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

Liping Di and Meixia Deng
Laboratory for Advanced Information Technology and Standards (LAITS)
George Mason University
9801 Greenbelt Road, Suite 316-317
Lanham, MD 20706
ldi@gmu.edu

LAITS

LAITS

Introduction

- Earth System Science (ESS) studies Earth as an integrated system.
- Satellite remote sensing is one of major methods for acquiring data required for the ESS research, especially for continental and global scale research.
- Handling a large volume of remote sensing data with computers in scientific models are the essential skill that ESS researchers must master.
- ESS education has to prepare students for handling data-intensive nature of the ESS.

LAITS

The Features of ESS Research

- The research is multi-disciplinary;
- The research needs the great amount of data and information and may be computational intensive;
- The research regions may be micro (e.g., a leaf), field, local, region, continental, or global.

LAITS

Process of Learning and Knowledge Discovery in Data-Intensive ESS

1. Find a real-world problem to solve
2. Develop/modify a hypothesis/model
3. Implement the model/develop analysis procedure at computer systems.
   - Determine the data requirements.
4. Search, find, and order the data from data providers.
5. Preprocess the data into the ready-to-analysis form
   - reprojection, reformatting, subsetting, subsampling, geometric/radiometric correction, etc.
6. Execute the model/analysis procedure to obtain the results.
7. Analyze and validate the results
8. Repeat steps 2-7 until the problem is solved.

LAITS

ESS Data Available at NASA

- The NASA Earth Observing System (EOS) collects more than 2Tb of remote sensing data/ day.
- Currently NASA Active Archive Data Centers (DAACs) have archived multiple peta bytes of data from EOS and pre-EOS era.
  - Significant part of the data archives have never been analyzed once.
- All of those data are free to all data users.

LAITS

NASA ESS Data Environment

- The EOS data and information system (EOSDIS) is designed to manage, archive, analyze, and distribute the ESS data.
  - Originally designed for supporting NASA funded scientists.
  - Based on technologies of 20 years ago.
  - Mainly for supporting well-funded NASA ESS research projects.
  - Not considering the small data users and educators.
- The standard data format in EOSDIS is HDF-EOS.
- EOSDIS distributes data in granules, which may cover large geographic regions.
- No data services provided.
- Technology insertion continues to improve EOSDIS

LAITS

1

Problems in Data-intensive ESSE

- Difficulty to access the huge volume of EOS data.
  - Take weeks to order and obtain a large volume of EOS data.
- Difficulty to use the data.
  - Significant time, resources, and data/IT knowledge are required for preprocessing the multi-source data into a ready-to-analyze form.
  - The ESSE faculty normally does not have enough knowledge in the data/IT knowledge.
- Lack of enough resources to analyze the data.
  - Few universities have the hardware/software resources to handle multi-terabytes of data in simulation and modeling for solving global-scale problems.

LAITS

Current Use of EOS Data in ESSE Classes

- Only use of samples of EOS data
  - Professors take weeks or months to obtain various samples of EOS data, then georectify, reproject, and reformat the data to the form acceptable by the in-house analysis systems.
  - The sample dataset normally cover a small geographic region
- All students share the same dataset for the class exercise.
- The same sample datasets are used in many semesters.
- Insufficiency on the software license and computer resource prevents students from free exploration of the data.
- Students are never exposed to richness of EOS data and will never learn how to use this vast amount of data in the real-world applications.

LAITS

The Objectives of the Research

- To enable the students and faculty of higher-education institutes easily accessing, analyzing, and modeling with the huge volume of NASA EOSDIS data for teaching and research just like they possess such vast resources locally at their desktops.
  - Enable the education users to handle vast NASA EOS data and computing resources like their local ones.
  - Develop/enhance courses that fully utilize the environment for Earth System Science/Geospatial education
- To realize this goal, we will develop an open, standard-based interoperable web geospatial information system called *GeoBrain* and operate it on top of NASA EOSDIS on-line data resources
  - Develop geospatial web service and knowledge management technologies for NASA EOS data environment.
  - Implement them in an open, standard-based, distributed, interoperable web service system.
  - It is a geospatial modeling and knowledge building system
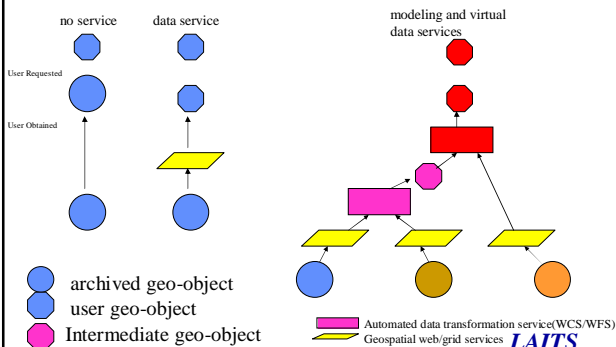
LAITS

Expected Significances

- The GeoBrain system will give ESSE institutes a geospatial data-rich learning and research environment that was never available to them before.
- The environment will enable students interactively, through their desktop computers, explore answers to the scientific questions by mining the peta-bytes of EOSDIS data.
- The technology also provides the interactive collaboration among students worldwide on scientific modeling, knowledge exchanges, and scientific criticism.
- Such an environment will inspire students' curiosity on sciences and enable faculties and students doing many new studies that could not be done before.
- It will also provide educators with unique teaching tools and compelling teaching experiences that they never have experienced and that only NASA can offer.

LAITS

Geo-object, Geo-tree, Virtual Dataset, Geospatial Models

- archived geo-object
- user geo-object
- Intermediate geo-object
- Automated data transformation service(WCS/WFS)
- Geospatial web/grid services

LAITS

The Infrastructure Foundation

- NASA ESE is working on putting ESS data at DAACs on-line for rapid access through data pools
  - Most commonly requested and most recently acquired data currently.
  - 4 DAACs have data pools online already.
  - Eventually all data will be on-line.
- NASA ESE has excellent network infrastructure for data traffic
  - In most cases, 1Gb/second links between NASA DAACs/research centers.
- NASA ESE has huge computational resources.
- Make the vast data and computational resources available and easily accessible to ESSE institutions
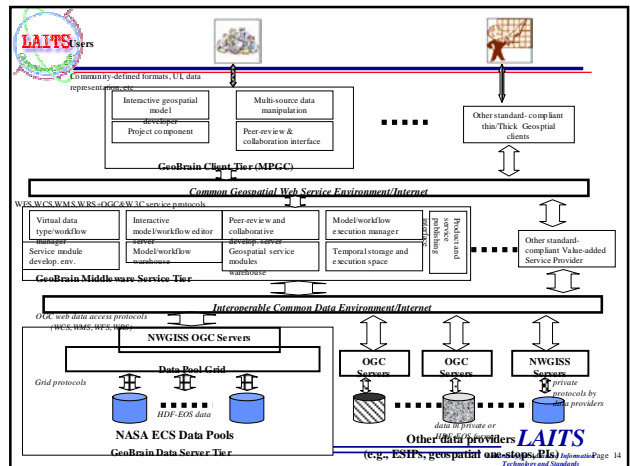
LAITS

The Technology Foundation

- The web-based geospatial interoperability technology.
  - Standards developed by FGDC, ISO, and OGC.
  - The common interfaces to data archives of different data providers for obtaining personalized ready-to-analyze dataset.
- The web service technology
  - The fundamental technology for E-commence.
  - Web Services are self-contained, self-describing, modular applications that can be published, located, and dynamically invoked across the Web.
  - Automatically and dynamically chaining individual services and connecting services to data for solving complex problems are the goal of semantic web.
- Grid technology
  - Securely share the geographically distributed data and computational resources.

LAITS

LAITS

System requirement at the user-side

- Any internet connected PC capable of running JAVA client of the system.
  - The client will be provided to any users for free.
- No fast network connection is required
  - all data reduction is done by the system at computers that users don't need to know.
  - Users only get the result back instead of all raw data.
- No powerful computer with large disk storage capability is needed
  - Basically the users possess the huge computational and data resources that the system can mobilize.
- No expensive analysis software is needed
  - Analysis and modeling capabilities are provided by the system

LAITS

System built by ESSE community for the community

- The GeoBrain system will be built by the ESS higher-education community for the community.
- The major tasks of system development will be:
  - Development of service framework that allows the automated execution of services and service chains.
  - Development of services modules and geospatial models.
- Any individuals can contribute both modules and models.
- A peer-review panel will be set up to review and validate the modules and models contributed by the community.

LAITS

Involvement of ESSE Community

- As the users of the system.
  - Provide the requirements
  - Evaluate the systems
  - Develop new curriculums and research around the newly available capabilities.
- Participate in the system development
  - Develop individual service modules
  - Contribute the geospatial modules

LAITS

Evolution and Self-enhancement of the System

- Beside the computational and network capacity and the data holdings in various distributed archives, the power of the system relies on the availability of the service modules and geospatial models.
- With more and more contributions of modules and models from the user community, the system will become more and more powerful and knowledgeable.
- The inclusions of the modules and models into the system will be subjected to rigorous peer review and testing.

LAITS

3

How Can College-level ESSE take advantage of the research

- The vast data and computational resources will be available and easily accessible on-line by any Internet connected desktop computers.
- Rapid modeling and analysis on vast data archives will become possible.
  - Many more research can be conducted that cannot be conducted before because of lack of resources.
- Students can explore the vast amount of data and computational resources and the analysis capability provided by the system freely.
- The new/enhanced courses utilizing the data-rich environment will be beneficial to ESSE.
- The software and tools will be freely available to user community.

LAITS

NASA EOS Higher-Education Alliance (NEHEA)

- In order to integrate this new learning environment into ESS teaching and research, we formed a NASA EOS Higher Education Alliance (*NEHEA*).
- The core of NEHEA consists of
  - The GeoBrain development team led by GMU.
  - Funded Education partners- a group of Earth science educators funded from an open RFP process.
- NEHEA welcomes Earth science educators around the world to join NEHEA as non-funded education partners.
- NEHEA members are incorporating the data enhanced learning environment into their teaching and on-going research and will develop new courses for taking advantages of the environment.
- All NEHEA education partners will have accessed to all NASA data available through GeoBrain, all GeoBrain software, and technical support for free.

LAITS

The Current NEHEA Team

- Development Team
  - George Mason University (PI Institute)
  - City University of New York
  - Northern Illinois University
  - University of Texas – Dallas
- Education partners
  - In the first three years of the project, three education partners will be selected in each year through a RFP process (Total 9 partners).
  - We have selected three funded education partners in 2004.
    - Penn State University, University of Central Florida, and CUNY-Hunter College.
  - Each partner will be provided two years of funds to develop new/enhanced courses based the capabilities, promote the use of the system in the peers, and provide feedback to the development team.
  - Faculty and students from any higher-education institutes are welcomed to use the system and participate in the development.

LAITS

Current Status and Future Development

- Data
  - 10 Tb of Landsat TM, ETM, ASTER, MODIS, DEM etc. on line now.
  - 200 Tb of NASA EOSDIS data pools data will be available through GeoBrain as on-line data resource in summer of 2005.
    - Most current EOS data will be provided in this way
  - All EOSDIS near-line archives will be available to users as the GeoBrain data resources through machine-to-machine gateway of EOSDIS.
- Services
  - All data are available through WCS data download, enabling the data services provided by WCS server (e.g., subsetting, resampling, etc)
  - Data are also provided to WCS and WMS clients, and searchable through OGC CS/W client.
  - Chainable geospatial web service modules are being developed by converting Open Source GRASS functions into web services.
  - A BPEL service engine has been developed and geospatial processing modeling capabilities are being implemented yet.
- Software
  - All executables of GeoBrain software are available for free downloading.
  - You can use those software to set up your geospatial web services.
- GeoBrain data and service resources can be found at http://geobrain.laits.gmu.edu

LAITS

Acknowledge

- This research is funded by a grant from NASA Earth Science Research, Education, and Applications Solution Network (REASoN) program.

LAITS