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)

- VMWare Virtual Machines
  - File Server
  - Map Server
  - Database Server
  - Web Server

- VMWare Virtual Hosts

Tiered Storage

Operational Data Management (GAMUT Workflow)

Software for Data Sharing, Publication, and Visualization

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
  o Not flexible (enough)
  o Hard to maintain
  o No real collaboration capabilities
  o 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/
• 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

And describe the model program as a resource

Storm Water Management Model (SWMM)

Author: Lewis Rossmann, Trent Schade, Daniel Sullivan, Robert Dickinson, Carl Chan, Edward Burgess
Owners: Mohamed Morsey, Anthony Castronova
Resource type: Model Program Resource
Created: June 3, 2015, 7:17 a.m.
Last updated: June 2, 2016, 7:12 p.m. by Mohamed Morsey

Abstract

The EPA Storm Water Management Model (SWMM) is a dynamic rainfall-runoff simulation model used for single event or long-term (continuous) simulation of runoff quantity and quality from primarily urban areas. The runoff component of SWMM operates on a collection of subcatchment areas on which rain falls and runoff is generated. The routing portion of SWMM transports this runoff through a conveyance system of pipes, channels, storage/treatment devices, pumps, and regulators. SWMM tracks the quantity and quality of runoff generated within each subcatchment, and the flow rate, flow depth, and quality of water in each pipe and channel during a simulation period comprised of multiple time steps. SWMM was first developed back in 1971 and has undergone several major upgrades since then. The current version, Version 5.5, is a complete rewrite of the previous release. Running under Windows, EPA SWMM 5 provides an integrated environment for editing drainage area input data, running hydraulic and water quality simulations, and viewing the results in a variety of formats. These include color-coded drainage area maps, time series graphs and tables, profile plots, and statistical frequency analyses.

How to cite

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.
The Scientific Workflow

- Creates dataset
- Uploads to Hydroshare
- Shares with colleagues
- Iterates and versions using Python
- Shares code and iterates with colleagues
- Creates and verifies metadata
- Publishes dataset
- Cites dataset
- Writes paper
- Cites code
iUTAH Resources in HydroShare

GAMUT Raw and QC1 Datasets

GAMUT Discharge Rating Curves

Individual Investigator Datasets
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
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