Geospatial-enabled Data Exploration and Computation through Data Infrastructure Building Blocks

Carol Song
NSF DIBBS PI
Senior Scientist, Rosen Center for Advanced Computing
Purdue University

Best Practices in Data Infrastructure Workshop
Pittsburgh, PA
May 17-18, 2016
GABBS: Geospatial Data Building Blocks

• Needs driven by broader participation

Computing Community Consortium (CCC)
Spatial Computing 2020 Workshop (2012)

Table 2.1: Challenges to spatial computing

<table>
<thead>
<tr>
<th>Late 20th Century</th>
<th>The New Reality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maps were produced by a few highly trained people in government agencies and surveying companies</td>
<td>Everyone is a mapmaker and many phenomena are observable.</td>
</tr>
<tr>
<td>Only sophisticated groups (e.g., Department of Defense, oil exploration groups) used GIS technologies</td>
<td>Everyone uses location-based services</td>
</tr>
<tr>
<td>Only specialized software (e.g., ArcGIS, Oracle SQL) could edit or analyze geographic information</td>
<td>Every platform is location aware</td>
</tr>
<tr>
<td>User expectations were modest (e.g., assist in producing and distributing paper maps and their electronic counterparts)</td>
<td>Rising expectations due to vast potential and risks</td>
</tr>
</tbody>
</table>
Research Data Needs

Technology Maturity

MultiSpec ©

iRODS Consortium

globus
GABBS: Geospatial Data Building Blocks

Lower the **barrier**
Make it easy to **visualize** geospatial data
Make it easy to **share** geospatial/georeferenced data
Open **source**, **community** driven

---

**Broaden** participation
**Faster** dissemination
**Enhance** learning
User Community

• Essentially anyone who needs open source way of sharing, publishing geospatial data and creating tools that deal with geospatial data

• The project is driven by use cases from:
  – Hydrology modeling
  – Climate impact research
  – Weather research
  – Disaster related data needs
  – Applied economics (global trade analysis)
  – Social sciences (communications, nature and human interaction, etc.)
Weather observational data

- No programming
- Almost instant visualization & data sharing
Spatial data and computation

Locks up my laptop!
Lost track of multiple datasets, results
How do I display my data on a map?
How do I set it up for my collaborators to run?
Now my paper has been published. Can I link the figure in the paper to this software?

Ag economists studying cropland supply and land transformation elasticity
Tool using GABBS-extended Rappture Toolkit, developed by Jing Liu & Nelson Villoria, Agricultural Economics
Users need more than maps

Systems to support various computation paradigms

Plots, bar graphs

Spreadsheets

Spatial-temporal data

Image galleries
GABBS at a high level

Geospatial data and computing building blocks – NSF DIBBS project

- Geospatial data processing, analysis and visualization support inside HUBzero
- Map library, Rapid Tool Development API (Rappture) with geospatial extension for developing online applications without web programming
- Online data management system linked to user tools
- DIY online interactive tool and data publishing (with DOI), publications linked to viewers and interactive tools

Platform for Scientific Collaboration

Modeling & Analysis

Explore and visualize data

Share & Publish
HUBzero Platform

HUBzero

Joomla! Framework

Linux / Apache / MySQL / PHP

Web Server

Hub Installation

Execution Host

Execution Host

Secure Container

Cloud/Cluster Computing Resources

User

Rappture Toolkit

http://hubzero.org
GABBS rendering architecture

MapBuilder

MyGeoHub.org

Map library

Rappture Toolkit

File Transfer

Datasets

User’s Home Directory

Rendering Server

PostGIS

WMS/TMS/WFS

Google / Bing / MapQuest

Submit Server

Web Browser

MyGeoHub.org

MapBuilder

Mapping Component

Datasets

User’s Home Directory

Rendering Server

PostGIS

WMS/TMS/WFS

Google / Bing / MapQuest

Submit Server

Web Browser

MyGeoHub.org

MapBuilder

Mapping Component

Datasets

User’s Home Directory

Rendering Server

PostGIS

WMS/TMS/WFS

Google / Bing / MapQuest

Submit Server

Web Browser

MyGeoHub.org

MapBuilder

Mapping Component

Datasets

User’s Home Directory

Rendering Server

PostGIS

WMS/TMS/WFS

Google / Bing / MapQuest

Submit Server
GABBS Deliverables – Visualization Toolkits

• Geospatial data toolkits
  – Client side library
  – GeoVis server for high performance needs
  – APIs and widgets for tool developers
GABBS Deliverables – Tool Builders

- GeoBuilder – creating geospatial data viewer/expplorer without coding
- Rappture Builder – Drag’n’drop style tool builder
GABBS Deliverables – Spatial data support

- Online data management (iRODS based)
  - Automatic metadata extraction and other microservices on data
  - Data quick view (including shapefile, raster)
  - “Open with” tools
  - Geospatial search
  - Publish data when ready

Project space for collaboration
Season-wise irrigated and rainfed crop areas for India around year 2005

By Gang Zhao\(^1\), Stefan Siebert\(^1\)

University of Bonn, Germany

Crop growing area and irrigated fraction for 21 crops in Kharif, Rabi and Zaid seasons for India around year 2005 in 500 m spatial resolution.

Listed in Datasets | publication by group Geoshare
GABBs Development

Developer
VMs
(development, unit test)

Changes
Additions

Stage server
(integration, test, quality control)

Production
server
(mygeohub.org)

Feedback
Requests
Use cases, interoperability

- Hydrologic modeling and model coupling
- AgMIP data aggregator and access
- CMIP5 data aggregator
- MultiSpec image processing remote sensing and spatial images
- Weather data explorer
- Land supply and transformation elasticity computation
- Educational modules, tutorials, class use
- Interoperability
  - Hydroshare
  - iRODS
  - Globus
  - TerraPop
Working with broader communities

Tools for sharing data and models
Tools for publishing data & models (DOI)
Computing resources
Training & learning platform
Collaboration platform
Tools for navigate and access data

MyGeoHub.org

Hosted service
AWS Instance
Open Source release

10/2013 – 9/2017
Partial releases 2015-2016
Full release: late 2016
2nd release: 2017
Team

Team of 10 faculty, scientists and professional developers
6+ graduate students
6+ undergraduate students (summer)
Various user groups and community (production site has > 4000 users)

This work is supported in part by the NSF grant #1261727
Thank you!

Questions, suggestions, and feature requests are welcome!

carolxsong@purdue.edu
NSF project info: http://mygeohub.org/groups/gabbs

Join us on MyGeoHub.org!