



G · A · B · B · S
geospatial data analysis building blocks

Project Overview

Carol Song, PI, NSF DIBBs award

Rosen Center for Advanced Computing

Purdue University

DIBBS PI Meeting, August 11, 2014

Overarching goal

Making it easy for scientists to share geospatial data and tools

Reach broad user community

- Anyone can create an online app and share
- Anyone can share geospatial data

Building for self service (DIY) – Leverage successful software – Develop building blocks



Project goals

- Data and Tool are one - tight integration of data capabilities with HUBzero tools
- Support geospatial data processing, analysis and visualization
 - Data services interface
 - Rapid tool creation APIs
 - Map and image renderers for online tools
 - Enabling geospatial data driven workflows
- All of these integrated with HUBzero core
 - Open source release
 - Hosting



Funding

- A National Science Foundation grant
- Data Infrastructure Building Blocks (DIBBs) program
- GABBs: 1 of 4 implementation awards in 2013
- \$4.5M, four years
- Collaboration with other DIBBs and DataNet awards



Team (11+)

Carol Song, PI & Project Director

Larry Biehl (remote sensing, GIS)

Venkatesh Merwade (hydrology, Civil Eng)

Nelson Villoria (global geospatial data, Ag Econ)

Ed Lee (project manager)

Michael McLennan (HUBzero architect)

Rob Campbell (sr developer, tool development)

Leif Delgass (sr developer, visualization)

George Howlett (sr developer, RAPPTURE Toolkit)

Lan Zhao (research scientist, geospatial applications, data management)

Rajesh Kalyanam (GIS data processing, management)



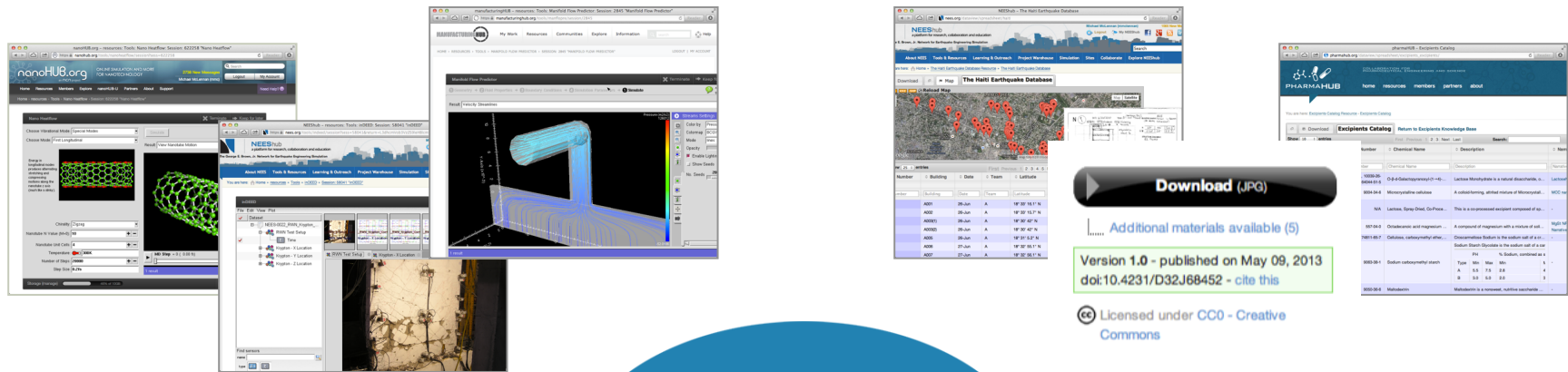
Geospatial hub projects

- Efforts in developing integrated geospatial data/modeling capabilities using HUBzero
 - Drinet hub (<http://drinet.hubzero.org>)
 - Geoshare hub (<http://geoshareproject.org>)
 - Water hub (<http://water-hub.org>)
 - Useful to Useable (u2u)
<http://drinet.hubzero.org/groups/u2u>

- Many hubs can make use of GABBs, such as NEES (network of earthquake engineering simulation), GENI (k-12 education), PURR (Purdue Research Repository), etc.

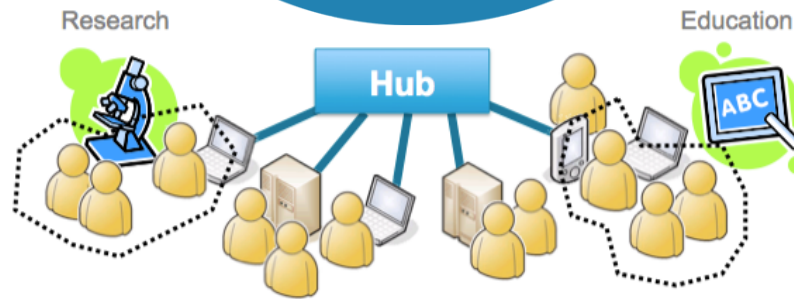


Platform for Scientific Collaboration



Computational Tools

Databases / Publications



Group/Project Collaboration

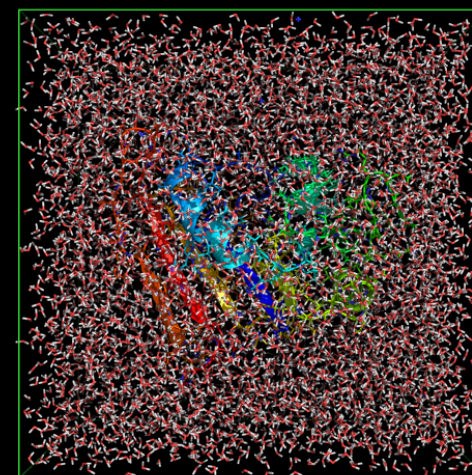
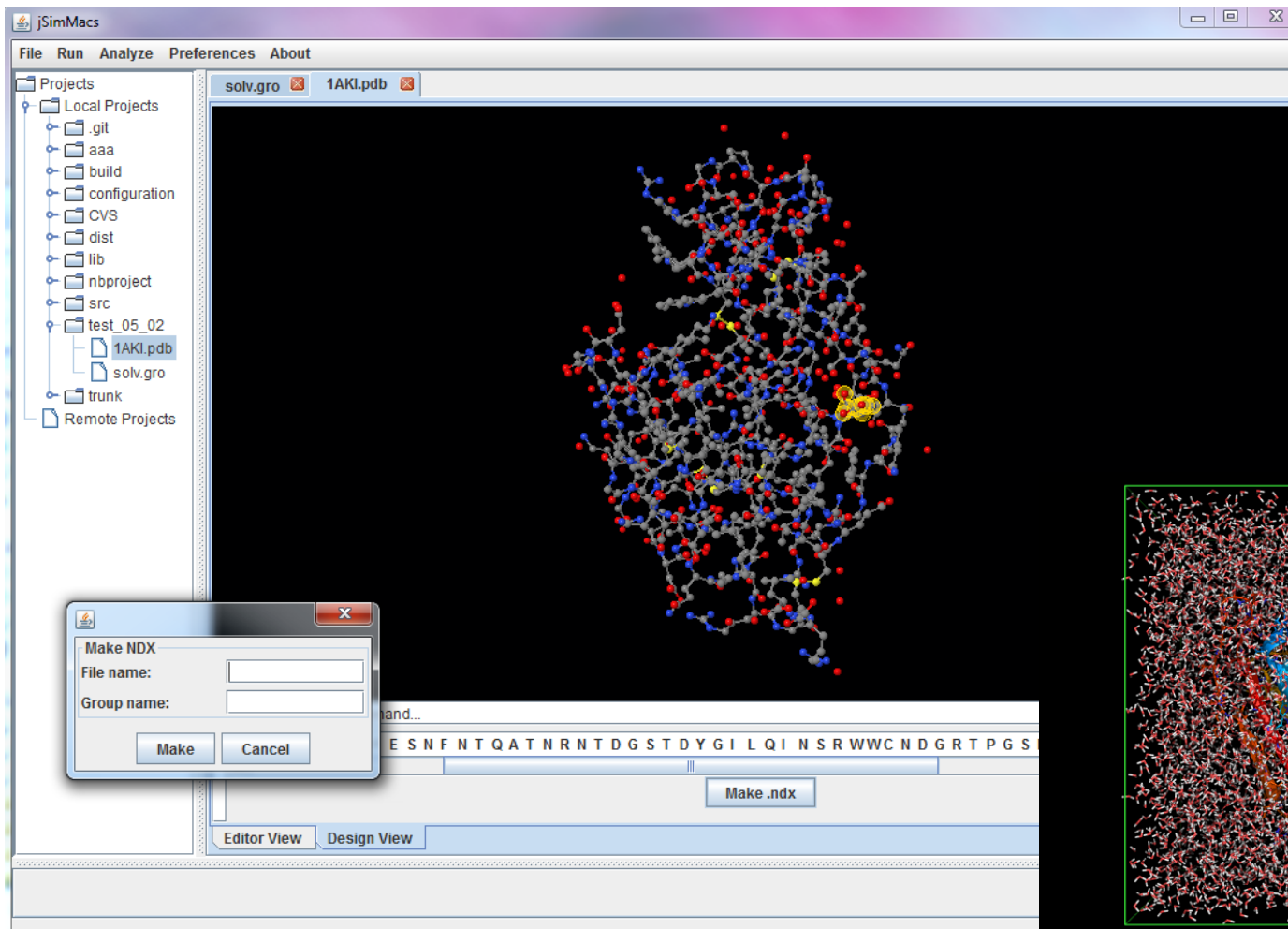


HUBzero for Scientific Collaboration

- An open source collaborative cyberinfrastructure for research and education
- Provides out-of-the-box support for developing web portals with content contribution, tool development workflows, user groups, wikis, ticketing systems, etc.
- Transparent access to large scale computation resources from online tools (no download, installation)
- Adopted by more than 50 science gateways cross many disciplines
- Serves more than 1,000,000 unique visitors in past 12 months
- Currently no integrated geospatial capabilities



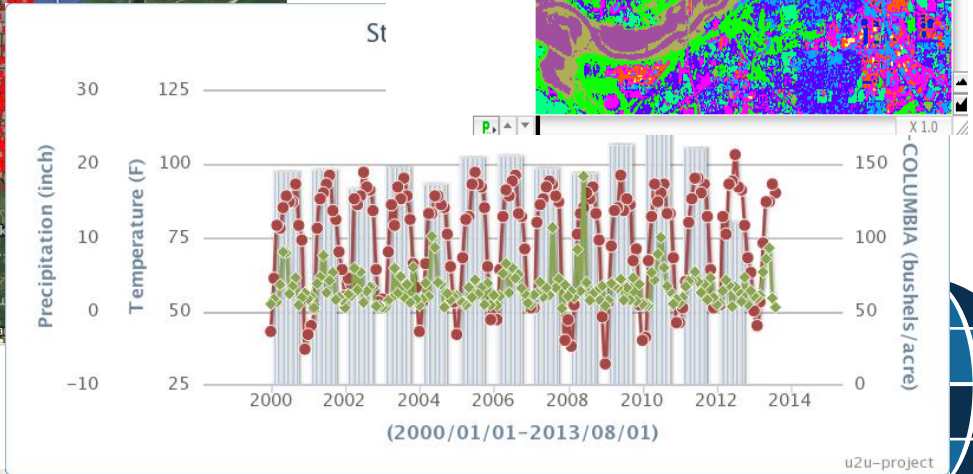
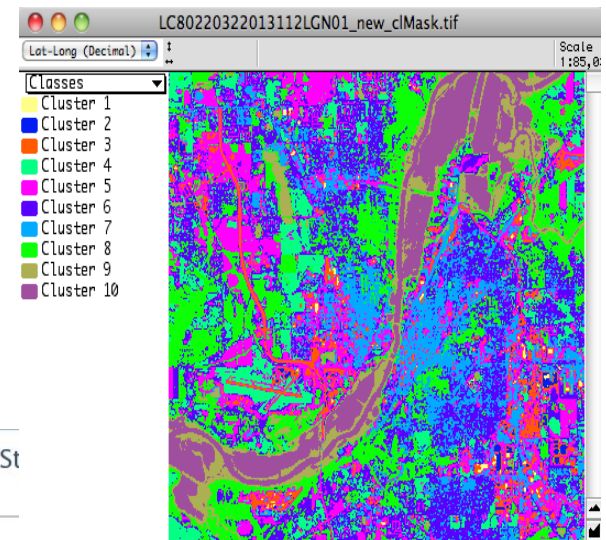
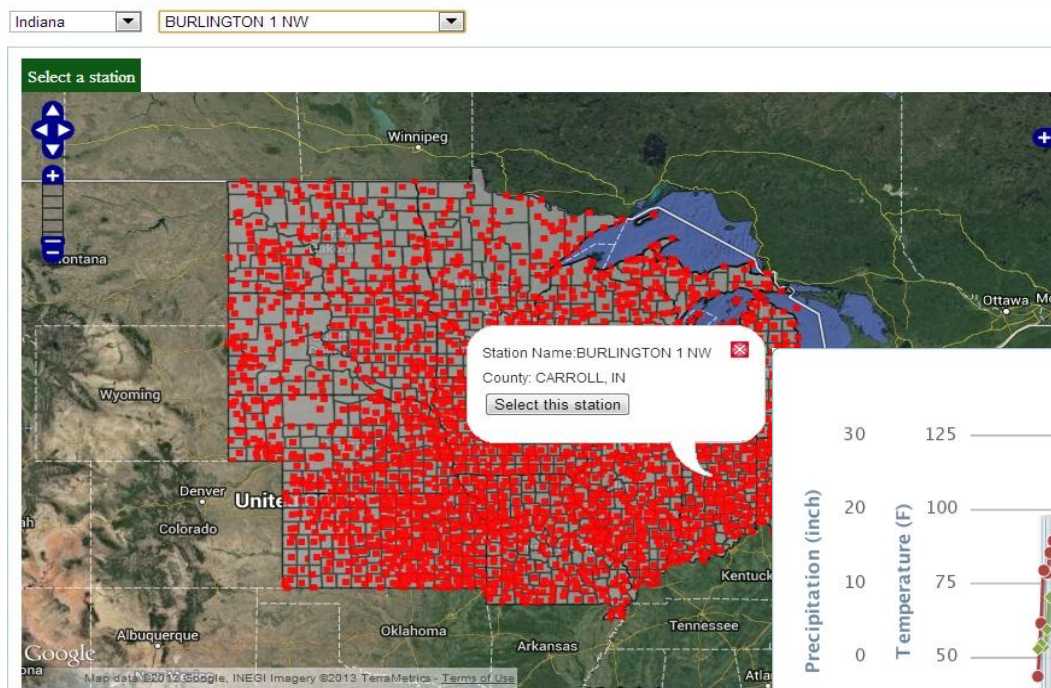
Online tools



Driving examples

- Easy deployment of geospatial tools

Please select a station from the drop down list or directly on the map.



Driving example

The screenshot displays the iDATA web application interface. At the top, the logo reads "iDATA Manage ge Publish, Browse & Discover". A navigation bar contains icons for Home, Add, Remove, Metadata, Refresh, Search, Setting, Download, and Log. The main interface is divided into several sections:

- My Collections / Shared Collections:** A list of collections with checkboxes and columns for Name and Description. Collections include MyColl1, MyTabColl1, MyFileColl1, and a subregion named "stJoseph".
- Current Path:** "stJosephRiv2".
- File View / Map View / Data View:** A tree view showing the current path structure: stJoseph > Riv2 > boundary, county83, outlets1, tempg1.
- Map:** A map showing a river network with red dots at outlets and yellow dots at subregions. A "Layers" panel on the right lists: tempg1, outlets1, boundary, and Riv2.
- Map Controls:** A vertical toolbar with navigation icons (compass, zoom in, zoom out, home, full screen).
- Metadata Table:** A small table at the bottom right showing details for subregion 17: name: Democratic Repu, tps: CG.

Buttons for "Save map" and "Add a layer" are located at the bottom right of the map area. The map shows a geographical area with a river network, and a small inset map of the United States is visible in the bottom left corner.

Driving example

- Multi-scale and multi-disciplinary data and modeling for addressing hydrologic and ag economic issues

The screenshot displays the Pegasus Tool [1.0] web interface. The main map shows a drought level overlay on a map of Iowa, with a legend indicating 'D0 Abnormally Dry'. A table of model names and their shared status is visible, with 'Upper_Iowa_R' selected. The interface includes a menu, a calendar, a log and statistics panel, and a list of output files. A 'Drought Level' legend shows a yellow square for 'D0 Abnormally Dry'. The map shows various cities in Iowa, including Saint Cloud, Saint Paul, Rochester, La Crosse, Richland Center, Madison, Milwaukee, Janesville, Rockford, Egan, Cedar Rapids, Des Moines, Iowa City, Davenport, and Joliet. The interface also features a 'My Models' section on the right with a list of shared models, including 'NorthRaccoon_Sensitivity', 'North_Raccoon_River_J', 'North_Raccoon_River_C', 'Sangamon_Calibrated', 'Sangamon_River_at_Fi', 'Sangamon_Sensitivity', 'SouthBranch_calbratio', 'SouthForkSangamon_C', 'SouthFork_sensitivity', 'SOUTH_FORK_SANGAM', 'SpoonRiver_Sensitivity', 'SPOON_RIVER_AT_LON', 'SPOON_RIVER_AT_LON', 'SugarRiver_Sensitivity', 'SUGAR_RIVER_calibrati', 'SUGAR_RIVER_NEAR_BI', and 'testflatriver'. The interface also includes a 'Download Model' and 'Download Output' button, and a 'Download File' and 'Download Archive' button. A note at the bottom states 'all tiles can also be downloaded.'

Model name	isShared
UpperIowa_ca	Yes
UpperIowa_Se	Yes
Upper_Iowa_R	Yes
Upper_Iowa_R	No

Property	Value
User ID	tomotoso
Model Name	Upper_Iowa_River_nea
Model Type	normal
Version	SWAT2009
HUC ID	07060002
Country	United States
State	IA

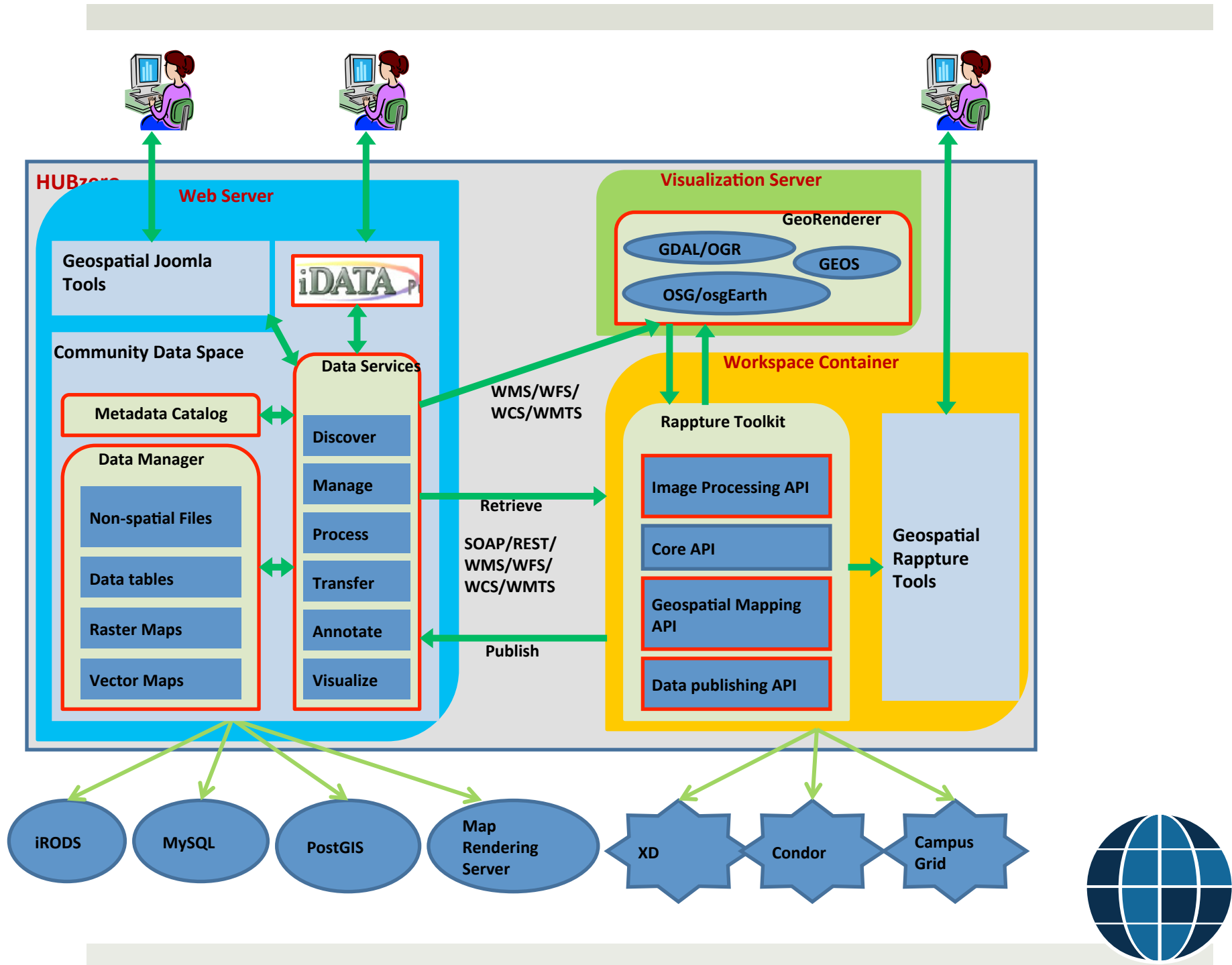
Attr.	Value
perimeter	145169.66613
area	1.30814367908914E
ncapc	54

My Models
Shared Models
NorthRaccoon_Sensitivity
North_Raccoon_River_J
North_Raccoon_River_C
Sangamon_Calibrated
Sangamon_River_at_Fi
Sangamon_Sensitivity
SouthBranch_calbratio
SouthForkSangamon_C
SouthFork_sensitivity
SOUTH_FORK_SANGAM
SpoonRiver_Sensitivity
SPOON_RIVER_AT_LON
SPOON_RIVER_AT_LON
SugarRiver_Sensitivity
SUGAR_RIVER_calibrati
SUGAR_RIVER_NEAR_BI
testflatriver
Other Models

Building on prior work

- ▣ **HUBzero** (Rappture, graphics rendering, collaborative web tools)
- ▣ **iData** (tool for self service data publishing and management)
- ▣ **Multispec** (tool for analyzing multispectral/hyperspectral image data)
- ▣ **Geospatial hub projects** (DRINET, Geoshare, WaterHUB, U2U etc)
- ▣ Leveraging software developed elsewhere
 - ▣ iRODS
 - ▣ Globus data transfer





Outcome

- The rapid tool development library RAPPTURE will support
 - geo-referenced data objects (maps, images, etc)
 - Easy way to share geospatial data, both in raw data, and visually and interactively
 - Easy way to share interactive tools that uses, and produces geospatial data
- Tool builder that supports geospatial data to further lower the barrier of creating interactive online tools
- Service interfaces to upload and share geospatial and other types of data in HUBzero
- Service interfaces to link tools and data
- Geospatial capabilities as part of core HUBzero open source release



Project Management

- Science driven
 - Scientists on the project team
 - Use cases in the proposal
 - Broad engagement of domain researchers
 - New use cases to help refine requirements
- Highly iterative development process
 - Use case broken into smaller building blocks
 - Prototyping, examples, and refactoring into more general codes
- Integrated with HUBzero software stack
- Engage potential collaborations early on (so we can design in from the beginning).



Challenges

- ❑ Dealing with large data sets
- ❑ Adapting the existing RAPPTURE model to support the new requirements of geospatial data and interactivity
- ❑ Map rendering in hub VM workspace
- ❑ Service interfaces
- ❑ Interfacing with other systems (Google drive, Dropbox, GIS servers)



Collaborations

- Current
 - iRODS
 - Globus
- Potential partners
 - DataOne
 - Other DIBBs



More to come

- New hub: <http://MyGeoHub.org>
- HUBbub 2014 – annual HUBzero conference

