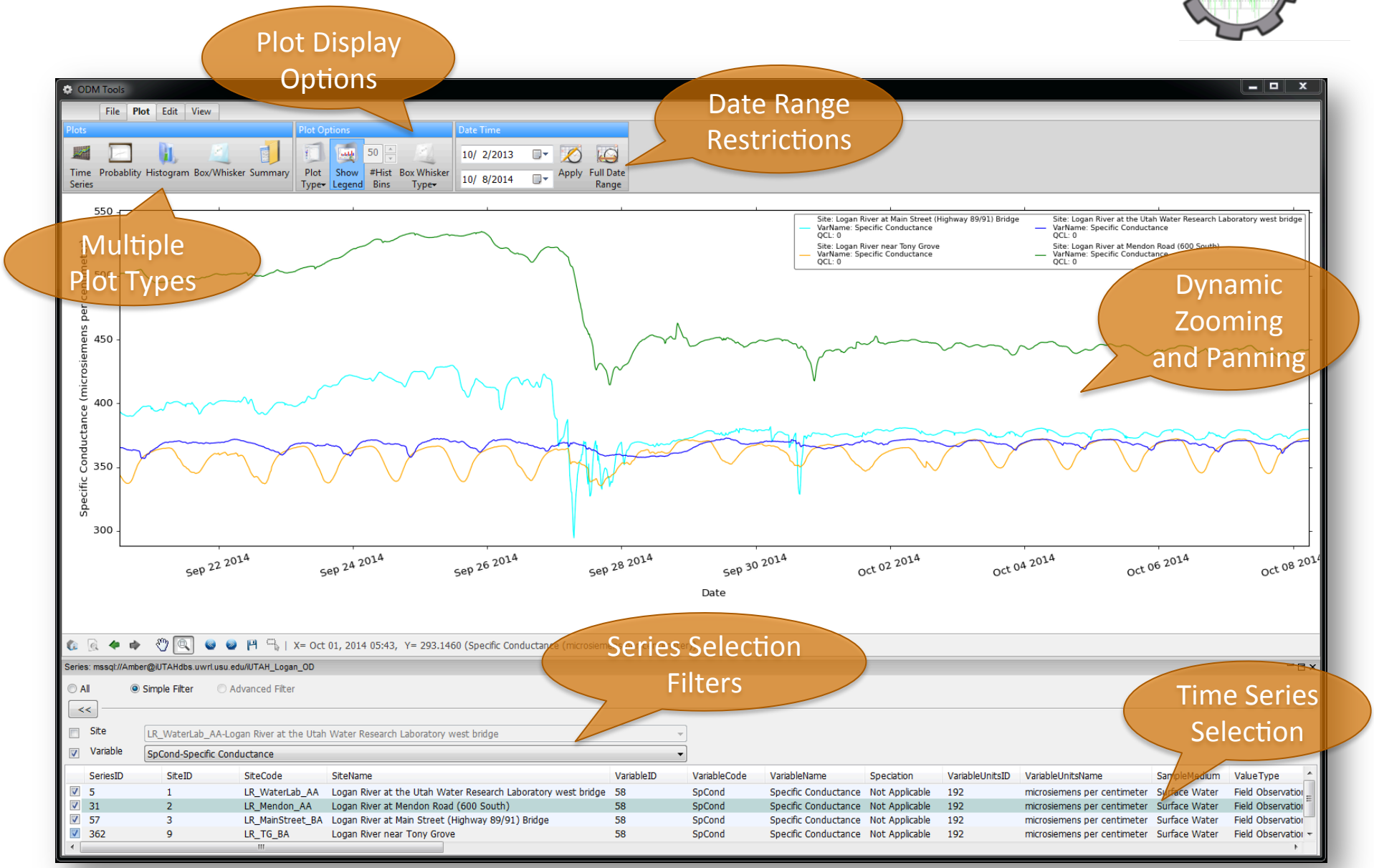
Sensor Data QAQC



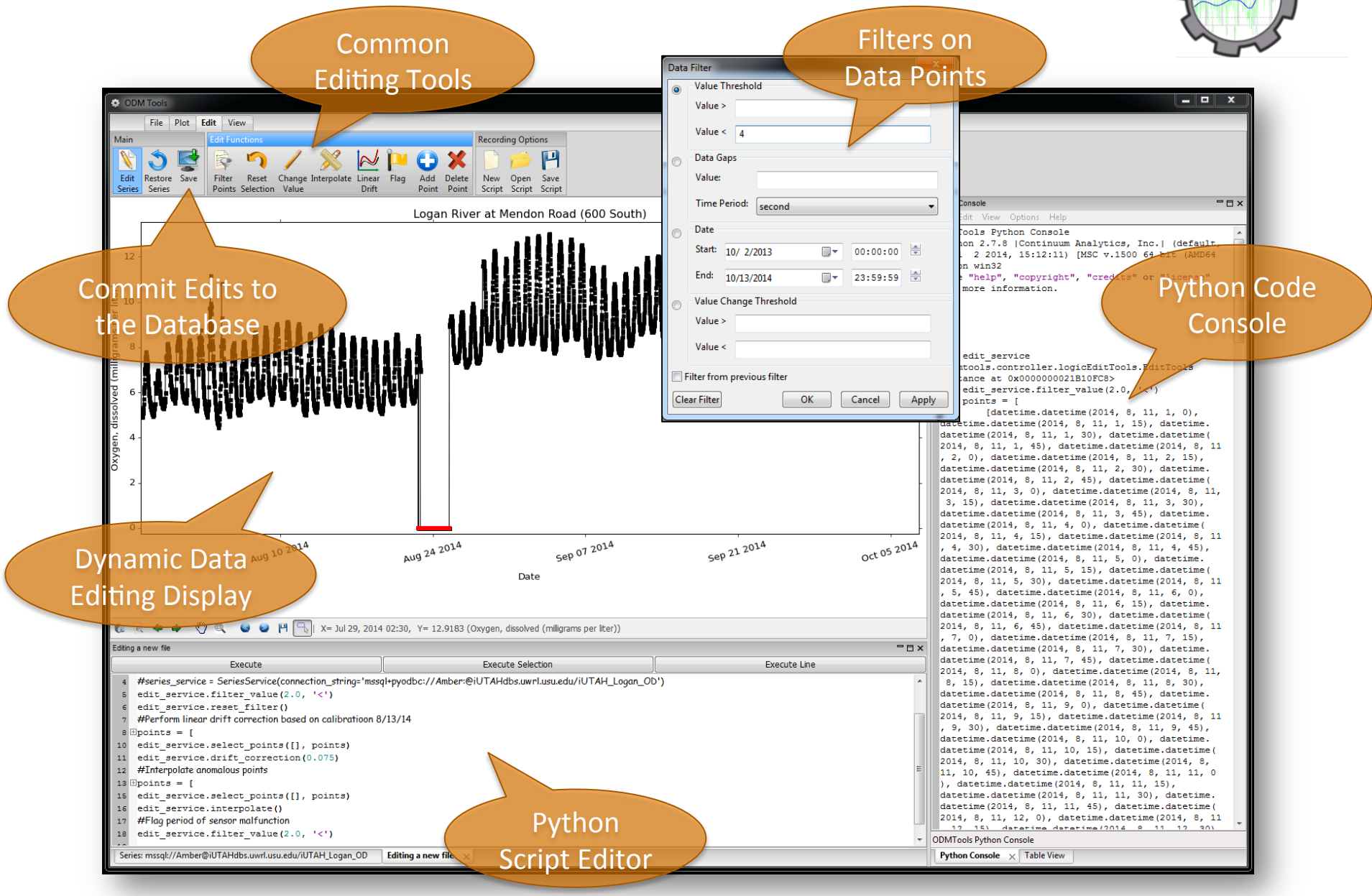




Sensor Data QAQC



QAQC: Post Processing



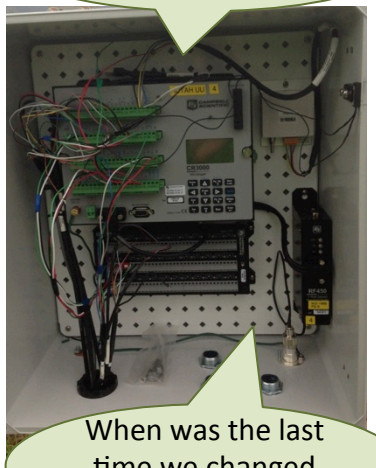
Managing Physical Infrastructure

Has this turbidity sensor been serviced at the factory?

When was the last time we cleaned this sonde?

What were the field conditions of our discharge measurement?

How did the other technician store calibration files?



Who programmed this datalogger?

When was the last time we changed desiccant?

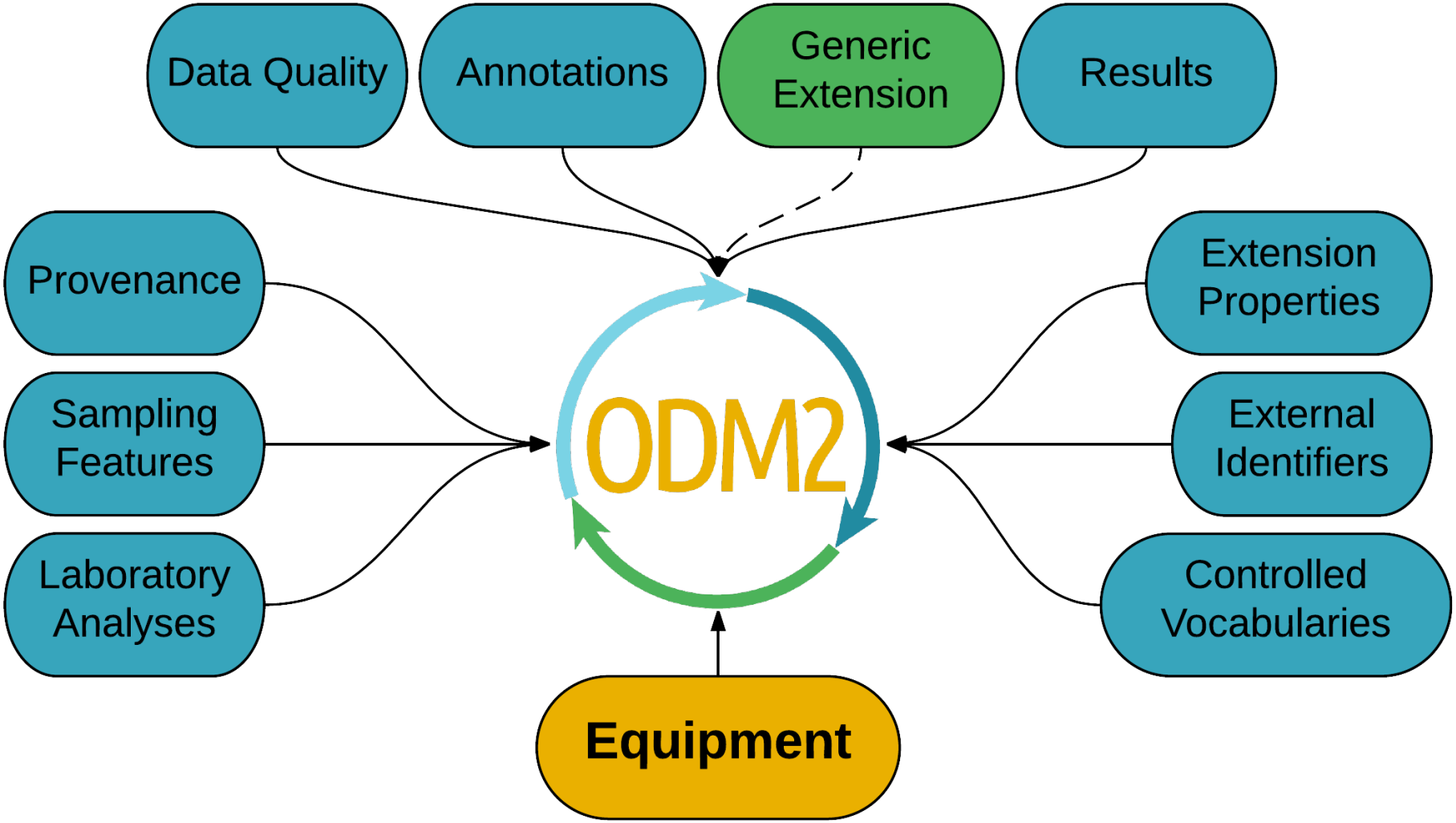
Which standards did we use for that calibration?

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

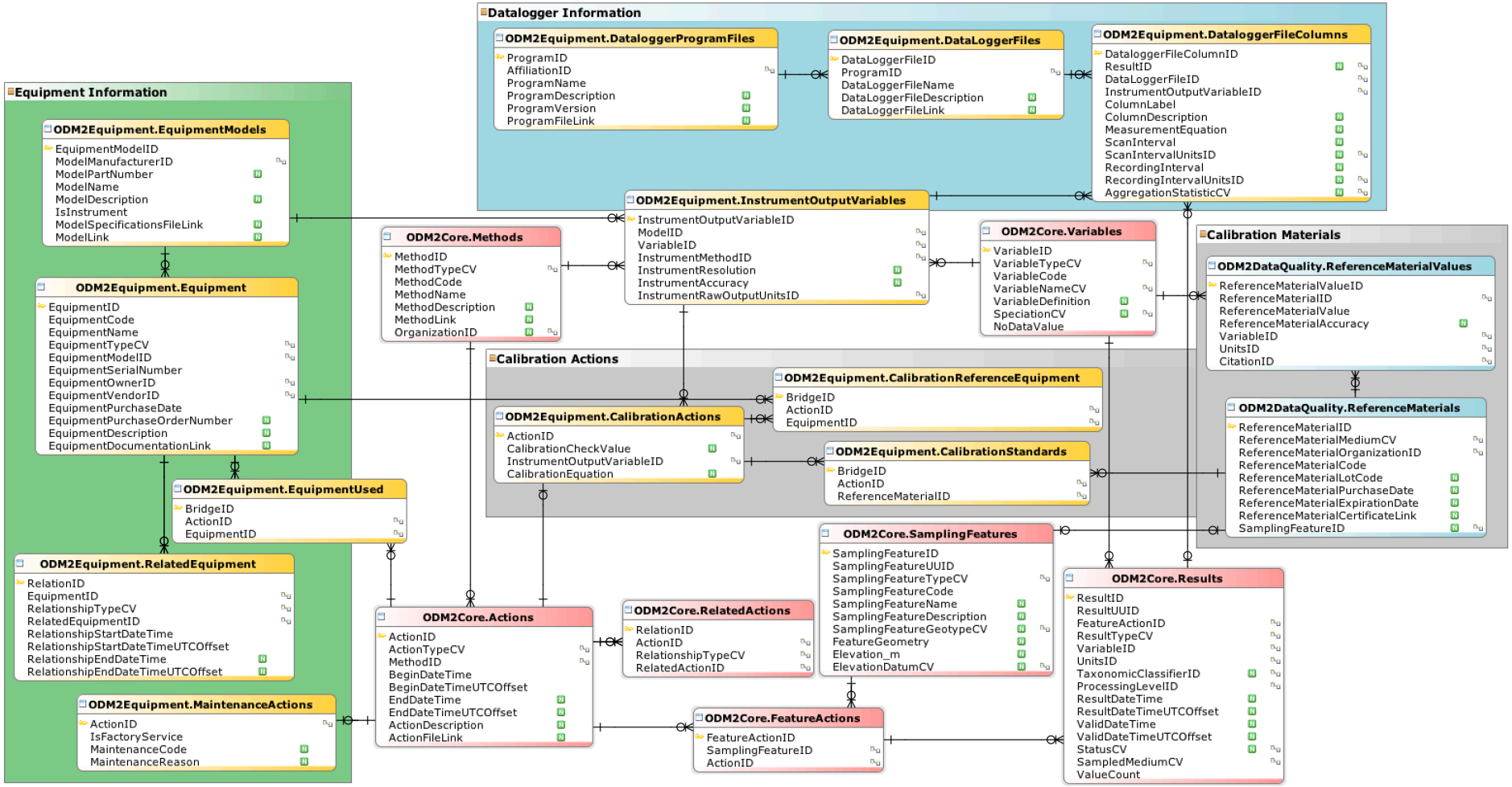
What is the calibration history of this fDOM sensor?

How long has that battery been deployed?

Managing Physical Infrastructure



Managing Physical Infrastructure



Managing Physical Infrastructure

Create Site Visit

Site: Knowlton Fork Basic Aquatic *

Begin Date Time: 2015-12-01 14:00 *

Begin UTC Offset: -7:00 MST *

End Date Time: 2015-12-01 15:30

End UTC Offset: -7:00 MST

Description: Cleaned sensors, calibrated aquatic sensors.

Crew: ×University of Utah: Dave Eiriksson ×University of Utah: Steven Hall *

- Remove Action

Action Type: Instrument calibration *

Begin Date Time: 2015-12-01 14:00 *

Begin UTC Offset: -7:00 MST *

End Date Time: 2015-12-01 14:30

End UTC Offset: -7:00 MST

Description: Calibration of pH sensor

Action File: Choose File No file chosen

Method: Calibration performed to pH reference standards of 7 and 10 *

Equipment Used: ×617: 12L100803 (Sensor, EXO2pH) *

Calibration Standards: ×Liquid aqueous : 1 3AH451 ×Liquid aqueous : 2 3AH355

Reference Equipment:

Instrument Output Variable: EXO2pH: pH

Calibration Check Value: 7.1

Calibration Equation:

+ Add Action

Cancel Create

iUTAH Home Sites Actions Inventory People Vocabularies Amber Jones Logout

Site Visit Details

Edit Site Visit Description | View Full Visit History for Site | Delete

Location

Site Code: LR_FB_BA

Site Name: Logan River near Franklin Basin

Description

People: Chris Cox

Begin Date Time: Aug. 14, 2014, 9:56 a.m.

End Date Time: Aug. 14, 2014, 11:47 a.m.

UTC Offset: -7

Action Description: Purpose: Cleaning and calibration, change pH tip
Type: Planned, regular

Related Actions

Add New Calibration | Add New Deployment | Add Another Action | Add New Site Visit

Show 50 entries Search:

Action Type	Begin Date Time	End Date Time	Description
Equipment maintenance	Aug. 14, 2014, 10:13 a.m.	Aug. 14, 2014, 11:01 a.m.	None
Equipment retrieval	Aug. 15, 2014, 11:16 p.m.	Aug. 15, 2014, 11:17 p.m.	None
Instrument calibration	Aug. 14, 2014, 11:23 a.m.	Aug. 14, 2014, 11:40 a.m.	None
Instrument calibration	Aug. 14, 2014, 11:23 a.m.	Aug. 14, 2014, 11:40 a.m.	None
Instrument calibration	Aug. 14, 2014, 11:23 a.m.	Aug. 14, 2014, 11:40 a.m.	None
Instrument deployment	Aug. 14, 2014, 11:16 a.m.	present	None

Showing 1 to 6 of 6 entries Previous 1 Next

← Previous Next→

iUTAH Home Sites Actions Inventory People Vocabularies Amber Jones Logout

Deployments for Todds Meadow Climate

New Deployment

Show 50 entries Search:

Date	Site Code	Site Name	Site Group	Equipment Serial Number	Equipment Type	Manufacturer Name	Model Part Number	Edit/Delete
Jan. 6, 2014, 3 p.m. to present	RB_TM_C	Todds Meadow Climate	Red Butte Creek	13762	Sensor	Apogee Instruments	--	
Jan. 6, 2014, 3 p.m. to present	RB_TM_C	Todds Meadow Climate	Red Butte Creek	61012159	Sensor	Rotronic	27725-6	
Jan. 6, 2014, 3 p.m. to present	RB_TM_C	Todds Meadow Climate	Red Butte Creek	67513	Sensor	Geonor	--	
Oct. 21, 2013, 3 p.m.	RB_TM_C	Todds Meadow Climate	Red Butte Creek	4373	Sensor	Apogee Instruments	--	

Managing Physical Infrastructure

iUTAH EPSCoR | Home | Sites | Actions | Inventory | People | Vocabularies | Amber Jones | Logout

Equipment

New Equipment

Show 50 entries | Search:

Equipment Serial Number	Equipment Type	Equipment Code	Equipment Model	Manufacturer Name	Owner	Edit/Delete
53253	Sensor					
53252	Sensor					
53251	Sensor					
6500934	Sensor					
6500933	Sensor					
6500932	Sensor					
6500931	Sensor					
6500930	Sensor					
6500929	Sensor					
6500928	Sensor					

iUTAH EPSCoR | Home | Sites | Actions | Inventory | People | Vocabularies

Equipment Details

Calibration History | Factory Service History | Deployment History

iUTAH EPSCoR | Home

Calibrations

Show 50 entries

Date
Sept. 5, 2014, 12:02 p.m.
Sept. 4, 2014, 1:11 p.m.
Aug. 14, 2014, 11:23 a.m.
July 30, 2014, 1:16 p.m.
July 8, 2014, 1:53 p.m.

Showing 1 to 5 of 5 entries

iUTAH EPSCoR | Home | Sites | Actions | Inventory | People | Vocabularies | Amber Jones | Logout

Person Details

Delete | Edit Person Description

Name:	Chris Cox
Organization:	Utah State University
Address:	chris.cox@usu.edu
Phone:	505-250-5885
Email:	chris.cox@usu.edu

Equipments owned by this person

Show 50 entries | Search:

Serial Number	Model	Type	Manufacturer
0651301-G-00243	BSP65-12	Solar panel	Northern Arizona Wind and Sun
0651301-G-00244	BSP65-12	Solar panel	Northern Arizona Wind and Sun
1028	ST-110	Sensor	Apogee Instruments
120519010198	SPM020P	Solar panel	Northern Arizona Wind and Sun
120519010205	SPM020P	Solar panel	Northern Arizona Wind and Sun
120519010206	SPM020P	Solar panel	Northern Arizona Wind and Sun
120519010208	SPM020P	Solar panel	Northern Arizona Wind and Sun
120519010210	SPM020P	Solar panel	Northern Arizona Wind and Sun
120519010212	SPM020P	Solar panel	Northern Arizona Wind and Sun
120519010213	SPM020P	Solar panel	Northern Arizona Wind and Sun

Managing Physical Infrastructure

Action Type: Instrument deployment *

Begin Date Time: 2015-12-01 14:00

Begin UTC Offset: -7:00 MST

End Date Time: 2015-12-01 14:00

End UTC Offset: -7:00 MST

Description: Deployed DO sensor

Action File: Choose File No file chosen

Method: Choose a Method

Equipment Used: ×578: 12L100319 (Sensor, EXO2DOProbe)

Results

Instrument Output Variable: EXO2DOProbe: Oxygen, dissolved percent of saturation

Units: percent

Processing Level: Raw

Sampled Medium: Liquid aqueous

+ Add Result

+ Add Action

Cancel | Create

iUTAH | Home | Sites | Actions | Inventory | People | Vocabularies | Amber Jones | Logout

Results

Show 50 entries | Search:

Result Date	Site Code	Site Name	Variable Code	Variable Name	Result Type
May 24, 2013, 10:23 a.m.	RB_GIRF_C	Green Infrastructure Climate	AirTemp_Avg	Temperature	Time series coverage
May 24, 2013, 10:23 a.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 10:23 a.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 10:23 a.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate			
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate	PARIn_Avg	Radiation, incoming PAR	Time series coverage
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate	PAROut_Avg	Radiation, outgoing PAR	Time series coverage
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate	PAROut_Avg	Radiation, outgoing PAR	Time series coverage
May 24, 2013, 1 p.m.	RB_GIRF_C	Green Infrastructure Climate	TargTemp_SI111_Avg	Temperature	Time series coverage

Results Details

Begin DateTime:	May 24, 2013, 10:23 a.m.
UTC Offset:	-7
Site Code:	RB_GIRF_C
Site Name:	Green Infrastructure Climate
Variable Code:	AirTemp_Avg
Variable Name:	Temperature
Units:	degree celsius
Action Description:	None

Equipment Used

Equipment Serial Number:	61052819	**Equipment Type:**	Sensor
Equipment Code:	373	**Equipment Name:**	373
Equipment Model Name:	HC2S3	**Manufacturer Name:**	Rotronic
Owner Name:	Dave Eiriksson		

Open Source Code

- WEBTSA – Time series data visualization
  - https://github.com/UCHIC/WEBTSA
- ODM Streaming Data Loader
  - https://github.com/ODM2/ODM2StreamingDataLoader
- ODM Tools Python – Sensor Data Management
  - https://github.com/ODM2/ODMToolsPython
- ODM2 Sensor – Sensor equipment management
  - https://github.com/UCHIC/ODM2Sensor

GitHub

Conclusions

- Researchers are using *in situ* sensors at multiple sites more commonly resulting in large datasets
- Much of the data management workflow can (and should!) be automated
- Visualization access can be made broadly available
- Reproducible QC removes the requirement that only one person perform post processing and increases confidence in the process
- 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

Questions?

Utah Water Research Laboratory
UWRL

UtahState University

Amber Jones
amber.jones@usu.edu

Jeff Horsburgh
jeff.horsburgh@usu.edu

http://data.iutahepscor.org

NSF

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.