HydroShare tools and recommended practices for sharing and publishing data and models in support of collaborative reproducible research

David Tarboton[1], Ray Idaszak[2], Jeffery Horsburgh[1], Daniel Ames[3], Jonathan Goodall[4], Alva Couch[5], Pabitra Dash[1], Hong Yi[2], Christina Bandaragoda[6], Anthony Castronova[7], Martyn Clark[8], and Shaowen Wang[9]

[1]Utah State University
[2]Renaissance Computing Institute
[3]Brigham Young University
[4]University of Virginia Main Campus
[5]Tufts University
[6]University of Washington
[7]Consortium of Universities for the Advancement of Hydrological Science
[8]NCAR
[9]University of Illinois at Urbana Champaign

November 21, 2022

Abstract

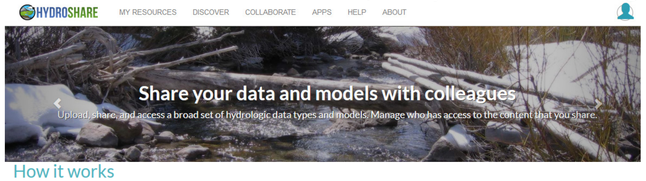
HydroShare is a domain specific data and model repository operated by the Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI) to advance hydrologic science by enabling individual researchers to more easily share products resulting from their research. The community platform supports, not just the scientific publication summarizing a study, but also the data, models and workflow scripts used to create the scientific publication and reproduce the results therein. HydroShare accepts data from anybody, and supports Findable, Accessible, Interoperable and Reusable (FAIR) principles. HydroShare is comprised of two sets of functionality: (1) a repository for users to share and publish data and models, collectively referred to as resources, in a variety of formats, and (2) tools (web apps) that can act on content in HydroShare and support web based access to compute capability. Together these serve as a platform for collaboration and computation that integrates data storage, organization, discovery, and analysis through web applications (web apps) and that allows researchers to employ services beyond the desktop to make data storage and manipulation more reliable and scalable, while improving their ability to collaborate and reproduce results. This presentation will describe the capabilities developed for HydroShare to support the full research data management life cycle. Data can be entered into HydroShare as soon as it is collected, and initially shared only with the team directly working on the data. As analysis proceeds, tools, scripts and models that act on the data to produce research results may be stored in HydroShare resources alongside the data. At the time of publication these resources may be permanently published and receive digital object identifiers and cited in research papers. Resources may themselves include citations to the research papers, thereby linking the publications to the supporting data, scripts and models. HydroShare design choices and capabilities for establishing relationships and versioning, based on simplicity, and ease of use, and some of the challenges encountered, will be discussed.

IN53E-0656: HydroShare tools and recommended practices for sharing and publishing data and models in support of collaborative reproducible research

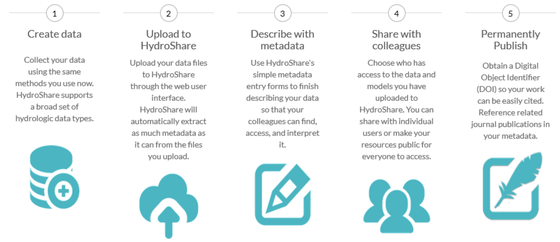
David Tarboton, Ray Idaszak, Jeffery Horsburgh, Dan Ames, Jon Goodall, Alva Couch, Pabitra Dash, Hong Yi, Christina Bandaragoda, Tony Castronova, Martyn Clark, Shaowen Wang and HydroShare Development Team

What is HydroShare ?

An online hydrologic information system for sharing data, models and code to enable more rapid advances in hydrologic understanding through collaborative research, analysis and modeling.



How it works



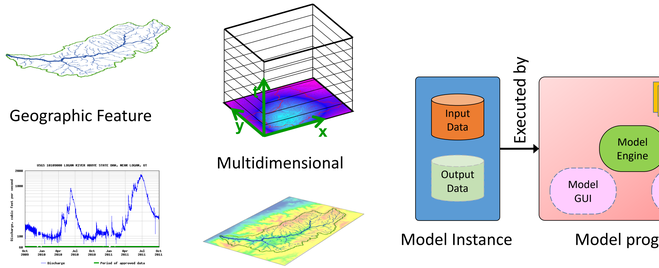
What you can do with HydroShare

- Share your data and models with colleagues
- Manage who has access to the content that you share
- Share, access, visualize and manipulate a broad set of hydrologic data types and models
- Use the web services API to program automated and client access
- Publish data and models to meet the requirements of your data management plan
- Discover and access data and models published by others
- Use web apps to visualize, analyze and run models on data in HydroShare

www.hydroshare.org

What can you store in HydroShare ?

1. In HydroShare, data and model files are stored as **resources**.
2. HydroShare supports any file, including several specific data formats.
3. Content "aggregations" hold data formats common in hydrology and support description with additional content specific metadata. Apps can act on specific content types.
4. Collections group together multiple resources related to a project or study.
5. Model Programs and Model Instances hold specific hydrologic models and associated data for application at a location.



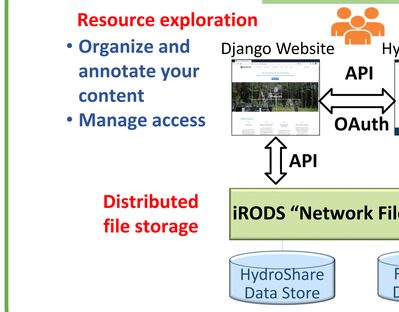
Why HydroShare ?

Collaboration: Share your data and model files; integrate information from multiple sources; organize individual, team, and group work.
Reproducibility, transparency and trust: Publish your work in any format, including data and models with a citable digital object identifier (DOI).
Do Science: Run Apps and models from a browser without installing software; access computational services for your big data and model analysis.
Learning: Use a platform where all students have access to the same functionality regardless of their computer.

HydroShare is a system to advance hydrologic science by enabling the community to more easily and freely share products resulting from their research, not just the scientific publication summarizing a study, but also the data and models used to create the scientific publication.

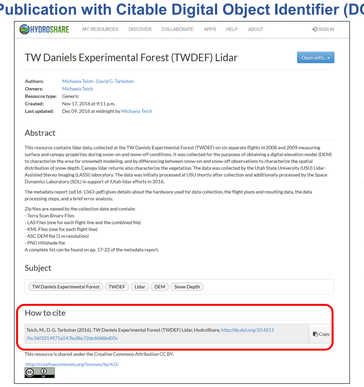
- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**eusable

Design

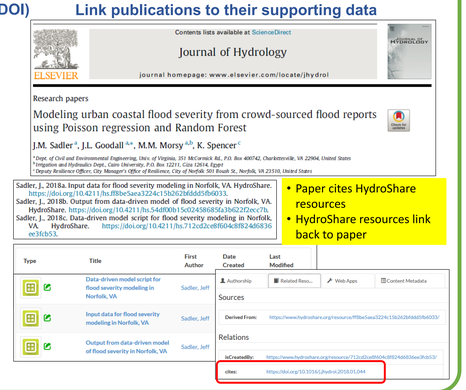


Publishing data and models

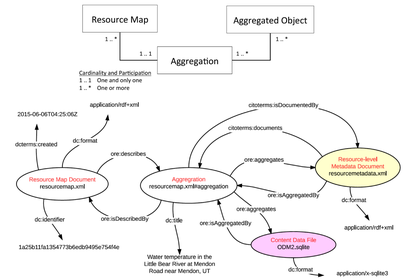
Publication with Citable Digital Object Identifier (DOI)



Link publications to their supporting data



OAI-ORE standard based Resource Data Model



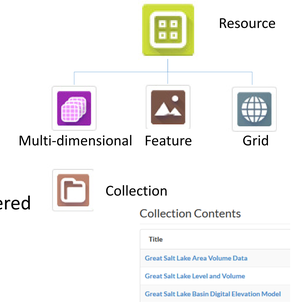
Dublin Core machine readable metadata and data model to make data in HydroShare, Findable, Accessible, Interoperable, Reusable

Resources, comprised of data and models, are framed as social objects, the basis for collaboration and interaction

Horsburgh, J. S., et al., (2016), "Hydroshare: Sharing Diverse Environmental Data Types and Models as Social Objects with Application to the Hydrology Domain," JAWRA, http://dx.doi.org/10.1111/1752-1688.12363.

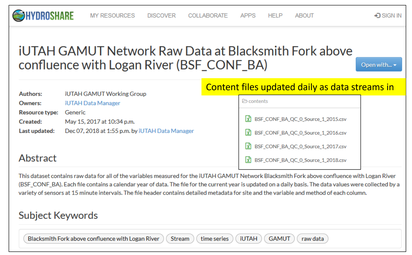
Resource Organization

- A resource can hold multiple aggregations
  - Each being a different type of data with its own set of metadata
  - Managed as one discoverable resource
  - One set of access controls (Owners, Editors etc.)
  - One unique identifier
  - One set of resource level metadata
- A collection can hold multiple resources
- Collections and their members may each be discovered separately
- Unique keyword tags form informal collections (e.g. "AGU2018")

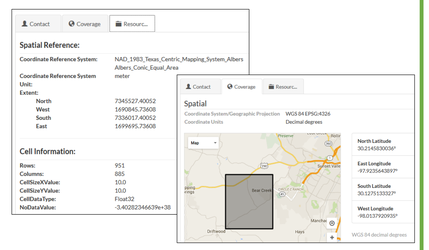


Key Functionality

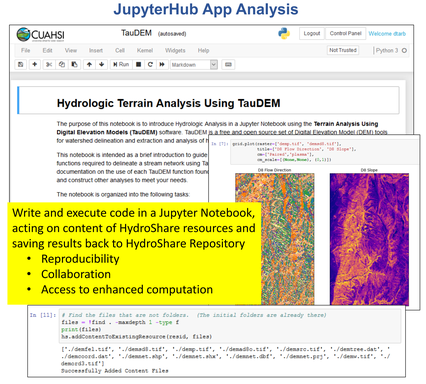
Data streamed into HydroShare as soon as it is collected



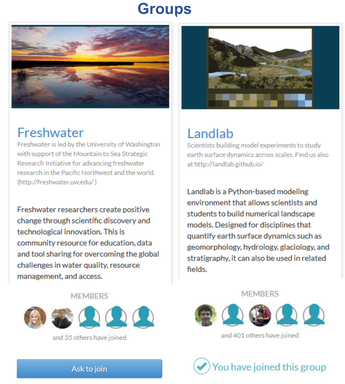
Metadata harvested automatically or captured via simple web page editing



JupyterHub App Analysis



Groups



OAC-1664061
OAC-1664018
OAC-1664119

HydroShare is operated by CUAHSI with ongoing development through a collaborative project among Utah State University, RENCI University of North Carolina, Brigham Young University, CyberGIS Center University of Illinois, Tufts, University of Virginia, NCAR, and University of Washington.