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)

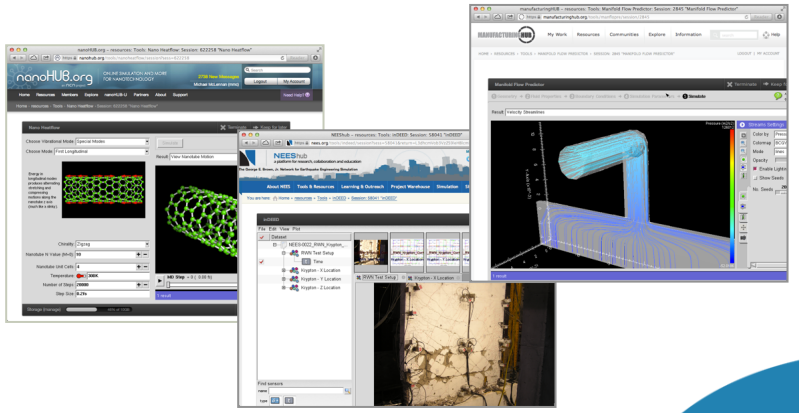
Lan Zhao (research scientist, geospatial applications, data management)

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

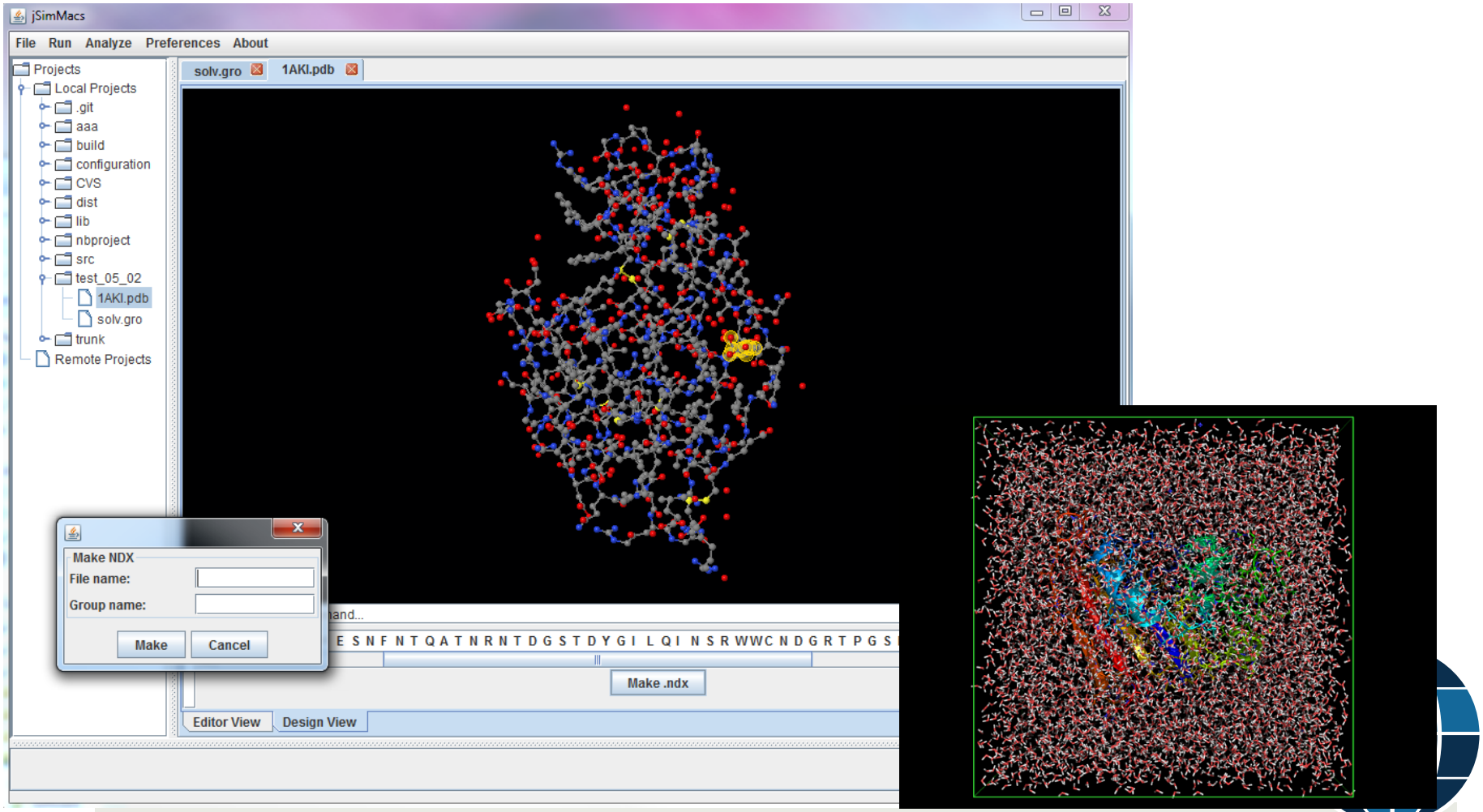
hubzero

Group/Project Collaboration

HUBzero for Scientific Collaboration

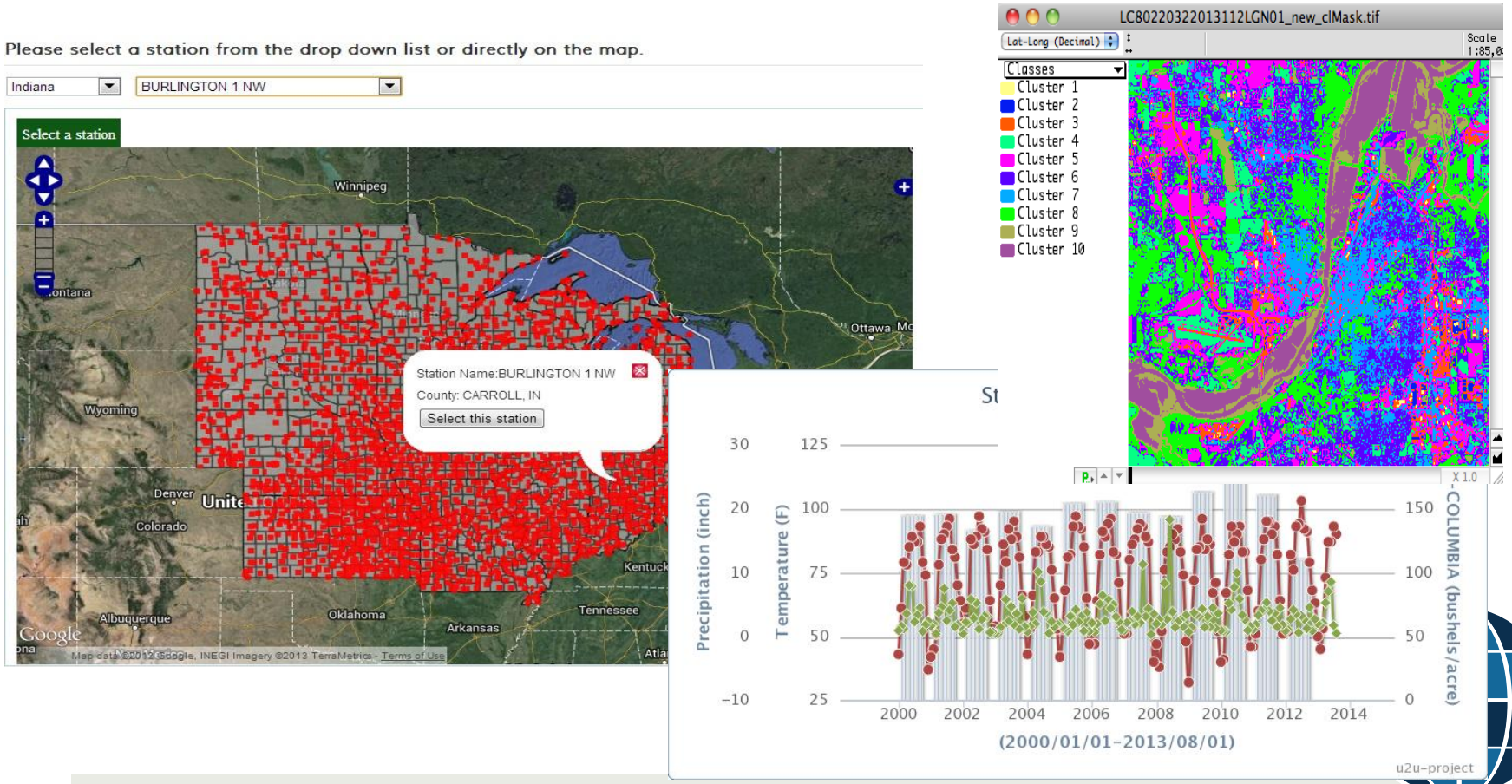
- An open source collaborative cyberinfrastructure for research and education
- Provides out-of-the-box support for developing web portals with content contribution, tool development workflows, user groups, wikis, ticketing systems, etc.
- Transparent access to large scale computation resources from 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

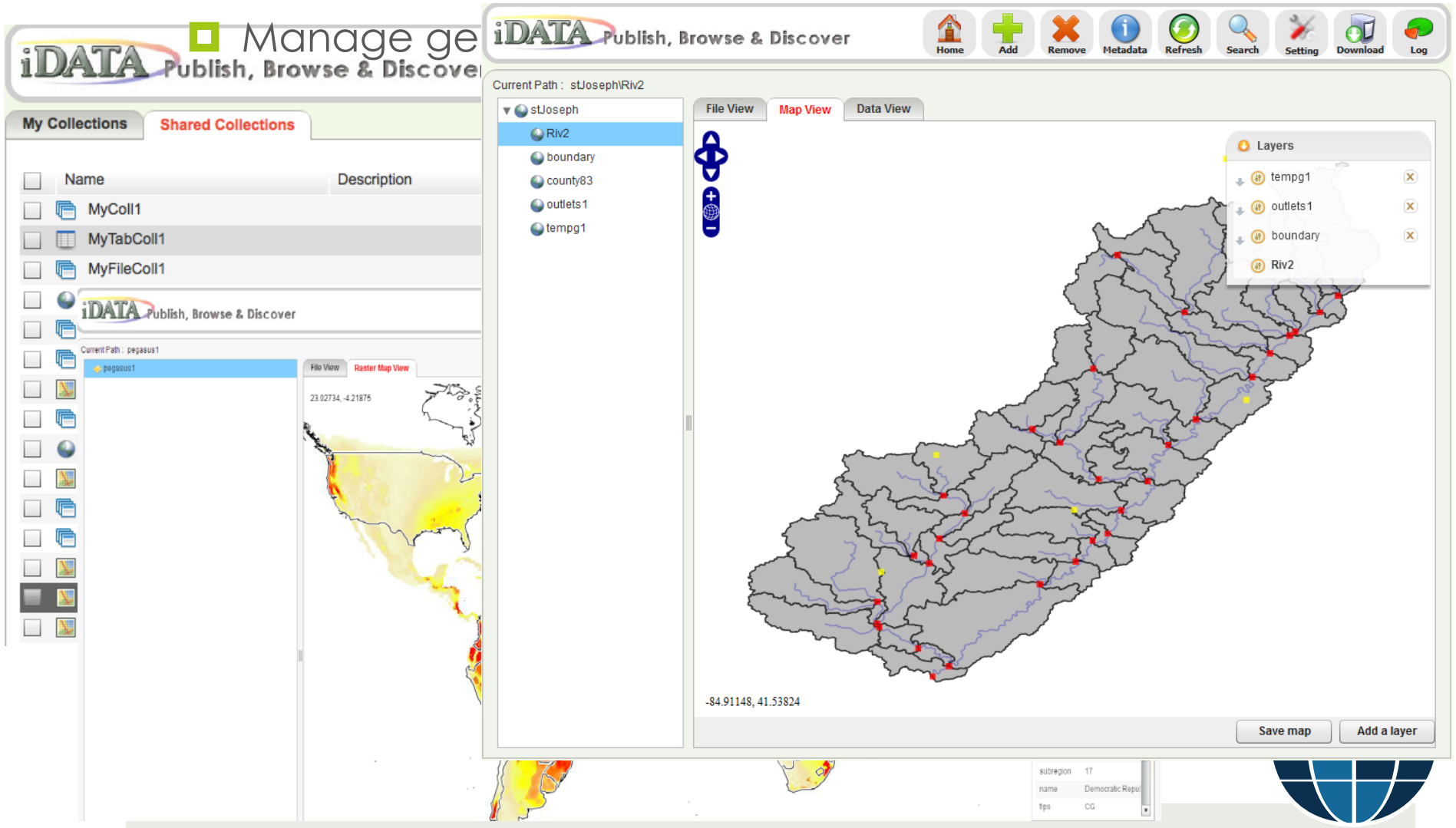


Driving examples

- Easy deployment of geospatial tools

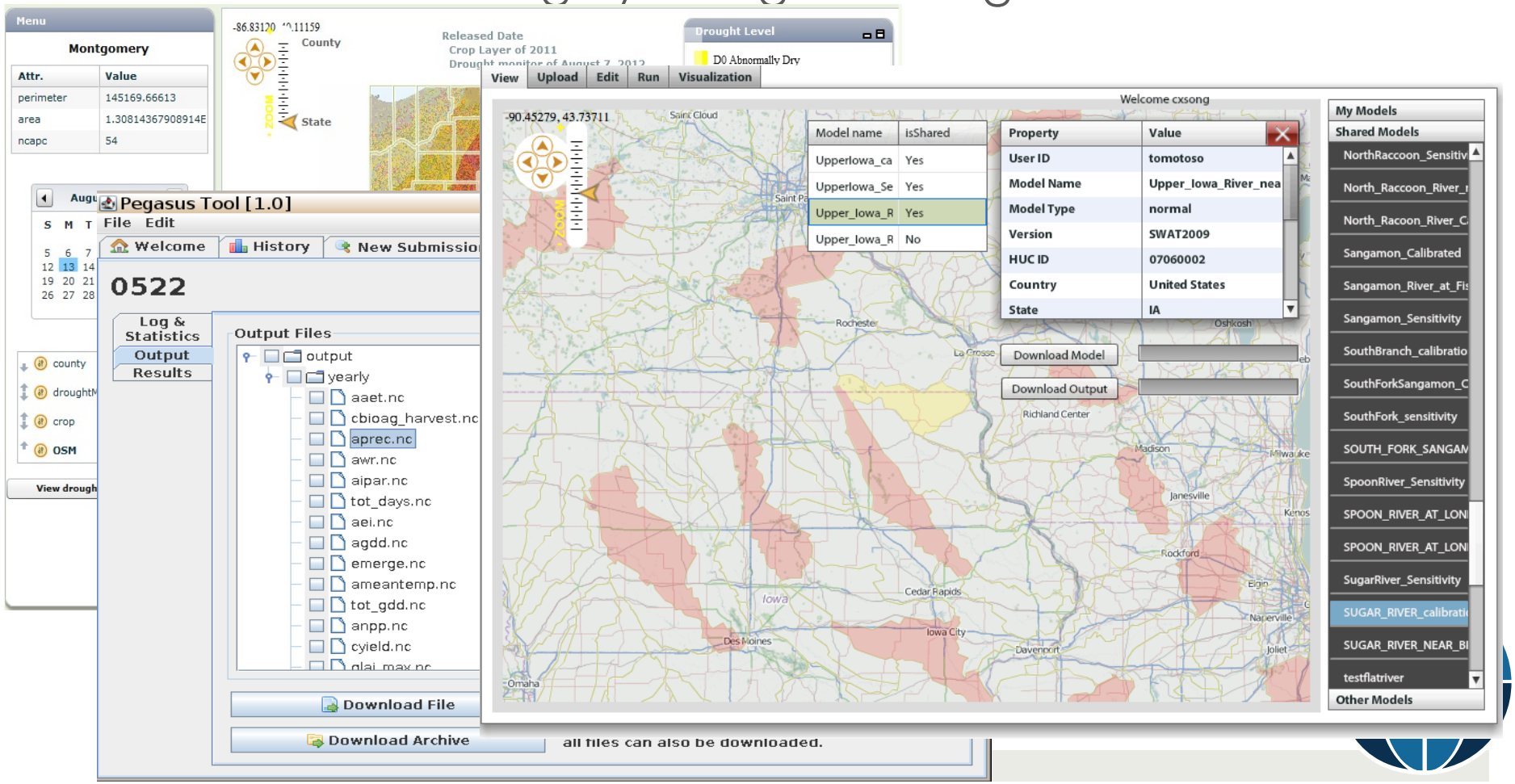


Driving example



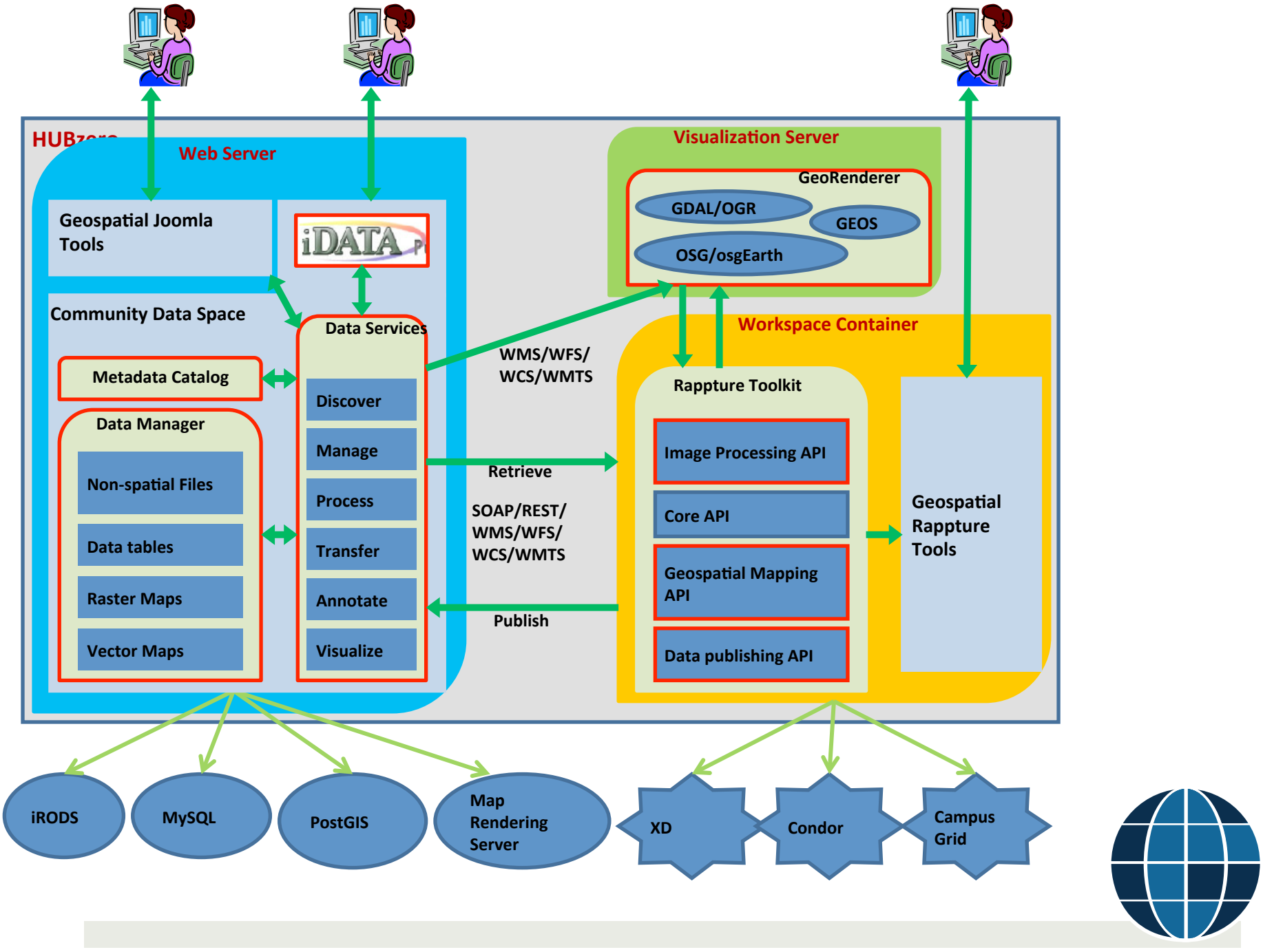
Driving example

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



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


HUBzero
Web Server
Geospatial Joomla Tools
iDATA
Community Data Space
Data Services
Metadata Catalog
Data Manager
Non-spatial Files
Data tables
Raster Maps
Vector Maps
Discover
Manage
Process
Transfer
Annotate
Visualize
WMS/WFS/ WCS/WMTS
Retrieve
SOAP/REST/ WMS/WFS/ WCS/WMTS
Publish
Visualization Server
GeoRenderer
GDAL/OGR
GEOS
OSG/osgEarth
Workspace Container
Rappture Toolkit
Image Processing API
Core API
Geospatial Mapping API
Data publishing API
Geospatial Rappture Tools
iRODS
MySQL
PostGIS
Map Rendering Server
XD
Condor
Campus Grid

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