

STEM

Developing Open Source Geospatial Data Analysis Building Blocks Software

Student researcher: **Jingbo Wu, Junior**

Data, data, data. The hottest buzzword right now is “big data.” Due to the exponential growth of geoscience information and data, more and more researchers are required to work together. This data boom and increased need for collaboration led to the development of a new online platform, myGeoHub, and then to the iData tool, which aims to provide a myGeoHub interface that allows geoscientists to upload, describe, and share their data.

My intent with this project was to develop a C-based API client library for iData that would allow users to access information from the online database and to download and upload their data using their own applications and tools. I chose to use C because it is one of the most popular programming languages used by scientists to create data analysis and modeling tools.

My work began with choosing suitable third-party libraries that would be sufficient for the client. The ones I considered were the *cURL* library, for transferring data with URL syntax, and *Jansson*, an open source C library for encoding, decoding, and manipulating JSON response data. Both of the libraries were well developed and would provide powerful functions for the client. During the coding there were several challenges that I tried my best to overcome. For example, the requirement for one of those functions was that once a user defined necessary input parameters, the client had to get all information recursively—that is, the function kept traveling to all the subfolders until each of them was visited at least once. This would easily lead to memory leakage. Also, for PHP there was a file size restriction of 128 MB. I finally solved this problem by chunking the big file and uploading each smaller file separately. I also developed a test tool for the client, based on a virtual machine.

Research advisor Rajesh Kalyanam writes: “The HUBzero platform has been instrumental in fostering a broad community-based research environment where researchers around the world can contribute scientific data and analysis tools. Simplifying data management and interoperability is vital in making the most of this platform. Jingbo’s work on client APIs is an important step toward simplifying how data and tools interact. He demonstrated admirable initiative and perseverance in determining the best libraries for the task and in solving some of the thornier issues with large file transfers.”

```

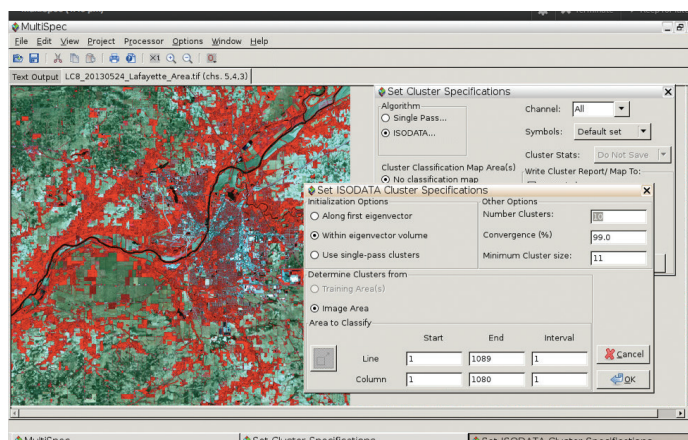
hostname was not found in DNS cache
* Trying 128.46.21.63...
* Connected to mygeohub.org (128.46.21.63) port 80 (#0)
> POST /api/iData/collection/upload?id=97&path=&ovner_id=2726&filename=testbook2.pdf&filesize=1368388 HTTP/1.1
Host: mygeohub.org
Accept: */*
Content-Length: 1368722
Content-Type: multipart/form-data; boundary=-----41802912586c0323

HTTP/1.1 201 Success
Date: Fri, 24 Jul 2015 13:53:42 GMT
Server: Apache/2.4.18 (Ubuntu)
Server: Apache
Set-Cookie: 8cb975bb43c7b4c65a983103b91c5f42=948954155d6280fa97475a7d4f4cb466; path=/; HttpOnly
Set-Cookie: 8cb975bb43c7b4c65a983103b91c5f42=47b2df6cedcc5147c1f08f4f4523a88; path=/; HttpOnly
X-RunTime: 0.43967795372
Expires: Mon, 1 Jan 2001 00:00:00 GMT
Last-Modified: Fri, 24 Jul 2015 13:53:45 GMT
Cache-Control: no-store, no-cache, must-revalidate
Cache-Control: post-check=0, pre-check=0
Pragma: no-cache
Connection: close
X-XSS-Protection: 1; mode=block
X-Content-Type-Options: nosniff
X-Permitted-Cross-Domain-Policies: master-only
X-Frame-Options: sameorigin
Content-Length: 33
Content-Type: application/json

Closing connection 0
message: file upload successful message: file upload successful

```

Screenshot showing the upload of an iData file using the client library Jingbo developed.



Example of a geospatial data analysis tool, MultiSpec, which will use the C iData library created from this project, extending its capability for sharing data on the Internet.