



G · A · B · B · S

geospatial data analysis building blocks

Broadening Access to Geospatial Capabilities

Carol Song, Ph.D.

Senior Research Scientist

Rosen Center for Advanced Computing/ITaP

Purdue University

Purdue GIS Day
November 7, 2014

Great opening!

80/20

Open

Virtual environment

Collaborative environment

Humanities 2.0

Toolkit

Participatory

Volunteer (VGI)

Multidisciplinary

Broader community needs



Acknowledgement

- Venkatesh Merwade
- Nelson Villoria
- Lan Zhao
- Larry Biehl
- Mike McLennan
- George Howlett
- Rob Campbell
- Rajesh Kalyanam
- Many others



Outline

- ▣ Needs from user community
- ▣ Introduction of the GABBs project
- ▣ Technology
 - ▣ HUBzero
 - ▣ New geospatial additions
- ▣ Examples

Why do you care?



Driving Use Cases

- Easy deployment of geospatial tools

The screenshot displays a web-based geospatial application. At the top, a browser window title reads "LC80220322013112LGN01_new_clMask.tif". The application header includes the "iDATA" logo and the text "Publish, Browse & Discover". A toolbar on the right contains icons for Home, Add, Remove, Metadata, Refresh, Search, Setting, and Upload. The main interface is divided into several sections:

- Left Panel:** A "Select a station" sidebar with a map of the United States and a vertical navigation bar.
- Top Left:** A dropdown menu showing "Indiana" and a "BU" button.
- Center:** A world map with a heatmap overlay. A popup window is open over the Democratic Republic of the Congo, displaying the following metadata:

```
iso3: COD
lon: 23.654
iso2: CD
subregion: 17
name: Democratic Republic of the Congo
tps: CG
pop2005: 56740547
region: 2
lat: -2.876
un: 180
area: 220705
Value: 20.00000
```
- Right Panel:** A "Layers" panel showing a list of layers including "raster.worldmap" and "geoshare.pegasus1_a.mean_1004_b1". Below this is a table with "Property" and "Value" columns:

Property	Value
iso3	COD
lon	23.654
iso2	CD
subregion	17
name	Democratic Repu
tps	CG

At the bottom of the interface, the text "Created 11/7/2014" is visible. A globe icon is located in the bottom right corner.

Driving Use Cases (cont'd)

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

The screenshot displays a web-based hydrologic modeling interface. The main window shows simulation results for a model named "CedarShort2" over the period 2009-2009. The visualization type is set to "Spatial".

Step 2: Select variables

- PRECIP(mm): Precipitation
- SNOMELT(mm): Snow or ice melt
- PET(mm): Potential evapotranspiration
- ET(mm): Actual evapotranspiration
- SW(mm): Soil water content
- PERC(mm): Water that percolates past the root zone
- SURQ(mm): Surface runoff contribution to streamflow
- GW_Q(mm): Groundwater contribution to streamflow
- WYLD(mm): Water yield
- SYLD(t/ha): Sediment yield
- ORGN(kg/ha): Organic N yield
- ORGP(kg/ha): Organic P yield
- NSURQ(kg/ha): Nitrate transported by the surface
- SOLP(kg/ha): Soluble P yield
- SEDP(kg/ha): Mineral P yield

Step 3: Set data range

Date range: 1 / / 2009 ~ 12 / / 2009

Total Monthly Precipitation

Month	Precipitation (mm)
1/09	30
2/09	45
3/09	60
4/09	45
5/09	80
6/09	75
7/09	100
8/09	120
9/09	45
10/09	55
11/09	130
12/09	125

Surface Runoff (2009-11)

The runoff map shows a watershed area with 13 sub-catchments. The runoff values range from approximately 10.0 mm to 30.0 mm. The map is overlaid on a geographic coordinate system with latitude from 39.8 to 40.4 and longitude from -92.6 to -92.2.

Download Archive

its contents or to download the file. Archive or all files can also be downloaded.

The GABBs Project

Geospatial Modeling and Data Analysis Building Blocks

Making it easy to:

- ▣ create geospatial tools
- ▣ share these tools
- ▣ share geospatial data
- ▣ use other people's data



Key Outcomes

- 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



Funding

- A National Science Foundation grant
- Data Infrastructure Building Blocks (DIBBs) program
- GABBs: 1 of 4 implementation awards in 2013
- \$4.5M, 4 years (10/2013 – 9/2017)
- Collaboration with other DIBBs and DataNet awards



Team (14+)

Carol Song, PI & Project Director

Larry Biehl (CoPI, image processing and visualization)

Venkatesh Merwade (CoPI, hydrologic modeling)

Nelson Villoria (agricultural economics, sustainability)

Betsy Hillery (project manager)

Michael McLennan (HUBzero architect)

Rob Campbell (sr developer)

Leif Delgass (sr developer)

George Howlett (sr developer)

Lan Zhao (research scientist)

Rajesh Kalyanam (graduate student)

Hou-Jen Ko (graduate student)

Graduate students in scientific domains

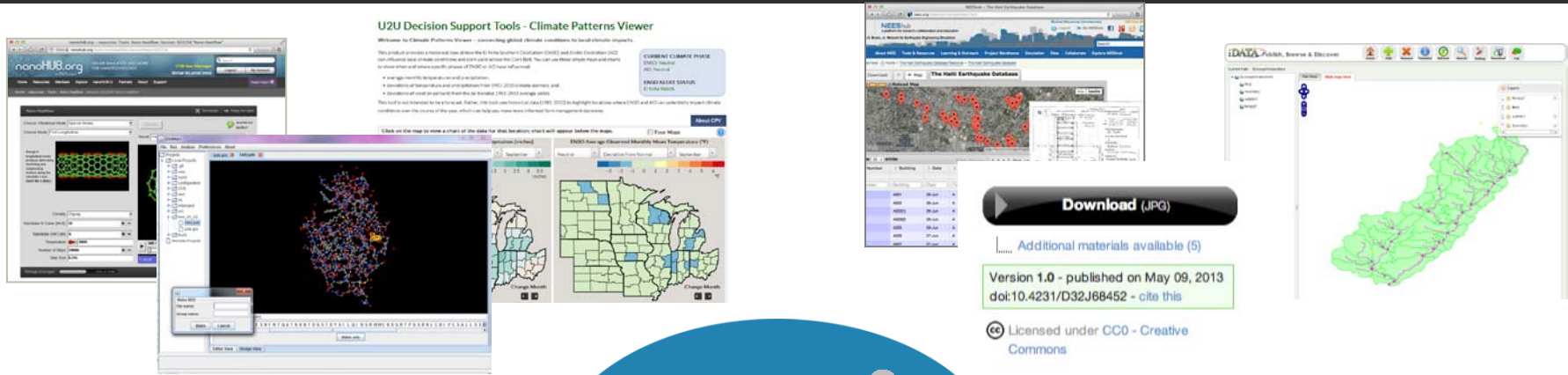


Building on prior work

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

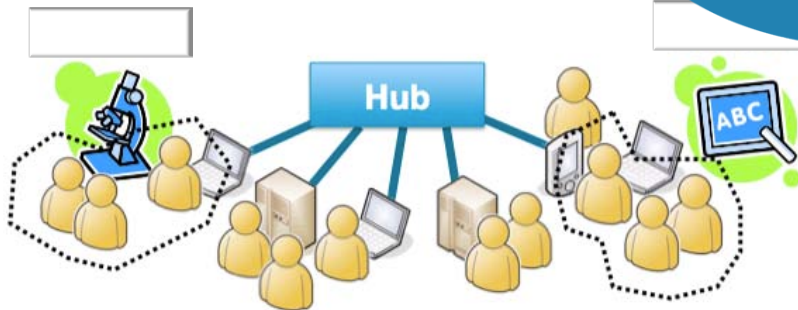


Platform for Scientific Collaboration

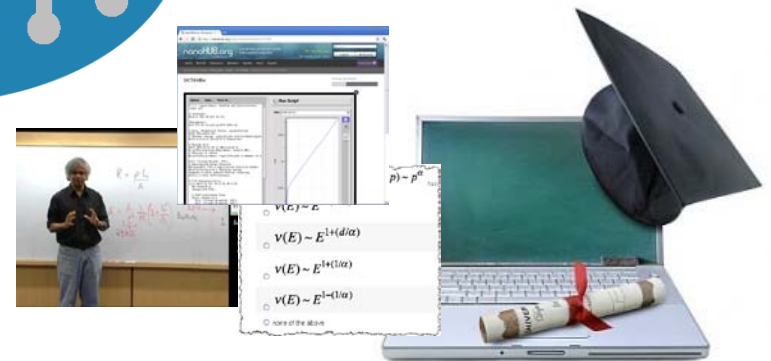


Computational Tools

Databases / Publications



Group/Project Collaboration



Learning Management



Who's Using HUBzero?

Supporting Purdue's largest research projects:



NEES: NSF \$105M - earthquake engr data (Ramirez)



NCN: NSF \$18M - nanotechnology (Klimeck/Lundstrom)



C3Bio: DoE EFRC \$20M - biofuels (McCann)



PRISM: DoE \$17M - mems devices (Murthy/Strachan)

Supporting Purdue infrastructure



Purdue University Research Repository (PURR) – data mgmt



PurdueNExT / nanoHUB-U – online education

Supporting many other Purdue Projects



Outside Institutions



60+ Hubs for many disciplines

~1,500,000
visitors total

visitors	users
689,743	330,251
343,350	112,862
64,131	32,763
59,517	4,669
56,355	14,646
47,967	23,088
46,710	12,643
44,723	5,372
41,689	5,396
40,289	8,207
39,188	6,362
39,134	7,933



nanoHUB.org



nees.org



pharmaHUB.org



HABRIcentral.org



vhub.org



GlobalHUB.org



cceHUB.org



PURR



iemhub.org



StemEdHub.org



ciHUB.org



molecularHUB.org



Tool page

AgMIP Tool: A GEOSHARE tool for aggregating outputs from the AgMIP's Global Gridded Crop Modeling Initiative (Ag-GRID)

By [Nelson Benjamin Villoria](#), [Joshua Elliott](#), [Christoph Müller](#), [Jaewoo Shin](#)¹, [Lan Zhao](#)¹

¹ [Purdue University](#)

Aggregate the yield shocks provided by the AgMIP project from their original 30x30 min resolution to any user specified level.

[Launch Tool](#)

Version **1.2.4** - published on 05 Nov 2014

Open source: [license](#) | [download](#)

[View All Supporting Documents](#)

[40 users, detailed usage](#)

Share: [f](#) [t](#) [g+](#) [...](#)

[0 review\(s\)](#) ([Review this](#))

[0 questions](#) ([Ask a question](#))

[0 wish\(es\)](#) ([Add a new wish](#))

[About](#) [Usage](#) [Versions](#) [Reviews](#) [Questions](#) [Wishlist](#) [Supporting Docs](#)

Category

Published on

[Tools](#)

05 Nov 2014

Abstract

As part of the Agricultural Model Intercomparison Project (AgMIP), the Global Gridded Crop Model Intercomparison group (fast track phase) have estimated historical and future changes in yields for several crops under several distinct climate change scenarios. For a detailed description of these data and its uses please refer to Rosenzweig, C. et al. (2014), Elliott, J. et al. (2014), Nelson, J. et al. (2014), and Muller and Robertson (2014).

The output archive comprises time series (1971-2099) generated by seven crop models (EPIC, GEPIC, pDDSAT, LPJmL, IMAGE-AEZ, PEGASUS, LPJ-GUESS), under a number of temperature trajectories from a suite of five global climate models (HadGEM2-ES, IPSL-CMSA-LR, MIROC-EXM-CHEM, GFDL-ESM2M, NorESM1-M) and five representative

Geoshare



See also

No results found.

Rappture Builder

The screenshot displays the Rappture Builder interface for a simulation. The main window is titled "CNTbands" and includes a "Simulate" button and a "Keep for later" option. The "Structure" dropdown is set to "Carbon Nanotube" and the "Simulation Method" is "Pz orbital". A tooltip is visible over the "Simulation Method" dropdown, providing details about the two methods: "Pz orbital" (faster but less rigorous) and "Extended Huckel Theory" (more rigorous but more time-consuming). The "Chirality (n,m)" section shows "n: 7" and "m: 5". The "Model parameters" section includes "Tight Binding Energy: 3eV" and "Carbon-carbon spacing: 1.42A". The "Length in 3-D view" is set to "40". The main visualization area shows a 3D model of a carbon nanotube structure. The bottom status bar indicates "1 result" and a "Clear" button.

CNTbands

Structure: Carbon Nanotube

Simulation Method: Pz orbital

Simulate

Keep for later

About this tool Questions?

Result: Molecular structure: overall

Determine the simulation method.

Pz orbital:
The Pz Orbital model uses 1 Pz orbital/atom as the basis set. Of the two simulation methods, this has the advantage of being the faster, but the disadvantage of being the less rigorous.

Extended Huckel Theory:
The Extended Huckel Theory model uses 4 orbitals (S, Px, Py, Pz)/atom as the basis set. Of the two simulation methods, this has the advantage of being the more rigorous, but the disadvantage of being the most time consuming.

Chirality (n,m)

n: 7

m: 5

Model parameters

Tight Binding Energy: 3eV

Carbon-carbon spacing: 1.42A

Length in 3-D view: 40

1 result

Clear



Publication

Localization of MLL1 morphemes in mouse mm9 genomic DNA

By Minou Bina¹, Phillip Wyss¹, Daidong Wang¹, Xiaohui Song²


1. *Purdue University, Department of Chemistry* 2. *Purdue University, Academic Technologies and Rosen Center for Advanced Computing*

Supplementary materials for the publication entitled "Discovery of MLL1 binding units, their localization to CpG Islands, and their potential function in mitotic chromatin." Bina M, Wyss P, et al. *BMC Genomics*. 14:927 (2013)

Listed in [Datasets](#)

[Download \(BED\)](#)

Version 1.0 - published on Mar 31, 2014
doi:10.4231/R7KW5CXF - [cite this](#)

 Licensed under [CC0 1.0 Universal](#)

About

Supporting Docs

Versions

Reviews

Questions

Citations

Abstract Mixed Lineage Leukemia 1 (MLL or MLL1) plays central roles in the regulation of protein-coding genes. Previous studies of mice have established that MLL1 controls key developmental pathways including the formation of body plan during the early stages of embryogenesis. Even though MLL1 is best known for catalyzing methylation of lysine 4 in histone H3 (H3K4), reported data indicate that MLL1 performs histone methyltransferase-independent functions including gene-bookmarking during mitosis, favoring genes that were highly transcribed during interphase, to accelerate reactivation of transcription following mitotic exit (Blobel *et al.*, *Mol Cell*. 2009).

By analyzing results of reported DNA binding and chromatin immunoprecipitation assays, we have uncovered the DNA sequence elements that bind the MT-domain in MLL1: [Bina et. al, BMC Genomics 14:927 \(2013\)](#). We describe these elements as minimal units or morphemes: the smallest 'words' in DNA that selectively bind the MLL1 MT-domain. This publication provides a file (BINA_MLL_morphemes_mm9.bed) to display the position of MLL1 morphemes in the genome browser at UCSC. You can obtain a copy of the bed file (in text format) by clicking on download. Alternatively, you can use the link below to view the position of MLL1 morphemes directly on a track in the browser at UCSC.

While using the genome browser, select Dense to view morpheme positions in genomic DNA. Select pack when zooming to short DNA segments. If you use the data in your research, please cite the primary publication by Bina et al., published in *BMC Genomics* 14:927 (2013).

[Open the position of MLL1 morphemes in the mouse built mm9 genomic DNA:](#)

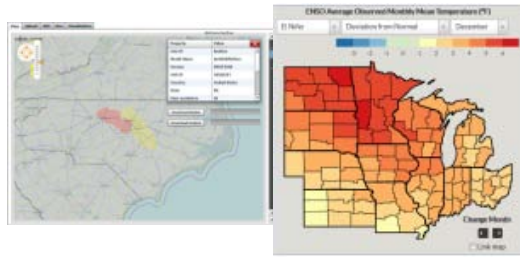


Building on HUBzero

- Extend RAPPTURE Toolkit
- Integrate map and spatial data rendering
- Adding new data objects to Rappture Builder
- Support geospatial data viewing and metadata in iData
 - Data management
 - Data sharing
 - Data quick view
 - Common processing
- Link data and tools in Hubzero
 - Dataset -> invoke tools
 - Tools -> result published and shared



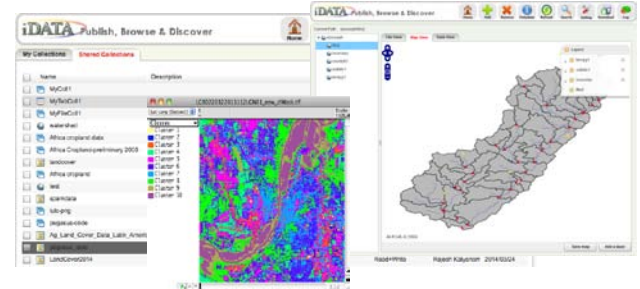
End User



Geospatial computation



Data presentation



Data Sharing

Building Blocks

Data management

Rendering

Tool builder

Remote servers

Layers

Geo-processing

Data-Tool linkage

Maps

Data sharing

Data presentation

HUBzero



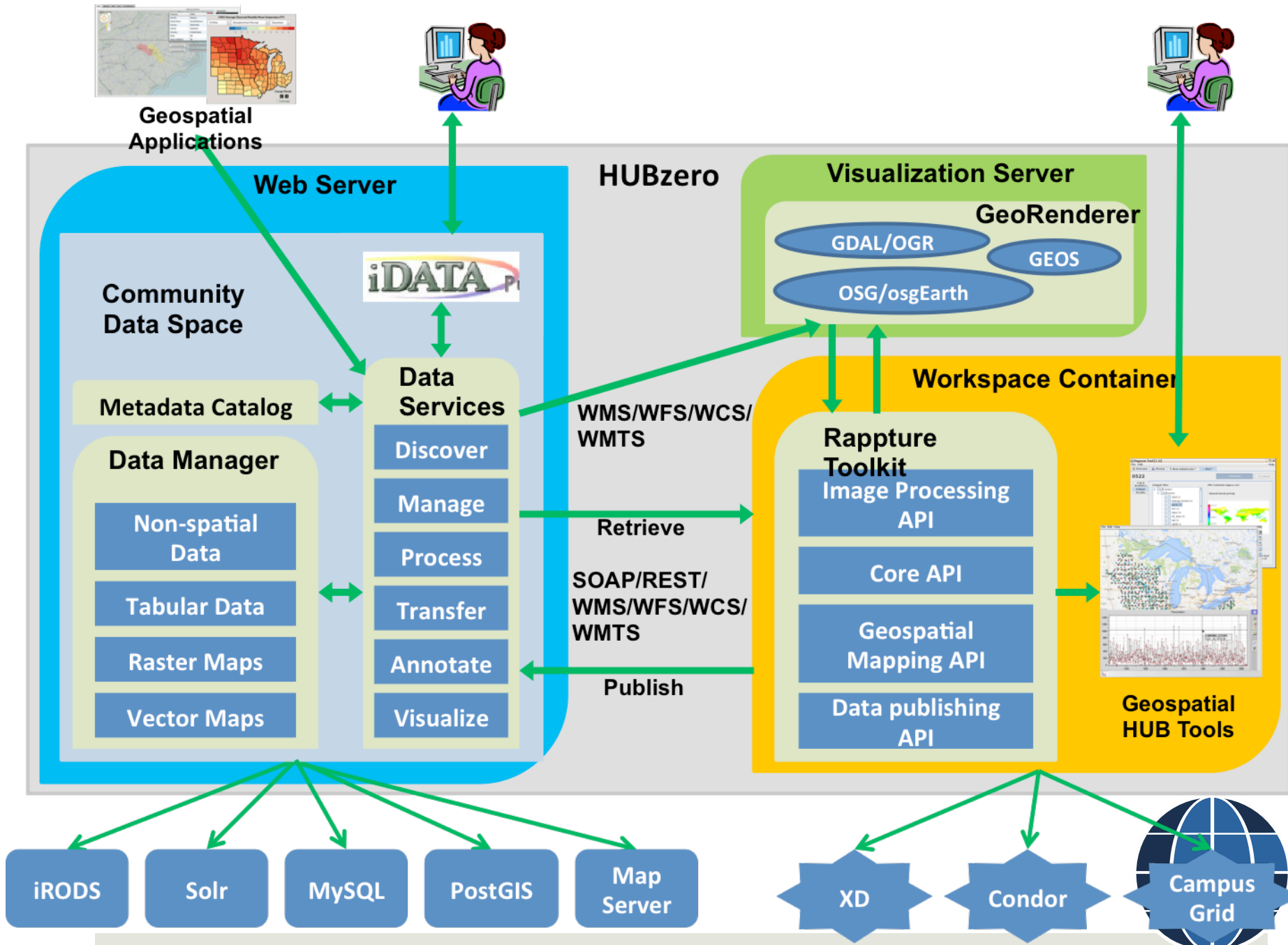
Computation tools and online databases, Content publishing, Collaboration (group, project), Learning (courses, self-help), Support (tickets, Q&A), Community (forum, review, calendar)



Resources

Supercomputers, storage, map servers, geo-processing engines





Examples

- Accessible via web browser
- Interactive
- Demonstrate capabilities



Chile Earthquake data

The image displays a software interface for analyzing earthquake data. It is divided into several sections:

- Map View (Top Left):** A map of Chile showing major cities including Valparaíso, Santiago, San Antonio, and Rancagua. A red dot indicates the location of the earthquake site.
- Photograph (Bottom Left):** A photograph showing the interior of a damaged building, likely the Torres del Sol, with significant structural damage to the walls and ceiling.
- Structural Plan (Right Panel):** A detailed floor plan of the 'TORRES DEL SOL' building. The plan shows a grid of columns and beams with dimensions and labels. The caption below the plan reads: "Fig. 7.18 (cont.) Floor Plan - Torres del Sol".
- Site Information (Right Panel, Top):** A text box labeled "Site:" containing the text "TORRES DEL SOL (Maule.Chile.2010)".

The interface includes standard menu bars (File, Edit, View, Help) and toolbars for navigation and analysis. A toolbar at the bottom contains icons for various data visualization and analysis tools.



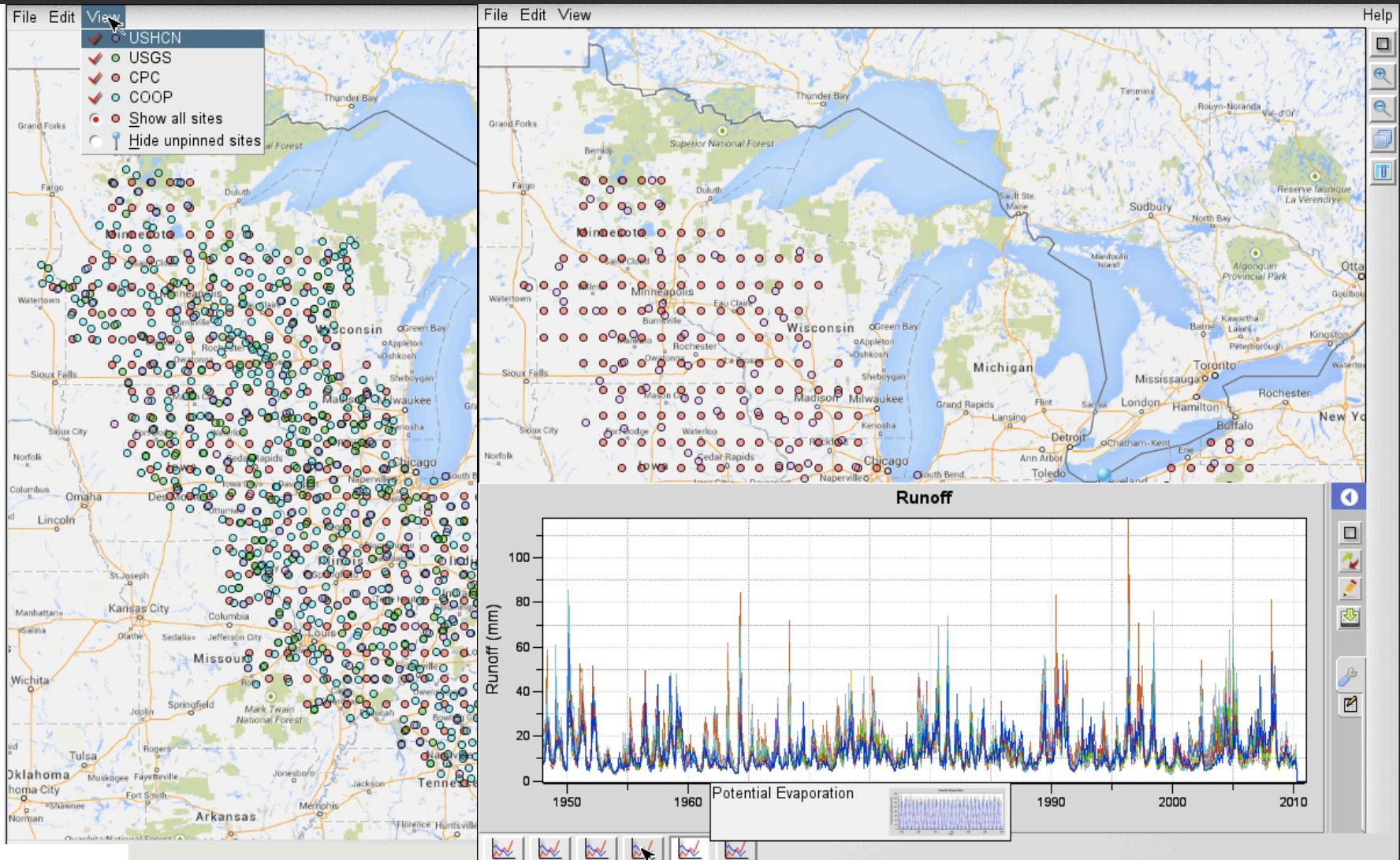
Time series precipitation data

The screenshot displays a GIS application interface with the following components:

- Station List Table:** A table listing various precipitation stations with their IDs, names, and coordinates. The 'Pre' column contains links to data pages.
- Map (Top Left):** A map of the Chicago metropolitan area showing major highways and city locations.
- Map (Bottom):** A map of the Lebanon, Indiana area with a pop-up window for 'LEBANON WATERWORKS' showing its ID (124908) and precipitation data for August 18, 2002.
- Time Series Chart:** A bar chart showing precipitation data from 1950 to 1990. The y-axis represents precipitation in inches, ranging from 0 to 6. Multiple colored bars represent different data series.

Station ID	Station Name	Latitude	Longitude	Pre
<input type="checkbox"/>	120132 ALPINE 2 NE	39.56666667	-85.16666667	/apps/geoexpl...
<input checked="" type="checkbox"/>	120177 ANDERSON SEWAGE PLT	40.1	-85.71666667	/apps/geoexpl...
<input type="checkbox"/>	120200 ANGOLA	41.63333333	-84.98333333	/apps/geoexpl...
<input type="checkbox"/>	120331 ATTICA 2 E	40.28333333	-87.18333333	/apps/geoexpl...
<input type="checkbox"/>	120482 BATESVILLE WATERWORK	39.3	-85.21666667	/apps/geoexpl...
<input type="checkbox"/>	120830 BLUFFTON 1 N	40.75	-85.16666667	/apps/geoexpl...
<input type="checkbox"/>	120922 BRAZIL	39.51666667	-87.11666667	/apps/geoexpl...
<input checked="" type="checkbox"/>	121147 BURLINGTON	40.48333333	-86.4	/apps/geoexpl...
<input type="checkbox"/>	121212 CAGLES MILL DAM	39.48333333	-86.91666667	/apps/geoexpl...
<input type="checkbox"/>	121256 CANNELTON	37.9	-86.63333333	/apps/geoexpl...
<input type="checkbox"/>	121415 CHALMERS	40.66666667	-86.88333333	/apps/geoexpl...
<input type="checkbox"/>	121628 CLINTON 2 W	39.66666667	-87.43333333	/apps/geoexpl...
<input type="checkbox"/>	121739 COLUMBIA CITY	41.15	-85.48333333	/apps/geoexpl...

Upper MS River Basin & OH River Basin Data



Start tools from datasets

The screenshot shows the iData web interface. The browser address bar displays the URL: <https://gabbs3.hubzero.org/idata/?controller=collection&task=browse&gid=22&path=Data>. The page title is "iData Publish, Browse, and Discover". Below the title, it says "You are here: Home". The main content area shows the "Contents of: Collections :: 2014 watershed: / Data". There is a toolbar with buttons for "Create Folder", "Upload File", "Parent Folder", "Download", "Rename", and "Search For Files". On the left, a sidebar shows a tree view of the file structure: Collections > 2014 watershed > Data > St Joe w. The main area contains a table with the following columns: Name, Size, Created, Modified, and Options. A blue arrow points to the "Data Explorer" button in the Options column of the row for "St Joe watershed".

<input type="checkbox"/>	Name	Size	Created	Modified	Options
<input type="checkbox"/>	St Joe watershed		9/8/2014 5:40 PM	9/8/2014 5:40 PM	Data Explorer



Publish data from tools

Please select the case from model list at your left and press run.

Run

Refresh

Job ID	Model Name	Job Type	Job Status	Submission Time	Actions
1394	flatriver	normal	DONE	2014-07-14-16:12:04 EDT	Output Delete Publish
1393	flatriver	normal	DONE	2014-07-14-15:12:41 EDT	Output Delete Publish
1392	flatriver	normal	FAILED	2014-07-14-14:52:34 EDT	Output Delete Publish
1382	flatriver	normal	DONE	2014-07-14-09:22:51 EDT	Output Delete Publish
1379	flatriver	normal	FAILED	2014-07-13-22:31:15 EDT	Output Delete Publish

Log Information

2014-07-14-16:12:04 EDT : Log history entry has been inserted to database.
2014-07-14-16:12:04 EDT : Credential File location is /group/ssg/waterhub/apps/swat/run/x509up_u318864
2014-07-14-16:12:04 EDT : GlobusCredential has been created.
2014-07-14-16:12:04 EDT : Created GSS credential.
2014-07-14-16:12:04 EDT : Created the RSL file for globus run
2014-07-14-16:12:04 EDT : Resource Specification RSL String =

```
&(executable=/group/ssg/waterhub/apps/swat/run/swat_run_zip.csh)(arguments = '/group/ssg/waterhub/data/hydroshare-dev/swat/users/lanzhao/input/flatriver.zip' '/group/ssg/waterhub/data/hydroshare-dev/swat/users/lanzhao/output/flatriver-1394-out.tar' '/group/ssg/waterhub/data/hydroshare-dev/swat/users/lanzhao/jobs/flatriver' 'SWAT2009' 'flatriver' '/group/ssg/waterhub/data/hydroshare-dev/swat/users/lanzhao/output/flatriver-out.tar' 'normal' '/group/ssg/waterhub/data/hydroshare-dev/swat/users/lanzhao/jobs/flatriver/swat-1394.out' '/group/ssg/waterhub/data/hydroshare-dev/swat/users/lanzhao/jobs/flatriver/swat-1394.err')(stdout=/group/ssg/waterhub/data/hydroshare-dev/swat/users/lanzhao/jobs/flatriver/swat-pbs.out)(stderr=/group/ssg/waterhub/data/hydroshare-dev/swat/users/lanzhao/jobs/flatriver/swat-pbs.err)(proiect="TG-ATM090060")
```

My Models

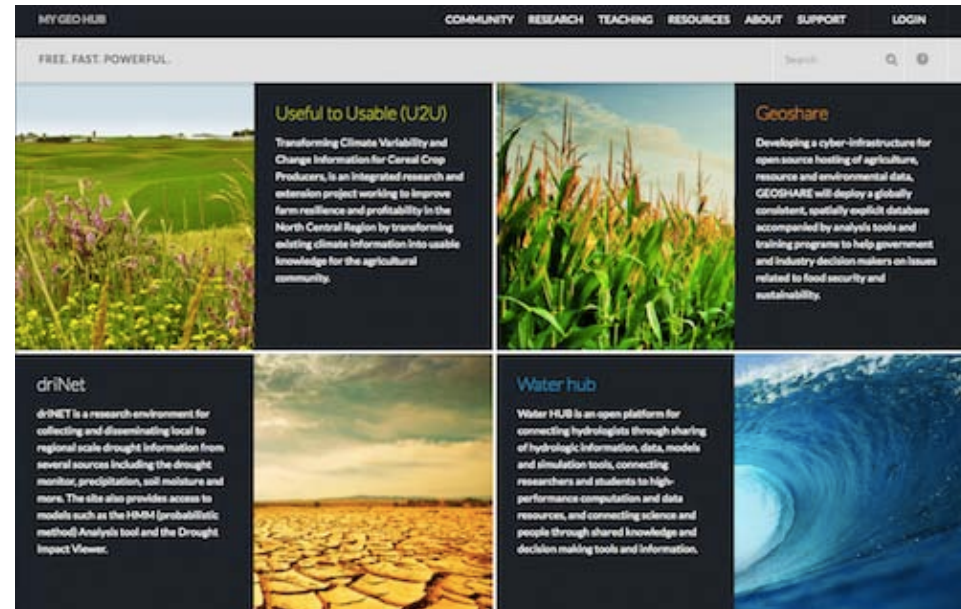
flatriver

Shared Models

Other Models

MyGeoHUB.org

- A hub for geospatial modeling, data analysis and visualization
- Hosting tools, datasets, groups and projects
- Free to sign up
- GABBs software will be installed on MyGeoHub.org
- Demonstration applications will be hosted
- User feedback and wishes are welcome!
- Build your own project, group, and community with Supergroups!



Super Groups



DECISION DASHBOARD MEDIA CENTER NEWSLETTER ABOUT US

powered by mygeohub

Water Hub
Platform for water education, research, data access, partnership and collaboration

HOME TOOLS RESOURCES EXPLORE ABOUT WATER HUB

SWATShare

Upload, share, run and analyze SWAT models online using SWA

Helping producers m

Weather and climate patterns in cropping systems. Useful to U working to improve farm resili existing climate data into usat producers make better long-t manage crops for maximum y



G · E · O · S · H · A · R · E

GEOSPATIAL DATA HOSTING FOR DISCOVERY AND DECISION MAKING

HOME RESC

Agriculture

NEWS ABOUT GABBS RESOURCES COMMUNITY



Enabling scientists, students and educators to create and share interactive tools and models for processing, analyzing and visualizing geospatial data

Welcome to Water J

GEOSPA

GABBS is an NSF-funded project to create a powerful Web-based system that will allow researchers worldwide to manage, curate, share, analyze and visualize geospatial data for purposes ranging from predicting damaging floods to projecting climate change effects on the poor. The project will build geospatial data hosting, processing and sharing capabilities into Purdue's HUBzero platform. This should open the way for easy development of a variety of Web-enabled tools for probing and presenting geospatial data in ways that can, among other things, help policymakers address pressing issues in the U.S. and around the globe.

Latest News



Collaboration

- Use cases
 - Datasets
 - applications
- Feedback, suggestions
- Participation
 - Share your work on [MyGeoHub.org](https://mygeo.org)
- Ask questions

