Cyberinfrastructure to Support Water Science Education and Reproducible Science

TONY CASTRONOVA
HYDROLOGIC SCIENTIST
CUAHSI
CUAHSI Cyberseminar Series

**April 13:** Data Archiving and Dissemination Tools to Support Water Science Research  
  ◦ Liza Brazil – Product Manager

**April 20:** Cloud-hosting Water Science Data for Collaborative Research  
  ◦ Mark Henderson – Senior Software Engineer

**April 27:** Cyberinfrastructure to Support Water Science Education and Reproducible Science  
  ◦ *Anthony Castronova – Hydrologic Scientist*

https://www.cuahsi.org/education/cyberseminars/archives/  
https://www.hydroshare.org/resource/1846fd3dce53418e8f33e95b964dcdd8/
Outline

I. Brief Introduction to CUAHSI
II. HydroShare Recap
III. CUAHSI JupyterHub Architecture
IV. Getting Started with CUAHSI JupyterHub
V. Reproducible Workflows
VI. Next Steps
Who are we?

• CUAHSI is a 501(c)3 Non-Profit Consortium of about 130 U.S. Academic Institutions, Non-Profits, and International Universities; also open to private organizations

CUAHSI's mission is to shape the future of hydrologic science by:
  ◦ Strengthening collaboration
  ◦ Developing and delivering data, models, instrumentation and technologies
  ◦ Promoting education

• Key Activities
  • **Community Services** – workshops, community meetings, training, etc.
  • **Data and Model Services** – software to support science, including HIS and HydroShare
Community Services

Resources and training to build capacity and extend capabilities

- Learn a new instrumentation method
- Engage in community on water issues
- Develop or improve data tools
- Add a field site to existing research
- The Summer Innovators program - contribute to the National Water Center goals
- Collaborate with researchers at other institutions
2018 CUAHSI Biennial Colloquium

July 29 – August 2 at the National Conservation Training Center in Shepherdstown, WV

- Hydrologic Connections: Climate, Food, Energy, Environment, and Society
- Workshops on UAVs and NASA SWOT mission
- Sessions on water and climate, hydrologic feedbacks, energy-water systems, etc.
Data Services

Access to data and models to facilitate research and education.

- Tools to enable science
- Share, discover, analyze, collaborate
- Free and open-source software
- Fulfill data management plans
- Research and classroom applications
- Driven by community needs
Outline

I. Brief Introduction to CUAHSI

II. HydroShare Recap

III. CUAHSI JupyterHub Architecture

IV. Getting Started with CUAHSI JupyterHub

V. Reproducible Workflows

VI. Next Steps
HydroShare Overview

How it works

1. Create data
   Collect your data using the same methods you use now. HydroShare supports a broad set of hydrologic data types.

2. Upload to HydroShare
   Upload your data files to HydroShare through the web user interface. HydroShare will automatically extract as much metadata as it can from the files you upload.

3. Describe with metadata
   Use HydroShare's simple metadata entry forms to finish describing your data so that your colleagues can find, access, and interpret it.

4. Share with colleagues
   You choose who has access to the data and models you have uploaded to HydroShare. You can share with individual users or publish your resources for everyone to access.
Outline

I. Brief Introduction to CUAHSI

II. HydroShare Recap

III. CUAHSI JupyterHub Architecture

IV. Getting Started with CUAHSI JupyterHub

V. Reproducible Workflows

VI. Next Steps
What is JupyterHub?

Web based version of Jupyter Notebooks
- [http://jupyter.org/](http://jupyter.org/)
- create and share documents that contain live code, equations, visualizations and explanatory text

Cloud-hosted
- [https://github.com/jupyterhub/jupyterhub](https://github.com/jupyterhub/jupyterhub)
- Multi-user Hub which spawns, manages, and proxies multiple instances of the single-user Jupyter notebook (IPython notebook) server
- Docker to manage separate user execution environments
- Pre-configured environment for the execution of scientific toolchains.
Design Criteria

- Manage Annoyingly Large Data – "Big Data"
  - Large datasets that cannot be moved easily
  - Workflows that download and process lots of data

- Support Computationally Intensive Toolchains
  - Long-term or large-scale simulations
  - Model Calibrations

- Educate and Disseminate
  - Illustrate scientific processes and algorithms to students and/or colleagues
  - Communicate findings with non-scientists

- Enable Reproducible Science
  - Reproduce publish findings
  - Share model simulations and/or workflows with
Software Architecture
Outline

I. Brief Introduction to CUAHSI
II. HydroShare Recap
III. CUAHSI JupyterHub Architecture
IV. Getting Started with CUAHSI JupyterHub
V. Reproducible Workflows
VI. Next Steps
Getting Started

Welcome to the HydroShare Python Notebook Server

You've arrived at the HydroShare Python Notebook Server. This environment provides web-based programming and model execution functionality for HydroShare users. It has been configured with many scientific libraries, tools, and models to enable easy manipulation of HydroShare data using the Python programming language. You can also save notebooks back into HydroShare or launch existing notebooks from HydroShare into this environment. This provides a unique solution to sharing preliminary and published research with collaborators and colleagues.

Jupyter notebooks enable scientists to explore, modify, and interact with data inside a remote execution environment by using the Python programming language. A Jupyter notebook is an enhanced computational environment that combines rich text and code execution into a single script-like container. The HydroShare Python Notebook Server aims to combine this standard notebook functionality with the HydroShare data hosting platform to provide a rich computational environment for water scientists.

The following example notebooks illustrate the functionality of this web application.

- Raster Processing using TauDEM
- LandLab Landslide Calculations
- Basic Time Series Analysis - GenericResource
- Basic Time Series Analysis - TimeSeriesResource
- Working with Large Files using iRODS
- WaterML R Package
Outline

I. Brief Introduction to CUAHSI

II. HydroShare Recap

III. CUAHSI JupyterHub Architecture

IV. Getting Started with CUAHSI JupyterHub

V. Reproducible Workflows

VI. Next Steps
Reproducible Workflows
Outline

I. Brief Introduction to CUAHSI

II. HydroShare Recap

III. CUAHSI JupyterHub Architecture

IV. Getting Started with CUAHSI JupyterHub

V. Reproducible Workflows

VI. Next Steps
Next Steps

- Create a HydroShare Account
  - http://www.hydroshare.org
- Explore Examples
  - http://www.cuahsi.org/jupyter
- Upload Existing Code
- Execute Remotely
- Collaborate
- Cite
Thank you

Tony Castronova
Hydrologic Scientist
acastronova@cuahsi.org
339-933-4127

http://www.hydroshare.org
http://www.cuahsi.org/jupyterhub
https://www.cuahsi.org/education/cyberseminars/archives/
https://www.hydroshare.org/resource/1846fd3dce53418e8f33e95b964dcdd8/