The Integration of Grid Technology with OGC Web Services (OWS) in NWGISS for NASA EOS Data

Liping Di, Aijun Chen, Wenli Yang and Peisheng Zhao
Laboratory for Advanced Information Technology and Standards (LATIS)
George Mason University (GMU)
9801 Greenbelt Road, Suite 316-317, Lanham, MD 20706
Email: ldi@mason.gmu.edu , achen@laits.gmu.edu

Abstract: Grid technology provides secure, fundamental methods for advanced distributed computing and data sharing over the Internet. The technology has been used in applications in many disciplines. However, applications in the geospatial disciplines are just beginning. This paper describes a project that applies Grid technology to the Earth observation environment through the integration of the Globus Toolkit with the NASA Web GIS Software Suite (NWGISS). As one of the implementations of Open GIS Consortium's (OGC) Web Services (OWS) technology, NWGISS is a web-based, multiple OGC-standard compliant geospatial data distribution and service system. It provides geospatial data access services to data users for Earth observing data archived in individual data centers (e.g., NASA Distributed Active Archive Centers, DAACs). But it does not provide secure on-demand distributed geospatial data processing and transfer among the data centers. In this project, Grid technology has been used successfully to solve these problems and a prototype Grid-enabled OGC-compliant NWGISS system was produced. This prototype is part of a large Committee on Earth Observation Satellites (CEOS) testbed for evaluating the usability of Grid technology in the geospatial disciplines, especially in Earth observations and remote sensing.

Keywords: Grid, Globus, OGC, OWS, Geospatial Data, NWGISS

1 Introduction

Grid is a promising technology for easily sharing distributed heterogeneous computing resources. It brings together geographically and organizationally dispersed computational resources, such as CPUs, storage systems, communication systems, data and software sources, instruments, and human collaborators to provide advanced distributed high-performance computing to users [1][2]. Globus, consisting of a set of services and software libraries, is the key middleware that provides core Grid capabilities. The Globus Toolkit facilitates the creation of usable Grids, enabling high-speed coupling of computers, databases, instruments, and human expertise. With Globus, scientists can run their gigabyte-per-time-step job on multiple high-performance machines at the same time, even though the machines might be located far apart and owned by different organizations. Grid technology helps scientists to deal with very large datasets and carry out complex remote collaborations. It can be used for large distributed computational jobs, remote instrumentation, remote data transfer, and shared immersive storage spaces [3].

Geospatial data are those that can be located on Earth. Earth observations through remote sensing are by far the largest source of geospatial data. The NASA's Earth Science

Enterprise (ESE) is generating a huge volume of remote sensing data daily for supporting Earth system science and application research through its Earth Observing System (EOS) project. Most of the EOS data are in HDF-EOS, the standard format for NASA's Earth Observing System (EOS) Data and Information System (EOSDIS). These data are vitally important to studying global changes [4]. Therefore, making those data widely, easily, and freely accessible to users will greatly facilitate the global change research.

OWS is one of OGC's many initiatives for addressing the lack of interoperability among systems that process georeferenced data. In the past several years, OGC has successfully executed several phases of the OWS initiative, including Web Mapping Testbed I (WMT-I), WMT II, OWS-1.1, and OWS 1.2. Those initiatives have produced a set of web-based data interoperability specifications, such as the OGC Web Mapping Specification (WMS), which allows interactively assembling maps from multiple servers, and the OGC Web Coverage Service (WCS) Specification, which defines interoperable interfaces for accessing geospatial data, especially those from remote sensing, from multiple coverage servers [5].

As a member of OGC and a participant of those OGC interoperability initiatives, LAITS at GMU has developed an OGC-specification compliant software package called the NASA Web GIS Software Suite (NWGISS). It is a web-based, multiple OGC-specification compliant, geospatial data distribution and service system for delivering NASA EOS data to broad user communities. Currently, NWGISS consists of a Web Map Server (WMS), a Web Coverage Server (WCS), a Web Catalog Server (WCAT), a Multiple-Protocol Geoinformation Client (MPGC), and a toolbox [4][9].

The Committee on Earth Observation Satellites (CEOS) is an international organization responsible for coordinating international civil spaceborne missions designed to observe and study the planet Earth. Its membership encompasses the world's government agencies responsible for civilian Earth Observation (EO) satellite programs, along with agencies that receive and process data acquired remotely from space. Inspired by the successful applications of Grid technology in other disciplines, the CEOS Working Group on Information Systems and Services (CEOS WGISS) started a CEOS Grid testbed in September 2002 to evaluate the feasibility and applicability of Grid technology to the EO community. The testbed, currently consisting of applications from NASA, USGS, NOAA and ESA, aims to address the use of grid technology for efficient support to diverse users worldwide for easy access and applications of EO data and to EO data providers for improving their efficiency of operation and maximizing the usefulness and benefit of the EO data which they gather [6].

As one of the NASA representatives to the CEOS Grid testbed, LAITS is contributing to the testbed by integrating OGC technology with Grid technology through the development of a Grid-enabled NWGISS. This paper describes our contributions.

2 Background

The Globus project team has just released Globus 3.0 alpha, the first major implementation of new Open Grid Service Architecture (OGSA). However, in this project, we use Globus 2.2.2, the last release before Globus 3.0 alpha. Globus 2.2 provides the following function modules: the Globus Resource Allocation Manager (GRAM) for providing a common user interface to submit a job to dispersed multiple machines, the Monitoring and Discovery Service (MDS) for providing information services through soft state registration, data modeling and a local registry, the Grid Security Infrastructure (GSI) for providing generic security services such as authentication, authorization and credential delegation for applications that will be run on the Grid, the GridFTP for providing a standard, reliable, high-speed, efficient and secure data transfer service, Metadata Catalog Service (MCS) for providing a mechanism for storing and accessing metadata of geospatial data, Replica Location Service (RLS) for maintaining and providing access to mapping information from logical names for data items to target names which may represent physical locations of data items, and other modules such as simple Certificate Authorization (CA) etc[3].

The Hierarchical Data Format (HDF) is a widely used scientific data format because of its portability and multiple data model support. NASA's EOS project extended HDF to HDF-EOS by adding three new EOS specific data models – point, swath, and grid. HDF-EOS is the standard format for EOS data and products. The EOS project is producing a tremendous amount of data in HDF-EOS format at the rate of more than 2 Tb/day. These data are highly demanded by a broad range of research and application communities. However, currently most geographic information systems (GIS) cannot easily ingest HDF-EOS data. Therefore, the development of the capability for making EOS data directly accessible by users' GIS systems through the Internet for analysis will greatly enhance the interoperability and public use of EOS data [8]. NWGISS aims to provide such capability, based on OGC Web Service specifications. Currently NWGISS works with all generic HDF-EOS files and is easy to use and easy to setup. It makes NASA EOS data immediately available and accessible to GIS developed by major GIS software vendors. Hence, it significantly increases the accessibility, interoperability, and inter-use of HDF-EOS data, improves data analysis and visualization, promotes the use of HDF-EOS data not only in global change research but also in the issues of public concern such as environment and resource management, education, and community planning [9].

The Open GIS Consortium, Inc. is a not-for-profit international membership-based organization founded in 1994 to address interoperability among systems that process georeferenced data. Since 1999 OGC has successfully implemented four web-based geospatial interoperability programs: Web Mapping Testbed I (WMT-I) in 1999, WMT II in 2000, OGC Web Service Initiative 1.1 (OWS 1.1) in 2001, and OWS 1.2 in 2002, and produced a set of web-based data interoperability specifications. WMT I produced the OGC Web Mapping Specification (WMS) version 1. WMS allows interactively assembling maps from multiple servers. WMT II produced a set of new interoperability specifications. One of the most important specifications for data access from WMT II is the draft Web Coverage Service (WCS) Specification. It allows a WCS client to access real multi-dimensional, multi-temporal data from coverage servers. WCS provides an interoperable way of accessing geospatial data, especially those from remote sensing. OWS-1 further developed WCS specification and produced three draft versions of WCS specification, v0.5, v0.6, and v0.7 [5]. Currently WCS 0.7 is going through OGC specification process to become a formal OGC interoperability specification.

In order to evaluate the applicability of Grid technology to the EO community, an interim CEOS Grid Task Team was established in May 2002 in Frascati, Italy by CEOS WGISS subgroup. In Jun 2002 the Team held a workshop at College Park, Maryland to focus on preparing a preliminary draft plan for the CEOS Grid testbed and to discuss some Grid-related projects. The draft plan was approved by WGISS in September 2002 and the testbed was officially started then. Currently, the testbed consists of five Grid demonstration projects, including the NOAA Operational Model Archive and Distribution System (NOMADS), USGS EDC's Data Delivery, ESA ESRIN Ozone, NASA GFSC's Advanced Data Grid, and NASA EOSDIS Data Pools [6][7]. As the executor of the Data Pools demonstration project, we have established a prototype of data pools based on Grid and OGC OWS technologies at GMU LAITS.

3 Integration of Grid Technology with OWS

OGC Web Service (OWS) Specifications provide standards for implementing interoperable, distributed geospatial information processing software systems (e.g. GIS), by which a user can easily share geospatial data, information, and services with others. However, OGC technology consists mainly of interface specifications and content standards. They do not provide a mechanism for securely sharing the distributed computational resources. Meanwhile, because of the large volumes of EO data and geographically scattered receiving and processing facilities, the EO data and associated computational resources are naturally distributed. The multi-discipline nature of global change research and remote sensing applications requires the integrated analysis of huge volume of multi-source data from multiple data centers. This requires sharing of both data and computing powers among data centers. Grid technology, because of its security and distributed resource sharing capabilities, is the ideal technology for filling the technology gap. Therefore, the integration of Grid technology with OWS will greatly benefit the EO community.

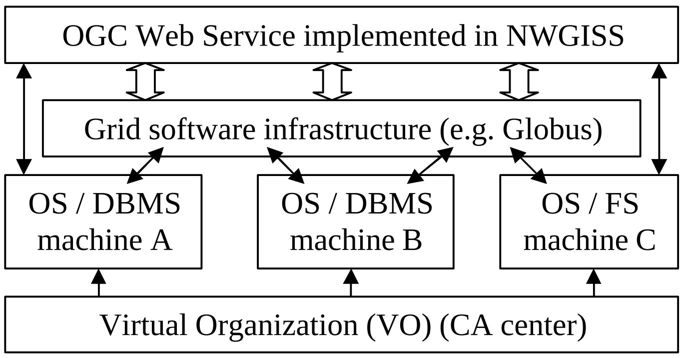


Fig. 1 Integrating Grid technology with OWS I

The information model used for the integration is shown in Figure 1. The Grid software works as the security and information infrastructures on top of the computer Operating System (OS) and its File System (FS) and Database Management System (DBMS), while

OWS software works as the application on top of the Grid security and information infrastructures to execute the secure sharing of geospatial data, information, and services. The Virtual Organization (VO) manages all sharable computers and associated resources. All machines and data users in the VO must be authorized with certifications from VO's CA (Certificate Authorization) center. OWS services will first be authenticated by the Grid security infrastructure as an authorized user and then use the Grid information infrastructure to complete the user requests. If all required resources for filling a user's request are available at the machine where the OWS service request is received, the services will be provided at the local machine without using the Grid infrastructure. Doing so will reduce the response time for filling user requests.

In addition to the integration of the OWS and Globus for secured data access and processing, we also integrated the geospatial metadata information model with Globus MCS/RLS information model [10][11] to facilitate the Grid-enabled geospatial data finding. Based on the metadata available at HDF-EOS data files [12], and referring to ISO 19115 Geographic Information- Metadata [13] and FGDC Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata [14], we proposed a general geospatial metadata information model to describe the HDF-EOS datasets. At the database level, we integrated this geospatial metadata information model with the MCS, making geospatial data retrieval Grid enabled. Users first search the geospatial metadata and MCS/RLS to get the accessible physical file name of data items. Then, users retrieve the data items by the physical file names. Details on the data retrieval process can be found on section 4.3 of this paper.

4 System Architecture and Data Flow

As a part of our initial development effort, we first set up our development environment by establishing our own Virtual Organization (VO) and Certificate Authorization (CA) center at LAITS (Figure 2), based on the Globus Toolkit. Then we integrated the Globus Toolkit with NWGISS on this VO environment to make NWGISS Grid enable. After development and proper testing, the Grid-enabled NWGISS will be installed at NASA EOS Data Pools for operational use. The following subsections will discuss the system architecture and the simplified data flow.

4.1 GMU LAITS Virtual Organization (VO) and Certificate Authorization (CA)

Currently, we have established our own VO and CA center, shown in Figure 2.

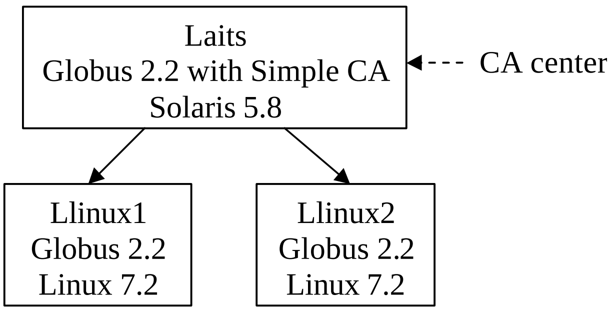


Fig. 2 Laits' VO and CA center

Globus Toolkit 2.2 was installed on a Unix server named Laits, with Solaris 5.8 and two DELL PC machines named Llinux1 and Llinux2, with Linux 7.2. We issued Globus certificates from the LAITS CA center to members of our own VO. Members in the VO can share their resources with their certificates. Our CA center's ID is "O=Grid, OU=GlobusTest, OU=simpleCA-laits.gmu.edu, CN=Globus Simple CA". Taking Laits machine's Certificates as example: Host Certificate's subject is "O=Grid, OU=GlobusTest, OU=simpleCA-laits.gmu.edu, CN=host/laits.gmu.edu", User Certificate's subject is "O=Grid, OU=GlobusTest, OU=simpleCA-laits.gmu.edu, OU=laits.gmu.edu, CN=Aijun Chen" and Service Certificate's subject is "O=Grid, OU=GlobusTest, OU=simpleCA-laits.gmu.edu, CN=ldap/laits.gmu.edu".

4.2 System Components

Figure 3 shows the architecture of the integration. The figure also shows the data flow between two machines within the VO. NWGISS consists of a map server, a coverage server, a catalog server, a geoinformation client, and a toolbox. All NWGISS components can work both independently and collaboratively, and the Globus Toolkit consists mainly of GSI, GRAM, MDS and GridFTP.

The OWS interface specifications implemented in NWGISS include the Web Map Service (WMS)[15], Web Coverage Service (WCS)[16], and Catalog Inter-operability Specification (CIS). The following paragraphs provide brief descriptions of each component. Detailed information can be found in [4].

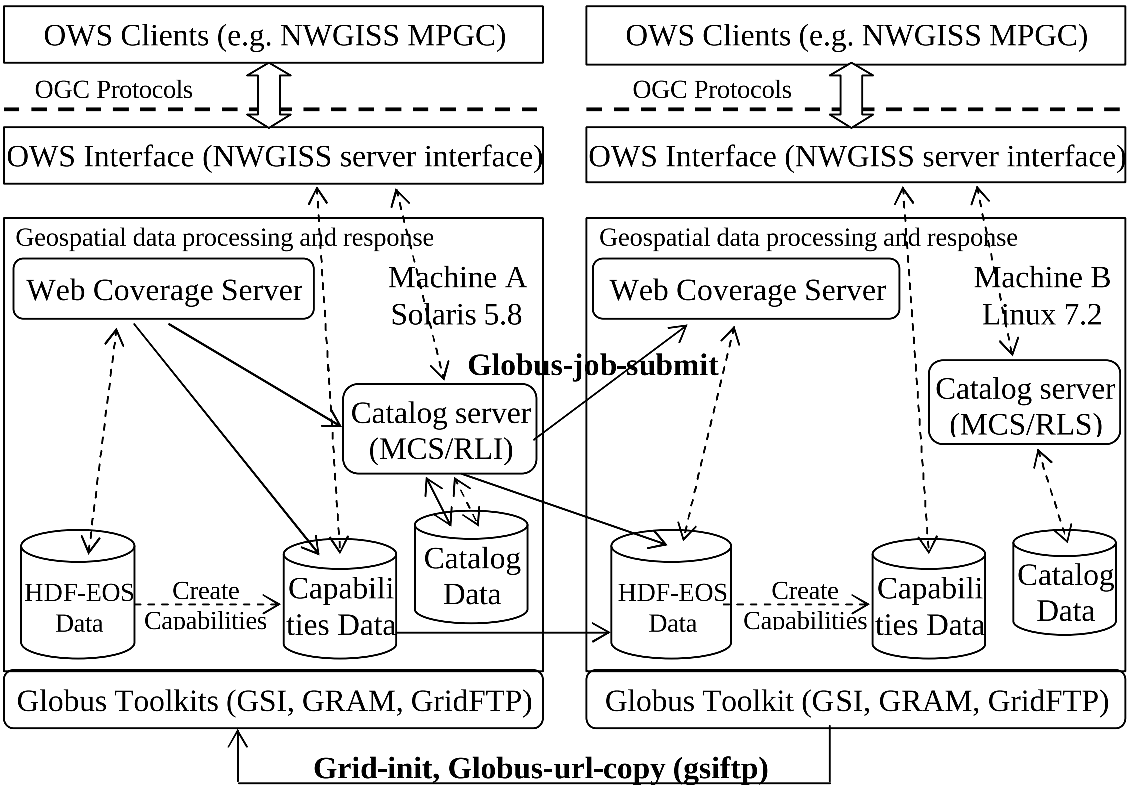


Fig. 3 System architecture and the simplified data flow demonstrating the integration (request from machine A to machine B).

- - - - - Broken lines show internal requests of NWGISS.

——— Solid lines show requests related to Globus.

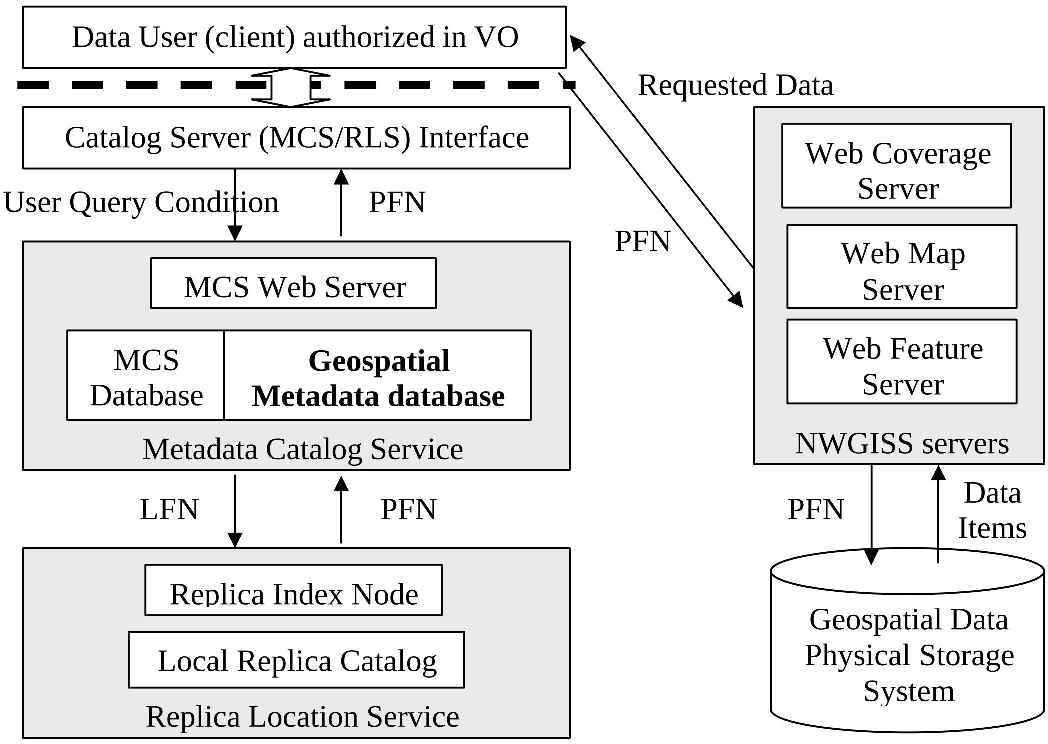
A. *The Map Server*: The map server enables GIS clients to access HDF-EOS data as maps. Currently the NWGISS map server complies with OGC WMS version 1.1.0. The OGC specification defines three interfaces, namely GetCapabilities, GetMap, and GetFeatureInfo. All three interfaces have been implemented and all three HDF-EOS data models (Grid, Point, and Swath) are supported.
B. *Coverage Server*: the OGC Web Coverage Service (WCS) specification is designed to enable GIS clients to access multi-dimensional, multi-temporal geospatial data. WCS defines three interface protocols: getCapabilities, getCoverage, and describeCoverageType. The NWGISS coverage server has implemented both versions 0.5 and 0.7 of the draft WCS specification. Three formats for encoding user-requested coverages are available in NWGISS, namely, HDF-EOS, GeoTIFF, and NITFF.
C. *Catalog Server*: Both WCS and WMS have the GetCapabilities protocol for clients to find geographic data/map and services available at servers. This protocol works nicely when a server has a small data archive. If the server has a lot of data, the capabilities description, which basically is a data catalog, is become very large. The catalog server allows GIS clients to search and find available geographic data and services in a NWGISS site based on the OGC catalog interoperability specification (CIS). The integration of this OGC catalog server with the MCS and RLS of Globus Toolkit makes the retrieval of geospatial data and services Grid enabled.
D. *Toolbox*: It contains tools for automated data ingestion and catalog creation. Currently, two types of tools are provided: the format conversion tools and XML capabilities creation tools. A third type of tools, the catalog creation tools, will be provided later.
E. *The Multi-protocol Geoinformation Client (MPGC)*: It enables the access of multi-dimensional and multi-temporal geospatial data from multiple coverage servers in the form that exactly matches users' requirements, regardless of the original forms of the data in the servers. The easy-to-use and easy-to-install web Java client enables GIS users to easily access the HDF-EOS data as well as other formatted data in a unified way, greatly enhancing the interoperability and public use of EOS data. Coupled with the NWGISS Server packages, MPGC makes HDF-EOS data available to GIS users based on the Open GIS Consortium's (OGC) interoperability protocols

The details about the Globus Toolkit can be found on the website: http://www.globus.org.

4.3 Data Requests and Data Flow Description

In our development and testing environment at LAITS, all three machines in the VO have been installed with Globus Toolkit and populated with geospatial data in HDF-EOS format. The Grid-enabled WCS, WMS and Catalog servers of NWGISS are installed at Laits and Llinux1. Users can send data and service requests from any OGC-compliant client (e.g., MPGC) through the Internet. NWGISS servers receive user requests and process the data on archives to produce data products on users' behavior. The most common data service functions include spatial, temporal, and parameter-based subsetting and subsampling, georectification, reprojection, reformatting. Those functions can be combined to produce data products that exactly match user requirements on data.

In our implementation, NWGISS servers at the machine that receives the user request play the major role in responding to user requests. Firstly, a user sends a query to the catalog server for searching and finding data through a client. User's requirements on the data are expressed in the query. The catalog server executes the query by searching the geospatial metadata database and MCS database to obtain matched logical file names (LFN). Then, the catalog server obtains the physical file names (PFN) through sending the logical file names to the RLS. The matched physical filenames and related metadata are sent back the client.



LFN: Logical File Name; PFN: Physical File Name

Fig. 4 Integration mechanism of Globus MCS with geospatial metadata

Secondly, after the user receives the search results, the user may choose to obtain the data by sending a data access request to one of NWGISS servers (e.g., WCS server). The WCS server firstly parses query to find the physical file name contained in the query. Then it uses the name to judge if the datasets reside in the local machine. If datasets exist locally, the data services of NWGISS at Laits will process the geospatial data sets locally, and return the results to user. If the datasets are not available locally, but available on any other Grid nodes in the VO, then the server on Laits will send a data request by this physical file name to the Grid nodes which host the data. Some of these Grid nodes may have only data but no services while others may have both. If the node (e.g. Llinux1) contains only the requested data, not the service, the data service on Laits will initialize Grid (grid-proxy-init) both at Laits and at Llinux1. Then it invokes the Grid secure copy function (globus-url-copy) at Laits to copy data from Llinux1 to Laits. After the data are copied, the NWGISS server at Laits will complete the user requests. If the site (e.g. Llinux2) has both data and requested services, the server on Laits will initialize Grid at both Laits and Llinux2, and then call the Grid job submitting function (globusrun) to submit user request to Grid services at Llinux2. Grid services will invoke the NWGISS data services at Llinux2 to complete the user request. The results will be copied back to Laits by using the Grid secure copy mechanism. The second case will significantly reduce the data traffic since the data reduction is done at the machine where the original dataset resides. All of the above processing is transparent to user. Figure 4 shows the data and the request/response flows.

5 Conclusions and Future Work

Our preliminary work and initial success have extended the applications of Grid technology to the EO community and also made OGC technology Grid enabled. Our experience with the project indicates that Grid technology has great potential for applications in the geospatial disciplines. By using Grid software as the foundation for geospatial data infrastructure, we achieve secure sharing of geospatial data and associated processing while providing common OGC interfaces and services to users. This integration takes advantages of both Grid and OGC technologies. It provides the user community a standard, disciplinary specific access to a huge volume of geospatial data available at space agencies and managed by Grid while shielding the details of Grid infrastructure underneath. As one of the technology prototypes for NASA ESE, the results from this project will be used by EOS Data Pools at NASA DAACs for securely sharing geospatial data and resources. Also, as one of the NASA application projects for CEOS Grid testbed, our experience and software will be shared with other participating space agencies. The NWGISS software mentioned in this paper can be downloaded at http://www.laits.gmu.edu.

The next step in our research will be the integration of OGC Web Registry Service (WRS) and Grid catalog systems for providing geospatial-specific OGC-compliable and Grid-enabled catalog services and enabling the virtual geospatial data services. Because of the nature of geospatial data and services, the current Grid metadata catalog system is not good enough for geospatial applications. Integration will provide a better, geospatial-specific catalog services for effective search and discovery of geospatial data and services managed by Grid.

Acknowledgement

This project was supported mainly by grants from the NASA Earth Science Data and Information System Project (ESDISP) and NASA Earth Science Technology Office (ESTO). Additional funding was provided by the Open GIS Consortium (OGC) for the development of NWGISS coverage server as a part of OGC WMT II and OWS-I.

References

[1] Ian Foster, Carl Kesselman and Steven Tuecke. The Anatomy of the Grid – Enabling Scalable Virtual Organizations. Intl. J. of High Performance Computing Applications, 15(3), 200-222, 2001.

[2] Ian Foster, Carl Kesselman, Jeffrey M. Nick and Steven Tuecke. The Physiology of the Grid: An open Grid services architecture for distributed systems integration. Feb. 17, 2002.
[3] Globus Toolkit. http://www.globus.org, Jun. 2003.
[4] Liping Di, Wenli Yang, Meixia Deng, D. Deng and K. McDonald. The prototype NASA HDF-EOS Web GIS Software Suite (NWGISS), Proceedings of the NASA Earth Science Technologies Conference. Greenbelt, Maryland. August 28-30, 2001 (CD-ROM, 4pp).
[5] OGC. OGC Interoperability Program. http://www.opengis.org/ogcInterop.htm, Mar. 2003.
[6] CEOS-Grid Task Team. V1.0 CEOS-Grid Interim Task Team Project Plan, Sep. 20 2002.
[7] CEOS Grid Tech Wiki. http://grid-tech.ceos.org/gridwiki, Feb. 2003.
[8] Liping Di and Ken McDonald. 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, p92-101. Nov. 1999.
[9] Liping Di, Wenli Yang, Meixia Deng, D. Deng and K. McDonald. Interoperable Access of Remote Sensing Data through NWGISS, In Proceedings of IGARSS 2002. Toronto, Canada. Jun. 2002.
[10] The Metadata Catalog Service (MCS). http://gaul.isi.edu/mcs/, Jun. 2003.
[11] The Replica Location Service (RLS). http://www.globus.org/rls/, Jun. 2003.
[12] NASA. Proposed ECS Core Metadata Standard Release 2.0. http://edhs1.gsfc.nasa.gov/waisdata/docsw/pdf/tp4200105.pdf, Dec. 1994.
[13] OGC. The OpenGIS™ Abstract Specification, Topic 11: OpenGIS(tm) Metadata (ISO/TC 211 DIS 19115). http://www.opengis.org/techno/abstract/01-111.pdf, May 2001.
[14] FGDC. Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata. http://www.fgdc.gov/standards/status/ csdgm_rs_ex.html, Dec. 2002.
[15] OGC. Introduction to OGC Web Services. A. Doyle and C. Reed eds. 2001. http://ip.opengis.org/ows/index.html.
[16] OGC. Web Coverage Service (WCS), versions 0.5, 0.6, 0.7. J. Evans eds. 2000, 2001, 2002. http://www.opengis.org/.