Serving NASA EOS Data to the GIS Community Through the OGC-standard Based NWGISS System

Kenneth R. McDonald
NASA Goddard Space Flight Center
Greenbelt, Maryland 20771, USA
Kenneth.R.McDonald@nasa.gov

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

Abstract
The NASA Earth Science Enterprise (ESE) is generating more than 2 Tb of remote sensing data per day through its Earth Observing System (EOS). Those data are widely used in the global change research. However, the potential of EOS data in GIS applications has not been fully explored. The major problem is that it is very difficult currently to obtain the needed data from EOSDIS and preprocess them into a form ready to be analyzed by users' in-house GIS for analysis. Therefore, the development of capabilities for delivering GIS-ready EOS data directly to a variety of users' GIS systems through the Internet, based on users' requirements will greatly enhance the interoperability and increase public use of EOS data. In the past several years, NASA ESE has funded several research projects to reduce the difficulty of accessing and using EOS data. One of these projects is the development of a NASA HDF-EOS Web GIS Software Suite (NWGISS). NWGISS is a client and server system, compliant with multiple OGC specifications, that provides standards-based access and services to NASA EOS data for the GIS user community. This paper introduces the NWGISS functions and components, its integration with Grid technology for the purpose of sharing data and computing resources among NASA data centers, and the planned implementation of geospatial web services.

Key Words: NWGISS, Web GIS, Web Service, EOS, Data, Remote Sensing

I. Introduction
Geospatial data is the dominant form of data in terms of data volume. It is estimated that more than 80% of data that human beings have collected are the geospatial data. Geospatial data has been widely used in all aspects of human activities, from environmental monitoring to military actions. Because of the importance of geospatial data, both the public and private sectors around the world have invested significant resources to collect geospatial data and a huge volume of such data has been collected. The majority of geospatial data is collected by remote sensing in imagery form. It is estimated that total amount geospatial data in the archives around the world has approached exabytes. With the rapid advance in sensor technologies, the ability to collect

geospatial data through remote sensing has increased significantly. How to make geospatial data easily accessible is one of the major tasks in geospatial data engineering.

As one of the contributions to U.S. national global change research, since late 1980's NASA has been developing the Earth Observing System (EOS), an integrated, multi-satellite, long-term program designed for monitoring Earth's land, oceans, and atmosphere as an integrated system. Currently, EOS is generating more than 2 Tb of remote sensing data per day. The data collected by EOS are processed, distributed, archived, and managed by the EOS Data and Information System (EOSDIS), distributed at nine Distributed Active Archival Centers (DAACs). Most of the data are in HDF-EOS, the standard format for EOSDIS.

Although EOS data are intended to be used for studying global climate and environmental changes, they have been increasingly used in local and regional applications. There are two typical groups of EOS data users. The first group of users is a small number of well-funded scientists who require a huge amount of data for their research and have significant resources in-house to process large volumes of HDF-EOS data. For this group of users, EOSDIS provides special access interfaces with a fast dedicated network connection. The second group of users is a large number of small data users with limited resources. They typically use a GIS to analyze data from EOS for a variety of applications. For most of them, obtaining the needed data from EOSDIS and preprocessing them into a form ready to be analyzed by their in-house GIS represents a major technical challenge that takes a significant amount of their time and resources. In addition to those two groups of existing users, there are many more potential users who would like to use geospatial information in their applications but don't have either the resources or the expertise to derive information from the input data available at the data centers. Therefore, the development of capabilities for delivering GIS-ready EOS data directly to a variety of users' GIS systems through the Internet based on users' requirement, and for providing services that makes user-specific products on demand will greatly enhance interoperability and increase the public use of EOS data. This paper describes a NASA-funded project to develop such capabilities in a system called NASA Web GIS Software Suite (NWGISS). The paper first presents the functionality and components of the current system. Then the paper discusses two efforts on extending NWGISS, the integration of NWGISS with Grid technology and the geospatial Web service middleware development.

II. The Current NWGISS Components and Their Functionalities

Currently, NWGISS consists of following components: a web map server, a web coverage server, a catalog server, a web registries server, a multi-protocol geoinformation client (MPGC), and a toolbox. All NWGISS components can work both independently and collaboratively.

In order for as many as possible users to access data served by NWGISS servers, several Open GIS Consortium (OGC) interface specifications have been implemented. The OGC interface specifications implemented in NWGISS servers include the Web Map Service (WMS) [de La Beaujardière, 2002], Web Coverage Service (WCS) [Evans, 2002], Web

Registries Service (WRS) [Reich, 2001], and Catalog Inter-operability Specification (CIS). On the client side, WRS, WCS, WMS, and Web Feature Services (WFS) [Vretanos, 2002] specifications have been implemented. Figure 1 shows the overall architecture of NWGISS. The following paragraphs provide brief descriptions of each component. The detailed information can be found in [Di, et al, 2001].

The Web Map Server: The map server enables GIS clients to access HDF-EOS data as maps. Currently NWGISS map server complies with OGC WMS version 1.1.0 [La Beaujardière, 2002]. 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.

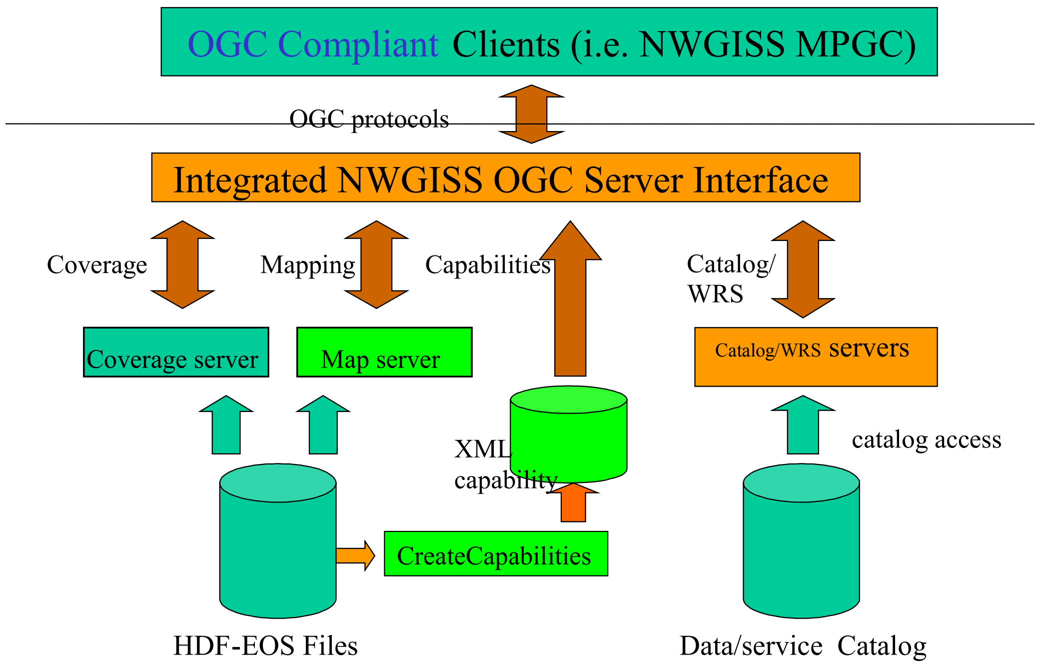

Figure 1. The high level architecture of NWGISS at present

The Web Coverage Server: OGC Web Coverage Service (WCS) specification is designed to enable GIS clients to access multi-dimensional, multi-temporal geospatial data from WCS servers. WCS defines three interface protocols, namely getCapabilities, getCoverage, and describeCoverageType. The NWGISS coverage server includes implementations of the 0.5, 0.6, and 0.7 versions of the draft WCS specification. Three formats for returning coverage are supported in NWGISS, namely, HDF-EOS4 [NASA, 2002], GeoTIFF [Ritter and Ruth, 1999], and NITFF [NIMA, 1999]. Currently we are implementing the WCS version 1.0 that is the first official version of the WCS specification.

The 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, becomes very large. The catalog server allows GIS clients to search and find available geographic data and services in a NWGISS site based on OGC catalog interoperability specification version 1.0 (CIS) [OGC, 2000]. Both state-full and state-less OGC CIS have been implemented in the NWGISS catalog server, which reuses part of Data and Information Access Link (DIAL) catalog server [Di et al., 1999].

The Web Registries Server: The draft OGC Web Registries Service implementation specification defines the interfaces for accessing both data and service catalogs. It also allows one to search for data based on available services and to search available services for a given data product. This specification is one of the keys for service discovery in the web service environment. Currently we have implemented a prototype WRS server as a part of NWGISS that uses an Oracle database at the backend.

The Multi-protocol Geoinformation Client (MPGC): MPGC is a comprehensive OGC client. Currently the OGC WRS, WMS, WFS, and WCS have been implemented in the client. For the coverage access, the client can interactively communicate with all OGC-compliant coverage servers (not only with NWGISS) to access multi-dimensional geospatial data and handles HDF-EOS, GeoTIFF and NITF coverage encoding formats. Besides performing basic WCS client-server communication, coverage access, visualization, and user interaction, the client also provides georectification, reprojection, and reformatting functions. The execution of those functions is automatically arranged based on the user's data requirement and the information about the data in servers. The interaction between NWGISS WCC and OGC compliant web coverage servers provides interoperable, personalized, on-demand data access and services for remote sensing data. In addition to the coverage access, the client can also access feature and map data and overlay them with coverage for data analysis [Deng et al., 2002]. All data accessed are stored and managed in a “project” folder for analysis later. Currently we are implementing WCS 1.0 in the client.

Toolbox: It contains tools for automated data ingestion and catalog creation. Currently, two types of tools are provided: the format conversion tools [LAITS, 2002] and XML capabilities creation tools. A third type of tools, the catalog creation tools, will be provided later.

III. The Integration of NWGISS with Grid Technology

Grid is a rapid developing technology, originally motivated and supported by science and engineering projects requiring high-end computing, for sharing geographically distributed high-end computing resources [Foster, 2001]. The vision of the Grid is to enable resource sharing and coordinated problem solving in dynamic, multi-institutional virtual organizations. It provides on-demand, ubiquitous access to computing, data, and services and constructs new capabilities dynamically and transparently from distributed services. With grid technology, new applications such as distributed collaboration, distributed data access and analyses, and distributed computing, are enabled by the coordinated use of geographically distributed resources. Figure 2 shows the Grid architecture.

Currently, dozens of major Grid projects around the world in scientific and technical computing for research and education have been either deployed for operational use or as technical demonstrated. Considerable consensus on key concepts and technologies has been reached. The key for the GRID success is the open source middleware called Globus Toolkit. It has become a de facto standard for all major Grid implementations. Although far from complete or perfect the Grid technology is out there, evolving rapidly, and has large tool/user base. The Global Grid Forum is a significant force that coordinates the development of the technology in the world.

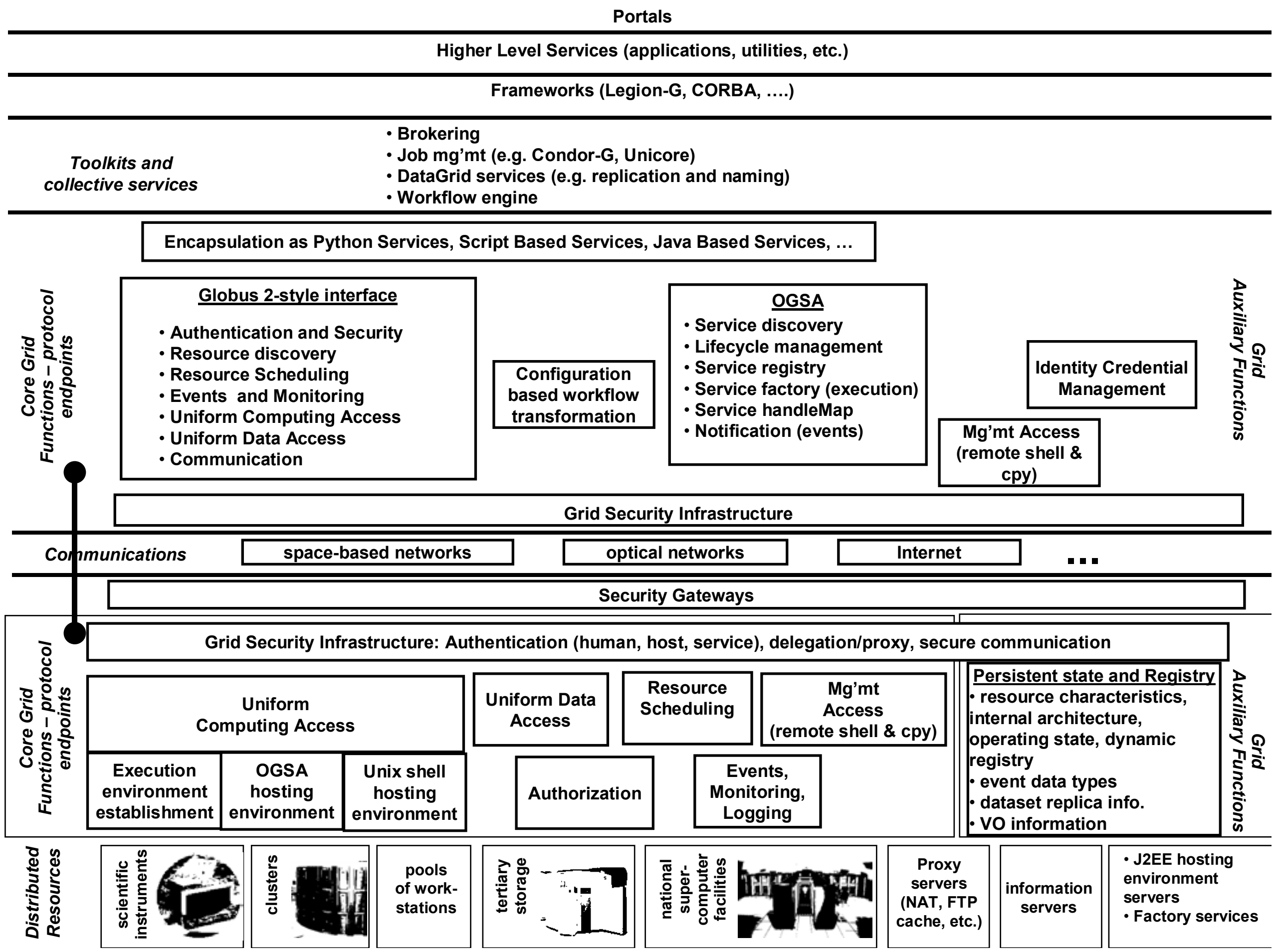

Figure 2. The Grid Architecture

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 geospatial 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.

Currently, we are working on the Grid technology with NWGISS so that all OGC standard compliant GIS clients can access Grid-managed geospatial data. The main work of this project is to integrate Grid and OGC technologies. The integration is taking place between the backend of the NWGISS OGC servers and front-end of data Grid services. The key is to make Grid-managed data accessible through NWGISS OGC servers. Figure 3 shows the proposed architecture of the integrated system. More detailed information about the integration can be found at Di et al, 2003.

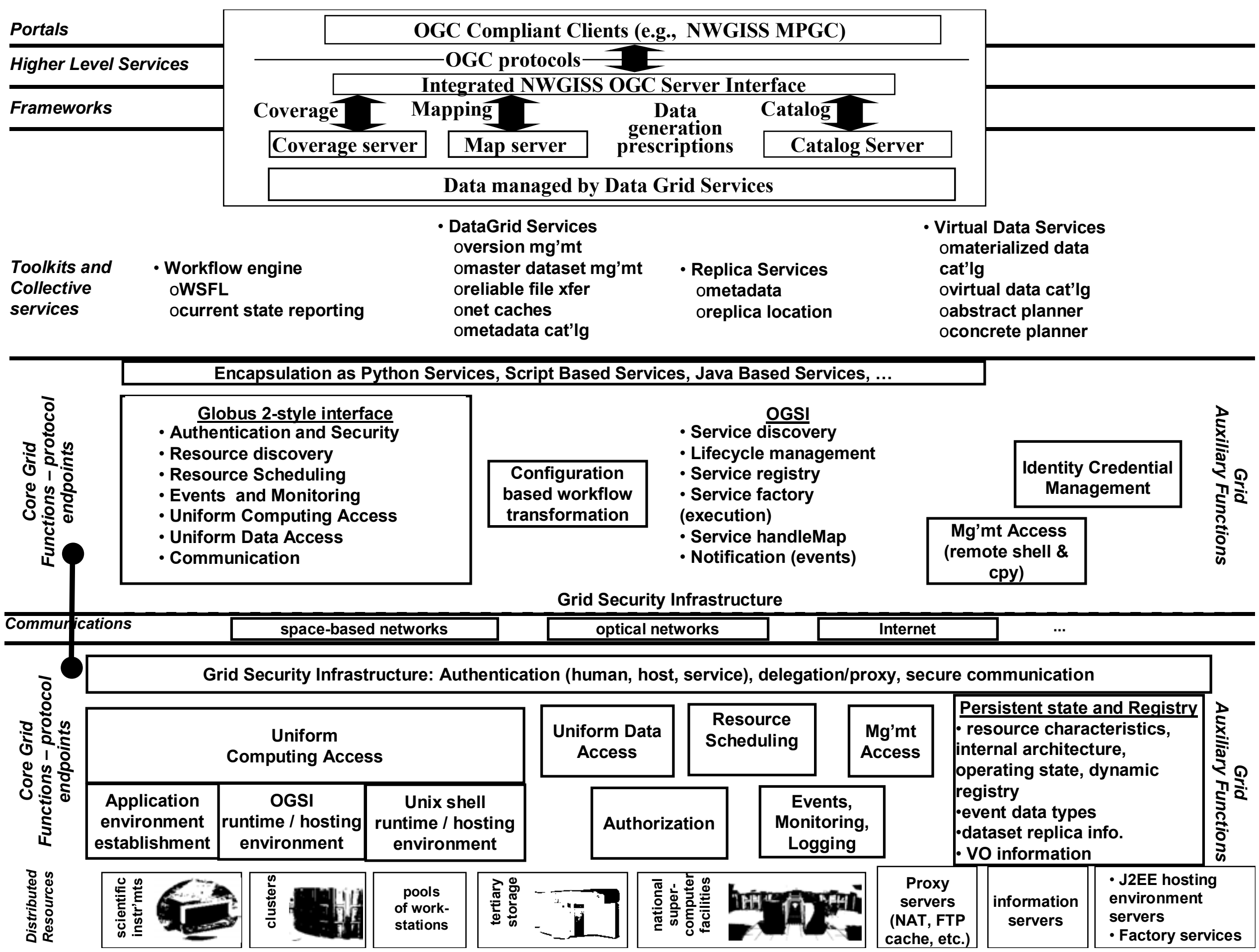

Figure 3. The integrated architecture of a geospatial-enabled, OGC-compatible Data Grid.

IV. The Development of Geospatial Web Services in NWGISS

Currently NWGISS servers provide the standard interfaces for accessing remote sensing data. Combined with the multi-protocol geoinformation client (MPGC), NWGISS allows data users to get data from data centers in the form that exactly matches the users' requirement. However, in many cases, the data products or information the user requests may not be readily available at the data centers. The traditional way to solve this problem is for the user to obtain all necessary raw data and then process the raw data locally to obtain the desired information. However, not all users have the capability or resources to do that and this has significantly limited the wider use of geospatial information. In order to solve this problem, we are developing interoperable geospatial service technologies. The development is based on the geo-object and geo-tree concept described in Di and McDonald 1999 and is implemented in the OGC web service environment [Lieberman 2003]. Figure 4 shows the architecture for the services extensions.

As discussed in the previous section, currently NWGISS is a two-layer client-server system. The servers will directly talk to HDF-EOS data in the archives to provide on-demand data access. As an extension, we will develop a middle service layer so that the NWGISS will become a three-tier system. The three tiers are the interoperable data server tier, the middleware geospatial web services and knowledge management tier, and the integrated geoinformation client tier. In the figure, all standards-based system-wide interfaces are shown as white arrows. Any other systems using the same standard interfaces as NWGISS can be interoperable and could be federated with NWGISS. All other types of arrows represent either private/internal interfaces or user community defined interfaces. The size of arrows represents the size of data traffic. Since the system provides customized data product and information to the end-users by performing the data reduction and processing at the data sever and middleware tiers, the size of data traffic at the end-users side is much smaller than those at data server side.

The interoperable data provider tier. This tier consists of data servers providing data to the geospatial service middleware, the application & data analysis systems, application clients, and human users (called *requestors* hereafter) through a common data environment. The NWGISS data servers are used in data provider's tier. *The common data environment* is a set of standard interfaces for finding and access data in diverse data archives, ranging from small data providers to multiple-petabyte NASA EOS data archives. The environment allows geospatial services and value-added applications to access diverse data provided by different data providers in a standard way without worrying about their internal handling of data.

The interface standards for the common data environment are OGC Web Data Services Specifications, including Web Coverage Services (WCS), Web Feature Services (WFS), Web Map Services (WMS), and Web Registries Services (WRS). The specifications allow seamless access to geospatial data in a distributed environment, regardless of the format, projection, resolution, and the archive location.

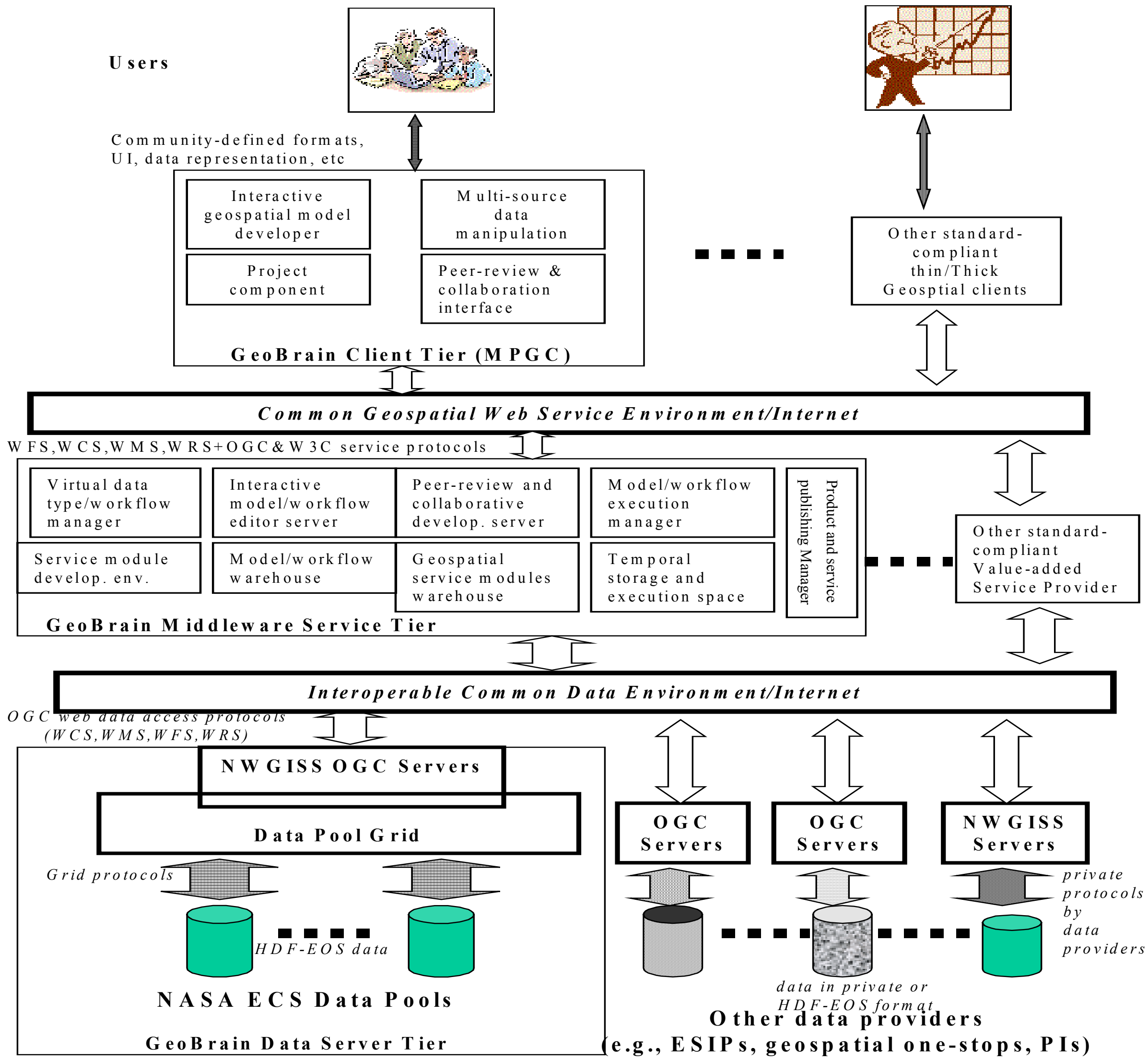

Figure 4. The Web Service Extensions of NWGISS

The middleware geospatial service and knowledge management tier. The middleware tier consists of multiple components that perform geospatial data processing, information extraction, and knowledge management. This tier includes the following components:

- *A Geospatial service module warehouse* containing individual geospatial web services executables.
- *A Geospatial model/workflow warehouse* containing geo-trees that describe geospatial models encoded in a workflow language. The geo-trees capture the knowledge of the geospatial process and modeling.
- *Virtual data type/workflow manager*: This component serves two purposes: to external users it is an OGC WRS server that helps requestors find both data instances and data types. Internally, it manages the model warehouse, working as the client of WRS servers.

- *The model/workflow execution manager* acts as a WCS, WFS, or WMS server, depending on the type of geoinformation requested. If the requested geoinformation is non-virtual, the manager will retrieve the product from the data provider's server and deliver it to the requestor. If it is virtual, the manger will manage the execution of the workflow and deliver the materialized virtual geoinformation to the requestor.
- *Interactive model/workflow editor server* provides the interactive modeling environment allows users to construct and test their models through the NWGISS client. Models will be constructed graphically in the client. The editor server will provide all available web service modules classified by their service categories, and both virtual data types represented as workflows and real data types to the client. The editor server will allow the user to instantiate, run, modify, debug, save the model and submit it to peer-review server for review.
- *Service module development environment* encourages end-users to develop additional service modules by providing the service module development environment. The development environment includes a set of libraries for handling the interface protocols, data encoding and decoding, and general utilities functions. By using those libraries, web-service modules developed by users will be standards-compliant and interoperable.
- *Peer-review and collaborative development server*. Any user developed web-service modules and the geospatial models are subject to peer review before being inserted into NWGISS as the operational capabilities. The peer-review server will facilitate the processing by providing the common environment for reviewers to run and evaluate the submission through the MPGC.
- *Product and service publishing interface.* This component will publish the product and services available to external registries so that they also can be found though the registries.

The integrated multiple-protocol geoinformation client (MPGC) tier. The client provides not only access to all virtual and real data/information provided by NWGISS servers and all other OGC-compliant providers but also the geospatial modeling/ workflow interfaces, peer-review interfaces, and the collaborative development interfaces.

V. Conclusions

Currently NWGISS has been used in many different testbeds and technology evaluation programs including OGC Web Service (OWS) Initiatives, the CEOS WGISS Test Facilities (WTF), and NASA EOSDIS data pools. Those tests show that NWGISS approach can significantly improve the access to NASA EOSDIS data for GIS user community. The current work on the further development of NWGISS will definitely enhance its functionality and applicability. The Grid extension will enable NWGISS to work in Grid environment for construction of the geospatial information Grid. The Web service middleware development will enable many potential users, who are not able to use EOS data currently, to use EOS data in their applications.

VI. Acknowledgements

This project has been supported by research grants from the NASA Earth Science Data and Information Project (ESDISP), the NASA Earth Science Technology Office (ESTO), and the Open GIS Consortium (OGC).

VII. References

Deng, D., M. Deng, P. Zhao, and L. Di, 2002. "A Thick Web Coverage Client (WCC) of NASA HDF-EOS Web GIS Software Suite (NWGISS)." In Chen and Jiang, eds, *The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Vol. XXXIV, Part 2, Commission II, p 87-110.

Di, L., A. Chen, W. Yang, and P. Zhao, 2003. "The Integration of Grid Technology with OGC Web Services (OWS) in NWGISS for NASA EOS Data," Proceedings of the Eighth Global Grid Forum (GGF8), 24-27 June 2003, Seattle, WA, USA. 10pp.

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.

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.

Di, L., R.Suresh, K. Doan, D. Ilg, and K. McDonald, 1999. "DIAL-an Interoperable Web-based Scientific Data Server". In M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman edit, *Interoperating Geographic Information Systems*, Section 4. System Experiences. Kluwer Academic Publishers, Norwell, MA. pp 301-312.

Evans, J. eds, 2002. Web Coverage Service Implementation Specification, OpenGIS® Project Document 02-024, http://www.opengis.org/techno/discussions/02-024.pdf.

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

J. de La Beaujardière, eds, 2002. "Web Map Service Implementation Specification, Version 1.1.0," http://www.opengis.org/techno/specs/01-047r2.pdf

LAITS, 2002. HDF-EOS GIS translators, http://heineken.gsfc.nasa.gov/eosgis/eghome.html

Lieberman, J., editor, 2003. OpenGIS Web Services Architecture. OGC 03-025. http://www.opengis.org/techno/discussions/03-025.pdf.

NASA, 2002. HDF-EOS4(HE4), http://hdfeos.gsfc.nasa.gov/hdfeos/he4.cfm

NIMA, 1999. National Imagery Transmission Format Version 2.1, National Imagery and Mapping Agency (NIMA), Bethesda, Maryland.

OGC, 2000. - Catalog Interface Implementation Specification (version 1.0), http://www.opengis.org/techno/specs/99-051.pdf

Reich, L. editor, 2001. Web Registry Server Discussion Paper. OpenGIS Project Document 01-024r1. http://www.opengis.org/techno/discussions/01-024r1.pdf.

Ritter N., and M. Ruth, 1999. GeoTIFF Format Specification GeoTIFF Revision 1.0, http://remotesensing.org/geotiff/spec/geotiffhome.html

Vretanos, P. A. editor, 2002. Web Feature Service Implementation Specification, Version 1.0.0 OGC 02-058. http://www.opengis.org/techno/specs/02-058.pdf.