The Implementation of Geospatial Web Services at GeoBrain

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

***Abstract* – The LAITS' REASoN project, entitled "NASA EOS High Education Alliance (NEHEA)," intends to mobilize NASA EOS data and information through Web service and knowledge management technologies for higher-education teaching and research. The technology component of the project is to develop and deploy a standard-compliant, open, distributed, three-tier web-service-based geospatial knowledge system called *GeoBrain*. This paper discusses recent progresses on GeoBrain, including the developments of multiple standard-compliant chainable geospatial processing services, a OGC CSW catalog service, and a workflow execution engine named BPELPower. The alpha version of GeoBrain, which combines the above new developments with NWGISS data access system, has been deployed at an Apple G5 based server cluster for testing and evaluation. The server cluster has been populated with large amount of typical EOS data for users to generate personalized, on-demand, value-added products.**

I. Introduction

NEHEA project is one of projects funded by NASA REASoN program. The project intends to mobilize NASA EOS data and information through Web service and knowledge management technologies for higher-education teaching and research. The technologies, based on geo-object and geo-tree concepts [1][2], are being implemented in a standard-compliant, open, distributed, three-tier web information system named *GeoBrain*. The system will make peta-bytes of NASA EOS data and information in both on-line data pools and near-line storages, easily accessible to higher-education users as if they have such resources locally.

The NASA EOS Higher Education Alliance (*NEHEA*), led by GMU and consisting of the development team from four universities, funded Earth science educators from nine universities, and non-funded educational partners, is responsible for the development, maintenance, and operation of GeoBrain. NEHEA is also responsible for promoting the use of the system in teaching and research of Earth system science (ESS) at higher-education institutes. NEHEA members will incorporate the data enhanced environment into their existing courses and on-going research and will develop new courses for taking advantages of the environment. The system, developed by the end-user community for the community, is bridging the gap between EOSDIS data archives and higher-education end-users. Di in [2] discussed the overall concepts of geo-object and geo-tree, the use of the concepts for construction of geospatial processing models and management of geospatial knowledge, the geospatial web service and interoperability technology, and the GeoBrain architecture. This paper discusses the recent progresses made in the development and deployment of the GeoBrain system.

II. The Development of GeoBrain

GeoBrain is a web-services based geospatial knowledge system for providing value-added geospatial service and modeling capabilities to ESSE user community. The development of GeoBrain is the core of the technology development in the NEHEA project. During the past year, significant progresses have been made on GeoBrain, including the development and deployment of several geospatial web services and a workflow execution engineer (BPELPower), and the enhancement of Multiple Protocol Geoinformation Client (MPGC). The following subsections will discuss those new developments.

A. Geospatial Web Services

Individual geospatial web services are the building blocks for constructing a geospatial processing workflow. The power of GeoBrain relies on the availability of large numbers of standard compliant, chainable services. Therefore, the project has spent a fair amount of time and resources to develop individual geospatial web services. The following paragraphs provide descriptions on the services. The URLs for those services can been found at http://geobrain.laits.gmu.edu or at http://laits.gmu.edu.

- Web Coverage Service (WCS): OGC WCS supports the networked access to multi-dimensional and multi-temporal geospatial data as "coverage" through "*getCapabilities*", "*describeCoverage*" and "*getCoverage*" interfaces [3]. WCS provides intact

This project is funded by NASA REASoN Program (NNG04GE61A)

geospatial data products encoded in HDF-EOS, NITF and GeoTIFF to meet the requirements of client-side rendering, multi-source integration and analysis, and inputs to scientific models and other clients beyond simple viewers. The official version of OGC WCS specification is 1.0. However, WCS 1.0 has problems associated with access to multiple-dimension remote sensing data. During OGC Web Service Initiative 2.0, proposal on revising WCS was developed and incorporated into a draft version that the WCS revision working group was working on. The draft version is WCS 1.0.2 that greatly simplifies the access to multi-dimension remote sensing data. The WCS 1.0 was implemented in NWGISS [4][5]. The GeoBrain project extends the service to support 1.0.2. In addition, WSDL description and SOAP binding were added to both versions of the servers so that they become chainable services. Both versions of services can work with all NASA EOS datasets in standard HDF-EOS format.

- Web Map Service (WMS): OGC WMS supports the networked interchange of geospatial data as "map" which is generally rendered in a spatially referenced pictorial image, such as PNG, GIF or JPEG, dynamically from real geographical data. NWGISS contains a WMS 1.1.1 server that works with NASA HDF-EOS data [5]. Similar to NWGISS WCS server, the server is enhanced to become a chainable service. The service has been deployed at an Apple cluster server as a part of GeoBrain Map services.
- Web Coordinate Transformation Service (WCTS): An OGC WCTS service has been developed for re-projecting a coverage encoded in HDF-EOS Grid from one map projection to another one. The service is based on the OGC WCTS draft specification [6]. Currently the service supports all of the map projections defined in GCTP and used in HDF-EOS. The required input parameters for using the service include the URL of source HDF file, target projection type, resampling method, bounding box and the 15 float-type projection parameters defined by GCTP, and the result is a URL point to the resulted HDF file.
- Web Image Classification Service (WICS): An unsupervised image classification service was developed, based on an open-source GIS software package called GRASS. The service takes two parameters: "*initialNumberOfClusters*" and "*imageURL*". The unsupervised classification modules of GRASS have other parameters with default values. We omitted them to make the input simpler. The return of this service is "classifiedImageURL" point to classified image. The service is deployed at http://data.laits.gmu.edu:8099/wics/classifiers/. The service was a part of the development of OGC Web Service Initiative 2.0 and resulted in an OGC discussion paper on Web Image Classification Service [7].
- Web Image Cutting Service (IMCS): IMCS allows user to get data for an area specified by a state name or a series of coordinates that defines a polygon. Not like the bounding box that only allows for a rectangular region, the cutting service will cut out the image in any shape specified by the set of coordinates. A filling value is given for the area outside of the polygon. The filling value (normally zero) is recorded in metadata. The service can be easily chained with other services, such as WCS to provide access to data for a specific non-rectangular administrative or geographic region.
- GRASS Web Services**:** The power of GeoBrain relies on having a large number of interoperable, web-based geospatial processing services. It is very time consuming to develop such services from scratch and the project cannot afford to develop services in this way. Our strategy is to make existing geospatial processing packages web-service enabled. The software package we chose is an open source GIS package called GRASS [8]. We are adding the SOAP based web service interfaces to more than 200 geospatial processing modules in GRASS. A detailed description of the GRASS Web services can be found at the web page: http://geobrain.laits.gmu.edu:8099/webservices/grass/Grass_Webservices.html
- Catalog Service for Web (CSW): OGC CSW specification supports the registry and discovery of geospatial information resources [9]. It plays a "directory" role in the open, distributed Web service environment. Data and services providers register their capabilities using metadata, and users can then query the metadata to discover interesting information. Because of importance of catalog services in GeoBrain, a catalog service has been developed [10], which implements OGC Registry Information model (OGC-RIM) and allows users to discover distributed service and data by a very simple user interface. The detailed descriptions, samples and usages of CSW server can be found at http://geobrain.laits.gmu.edu/CSW/discovery/

B. Multi-Protocol Geoinforamtion Client (MPGC)

During the past year, enhancement to NWGISS MPGC has been done to make it more reliable and to work with services available at GeoBrain. The overall goal of the enhancement is to make MPGC a GeoBrain services and modeling client. At the end of January 2005, MPGC 1.0 was released to NEHEA members for testing and evaluation. MPGC provides an interoperable way of accessing geospatial Web services, especially those from the Open Geospatial Consortium (OGC), for integrating and analyzing distributed heterogeneous Earth science data. MPGC complies with the OGC Catalog Service for Web (CSW) specification [9], allowing the registry, discovery and access of geospatial information resources that are distributed over Internet. In addition, MPGC 1.0 also implemented the latest protocols of the OGC Web Feature Service (WFS)[11], Web Map Service (WMS)[12] and Web Coverage Service (WCS)[4]. It provides a single point of entry for accessing to OGC-compliant data and services around the world and for requesting any subsets of multi-dimensional and multi-temporal geospatial data for a user-specified geographic region. MPGC also can access Web geospatial processing services, such as the Web Image Classification Service (WICS) and the Web Coordinate Transformation Service (WCTS), to generate value-added data products on demand. Moreover, at the local machine where MPGC is running it also can 1) reformat the retrieved dataset into a data format specified by the user; 2) provide robust visualization and analytical tools for multi-source geospatial data; and 3) support multiple data formats: HDF, GeoTiff, GML, JPG, PNG, and GIF. The website for downloading MPGC and getting help information is at http://geobrain.laits.gmu.edu/mpgc.

C. Workflow Execution Manager (WEM) – BPELPower

In order for multiple web services to work together to produce user-defined on-demand geospatial products, individual services have to be organized into a service chain based on service logics. The service chain is then executed by a workflow execution manager to produce the on-demand product. The GeoBrain development team has developed a workflow execution manager called BPELPower. It is based on the mainstream standards of Web services, including BPEL, WSDL, WSIF, Xalan, Xerces, UDDI, AXIS, SOAP, JNDI, J2EE (servlets/EJBs/JSPs), Jetspeed (Portlets) and JMX. It runs on top of popular application servers, such as Tomcat, J2EE, JBoss, Weblogic and WebSphere. It supports BPEL-based web service chain completely. First, BPELPower supports "deploy it". WSDL-based web services can be deployed in BPELPower, where their validations are checked. Second, it supports "see it". WSDL-based web services can be displayed in BPELPower in different ways. Third, it supports "try it". BEPLE-based web services chain can be executed in BPELPower dynamically. Different invocations (e.g., HTTP POST/GET, SOAP document/rpc, etc.) are all supported. BPELPower was demonstrated at OGC OWS 2.0 demo event in September 2004 to show the generation of virtual geospatial products on demand and the power of service chaining. The detailed description of BPELPower can be found at http://bpel.laits.gmu.edu.

III. Deployment and Operation of GeoBrain

A. Deployment of GeoBrain at an Apple Cluster Server

One of keys for GeoBrain to work is to have high-performance servers to provide operational value-added services to education users of Earth system science. By combining resources from other funding projects, this project was able to setup an Apple G5 cluster server with all necessary GeoBrain service software for providing on-demand geospatial data access and value-added services to users. The cluster is running as the operational platform for GeoBrain. The following is the hardware, software, and network specifications of the system.

- *Hardware*
  - A) One Apple 64-bit Xserver G5 head node with dual 2.0GHz G5 processors, 4GB of DDR SDRAM and 2 * 250GB hard drive.
  - B) Five Apple 64-bit Xserver G5 cluster nodes, two with dual 2.0GHz G5 processors, and other three with 2.3 GHz G5 processors. And 2GB of DDR DSRAM and 80GB hard drive for each node.
  - C) Three Xserver RAID systems with total capacity of 21.7TB.
  - D) 2 Gbps Fiber Switch for exchanging data between cluster node and RAID system.
  - E) 8 Port Gigabit Ethernet Switch for system administration online.
  - F) 3000VA UPS for power protection and supply.
  - G) 25U Enclosure for racking cluster nodes and RAID system.
  - H) PowerBook G4 for system administration.
  - I) Other peripheral equipments.
- *Software*
  - A) Pre-installed Apple Mac OS X server V10.3 Unlimited-Client on head node.
  - B) Pre-installed Apple Mac OS X server V10.3 10-Client on cluster node.
  - C) Pre-installed Apple Mac OS X Panther V10.3 on PowerBook G4.
  - D) Installed Apple Remote Desktop 2 on cluster node and PowerBook G4.
  - E) Installed Apple Xcode Tools.
  - F) Installed other GNU tools.
  - G) Installed HDF, HDF-EOS tools and development library
  - H) Installed Globus Toolkit 3.2.1
  - I) Installed components of the GeoBrain system, including the WCS, WMS, and CS/W servers.
- *Total System Performance and Capacity:*
  - A) Three dual 2.0GHz CPU and three dual 2.3GHz CPU G5 processors.
  - B) 14GB DDR SDRAM.
  - C) 21.7 TB of RAID system plus additional 900GB hard drives.
  - D) 2Gbps data exchange between cluster node and RAID storage system.
  - **E)** 1 Gbps network link to Internet II and 100 Mbps network link to Internet I.

B. Data Products

The system has been populated with some key Earth remote sensing datasets. Currently, the following Landsat datasets are ready in the GeoBrain system:

Name	Year	Coverage	Description
MSS	1975	North East Europe, West Russia	304 scenes, 4 bands in each scene, total size is 19.22GB
TM	1990	Global	8473 scenes, 7 bands in each scene, total size is 4.07TB
ETM+	2000	Global	7935 scenes, 9 bands in each scene, total size is 13.52TB

In addition, more than 200GB of typical EOS data product samples as well as the Global Digital Elevation Model at 90-meter spatial resolution are deployed in the GeoBrain system. All data products in the machine are registered and served by GeoBrain. Users can access the data on-line with GeoBrain through OGC standard interfaces. Valued added services, such as subsetting, resampling, reformatting, georectification, and re-projection, are automatically provided based on the differences between what user requested and what the data in the server.

The data sets populated at the GeoBrain Apple server cluster has been registered at the CSW server. User can access to the data discovery service at http://geobrain.laits.gmu.edu:8099/CSWDiscovery/Query

C. User Statistics

To provide better and easier access and use of vast NASA EOS data resources to ESSE and general user community is the overall goal of this project. In order to better understand who are using the GeoBrain services, we installed a software package called Advanced Web Statistics 6.2 (build 1.783) on January 19, 2005 to log the web services access. From January 19, 2005 to January 31, 2005, 270 unique users accessed GeoBrain services. For the period from February 1, 2005 to February 28, 2005, 510 unique users used GeoBrain services.

	January 19-31, 2005	February 1-24, 2005
Unique Users	270	510
Foreign domain	35	52
U.S. Government domain	2	6
U.S. Academy domain	19	18

The software used to log the data access currently is not able to log the amount of data accessed. We are currently implementing data access log so that such information will be available in future. With more data and services becoming available at GeoBrain, it is expected that significant increase of GeoBrain users will happen in the coming year.

IV. Conclusions and Future Work

Web service technology is a very promising technology for solving problems related to geospatial interoperability and knowledge discovery. The initial deployment and operation of GeoBrain have demonstrated this point and also shown the flexibility and extensibility of the service-oriental architecture (SOA) and standard-based geospatial web services. Our work on GRASS proves that it is not very difficult to turn the traditional, stand-alone geospatial analysis packages into chainable web services. Therefore, it is expected in the near future, many such packages will be available as web services.

In the next several years, we will concentrate on developing and refining individual components of GeoBrain. Although the current capability of GeoBrain is rather limited compared to what we have planned for the final system. However, even the limited functionality currently in GeoBrain is very useful to the user community if more data become available through GeoBrain. In the near term, the first priority will be given to develop a CSW portal that will be able to search NASA ECHO [13] to find NASA EOS datasets at EOSDIS on-line data pools and near-line storages. Value-added services will be provided to those EOS data through GeoBrain.

References

[1] Di, L., and K. McDonald, 1999 "Next Generation Data and Information Systems for Earth Sciences Research", in *Proceedings of the First International Symposium on Digital Earth*, Volumn I. Science Press, Beijing, China. 8 pp.

[2] 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).

[3] Evans JD (ed) (2003) Web Coverage Service (WCS), Version 1.0.0. OpenGIS© Implementation Specification. Open GIS Consortium Inc. http://www.opengis.org/docs/03-065r6.pdf

[4] Di, L., W. Yang, M. Deng, D. Deng, and K. McDonald, 2001. "The Prototypical NASA HDF-EOS Web GIS Software Suite (NWGISS)", Proceedings of NASA Earth Science Technology Conference, College Park, MD. CD-ROM.

[5]. Di, L., and K. McDonald, 2005. "The NASA HDF-EOS Web GIS Software Suite (NWGISS)." In Qu, J. etc eds, *Earth Science Satellite Remote Sensing*. Springer-Verlag (to appear).

[6]. Whiteside, A., M. U. Müller, S. Fellah, and F. Warmerdam, 2005. Web Coordinate Transformation Service (WCTS). https://portal.opengeospatial.org/files/?artifact_id=8847

[7] Yang, W., A. Whiteside, 2005. Web Image Classification Service, OGC discussion paper. http://portal.opengeospatial.org/files/?artifact_id=8981

[8] GRASS homepage, http://grass.baylor.edu.

[9] Open Geospatial Consortium. OpenGIS® Catalogue Services Specification (04-021r2), version 2.0, http://portal.opengeospatial.org/files/files/?artifact_id=5929, May 11, 2004.

[10] Y. Wei, L. Di, B. Zhao, G. Liao, A. Chen, Y. Bai, Y. Liu, 2005. The Design and Implementation of a Grid-enabled Catalogue Service. In Proceedings of IGRASS'05. 25-29 July 2005, Seoul, South Korea. 4 p.

[11] Vretanos PA (ed) (2002) Web Feature Service Implementation Specification, Version 1.0.0. OGC 02-058. Open GIS Consortium Inc. http://www.opengis.org/docs/02-058.pdf.

[12] 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.

[13] NASA, 2005. Earth Observing System Clearinghouse (ECHO). http://www.echo.eos.nasa.gov/