Support Big Environmental Data Sharing and Analyses with Advanced Computing Technologies

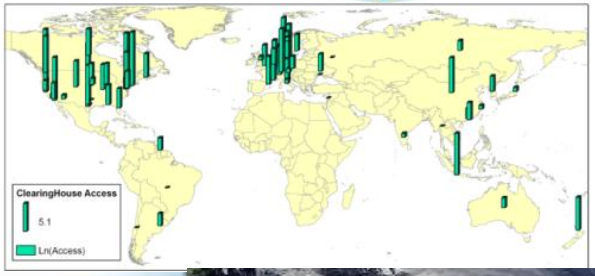
Chaowei Yang, Co-Director
Center of Intelligent Spatial Computing for Water/Energy Science
George Mason University

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

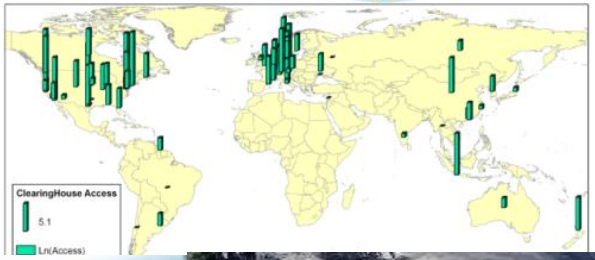
GEOSS Clearinghouse Objective

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**

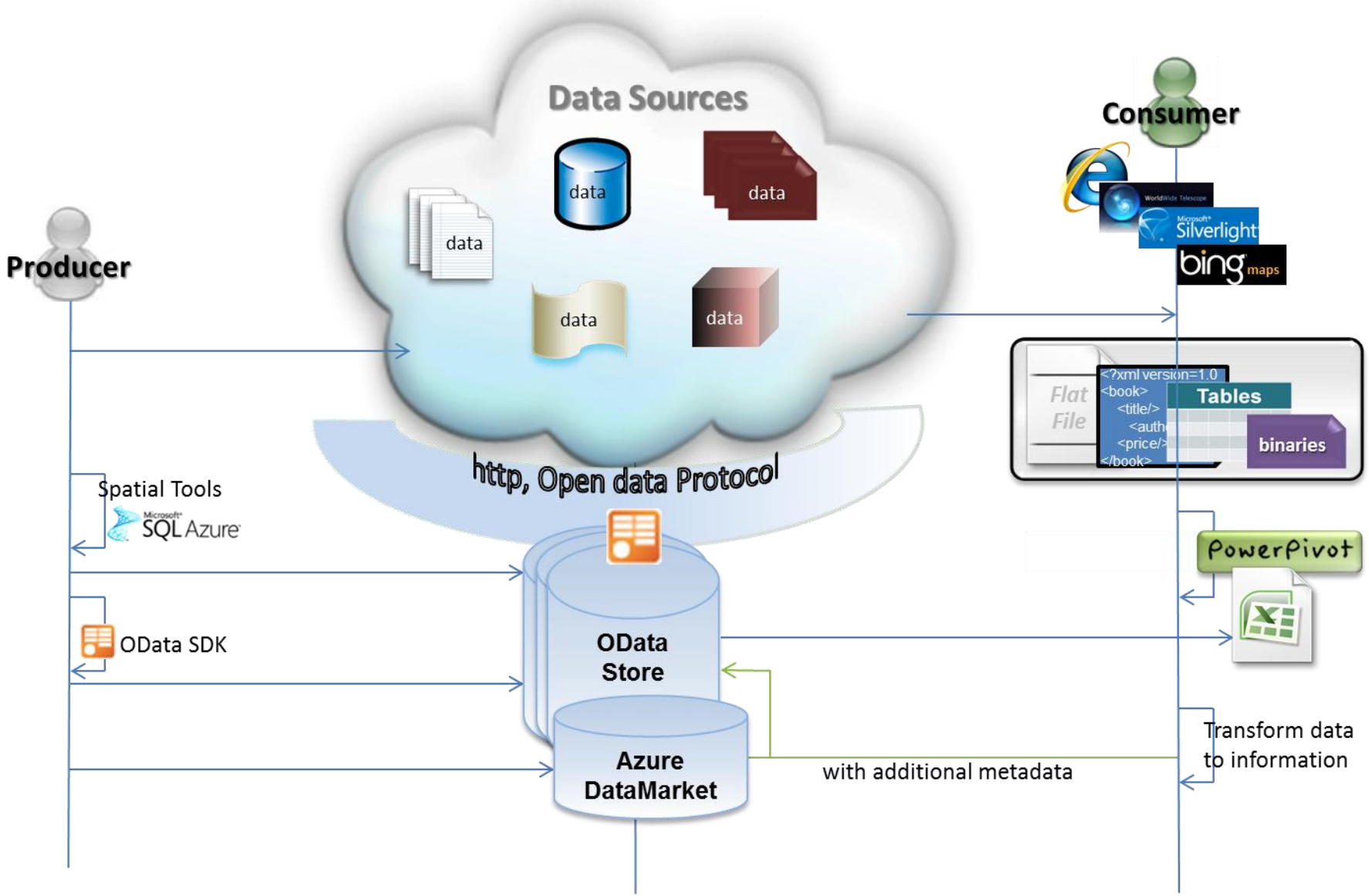


GEOSS Clearinghouse Challenges

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



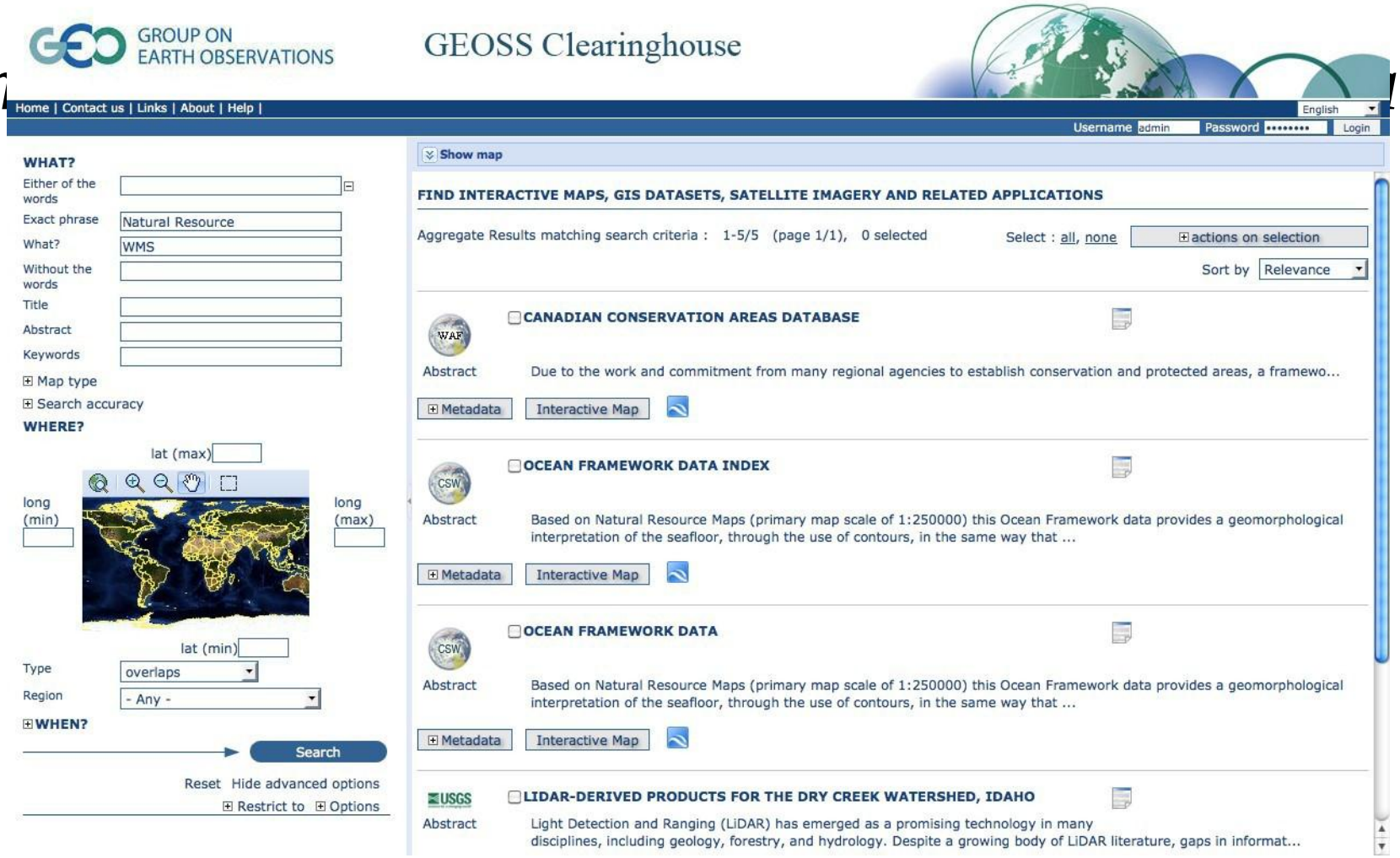
Environmental Informatics Framework



Vocabularies and Semantics

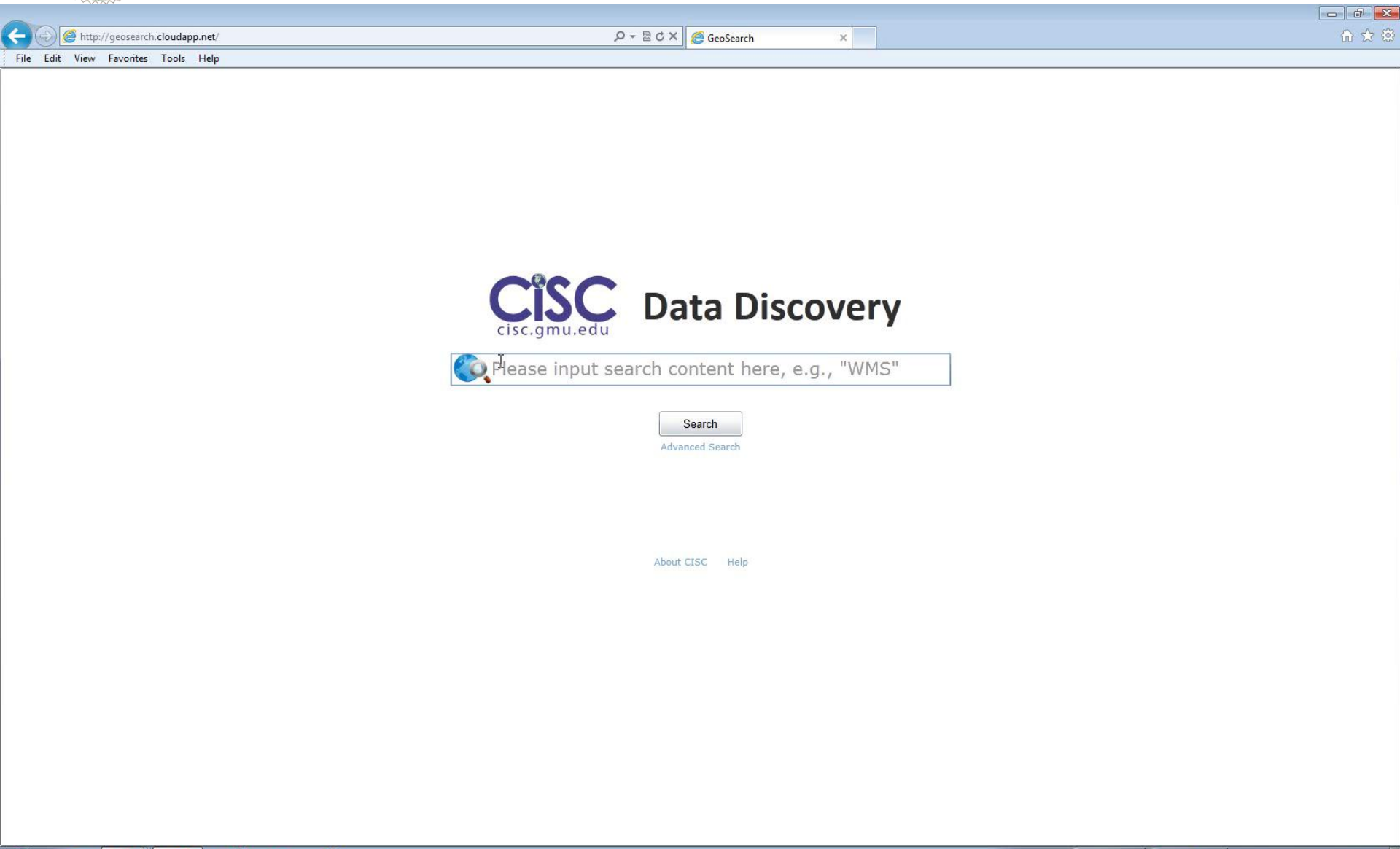
- Dynamic Distributed Search for Federated Catalogs
  - Rank Results

Sim *(p, q)*

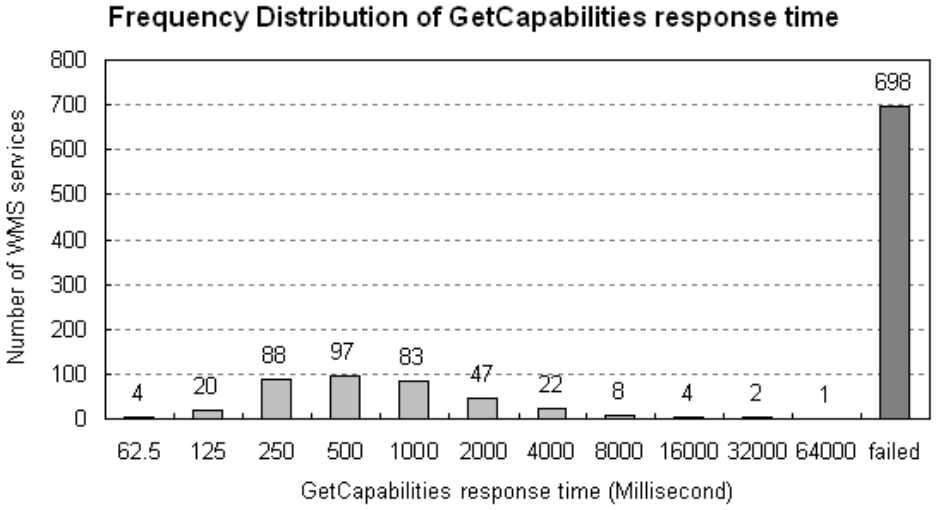


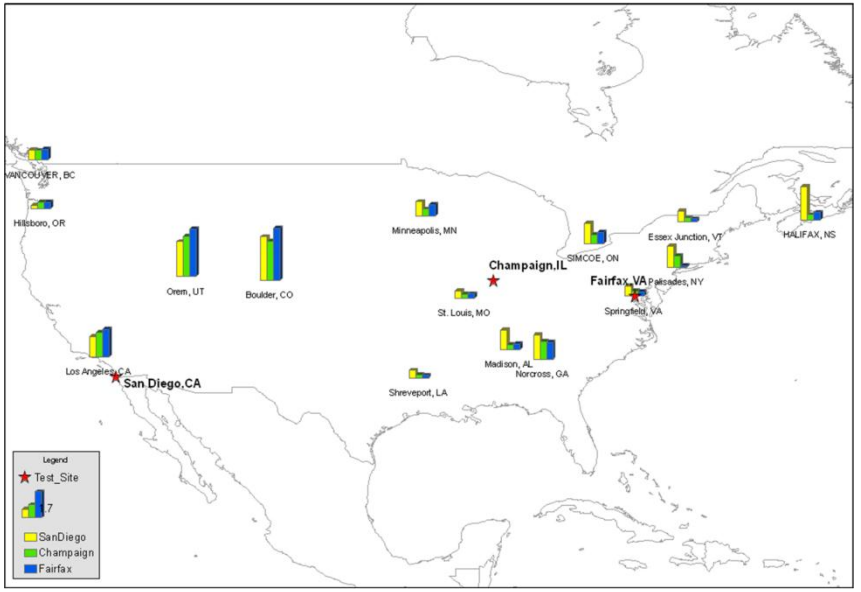
Relevance Demo

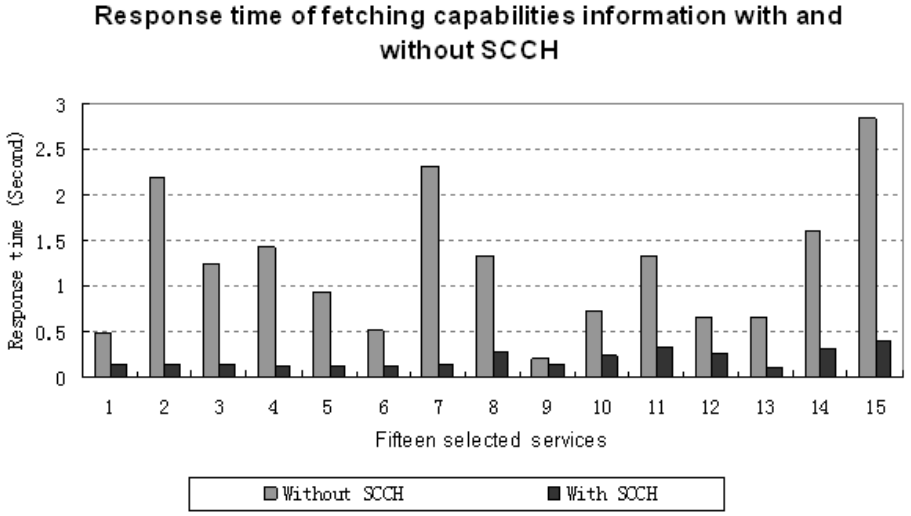
http://geosearch.cloudapp.net/



Federate Systems' Performance Monitoring and Utilization



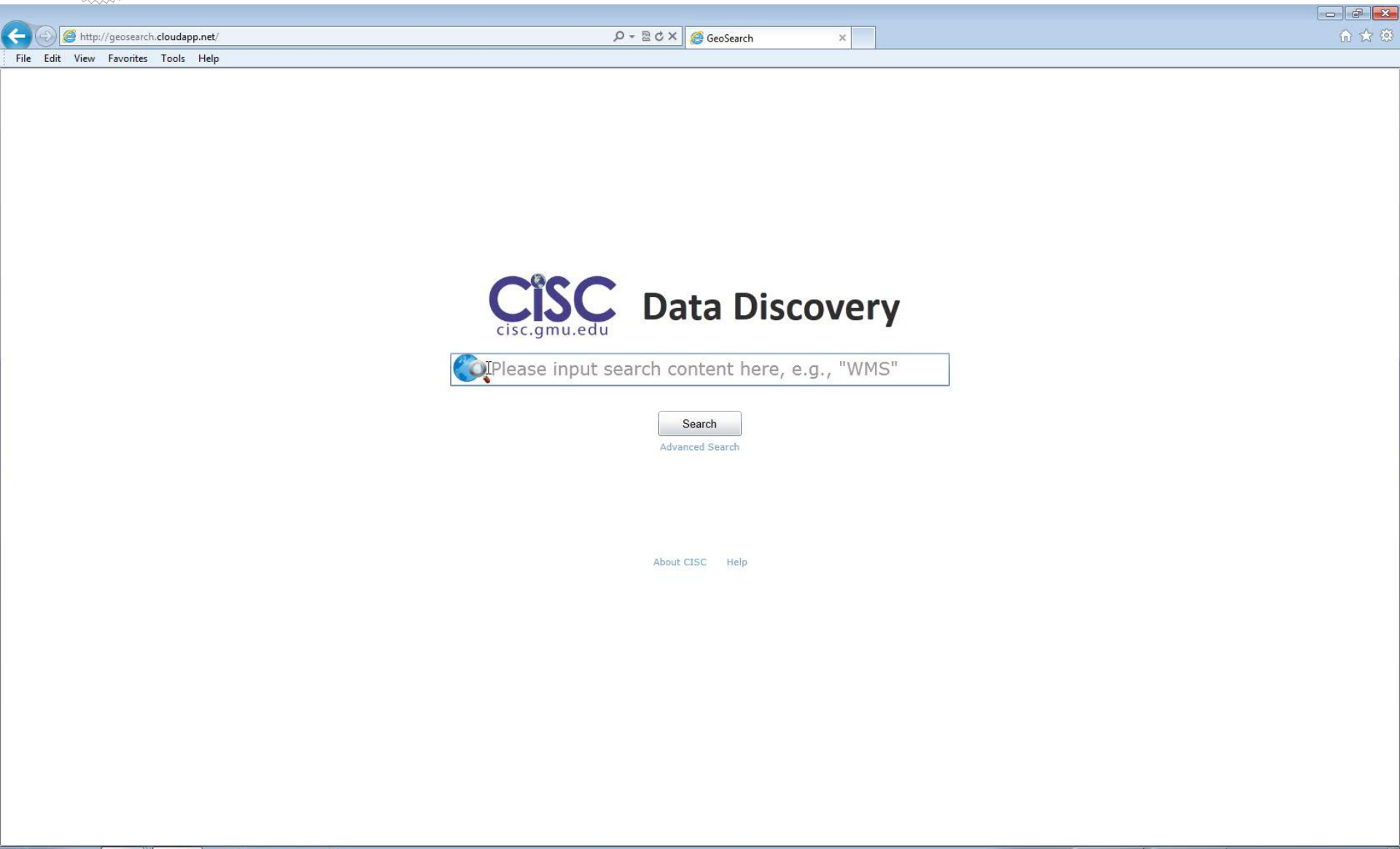




Li et al., 2011

Federate Systems' Performance Monitoring and Utilization

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

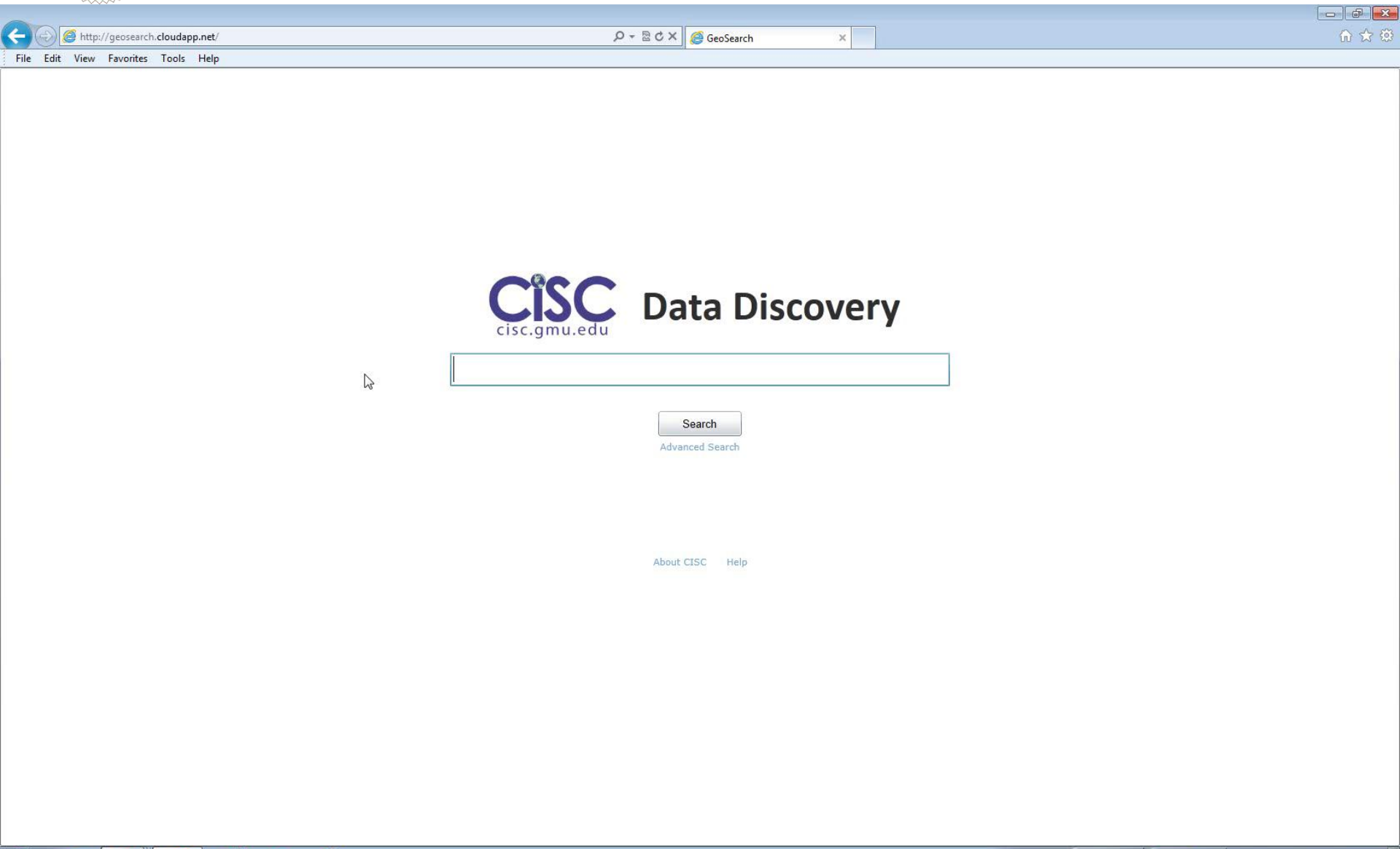


Categorization of Earth Observation Data

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

Categorization of Searching Results

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

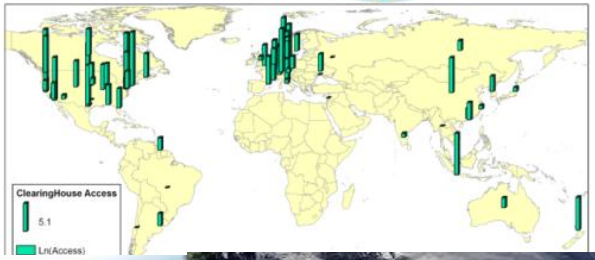


Data Exploration

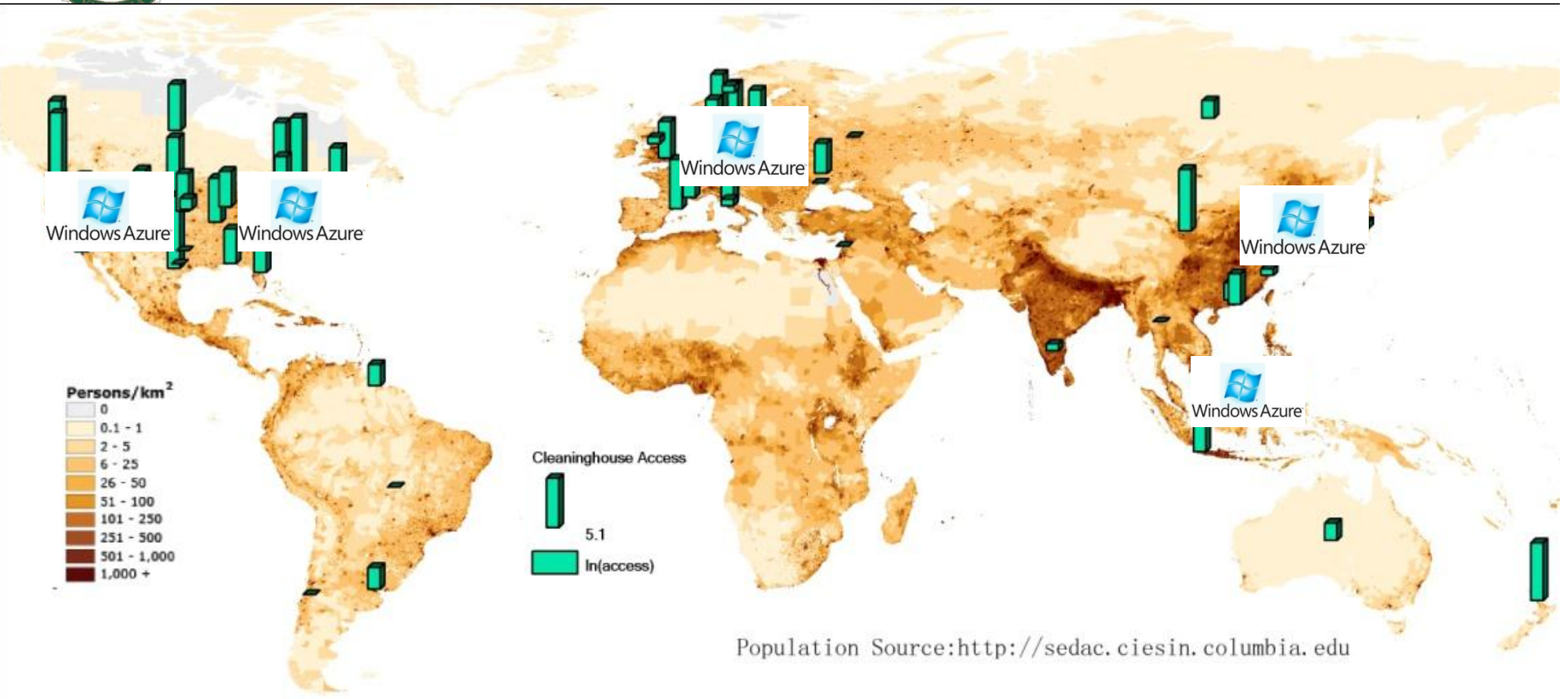
http://geosearch.cloudapp.net/

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



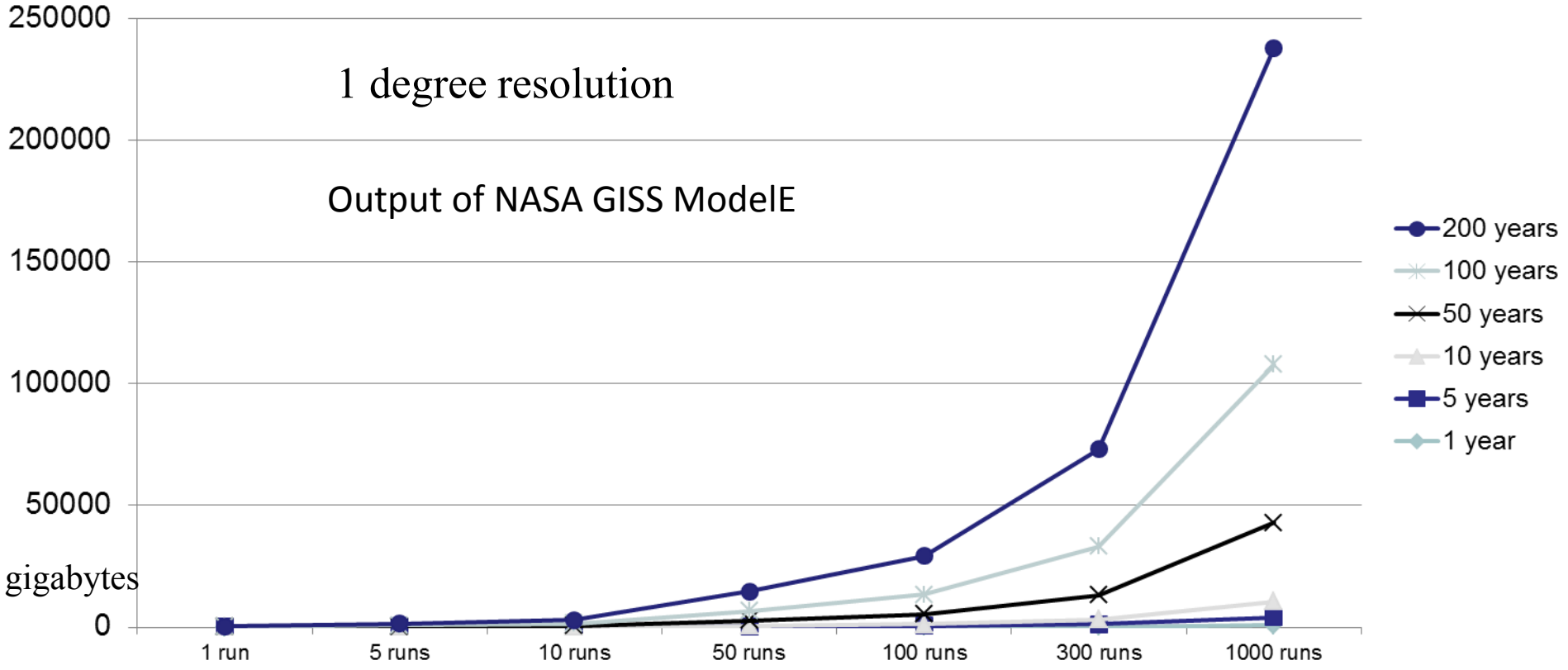
However, the user distribution is dynamic, we need spatiotemporal index to enable global operation (spatiotemporal distribution)

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

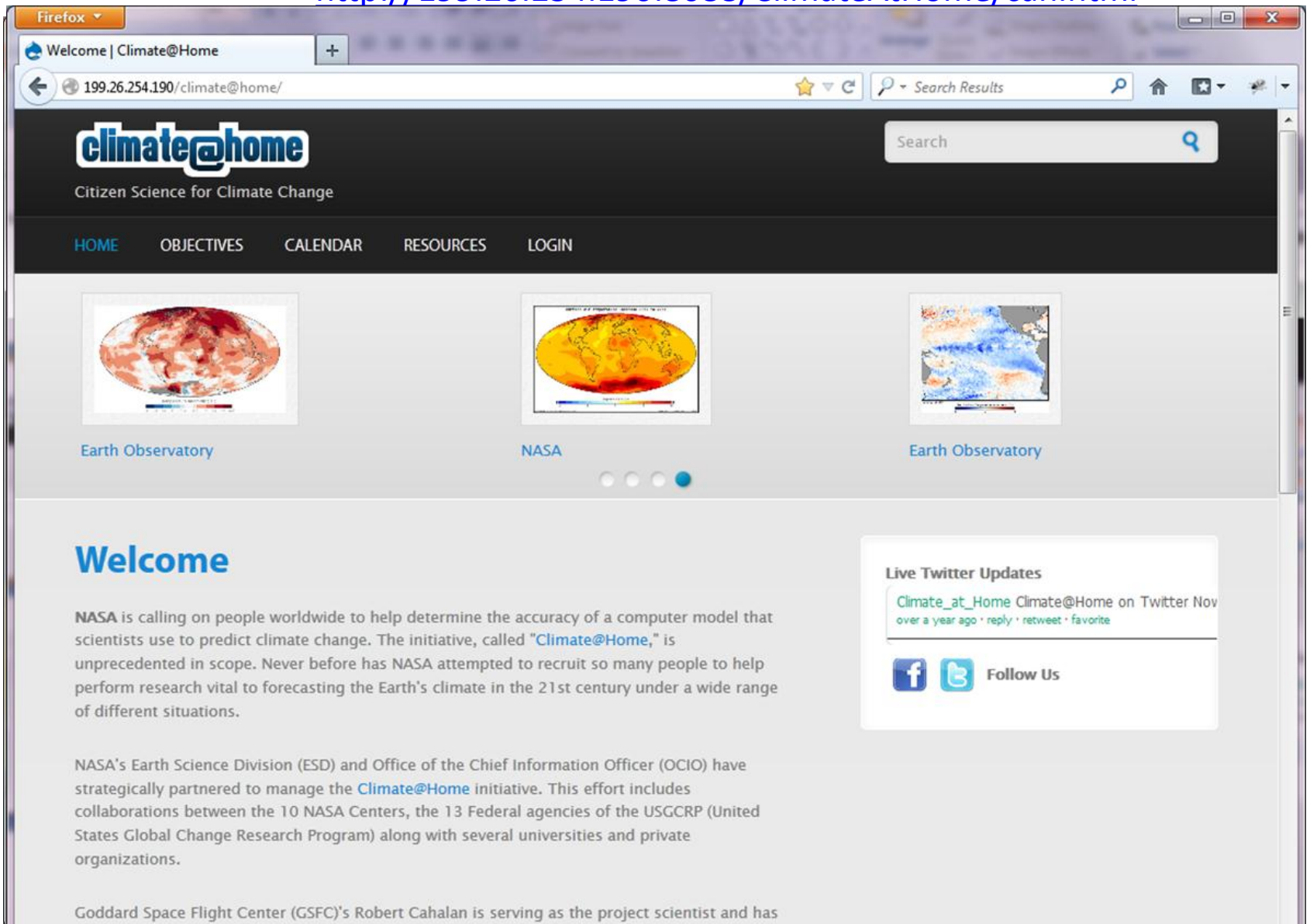
Climate Studies - Big Data

- 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



2D/3D Visualization | Multi-D Visualization | Statistics&Analyis | 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

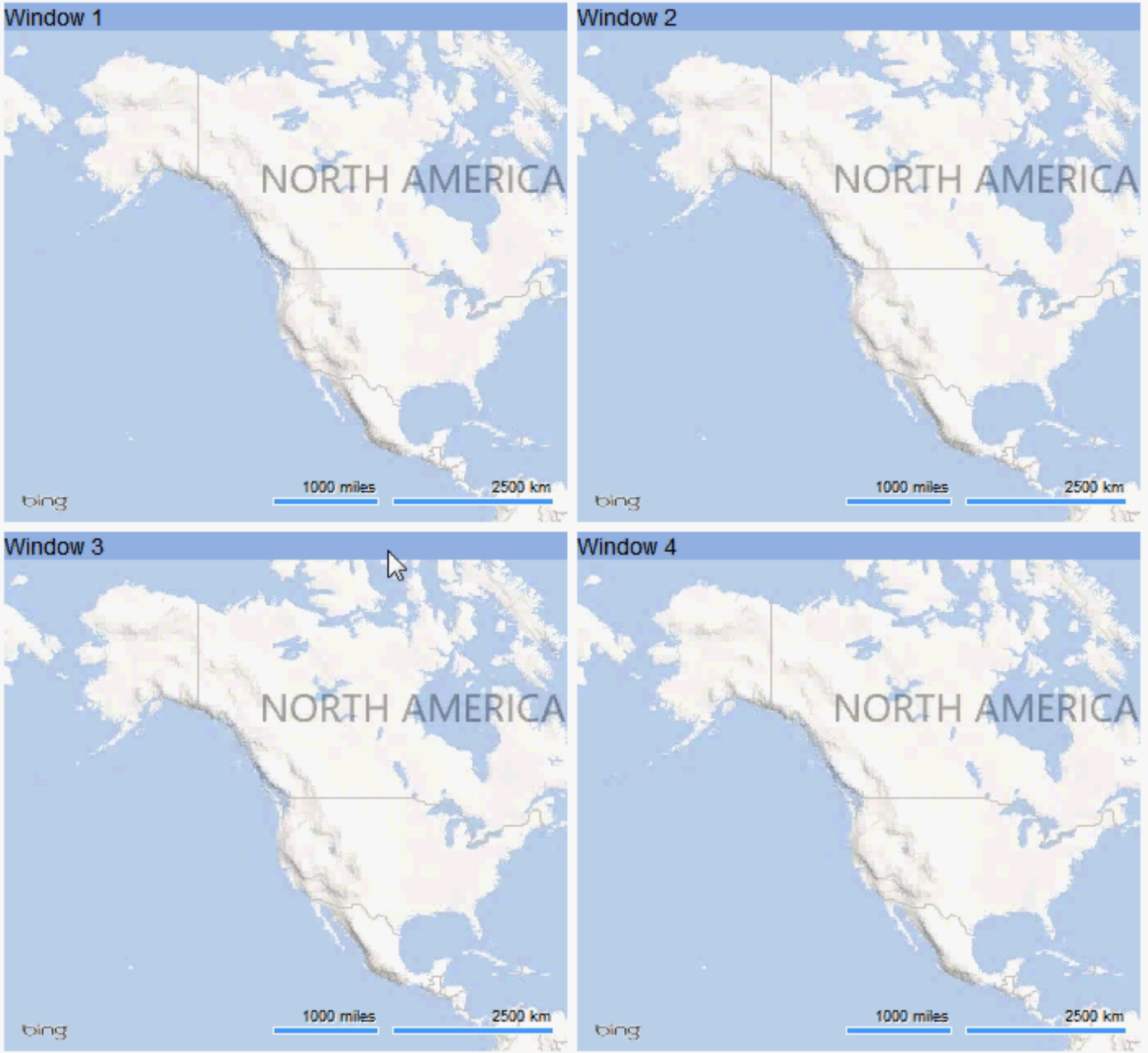
E4M20a_000029
E4M20a_000114
E4M20a_000083
E4M20a_000125
E4M20a_000055
E4M20a_000009
E4M20a_000221
E4M20a_000149
E4M20a_000010
E4M20a_000295

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



2D/3D Visualization | Multi-D Visualization | Statistics&Analyis | Pivot Viewer | Resources

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

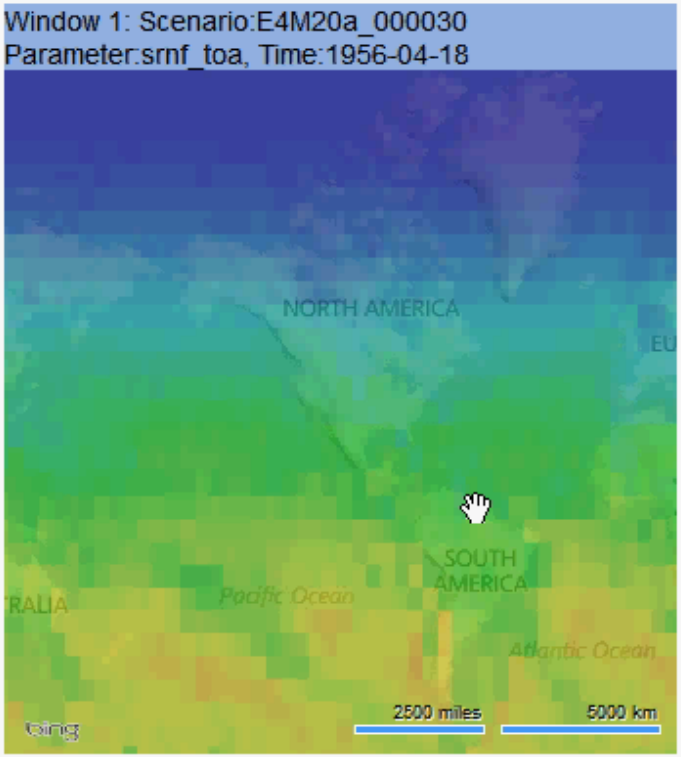
E4M20a_000030
E4M20a_000163
E4M20a_000233
E4M20a_000082
E4M20a_000036
E4M20a_000028
E4M20a_000022
E4M20a_000044
E4M20a_000084
E4M20a_000067

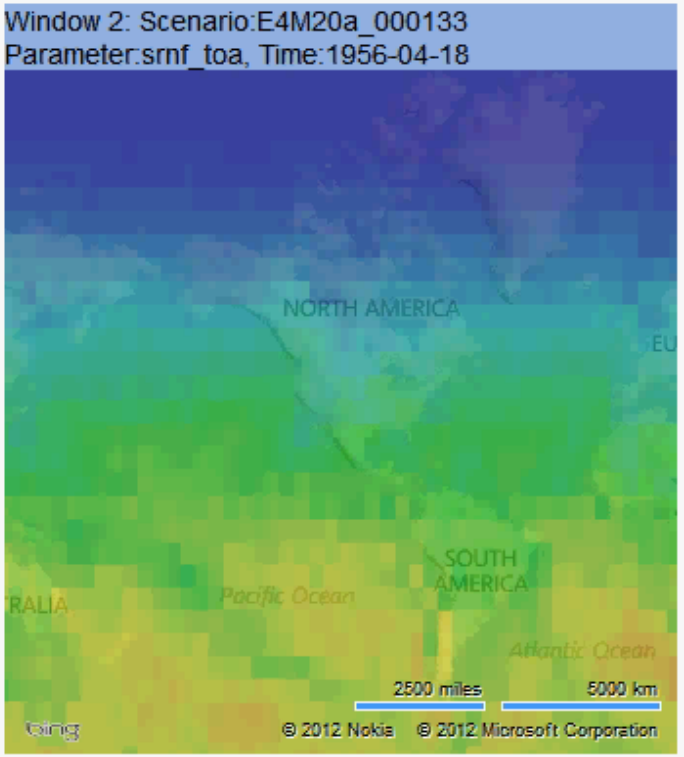
Select an variable

prec
trnf_toa
srnf_toa
tsurf
net_rad_planet
cldtpp_isccp
---
icij_rsi
icij_usi
icij_vsi

Select a date

1956-04-18





2D/3D Visualization | Multi-D Visualization | Statistics&Analyis | Pivot Viewer | Resources

Diagnostics

- [ ] net radiation planet
- [ ] net thermal radiation TOA
- [ ] surface air temperature
- [x] net solar radiation
- [x] Precipitation
- [x] Cloud top pressure

Input Parameters

- [x] funio_denominator
- [ ] autoconv_multiplier
- [x] radius_multiplier
- [ ] entrainment_cont1
- [x] entrainment_cont2
- [ ] U00a
- [x] U00b

check all

uncheck all

reverse

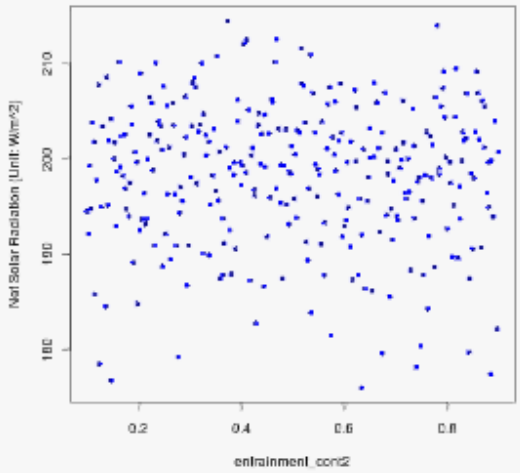
Submit

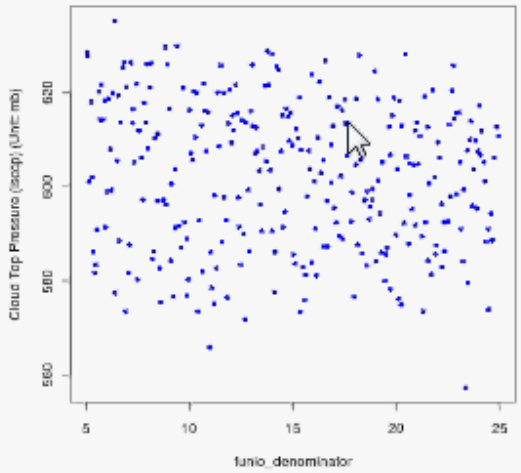
Compare results

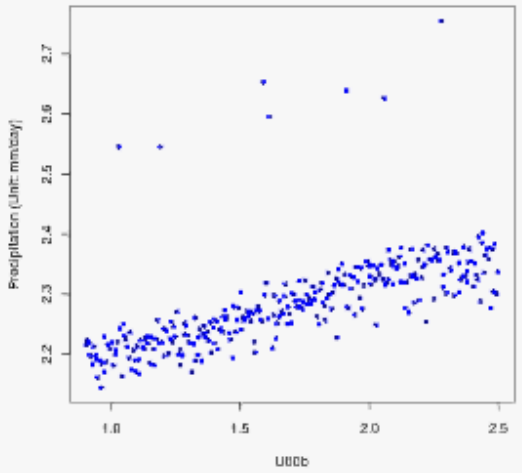
Select Plot1 e2_srnf_toa

Select Plot2 f_cldtpp_isccp

Select Plot3 Ub_prec







2D/3D Visualization | Multi-D Visualization | Statistics&Analyis | Pivot Viewer | Resources

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

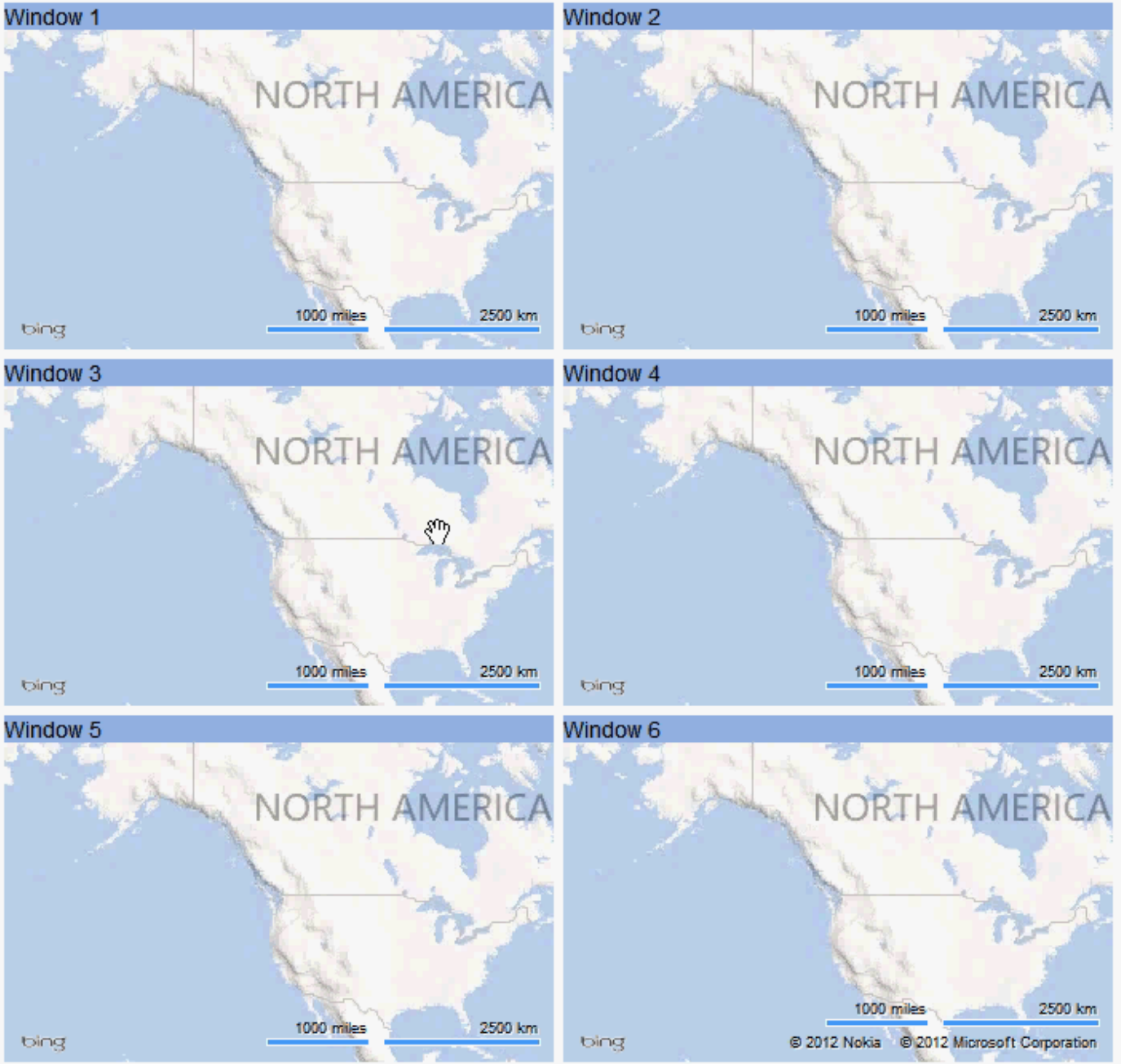
E4M20a_000030
E4M20a_000163
E4M20a_000233
E4M20a_000082
E4M20a_000036
E4M20a_000028
E4M20a_000022
E4M20a_000044
E4M20a_000084
E4M20a_000067

Select an variable

prec
trnf_toa
srnf_toa
tsurf
net_rad_planet
cldtpp_isccp
---
icij_rsi
icij_usi
icij_vsi

Select a date

1956-04-18



Future: Spatiotemporal Statistics

Validate the results at different spatial and temporal scales

- Spatial accuracy
- Temporal accuracy
- Sensitive to events, such as earthquake and vocalic 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?

Collaborate on projects

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/

Acknoledgements

- 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)