CUAHSI Services to Enable Data Fusion

Tony Castronova acastronova@cuahsi.org
Hydrologic Scientist
CUAHSI

04-23-2018
CUAHSI in a Nutshell

- 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, such as workshops, community meetings, training, etc.
  - Data and Model Services software to support science, including HIS and HydroShare
Community Services
Early Career Services
Resources and training to build capacity and extend capabilities

○ Learn a new instrumentation method
○ Engage your local community on water issues
○ Develop or improve data tools
○ Add a new field site to existing research

○ The Summer Innovators program - contribute to the National Water Center goals
○ Collaborate with researchers at other institutions

Tony Castronova acastronova@cuahsi.org  Hydrologic Scientist  CUAHSI
Community Meetings

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 drones and NASA SWOT mission
- Sessions on water and climate, hydrologic feedbacks, energy-water systems, etc.
Summer Innovators Program
Training the next-generation of water scientists

- Partnership between CUAHSI, NWC, and the NWS
- Involvement of academic community, NCAR, DHS, US ACE, USGS
- Engages the academic research community in the enhancement of the National Water Model
- Fourth summer institute emphasis on groundwater and channel processes
Summer Innovators Program
Outcomes and broader engagement

- More than 100 students trained on NWM creating next generation of researchers and users, more than half of their advisors engaged in the process
- Advances in the NWM functionality and improved understanding of performance
- Growing engagement of academic community
- JAWRA Featured collections and CUAHSI – NOAA Technical Reports
- Transformational educational experience
Data Services
Data Services
Resources to access data and models for research and education

- Tools to enable science
- Share, discover, analyze, collaborate
- Free and open-source software
- Fulfill data management plans
- Research and classroom activities
- Driven by community needs
Hydrologic Information System
A general overview

- **HydroSever**: stores, organizes and publishes data
- **HIS Central**: database of metadata harvested from network
- **Clients**: interfaces for accessing data
HydroClient  
A CUAHSI web client for the HIS - http://data.cuahsi.org

- Discover, view, download
- Federal, university, community data
- Enhancements to search and discovery interface and backend cataloging via Solr
- Integration with HydroShare Apps
Advancing Hydrologic Understanding

- requires integration of information from multiple sources
- may be data and computationally intensive
- requires collaboration and working as a team/community
HydroShare
Dealing with data disparity

- Time series
- Geographic raster
- Geographic feature
- Multidimensional space/time
- Model programs
- Model instances
- Combinations of the above
HydroShare
An architectural overview

- 50 TB rolling window of NWM forecasts (2-3 TB/day)
- National HAND layer
- 50 TB store for other HydroShare data
HydroShare
An overview of key functionality

- Upload data to HydroShare
- Enhance scientific value with metadata
- Manage access and control
- Collaborate
- Publish
National Water Model Viewer
Access to daily forecasts
Welcome to the HydroShare Python Notebook Server

You have arrived at the HydroShare Python Notebook Server landing page. The cells below demonstrate HydroShare content inside this Python scripting environment. You will be asked to provide your HydroShare notebook server can establish a secure connection with your resources. When accessing resources, you will be provided hyperlinks to launch them in some changes from this session back into HydroShare using the hs-utils library.

Jupyter notebooks enable scientists to explore, modify, and interact with data using the Python programming language. A Jupyter notebook is an enhanced cell editor and code execution into a single script-like container. The HydroShare Python standard notebook functionality with the HydroShare data hosting platform to provide water scientists.

See what scientists are doing with notebooks:

- Raster Processing using TauDEM
- LandLab Landslide Calculations
- Basic Time Series Analysis - GenericResource
- Basic Time Series Analysis - TimeSeriesResource
- RHESSys Modeling

# define a function to download nwm results

```
def get_nwm(q, iolock, out_q, outdir, cnt):
    resid = q.get()
    if resid is None:
        break
    iolock = resid
    with iolock:
        cnt += 1
    if cnt % 10 == 0:
        print("channel rt forecast in parallel using iRGeD")
        in_q = mp.Queue(maxsize=NCORE)
        out_q = mp.Queue()
        iolock = mp.Lock()
        Pool = mp.Pool(NCORE, initializer=get_nwm,
                       initializer=(in_q, iolock, out_q, nwm_dir, cnt))
        for i in range:
            Pool.map(nwm_channel_rt_forecast, nwm_dir)
```

```
<table>
<thead>
<tr>
<th>feature_id</th>
<th>reference_time</th>
<th>streamflow</th>
<th>nudge</th>
<th>q_lateral</th>
<th>velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>2017-07-18 03:00:00</td>
<td>0.90</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>2017-07-18 06:00:00</td>
<td>0.95</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>2017-07-18 09:00:00</td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>count</th>
<th>max</th>
<th>mean</th>
<th>std</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.152410e+08</td>
<td>2.173518e+08</td>
<td>8.026310e+08</td>
<td>6.757011e-02</td>
</tr>
</tbody>
</table>
```

```
NWM Results by COMID

<table>
<thead>
<tr>
<th>date (y-m-d)</th>
<th>streamflow</th>
<th>nudge</th>
<th>q_lateral</th>
<th>velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017-07-18</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2017-07-18</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2017-07-18</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2017-07-18</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
```
Reproducible Science
Scientific Collaboration
Education
Computationally Intensive Research
Data Intensive Research

Import Libraries and Prepare Environment

```python
In [1]:
import os, sys

# unzip the python libraries
!unzip -o pythonlibs.zip

# add our lib directory to the path
sys.path.append('./pythonlibs')
sys.path.append('./pythonlibs/lib/python3.6/site-packages')
sys.path.append('./pythonlibs/bin')
```

Collect Input Data from HydroShare

- Connect to HydroShare
- Download simulation input data

```python
In [2]:
import celeryworker as c
from pyemma.Simulation import Simulation
from hydroshare import hydroshare

ha = hydroshare.hydroshare()

Please enter your HydroShare username: TonyCastronova
Enter the HydroShare password for user 'TonyCastronova': ********
Successfully established a connection with HydroShare

In [4]:
ha.getResourceFromHydroShare('6866d8f26ac4bco8ff4c1e55374c')
Download finished
Successfully downloaded resource 6866d8f26ac4bco8ff4c1e55374c

Found the following file(s) associated with this HydroShare resource.

- celia-simulation.zip
- These files are stored in a dictionary called ha.content for your convenience. To access a file, simply issue the following command where MY_FILE is one of the files listed above:
  ha.content['MY_FILE']

- Look at the content of the HydroShare resource that was downloaded

In [5]:
print(ha.content)
data_archive = ha.content['celia-simulation.zip']
```
Meeting the Needs of the Community

- CUAHSI is model agnostic
- Provide community tools such as HIS and HydroShare, along with documentation and API’s
- Provide community data sets for selected use cases through CUAHSI Water Data Services. (Harvey/Irma and Maria RAPID awards)
- Continue with Summer Institute and potentially similar activities
- Organize and host community meetings
Thank You

Tony Castronova
acastronova@cuahsi