

Support Big Environmental Data Sharing and Analyses with Advanced Computing Technologies

Chaowei Yang, Co-Director
Center of Intelligent Spatial Computing for Water/Energy Science
George Mason University



OUTLINE

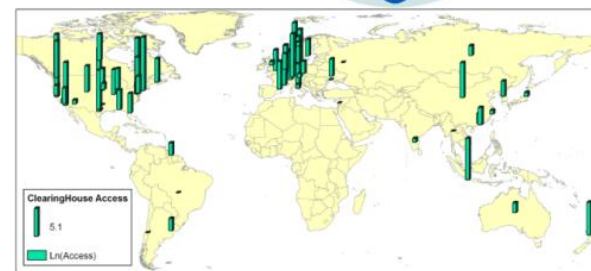
- **GEOSS Clearinghouse**
 - Background
 - Semantics and Results Sorting
 - Performance within a Federated Environment
 - Categorization
 - Data Exploration
- **Climate Studies and Big Data Analyses**
 - Project Background and the Big Data Dilemma
 - Online Analyses Requirements
 - Spatiotemporal Statistics Challenge
- **One Future Effort**



GEOSS Clearinghouse Objective



1. Share Global Earth Observation Data Among 140+ Countries to Address Global Challenges, such as Natural Hazards and Emergency Responses
2. Provide an entry point for Global Earth observation Data
3. Support End Users to **Discover, Access, and Utilize** of EO Data
4. Provide **Responses** to End Users in Seconds

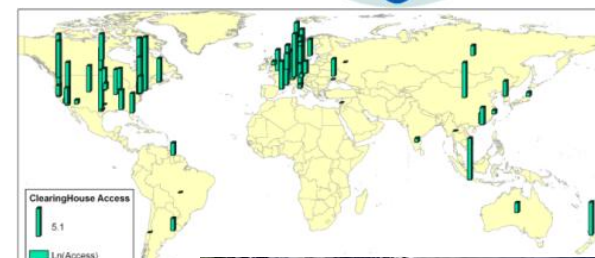




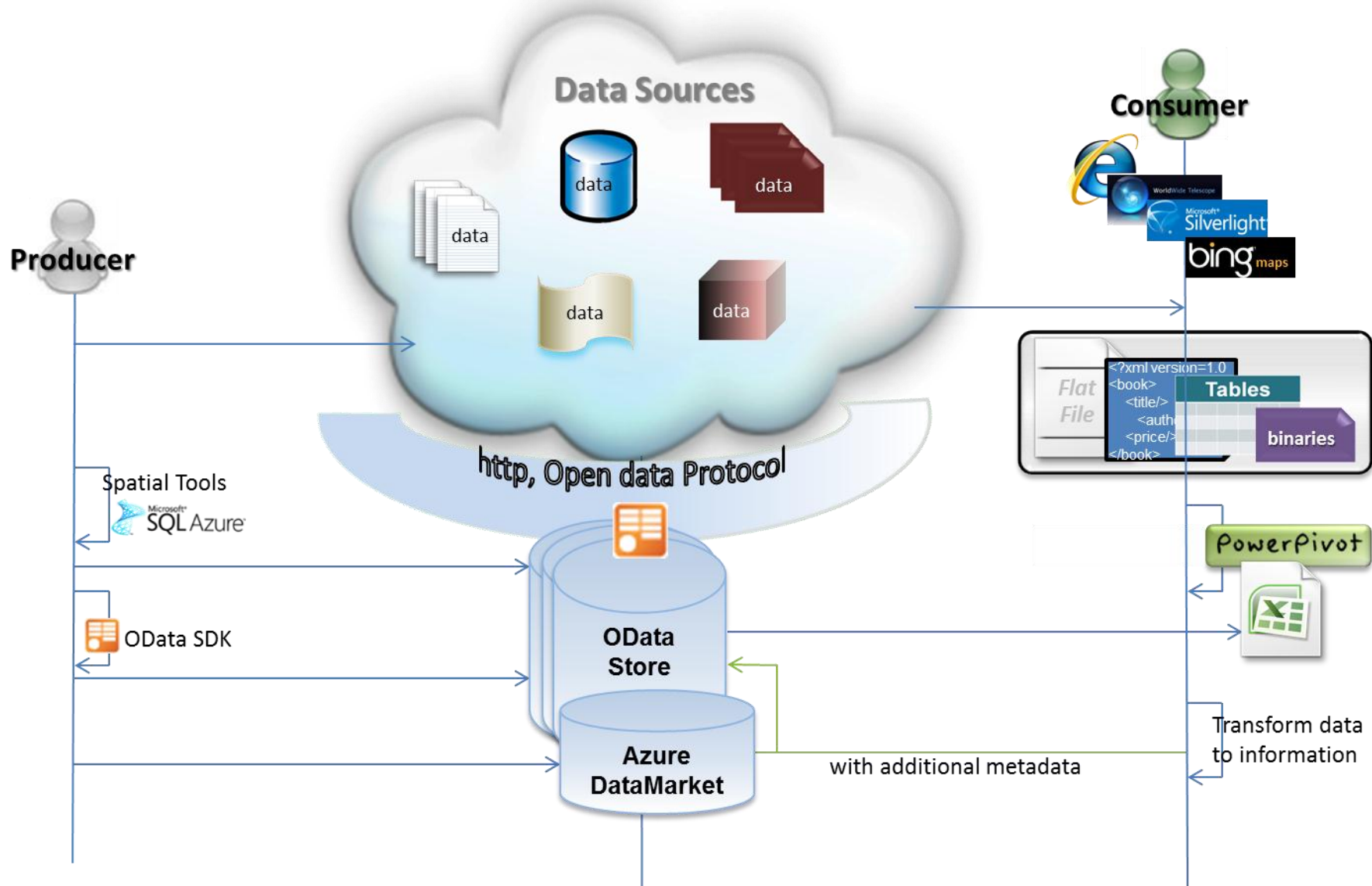
GEOSS Clearinghouse Challenges



1. Provide Same User Experience to Global Users
2. Be as Interactive as Possible and Provide Information based on Geospatial Framework Information
3. Support Different Societal Benefit
4. Provide Fast (seconds) Responses within a Federated Environment
5. Provide Accurate Results for Searching



Microsoft® Environmental Informatics Framework





Vocabularies and Semantics

❑ Dynamic Distributed Search for Federated Catalogs

❑ Rank Results

Sim

GO GROUP ON
EARTH OBSERVATIONS

GEOSS Clearinghouse



[Home](#) | [Contact us](#) | [Links](#) | [About](#) | [Help](#) |

English
Username Password [Login](#)

WHAT?

Either of the words

Exact phrase

What?

Without the words

Title

Abstract

Keywords

☒ Map type
☒ Search accuracy

WHERE?

lat (max)

long (min)

lat (min)

long (max)

Type

Region

WHEN?

[Search](#)

[Reset](#) [Hide advanced options](#)
☒ Restrict to ☒ Options

☒ Show map

FIND INTERACTIVE MAPS, GIS DATASETS, SATELLITE IMAGERY AND RELATED APPLICATIONS

Aggregate Results matching search criteria : 1-5/5 (page 1/1), 0 selected

Select : [all](#), [none](#)

[actions on selection](#)

Sort by



☐ CANADIAN CONSERVATION AREAS DATABASE

Abstract

Due to the work and commitment from many regional agencies to establish conservation and protected areas, a framework...

[Metadata](#)

[Interactive Map](#)



☐ OCEAN FRAMEWORK DATA INDEX

Abstract

Based on Natural Resource Maps (primary map scale of 1:250000) this Ocean Framework data provides a geomorphological interpretation of the seafloor, through the use of contours, in the same way that ...

[Metadata](#)

[Interactive Map](#)



☐ OCEAN FRAMEWORK DATA

Abstract

Based on Natural Resource Maps (primary map scale of 1:250000) this Ocean Framework data provides a geomorphological interpretation of the seafloor, through the use of contours, in the same way that ...

[Metadata](#)

[Interactive Map](#)



☐ LIDAR-DERIVED PRODUCTS FOR THE DRY CREEK WATERSHED, IDAHO

Abstract

Light Detection and Ranging (LIDAR) has emerged as a promising technology in many disciplines, including geology, forestry, and hydrology. Despite a growing body of LIDAR literature, gaps in informat...

l, q)



Relevance Demo

<http://geosearch.cloudapp.net/>

A screenshot of a web browser window displaying the GeoSearch application. The browser's address bar shows the URL <http://geosearch.cloudapp.net/>. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The browser's toolbar includes a back button, a forward button, a search button, and a tab labeled "GeoSearch".

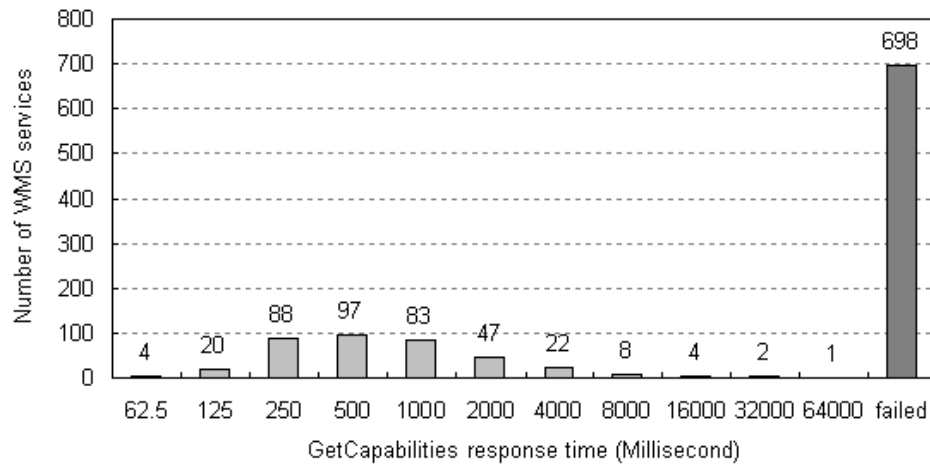
The main content area of the browser displays the "CISC Data Discovery" interface. The "CISC" logo is prominently displayed, with the URL "cisc.gmu.edu" below it. To the right of the logo, the text "Data Discovery" is displayed. Below the logo and text, there is a search input field with a placeholder text "Please input search content here, e.g., 'WMS'". A "Search" button is located below the input field. Below the "Search" button, there is a link labeled "Advanced Search". At the bottom of the page, there are two links: "About CISC" and "Help".

The Windows taskbar is visible at the bottom of the screen, showing the Start button, several application icons (including Internet Explorer, Google Chrome, and Firefox), and the system tray area which includes a clock showing 2:29 PM on 10/8/2012 and a battery level indicator showing 98%.

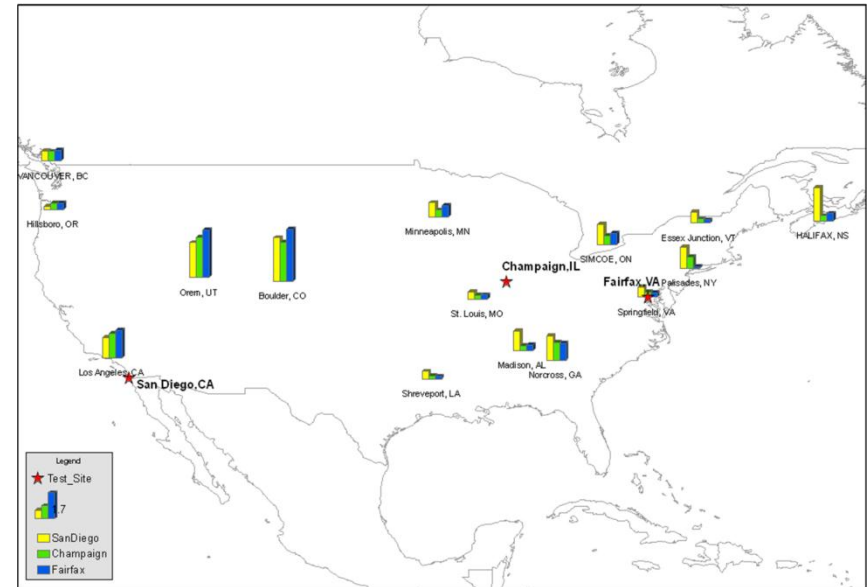
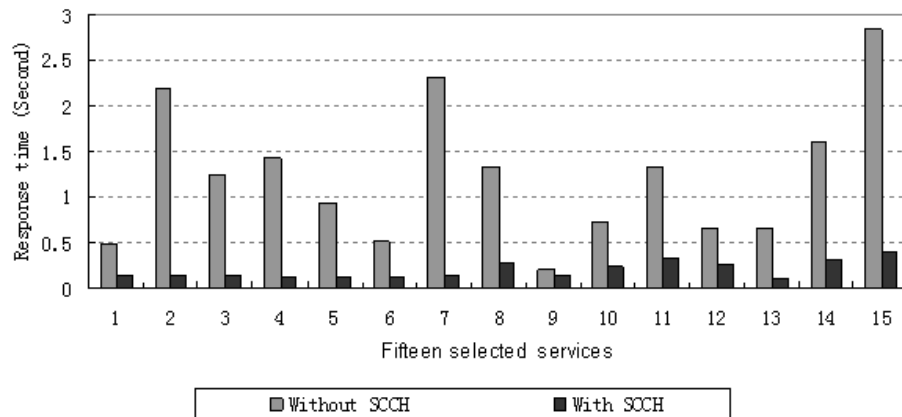


Federate Systems' Performance Monitoring and Utilization

Frequency Distribution of GetCapabilities response time



Response time of fetching capabilities information with and without SCCH



Li et al., 2011



Federate Systems' Performance Monitoring and Utilization

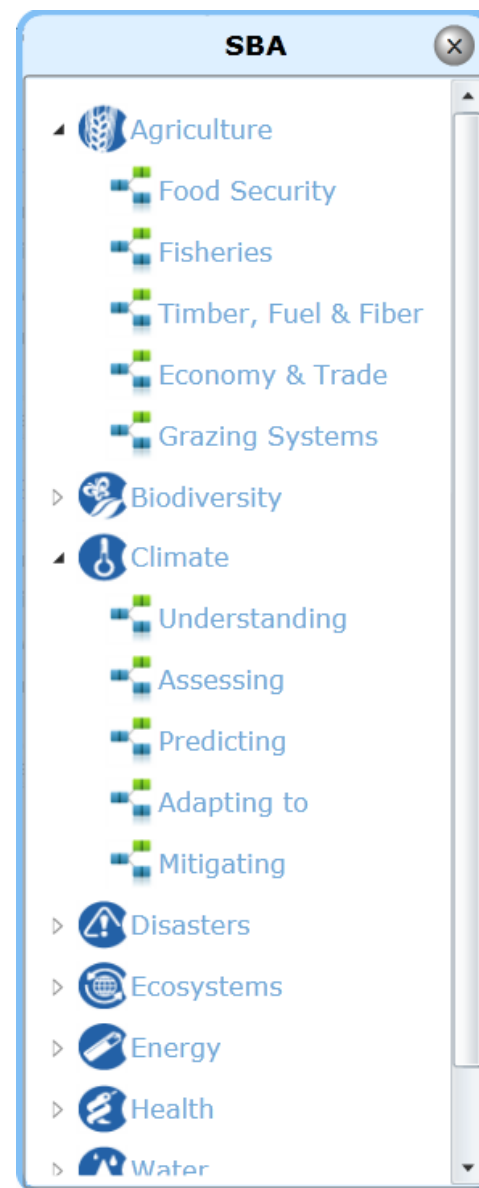
Demo: <http://geosearch.cloudapp.net/>

A screenshot of a web browser window displaying the GeoSearch application. The browser's address bar shows the URL <http://geosearch.cloudapp.net/>. The application interface includes the CISC logo (cisc.gmu.edu) and the title "Data Discovery". A search input field contains the placeholder text "Please input search content here, e.g., 'WMS'". Below the input field is a "Search" button and a link for "Advanced Search". At the bottom of the page, there are links for "About CISC" and "Help". The Windows taskbar at the bottom shows the system clock as 2:52 PM on 10/8/2012, along with icons for various applications and a battery level indicator at 98%.



Categorization of Earth Observation Data

Different Societal Benefit Areas
Different Performance Results
Different Source
Different Services
Etc.





Categorization of Searching Results

Demo using Pivot viewer: <http://geosearch.cloudapp.net/>

A screenshot of a web browser window displaying the GeoSearch application. The browser's address bar shows the URL "http://geosearch.cloudapp.net/". The application interface has a light blue header bar with the text "GeoSearch" and a close button. Below the header, the main content area is white and features the "CISC Data Discovery" logo, which includes the text "cisc.gmu.edu". A large, empty rectangular search input field is centered on the page. Below the input field is a "Search" button, and further down is a link for "Advanced Search". At the bottom of the page, there are links for "About CISC" and "Help". The browser's taskbar at the bottom shows various application icons, including Internet Explorer, Google Chrome, and Firefox, along with system icons for network, volume, and battery status. The system clock in the bottom right corner indicates the time is 2:24 PM on 10/8/2012.



Data Exploration

<http://geosearch.cloudapp.net/>

A screenshot of a web browser displaying the GeoSearch application. The browser's address bar shows the URL "http://geosearch.cloudapp.net/". The application's header includes the "CISC" logo with the URL "cisc.gmu.edu" and the title "Data Discovery". Below the header is a large search input field with a cursor. A "Search" button is positioned below the input field, with a link for "Advanced Search" underneath it. At the bottom of the page, there are links for "About CISC" and "Help". The Windows taskbar at the bottom shows the time as 6:59 PM on 10/8/2012, along with system icons for network, volume, and battery (98%).

http://geosearch.cloudapp.net/

File Edit View Favorites Tools Help

CISC
cisc.gmu.edu

Data Discovery

Search

[Advanced Search](#)

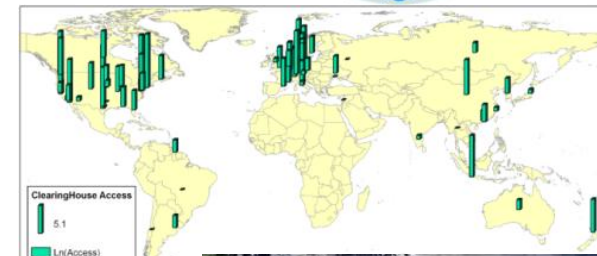
[About CISC](#) [Help](#)

6:59 PM
10/8/2012



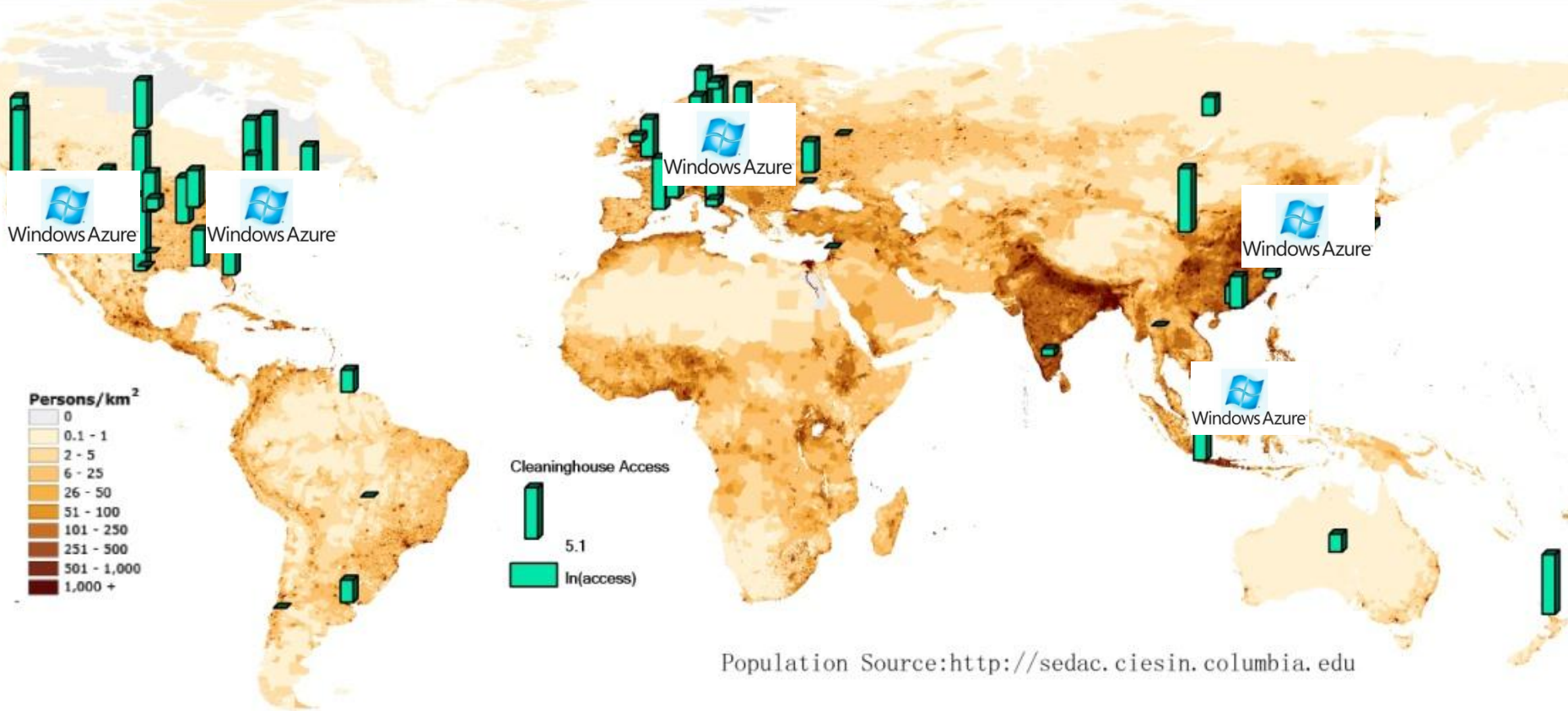
MS Advanced Computing Technologies

1. Cloud Computing (Azure) Responds to Spike Massive Concurrent End Users
2. Cloud DB (SQLAzure) Manages Millions to Billions of Metadata Records
3. WebGIS & BingMaps
4. Silverlight
5. Pivot Viewer





Next Step – Concurrent Intensive



However, the user distribution is dynamic, we need spatiotemporal index to enable global operation (spatiotemporal distribution)



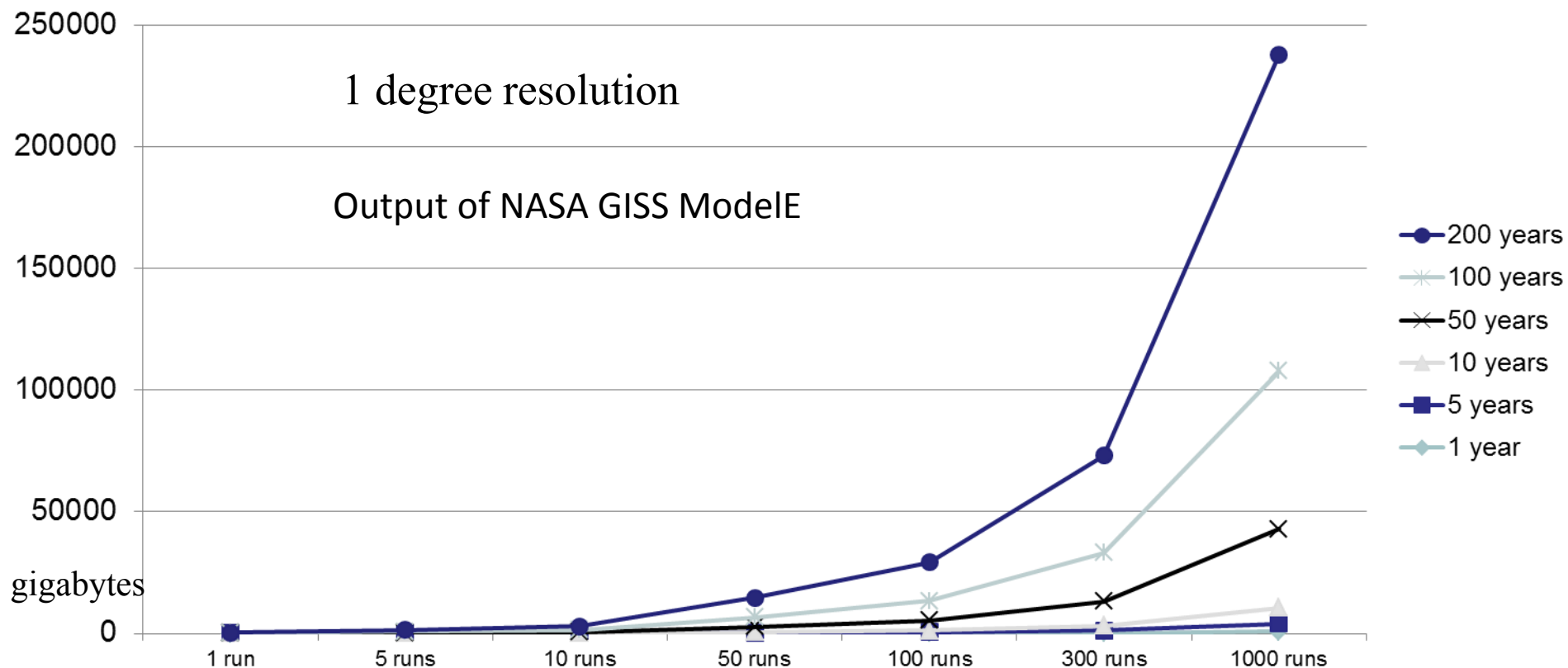
OUTLINE

- GEOSS Clearinghouse
 - Background
 - Semantics and Results Sorting
 - Performance within a Federated Environment
 - Categorization
 - Data Exploration
- **Climate Studies and Big Data Analyses**
 - Project Background and the Big Data Dilemma
 - Online Analyses Requirements
 - Spatiotemporal Statistics Challenge
- One Future Effort



Climate Studies - Big Data

- Hundreds of parameters
- Different granules: from daily, weekly, monthly, seasonal to yearly, or decadal
- Different resolution: global degree to regional 10KM (as high as possible)?
- Different combinations of model runs, thousands to millions and more





Data Management and Analyses Online

<http://199.26.254.190:8088/ClimateAtHome/cah.html>

A screenshot of a web browser displaying the Climate@Home website. The browser window has a title bar with "Firefox" and a tab labeled "Welcome | Climate@Home". The address bar shows the URL "199.26.254.190/climate@home/". The website header features the "climate@home" logo and the tagline "Citizen Science for Climate Change". A navigation menu includes links for "HOME", "OBJECTIVES", "CALENDAR", "RESOURCES", and "LOGIN". Below the menu, there are three image thumbnails: "Earth Observatory", "NASA", and "Earth Observatory". The main content area has a "Welcome" section with text about NASA's initiative. A sidebar on the right shows "Live Twitter Updates" with a tweet from "Climate_at_Home" and social media icons for Facebook and Twitter.

2D/3D Visualization

Multi-D Visualization

Statistics&Analysis

Pivot Viewer

Resources

Select the number of map windows for comparison

4 Windows

Select a map window to load image

Window 1

Window 2

Window 3

Window 4

Move to next window ☒

Select a scenario

E4M20a_000029

E4M20a_000114

E4M20a_000083

E4M20a_000125

E4M20a_000055

E4M20a_000009

E4M20a_000221

E4M20a_000149

E4M20a_000010

E4M20a_000295

Select an variable

prec

trnf_toa

srnf_toa

tsurf

net_rad_planet

cldtpp_isccp

icij_rsi

icij_usi

icij_vsi

Select a date

1956-12-25

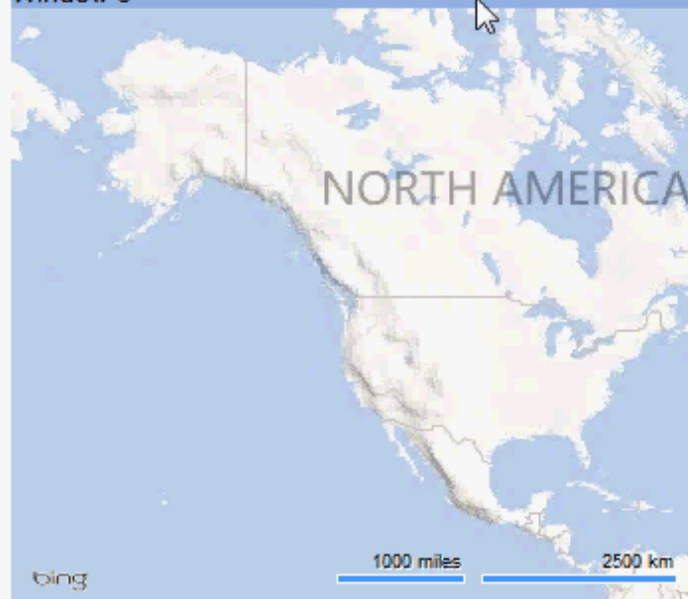
Window 1



Window 2



Window 3



Window 4



Select the number of map windows for comparison

2 Windows

Select a map window to load image

Window 1

Window 2

Move to next window ☒

Select a scenario

E4M20a_000030

E4M20a_000163

E4M20a_000233

E4M20a_000082

E4M20a_000036

E4M20a_000028

E4M20a_000022

E4M20a_000044

E4M20a_000084

E4M20a_000067

Select an variable

prec

trnf_toa

srnf_toa

tsurf

net_rad_planet

cldtp_p_isccp

icij_rsi

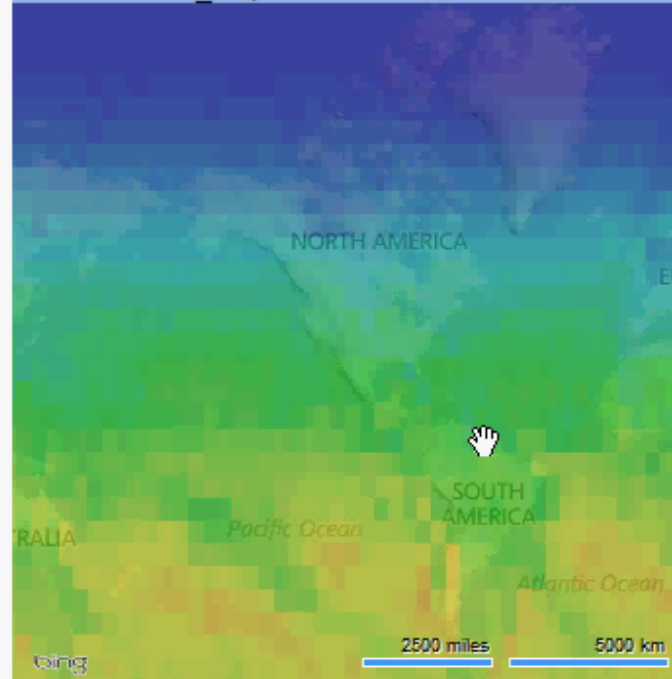
icij_usi

icij_vsi

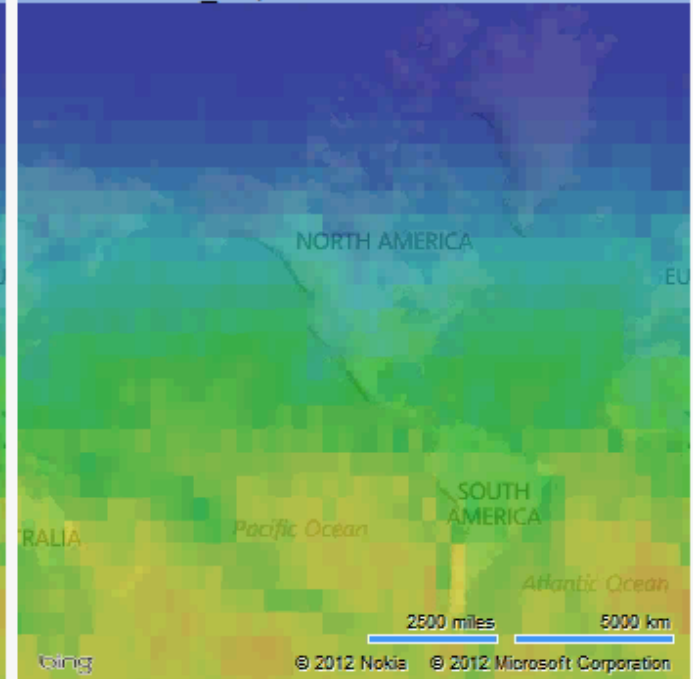
Select a date

1956-04-18

Window 1: Scenario:E4M20a_000030
Parameter:srnf_toa, Time:1956-04-18



Window 2: Scenario:E4M20a_000133
Parameter:srnf_toa, Time:1956-04-18



Diagnostics

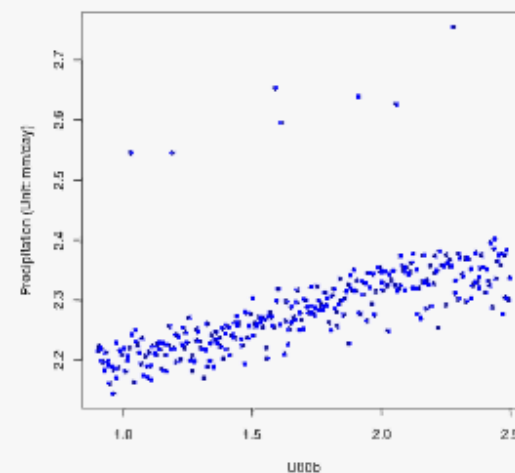
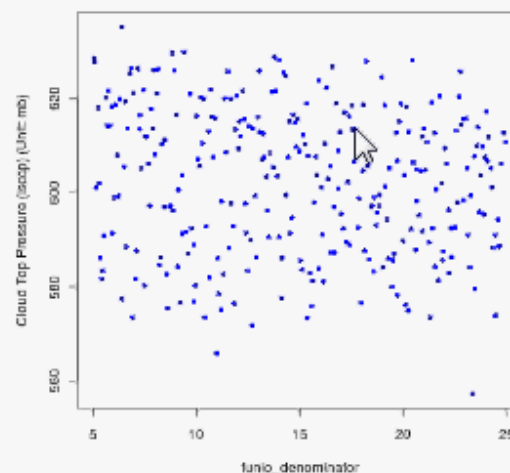
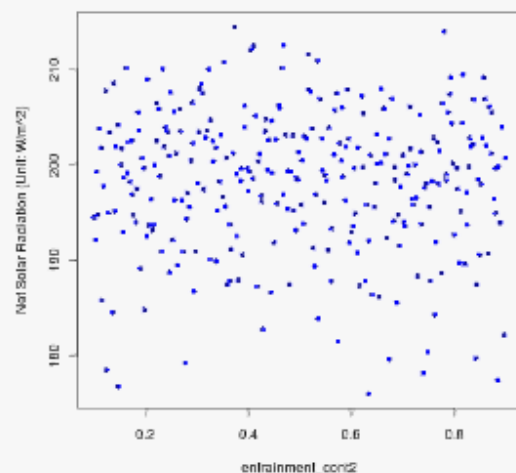
- ☐ net radiation planet
- ☐ net thermal radiation TOA
- ☐ surface air temperature
- ☒ net solar radiation
- ☒ Precipitation
- ☒ Cloud top pressure

Input Parameters

- ☒ funio_denominator
- ☐ autoconv_multiplier
- ☒ radius_multiplier
- ☐ entrainment_cont1
- ☒ entrainment_cont2
- ☐ U00a
- ☒ U00b

check alluncheck allreverseSubmit

Compare results

Select Plot1 e2_srnf_toaSelect Plot2 f_cldtpp_isccpSelect Plot3 Ub_prec

Select the number of map windows for comparison

6 Windows

Select a map window to load image

Window 1

Window 2

Window 3

Window 4

Window 5

Window 6

Move to next window ☒

Select a scenario

E4M20a_000030

E4M20a_000163

E4M20a_000233

E4M20a_000082

E4M20a_000036

E4M20a_000028

E4M20a_000022

E4M20a_000044

E4M20a_000084

E4M20a_000067

Select an variable

prec

trnf_toa

srnf_toa

tsurf

net_rad_planet

cldtp_iscp

icij_rsi

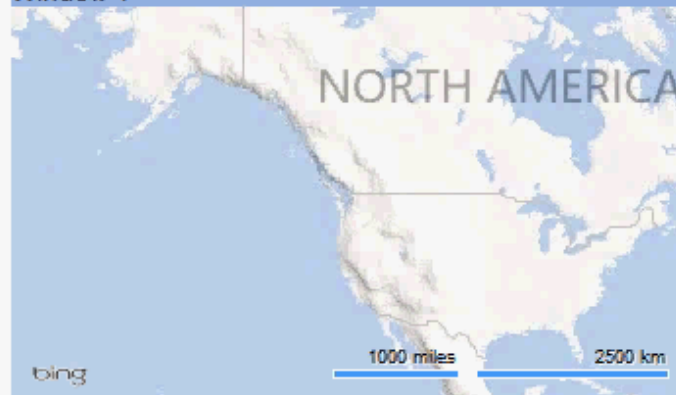
icij_usi

icij_vsi

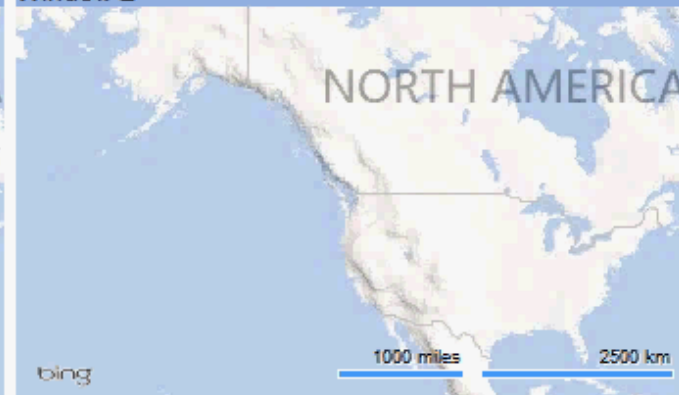
Select a date

1956-04-18

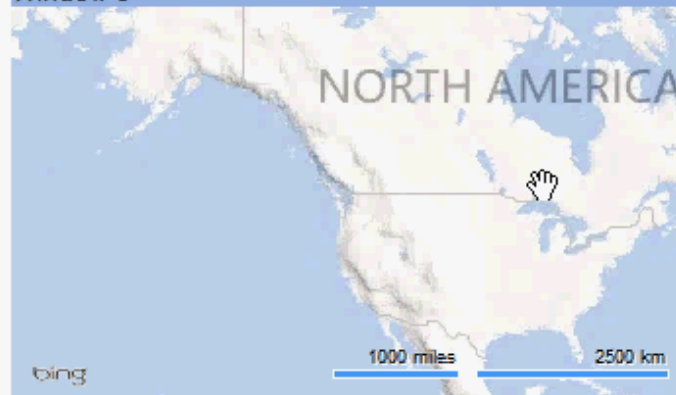
Window 1



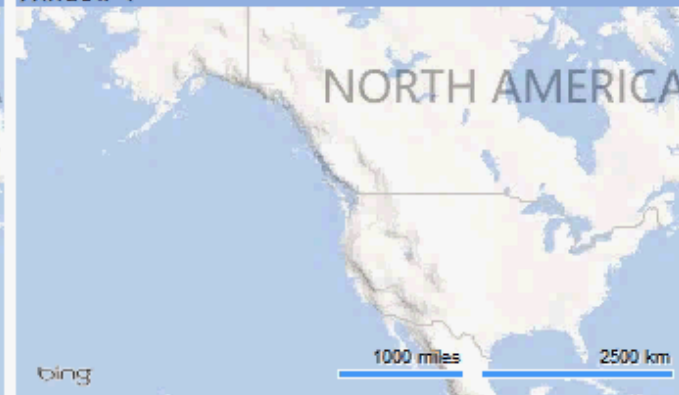
Window 2



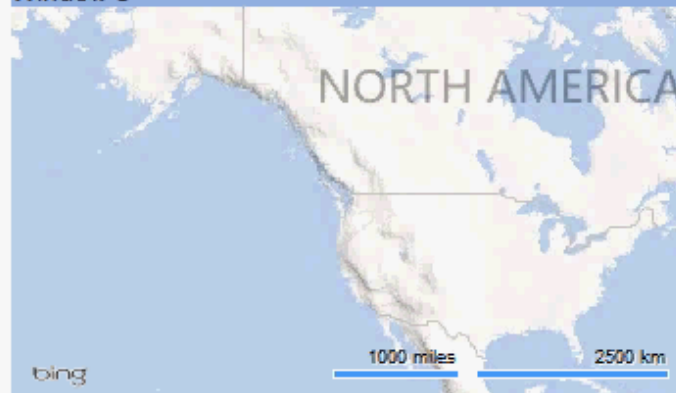
Window 3



Window 4



Window 5



Window 6





Future: Spatiotemporal Statistics

Validate the results at different spatial and temporal scales

- Spatial accuracy
- Temporal accuracy
- Sensitive to events, such as earthquake and volcanic eruptions
- How carbon/methane/etc. emissions impact spatiotemporally

Computational challenges

- New statistical methodologies
- Big data
- Computing intensive
- Easy to use GUI



OUTLINE

- GEOSS Clearinghouse
 - Background
 - Semantics and Results Sorting
 - Performance within a Federated Environment
 - Categorization
 - Data Exploration
- Climate Studies and Big Data Analyses
 - Project Background and the Big Data Dilemma
 - Online Analyses Requirements
 - Spatiotemporal Statistics Challenge
- **One Future Effort**



Future Activities?

NSF

UNIV MGT (VP, Dean, and

UNIVERSITY PERSONNEL

swp.gmu.edu/sttc/?q=node/10

Dashboard Content Structure Appearance People Modules Configuration Reports Help

OVERALL

STC-12-00-spatiotemporal-studies Evaluate

CROSSCUTTING TECHNOLOGY

GMU-12-01-spatiotemporal-index Evaluate

GMU-12-02-5d-visualization Evaluate

GMU-12-03-eod-data-discover Evaluate

GMU-12-07-sec-middleware Evaluate

UCSB-ec Visual analytics solutions for differential rate spatio-temporal dynamics Evaluate

02-Exploratory Interaction With Heterogeneous Geo-Data Evaluate

ENVIRONMENTAL SCIENCES

hydrology driven by remote sensing Evaluate

discover-model-framework Evaluate

st-carbon-tracking Evaluate

ations Evaluate

ENCES

social-media Evaluate

collaborate on projects

STC

10/17/2011

NSF

GEORGE MASON UNIVERSITY

HARVARD

UNIVERSITY OF CALIFORNIA



How are we addressing the issues?

1. Observing and identifying the **spatiotemporal principles/patterns** within and used to address the challenges in all other center projects
2. Producing new spatiotemporal products in collaboration amongst our members to **improve industry efficiency and competitiveness**
3. Hosting one to two training programs/special list meeting each year to discuss and share **spatiotemporal problems and knowledge**
4. Advancing human knowledge and intelligence by combining spatiotemporal principles and computational thinking to **form spatiotemporal thinking** as a new and innovative thinking process
5. Assessing the applicability of the methodologies to enable science discoveries and engineering designs, e.g., **next generation computing**
6. Organizing one to two science advisory committee meetings each year to refine knowledge and adjust/define **research agenda**



Make a Difference

With your participation

Together, we can make a difference and build the spatiotemporal scientific and engineering infrastructure to enable future science discovery and problem solving.

<http://swp.gmu.edu/sttc/>



Acknowledgements

- STC Partners: UCSB and Harvard.
- The GMU STC Team: Zhipeng Gui, Qunying Huang, Yunfeng Jiang, Jing Li, Kai Liu , Peter L. Lostritto, Abdelmounaam Rezgui, Min Sun, Jizhe Xia, Chen Xu
- Sponsors: NSF, NASA, USGS, Interior, EPA, NPS, Microsoft, Intergraph, Northrup Grumm, NGCC, NASG, and many others



Sponsors and Collaborators





References

1. **Yang C.**, Wong D., Yang R., and Kafatos M., Li Q., 2005. [Performance improving techniques in WebGIS](#), *International Journal of Geographical Information Science*, 19(3): 319-342.
2. **Yang P.**, Evans J., Cole M., Alameh N., Marley S., and Bambacus M., 2007. [The emerging concepts and applications of the spatial web portal](#), *Photogrammetry & Remote Sensing*, 73(6):691-698.
3. **Yang C.**, Cao Y., Evans J., 2007. [WMS performance and client design principles](#), *The Journal of GeoInformationScience and Remote Sensing*, 44 (4):320-333.
4. **Yang C.**, Li W., Xie J., and Zhou B., 2008. [Distributed geospatial information processing: sharing earth science information to support Digital Earth](#), *International Journal of Digital Earth*, 1(3): 259-278.
5. **Yang C.**, Raskin R., 2009. [Introduction to Distributed geographic information processing research](#), *International Journal of Geographic Information Science*, 23(5): 553-560.
6. **Yang C.**, Raskin R., Goodchild M.F., Gahegan M., 2010, [Geospatial Cyberinfrastructure: Past, Present and Future](#), *Computers, Environment, and Urban Systems*, 34(4):264-277.
7. **Li Z.**, **Yang C.**, Wu H., Li W., Miao L., 2011, [An optimized framework for OGC web service seamlessly integration to support Geospatial Sciences](#), *International Journal of Geographic Information Science*, 25(4): 595-613.
8. Wu H., **Li Z.**, Zhang H., **Yang C.**, 2011. [Quality of Service control for Web Map Service, Computer & Geoscience](#), 37(4): 485-494.
9. **Yang C.**, Wu H., **Li Z.**, Huang Q., Li J., 2011, Utilizing Spatial Principles to Optimize Distributed Computing for Enabling Physical Science Discoveries, *Proceedings of National Academy of Sciences*, 108(14): 5498-5503.
10. **Liu K.**, **Yang C.**, Xia J., Gui Z., 2012. Using ontology and similarity calculations to rank Earth science data searching results, *International Journal of Geospatial Information Applications*. (in press)