Review of Geospatial Data Systems' Support of Global Change Studies

Yuqi Bai[1] and Liping Di[2*]

[1]*Ministry of Education Key Laboratory for Earth System Modeling, and Center for Earth System Science, Tsinghua University, Beijing, China, 100084*
[2]*Center for Spatial Information Science and Systems, Geroge Mason University, Fairfax, VA, USA 22030.*

Authors' contributions

This work was carried out in collaboration between the two authors. Authors YB and LD jointly performed the review study. Author YB wrote the first draft of the manuscript and author LD revised the manuscript. All authors read and approved the final manuscript.

Review Article

ABSTRACT

Aims: Global change studies need to manipulate large volume of observation and prediction data, most likely from multiple sources. From the researchers' perspective, the whole research process consists of the follow stages: data discovery, data access, data processing, data analysis and result dissemination. The aim of paper is to review the state-of-the-art of geospatial data systems to reveal the way towards a better support of global change studies.
Methodology: This paper reviews the capabilities of exemplar geospatial data systems. It further analyzes the needs of manipulating large volume of diverse data when performing global change studies. By comparing the available capabilities with the real needs, this study shows the strengths and limitations of existing data systems when supporting global change studies.
Results: The analysis shows that data systems are helpful for researchers to fulfill data discovery and access, while most of them do not provide further functionalities to cover other stages in the whole research process. This suggests that a new generation of data systems is highly needed to provide efficient and enough support for scientists to perform global change studies. Instead of simply moving data from sources to researchers' local archives, it will enable more on-line data manipulation functionality and the interoperability of data and systems.

*Corresponding author: Email: ldi@gmu.edu;

Conclusion: Traditional geospatial data systems are designed to operate locally without built-in interoperability and sharing capabilities. Such systems are operated under the paradigm of "everything-locally-owned-and-operated". Conducting global change studies using such a system requires moving a large volume of data from providers' sites to researchers' site. Such a system does not provide strong support for the entire research process. Since climate research requires manipulating a huge volume of complex and diverse multi-source data, a new paradigm of "everything-shared-over-the-Web" is promising when designing a new generation of geospatial data systems, which are standard-based, interoperable, and sharable, for global change studies.

1. INTRODUCTION

The 20th century had experienced widespread global changes due to climatic variations, shifting demographics, land use changes, population migration, and economic development. As the world is being transformed at an unprecedented pace and in uncertain directions, these changes are expected to have a significant impact on the quantity and quality of land, water, air and ecosystem inextricably linked across the globe. For sound environmental assessment, dedicated data centers have been deemed essential to manage large volume of geospatial data, including remotely sensed imageries, to support various types of environmental assessment. For a sound spatial analysis, a salient example that can be noted is that petabytes of images had been collected and archived in several Distributed Active Archive Centers (DAAC) through National Aeronautics and Space Administration (NASA) Earth Observing program in the U.S. [1].

These data centers are providing two basic functionalities to researchers: data archiving and data distribution. Data centers usually provide several levels of capabilities to long-term archives of the data: such as CD-ROMs, Redundant Array of Independent Disks (RAID) hard drives, and tape storage systems. For distribution, data centers usually provide Web-based interfaces, i.e. portals, for users to search and download the data. An off-line order processing procedure may be needed for some data centers or for some data products.

From the researchers' perspective, the environmental assessment process usually consists of five steps: data discovery, data access, data customization, data analysis and results dissemination. Comparing the aforementioned functionalities that data systems could provide with the ones researchers really need during their research process, we can find that most of data systems are only helpful in two of the five steps, and they leave all other steps alone with researchers to fulfill.

This becomes a big problem when performing global change studies. For researchers in this field, downloading, processing, and analyzing terabytes of climate data locally is very time-consuming, and sometimes, even infeasible. For example, researchers at the Center for Earth System Science in Tsinghua University in China submitted the FGOALS-g2 [2] global climate model output data to the Coupled Model Intercomparison Project Phase 5 (CMIP 5) [3], and the data volume is about 40 TB. The data volumes of many climate model output submitted to the CMIP 5 are at this level. Actually, documenting the past behavior of the climate system and detecting changes and their causes require the use of data form

instrumental, paleoclimatic, satellite, and model-based sources. The volume and the complexity of worldwide climate data are expanding rapidly, creating challenges for both physical archiving and sharing, as well as for easy discovery of and access to the needed data [4]. Effectively manipulating petabytes of climate data are essential in facilitating the research to tackle the scientific quest for an understanding of how climate behaved in the past and will behave in the future.

The purpose of this article is to review the state-of-the-art of geospatial data systems to know their capabilities in making data publically available, to show their strengths and limitations when providing sustainable support to the global change studies, and to identify the way towards a new generation of geospatial data systems for global change studies.

The rest of the paper is organized as follows. Section 2 analyzes how Earth observation data are served at the institutional, agency, cross-agency and international levels respectively through four exemplar case studies: AeroStat [5], NASA's Earth Observing System (EOS) Clearinghouse (ECHO) [6], the Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems and Services (WGISS) Integrated Catalogue (CWIC) [7], and the Global Earth Observation System of Systems (GEOSS) [8]. Section 3 concentrates on the analysis of geospatial data systems' capabilities and the researchers' real needs when performing scientific research. A paradigm shift in global change studies is presented in section 4, followed by a case study featuring the new Web-and-service-centric paradigm. The findings are concluded in section 5.

2. EXAMPLAR GEOSAPTIAL DATA SYTEMS

2.1 AeroStat

AeroStat, developed at the NASA Goddard Earth Science Data and Information Services Center (GES DISC), is an online system for the direct statistical intercomparison of global aerosol parameters, in which the provenance and data quality can be easily accessed. It offers statistical analysis, visualization and downloadable products from aerosol data measured by satellites and Aeronet [9], Holben et al. [10] stations.

In particular, researchers are allowed to select one Aeronet ground location, one or more parameters of interest from three satellite aerosol data products and the date range of interest. AeroStat then retrieves spatially and temporally collocated satellite data and the corresponding ground-based observation data, intercompares them, and presents the results as plot(s) or downloadable csv file(s), as shown in Fig. 1.
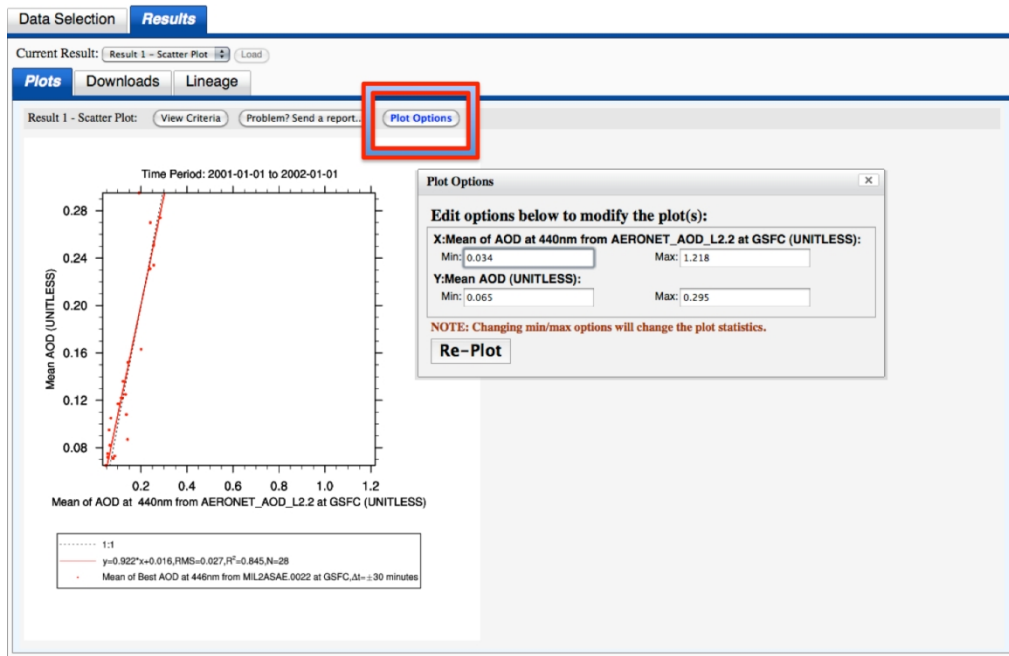
Fig. 1. Plotting service in AeroStat

In AeroStat, it is possible to create a custom data file containing one or more satellite-based aerosol data products of interest, sampled at a specific Aeronet ground location and at specific time duration. It is also possible to display the selected data on a plot, providing for a rapid data exploration and evaluation process.

2.2 NASA ECHO

NASA's Earth Observing System has established several Distributed Active Archive Centers (DAAC) to receive and archive the data collected by the space- and air-born remote sensing devices and field instruments. The differences among these centers are the scientific disciplines that they serve, and accordingly, the data that they receive and serve. To make the archived data publicly available, NASA used to deploy the same data gateway in each of the data centers. Using this gateway, each center was responsible for the generation and maintenance of the descriptive information, i.e. metadata, about archived data, for the data discovery and access.

Since 2003, NASA has changed this approach by introducing a centralized metadata clearinghouse and an order broker system, called ECHO. The reason for this effort is to enable a standardized and integrated data discovery and order for all the NASA data, and to enable more domain-specific client partners. As shown in Fig. 2, the data centers are still responsible for receiving, managing, and archiving Earth science data. They now routinely generate and remotely ingest the metadata into ECHO. On the front end, ECHO provides web service APIs to fulfill data discovery, which further enables different interfaces provided by client partners. For data order requests submitted by users, ECHO forwards them to the corresponding data centers to fulfill.
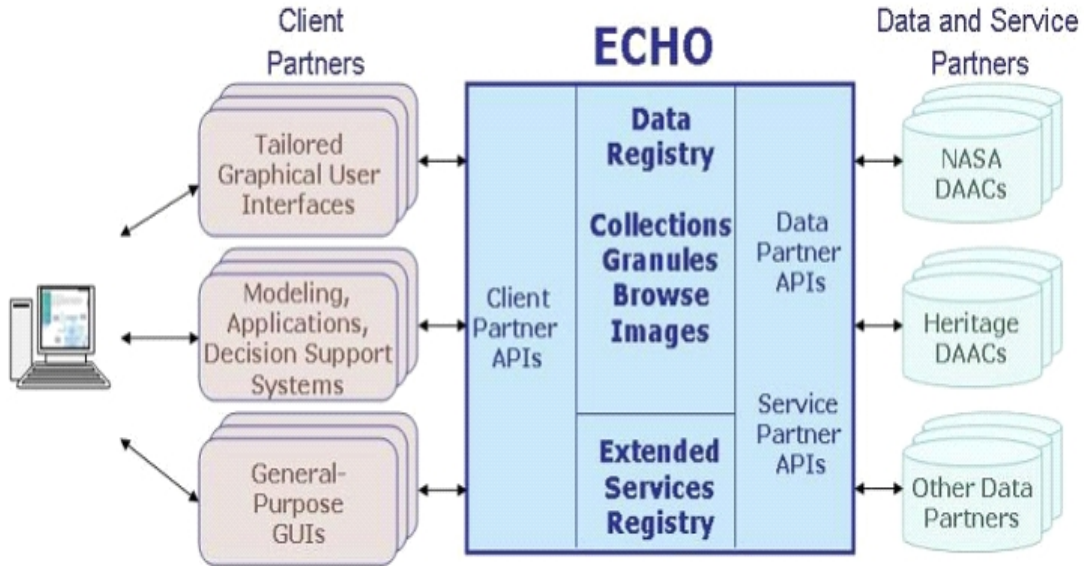
Fig. 2. NASA ECHO

2.3 COES CWIC

The Committee on Earth Observation Satellites (CEOS) coordinates civilian space-borne Earth observations. Its 52 members and associate members cover all major Earth observation agencies in the world, including NASA, National Oceanic and Atmospheric Administration (NOAA), U.S. Geological Survey (USGS), United Kingdom Space Agency, Instituto Nacional de Pesquisas Espaciais (INPE) of Brazil, Japan Aerospace Exploration Agency (JAXA), and the China Center for Resources Satellite Data and Application. Each of these members has their own systems to manage the data and metadata, and make them discoverable and accessible.

A new endeavor in CEOS is to provide a single access point to these members' data catalogues to enable efficient data discovery: researchers do not need to go to individual portals individually. Instead, they can go to this new CEOS-wide interface to search for the data of interest from CEOS member's data catalogues in one step.

Since June 2010, NOAA has funded a pilot project on integrating individual satellite data catalogues for CEOS. NOAA, NASA, USGS, INPE, three data centers in China, and researchers from the Center for Spatial Information Science and Systems (CSISS) at George Mason University (GMU) worked together on the design and implementations of CWIC till June 2011, and released the first version of CWIC at the WGISS-31 meeting. Fig. 3 shows the context diagram and the inner structure of CWIC 1.0 [11].
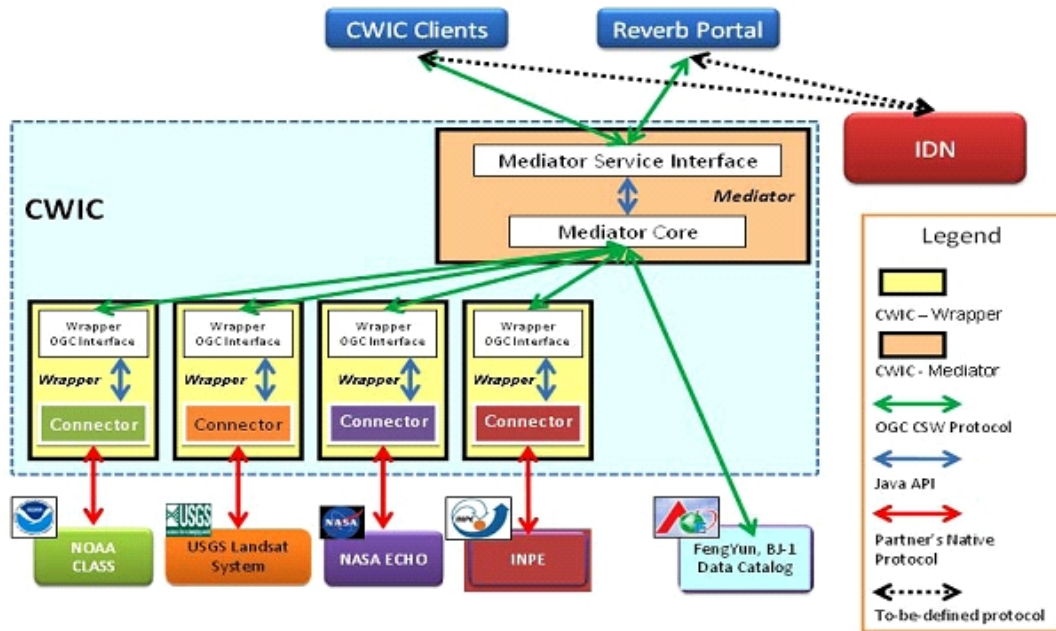
Fig. 3. System architecture of CWIC

Internally, CWIC employs mediator-wrapper architecture, where wrappers translate OGC catalogue protocols to the native protocols that CEOS members' satellite catalogues expose, and mediator provides an integrated access point to all the affiliated CEOS members' satellite catalogues.

CWIC exposes OGC Catalogue Service for Web (CSW) interfaces [12] on the front for client programs to perform distributed search against the affiliated satellite data catalogues. In particular, NASA Reverb [13], the next generation Earth Science and Discovery tool, provides a new instance, identified as Reverb portal in Fig. 3, to demonstrate how CWIC could be accessed through a graphical user interface.

The CEOS International Directory Network (IDN) [14] contains a list of available satellite data products maintained by individual CEOS members. CWIC is going to utilize this framework to fulfill data search at directory level, which was not available in the version 1 release.

2.4 GEOSS

The Group on Earth Observation (GEO) is a voluntary partnership of governments and international organizations. It is an effort by 88 countries, the European Commission, and 67 participating organizations to meld disparate remote-sensing tools and ground-based databases- 300 databases and counting- into a seamless Global Earth Observation System of Systems (GEOSS), which is expected to come fully online in 2015 [15].

Since the beginning of 2007, researchers from CSISS of GMU and technical staffs from the USGS have collaborated with other GEO partners on the design, implementation, and upgrade of GEOSS. In particular, they have introduced the GEOSS Common Infrastructure (GCI) to enable this blueprint.
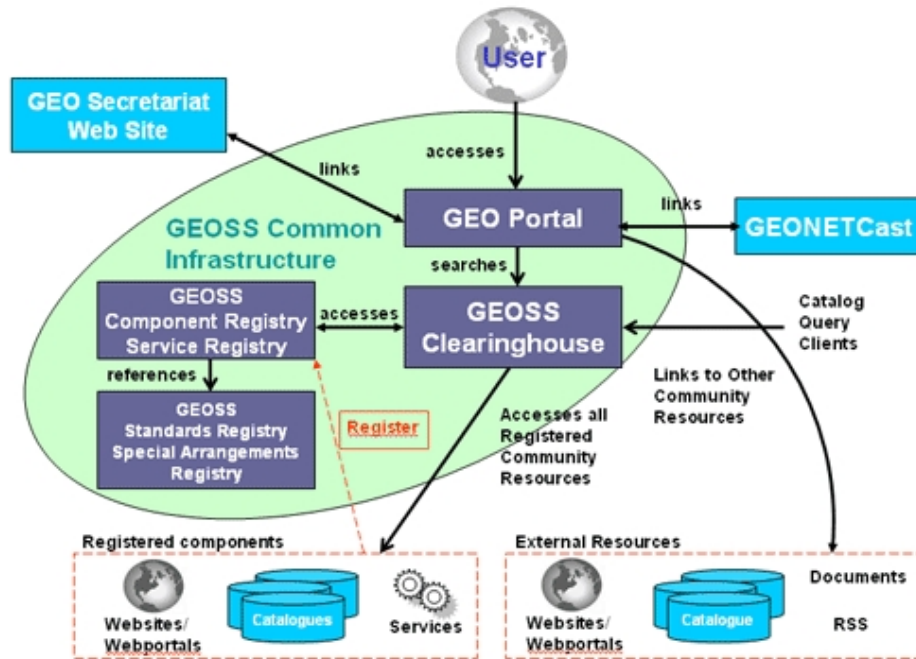
Fig. 4. GEOSS common infrastructure

As shown in Fig. 4, resource providers register their Earth observation resources in the Component and Service Registry [16]. The categories of resources that could be contributed to the GEOSS are listed in Table 1.

Table 1. GEOSS resource category

No	Resource category name
1	Datasets
2	Monitoring and Observation Systems
3	Computational models
4	Initiative or programme
5	Websites and documents
6	Analysis and visualization
7	Alerts, RSS and information feeds
8	Catalogues, inventories and metadata collections
9	Software and applications
10	Service interfaces

During the registration process, providers may reference the standards maintained in the Standards and Special Arrangements Registry [17] to further identify the characteristics of their resources, such as supported data format and available access protocol. Clearinghouse is responsible for indexing the registered resources, including the ones maintained in the external catalogues. It exposes a metadata search API interface to the GEO Portal [18], which presents a graphical user interface to users for searching available resources in GEOSS.

In particular, existing databases of EO resources, such as Global Change Master Directory [19], only need to be registered once with the Component and Service Registry as a record. Clearinghouse will retrieve the list of catalogues from the Component and Service Registry, and then either harvest metadata from these catalogues, or perform distributed search when needed.

3. COMPARING THE DATA SYSTEMS' CAPABILITIES WITH RESEARCHERS' REAL NEEDS

3.1 Analysis of Geospatial Data Systems' Capabilities

The above four exemplar systems represent typical geospatial data systems at the institutional, agency, cross-agency, and international levels respectively. The capabilities in making geospatial data publicly available vary at these levels.

Institution-level data systems, such as the one in NASA GES DISC, usually have established data systems for long-term archiving and management of Earth science data, and provided several ways for users to search and download the data. Since the scientific disciplines these data systems focus are few, the number of available data products is limited, and the heterogeneity among the data products are low, more advanced capabilities could be developed at these data systems, such as data customization, integrated data analysis, as exemplified by AeroStat.

Agency-level data systems, such as NASA ECHO, usually have multiple affiliated institution-level data systems to support. Though the focused scientific disciplines at this level may not be very broad, the number of data products is usually high and heterogeneities in the semantic meanings and syntactic encodings of archived data products exist. Therefore, data systems at this level can still enable data discovery and access for researchers by federating affiliated data systems together, but they may not be able to provide advanced data analysis and data customization functionalities for researchers. Taking NASA as an example, it is already a big achievement to provide a centralized metadata clearinghouse, and enable an API interface on top of that. Further enabling data customization and even data analysis for all the NASA satellite data holdings is difficult. Actually, the ways of customization and analysis may be highly dependent on the data products themselves: they vary from one data product to another.

Data systems at the cross-agency level are actually a federation of multiple data systems individually owned by different agencies Bai et al. [20], as exemplified by CWIC. Differences on the ways of data discovery and data access among these agencies' data systems are very common. These are barriers to effective data discovery and data access. Advanced capabilities in making data publicly available are usually not available at this level.

For the data systems at the international level, data discovery itself is already a big issue. In the case of GEOSS, though individual Earth observation resources could be registered directly with the Component and Service registry and then quickly exposed to the users, most of the existing databases of EO resources, in particular EO agency's data catalogues, are still not accessible. Metadata harvest and distributed search are two basic ways to solve this problem. However, for metadata harvest, many data systems do not prefer this one, as the metadata information may not be routinely updated once harvested. For distributed search, incentives for the GEO to provide dedicated wrappers for these catalogues

respectively or for the providers to upgrade their catalogues to be compliant with open standards are still not big enough. The way to get the data delivered to the users at this level is another big issue, where user registration, direct data access through URLs, and indirect data access through data ordering are major problems when fulfilling data download for researchers. Therefore, integrated capabilities of data customization and data analysis in data systems at the international level are not realized at this moment.

The capabilities of data systems at these four levels are summarized in Table 2 below, where "Yes" means the specific capability is usually available in data systems at the corresponding level, "No" means it is usually not available, and "Yes/No" means it is already available in some but not all of data systems at the level.

Table 2. The results of data system capabilities analysis

Data system level	Capabilities			
	Data discovery	Data access	Data customization	Data analysis
Institution level	Yes	Yes	Yes/No	Yes/No
Agency level	Yes	Yes	Yes/No	Yes/No
Cross-agency level	Yes	Yes/No	No	No
International level	Yes/No	Yes/No	No	No

3.2 Analysis of Researchers' Needs in Manipulating Scientific Data

In the guide of Dissertation Research in Education [21], the general research process was illustrated as a circle, as shown in Fig. 5. This process usually starts from choosing a topic of interest. After gaining a general overview, researchers need to narrow the subject into a specific research question. Decision on the types and amount of information needed is a necessary step before searching the data and downloading them into local archive. Examining the search results and performing analysis on them are needed to reveal new findings. Repeating these steps might be needed before finishing this piece of study and publishing the results.
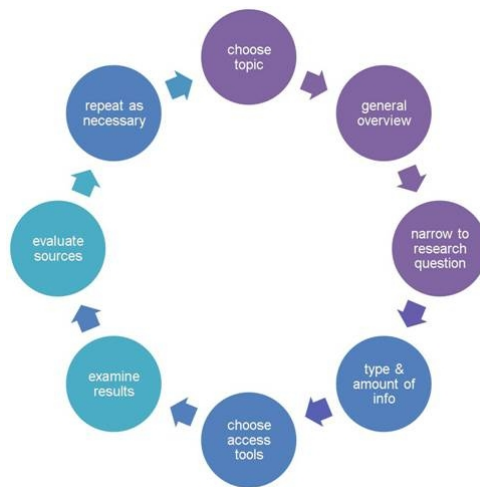

Fig. 5. Research Process

As shown in Fig. 6, Bearman and Trant [22] introduced another view on the research process, with a focus on the interactions between data providers and data consumers, i.e. researchers.
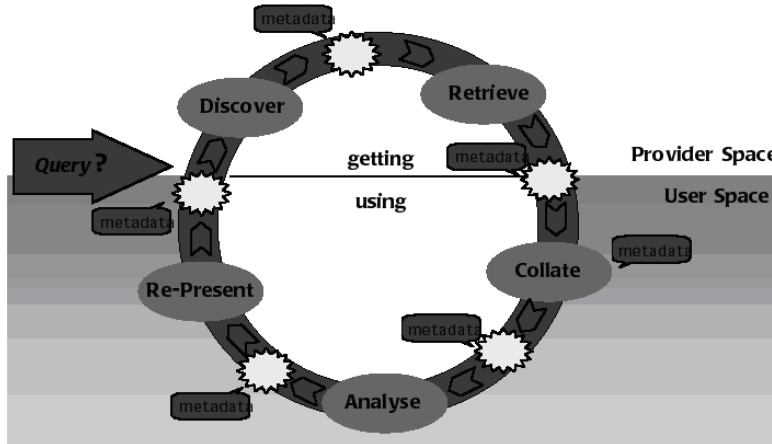

Fig. 6. Research Process with a focus on provider-user interaction

In this diagram, research process is also depicted as a circle: discovering the data is the very first step, which is basically fulfilled by the data providers. The next step is retrieving the data of interest, which is also fulfilled by the providers. After that, collating the data retrieved from multiple sources, analyzing them to have new findings, and finally re-presenting the research results are usually taken care of by researchers.

What highlighted in this view are the functionalities needed to manipulate the research data, and where they are usually fulfilled, in either provider space or user space. What is the same between this view and the state-of-the-art of geospatial data systems is that data discovery and data access are still fulfilled by providers, which makes sense, as it is where data are maintained.

4. A PARADIGM SHIFT IN GLOBAL CLIMATE CHANGE RESEARCH

4.1 From "Everything-Locally-Owned-and-Operated" to "Everything-Shared-Over-the-Web"

The provider-user interaction depicted in Fig. 5 actually reflects the old paradigm in the data-intensive Earth science research, of which the data manipulation capabilities and resources are owned and operated by researchers locally without interoperability and sharing with others. This paradigm is called "everything-locally-owned-and-operated" [23] [24]. With this paradigm, moving data from the provider site to the researcher site becomes necessary. Because a typical global change study requires integral analysis of huge volumes of multi-source data, it takes a lot of time to gather and download all required data and also requires the researchers to have tremendous local computing resources and analysis capabilities. Only a few researchers have such resources. Therefore, the old paradigm of research has significantly hampered the progress of global climate change research.

In the past decade, the technology development in standards and interoperability and policy change on data and resource sharing have promoted the shifting of research paradigm from "everything-locally-owned-and-operated" to "everything-shared-over-the-Web (or Web-and-Service Centric paradigm)" [25]. The state-of-the-art geospatial data systems built with this paradigm have demonstrated that such systems can dynamically pull standard-based interoperable data, computing, and analysis resources contributed by the community to perform a climate change research project entirely over the Web without requiring any local resources at the researcher's site [26,27]. In the new paradigm, downloading the data to local archives become unnecessary, and any researchers with a Web access can perform climate change research that was only able to be conducted by a few well-funded scientists.

In such a system, the analysis functions are decoupled from the data sources since an analysis may use data from different sources but the data customization functions can be implemented for a specific data source. Both analysis and customization functions interact with data sources through standard-based machine-to-machine interfaces. Moving data from sources to an analysis function is still needed but is done through standard-based machine-to-machine interface automatically without human intervention. Because climate change research deals with large amount of data, it is advantage to deploy data customization and analysis capabilities as close to data sources as possible to reduce the data traffic and improve the system efficiency.

4.2 Global Agricultural Drought Monitoring and Forecasting System (GADMFS): an Exemplar System Featuring the Web-and-Service Centric Paradigm

Global Agricultural Drought Monitoring and Forecasting Yagci et al. [28] is a NASA- and NOAA-funded research project (NA09NES4280007, NNX09AO14G, PI: Prof. Liping Di). It aims to provide a Web-based service system for researchers and policy makers to monitor and to forecast the global agricultural drought status (Fig. 7).

The system consists of the data providing component, middleware service component, and data dissemination portal component contributed by different organizations Deng et al. [29]. The data providing component is the NASA Land and Atmosphere Near real-time Capability for EOS (*LANCE*) system for near real time satellite data and MODIS data archiving system at the NASA's Land Processes Distributed Active Archive Center (LP DAAC). Both LANCE and the data system at LP DAAC are intended to provide data to many different human and machine users, not only for GADMFS. The middleware service component, developed by CSISS at GMU, provide geospatial services for computing agricultural drought indices [30], such as the Normalized Difference Vegetation Index (NDVI), Vegetation Condition Index (VCI), and Agricultural Drought Severity Classification. It also provides drought analysis functions as services. Finally, a Web-based data dissemination portal [31] allows the users to interact with the system to provide user requests and visualize and download analysis results. The three components of the system interact each other through the open standard machine-to-machine interfaces, mainly defined by the Open Geospatial Consortium (OGC).
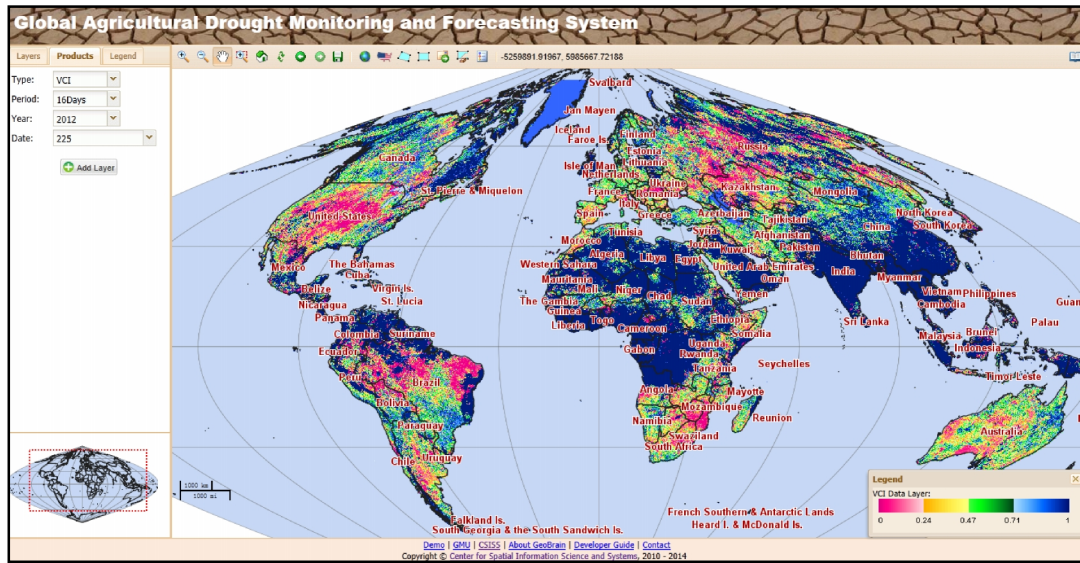
Fig. 7. Global agricultural drought monitoring and forecasting system

By adopting the "web-and-service centric" paradigm, the system has serval advantages over the traditional system. With the system, anyone in the world can monitor and analyze the current and past agricultural drought conditions for any part of the world without requiring any local resources other than the Web access. The system provides much more flexibility, reusability and scalability. Although currently the middleware service component is working with NASA data sources, it can also work with any other data sources that support the OGC standard interfaces. Any client or portal can invoke the services in the middleware service components and obtain the analysis results as long as the client/portal complies with the same interface standards as that employed by the services. Therefore, the services in the system can be repurposed dynamically and can be a component of other systems for different applications. Through the system, scientific research results could be easily discovered, customized, and also integrated with others by a broad range of users. The system also highlights the role that a data service provider could play in the whole research community, which has been gradually moving the research operation model from the traditional "data center – end user" model to the "data center – service provider- end user" model under the web-and-service centric paradigm.

5. GLOBAL CHANGE STUDIES CALL FOR THE WEB-AND-SERVICE CENTRIC PARADIGM

"Everything-locally-owned-and-operated" is the old research paradigm and outdated design principle for geospatial data system architectures. It clearly distinguishes the data systems and the user communities when supporting the whole research process: data systems are responsible for archiving the data, and enabling the data discovery and data access steps, while all other steps are fulfilled by researchers themselves locally. Therefore, moving data of interest from the archive environment on the provider side to the working environment on the researcher side is necessary.

"Everything-shared-over-the-Web" is the promising research paradigm and the state-of-art design principle for geospatial data system architectures. It focuses on supporting the whole research process by sharing resources contributed by the research community. This paradigm maximizes the use of the computational power, analysis capabilities, and the storage facility on government agencies, large institutions, and the entire research community to fulfill as many steps of climate change research process as possible. Therefore, researchers don't need to transfer the data of interest from the data provider's site to the local working environment and large local computing resources and analysis capabilities become unnecessary.

When performing global change studies, understanding the difference between these two paradigms is critical, as the need for effectively manipulating large volume of data is greater than ever before. Taking the Coupled Model Intercomparison Project Phase 5 (CMIP 5) for example, besides promoting a standard set of model simulations, another key work item during the CMIP 5 is to inter-compare the model outputs submitted by difference model develop teams. However, transferring 40 TB FGOALS-g2 global climate model output data from one place to another is a big demand for network bandwidth. Archiving and then analyzing it locally is another big demand for data storage and computational power. The current best practices of analyzing the model outputs is the data system provides subseting and analysis services so researchers can analyze the portion of data they are interested in online, instead of downloading the large volume of data locally.

By adopting the paradigm of "everything-shared-over-the-Web", a new generation of geospatial data systems can mobilize the resources in the entire research community for global change studies. Such mobilization increases the research efficiency, maximizes the utilization of existing research resources, enables more scientific studies and knowledge discovery, and democratizes the climate change research. And this is exactly the concept of cyberinfrastructure promoted by the US National Science Foundation [32].

6. CONCLUSIONS

In this paper, capabilities of geospatial data systems are analyzed through four exemplar systems. The general research process is then analyzed to show the researchers' needs in performing the global change studies. Comparing the available data systems' capabilities with researchers' needs show the strengths and limitations of data systems at different levels when supporting global change studies. This study reaches the following conclusions:

1. Generally, researchers need to fulfill data discovery, data access, data customization, data analysis, and result publication when performing scientific research.
2. The capabilities of geospatial data systems vary from one level to another. Data systems at the cross-agency level and the international level may only support data discovery and data access, while data systems at the institution level and the agency level tend to have a better support for researchers in data customization, analysis, and visualization.
3. Most of the existing operational geospatial data systems (such as NASA's ECHO) follow the "everything-locally-owned-and-operated" paradigm, where scientific data needs to be transformed from data systems to researchers. In the contrary, "everything-shared-over-the-Web" is a new paradigm and design principle for geospatial data systems to have more data manipulation capabilities available

through interoperability and sharing. Most of systems currently under the development, such as GOESS and CWIC, follow this paradigm although data analysis capabilities are not fully operational yet. Existing operational systems are moving towards this paradigm through adding standard-based service interfaces in their evaluation.

4. Adopting the paradigm of "everything-shared-over-the-Web" with the design principle of putting the processing capabilities as close to data sources as possible for a new generation of geospatial data systems in support of global change studies is very promising: the whole research process could be efficiently fulfilled, and global climate change studies can be conducted quicker, better, and less expensive.

5. This new paradigm is fully aligned with US NSF's cyberinfrastructure initiative, where the computational power, analysis capabilities, and institutional knowledge of the entire research community could be adequately leveraged to release scientists' time on data preparation and manipulation so that scientists can focus on the science.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. NASA's Earth Observing System Data and Information System. Accessed 18 January 2013.
   Available: http://earthdata.nasa.gov/.
2. Wang B, Liu M, Yu Y, Li L, Lin P, Dong L, et al. Preliminary Evaluations of FGOALS-g2 for Decadal Predictions. Advances in Atmospheric Sciences; 2013: (In press).
3. Coupled Model Intercomparison Project Phase 5. Accessed 18 January 2013.
   Available: http://cmip-pcmdi.llnl.gov/cmip5/.
4. Overpeck TJ, Meehl AG, Bony S, Easterling RD. Climate Data Challenges in the 21st Century. Science. 2011;331(6018):700-702. DOI: 10.1126/science.1197869.
5. AeroStat: Online Platform for the Statistical Intercomparison of Aerosols. Accessed 18 January 2013.
   Available: http://disc.sci.gsfc.nasa.gov/aerosols/services/aerostat/aerostat.html.
6. NASA's Earth Observing System (EOS) Clearinghouse (ECHO). Accessed 18 January 2013.
   Available: http://earthdata.nasa.gov/echo.
7. CEOS WGISS Integrated Catalog Project Page. Accessed 18 January 2013. Available: http://www.ceos.org/index.php?Itemid=225.
8. Global Earth Observation System of Systems (GEOSS). Accessed 18 January 2013.
   Available: http://www.earthobservations.org/geoss.shtml.
9. Aeronet: AErosol RObotic NETwork.
   Available: http://aeronet.gsfc.nasa.gov/.
10. Holben BN, Eck TF, Slutsker I, Tanré D, Buis JP, Setzer A, et al. AERONET - A federated instrument network and data archive for aerosol characterization, Remote Sensing of Environment. 1998;66:1-16.
11. Shao Y, Di L, Bai Y, Kang L, Wang H, Yang C. Federated Catalogue for Discovering Earth Observation Data. Journal of photogrammetry, remote sensing and geoinformation processing. 2013: (In press). DOI: 10.1127/1432-8364/2013/0157.

12. Nebert ND, Arliss W, Panagiotis V. Open GIS Catalogue Service Specification. Accessed 18 January 2013.
Available: http://www.opengeospatial.org/standards/specifications/catalog.

13. Reverb: the Next Generation Earth Science Discovery Tool. Accessed 18 January 2013.
Available: http://reverb.echo.nasa.gov/.

14. CEOS International Directory Network (IDN). Accessed 18 January 2013. Available: http://idn.ceos.org/.

15. Stone R. Earth-Observation Summit Endorses Global Data Sharing. Science. 2010;330(6006):700-702. DOI: 10.1126/science.330.6006.902.

16. Bai Y, Di L, Chen A, Wei Y, Cheng X, Shrestha R, et al. GEOSS Component and Service Registry: Design, Implementation and Lessons Learned. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS). 2012;5(6):1678-1686. DOI: 10.1109/JSTARS.2012.2215914.

17. GEOSS Standards Registry. Accessed 18 January 2012. Available: http://seabass.ieee.org/groups/geoss/.

18. GEO Portal. Accessed 18 January 2012. Available: http://www.geoportal.org/.

19. Global Change Master Directory. Accessed 18 January 2013. Available: http://gcmd.nasa.gov/.

20. Bai Y, Di L, Chen A, Liu Y, Wei Y. Towards a Geospatial Catalogue Federation Service. Photogrammetric Engineering and Remote Sensing. 2007;73(6):699-708.

21. Dissertation Research in Education. Accessed 18 January 2013. Available: http://mc.libguides.com/content.php?pid=47444&sid=483903.

22. Bearman D, Trant J. Unifying our cultural memory: Could electronic environments bridge the historical accidents that fragment cultural collections. Accessed 18 January 2013.
Available: http://www.archimuse.com/papers/ukoln98paper/index.html.

23. Di L, McDonald K. Next generation data and information systems for earth sciences research. In Proc. First International Symposium on Digital Earth, November 29 - December 2, 1999, Beijing. 1999;92-101.

24. Di L. GeoBrain - A Web Services based Geospatial Knowledge Building System. In Proc. NASA Earth Science Technology Conference, June 22-24, 2004, Palo Alto; 2004.

25. Di L. A framework for developing web-service-based intelligent geospatial knowledge systems. Journal of Geographic Information Sciences. 2005;11(1):24-28.

26. Di L, Zhao P, Han W, Wei Y, Li X. Web service-based GeoBrain online analysis system (GeOnAS). In Proc. NASA Earth Science Technology Conference, June 19-21, 2007, College Park. 2007;19-21.

27. Han W, Di L, Zhao P, Li X. GeoBrain Online Analysis System: A SOA-based Geospatial Web Portal. In: Zhao P, Di L, editors. Geospatial Web Services: Advances in Information Interoperability. IGI Global; 2011.

28. Yagci LA, Di L, Deng M, Yu G, Peng C. Global agricultural drought mapping: Results for the year 2011. In Proc. IEEE International Geoscience and Remote Sensing Symposium, July 22-27, 2012, Munich. 2012;3764–3767. DOI: 10.1109/IGARSS.2012.6350498.

29. Deng M, Di L, Yu G, Yagci, LA, Peng C, Zhang B, Shen D. Building an on-demand web service system for Global Agricultural Drought Monitoring and Forecasting. In Proc. IEEE International Geoscience and Remote Sensing Symposium, July 22-27, 2012, Munich. 2012; pp. 958-961. DOI: 10.1109/IGARSS.2012.6351392.

30. Shen D, Deng M, Di L, Han W, Peng C, Yagci LA, et al. Study on generation and sharing of on-demand global seamless data -Taking MODIS NDVI as an Example, Computers & Geosciences. 2013: *(in press)*. DOI: 10.1016/j.cageo.2012.11.011.
31. Global Agricultural Drought Monitoring and Forecasting System. Accessed 18 January 2013.
Available: http://gis.csiss.gmu.edu/GADMFS/.
32. Cyberinfrastructure Vision for 21st Century Discovery. Accessed 18 January 2013.
Available: http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

_____

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=180&id=10&aid=877