MyGeoHub Geospatial Gateway

R. Kalyanam*, L. Zhao, R. Campbell, D. Kearney, I. L. Kim, J. Shin, L. Biehl, W. Wan and C. X. Song, Purdue University
*Corresponding author address: Research Computing, Purdue University, West Lafayette, IN, 47907, USA; email: rkalyana@purdue.edu

Abstract: *MyGeoHub is a science gateway for researchers working with geospatial data. Based on the HUBzero cyberinfrastructure framework, it provides general-purpose software modules enabling geospatial data management, processing and visualization. Termed "GABBs" (Geospatial Data Analysis Building Blocks), these modules can be leveraged to build geospatial data driven tools with minimal programming and construct dynamic workflows chaining both local and remote tools and data sources. We will present examples of such end-to-end workflows demonstrating the underlying software building blocks that have also found use beyond the MyGeoHub gateway in other science domains.*

1. Introduction

The MyGeoHub science gateway [1] grew out of efforts to design a sustainability model to continue hosting data and tools from research projects with a cyberinfrastructure focus even after the initial project funding runs out. The model is based on a shared hosting paradigm where multiple research projects are hosted on the same science gateway sharing both resources and the cost of maintenance. Each project can have its own content group with distinct branding, while leveraging shared data management and processing tools including access to HPC resources. When the hosted projects all belong to the same science domain (as is the case with MyGeoHub), pooled funds can be used towards the development of general-purpose capabilities that are useful to all hosted projects.

Common feature requests and observations of typical usage patterns among researchers from the individual hosted projects on MyGeoHub led to a funding proposal and subsequent development of geospatial software building blocks under the NSF DIBBs program. These building blocks have subsequently enabled the rapid development of several geospatial data-driven tools for the hosted projects on MyGeoHub that previously required significant time and effort.

MyGeoHub utilizes the HUBzero cyberinfrastructure framework [2] that comes with user and community management, a wiki, ticketing system and DOI support for data publication out-of-the-box. HUBzero has been used in over 60 science gateways for domains ranging from earthquake engineering and cancer care engineering to HPC and the study of human-animal bonds. HUBzero provides a tool development environment and the ability to deploy containerized tools that can be launched and used right from the user's web browser. This enables researchers to bring their legacy scientific code online and share it with the broader community on a HUBzero gateway. In addition, researchers can collaborate and share data via various HUBzero collaboration and publication mechanisms.

The rest of this proposal is organized as follows: we briefly describe the GABBs building blocks that were developed and now underlie various MyGeoHub tools and typical usage patterns, and conclude with a description of two workflows that utilize several of these building blocks and which we believe will best demonstrate the MyGeoHub gateway at the Gateways 2017 conference.

2. Geospatial Data Management, Processing and Visualization Building Blocks

The GABBs building blocks enhance several HUBzero components while providing value addition to HUBzero gateways centered around

Presented at Gateways 2017, University of Michigan, Ann Arbor, MI, October 23-25, 2017.
https://gateways2017.figshare.com/

geospatial data. These new building blocks include the following:

(a) The default HUBzero data management and collaboration framework is extended with iData, a data management system that supports event-based processing for metadata extraction and quick previews of geospatial files. In addition, a seamless connection between iData and the hub tool environment simplifies typical workflows.

(b) A general-purpose image processing and analysis tool for hyper and multispectral image data in various file formats, MultiSpec [3] is now available as a hub tool.

(c) The GeoBuilder tool enables users to construct and share complex data views combining both geospatial and spreadsheet data with no programming.

(d) A Python mapping library, pyMapLib [4] provides basic map widgets and APIs for constructing map-based tools in a variety of programming languages including Python, Java and C++.

(e) The HUBzero tool development environment is extended to include map elements that can be incorporated into any hub tool.

(f) A REST API and OAuth authentication plugin enables interoperation with other science gateways so that data from one gateway can be used with tools in another gateway without having to maintain separate user accounts on both.

3. Demonstration Details

We propose to start our demonstration with an overview of basic features and the four research projects hosted on MyGeoHub. Next, we will present a typical workflow involving data management, tool invocation, simple geospatial processing and saving results on the hub for future use. This will illustrate both the seamless connection between data and tools as well as the automatic metadata extraction and quick preview capabilities. We will also demonstrate how users can construct complex data views combining spreadsheet and geospatial data using the GeoBuilder tool and save the configuration, enabling a reproducible experience for other users. Our second workflow demonstrates a more complex usage pattern involving multiple gateways. Hydrological modeling data from HydroShare (www.hydroshare.org, a gateway for managing hydrological data) is used to execute simulations on MyGeoHub with result data saved to iData. This model output can be subsequently accessed using the REST API in a web-based tool that performs further computation for in depth exploration and visualization. This workflow illustrates both the use of the OAuth plugin simplifying access across the two gateways and the use of the REST API for greater interoperability between data and other components on MyGeoHub.

Acknowledgments

This work was supported in part by by the National Science Foundation grant number 1261727.

References

[1] Zhao, L., Song, C.X., Biehl, Larry. MyGeoHub Science Gateway for Spatial Data and a Model for Sustainability. Gateways 2016, San Diego, CA, November 2-3, 2016

[2] M. McLennan, R. Kennell, HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering, Computing in Science and Engineering 12 (2) (2010) 48–52.

[3] L. Biehl, D. Landgrebe, MultiSpec - A Tool for Multispectral-Hyperspectral Image Data Analysis, Computers and Geosciences 28 (10) (2002) 1153–159.

[4] W. Wan, L. Zhao, J. Shin, S. Zhe, C. X. Song, PyMapLib for Rapid Development of Geospatial Data Analysis Tools, in: Proceedings of the Third International Conference on CyberGIS and Geospatial Data Science, CyberGIS16, 2016.