

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



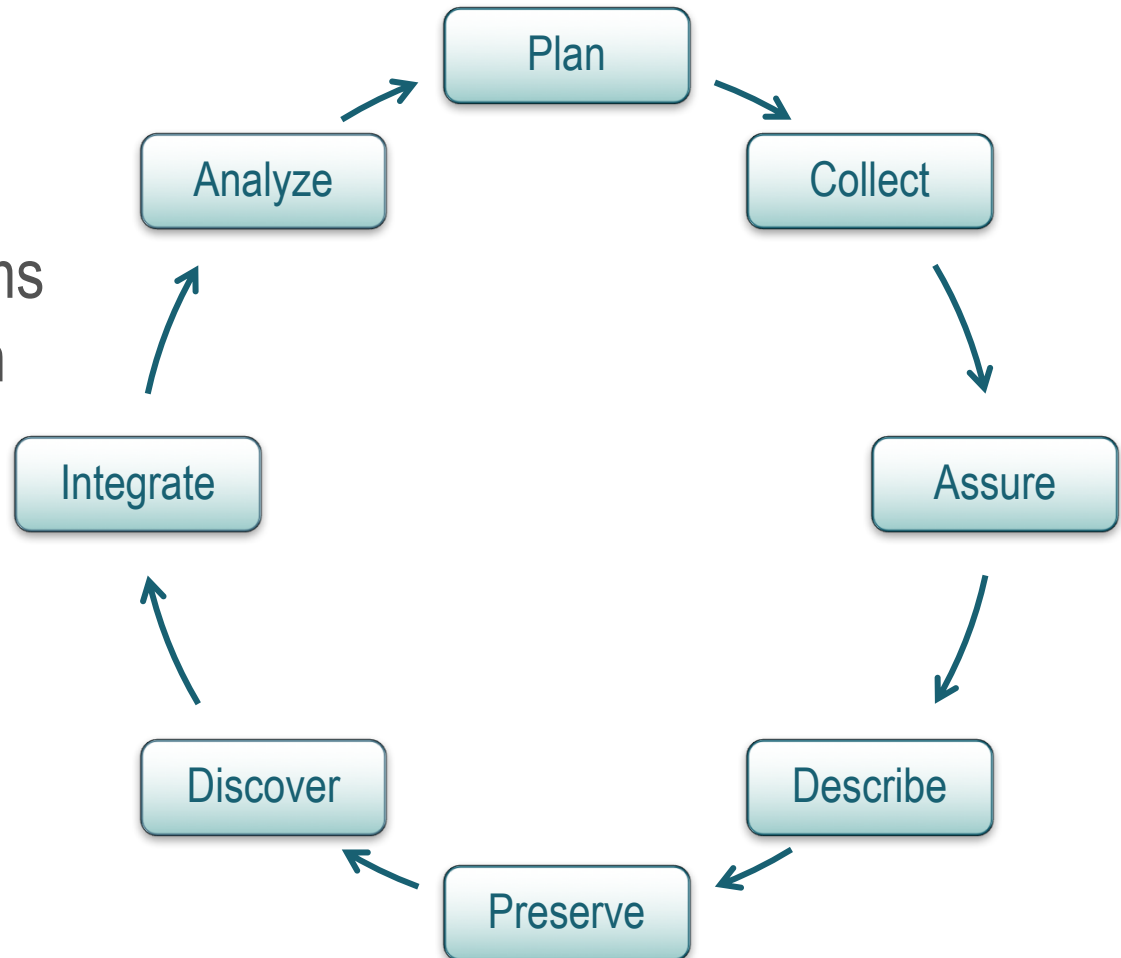
- Cyberinfrastructure needed to support:
 - Interdisciplinary and cross-institutional teams
 - Diverse data collection and modeling efforts
 - The full data life cycle
- Includes storage, software, networking, computational, and human resources.



Data Life Cycle



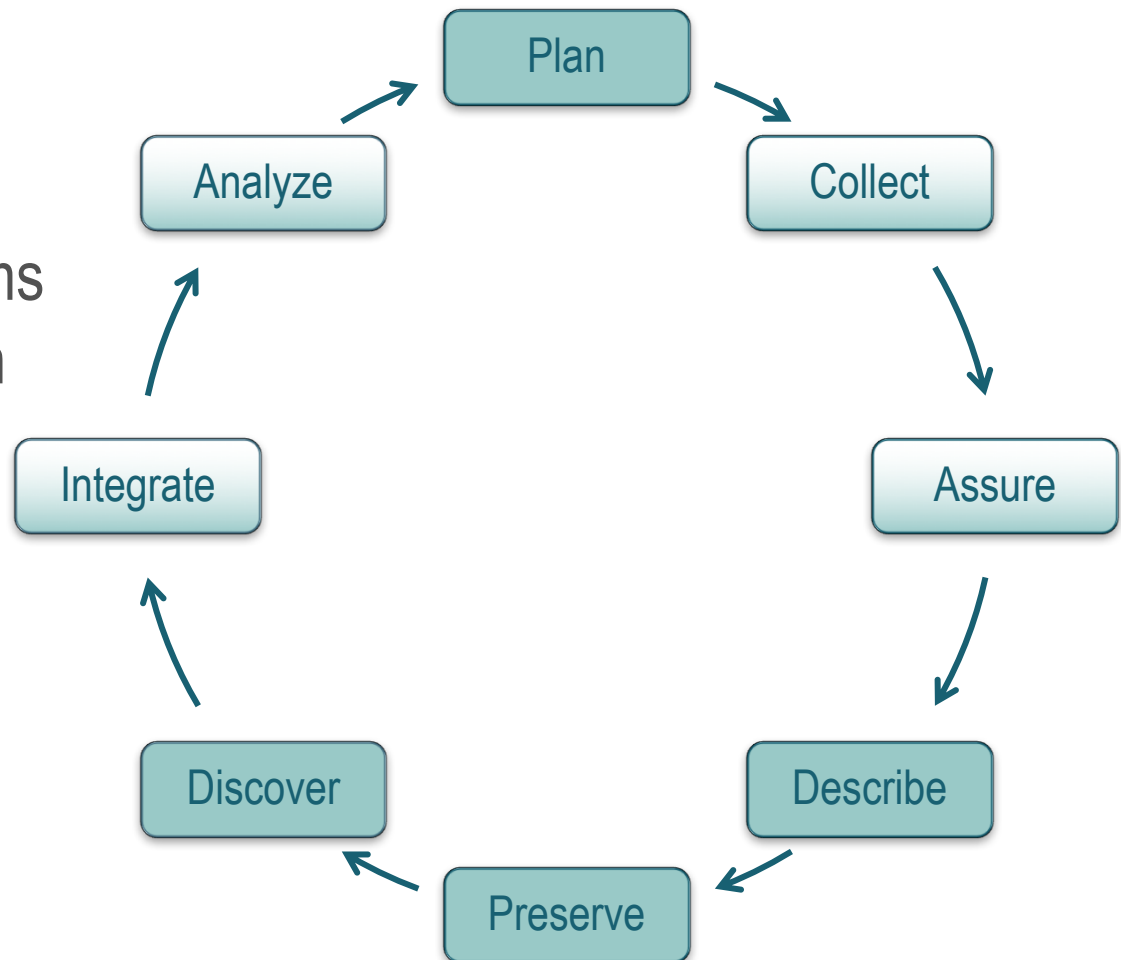
- Cyberinfrastructure needed to support:
 - Interdisciplinary and cross-institutional teams
 - Diverse data collection and modeling efforts
 - The full data life cycle
- Includes storage, software, networking, computational, and human resources.



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.



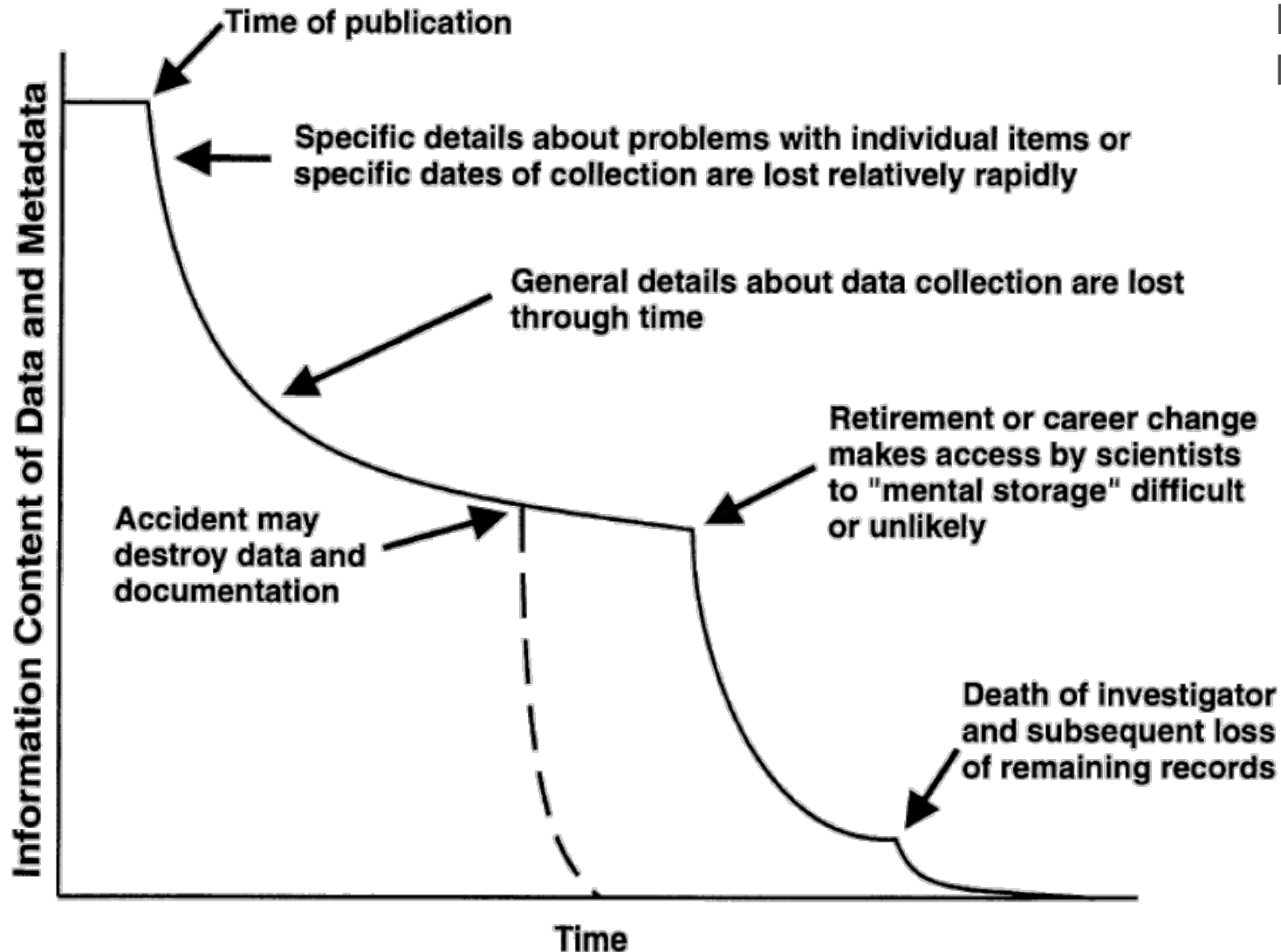
“All of the primary datasets collected as part of this project will be made freely and publicly available...”

- iUTAH Proposal Data Management Plan

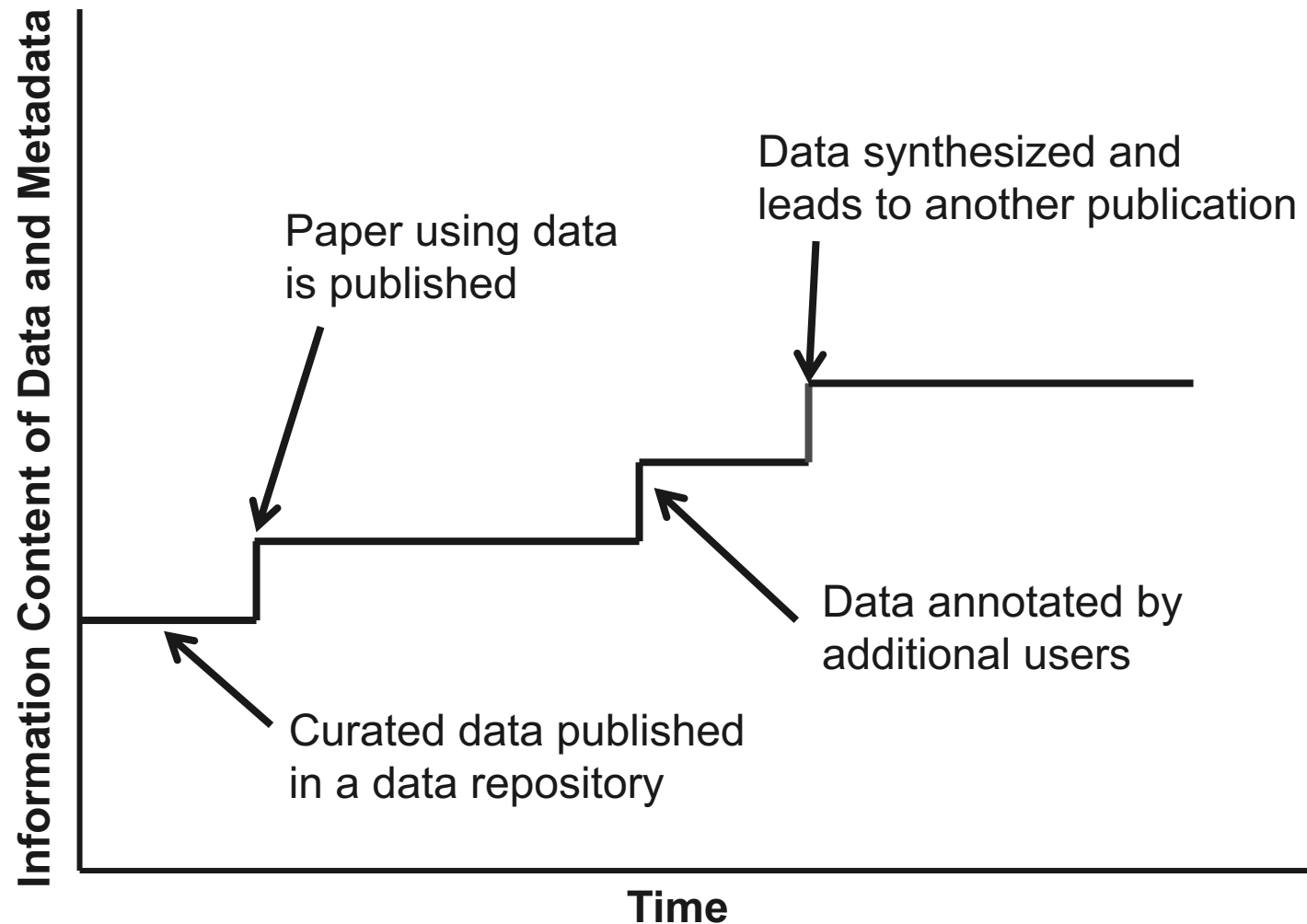
Information Entropy



Figure taken from
Michener (2006)



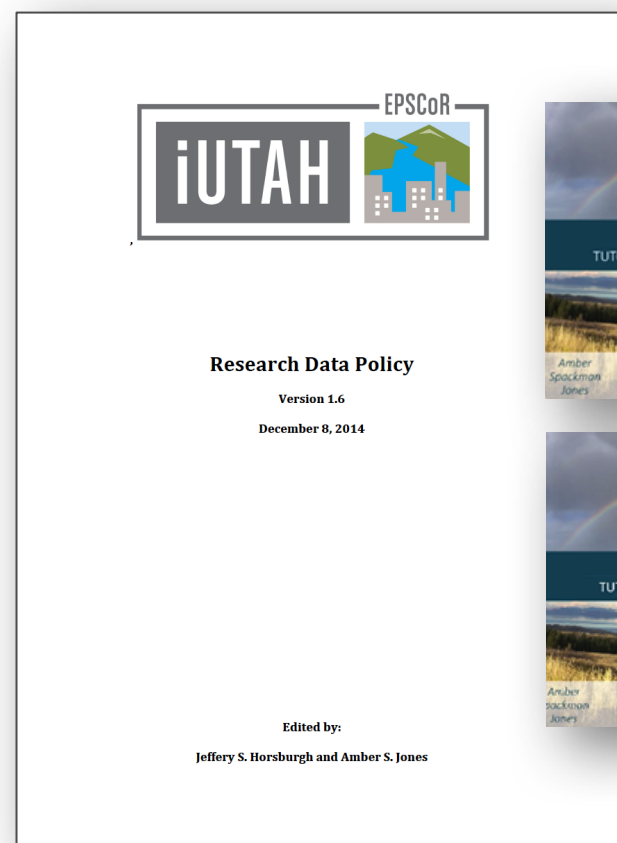
Information Entropy: What if instead...



Data Policy and Data Management Training

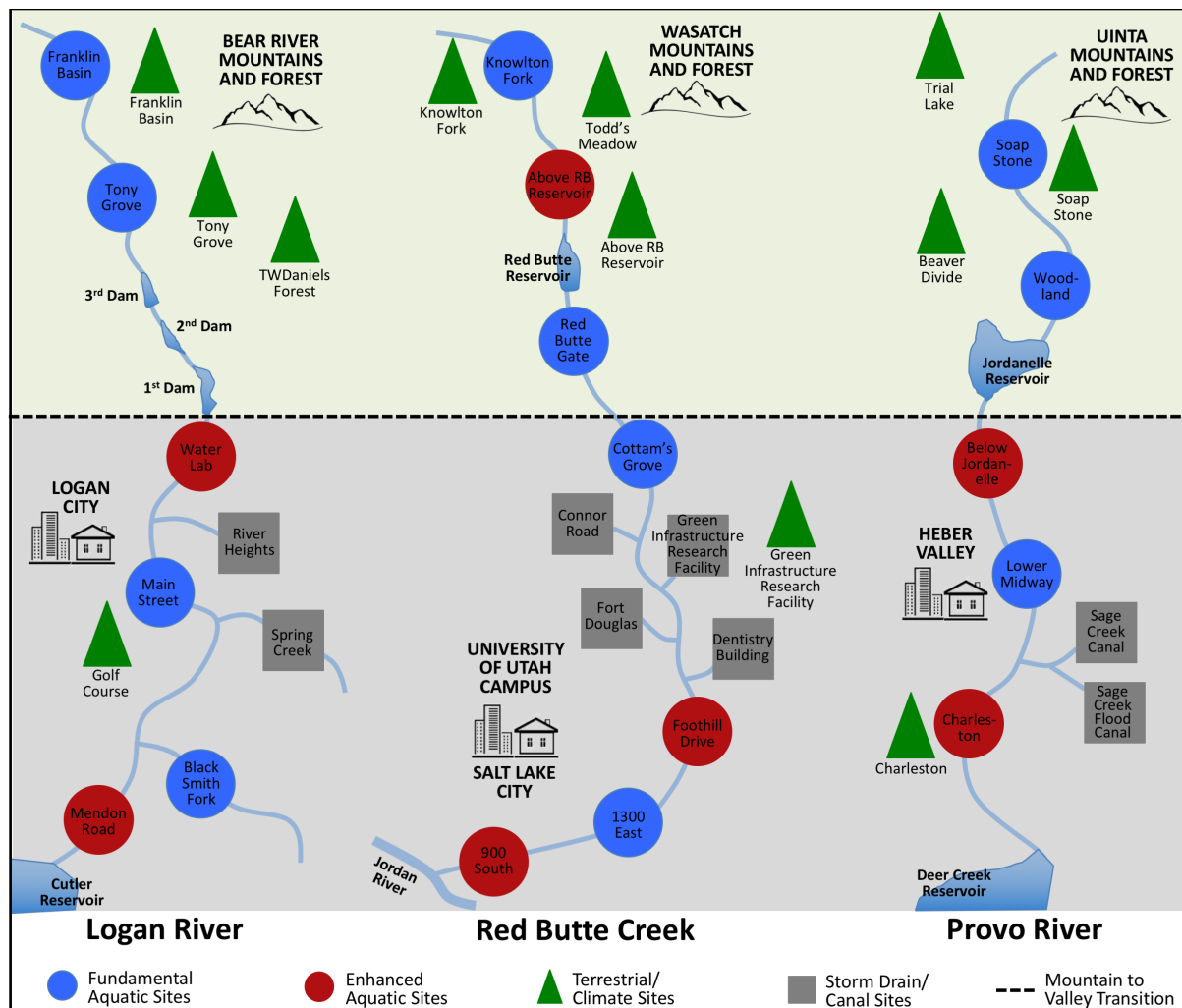


- Developed project data policy to codify guidelines and timeframes for data sharing
- Conducted training
- Developed data publication tutorials
- Review and curation of submitted datasets



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

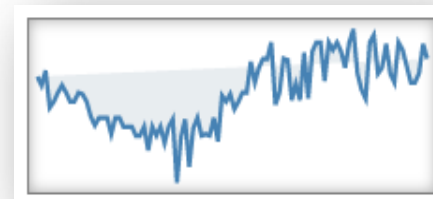
GAMUT: Gradients Along Mountain to Urban Transitions



40 sites

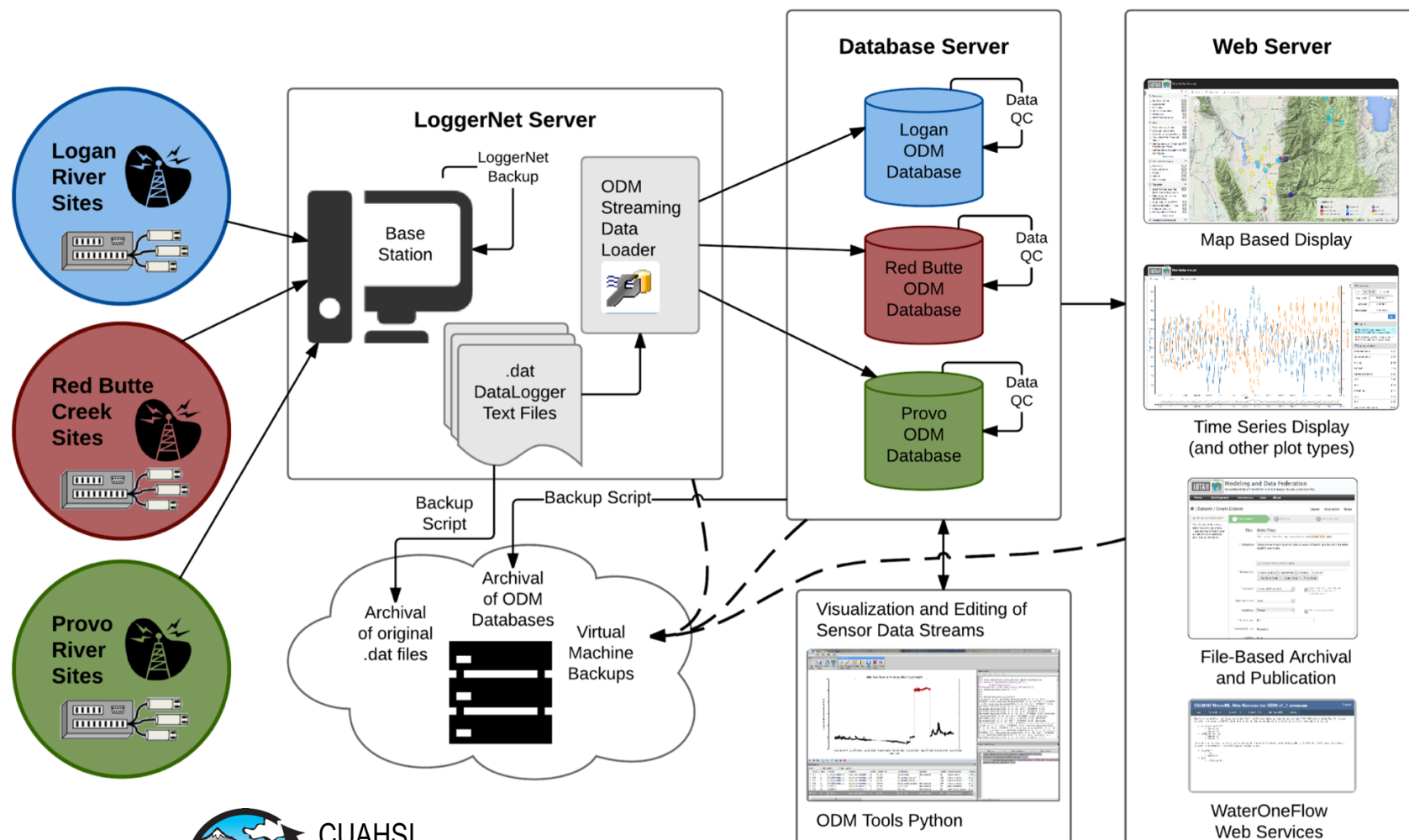



2000 data series



200 million+ values over 4 years

GAMUT: Data Work Flow










Open with...

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

Authors:
Owners:
Resource type:
Created:
Last updated:

iUTAH GAMUT Working Group
iUTAH Data Manager
Generic
July 19, 2016, 7:24 p.m.
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/af44ed6dc09a4070ac15a6ec0741f0d2>

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

 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
2.0 MB

 data/contents/iUTAH_GAMUT_RB_TM_C_RawData_2016.csv
12.9 MB



The screenshot displays the 'Time Series Analyst' web application interface. The top section features a map of the Logan River watershed in Utah, with various data points and features labeled. The bottom section shows a time series plot of Dissolved Oxygen (mg/L) and Temperature (°C) from 2014 to 2016. The plot shows seasonal fluctuations in both variables. The right sidebar contains options for data selection, visualization, and summary statistics.

Map Panel:

- Network:**
 - Logan River
 - Two Butte Creek
 - Provo River
 - US20 Groundwater
 - Northeast Field Canal
 - US20 Daily
 - US20 Interference
 - Green Infrastructure Research
- Quality Control Level:**
 - Raw data
 - Quality controlled data
 - Derived products
- Site:**
 - Comets Station at Logan River Golf Course
 - Climate Station at TMI Daniels Experimental Forest
 - Climate Station at Franklin Basin
 - Climate Station at Tony Grove
 - Logan River at the Utah Water Research Laboratory west bridge
 - Logan River at Mendon Road 800 South
 - Logan River near Tony Grove
 - Logan River at Main Street Highway 89/91 Bridge
 - Logan River near Franklin Basin
 - River Heights Bridge Storm Drain
 - Blacksmith Fork above confluence with Logan River
 - Spring Creek Storm Drain

Time Series Plot:

- Y-axis:** Dissolved Oxygen (mg/L) (left, 0-14) and Temperature (°C) (right, 0-24).
- X-axis:** Date (2014, April, July, October, 2015, April, July, October, 2016, April, July, October, 2017).
- Legend:**
 - DOO Oxygen, dissolved
 - LA_WaterLab_AA_Logan River at the Utah Water Research Laboratory west bridge
 - Water Series (DOO) Temperature
 - LA_WaterLab_AA_Logan River at the Utah Water Research Laboratory west bridge

Right Sidebar:

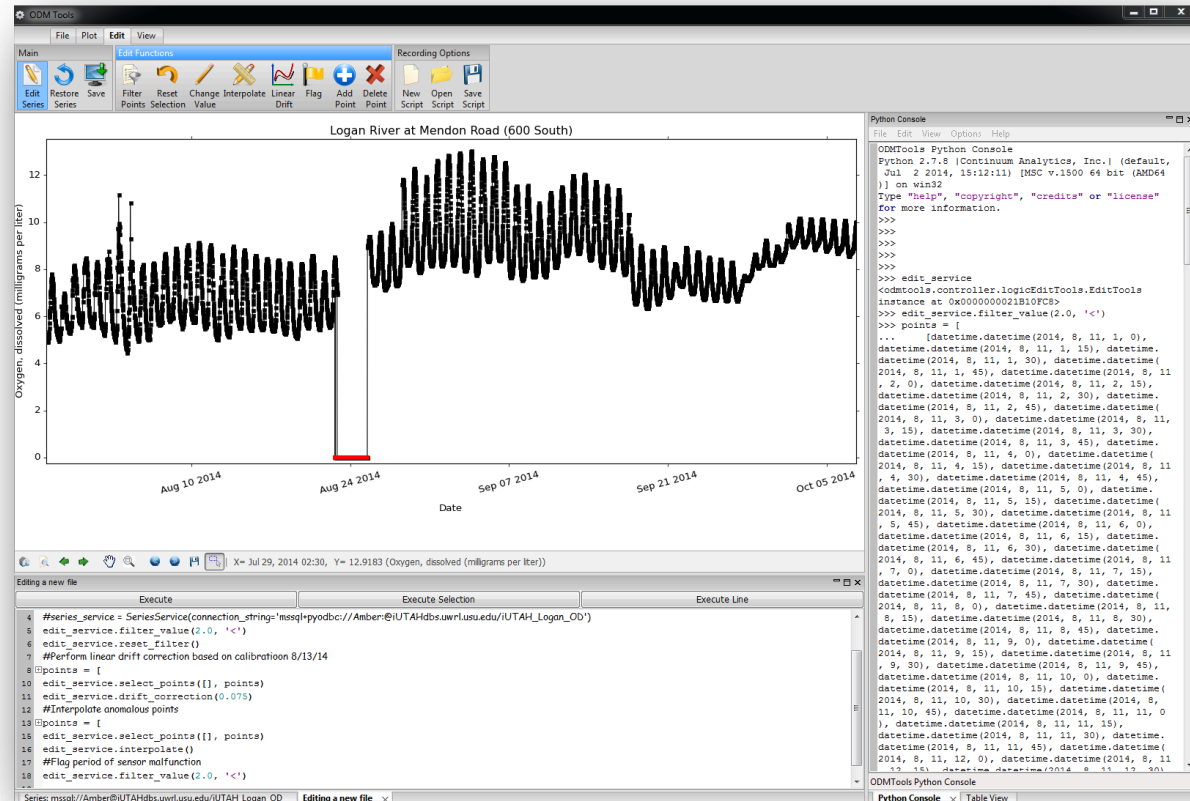
- Plot Options:**
 - All
 - Last Month
 - Last Week
 - Begin Date: 10/03/2013
 - End Date: 4/04/2016
 - Visualization: Time Series
- Legend:**
 - Analysis Site
 - Red Butte Creek
 - US20 Interference
 - Green Infrastructure Research
 - Climate Site
 - Logan River
 - Utah City
 - Other
 - Provo River
 - Groundwater
 - Northeast Field Canal
- Summary Statistics:**
 - Arithmetic Mean: 10.5
 - Geometric Mean: 5.1
 - Maximum: 14.0
 - Minimum: 0.0
 - Standard Deviation: 1.7
 - 10%: 10.0
 - Median: 50%
 - 75%: 10.0

[illegible]

GAMUT: Quality Control of Sensor Data



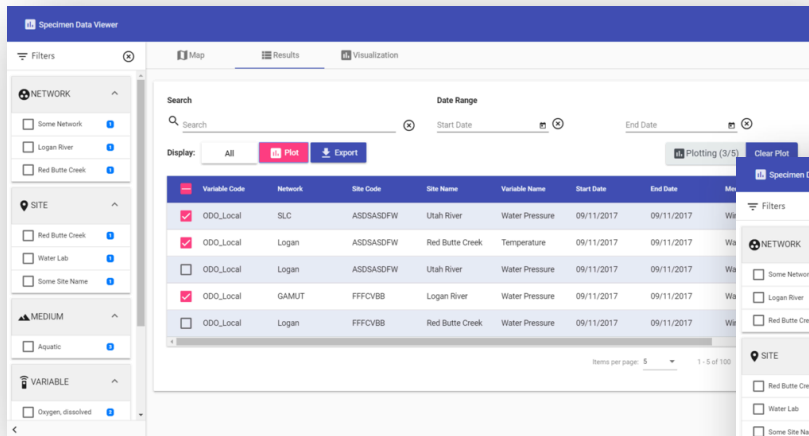
- Continuous, high frequency data require post processing
- Modifications to correct for common data errors
- Sensor Drift & Calibration
- Fouling
- Power Failure
- Icing
- Anomalies



GAMUT: Grab Sample Results



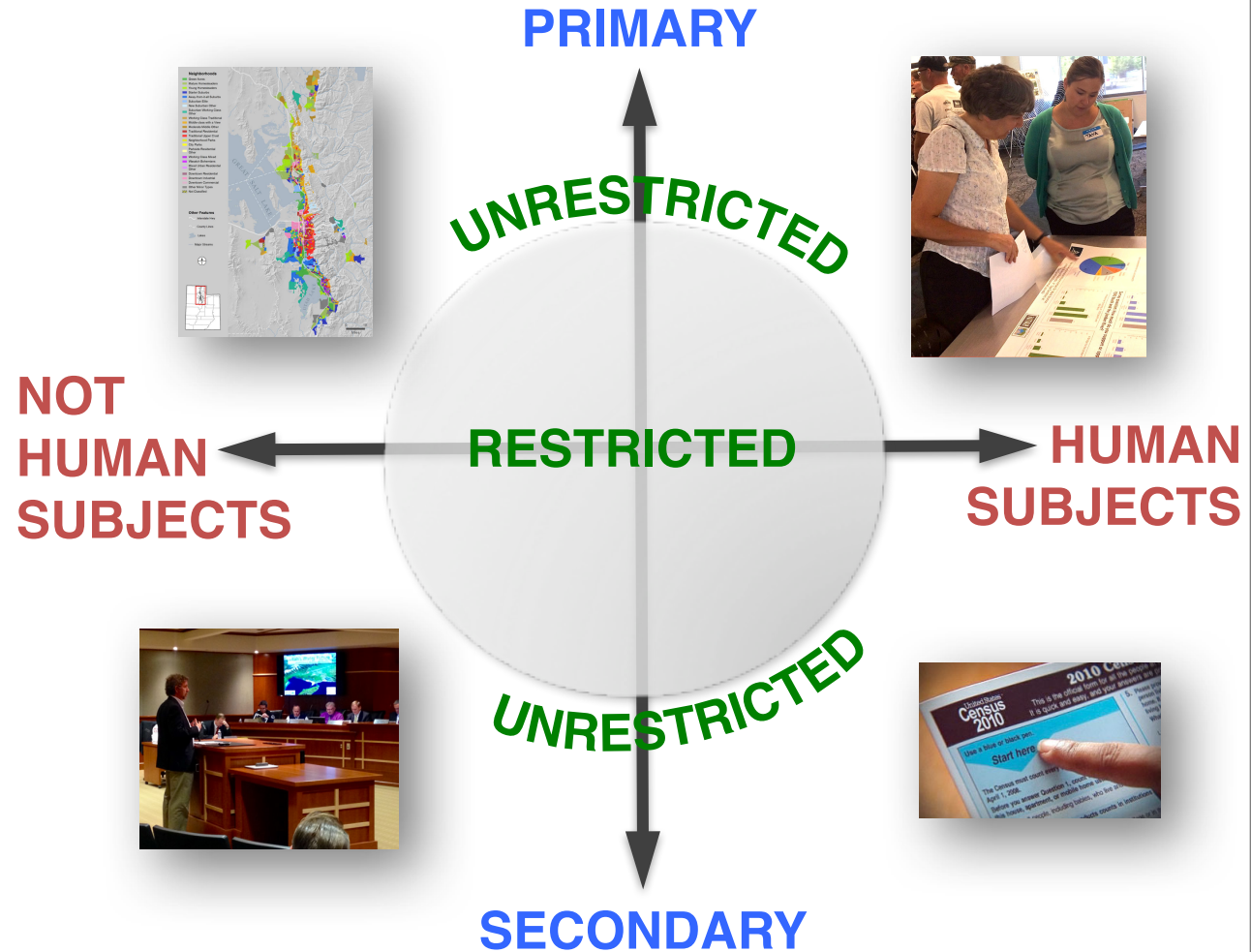
- Two efforts with different variable sets:
 - Monthly sampling: 2013-2014
 - Biweekly sampling: 2014-2016
- Variables: TSS, VSS, species of Nitrogen and Phosphorus, Total Coliform, E.coli, DOC, Fluorescence, Chlorophyll-a, Isotopes, Ions



Social Water Science Data: Considerations



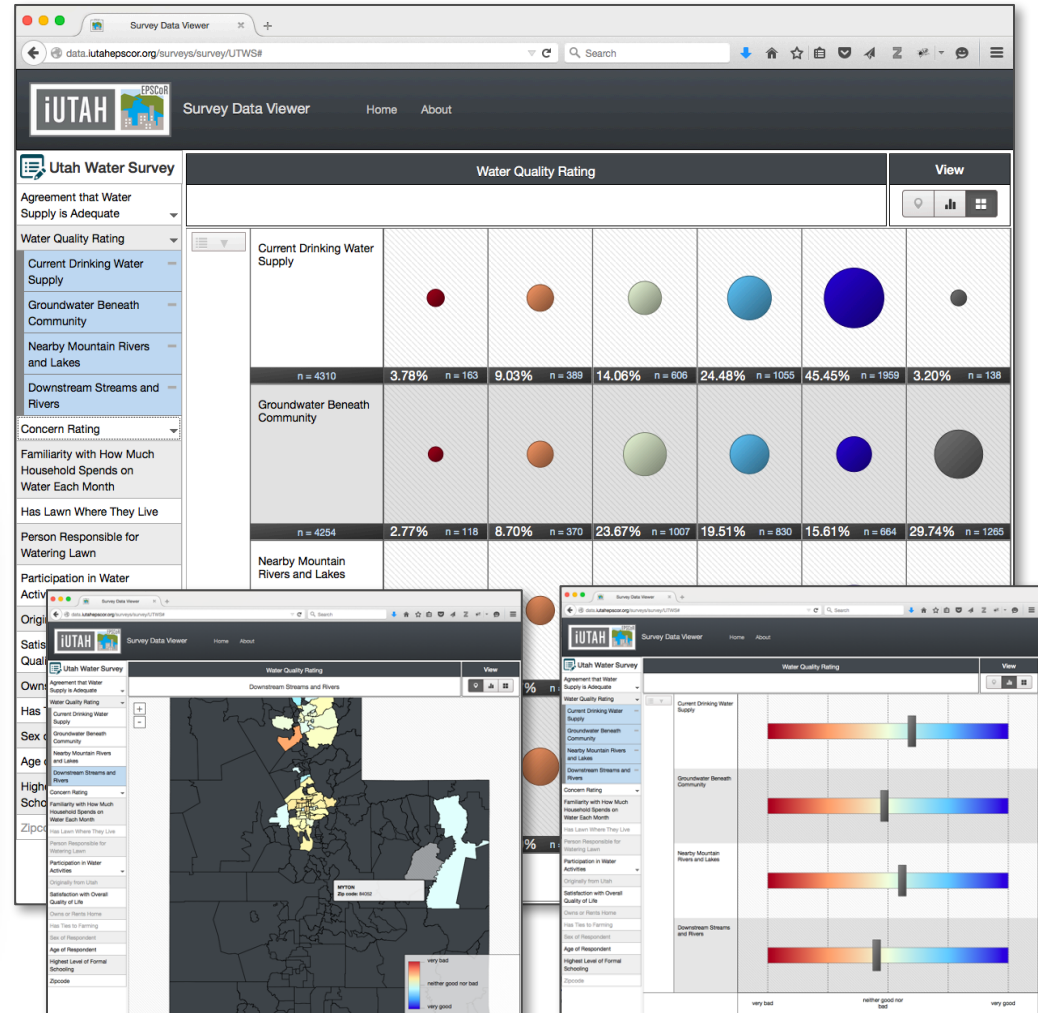
- Developed methods for categorizing social science data
- Dimensions help us understand mechanisms and restrictions for how social science data can be shared



Social Water Science Data: Visualization



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



<http://data.iutahepscor.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

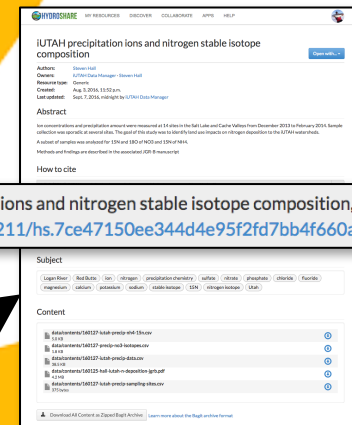
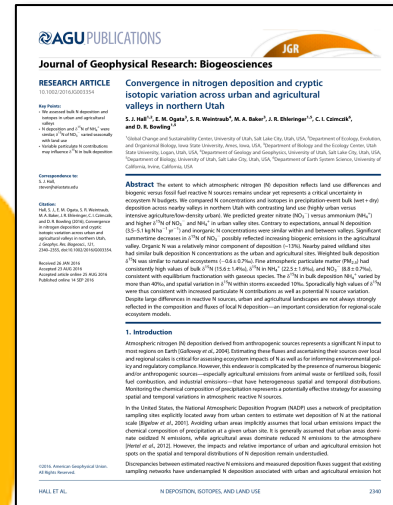
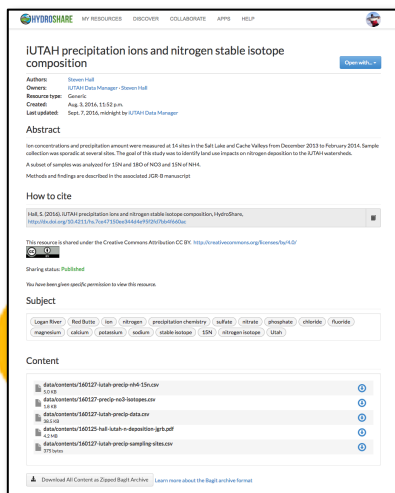


- An online, collaborative system for sharing and publishing various data types, models, and code
- Support for collaboration-sharing with individuals, groups, or publically
- Ability to formally publish with a DOI

The screenshot displays the HydroShare website interface. The top navigation bar includes links for 'MY RESOURCES', 'DISCOVER', 'COLLABORATE', 'APPS', and 'HELP'. The main banner features a river scene with the text 'Share your data and models with colleagues' and a sub-header 'Upload, share, and access a broad set of hydrologic data types and models. Manage who has access to the content that you share.' Below this, a sidebar lists 'What you can do' with green checkmarks for: 'Share your data and models', 'Manage who has access to your data', 'Share, access, visualize, and download models', 'Use the web services to access data', 'Publish data and models to the web', 'Discover and access data', and 'Use web apps to visualize data'. The main content area shows the title 'iUTAH GAMUT Network Raw Data at Logan River at the Utah Water Research Laboratory west bridge (LR_WaterLab_AA)' with an 'Open with...' button. Below the title, metadata is listed: 'Authors: iUTAH GAMUT Working Group - iUTAH Data Manager', 'Owners: iUTAH Data Manager', 'Resource type: Generic', 'Created: May 17, 2017, 7:28 p.m.', and 'Last updated: June 10, 2017, 11:17 p.m. by iUTAH Data Manager'. An 'Abstract' section follows, stating: 'This dataset contains raw data for all of the variables measured for the iUTAH GAMUT Network Logan River at the Utah Water Research Laboratory west bridge (LR_WaterLab_AA). 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 site and the variable and method of each column.' Below the abstract is a 'Subject' section with a list of tags: 'Logan River at the Utah Water Research Laboratory west bridge', 'Stream', 'time series', 'iUTAH', 'GAMUT', and 'raw data'. A 'How to cite' section provides the citation: 'iUTAH GAMUT Working Group, i. D. Manager (2017), iUTAH GAMUT Network Raw Data at Logan River at the Utah Water Research Laboratory west bridge (LR_WaterLab_AA), HydroShare, <http://www.hydroshare.org/resource/2b3afc29e11c412c84d5c9cd9b6279d6> Copy'. Below the citation is a Creative Commons Attribution CC BY license link: '<http://creativecommons.org/licenses/by/4.0/>'. The 'Sharing status' section shows a 'Public' button.



HYDROSHARE



Steven verified his data and metadata were correct but kept the data private

Steven submitted his paper for publication and responded to reviews

Steven published his data in HydroShare and received a DOI

With a little help, Steven deposited his dataset in the online HydroShare repository

Steven collected his data in the field and transformed into a sharable format



Steven published his paper and cited published data in HydroShare

Acknowledgments

Upon manuscript acceptance all data are publicly available online at the Hydroshare database: <http://dx.doi.org/10.4211/hs.7ce47150ee344-d4e95f2fd7bb4f660ac>. We thank four

The Steven Hall Story

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.”



THE UNIVERSITY OF UTAH®



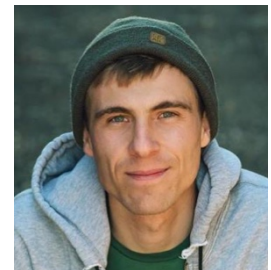
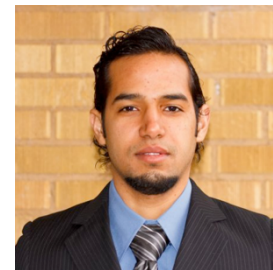
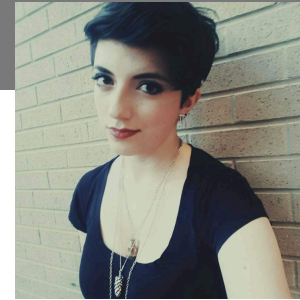
UtahState University®



BRIGHAM YOUNG UNIVERSITY

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?



Open Source Code Repositories

Jeff Horsburgh

jeff.horsburgh@usu.edu

435-797-946

Amber Jones

amber.jones@usu.edu

435-797-7147

- **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