Extending Gateways into the Field: Geospatial Extensions and Remote Data Services

Robert Campbell, Lan Zhao, Rajesh Kalyanam, Carol X. Song, Larry L. Biehl
Rosen Center for Advanced Computing, Purdue University, West Lafayette, IN, 47907, USA;
email: {rcampbel, lanzhao, rkalyana, cxsong, biehl}@purdue.edu.edu

Abstract: *Science gateways are generally accessed via web browsers from workstation or laptop computers. However, activities such as fieldwork and data collection have prompted an increasing call for new methods of access using a wide range of mobile and specialized devices. Our work to extend the HUBzero gateway framework with new access routes is described. Specifically, an emphasis is placed on enabling the flow of geospatial data and metadata between geo-enhanced HUBzero hubs and devices in the field.* [1]

1. Introduction

Most science gateways are web based applications accessed via web browsers providing integrated data management and/or analysis functions to a user community. However, this mode of access is not efficient for scientists who collect data in the field. With the widening adoption of smart devices, there is an increasing call for new ways of gateway access using mobile and specialized devices. Furthermore, there is a growing need for field workers to not only upload their data, but also annotate (typically with geolocation information) and preview the data, as well as to perform simple data processing, using their hand-held devices. In this paper, we present our work of extending the HUBzero gateway framework [1] with geospatial data management and analysis functions as well as a REST API to expand access to such devices. This enables third party applications, especially mobile apps, to dynamically connect to the gateway, access the data management system, and launch dynamic workflows.

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

2. About HUBzero

HUBzero provides an open source, configurable, almost turn-key solution for creating a science gateway (also known as a “hub”). Through the installation of open source software packages, hubs enable web-based community building, collaboration, and information dissemination. Two existing HUBzero features were particularly relevant to our work: tool hosting and project files.

A hub hosts applications, known as "tools" [2], that allow the author and other users to process, present, and explore data. Frequently, tools are developed to run simulations, either on standard servers or high performance computing resources. Originally, tools were implemented as desktop applications running on the server and presented to the user within VNC containers [3] via a web page interface. Recently, hubs are starting to add support for tools built in Jupyter Notebook and R Shiny. This allows tools written in the Python and R programming languages to present web-native user interfaces. Tools can be a valuable part of workflows involving fieldwork or data collection.

Hubs also provide a shared space for teams of users to manage data in the form of files. "Projects" [4] are access controlled areas within hubs that allow teams to collaborate. Inside each project, storage space is provided for members to create, upload, manage, and publish files.

3. Geospatial Enhancement to HUBzero

The original, core HUBzero framework does not have the built-in support for geospatial data management and analysis which is a key requirement for data collection from fieldwork. A recent NSF DIBBs project, "Geospatial Data Analysis Building Blocks (GABBs)", [5] has added geospatial data capabilities into the HUBzero framework to support collaboration around geospatial data analysis, visualization and workflows.

GABBs enhances hub projects' file storage with "iData", a system built on the iRODS data management system [6] that supports extended metadata for its managed files as well as highly granular, event-based file processing. In addition, this system is accessible via the FUSE mount mechanism, supporting standard, POSIX file operations. Hub tools do not need to be modified to use this system. iData leverages these iRODS capabilities to automatically extract and capture extended metadata for geospatial files and a user interface is provided for optional, manual edits.

Map-based searching and the ability to preview raster files and shapefiles were also implemented and integrated into iData. Finally, geospatial tools and tool libraries were created to provide useful examples and foster geospatial tool development with minimal programming.

All the GABBs components have been deployed and are publicly accessible on MyGeoHub (mygeohub.org) [7] where a number of tools and projects use its geospatial data management and toolkits. GABBs-enabled hubs will be supported on Amazon AWS in Q3/Q4 of 2017.

4. Open Access

Beyond standard web access, HUBzero hubs natively provide the ability to upload and download files via HTTP and SFTP. The GABBs project extended these features to allow HTTP and SFTP access to iData. Irrespective of the file upload mechanism used, iData uniformly enhances hubs' project file areas with automated metadata extraction, geospatial preview, and the ability to federate with external iRODS repositories. Metadata is stored in a database as key/value pairs. Initially, terms from the Dublin Core are created and automated determination of some values is performed when possible. This includes geospatial fields. Also, users are able to edit values and add key/value pairs. All metadata can be exported by tools into desired formats.

With SFTP access, file-based, remote software has the basic ability to exchange files with the hub. In order to further facilitate mobile access and support metadata-enhanced data collection demanded by researchers working in the field, an applications programming interface (API) for iData was developed to enable client-server applications.

This REST-compliant API extends the accessibility of GABBs-enabled hubs to other websites, mobile apps, and specialized devices. The iData API exposes lower level file and metadata operations. Using the API, remote software can authenticate and then read and write files with extended metadata support. For example, software consuming the API can create a folder (directory), create a file within that folder, write data to the file, and then record extended metadata for that file.

The term "extended metadata" is meant to indicate information beyond that which is normally recorded for a file (timestamp, author, access permissions, etc.) For example, geospatial field/value pairs such as projection and coverage (latitude and longitude bounds) are automatically extracted and associated with map layer files. In addition, the API allows for the creation of user-defined metadata field/value pairs.

Files and metadata transmitted via the API are then instantly accessible from a team's project storage area and can be processed by hub tools.

5. Use Cases

In this section, we describe the use of the GABBs functions and iData service API to demonstrate its utility in the field.

5.1 GrABBs Mobile App

To encourage and facilitate broader adoption of the HUBzero geospatial functions, we developed a general-purpose mobile app named "GrABBs" to demonstrate the use of the iData API. This example application is available as a working reference to anyone wanting to connect to a GABBs-enabled hub using the service API. The source code is available for both the Android and iOS platforms [8]. GrABBs implements the following features:

(a) authentication to hub (MyGeoHub) and authorization of project access based on user's membership
(b) browse/create/delete files and folders within iData of projects for which user is a member
(c) image and video capture to file using device camera
(d) audio capture to file using device microphone
(e) file upload and download to hub project's iData file area
(f) automated and manual metadata annotation of files
(g) display the geo-locations of collected data on a map

As-is, the app is useful to projects that simply need to capture, annotate, and upload files, images, audio or video.

GrABBs is also meant to act as a reference for--or even the basis of--custom apps developed by other investigators. They may need to present specialized data collection user interfaces or communicate with specific hardware and software on consumer or custom mobile devices. A specialized app may also invoke specific logic to preprocess collected data.

5.2 *Plant Phenotyping App and Workflow*

The GABBs geospatial functions and data API have been used in a plant phenotyping research project to enable field data collection and ingestion and subsequent collaborative analysis on a hub [9]. Led by Dr. Jian Jin in the Agricultural and Bioengineering Department at Purdue University, the research team developed a mobile device that images plants in the field to determine their health status.

The app is used for crop assessment in an agricultural context as a fast, non-destructive method for obtaining the physiological features of plants. It incorporates the use of physical optics, plant science models, and the image processing ability of smartphones to provide accurate feature predictions such as water, chlorophyll, and nitrogen levels. Farmers and agricultural researchers use the app to access real-time crop plant health data and to avoid the expensive, time consuming chemical analysis associated with conventional plant sampling.

Using the GrABBs source code as reference, the researchers rapidly developed new functions in their Android app to upload images and data to a hub project on MyGeoHub via the data service API. The solution, including the device and hub project data repository, is intended as a more affordable and compact alternative to current technologies. As soon as collected data is transmitted to the hub, an analysis tool processes and displays it as a set of maps and graphs. This web-based interface presents the user with collection location maps and corresponding value heat maps that indicate plant health metrics.

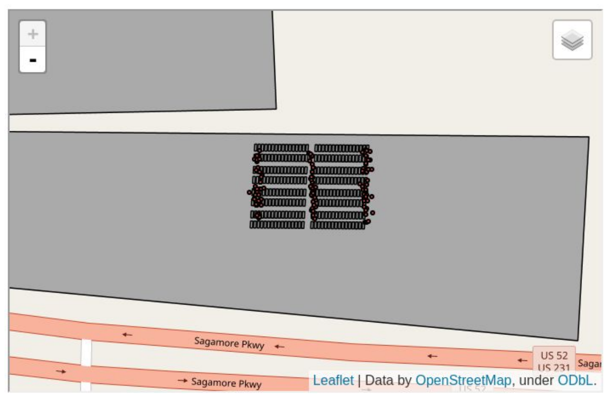

Figure 5.1 - Portion of tool screenshot showing map that indicates locations of sampling (filled circles) in relation to crop rows (hollow rectangles).

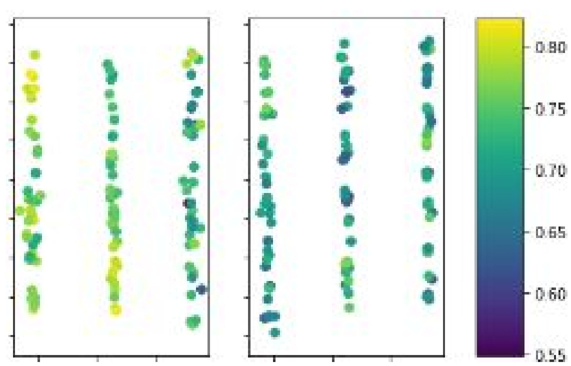

Figure 5.2 - Portion of tool screenshot showing heatmap that compares NDVI levels collected from

same crop area at different times.

Across science gateways, the collection, processing, and presentation of data does not fit a finite set of workflow patterns. Every group of investigators that collaborate may employ different processes and procedures. To fulfil this requirement, GABBs-enabled hubs intentionally support loosely coupled and flexible workflow options. This capability was applied to the plant phenotyping project and below is a collaborative workflow example that illustrates steps related to fieldwork use of the GABBs functions and API in this project. It does not attempt to encompass all available steps and possible scenarios. Note that different types of users will perform different roles within a team.

(a) Research group members register as new users on the hub and sign in using hub-specific ID and password or using a affiliated authentication ID.
(b) The project manager creates a "Plant Phenotyping" hub group. This will be used to define the team of collaborators that have access to project data.
(c) The project manager creates a "Plant Phenotyping" hub project. This provides areas for managing files, to-do lists, notes, publications, and databases. He invites research group members to join the project.
(d) A researcher goes to "iData Storage" under "Plant Phenotyping" project. He creates a new folder called "Phenotyping Box" within the project files area to contain collected data.
(e) Researchers go into the field, collect data, and upload the data from their devices. The Android app extracts geo-location information and automatically attaches additional, relevant metadata fields to the data.
(f) Back in the lab, researchers can log into MyGeoHub and use the map preview functions of iData to quickly verify supplementary files. They can further launch a Jupyter Notebook tool on the hub to process collected data and interactively display the plant health measurements as heatmaps or line plots over a regional shapefile within the browser. The tool also allows users to save a raster file summarizing map data.
(g) Researchers assemble and publish selected project files into a hub publication [10] complete with unique Digital Object Identifier (DOI) and metadata information which is available to the public.

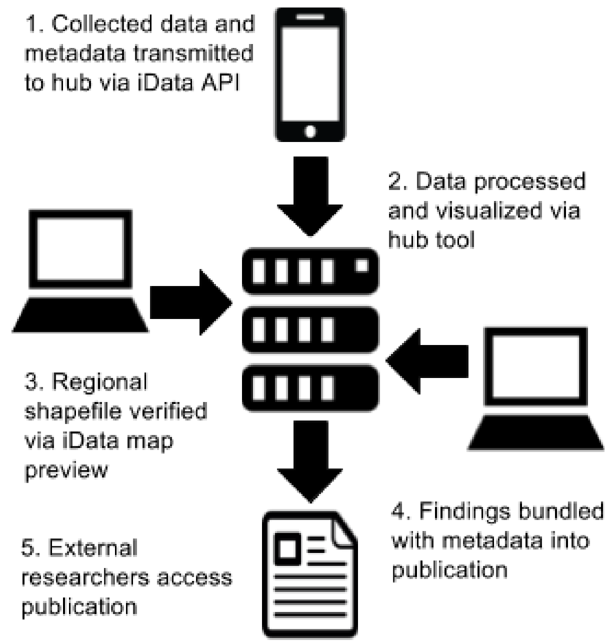

Figure 5.3 - Geospatial functions and APIs used in plant phenotyping workflow

6. Conclusion

This short paper summarizes our work to extend the HUBzero science gateway platform for fieldwork. Brief overviews of the HUBzero framework and its geospatial extension through the GABBs project were presented. Various methods for remote, programmatic access to GABBs-enabled hubs were listed and emphasis was placed on the use of the GABBs geospatial functions and data service API in a mobile app interface to a hub. Finally, a common research workflow in the plant phenotyping domain was illustrated, outlining the use of our geo-enabled HUBzero gateway through the entire research workflow from fieldwork to data publication.

7. Acknowledgments

This work has been supported in part by NSF award #1261727. The authors would also like to thank our colleague, Bekmurat Spayev, our collaborators on the HUBZero team, and Dr. Jian Jin and his team for their contribution to the project.

8. References

[1] M. McLennan, R. Kennell, "HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering," Computing in Science and Engineering, 12(2), pp. 48-52, March/April, 2010

[2] Tools - HUBzero feature documentation: https://hubzero.org/documentation/2.0.0/users/tools

[3] Richardson, T.; Stafford-Fraser, Q.; Wood, K. R.; Hopper, A. (1998), "Virtual Network Computing" IEEE Internet Computing, 2: 33. doi:10.1109/4236.656066

[4] Projects - HUBzero feature documentation: https://hubzero.org/documentation/2.0.0/users/projects

[5] Lan Zhao, X. Carol Song, Rajesh Kalyanam, Larry Biehl, Robert Campbell, Leif Delgass, Derrick Kearney, Wei Wan, Jaewoo Shin, I Luk Kim and Carolyn Ellis. GABBs - Reusable Geospatial Data Analysis Building Blocks for Science Gateways. 9th International Workshop for Science Gateways, Poznan, Poland, June 19-21, 2017.

[6] iRODS: https://irods.org/

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

[8] GrABBS: https://github.rcac.purdue.edu/ScientificSolutions/GABBs

[9] "Plant Phenotyping on Mobile Devices": http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1561&context=surf

[10] Publications - HUBzero feature documentation: https://hubzero.org/documentation/2.0.0/users/publications