

Coordinates to Move Your Science Forward

June 2021

Inside This Issue

- Message from the Principal Investigator
- Meet the Family
- Project Spotlight
- Cyberinfrastructure Focus
- Get Involved

Users with little or no programming experience will be able to create and build data pipelines that collect and process data into usable results

Message from the Principal Investigator

It's been an exciting year for the MyGeoHub team. We'd like to share a little more about what we've been up to. We hope that together we can find ways to make a difference in your day-to-day research workflows.

Carol Song, Purdue University

Meet the GABBs Family

MyGeoHub.org is a public science gateway supporting broad geospatial modeling, data analysis and visualization needs of research and education communities through hosting and management of groups, datasets, tools, training materials, and educational contents. Built on HUBzero, it features geospatial data capabilities developed under the GABBs project efforts.



GABBs is an NSF-funded project to create a powerful Web-based system that will allow researchers worldwide to manage, curate, share, analyze and visualize geospatial data for purposes ranging from predicting damaging floods to projecting climate change effects on the global population.

Project Spotlight – GABBs 2.0

In its second iteration, GABBs is now adding novel capabilities for building seamless connections among platforms, data and tools, making large scientific and social geospatial datasets directly usable in scientific models and tools. Users with little or no programming experience will be able to create and build data pipelines that collect and process data at multiple scales and convert such data into usable results.

Four case studies will demonstrate these novel capabilities:

- 1) flood hazard prediction – workflows will enable highly accurate multiscale urban and rural flood modeling and prediction
- 2) plant phenotyping – workflows will support real-time analysis of crop health from low-cost handheld sensor data
- 3) water quality monitoring – workflows will support citizen science data management for freshwater quality assessment in EPSCoR states
- 4) sustainable development – workflows will support novel interdisciplinary analysis of climate impacts on global food security

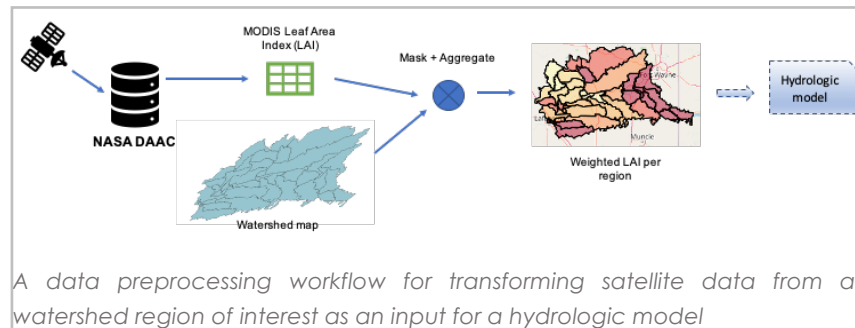
GeoEDF enables researchers to design plug-and-play workflows in a simple, high-level syntax composed out of community-contributed data wrangling building blocks

Cyberinfrastructure in Focus

A core component of GABBs 2.0 is GeoEDF, a Python-based extensible data framework designed to simplify data wrangling in geospatial research workflows. GeoEDF enables researchers to design plug-and-play workflows in a simple, high-level syntax composed out of community-contributed data wrangling building blocks.

We are excited to announce the release of version 1.0 of the **GeoEDF Workflow Engine** that takes a high-level workflow, automatically executes the corresponding data wrangling processes on the Halsted cluster at Purdue University. The workflow engine is available for public use in the Jupyter notebook environment on the MyGeoHub science gateway.

This process is illustrated below for an example data preprocessing workflow from hydrologic modeling:



```

$1:
Input:
  NASAInput:
    url: https://e4ftl01.cr.usgs.gov/MOTA/MCD15A3H.006/{filename}
    user: rkalyana
    password:
  Filter:
    filename:
      PathFilter:
        pattern: '%{dtstring}/MCD15A3H.*.h09v07*.hdf'
      dtstring:
        DateTimeFilter:
          pattern: '%Y.%m.%d'
          start: 07/16/2002
$2:
HDFEOSShapefileMask:
  hdfFile: $1
  shapefile: /home/earthcube/geoedf/files/watershed/subs1_projected_171936.shp
  datasets: [Lai]

```

The same workflow expressed in the high-level GeoEDF workflow syntax

```

from geoedfengine.GeoEDFWorkflow import GeoEDFWorkflow
workflow = GeoEDFWorkflow('/home/geoedfuser/workflows/mcd15.yml')
workflow.execute()

```

The steps for executing the above GeoEDF workflow using the Python GeoEDF Workflow engine library

In the **coming weeks and months**, we will be releasing example workflows from the various science use cases of the GeoEDF project. For more information on the workflow syntax and the development and contribution of the data wrangling building blocks, refer to the GeoEDF Documentation at <https://geoedf.readthedocs.io/en/latest/>

Get Involved

Interested in learning more, please contact: info@gabbs-project.org

Follow Us on Twitter: 

[@MyGeoHub_](https://twitter.com/MyGeoHub)

[@gabbs_project](https://twitter.com/gabbs_project)

GABBs Project:

mygeohub.org/groups/gabbs/

Look for us in upcoming Events:

EarthCube Demo, June 15-17, 2021

PEARC 2021

Gateways 2021