Cyberinfrastructure

Promoting Collaborative Publication, Interoperability, and Reuse of iUTAH Data and Research Products

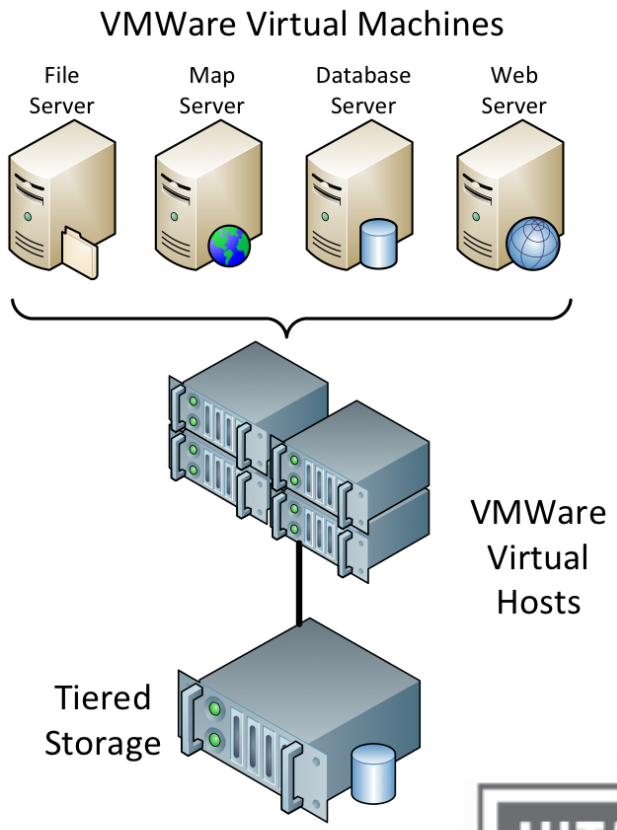
Jeffery S. Horsburgh

Amber Jones and the iUTAH CI Team

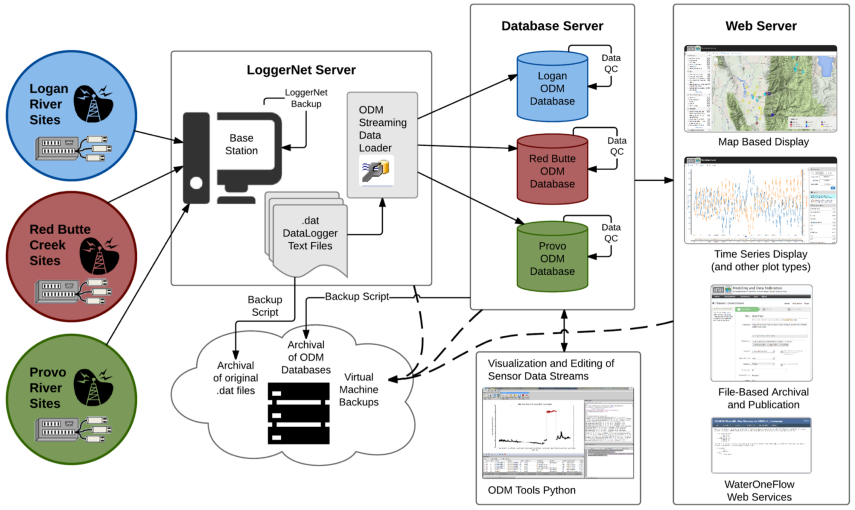
Fall 2016 All Hands Meeting
"Focus on Facilities"

Major CI Components

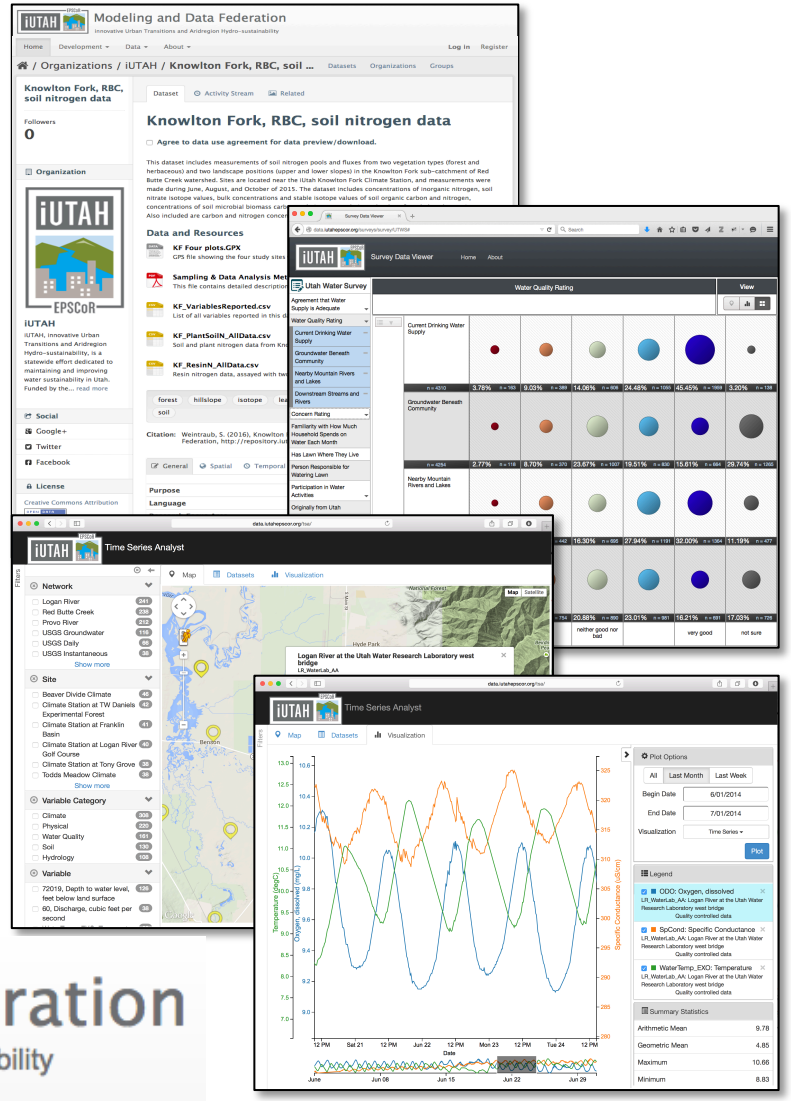
Hardware (Servers and Storage)



Operational Data Management (GAMUT Workflow)



Software for Data Sharing, Publication, and Visualization



iUTAH EPSCoR

Modeling and Data Federation

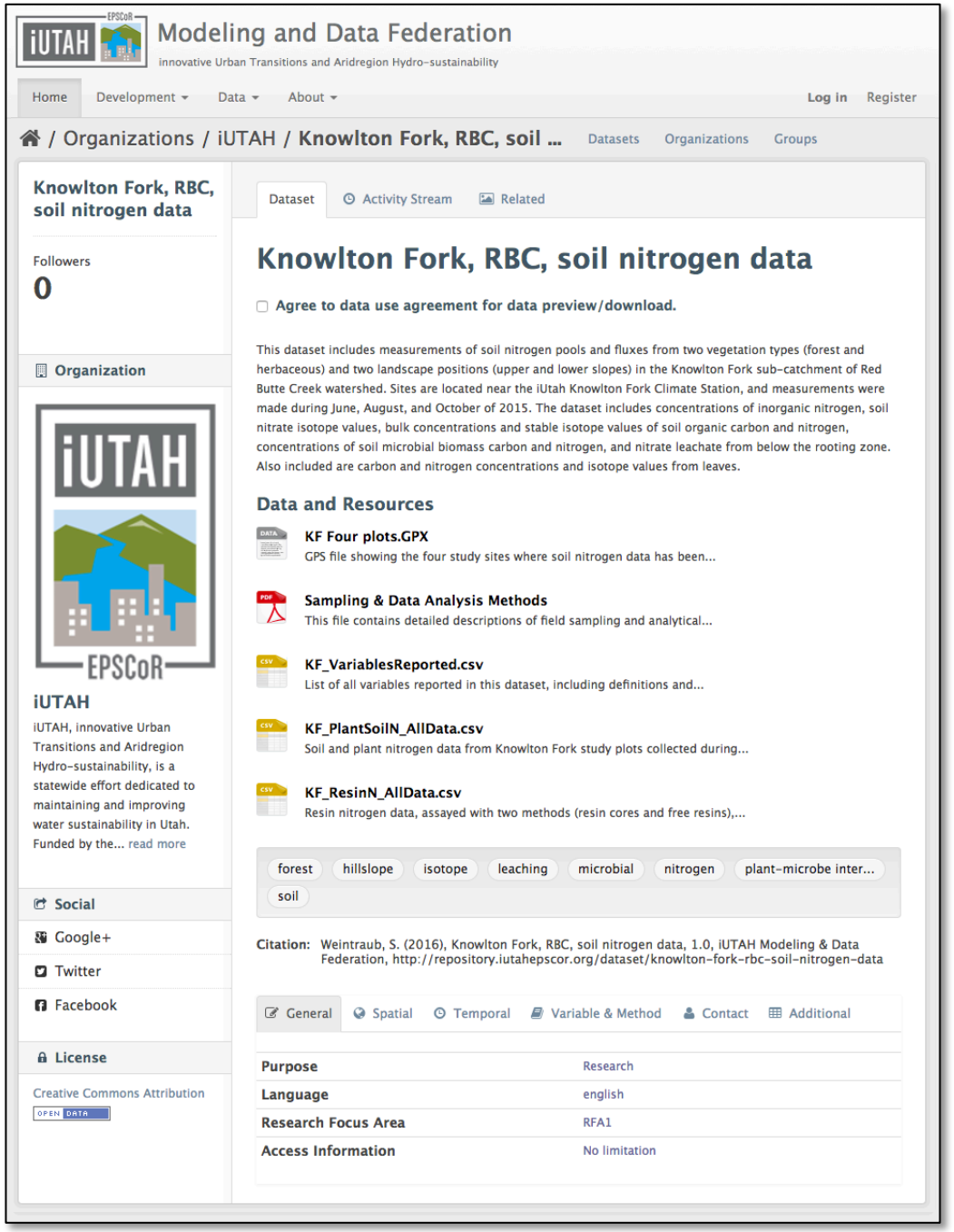
innovative Urban Transitions and Aridregion Hydro-sustainability

iUTAH Data Repository

http://repository.iutahepscor.org

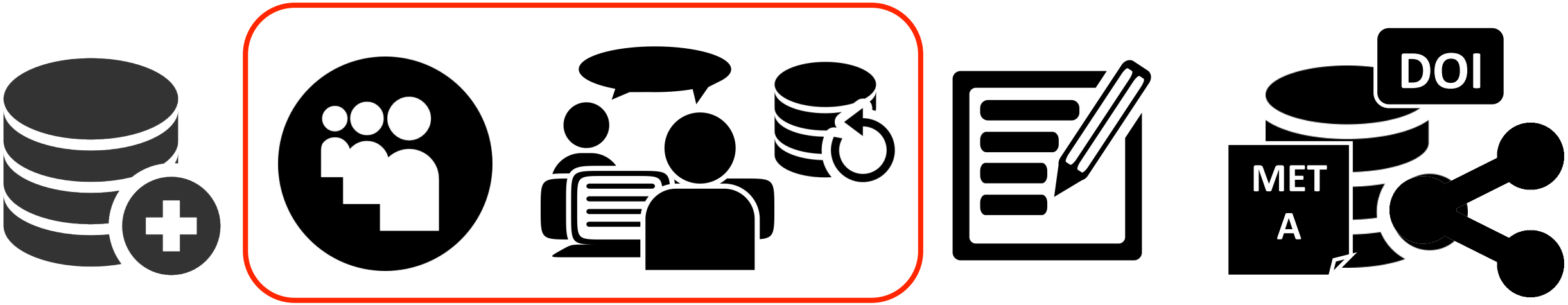
- Dataset upload and publication
- Dataset landing page
- Permanent URL and citation
- Limited curation process

- Limitations
  - Not flexible (enough)
  - Hard to maintain
  - No real collaboration capabilities
  - No DOI for published resources



What do we want to do?

- Easily create a digital instance of a dataset or model
- Quickly share it with colleagues (perhaps privately at first)
- Add value through collaboration, annotation, and iteration
- Describe with metadata
- Eventually…share publicly or formally Publish

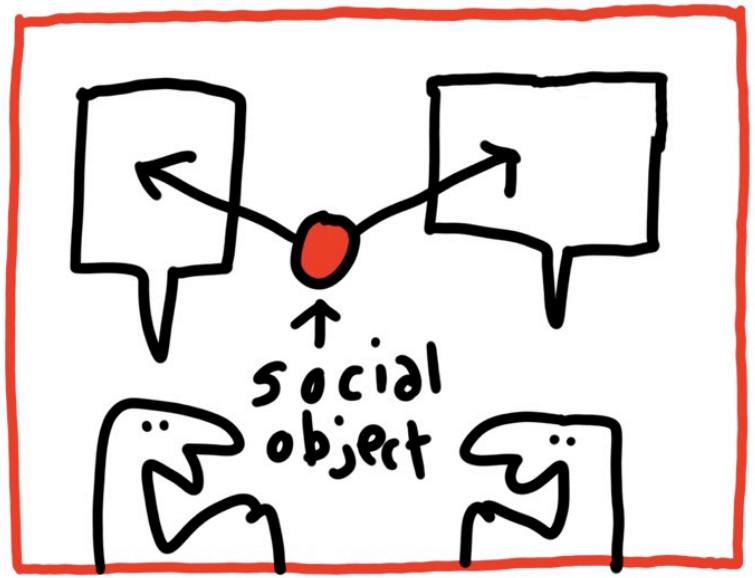


Data and models are “social objects” shared among scientists

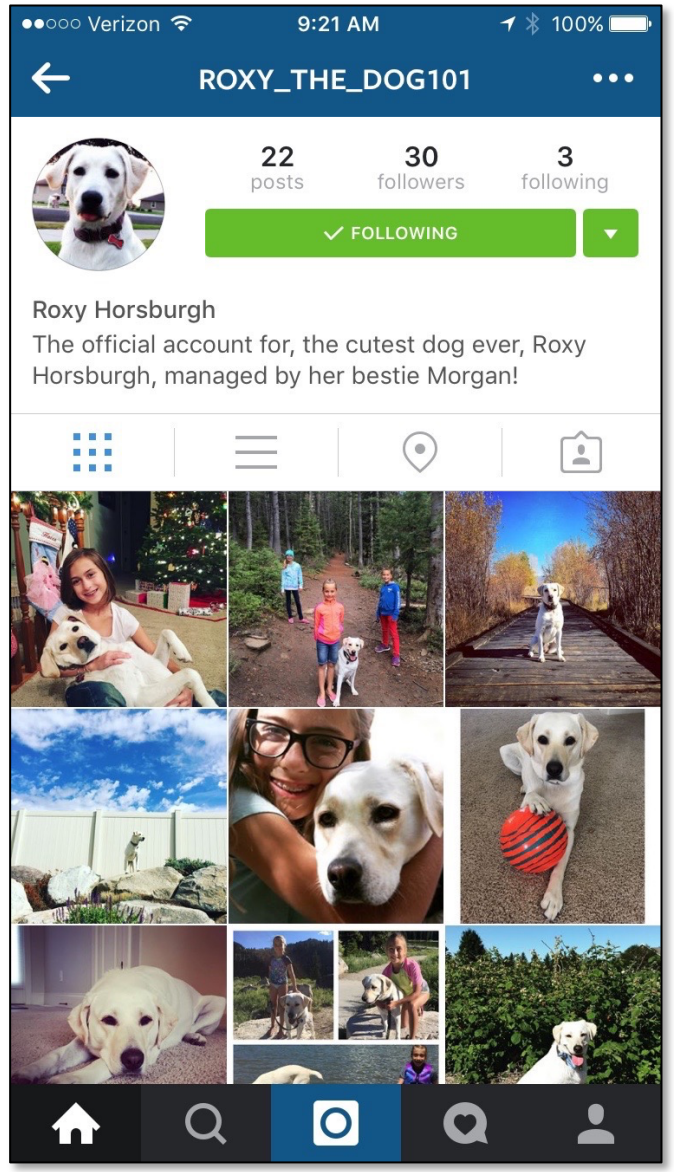
Social Objects

"Objects around which social networks form"

Jyri Engeström



Hugh MacLeod
http://sharinglab.dk/what-is-a-social-object/

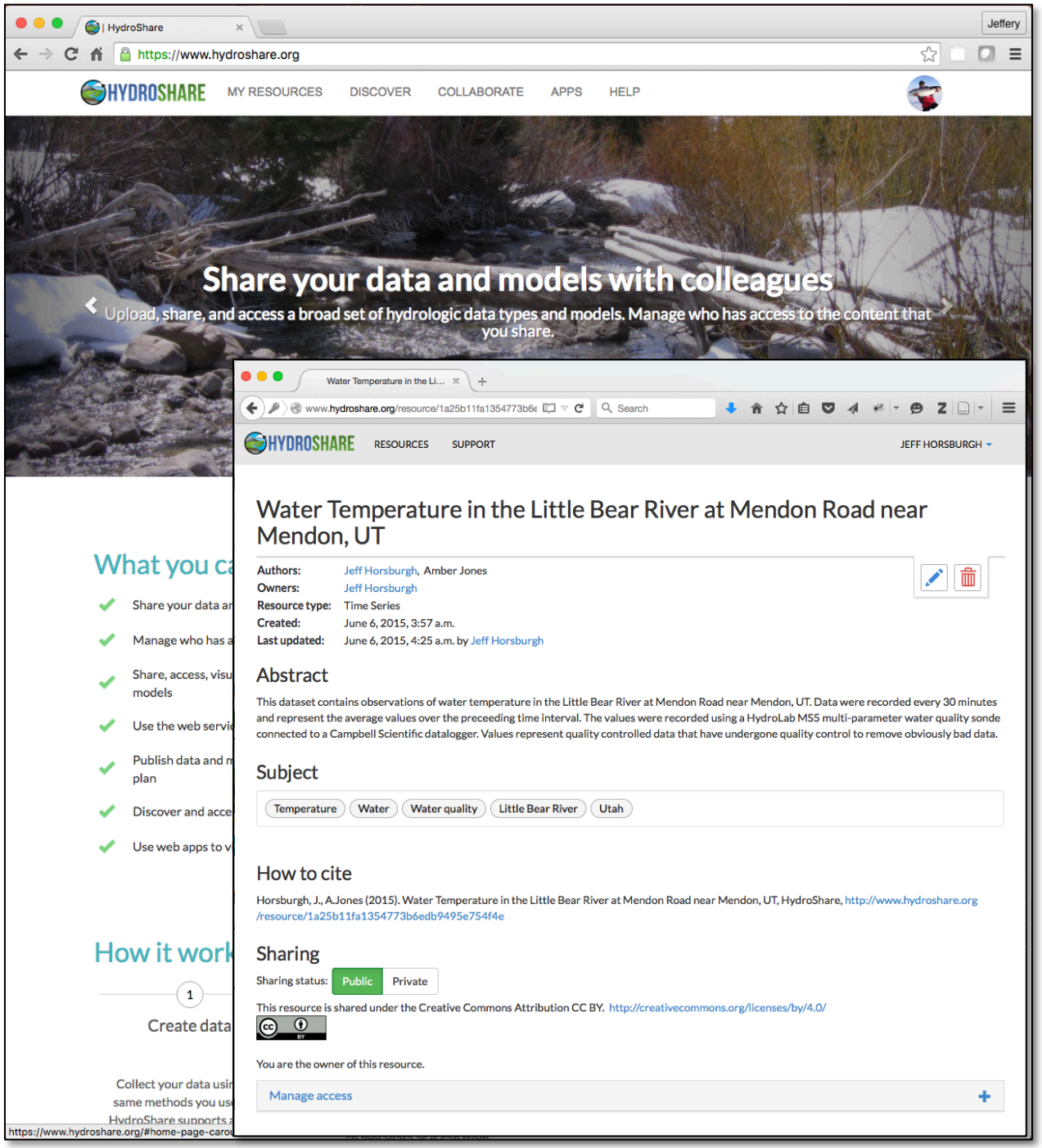


HYDROSHARE

http://www.hydroshare.org

- Collaborative development project that started about the same time as iUTAH
- Sharing and collaborating around diverse data types used by water scientists

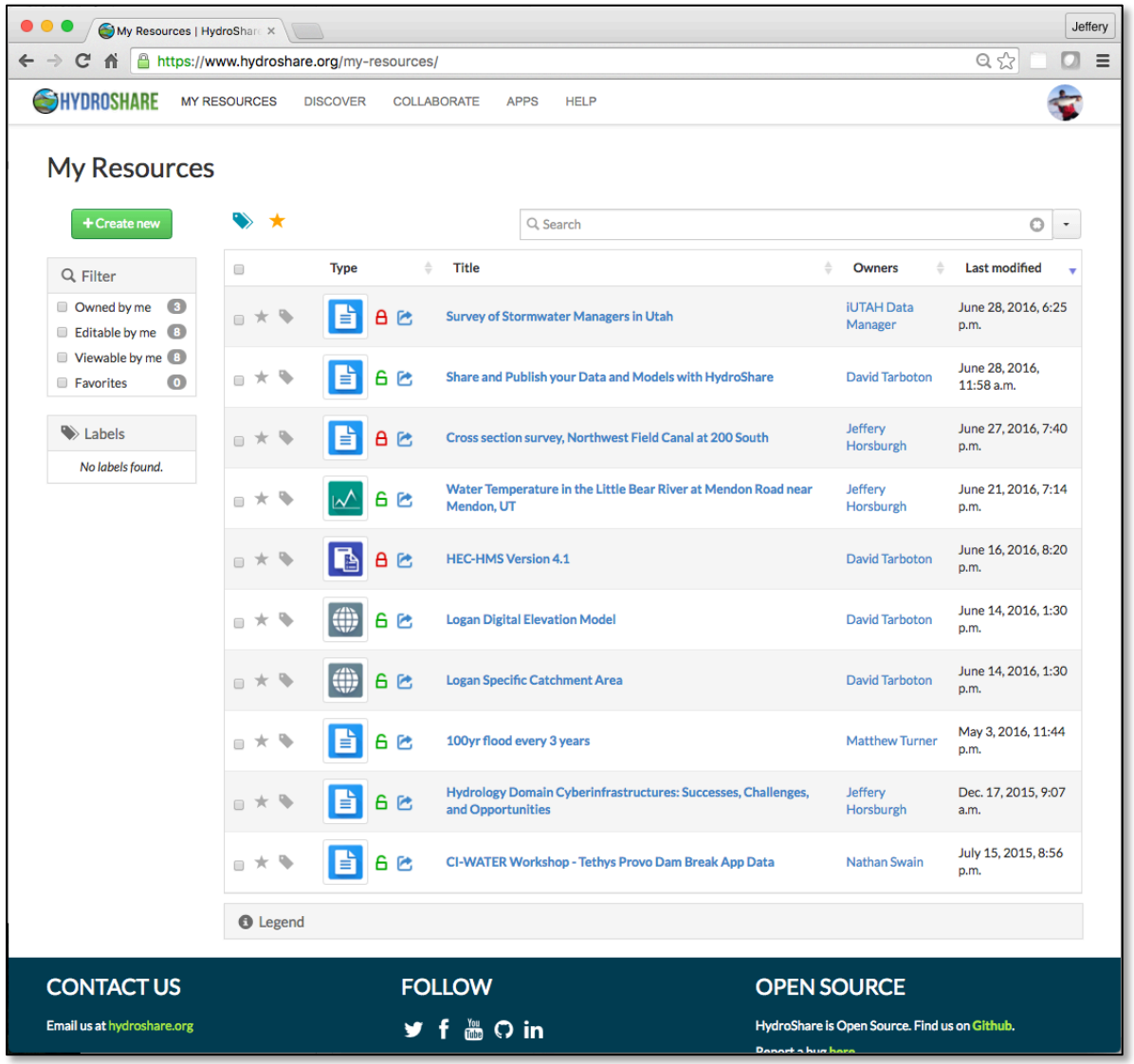
HydroShare's goal: Enable scientists to create **social objects** that add value



HydroShare "Resources"

- **Resource** = primary unit of digital content
  - Create
  - Share
  - Own
  - Access
  - Filter
  - Discover
  - Publish

Resources can be datasets, models, or other digital content

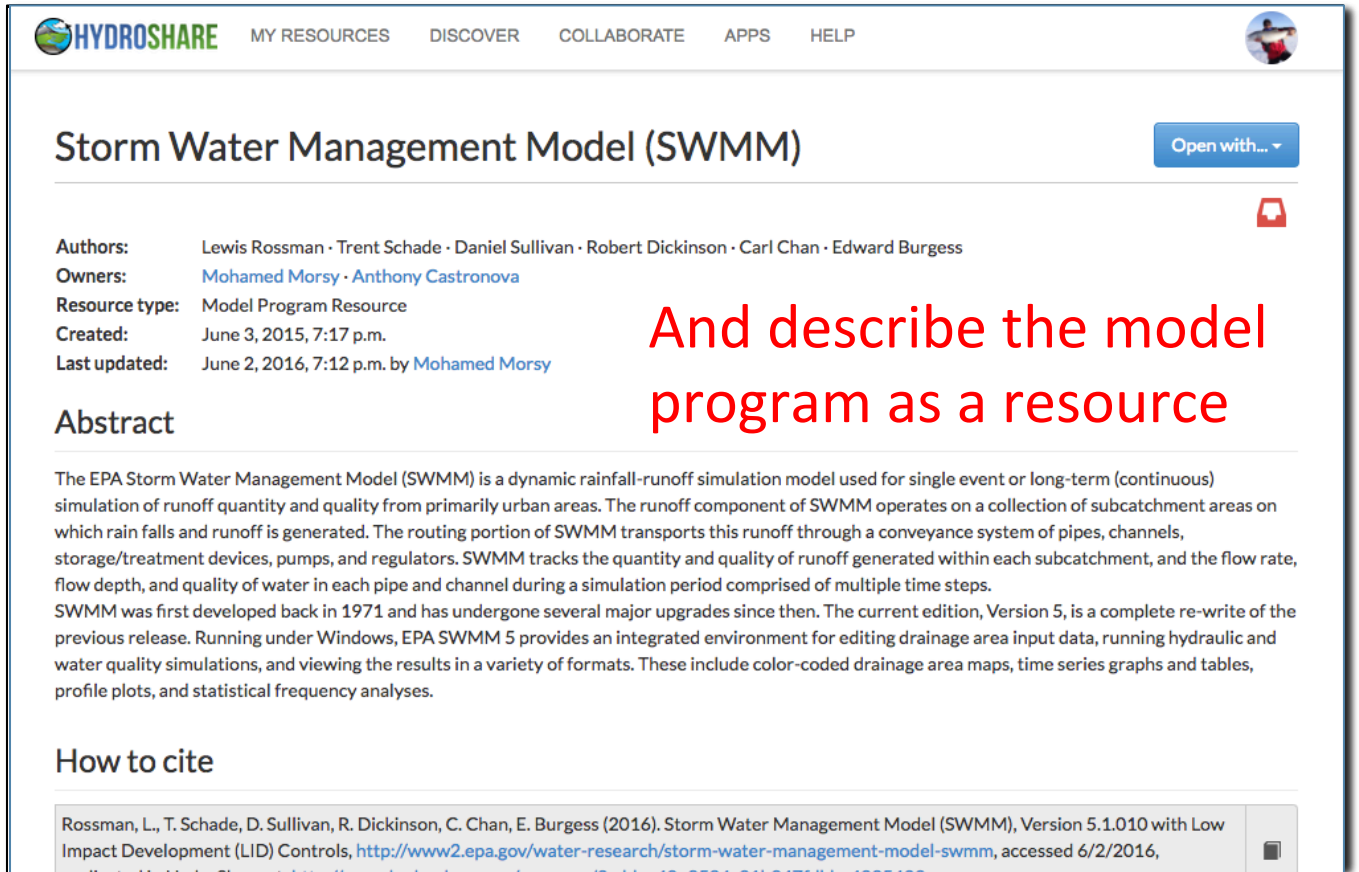


Resource Types in HydroShare

- Generic
- Multidimensional (NetCDF)
- Collections
- Geographic
  - Raster
  - Feature (ESRI Shapefiles)
- Time Series
  - Time Series
  - Referenced Time Series
- Modeling
  - Model Program
  - Model Instance
  - SWAT Model Instance
  - MODFLOW Model Instance
  - Script

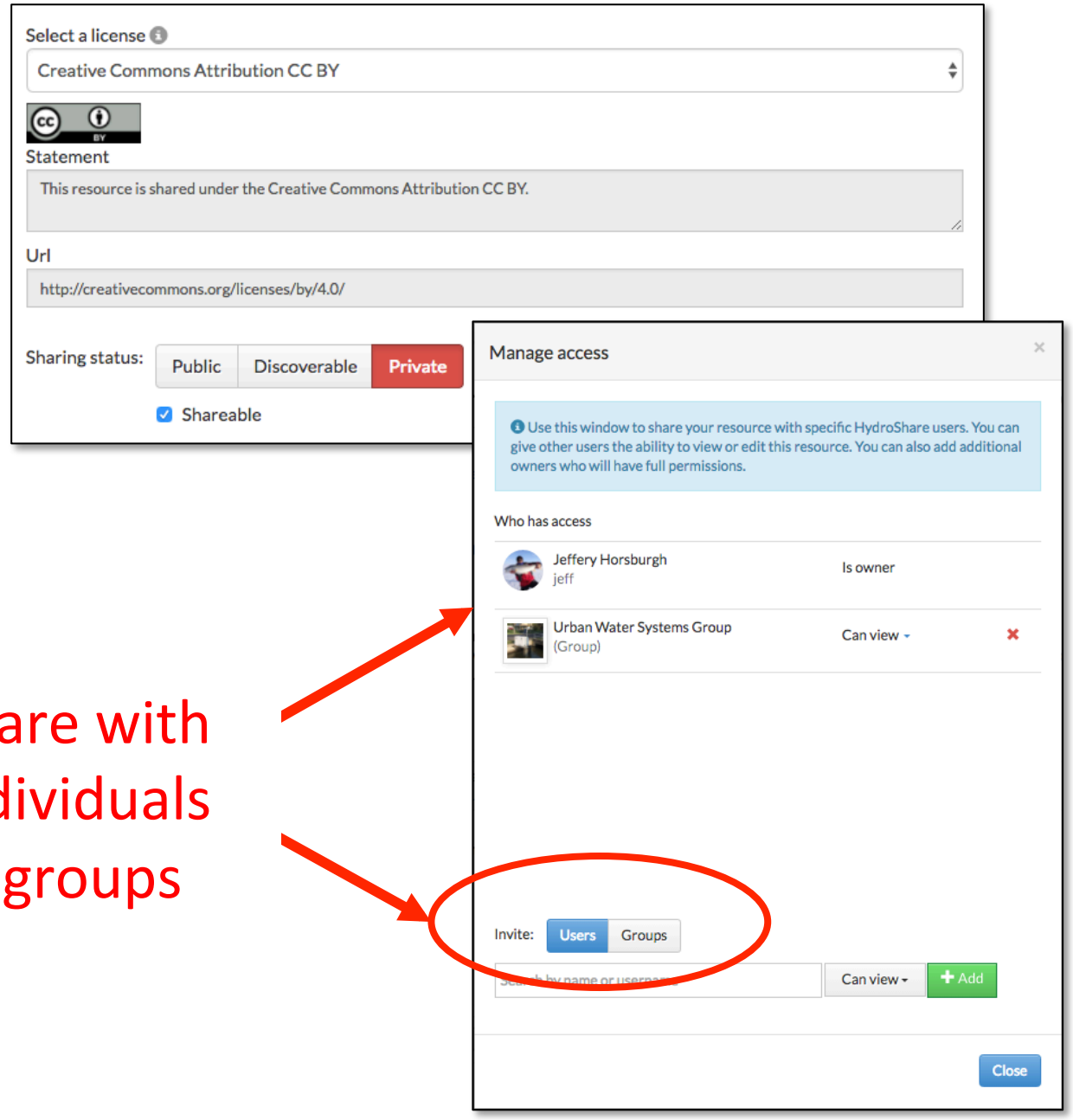
Specific Resource types have:

1. A specific metadata description
2. An expected file format and structure

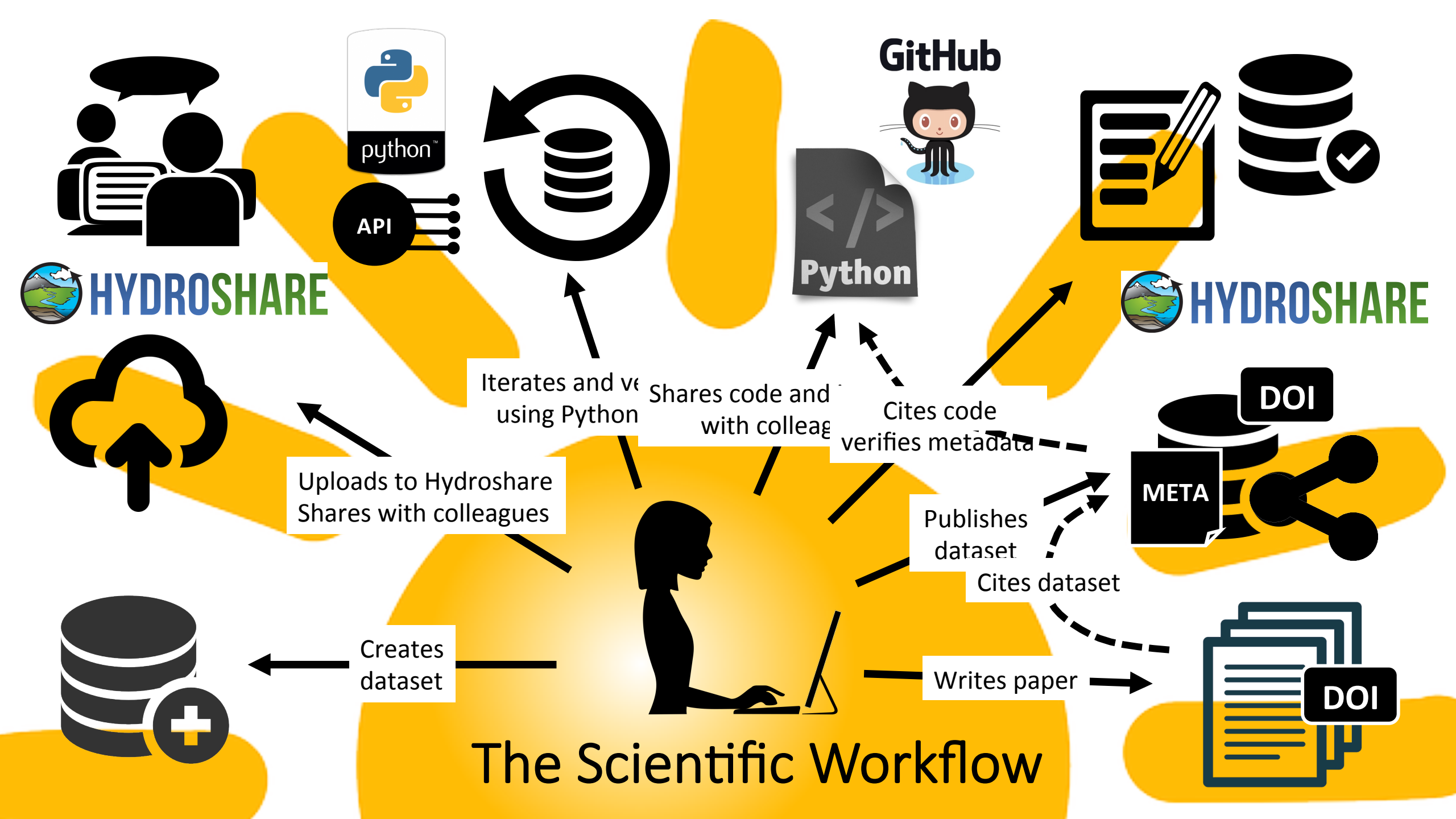


Resource Sharing in HydroShare

- You control who has access
- 5 Options:
  - **Private** – Only individual users you have given access can view metadata or access content
  - **Discoverable** – Anyone can view the metadata, but only users with permission can access content
  - **Public** – Anyone can view the metadata and access content
  - **Published** – Same as public, but the resource is made immutable and a DOI is assigned
  - **Shareable** – Other users can grant access at their same level of access



Share with individuals or groups


python
API
GitHub
Python
HYDROSHARE
HYDROSHARE
Iterates and ve
using Python
Shares code and
with colleag
Cites code
verifies metadata
Uploads to Hydroshare
Shares with colleagues
DOI
META
Publishes
dataset
Cites dataset
Creates
dataset
Writes paper
DOI
The Scientific Workflow

iUTAH Resources in HydroShare

GAMUT Raw and QC1 Datasets

HYDROSHARE MY RESOURCES DISCOVER COLLABORATE APPS HELP

iUTAH GAMUT Network Raw Data at Knowlton Fork Basic Aquatic Site (RB_KF_BA)

Open with...

Authors: iUTAH GAMUT Working Group
Owners: iUTAH Data Manager
Resource type: Generic
Created: July 8, 2016, 1:29 p.m.
Last updated: Oct. 28, 2016, 9:42 a.m. by iUTAH Data Manager

Abstract

This dataset contains raw data for all of the variables measured for the iUTAH GAMUT Network aquatic site on the Knowlton Fork tributary to Red Butte Creek (RB_KF_BA). 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 Knowlton Fork Basic Aquatic Site (RB_KF_BA), HydroShare, http://www.hydroshare.org/resource/dd1a2ffbc2e4422682a3fbd552620c32

This resource is shared under the Creative Commons Attribution CC BY. http://creativecommons.org/licenses/by/4.0/

Sharing status: Public & Shareable

You have been given specific permission to edit this resource.

Subject

Red Butte Creek | GAMUT | raw data | time series | water quality

Content

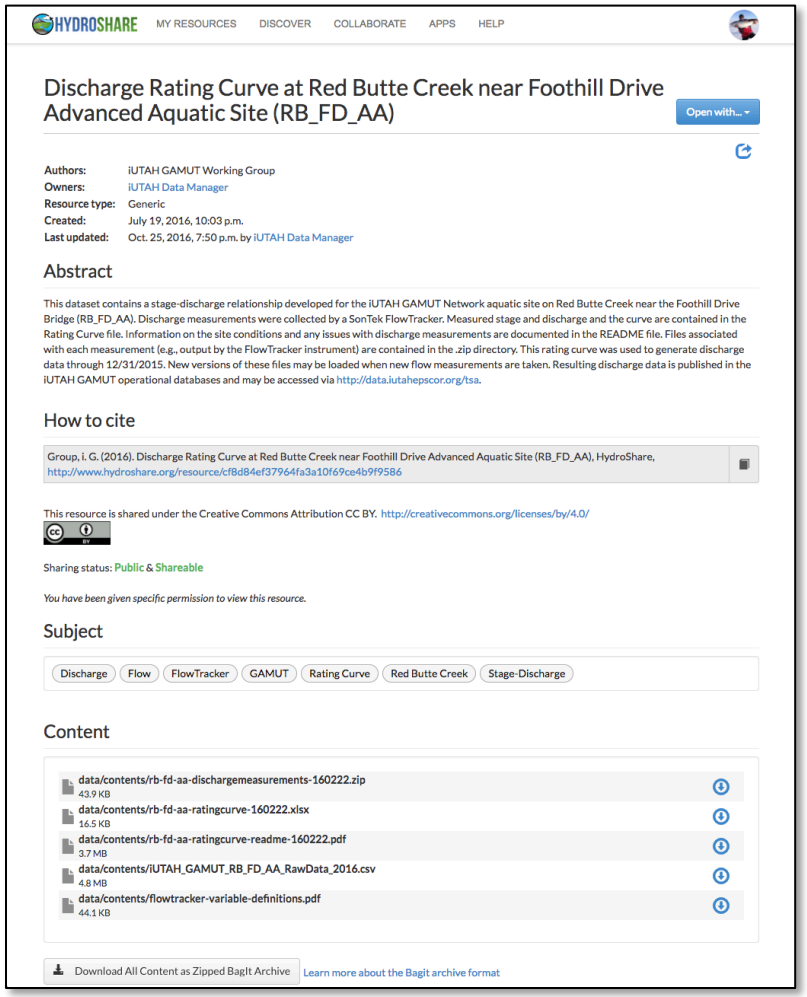
data/contents/iUTAH_GAMUT_RB_KF_BA_RawData_2013.csv 2.1 MB
data/contents/iUTAH_GAMUT_RB_KF_BA_RawData_2014.csv 6.5 MB
data/contents/iUTAH_GAMUT_RB_KF_BA_RawData_2015.csv 5.6 MB
data/contents/iUTAH_GAMUT_RB_KF_BA_RawData_2016.csv 2.8 MB

Download All Content as Zipped BagIt Archive Learn more about the Bagit archive format

Authorship | Coverage | Web Apps | Extended Met...

Authors

GAMUT Discharge Rating Curves



HYDROSHARE MY RESOURCES DISCOVER COLLABORATE APPS HELP

Discharge Rating Curve at Red Butte Creek near Foothill Drive Advanced Aquatic Site (RB_FD_AA)

Open with...

Authors: iUTAH GAMUT Working Group
Owners: iUTAH Data Manager
Resource type: Generic
Created: July 19, 2016, 10:03 p.m.
Last updated: Oct. 25, 2016, 7:50 p.m. by iUTAH Data Manager

Abstract

This dataset contains a stage-discharge relationship developed for the iUTAH GAMUT Network aquatic site on Red Butte Creek near the Foothill Drive Bridge (RB_FD_AA). Discharge measurements were collected by a SonTek FlowTracker. Measured stage and discharge and the curve are contained in the Rating Curve file. Information on the site conditions and any issues with discharge measurements are documented in the README file. Files associated with each measurement (e.g., output by the FlowTracker instrument) are contained in the .zip directory. This rating curve was used to generate discharge data through 12/31/2015. New versions of these files may be loaded when new flow measurements are taken. Resulting discharge data is published in the iUTAH GAMUT operational databases and may be accessed via http://data.iutahepscor.org/tsa.

How to cite

Group, i. G. (2016). Discharge Rating Curve at Red Butte Creek near Foothill Drive Advanced Aquatic Site (RB_FD_AA), HydroShare, http://www.hydroshare.org/resource/cf8d84ef37964fa3a10f69ce4b9f9586

This resource is shared under the Creative Commons Attribution CC BY. http://creativecommons.org/licenses/by/4.0/

Sharing status: Public & Shareable

You have been given specific permission to view this resource.

Subject

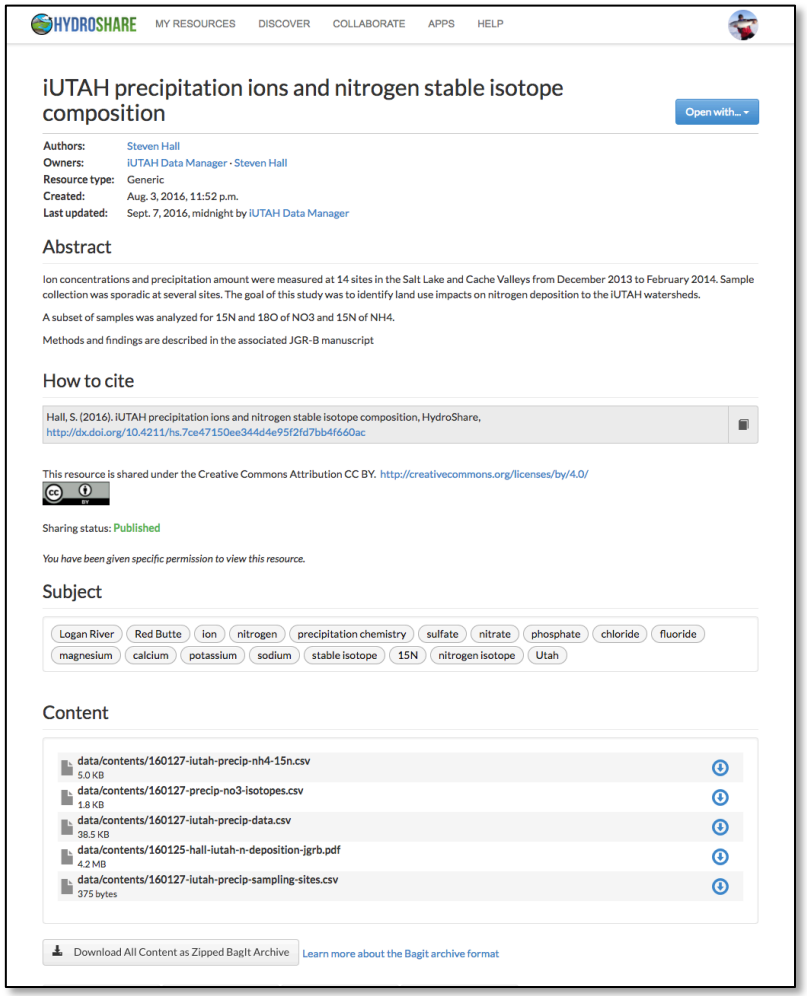
Discharge | Flow | FlowTracker | GAMUT | Rating Curve | Red Butte Creek | Stage-Discharge

Content

data/contents/rb-fd-aa-dischargemeasurements-160222.zip 43.9 KB
data/contents/rb-fd-aa-ratingcurve-160222.xlsx 16.5 KB
data/contents/rb-fd-aa-ratingcurve-readme-160222.pdf 3.7 MB
data/contents/iUTAH_GAMUT_RB_FD_AA_RawData_2016.csv 4.8 MB
data/contents/flowtracker-variable-definitions.pdf 44.1 KB

Download All Content as Zipped BagIt Archive Learn more about the Bagit archive format

Individual Investigator Datasets



HYDROSHARE MY RESOURCES DISCOVER COLLABORATE APPS HELP

iUTAH precipitation ions and nitrogen stable isotope composition

Open with...

Authors: Steven Hall
Owners: iUTAH Data Manager · Steven Hall
Resource type: Generic
Created: Aug. 3, 2016, 11:52 p.m.
Last updated: Sept. 7, 2016, midnight by iUTAH Data Manager

Abstract

Ion concentrations and precipitation amount were measured at 14 sites in the Salt Lake and Cache Valleys from December 2013 to February 2014. Sample collection was sporadic at several sites. The goal of this study was to identify land use impacts on nitrogen deposition to the iUTAH watersheds.

A subset of samples was analyzed for 15N and 18O of NO3 and 15N of NH4.

Methods and findings are described in the associated JGR-B manuscript

How to cite

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

This resource is shared under the Creative Commons Attribution CC BY. http://creativecommons.org/licenses/by/4.0/

Sharing status: Published

You have been given specific permission to view this resource.

Subject

Logan River | Red Butte | ion | nitrogen | precipitation chemistry | sulfate | nitrate | phosphate | chloride | fluoride | magnesium | calcium | potassium | sodium | stable isotope | 15N | nitrogen isotope | Utah

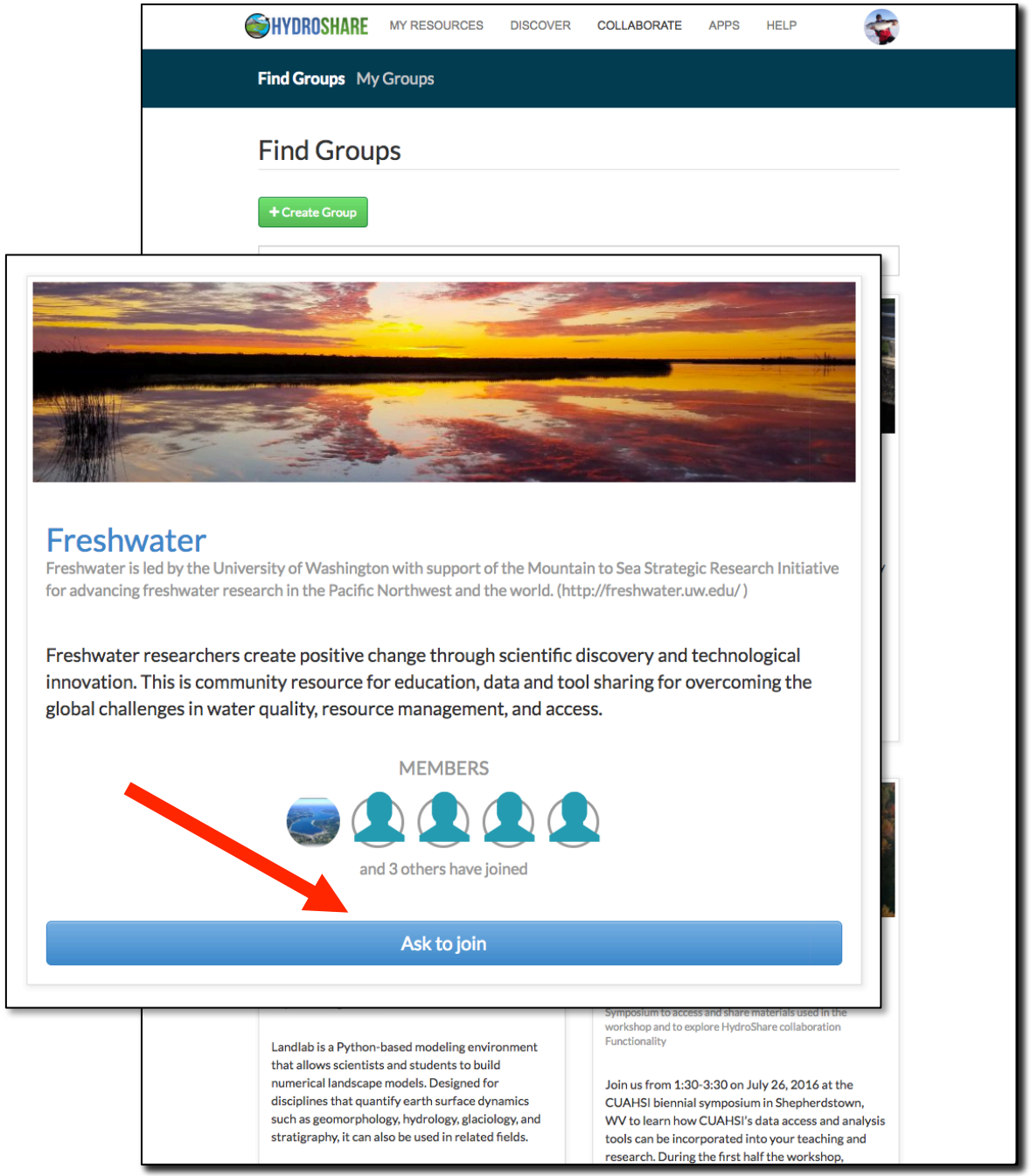
Content

data/contents/160127-iutah-precip-nh4-15n.csv 5.0 KB
data/contents/160127-precip-no3-isotopes.csv 1.8 KB
data/contents/160127-iutah-precip-data.csv 38.5 KB
data/contents/160125-hall-iutah-n-deposition-jgrb.pdf 4.2 MB
data/contents/160127-iutah-precip-sampling-sites.csv 375 bytes

Download All Content as Zipped BagIt Archive Learn more about the Bagit archive format

Collaborative Groups

- Create a collaborative group
- Share resources with a group
- Manage group membership
- Find groups you are interested in
- Request group membership



Collaborative Capabilities of HydroShare

- Dataset/model creation as a “Resource” – **the Social Objects!**
- Resource sharing (public and private)
- Collaborative groups and sharing within groups
- Resource versioning
- Formal Publication of resources and assignment of DOIs
- Rating and commenting on resources

Moving iUTAH Repository to HydroShare: Next Steps

- Your data collection plans are still valid and your datasets still need to be published
- Modelers – there is now a way for you to share your results!

- If you had a dataset in http://repository.iutahepscor.org - we have moved it to HydroShare for you
  - **You will receive an email from the iUTAH Data Manager** with specific instructions for:
    - Creating an account in HydroShare
    - Asserting ownership of your resource
    - Publishing your dataset
- If you have not submitted your data yet
  - Work directly with Amber to get it set up in HydroShare
  - We'll be making some instructional videos

Reusability of CI Components

- We have worked hard to ensure that what we have developed for iUTAH is reusable
- All of our source code is in open source code repositories in GitHub
- We have published papers about many of the things we have done

GitHub

Environ Monit Assess (2015) 187:348
DOI 10.1007/s10661-015-4594-3

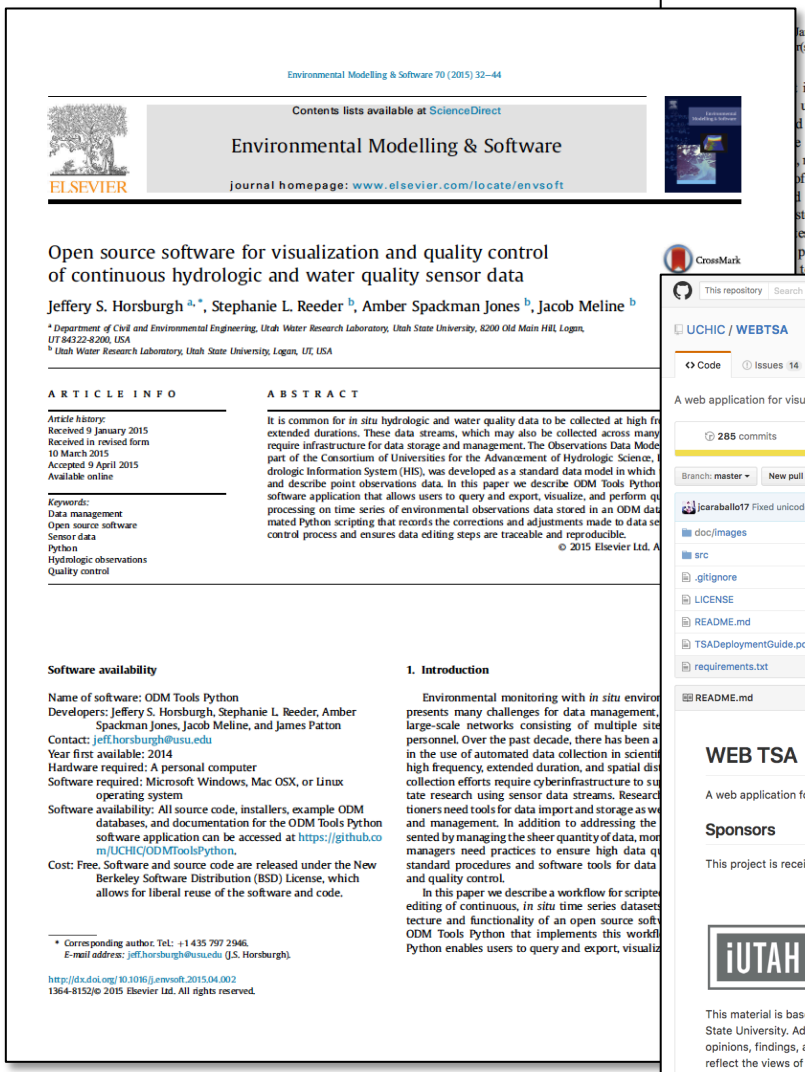
A data management and publication workflow for a large-scale, heterogeneous sensor network

Amber Spackman Jones · Jeffery S. Horsburgh · Stephanie L. Reeder · Maurier Ramírez · Juan Caraballo

anuary 2015 / Accepted: 5 May 2015
r(s) 2015. This article is published with open access at Springerlink.com

is common for hydrology researchers to using in situ sensors at high frequencies, d durations, and with spatial distributions e data volumes requiring infrastructure for management, and sharing. The availability of these data in addressing scientific ques- l to water availability, water quality, and sters relies on effective cyberinfrastructure es transformation of raw sensor data into products. It also depends on the ability of to share and access the data in useable

and terrestrial sites for continuous monitoring of common meteorological variables, snow accumulation and melt, soil moisture, surface water flow, and surface water quality. We present the overall workflow we have developed for effectively transferring data from field monitoring sites to ultimate end-users and describe the software tools we have deployed for storing, managing, and sharing the sensor data. These tools are all open source and available for others to use.



Environmental Modelling & Software 70 (2015) 32–44

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Open source software for visualization and quality control of continuous hydrologic and water quality sensor data

Jeffery S. Horsburgh [a, *], Stephanie L. Reeder [b], Amber Spackman Jones [b], Jacob Meline [b]

[a] Department of Civil and Environmental Engineering, Utah Water Research Laboratory, Utah State University, 8200 Old Main Hill, Logan, UT 84322-8200, USA
[b] Utah Water Research Laboratory, Utah State University, Logan, UT, USA

ARTICLE INFO

Article history:
Received 9 January 2015
Received in revised form
10 March 2015
Accepted 9 April 2015
Available online

Keywords:
Data management
Open source software
Sensor data
Python
Hydrologic observations
Quality control

ABSTRACT

It is common for *in situ* hydrologic and water quality data to be collected at high fr extended durations. These data streams, which may also be collected across many require infrastructure for data storage and management. The Observations Data Mode part of the Consortium of Universities for the Advancement of Hydrologic Science, I drologic Information System (HIS), was developed as a standard data model in which and describe point observations data. In this paper we describe ODM Tools Python software application that allows users to query and export, visualize, and perform qu processing on time series of environmental observations data stored in an ODM data mated Python scripting that records the corrections and adjustments made to data se control process and ensures data editing steps are traceable and reproducible.

© 2015 Elsevier Ltd. A

Software availability

Name of software: ODM Tools Python
Developers: Jeffery S. Horsburgh, Stephanie L. Reeder, Amber Spackman Jones, Jacob Meline, and James Patton
Contact: jeff.horsburgh@usu.edu
Year first available: 2014
Hardware required: A personal computer
Software required: Microsoft Windows, Mac OSX, or Linux operating system
Software availability: All source code, installers, example ODM databases, and documentation for the ODM Tools Python software application can be accessed at https://github.com/UCHIC/ODMToolsPython.
Cost: Free. Software and source code are released under the New Berkeley Software Distribution (BSD) License, which allows for liberal reuse of the software and code.

1. Introduction

Environmental monitoring with *in situ* environ presents many challenges for data management, large-scale networks consisting of multiple site personnel. Over the past decade, there has been a in the use of automated data collection in scientif high frequency, extended duration, and spatial dist collection efforts require cyberinfrastructure to su tate research using sensor data streams. Research tioners need tools for data import and storage as we and management. In addition to addressing the sented by managing the sheer quantity of data, mor managers need practices to ensure high data qu standard procedures and software tools for data and quality control.

In this paper we describe a workflow for scripte editing of continuous, *in situ* time series datasets tecture and functionality of an open source soft ODM Tools Python that implements this workfl Python enables users to query and export, visualiz

UCHIC / WEBTSA

Unwatch 13 Star 2 Fork 1

Code Issues 14 Pull requests 0 Projects 0 Wiki Pulse Graphs Settings

A web application for visualizing time series of environmental observations. — Edit

285 commits 2 branches 0 releases 4 contributors BSD-3-Clause

Branch: master New pull request Create new file Upload files Find file Clone or download

jcaraballo17 Fixed unicode decoding error for dataseries resource. Latest commit 89b33a2 9 days ago

doc/images	Initial repository setup.	3 years ago
src	Fixed unicode decoding error for dataseries resource.	9 days ago
.gitignore	Changed project structure for Django library upgrade.	a year ago
LICENSE	Update LICENSE	3 years ago
README.md	Added the project dependencies.	2 years ago
TSADeploymentGuide.pdf	Added deployment guide.	a year ago
requirements.txt	Changed project structure for Django library upgrade.	a year ago

README.md

WEB TSA

A web application for visualizing environmental time series data.

Sponsors

This project is receiving or has received support from a number of agencies and complementary efforts including:

iUTAH EPSCoR CUAHSI NSF

This material is based on work supported by the National Science Foundation Grant EPS 1208732 awarded to Utah State University. Additional support was previously provided by National Science Foundation grant EAR 0622374. Any opinions, findings, and conclusions or recommendations expressed are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Opportunities

- Better tools for working with heterogeneous data
- Software/apps/systems that enhance our “personal cyberinfrastructure”
- Freedom to move data and models from one platform/app/software to another
- Training a next generation of “cyber-savvy” engineers and scientists

An enhanced, cloud connected, and social scientific workflow