



G · E · O · S · H · A · R · E

An introduction to GEOSHARE's
cyberinfrastructure:
exchange and analysis of geo-spatial
data through HUBzero

Nelson Villoria, GEOSHARE and NSF-DIBBS Co-Principal Investigator

Department of Agricultural Economics

Purdue University

GLP 2nd Open Science Meeting, Berlin March 21, 2014

This presentation is based on work with the following collaborators:

- ▣ HongJun Choi (Purdue U.)
- ▣ Delphine Derying (University of East Anglia)
- ▣ Joshua Elliot (AgMip)
- ▣ Thomas Hertel (Purdue U.)
- ▣ Erich Huebner (Purdue U.)
- ▣ Rajesh Kalyanam (Purdue U.)
- ▣ Danny Plouffe (McGill U.)
- ▣ Navin Ramankutty (McGill U.)
- ▣ Carol Song (Purdue U.)
- ▣ Lan Zhao (Purdue U.)



GEOSHARE's Mission

GEOSHARE is a decentralized network of scientists in many disciplines which mission is to provide:

- ▣ a freely available, global, spatially explicit database on agriculture, land use, and the environment
- ▣ analysis tools for scientists, decision makers, and development practitioners



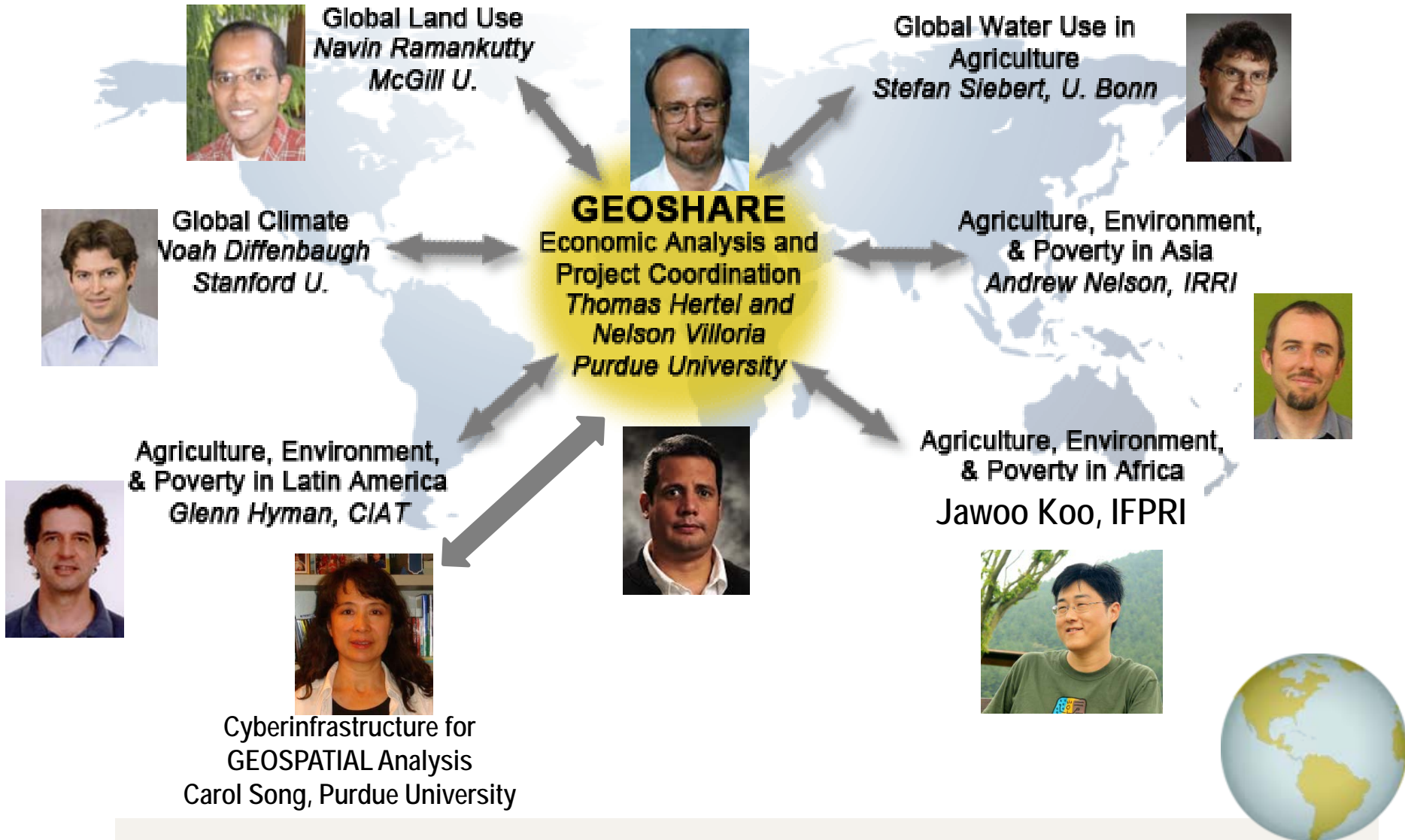
GEOSHARE: Vision of a Network

GEOSHARE envisions a vibrant global network:

- contributing to shared cyberinfrastructure
- enhancing capacity for geospatial analysis
- applying geospatial tools to guide decision making related to food security, land use, environmental sustainability and poverty reduction



GEOSHARE's Network Today (Pilot Project)



Why a shared cyberinfrastructure
for GEOSHARE?

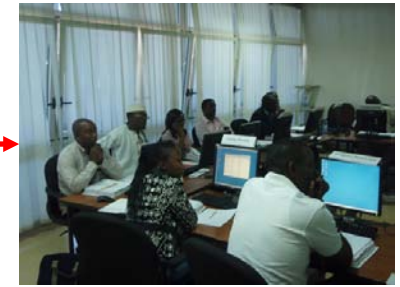


Different Worlds

Expert groups in educational and international institutions generate knowledge, data, and tools

Often, knowledge stays within disciplinary silos

Users elsewhere wish to use knowledge for decision making, learning, etc.



Crop Models



What Keeps Them Separate?



Commercial Software is expensive, requires powerful Hardware, and specialization
SAS ≠ STATA ≠ ARCGIS



Multidisciplinary collaborations to open the silos are costly



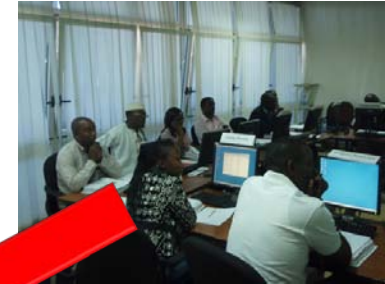
© Riccio Erlic



HUBzero technology seeks to break these barriers

Use the Web to share data, code, hardware, and spread the impact of research efforts and funds

No additional Software to install or learn



How to do this?



- HUBzero is a technology developed at Purdue University with NSF funds for creating dynamic websites for scientific research and educational activities:
 - It's fully open source
 - Whatever can be compiled in Linux, can be run in the Hub
 - Rapid deployment of Graphical User Interfaces
 - Access to cluster computer resources



Have this worked elsewhere?

- We have been inspired by nanoHUB.org
- nanoHUB.org's objective is to transfer the knowledge generated in extensive nanotechnology facilities to others in academy and industry
- Created by the NSF-funded Network for Computational Nanotechnology.
- Next slides from Prof. Gerhard Klimeck, nanoHUB.org Director



The nanoHUB Proposition and Emerged (and Busted) Myths

Proposition

- ▣ Be Developer Friendly

-
- ▣ Be Accessible

-
- ▣ Be User Friendly

Myth:

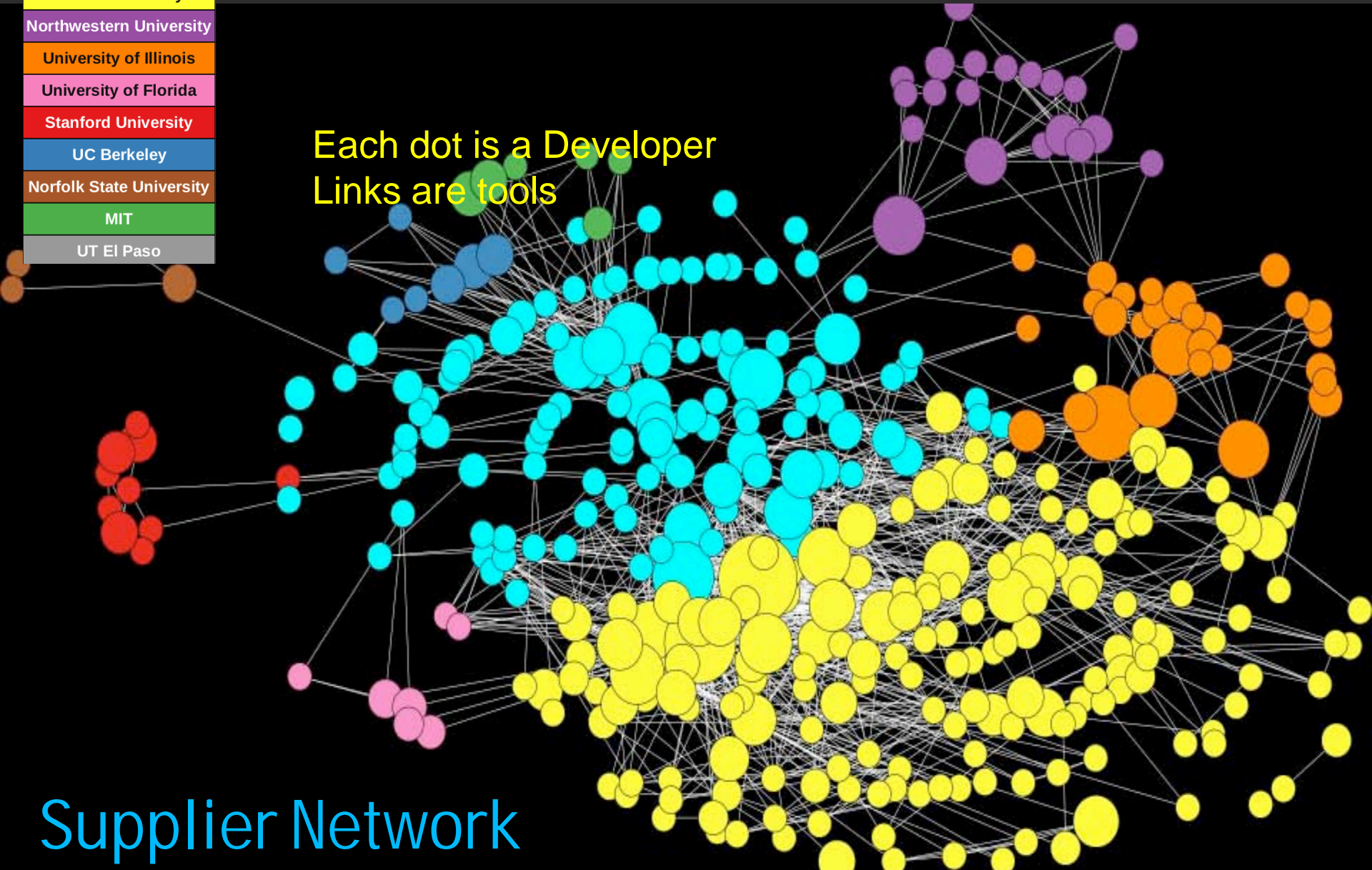
- ▣ Cannot use research codes for education
- ▣ Must write own code to do research
- ▣ Experimentalists cannot use research codes
- ▣ NO End-to-end Science Cloud Possible
- ▣ Building User Interfaces too Difficult
- ▣ Must rewrite code for web deployment
- ▣ There is no incentive to share codes



nanoHUB.org: Developer Collaboration Network

Purdue University
Northwestern University
University of Illinois
University of Florida
Stanford University
UC Berkeley
Norfolk State University
MIT
UT El Paso

Each dot is a Developer
Links are tools



Supplier Network

<http://nanoHUB.org>

Network members in 172 countries

Over 12,000 simulations/year

Over 300,000 users annually



A world map with a dark blue background. Numerous small, semi-transparent bubbles of varying sizes are scattered across the continents. The bubbles are color-coded: green, yellow, and red. The highest density of bubbles is in North America, Europe, and East Asia. The size of the bubbles varies, with larger ones appearing more frequently in North America and Europe. The map shows a global distribution of users, with significant concentrations in the Northern Hemisphere.

New Registrations
Simulation Users
Tutorial / Lecture Users

GEOSHARE Example 1: Pegasus



Pegasus 1.0 Predicting Ecosystem Goods And Services Using Scenarios

- Published by/in:
 - Deryng, D., W. J. Sacks, C. C. Barford, and N. Ramankutty. 2011. "Simulating the Effects of Climate and Agricultural Management Practices on Global Crop Yield." *Global Biogeochemical Cycles* 25 (May): 18 PP.
- A global crop model that integrates climate, the effect of planting dates and cultivar choices, irrigation, and fertilizer application on crop yield for maize, soybean, and spring wheat.
- Useful for studying adaptation to climate change.



Inputs when we started:

- ▣ Fortran Code with Model Equations:

```
author Mike  
This class will use the ghost particles which  
to insert a collection of particles which  
have no net charge. This is used to calculate  
chemical potential and activity coefficients  
class widom : public analysis {  
private:  
    average<double> expsum; ///  
protected:  
    int ghostin;           ///  
    long long int cnt;    ///  
    vector<particle> g;    ///  
public:  
    widom(int n=10);  
    string info();  
    void add(particle);  
    void add(container &);  
    void insert(container &, energybase &  
    void check(checkValue &);  
    double ma() { return -log(expsum);
```

- ▣ Few megabytes of data in NetCDF format to calibrate the model.



Now: Pegasus 1.0 at geoshareproject.org

Geoshare - resources: Tools: Pegasus To...
https://geoshareproject.org/tools/pegasus/session?sess=2244

G · E · O · S · H · A · R · E
GEOSPATIAL DATA HOSTING FOR DISCOVERY AND DECISION MAKING

SEARCH Geoshare

NELSON B VILLORIA (PURDUE) (NVILLORIA) | MY ACCOUNT | LOGOUT

HOME RESOURCES MEMBERS EXPLORE WORKSHOPS ABOUT SUPPORT HELP

Pegasus Tool [1.0]
File Edit Help
Welcome History

or completed runs.

My Runs

About Pegasus
Find out more about PEGASUS and related publications.

At A Glance...
Total Simulations: 1
Running Simulations: 0
Disk Space Used: 1.71 KB

Storage (manage) 38% of 10GB 894 x 670

Share session with: (supports usernames, user IDs, and e-mails)
This session is shared with:
(none)

Share with one of your Groups:
Select Group

zotero

- > No code was rewritten!
- > No Web developer was needed!
- > No additional extensions for the web server
- > For next Pegasus, just update the code!



On-the-fly visualization of results

The screenshot shows the Pegasus Tool [1.0] interface. The main window is titled "simtest" and has a status bar indicating "Finished". The interface is divided into several sections:

- Log & Statistics:** A sidebar on the left with buttons for "Log & Statistics", "Output", and "Results".
- Output Files:** A list of files with checkboxes. The file "cyield.nc" is selected and checked.
- File Contents (cyield.nc):** A panel showing the contents of the selected file, titled "Economic yield, in carbon units". It displays a world map with a color scale ranging from 0.00 (blue) to 1.05 (red).
- Download File:** A button at the bottom of the file list.
- Download Archive:** A button at the bottom of the file list.

At the bottom of the window, there is a storage indicator showing "38% of 10GB" and a resolution of "894 x 670".


Download output for further processing



GEOSHARE Example 2: Processing AgMIP's Aggregated Data



The AgMIP GRIDded crop modeling initiative

- Crop Modeling Teams grouped around  AgMIP
- The Global Gridded Crop Model Intercomparison (GGCMI):
 - Fast track phase:
 - 5 Crop Models
 - 5 Global General Circulation Models
 - 5 Representative Concentration Pathways
 - CO2 Fertilization and irrigation
 - 12 Crops
 - All the combinations ran from 1971 to 2099
- + 35,000 $0.5^{\circ} \times 0.5^{\circ}$ grids with crop yields



The Problem

- ❑ Economic modelers use data aggregated to the country level
- ❑ NetCDF and other spatial data delivery formats are foreign to non-specialists
- ❑ Download of large grids is challenging even with good bandwidth
- ❑ Storage is challenging for otherwise very good PCs
- ❑ How to make the data available to Users



The Solution

- ▣ Good all-around R Code:



- ▣ Reads data from Globus Online to the HUB



- ▣ Use `aggregate()` function
- ▣ Aggregates using mapping from XY coordinate to country, AEZs, up to the user.
- ▣ Calculate summary statistics (mean, etc.) or weighted averages using user provided weights (production, area, population, etc.)
- ▣ Wrap the R function around a Graphical User interface



The Result

Crop Data 1.0 CITATION WILL BE HERE Help

AgMIP Download | Aggregation | Visualization

Crop Data

Model	GCM	RCP	SSP	Crop
<input type="radio"/> EPIC	<input type="radio"/> HadGEM2-ES	<input type="radio"/> hist	<input type="radio"/> ssp2	<input type="radio"/> maize
<input type="radio"/> GEPIC	<input type="radio"/> IPSL-CM5A-LR	<input type="radio"/> rcp8p5	<input type="radio"/> CO2	<input type="radio"/> soy
<input type="radio"/> pDSSAT	<input type="radio"/> MIROC-ESM-CHEM	<input type="radio"/> rcp6p0	<input type="radio"/> co2	<input type="radio"/> wheat
<input type="radio"/> LPGmL	<input type="radio"/> GFDL-ESM2M	<input type="radio"/> rcp4p5	<input type="radio"/> noco2	<input type="radio"/> rice
<input type="radio"/> IMAGE-LEITAP	<input type="radio"/> NorESM1-M	<input type="radio"/> rcp2p6	<input type="radio"/> IRR	<input type="radio"/> managed_grass
			<input type="radio"/> noirr	<input type="radio"/> others
			<input type="radio"/> firr	<input type="radio"/> rapeseed
				<input type="radio"/> barley
				<input type="radio"/> millet
				<input type="radio"/> sorghum
				<input type="radio"/> sugarcane
				<input type="radio"/> sugar_beet

Path:

Log

Crop Data 1.0




Citations [min] [max] [close]

[Proceedings of the National Academy of Sciences of the United States of America, March 4, 2014; 111 \(9\)](#)

Please use the following citations when using these data:

- [Rosenzweig, C. et al. \(2014\). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. Proceedings of the National Academy of Sciences, 111 \(9\): 3268-3273.](#)
- [Elliott, J. et al. \(2014\). Constraints and potentials of future irrigation water availability on global agricultural production under climate change. Proceedings of the National Academy of Sciences, 111 \(9\): 3239-3244.](#)
- [Müller, Christoph, and Richard D. Robertson. "Projecting future crop productivity for global economic modeling." Agricultural Economics 45.1 \(2014\): 37-50.](#)
- [Nelson, J. et al. \(2014\). Climate change effects on agriculture: Economic responses to biophysical shocks. Proceedings of the National Academy of Sciences, 111 \(9\): 3274-3279.](#)

CSV [Browse] [Create] [Help]




Years

100
80
60
40
20
0

[Download]

Crop Data 1.0 [min] [max] [close] **Citations** [min] [max] [close]



Aggregation Possibilities

Crop Data 1.0

AgMIP CITATION WILL BE HERE Help

AgMIP Download **Aggregation** Visualization

Input Files

1. AgMIP Files:

2. Region Map:

3. Weight Map or Aggregation Function

Weight Map:

Aggregation Functions

Enable Functions MAX MIN MEAN SD

Log

Crop Data 1.0



Aggregation Possibilities

The screenshot shows the 'Crop Data 1.0' web application interface. The main window has three tabs: 'AgMIP Download', 'Aggregation', and 'Visualization'. The 'Aggregation' tab is active. The 'Input Files' section contains three main fields:

- 1. AgMIP Files:** A text input field with the placeholder 'Select agmip files' and a 'Browse' button.
- 2. Region Map:** A text input field containing the path '/apps/cropdatatool/r163/e...'. There are 'Reset' and 'Browse' buttons next to it.
- 3. Weight Map or Aggregation Function:** A section containing a 'Weight Map' field with the same path, and an 'Aggregation Functions' sub-section with a checked 'Enable Functions' checkbox and a radio button for 'MAX'. A 'Reset' button is located below this section.

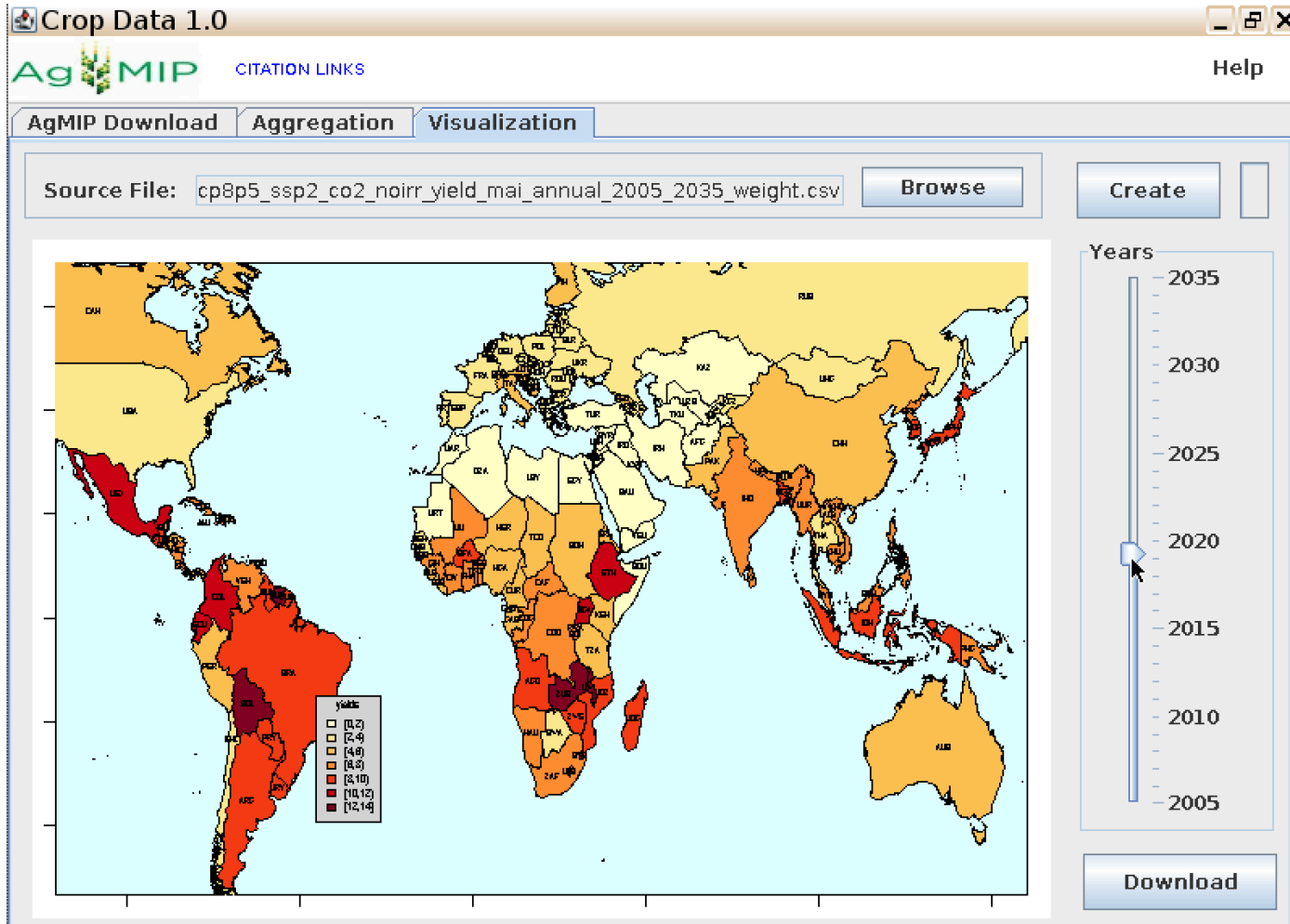
At the bottom left, there is a 'Log' section with an empty text area. A modal dialog box is overlaid on the right side of the application, titled 'Opening AgMIP_aggregation_20140320_143356.zip'. The dialog contains the following information:

- You have chosen to open:**
- AgMIP_aggregation_20140320_143356.zip** (with a zip icon)
- which is: Compressed (zipped) Folder (21.1 KB)
- from: <https://geoshareproject.org>
- What should Firefox do with this file?**
- Open with** Windows Explorer (default)
- Save File**
- Do this automatically for files like this from now on.**

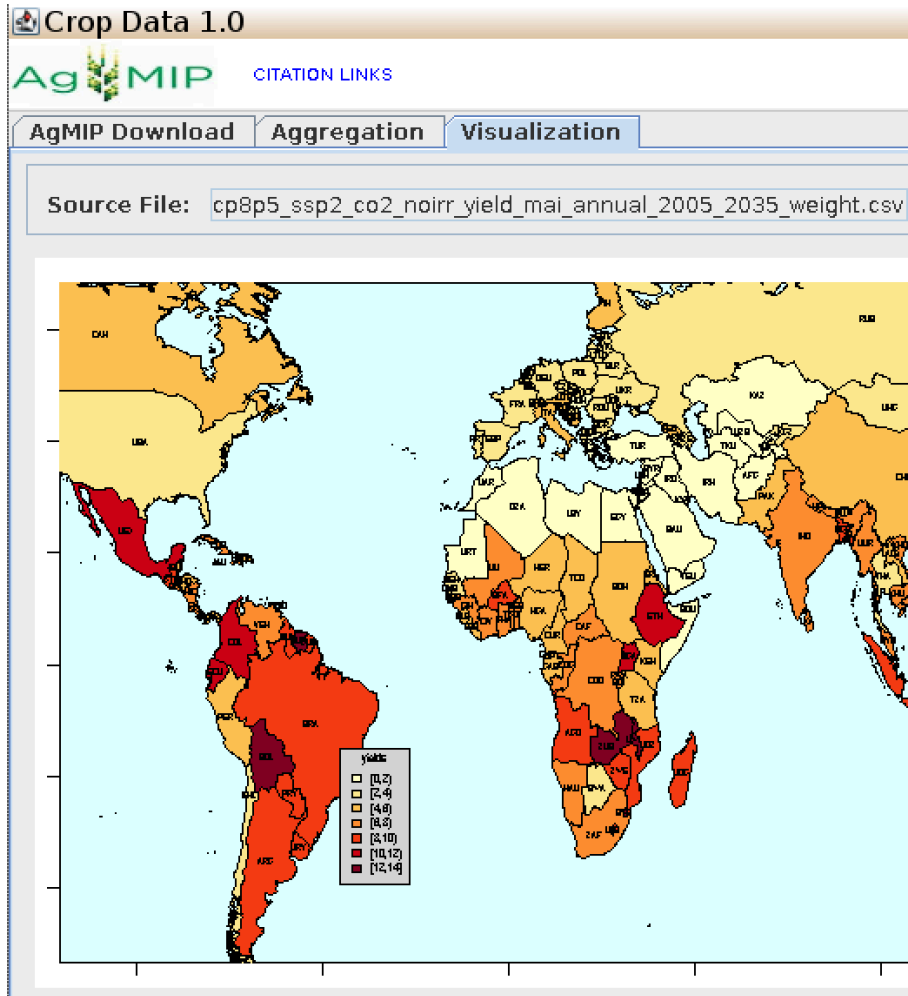
Buttons for 'OK' and 'Cancel' are at the bottom of the dialog. The application title bar at the bottom reads 'Crop Data 1.0'.



Visualization



Visualization



FILE HOME INSERT PAGE LAYOUT

From Access From Web From Text From Other Sources Existing Connections Ref AI

Get External Data

A1

	A	B	C	D
1	id	time	V1	
1442	1441	VEN	2051	2.375524
1443	1442	VEN	2052	2.293012
1444	1443	VEN	2053	2.258084
1445	1444	VEN	2054	2.380432
1446	1445	VEN	2055	2.236955
1447	1446	VEN	2056	2.362664
1448	1447	VEN	2057	2.220894
1449	1448	VEN	2058	2.322215
1450	1449	VEN	2059	1.946219
1451	1450	VEN	2060	2.153299

Download



Costs and Impacts

- ▣ Costs to the developers:
 - ▣ Write R function
 - ▣ Test it
 - ▣ Develop GUI

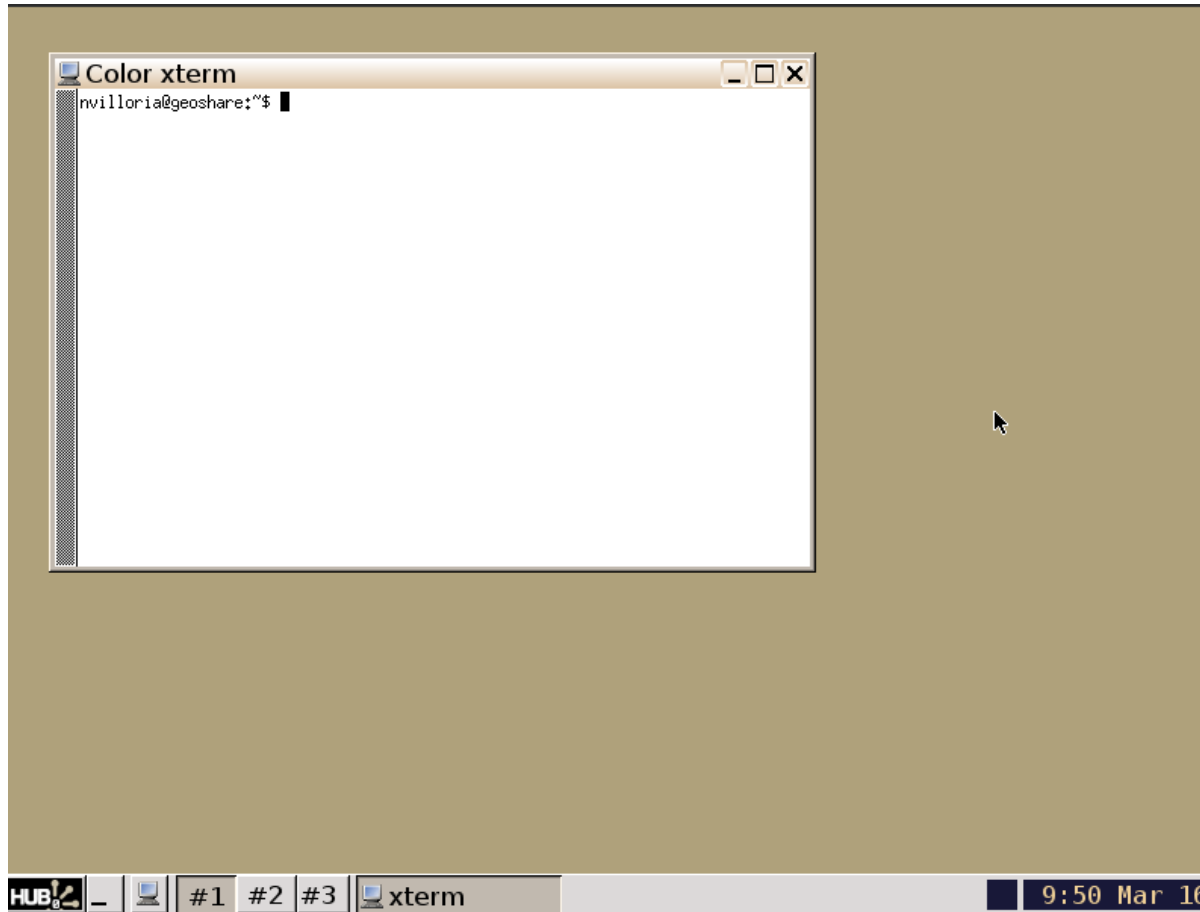
- ▣ Cost to the User:
 - ▣ None.

- ▣ Software Licenses, Hardware and Other Equipment
 - ▣ None.
 - ▣ All the processing on-the-fly and on university owned hardware

- ▣ Impacts:
 - ▣ Anyone interested in getting aggregated climate shocks for aggregated analysis can do so.



A peak inside the Hub: A Linux Workspace



If it can be run on Linux, it can be converted on a web application



Graphical User Interfaces

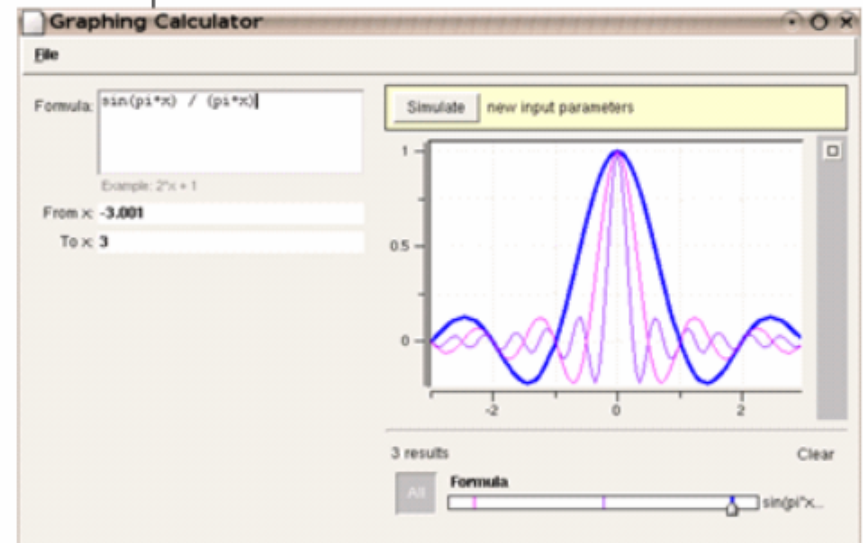
- ▣ HUBzero supports its own language, XML based, called RAPPTURE toolkit
- ▣ But also supports Python, Java, etc.
- ▣ User's knowledge and experience set boundaries.




```

<?xml version="1.0"?>
<run>
  <tool>
    <title>Graphing Calculator</title>
    <about>Press Simulate to view results.</about>
    <command>python @tool/graph.py @driver</command>
  </tool>
  <input>
    <string id="formula">
      <about>
        <label>Formula</label>
        <hints>Example: 2*x + 1</hints>
      </about>
      <size>30x5</size>
    </string>
    <number id="min">
      <about> <label>From x</label> </about>
      <default>0</default>
    </number>
    <number id="max">
      <about> <label>To x</label> </about>
      <default>1</default>
    </number>
  </input>
  <output>
    <curve id="result">
      <about> <label>Formula: Y vs X</label> </about>
    </curve>
  </output>
</run>

```

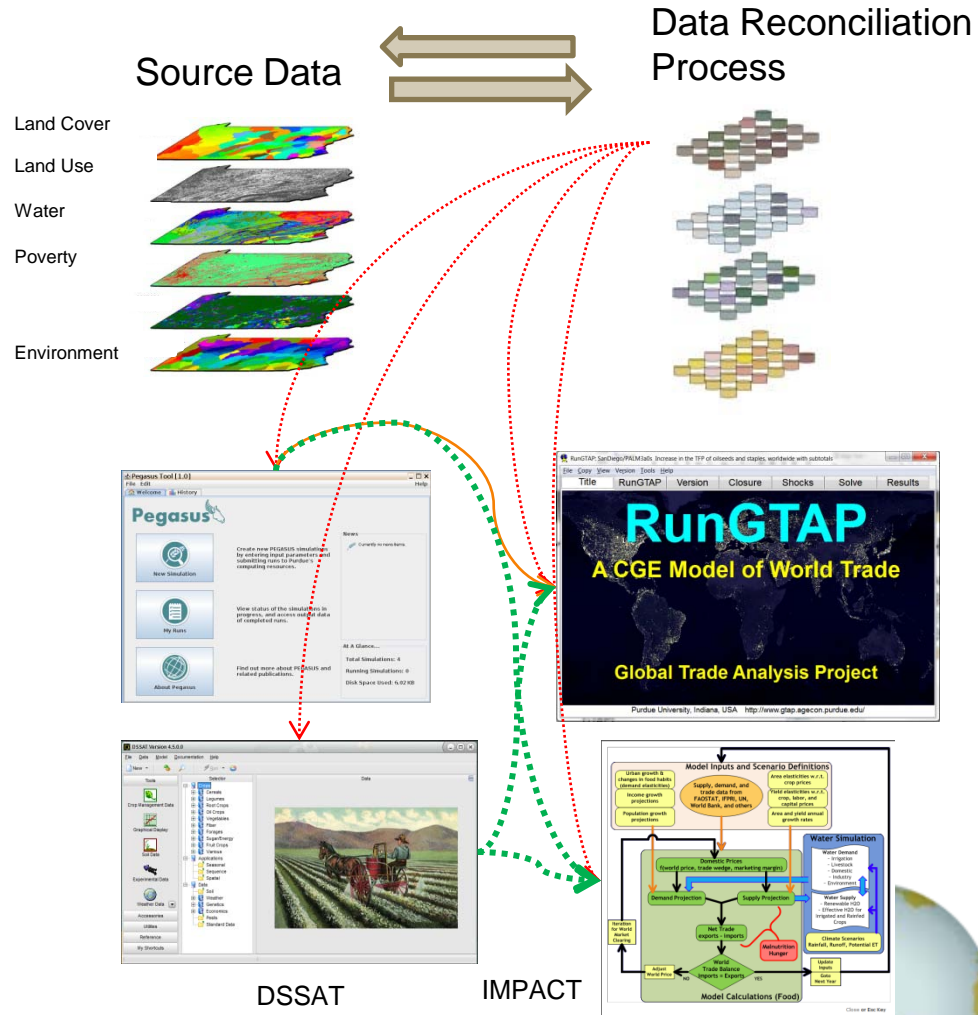


GEOSHARE Example 3: A workflow for assessing the value of improved data



Workflows: From data to analysis

- Gridded source data on land cover & use, water, poverty and environment - *flows into* -
 - Data reconciliation 'models', e.g. SPAM which produce usable data - *for use in* -
 - Biophysical and economic models (e.g., DSSAT, IMPACT, GTAP)
- Sensitivity of results at final stage determines value of improving quality of source data



Other considerations



GEOSHARE is free for all users and open to all

- ▣ Yes, users pay nothing.
- ▣ What about developers?
 - ▣ This is shared space
 - ▣ The cost of the Hub is around 50,000 per/year
 - ▣ Working on GEOSHARE governance cost-sharing model for self sustainability
 - ▣ But opportunities for economies of scale are huge.

Space is always available!



Recent Funding For Expanding HUBzero's Spatial Capabilities

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



Sister 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>



Overarching goals

- ▣ Enabling geospatial modeling and analysis online
- ▣ Anyone can create an online geospatial app and share
- ▣ Further support for geospatial data
- ▣ Building blocks can be used by other projects

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



▣ Questions

▣ Feedback?

▣ Thanks!

▣ nvillori@purdue.edu

