



Cyberinfrastructure to Support Large Scale, Collaborative Water Research in Utah: Critical Outcomes from the iUTAH Project



Amber Spackman Jones

Jeffery S. Horsburgh

and the iUTAH Cyberinfrastructure Team



iUTAH: innovative Urban Transitions and Aridregion HydroSustainability



- Cyberinfrastructure needed to support:
 - Interdisciplinary and cross-institutional teams
 - Diverse data collection and modeling efforts
 - The full data life cycle
- Includes storage, software, networking, computational, and human resources.



Data Life Cycle



- Cyberinfrastructure needed to support:
 - Interdisciplinary and cross-institutional teams
 - Diverse data collection and modeling efforts
 - The full data life cycle
- Includes storage, software, networking, computational, and human resources.

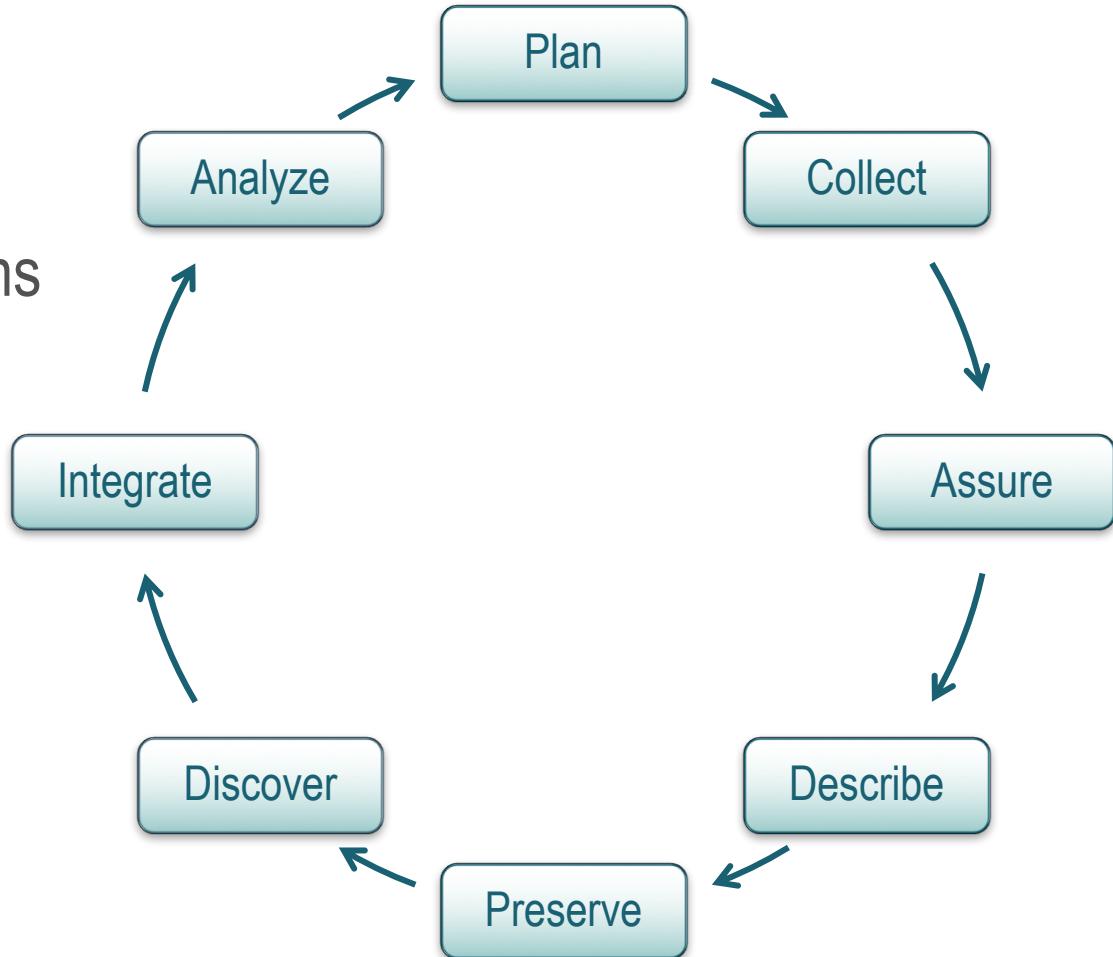


Figure courtesy DataOne

Data Life Cycle



- Cyberinfrastructure needed to support:
 - Interdisciplinary and cross-institutional teams
 - Diverse data collection and modeling efforts
 - The full data life cycle
- Includes storage, software, networking, computational, and human resources.

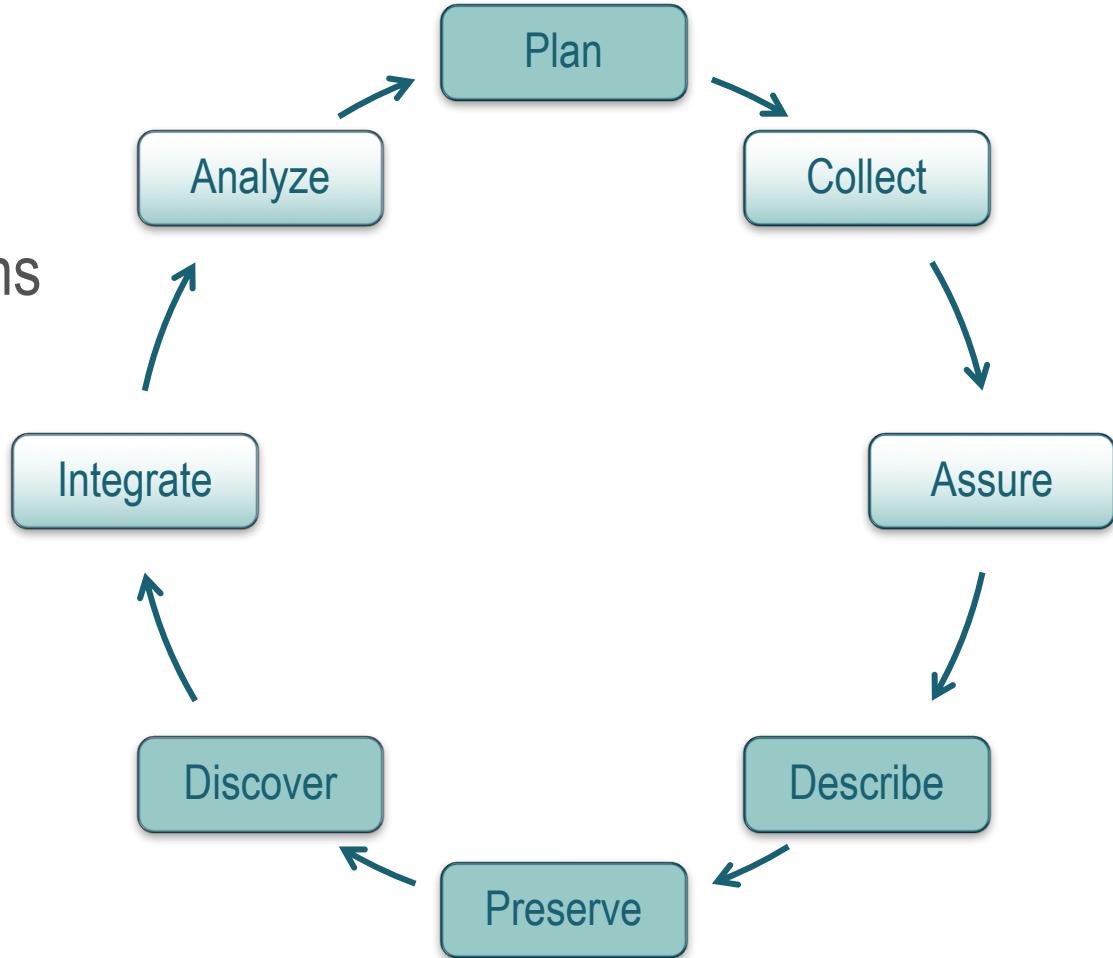


Figure courtesy DataOne

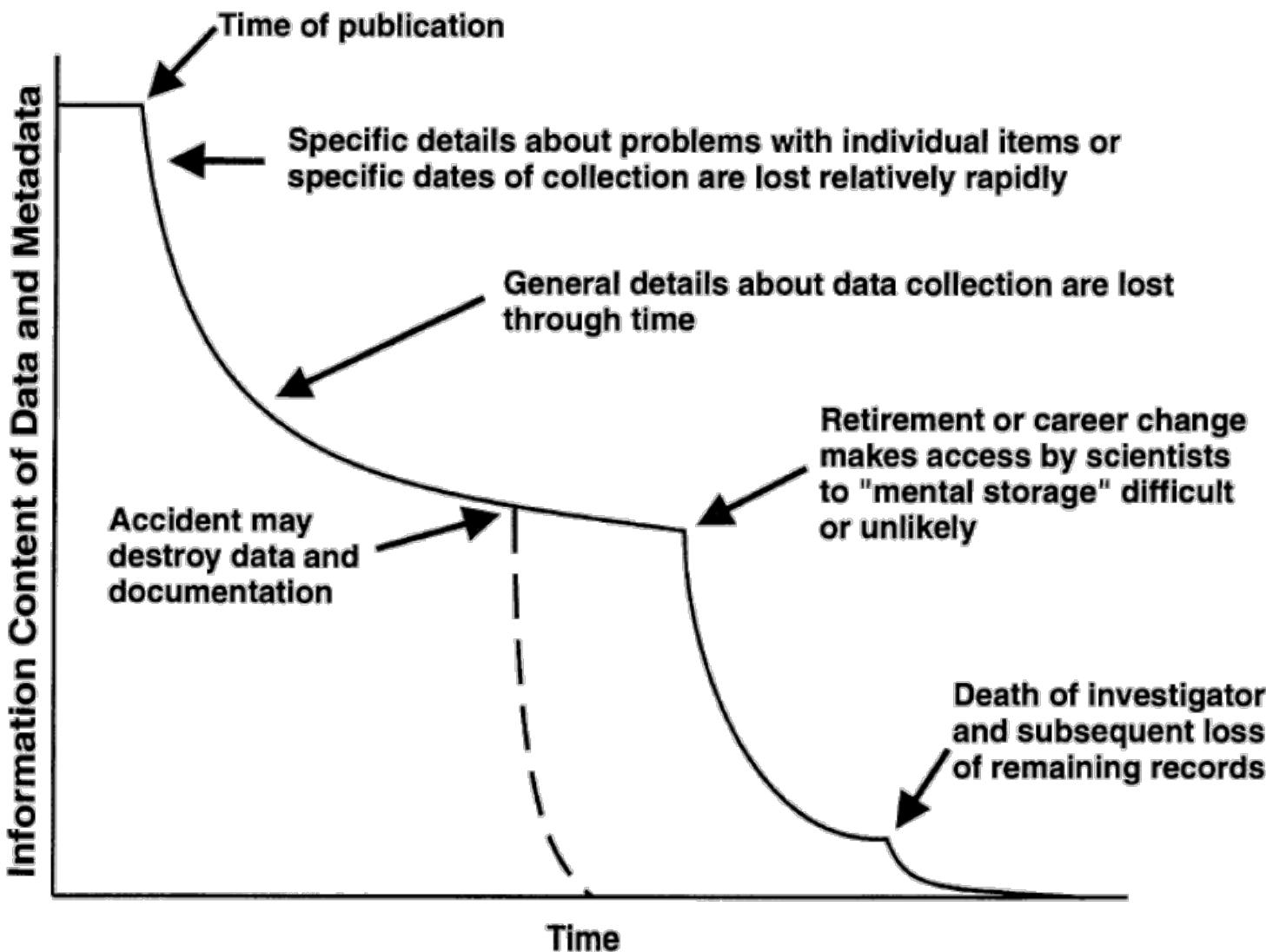
“All of the primary datasets collected as part of this project will be made freely and publicly available...”

- iUTAH Proposal Data Management Plan

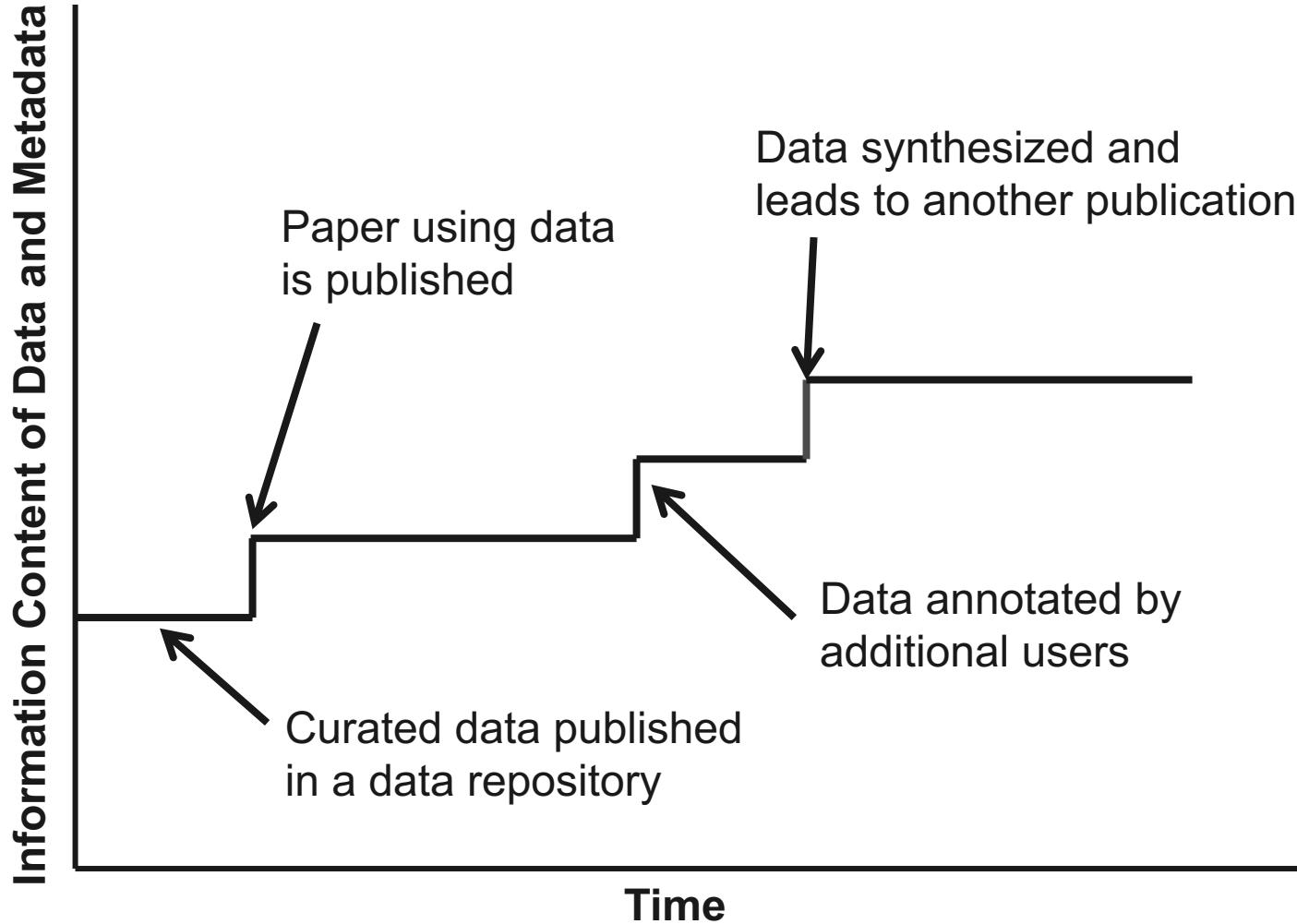
Information Entropy



Figure taken from
Michener (2006)



Information Entropy: What if instead...



Data Policy and Data Management Training

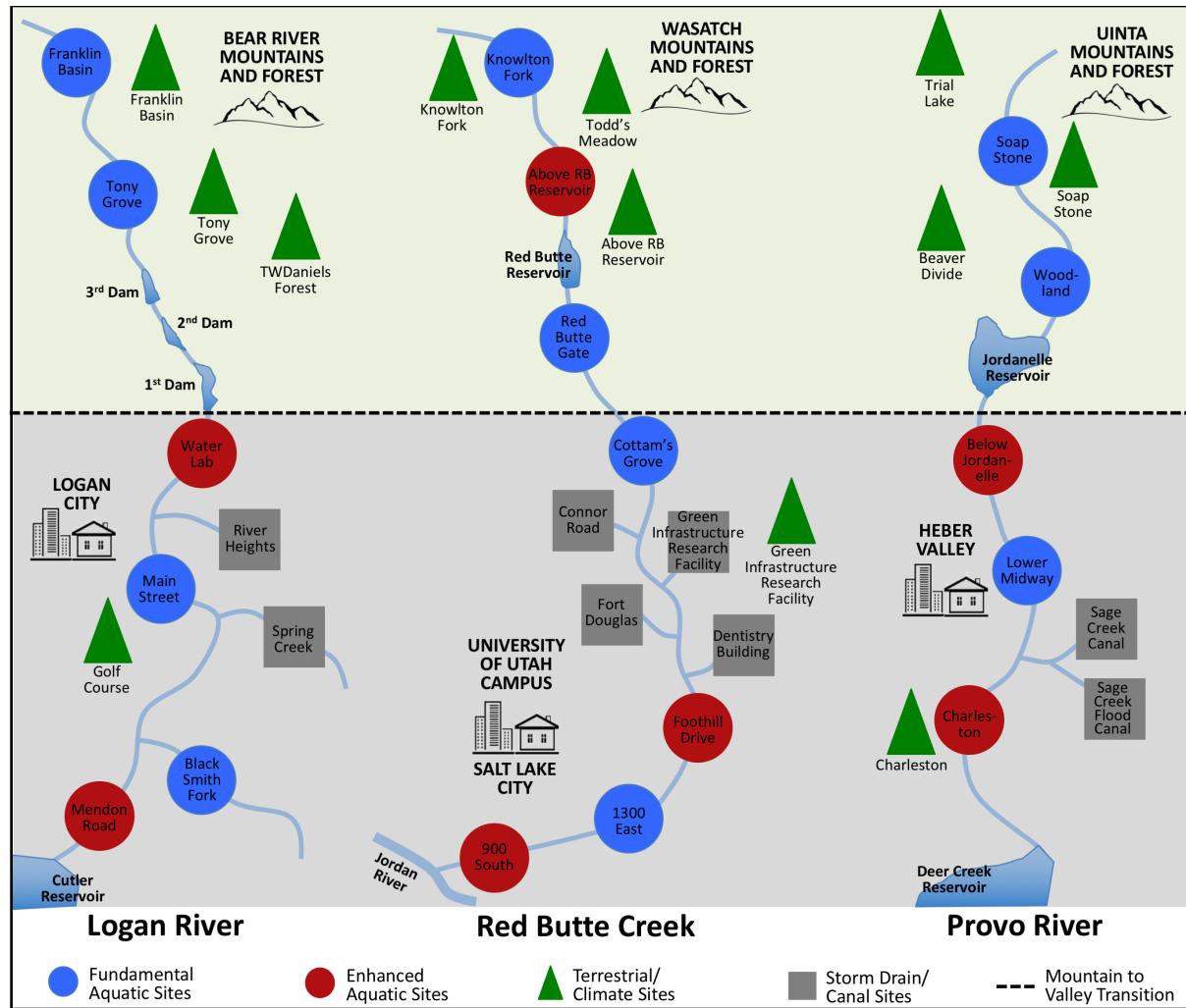


- Developed project data policy to codify guidelines and timeframes for data sharing
- Conducted training
- Developed data publication tutorials
- Review and curation of submitted datasets

The image shows a screenshot of a presentation slide. On the left, there is a logo for "iUTAH" with "EPSCoR" above it, accompanied by a graphic of green mountains and blue water. Below the logo, the title "Research Data Policy" is displayed, followed by "Version 1.6" and the date "December 8, 2014". At the bottom, it says "Edited by: Jeffery S. Horsburgh and Amber S. Jones". To the right of this main content area are two smaller, overlapping images. Both images feature a rainbow over a landscape and the text "HYDROSHARE Share and Collaborate TUTORIAL: Getting Started Creating a Resource" and "HYDROSHARE Share and Collaborate TUTORIAL: Sharing and Publishing Resources". The bottom right corner of the main slide area also contains the "HYDROSHARE" logo.

Changing the culture from “My Data” to “Our Data”

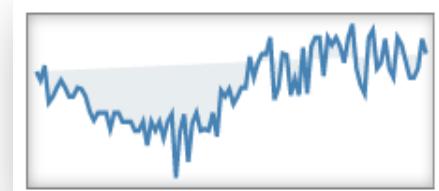
GAMUT: Gradients Along Mountain to Urban Transitions



40 sites

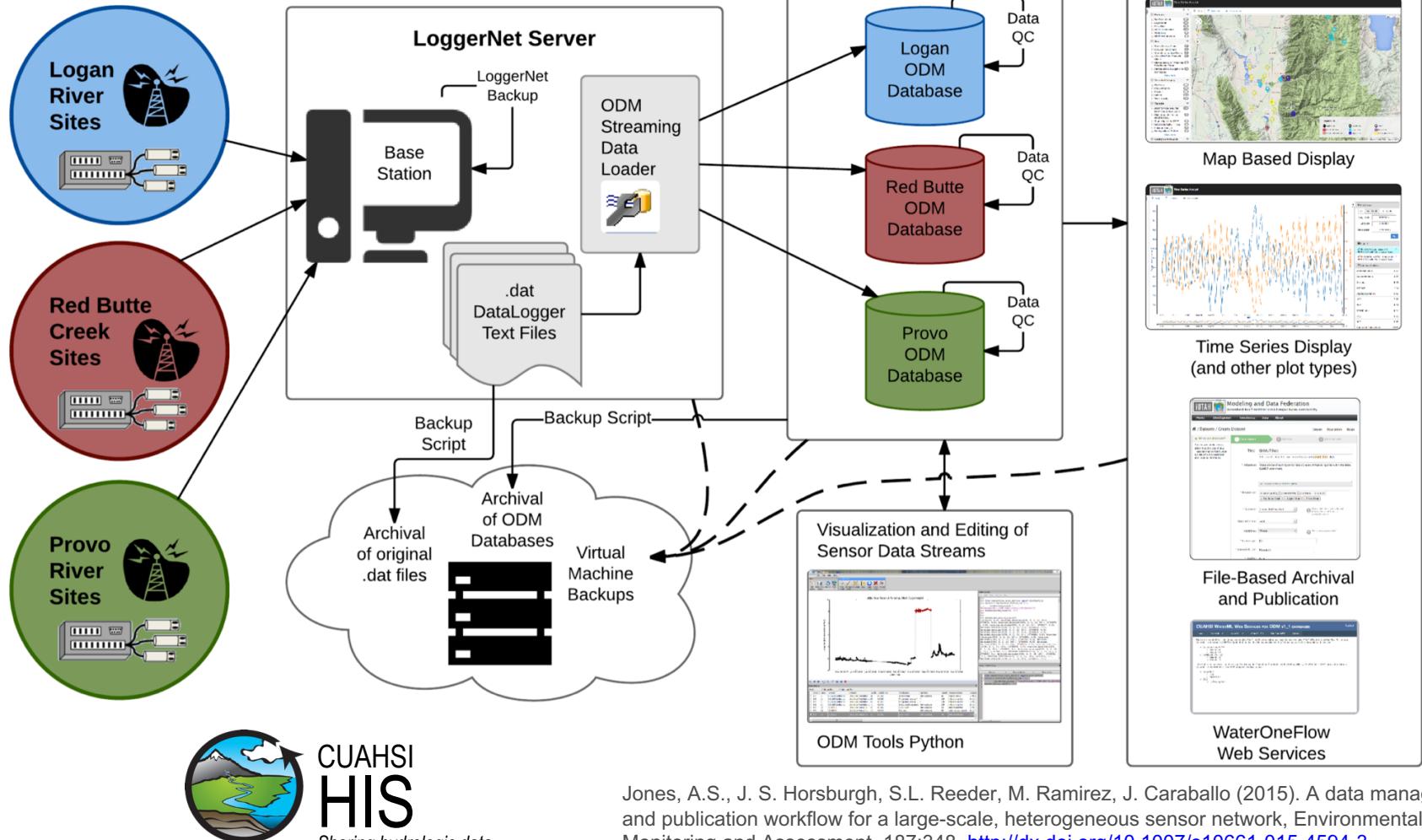


2000 data series



200 million+ values over 4 years

GAMUT: Data Work Flow



Jones, A.S., J. S. Horsburgh, S.L. Reeder, M. Ramirez, J. Caraballo (2015). A data management and publication workflow for a large-scale, heterogeneous sensor network, Environmental Monitoring and Assessment, 187:348, <http://dx.doi.org/10.1007/s10661-015-4594-3>.

GAMUT: Accessing Sensor Data



HydroShare:

Data available in flat files (csv) and updated daily

iUTAH GAMUT Network Raw Data at Todd's Meadow Climate Site (RB_TM_C)

[Open with...](#)

Authors: iUTAH GAMUT Working Group
Owners: iUTAH Data Manager
Resource type: Generic
Created: July 19, 2016, 7:24 p.m.
Last updated: Oct. 27, 2016, 9:46 a.m. by iUTAH Data Manager

Abstract

This dataset contains raw data for all of the variables measured for the iUTAH GAMUT Network climate site near Todd's Meadow (RB_TM_C). Each file contains a calendar year of data. The file for the current year is updated on a daily basis. The data values were collected by a variety of sensors at 15 minute intervals. The file header contains detailed metadata for the site and the variable and method of each column.

How to cite

Group, I. G. (2016). iUTAH GAMUT Network Raw Data at Todd's Meadow Climate Site (RB_TM_C). HydroShare, <http://www.hydroshare.org/resource/aff4e6dfc09a4070ac15a6ec0741fd02>

This resource is shared under the Creative Commons Attribution CC BY.

<http://creativecommons.org/licenses/by/4.0/>

Sharing status: Public Discoverable Private

Shareable

You are the owner of this resource.

Subject

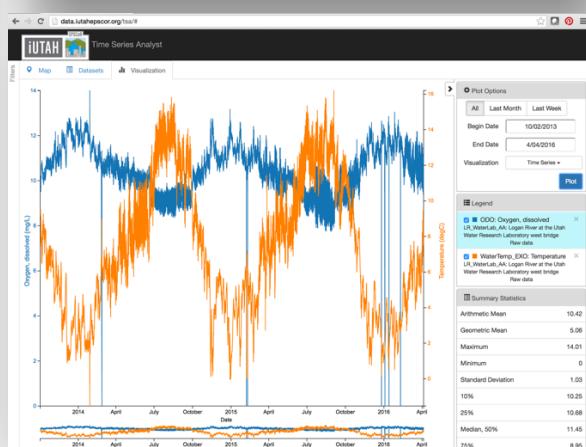
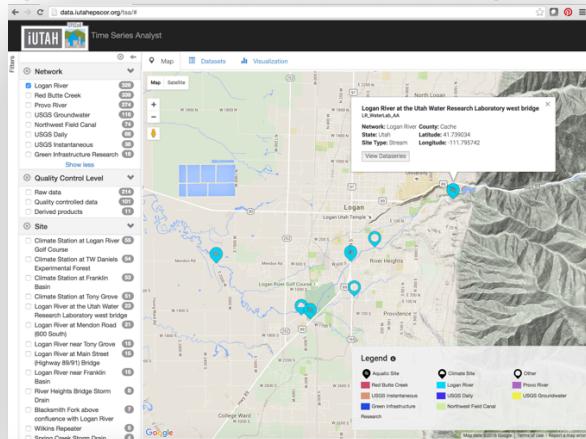
GAMUT Red Butte Creek climate precipitation raw data snow soil time series

Content

File	Size	Actions
data/contents/iUTAH_GAMUT_RB_TM_C_RawData_2015.csv	16.0 MB	
data/contents/iUTAH_GAMUT_RB_TM_C_RawData_2014.csv	18.1 MB	
data/contents/iUTAH_GAMUT_RB_TM_C_RawData_2013.csv	18.1 MB	
data/contents/iUTAH_GAMUT_RB_TM_C_RawData_2016.csv	12.9 MB	

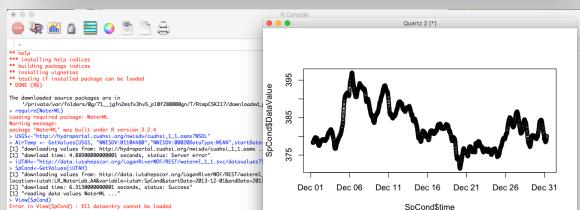
Time Series Analyst:

Web interface, map and
plot tools, and data export



Web Services:

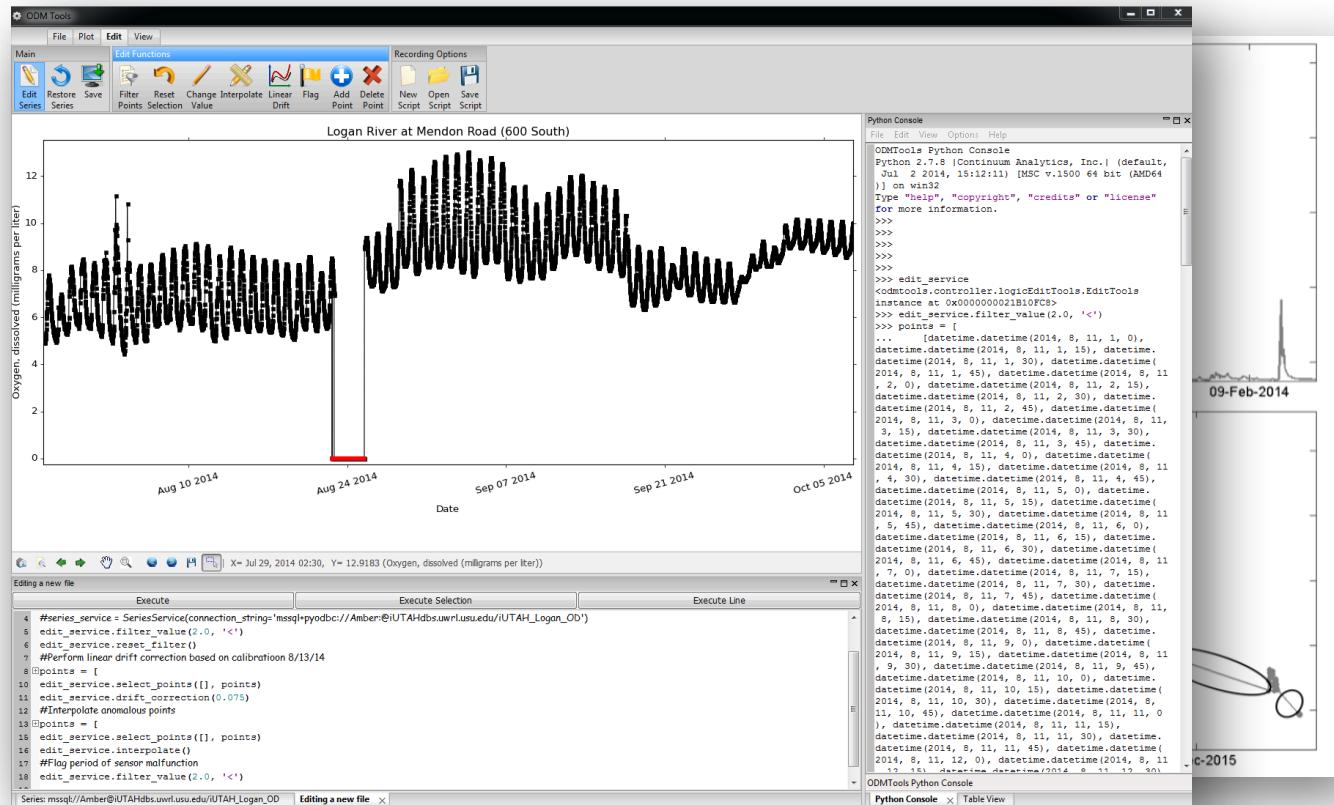
Programmatic data access



GAMUT: Quality Control of Sensor Data



- Continuous, high frequency data require post processing
- Modifications to correct for common data errors
- Sensor Drift & Calibration
- Fouling
- Power Failure
- Icing
- Anomalies



GAMUT: Grab Sample Results



- Two efforts with different variable sets:
 - Monthly sampling: 2013-2014
 - Biweekly sampling: 2014-2016
- Variables: TSS, VSS, species of Nitrogen and Phosphorus, Total Coliform, E.coli, DOC, Fluorescence, Chlorophyll-a, Isotopes, Ions



Specimen Data Viewer

Filters

NETWORK

- Some Network
- Logan River
- Red Butte Creek

SITE

- Red Butte Creek
- Water Lab
- Some Site Name

MEDIUM

- Aquatic

VARIABLE

- Oxygen, dissolved

Results

Search

Date Range

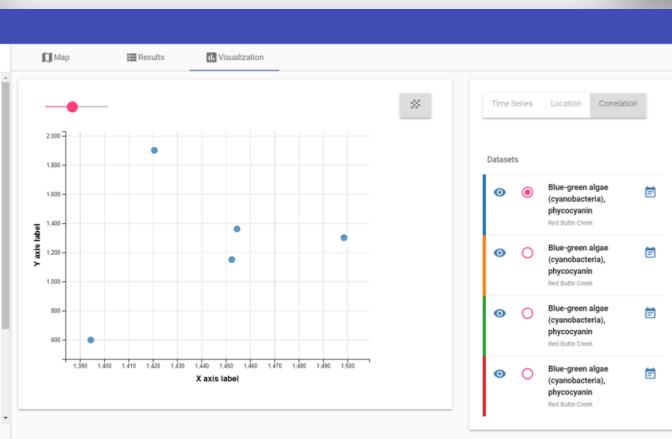
Display: All Plot Export

Plotting (3/5)

Clear Plot

Variable Code	Network	Site Code	Site Name	Variable Name	Start Date	End Date	Method
ODO_Local	SLC	ASDSASDFW	Utah River	Water Pressure	09/11/2017	09/11/2017	
ODO_Local	Logan	ASDSASDFW	Red Butte Creek	Temperature	09/11/2017	09/11/2017	
ODO_Local	Logan	ASDSASDFW	Utah River	Water Pressure	09/11/2017	09/11/2017	
ODO_Local	GAMUT	FFFCVBB	Logan River	Water Pressure	09/11/2017	09/11/2017	
ODO_Local	Logan	FFFCVBB	Red Butte Creek	Water Pressure	09/11/2017	09/11/2017	

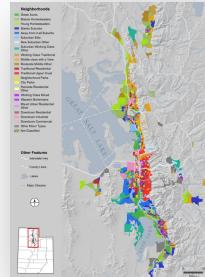
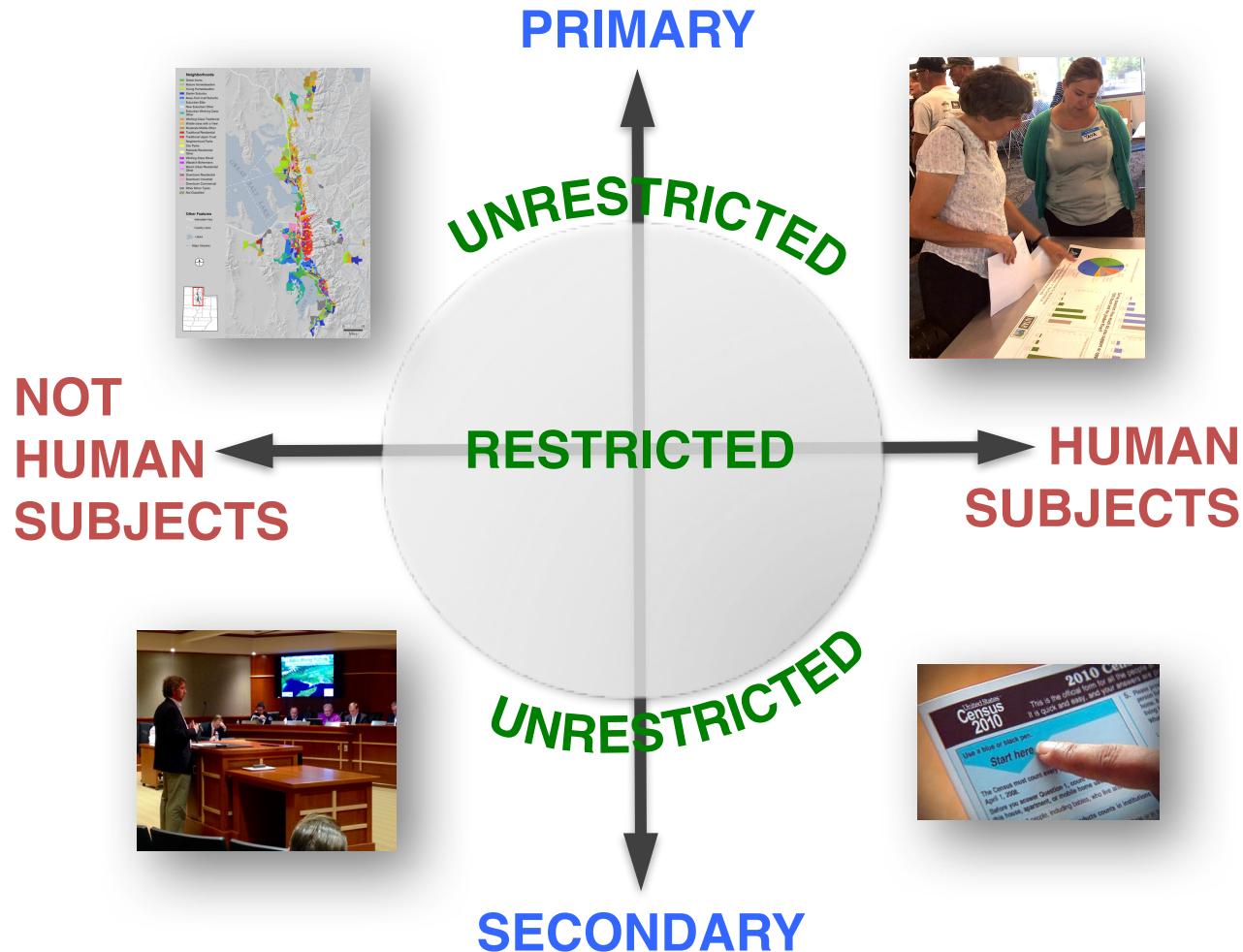
Items per page: 5 1 - 5 of 100



Social Water Science Data: Considerations



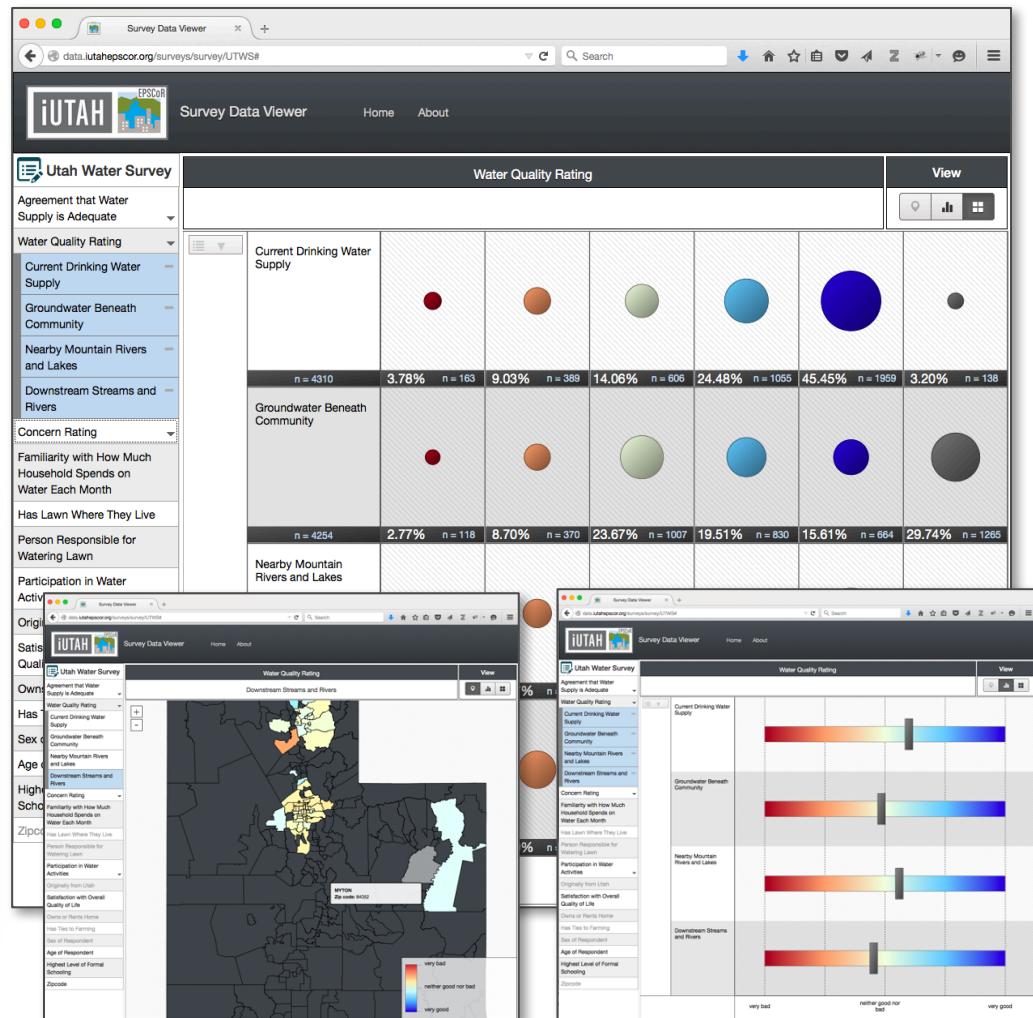
- Developed methods for categorizing social science data
- Dimensions help us understand mechanisms and restrictions for how social science data can be shared



Social Water Science Data: Visualization



- Visualization of public intercept survey results
- Generic and reusable survey template
- Open source code



<http://data.iutahpsc.org/surveys/>

Jones, A. S., Horsburgh, J. S., Jackson-Smith, D., Ramirez, M., Flint, C., Caraballo, J. (2016). A Web-based, interactive visualization tool for social environmental survey data, *Environmental Modelling & Software*, 84, 412-426, doi:10.1016/j.envsoft.2016.07.013.

Sharing Diverse Data



- An online, collaborative system for sharing and publishing various data types, models, and code
- Support for collaboration-sharing with individuals, groups, or publically
- Ability to formally publish with a DOI

The screenshot shows a dual-monitor setup. The left monitor displays the HydroShare homepage with a large search bar and navigation links for 'MY RESOURCES', 'DISCOVER', 'COLLABORATE', 'APPS', and 'HELP'. The right monitor shows a specific dataset page for 'iUTAH GAMUT Network Raw Data at Logan River at the Utah Water Research Laboratory west bridge (LR_WaterLab_AA)'. The page includes a large thumbnail image of a river scene, a summary section with a 'Share your data and models with colleagues' button, a 'What you can do' list (Share your data and models, Manage who has access, Share, access, visualize models, Use the web services API, Publish data and models plan, Discover and access data, Use web apps to visualize data), and sections for 'How it works' and 'Create data'. The right monitor also shows the browser's address bar with the URL <https://www.hydroshare.org/resource/2b3afc29e11c412c84d5c9cd9b6279d6/>.



Hydroinformatics Course



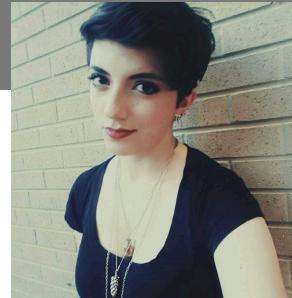
- 3 to 5 partner universities
- 30 - 45 students total across the campuses
- We focus on:
 - Data and the data life cycle
 - Databases and data models
 - Data visualization, transformation, analysis, and modeling
- Technologies we use:
 - Relational database management systems
 - Structured query language
 - Python Programming
 - R Statistical Computing

"My team used basic concepts from almost every class period and topic section in our term project. It was cool to see how all the individual skills added up to help us create and maintain hydrologic information."



Undergraduate Training

- Creating the next generation of “Cyber-savvy” engineers and scientists
- Prototyping and developing new software applications
- Collaborating with iUTAH scientists
- Co-authoring CI-related papers
- Gaining practical experience and improving job prospects



Questions?



Jeff Horsburgh

jeff.horsburgh@usu.edu

435-797-946

Amber Jones

amber.jones@usu.edu

435-797-7147

Open Source Code Repositories

- WEBTSA – GAMUT time series data visualization
 - <https://github.com/UCHIC/WEBTSA>
- ODM Streaming Data Loader
 - <https://github.com/ODM2/ODM2StreamingDataLoader>
- ODM Tools Python – Sensor data management and QC
 - <https://github.com/UCHIC/ODMToolsPython>
- ODM2 Sensor – Sensor equipment management
 - <https://github.com/UCHIC/ODM2Sensor>
- iUTAH Utilities – Automated alerts, etc.
 - <https://github.com/UCHIC/iUtahUtilities>
- iUTAH Survey Data Viewer – Visualization of survey data
 - <https://github.com/UCHIC/SurveyDataViewer>
- iUTAH Data – Modeling and Data Federation Website
 - <https://github.com/UCHIC/iUTAHData>



GitHub