NEHEA and *GeoBrain* – An Organization and System for Data-Intensive Earth System Science (ESS) Education and Research at Colleges around the World

Meixia Deng (mdeng@gmu.edu), Liping Di (ldi@gmu.edu)
Laboratory for Advanced Information Technology and Standards (LAITS)
Center for Spatial Information Science and Systems (CSISS)
George Mason University
6301 Ivy Lane, Suite 620
Greenbelt, MD 20770

Abstract

This paper describes a data-enhanced Earth system science (ESS) learning and research environment enabled by a Web-service-based geospatial information and knowledge system called *GeoBrain*. The system allows students, faculty and other researchers from institutes all over the world to easily access and analyze on-line the large volume of NASA EOS data. The paper also introduces the NASA EOS Higher Education Alliance (NEHEA), an open organization with members from the ESS higher-education community for developing, managing, operating, maintaining and utilizing the GeoBrain system. In addition, it discusses how GeoBrain can significantly reduce some current problems in ESS education.

1. Introduction

Earth System Science (ESS) studies the Earth as an integrated system. It is a multi-disciplinary, system-focused Earth science with the goal of providing a comprehensive understanding of the changing Earth. The focus of ESS is on the systems concept, viewing Earth as a complex system with interacting components, including the atmosphere, biosphere, geosphere, hydrosphere, and anthroposphere, and understanding these individual components and the interactions among them. The overall purpose of ESS is to observe, understand, predict, and act on the changes of the Earth system (Kump, Kasting, and Robert Crane, 2002).

Three significant features distinguish contemporary ESS research and education from those scientific endeavors in traditional geo (Earth) science or other scientific domains:
1) ESS research and education is multi-disciplinary,
2) ESS research and education needs a large amount of data and information and may be computationally intensive; and
3) The study regions span multiple spatial scales (e.g., a leaf, a field, or an area of local, regional, continental, or global extent) (Di and McDonald, 1999). These features make many aspects of ESS, such as collecting data or information, obtaining computing resources, and becoming familiar with a wide field of knowledge, unique and challenging for both educators and researchers. Therefore, it is not a surprise that there exist many obstacles in current ESS related education and research and these problems have significantly hampered the associated scientific productivity (Di, 2004).

With this insight into ESS education and research, the NASA Earth Science Research, Education, and Applications Solution Network (REASoN) program funded a 5-year project named *NASA EOS Higher Education Alliance: Mobilization of NASA EOS Data and Information through Web Services and Knowledge Management Technologies for Higher Education Teaching and Research* in 2004. For convenience, this project is more generally called as NEHEA and GeoBrain, while NEHEA standing for NASA EOS Higher Education Alliance, and GeoBrain being the system to be developed for mobilization of NASA EOS data and information. The objective of this paper is to give readers some basic ideas about how this project can help the ESS research and education communities by providing targeted solutions to some current problems hampering ESS productivity and showing how researchers and educators can take advantage of the results of the project to fulfill their tasks better and more easily or to conduct some work that was previously impossible. Infrastructure, technology and other related details of this project will not be discussed here. Those interested in can go to http://geobrain.laits.gmu.edu/ for more information.

2. NEHEA and GeoBrain: the Organization and System

The NEHEA and GeoBrain project has formed a NASA EOS Higher Education Alliance (NEHEA) to undertake the development, operation, maintenance, evaluation, and promotion of the GeoBrain system. The project team consists of system developers from four universities and funded education partners from multiple higher-education institutes. At least three partners are selected by an open Request for Proposal (RFP) solicitation each year in the first three years of the project. George Mason University is the PI institute for this project. NEHEA is an open and free alliance with the project team to be the core members. Any interested institutes and individuals in the world are welcome to join as associate members at no cost (http://geobrain.laits.gmu.edu/partners.htm).

The goal of the NEHEA and GeoBrain project is to provide solutions for some current problems in and innovative technologies for the ESS education and research activities in higher-education institutes. The goal will be reached through collaboration between the development team and the education partners in NEHEA.

The development team is responsible for developing *GeoBrain*, a geospatial Web-service system, based on the geo-object and geo-tree concepts discussed in Di and McDonald, 1999. GeoBrain is being implemented as a standards-compliant, open, distributed, interoperable, three-tier, self-evolutionary Web information system. It is also a geospatial modeling and knowledge building system (Di, 2004). GeoBrain provides a data-rich learning and research environment that allows students, faculty and other researchers from institutes all over the world easily accessing, analyzing, and modeling with the large amount of NASA EOS data on-line through their desktop computers. The system leverages NASA EOS data and computational resources and the latest geospatial information technologies, including: (1) the web-based geospatial interoperability standards and protocols developed by FGDC (Federal Geographic Data Committee),

International Standards Organization (ISO), and Open GIS Consortium (OGC) (Di, 2003); (2) the Web service and semantic Web (Di, 2004b); (3) geospatial Grid (Chervenak et al., 2001; Foster, Kesselman, and Tuecke, 2001; Di 2005). In the past two years, many GeoBrain functionalities have been developed, including customized data access to NASA EOS data archives, on-line data analysis, and virtual geospatial product generation. The details of current GeoBrain functionalities were described in Di, 2005b.

The NEHEA education partners are responsible for providing requirements to the system, evaluating the system for debugging and improvements, and developing new curricula for fully utilizing the new capabilities enabled by GeoBrain. In order to facilitate the exchange of experiences gained from the uses of GeoBrain in teaching and research and bring the common interests and resources of education partners together, NEHEA has formed a course development panel for the education partners. The panel members are working together to develop teaching materials that fully utilize GeoBrain capabilities. Currently, the panel is working on teaching materials for land use/land over change both in urban and rural areas using GeoBrain on-line resources. The materials will be available on-line once the development is finished.

3. GeoBrain Solution to Common Problems in ESS Education and Research

Currently, ESS research routinely involves collecting and manipulating large volumes of geospatial data. Satellite remote sensing is one of the major sources for such data (Asrar and Dozier, 1994). NASA's Earth Observing System (EOS) collects more than four TB of remote sensing data per day, and NASA's Earth science data centers have archived multiple Pbytes of data from EOS and the pre-EOS era (Asrar and Greenstone, 1995; McDonald, Di, and Wharton, 2002; King et al., 2003). Those EOS data are very attractive and useful to ESS educators and researchers because of the variety of data sources, the multi-disciplinary coverage, science orientation of the data collection, and free or low cost nature of the data. Therefore, EOS is the first choice of data sources for most ESS researchers and educators. A qualified ESS researcher must be able to deal with the large volume of EOS data. In order to prepare students better for understanding the Earth system and global changes, higher-education institutes engaging in ESS education must train students to deal with the large data volumes and the diversity of data and information sources. However, in current stage there are great difficulties associated with using large volumes of EOS data in classroom teaching and student research. The following subsections will discuss how GeoBrain can contribute solutions for overcoming those difficulties.

3.1. Data Access

The difficulty in accessing the huge volumes of EOS data is not the availability of EOS data but the complicated process to get the data and pre-process the data into a ready-to-use form. Currently a typical way for a user to obtain EOS data for their applications is to place a data request at the NASA EOS data and information system (EOSDIS). If the volume of requested data is large, users sometime need to wait for days or even weeks for actually getting the data. And the data users obtained, in most cases, are still in archive

form, which may be significantly different from the form needed by the users in terms of data form, spatial/temporal coverage, and projection. For example, if a user conducts a study covering the state of Virginia using MODIS data, what the user can obtain from EOSDIS may be MODIS data granules covering almost half the USA.

This situation causes problems in using EOS data for research. First, many studies requiring real or near-real time data cannot be conducted. Second, because users cannot obtain the data in a user-specified form, they need to spend a significant amount of time and resources in pre-processing the data. Third, many users are unfamiliar with HDF-EOS format used for archiving all NASA EOS data. Few existing analysis software packages can ingest or are generic enough to handle small differences in the HDF-EOS data produced by different producers. Other problems include differences in map projection, the need for georectification of swath data before use, etc. Those problems are common not only to university users, but also to most other EOS data users. As a result of all these problems, it may take months for a user to get the EOS data into their analysis system for analysis or modeling. If researchers want to modify models or to apply models to new data, they need to go through all the steps again. Huge efforts and resources are wasted in repetitive tedious data preparation work.

GeoBrain provides a solution to those problems by offering **interoperable**, **personalized**, **on-demand data access to petabytes of NASA EOS data.** The system allows users to retrieve instantly any part of the NASA EOS on-line or near-line data in the format, projection, and geographic location specified by the user, so that the data are ready for analysis at the user's system. It may only take seconds or minutes to finish the work that previously took months to complete. Such a capability of GeoBrain is built on huge NASA computational resources. These resources support standard-based open data discovery and access. NASA EOS data centers are in the process to provide OGC standard compliant services (e.g., Catalog Service for Web-CSW (Nebert, 2004), and Web Coverage Service-WCS (Evans etc., 2003)) on the data pools. Those on-line standard-compliant resources allow GeoBrain to obtain the user requested data through those interfaces and apply value-added services to those data so that the user will obtain the data in the user-specified form. The value-added geospatial data services include subsetting, resampling, reformatting, reprojection, and georectification.

This capability of GeoBrain is operational currently and ready for all-purpose use. There are two ways to exploit this capability: 1) accessing the on-line data products through Web interface http://geobrain.laits.gmu.edu:8099/cswquery-csf/ directly, and 2) installing a Multiple Protocol Geoinformation Client (MPGC) at a local machine and using the client to access the capability. The client can be downloaded from http://geobrain.laits.gmu.edu/mpgc/.

3.2. Data Analysis

As we mentioned, ESS research is data- and computing- intensive. Not many universities have software systems or computing resources sufficient for each student to conduct their own study of interest in ESS. Currently, a professor who uses EOS data in the class

typically spends weeks or months in obtaining various samples of EOS data and in georectifying, reprojecting, and reformatting the data into the form acceptable by the in-house analysis systems. The sample dataset normally covers a small geographic region, and all students share the same dataset for the class exercise. The same sample dataset is normally used semester after semester. In addition, lack of the software licenses and insufficiency of computers resource prevents students from free exploration of the data. As a result, students are never exposed to the richness of the EOS data and will never learn how to use this vast amount of data in real-world applications.

The GeoBrain system provides **on-line analysis capabilities** so that a user can conduct ESS projects without the need to have large computing facilities locally. Numerous geospatial processing functions for on-line analysis of geospatial data are implemented as Web services in GeoBrain. The services allow users to process and analyze the NASA EOS data on-line and get only the results back. Currently GeoBrain has implemented more than 200 Web geospatial processing services. Most of them are converted from standalone open source GIS package called GRASS (GRASS 2005).

3.3. Geospatial Modeling and Knowledge Sharing

In addition to the difficulties and problems discussed above, there is no existing Web-service software system in the ESS world that provides an open, interoperable platform for users to construct complex Web-executable geospatial models, run the models on line, and share the models with peers.

GeoBrain provides not only customized EOS data to users with greater easiness, but also a standard framework and platform for users to construct and share service modules and geospatial models. All analysis functions in GeoBrain are implemented as chainable Web services. GeoBrain allows users to construct complex geospatial processing models by chaining those service functions together. GeoBrain then executes the model and sends the result back to users. Users do not need to worry where the input data are located and in what forms and how the model is executed since GeoBrain will automatically take care of those issues. The geospatial processing functions and models represent the knowledge in the geospatial/ESS domain. Students and researchers can submit their models and functions to GeoBrain. After proper review and validation, those functions and models will become the operational capabilities of the system, available to any GeoBrain user. This functionality enables community development of the system. The system will evolve and grow with more and more contributions from users.

Currently, this capability is still under development; only a demonstration prototype is available. The prototype modeling capabilities are provided allowing users to construct geospatial processing models and executed them at the GeoBrain workflow engine, the BPELPower.

4. Use of GeoBrain in Classroom Teaching and Research

The GeoBrain system essentially gives students and faculty a data-rich geospatial learning and research environment that was never available to them before. Though the technology developed in GeoBrain can be used for other purposes, the GeoBrain system is intended to serve higher-education users as 1) an unlimited global geo-data source, 2) an on-line data analysis system, 3) an on-line platform for geospatial processing and modeling, and 4) a platform for sharing geospatial knowledge. With the functionality and capacity of the GeoBrain system, requirements at the users' side become very light. Since the client software component is free, users can conduct complex ESS research with an Internet- connected PC capable of running a JAVA client. Fast Internet connection is not necessary at the user site, because GeoBrain does the data reduction and analysis at server side and users need only retrieve the result rather than all the raw data.

GeoBrain could potentially be a valuable tool for classroom teaching and research. In the past two years, the main focus of the project was to develop the baseline functionalities of GeoBrain so that education partners can start to evaluate the system and incorporate the capability into classroom teaching and research. The baseline system has been fully functional since October 2005. Currently twelve graduate and undergraduate courses in nine universities are evaluating GeoBrain capabilities for classroom teaching. Those courses covers diverse spectrum of the Earth system science, from meteorology, geology to remote sensing of land surface. The evaluation and feedback will be valuable for guiding the further development of GeoBrain. In addition, over ten federal and state-funded ESS research projects in the institutes of NEHEA core members are evaluating the usability of GeoBrain as an on-line data source and research tool. Eleven graduate students in those institutes are also involved in developing components of the GeoBrain system. Preliminary feedbacks from these evaluations are mostly positive although a number of urgent improvements are identified, such as better user interfaces, better documentation, more projection options, and data discovery through Landsat World Reference System (WRS). The development team is currently implementing those improvements and it is expected the first round of improvements will be finished by the end of May 2006.

GeoBrain is configured for open access through Web. There is no charge to use any GeoBrain services even the user is not a NEHEA member. Currently a lot of non-NEHEA members are using the system for various proposes, such as technology demonstration, interoperability testbed, data sources for research, etc. The GeoBrain system is also a part of Open Geospatial Consortium Network, the Global Earth Observation System of Systems (GEOSS) Service Network, and Committee on Earth Observation Satellites (CEOS) testbed. In higher education, the system has recorded users from more than 100 universities other than the NEHEA member institutes around the world using the GeoBrain data and services. So far, we have not received any feedback from those users on the services provided by the systems.

5. Summary

GeoBrain provides a new data-rich learning and research environment enabled by the latest geospatial Web service and interoperability technologies. The environment makes

vast amount of geospatial data and computational resources available and easily accessible on-line to any Internet connected desktop computer and enables rapid modeling and analysis of data. It saves ESS researchers a significant amount of time and computing resources when conducting application studies and enables them to perform some formerly impossible tasks. It allows students in any university to freely explore the vast amount of data, computational resources, and, analysis capability in ESS for use in learning and research. The whole ESS higher-education community will benefit from new or enhanced courses utilizing the GeoBrain environment.

The technologies and functionalities implemented in GeoBrain will dramatically reduce the current problems associated with conducting ESS education and research at colleges and universities and help to train students on how to effectively use large amounts of remote sensing data in ESS research and applications. Although the system is still in the development stage, a large number of higher-education users around the world have already used the services provided by the initial version of GeoBrain. This indicates the value of the system in the ESS education. When additional service modules, geospatial models, and EOS data become available in the system, it is expected the user base of the system will expand rapidly.

The GeoBrain system is designed to be a self-evolutionary system built by the ESS higher-education community for the community. The advice and feedback from the NEHEA higher-education partners ensure that the system will meet the needs and expectations of the community. It is also expected that the ESS higher-education community will have great involvement in the system development and evolution by contributing significant numbers of geospatial services modules and models to the system.

Acknowledgements

This research is funded by a grant from NASA Earth Science Research, Education, and Applications Solution Network (REASoN) program (NNG04GE61A, PI: Dr. Liping Di).

Reference

Asrar, G., and Dozier, J., 1994. *EOS: Science Strategy for the Earth Observing System*. American Institute of Physics Press, Woodbury, New York 11797.

Asrar, G., and Greenstone, R., 1995. *1995 MTPE EOS Reference Handbook*. NASA Goddard Space Flight Center, Greenbelt, MD 20771.

Chervenak A., I. Foster, C. Kesselman, C. Salisbury, S. Tuecke, 2001. The Data Grid: Towards an Architecture for the Distributed Management and Analysis of Large Scientific Datasets, *Journal of Network and Computer Applications*, 23:187-200.

de La Beaujardière J (ed) (2001) Web Map Service Implementation Specification, Version 1.1.1. OGC 01-068r2, Open GIS Consortium Inc. http://www.opengis.org/docs/01-068r2.pdf

Di, L., 2005. "The Geospatial Grid." In S. Rana and J. Sharma eds, *Frontiers of Geographic Information Technology*, Springer-Verlag.

Di, L., 2005b. "The Implementation of Geospatial Web Services at GeoBrain." Proceedings of 2005 NASA Earth Science Technology Conference. June 28-30, 2005. College Park, MD. 7p. (CD-ROM).

Di, L., 2004b. Distributed Geospatial Information Services-Architectures, Standards, and Research Issues. *The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Vol. XXXV, Part 2, Commission II, 7p.

Di, L., 2004. GeoBrain-A Web Services based Geospatial Knowledge Building System. Proceedings of NASA Earth Science Technology Conference 2004. June 22-24, 2004. Palo Alto, CA, USA. (8 pages. CD-ROM).

Di, L., 2003. "The Development of Remote-Sensing Related Standards at FGDC, OGC, and ISO TC 211." *Proceedings of IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2003), July 21-25, Toulouse, France.* 4p (Invited).

Di, L., and K. McDonald, 1999 "Next Generation Data and Information Systems for Earth Sciences Research", *Proceedings of the First International Symposium on Digital Earth*, Volumn I. Science Press, Beijing, China. Pp. 92-101.

Di, L., W. Yang, D. Deng, and Ken. McDonald, 2002. "Interoperable, Personalized, On-demand Geospatial Data Access and Services Based on OGC Web Coverage Service (OWS) Specification", *Proceeding of NASA Earth Science Technology Conference*, CDROM, Pasadena, California. 3pp.

Evans, J. B. Bernhard, L. Di, etc., 2003. Web Coverage Service (WCS), Version 1.0.0, OGC 03-065r6. OpenGIS© Implementation Specification. Open Geospatial Consortium. https://portal.opengeospatial.org/files/?artifact_id=3837 (101p).

Foster I., C. Kesselman and S. Tuecke, 2001. The Anatomy of the Grid – Enabling Scalable Virtual Organizations. Intl. J. of High Performance Computing Applications, 15(3), 200-222.

GRASS 2005. GRASS homepage, http://grass.baylor.edu.

King, M.D., Closs, J., Spangler, S., and Greenstone, R., (eds), 2003. *EOS Data Products Handbook*. National Aeronautics and Space Administration, Washington D.C., NP-2003-4-544-GSFC.

Kump L., J. Kasting, and R. Crane, 2004. *The Earth System*, Second edition. Culinary and Hospitality Industry Publications Services. 420p.

McDonald, K., L. Di, and S. W. Wharton, 2002. "Strategic Evolution of NASA ESE Data Systems (SEEDS)." In Chen and Jiang, eds, *The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Vol. XXXIV, Part 2, Commission II, pp 327-329.

Nebert D., eds, 2004. *OpenGIS Catalog Service Specification*. (OGC 04-021r2). version 2.0, http://portal. opengeospatial.org/files/files/?artifact_id=5929.