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

Amber Spackman Jones

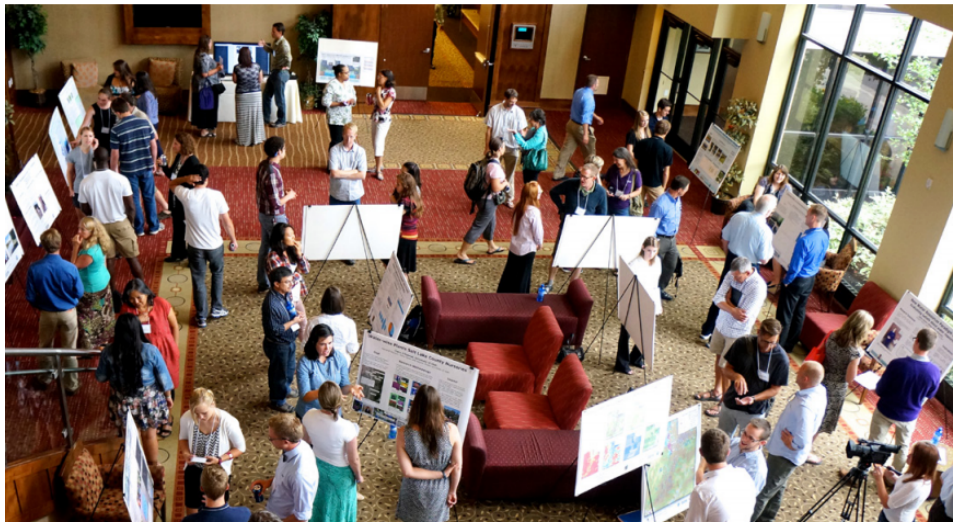
Jeffery S. Horsburgh

and the iUTAH Cyberinfrastructure Team

EPSCoR iUTAH

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.

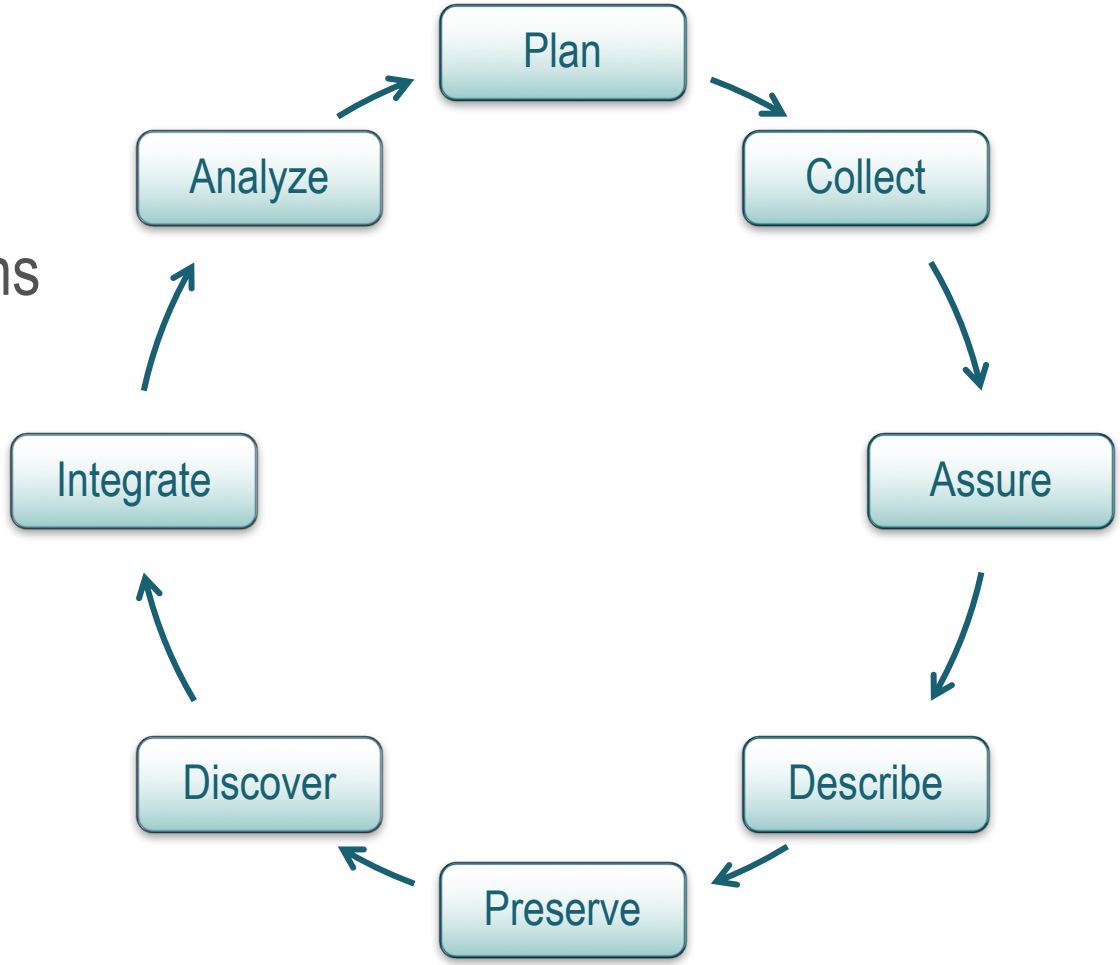


Figure courtesy DataOne

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.

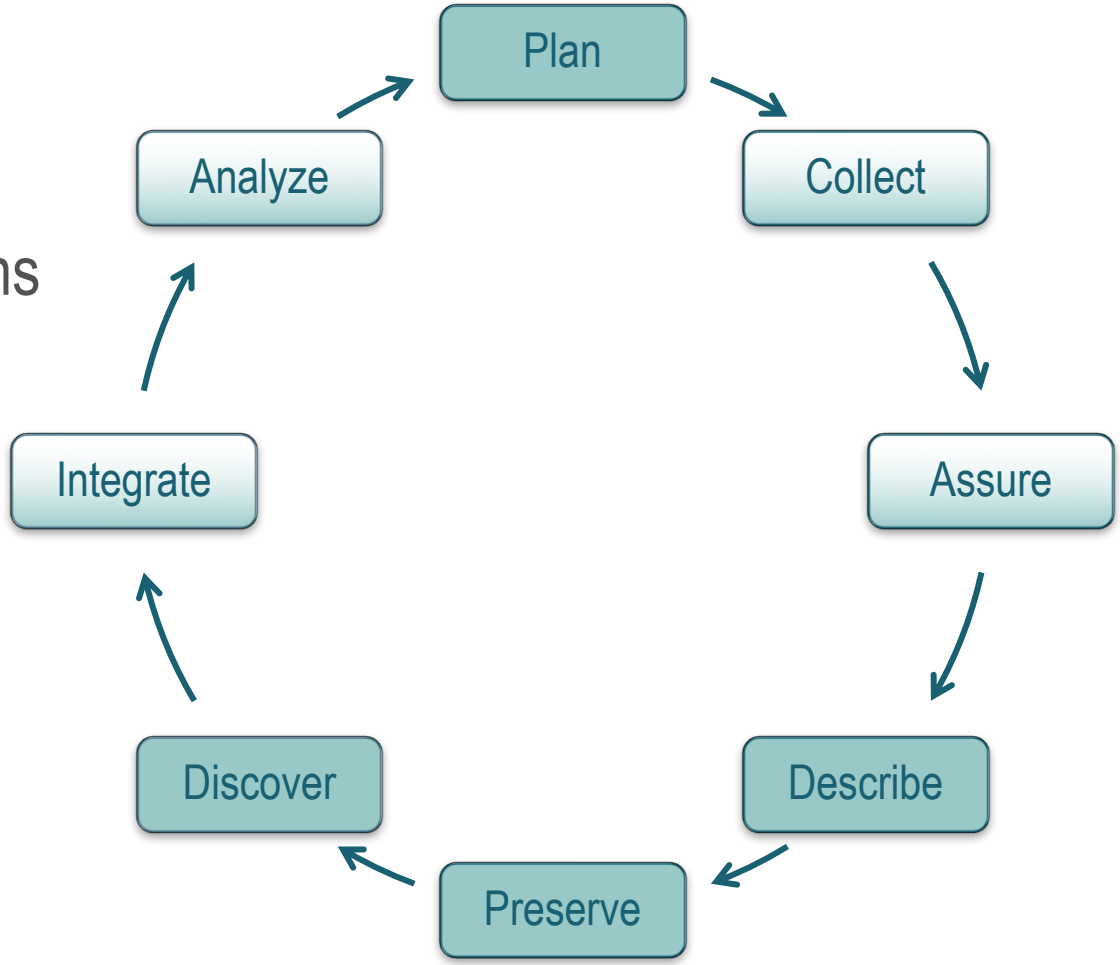


Figure courtesy DataOne

“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

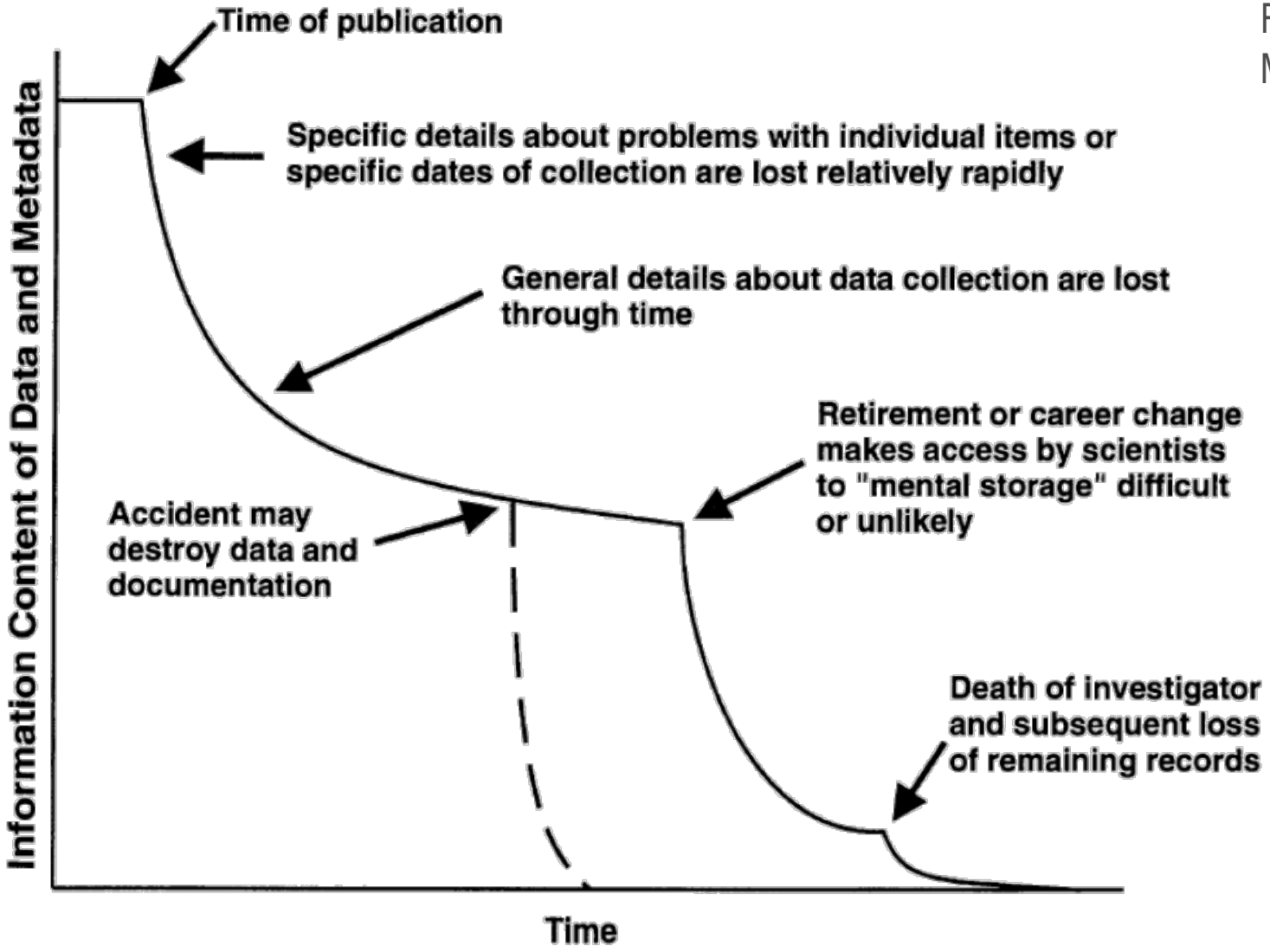
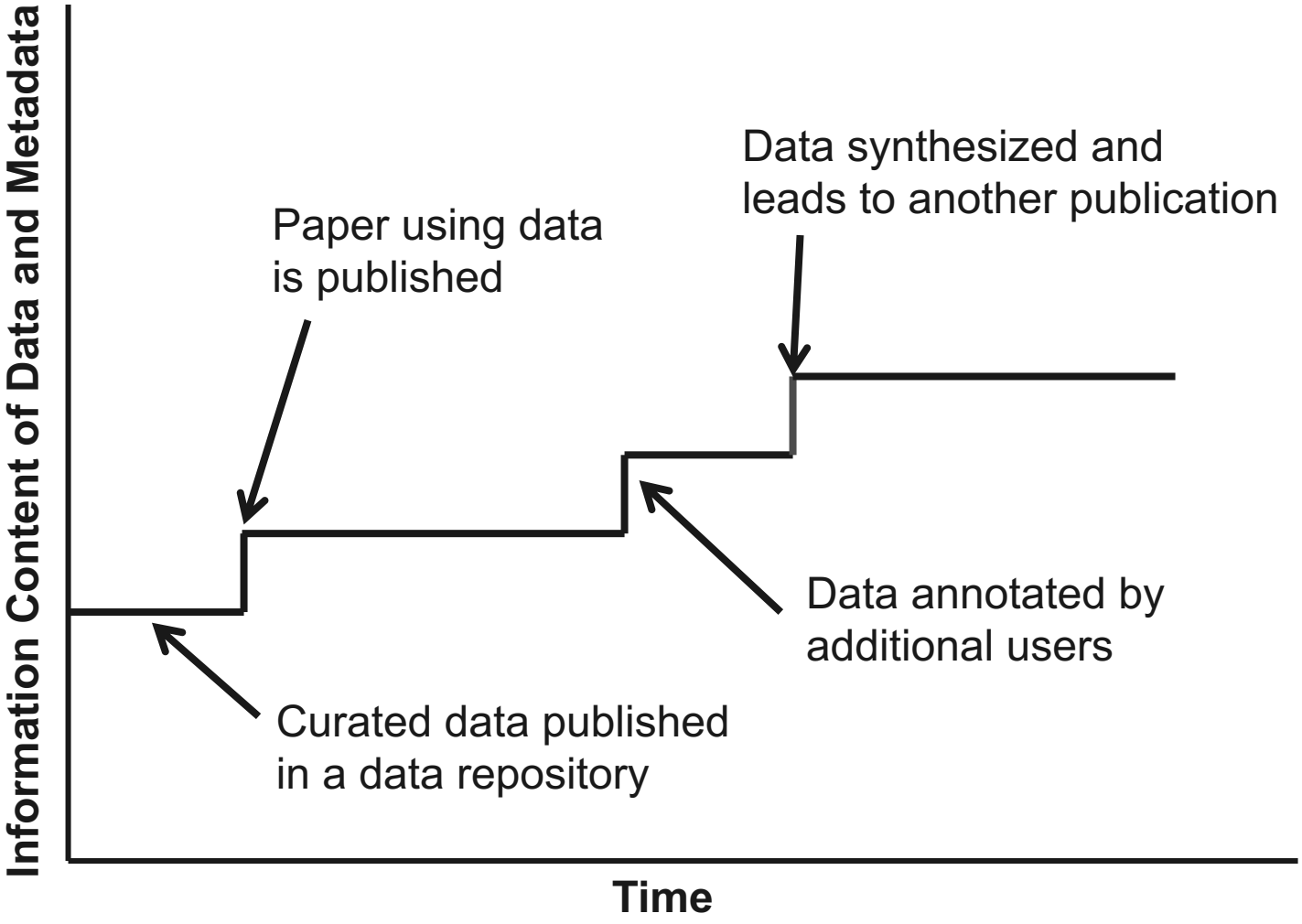


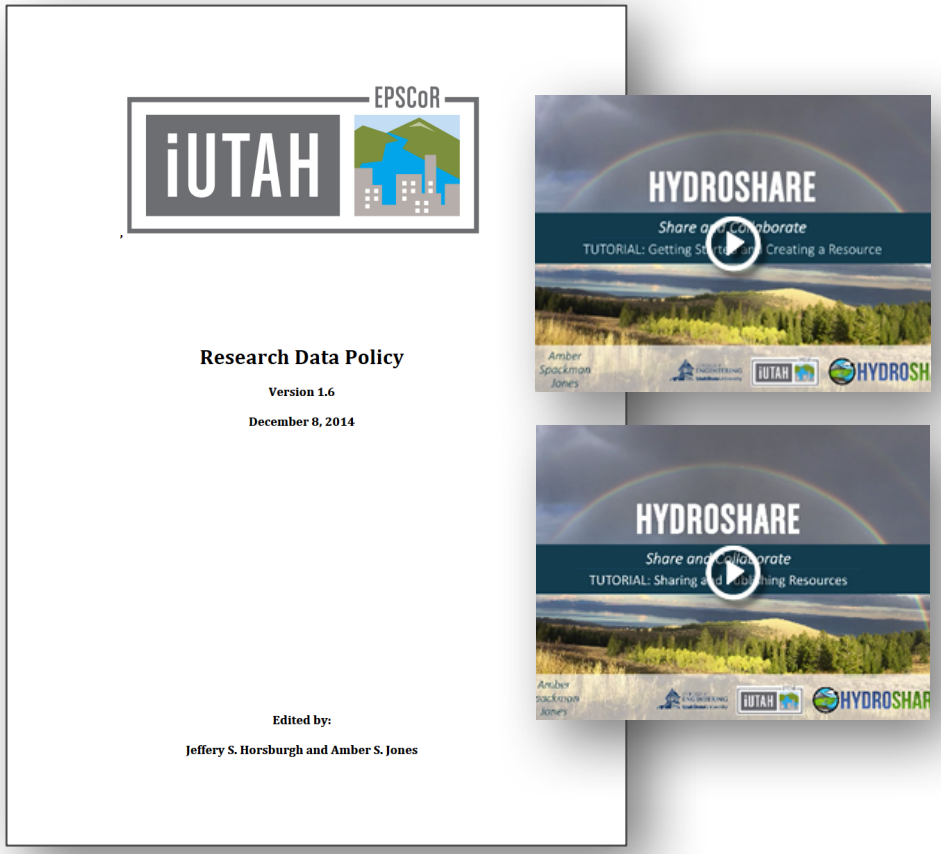
Figure taken from Michener (2006)

Information Entropy: What if instead…



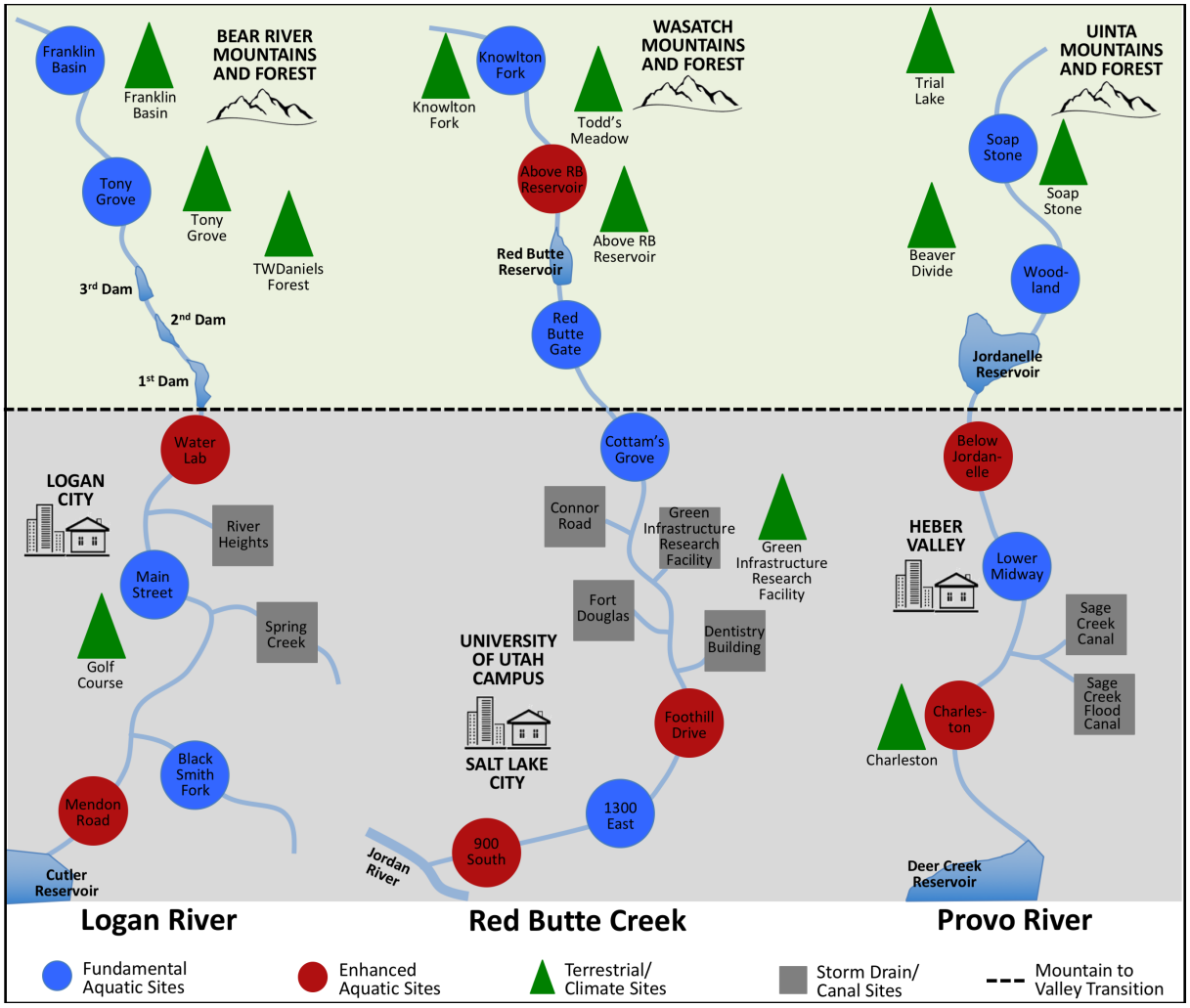
Data Policy and Data Management Training

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

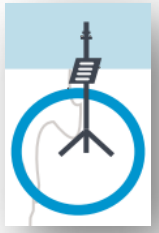


Changing the culture from "My Data" to "Our Data"

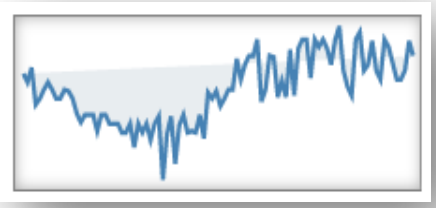
GAMUT: Gradients Along Mountain to Urban Transitions



40 sites



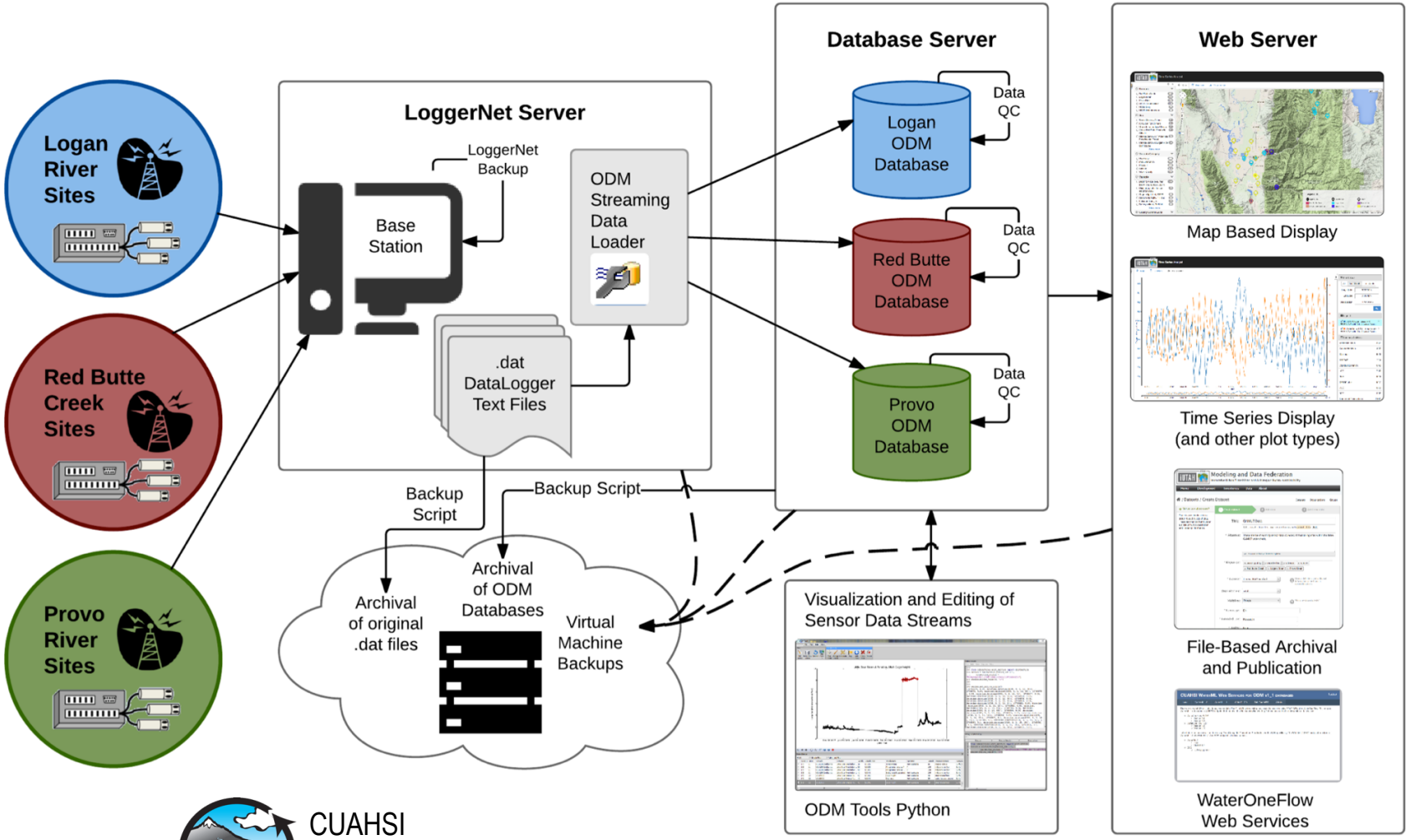
2000 data series



200 million+
values over 4 years

Jones, A.S. , Z.T. Aanderud, J.S. Horsburgh, D.P. Eiriksson, D. Dastrup, C. Cox, S. B. Jones, D.R. Bowling, J. Carlisle, G.T. Carling, M.A. Baker, 2017. Designing and Implementing a Network for Sensing Water Quality and Hydrology across Mountain to Urban Transitions. Journal of the American Water Resources Association (JAWRA) 53(5):1095–1120. https://doi.org/10.1111/1752-1688.12557

GAMUT: Data Work Flow

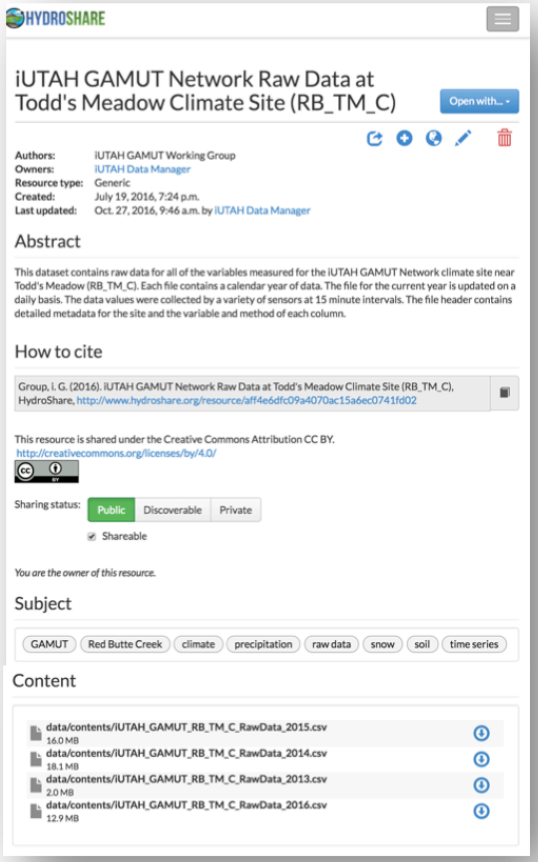


CUAHSI HIS
Sharing hydrologic data

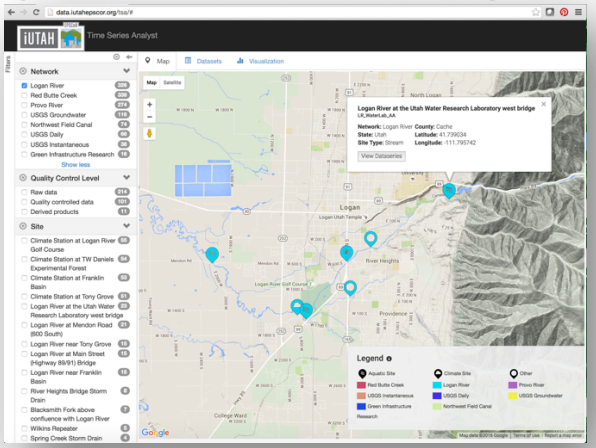
Jones, A.S., J. S. Horsburgh, S.L. Reeder, M. Ramirez, J. Caraballo (2015). A data management and publication workflow for a large-scale, heterogeneous sensor network, Environmental Monitoring and Assessment, 187:348, http://dx.doi.org/10.1007/s10661-015-4594-3.

GAMUT: Accessing Sensor Data

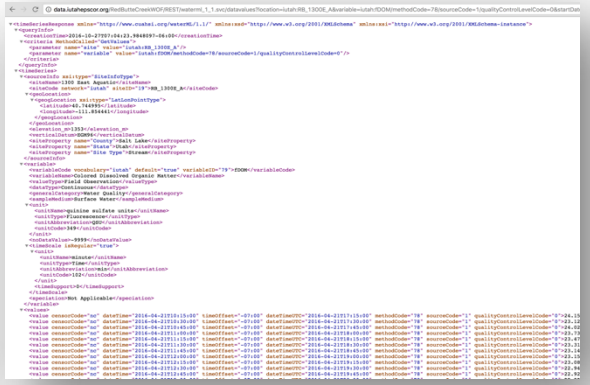
HydroShare:
Data available in flat files (csv) and updated daily

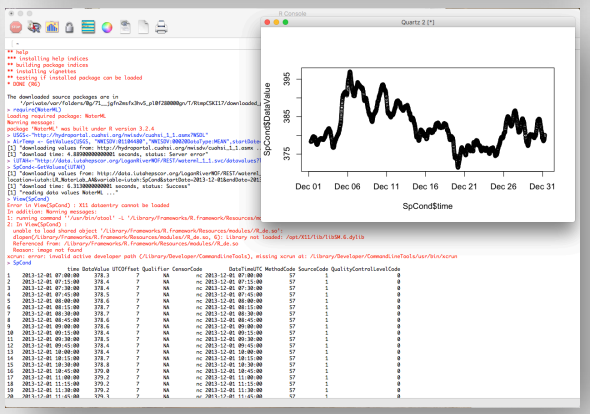


Time Series Analyst:
Web interface, map and plot tools, and data export



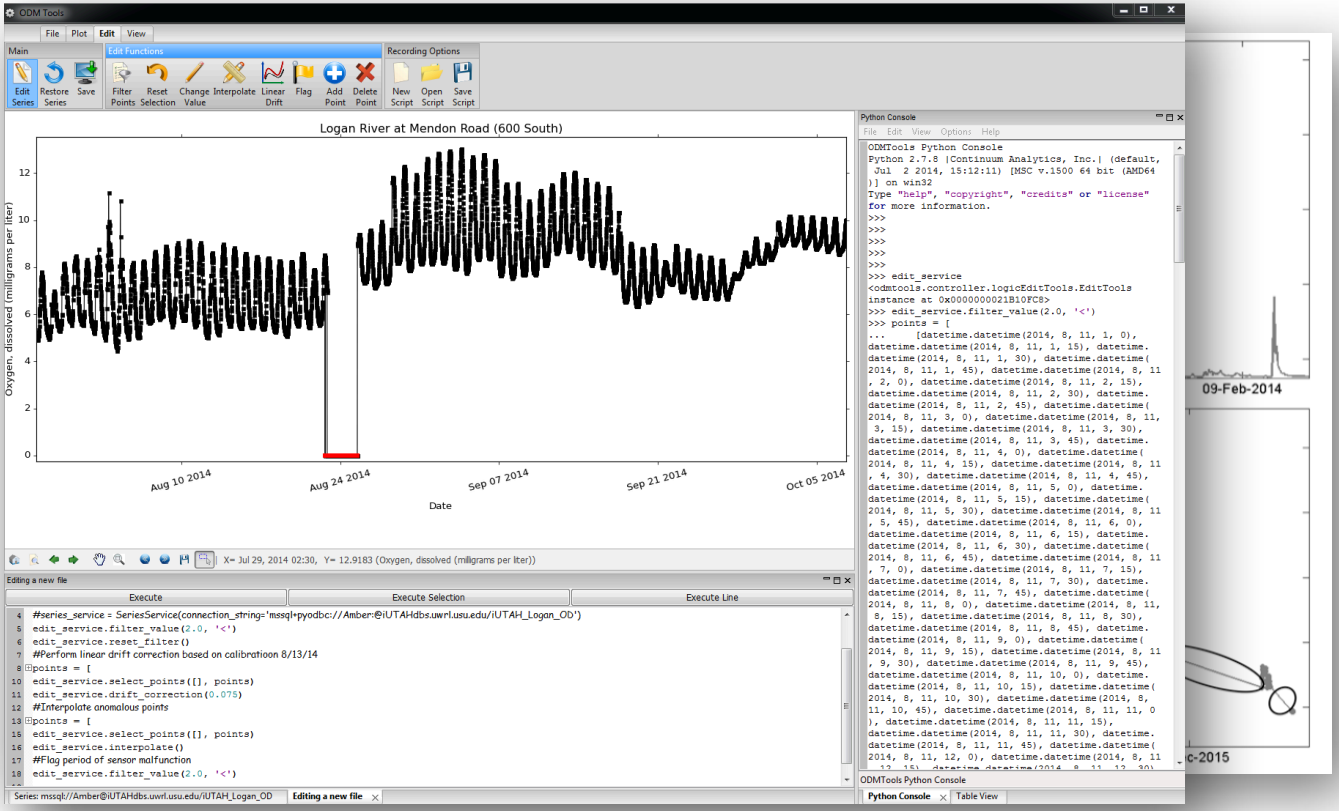
Web Services:
Programmatic data access





GAMUT: Quality Control of Sensor Data

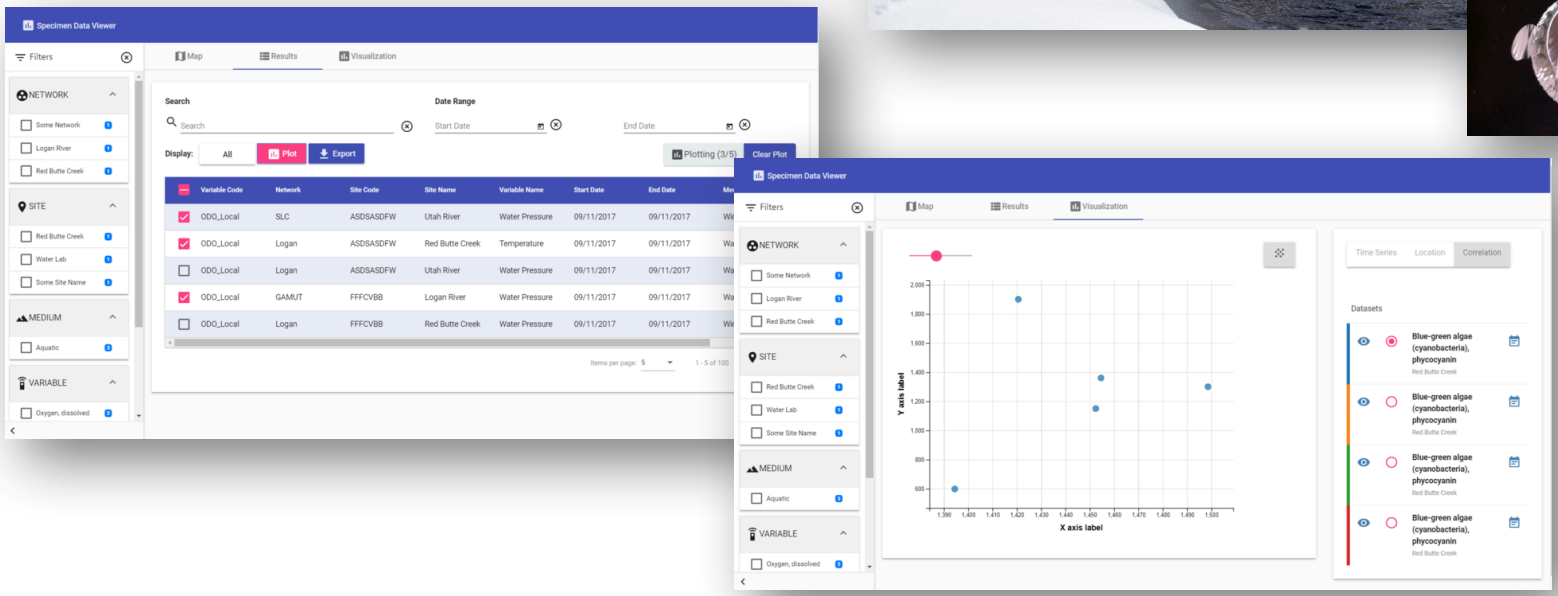
- Continuous, high frequency data require post processing
- Modifications to correct for common data errors
- Sensor Drift & Calibration
- Fouling
- Power Failure
- Icing
- Anomalies



Horsburgh, J. S., Reeder, S. L., Jones, A.S., Meline, J. (2015). Open source software for visualization and quality control of continuous hydrologic and water quality sensor data, Environmental Modelling & Software, 70, 32-44, doi:10.1016/j.envsoft.2015.04.002.

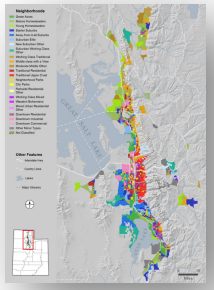
GAMUT: Grab Sample Results

- 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, Chlorphyll-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



PRIMARY

UNRESTRICTED

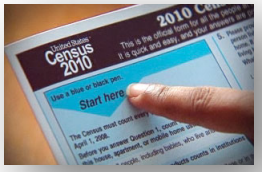
NOT HUMAN SUBJECTS

RESTRICTED

HUMAN SUBJECTS

UNRESTRICTED

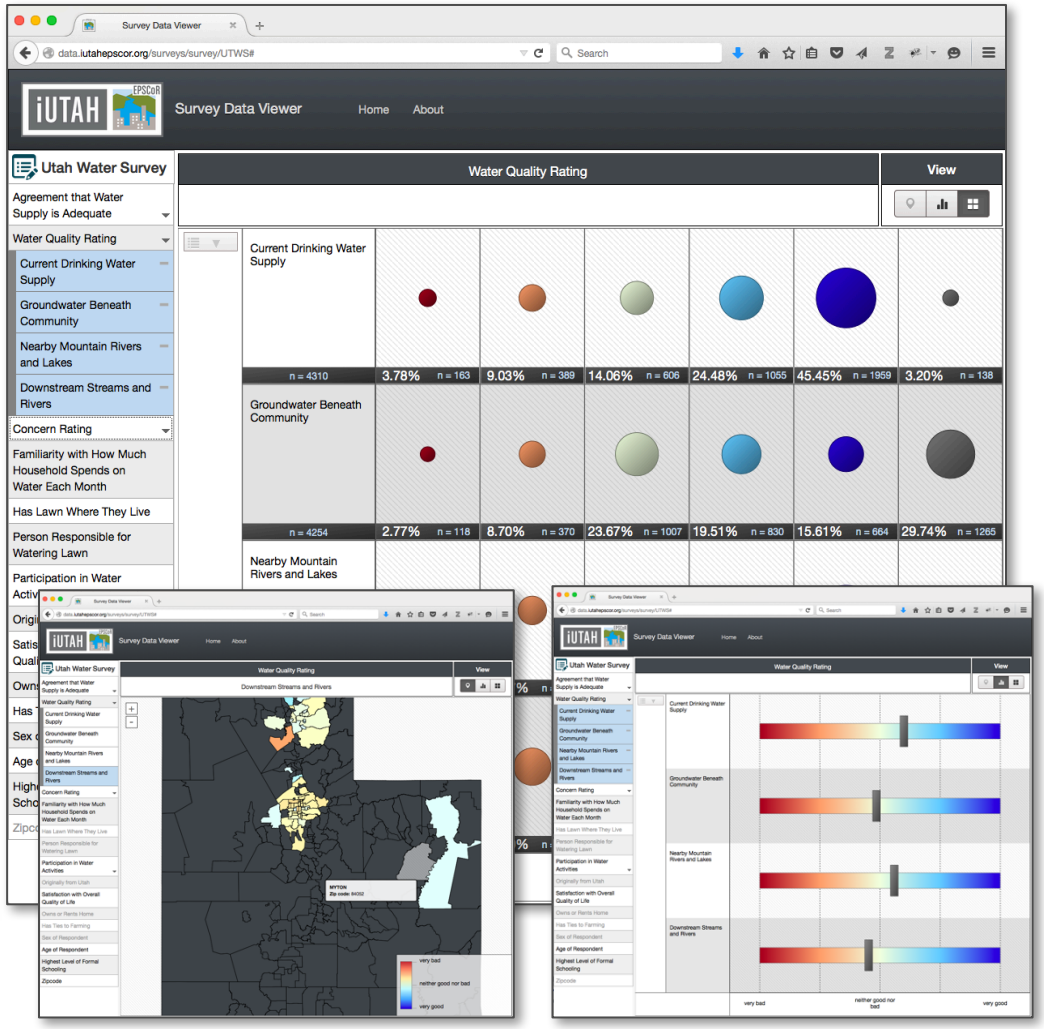
SECONDARY



Flint, C.G., A.S. Jones, J.S. Horsburgh, 2017. Data Management Dimensions of Social Water Science: The iUTAH Experience. Journal of the American Water Resources Association (JAWRA) 53(5):988-996. https://doi.org/10.1111/1752-1688.12568

Social Water Science Data: Visualization

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



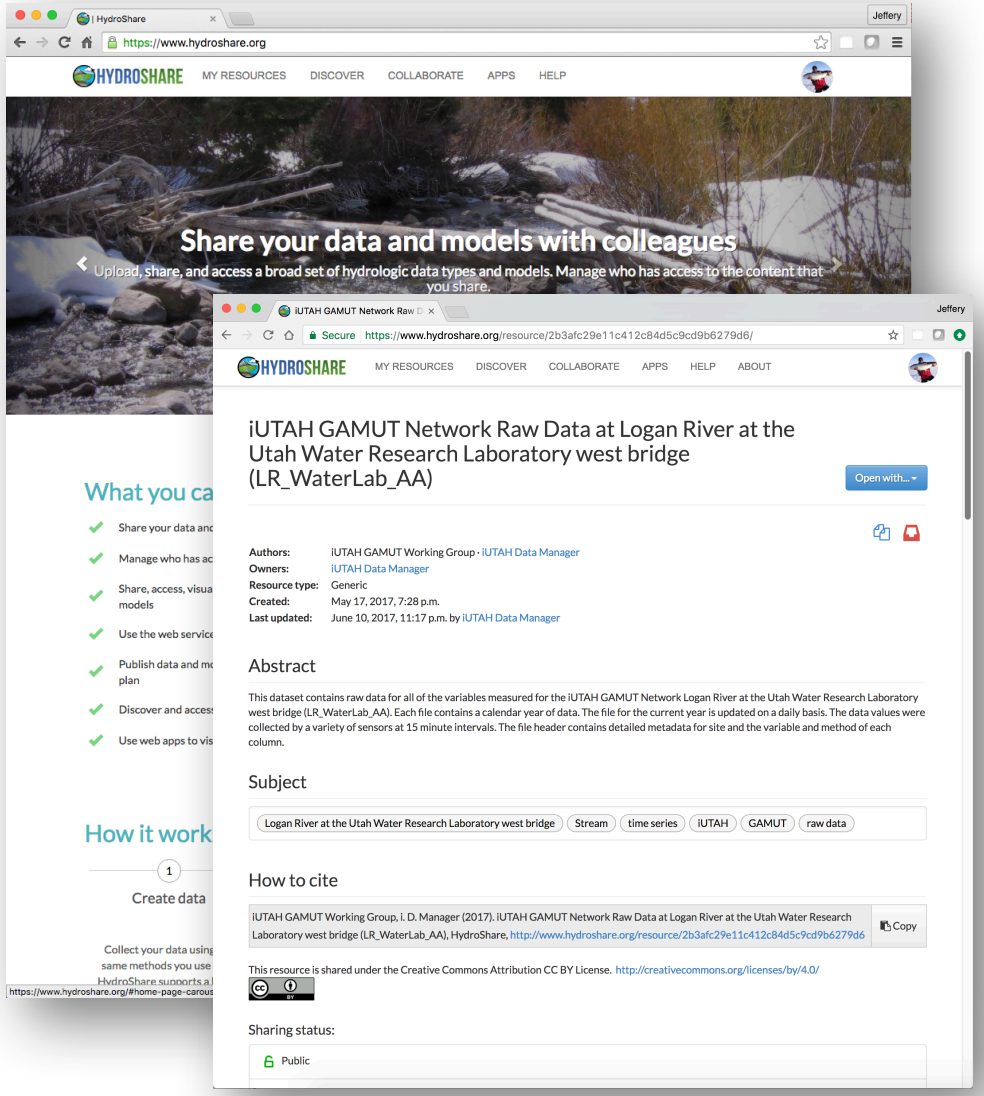
http://data.iutahepscor.org/surveys/

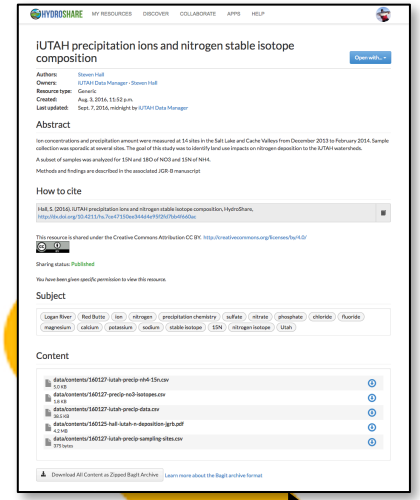
Jones, A. S., Horsburgh, J. S., Jackson-Smith, D., Ramirez, M., Flint, C., Caraballo, J. (2016). A Web-based, interactive visualization tool for social environmental survey data, *Environmental Modelling & Software*, 84, 412-426, doi:10.1016/j.envsoft.2016.07.013.

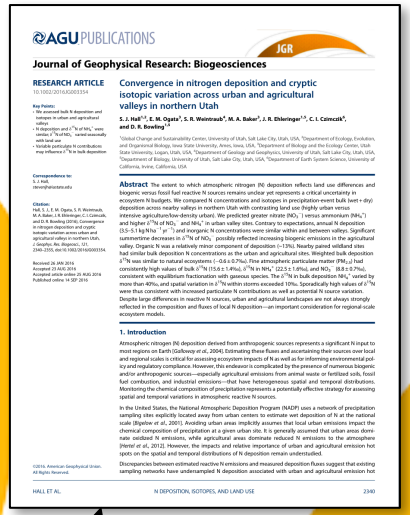
Sharing Diverse Data

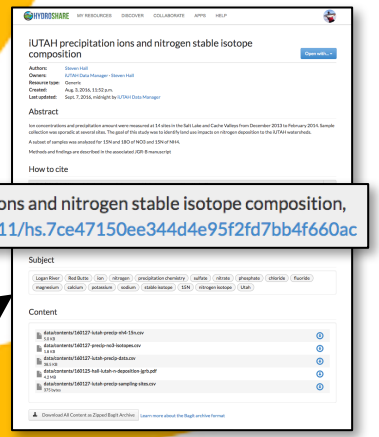
HYDROSHARE
http://www.hydroshare.org

- 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









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

Steven submitted his paper for publication and responded to reviews

Hall, S. (2016). iUTAH precipitation ions and nitrogen stable isotope composition, HydroShare, http://dx.doi.org/10.4211/hs.7ce47150ee344d4e95f2fd7bb4f660ac

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

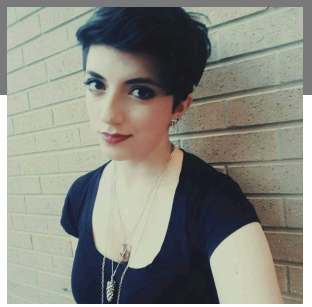
Hydroinformatics Course

- 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."

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