Discover, Visualize, and Deliver Geospatial Data through OGC Standards-based WebGIS System

Yaxing Wei, Suresh-Kumar Santhana-Vannan, Robert B. Cook
Environmental Sciences Division (ESD), Oak Ridge National Laboratory (ORNL)
P.O. Box 2008, MS 6407, Oak Ridge, TN 37831, USA
weiy@ornl.gov, santhanavans@ornl.gov, cookrb@ornl.gov

***Abstract*—Geospatial data are important to understand the Earth – ecosystem dynamics, land cover changes, resource management, and human interactions with the Earth to name a few. One of the biggest difficulties users face is to discover, access, and assemble distributed, large volume, heterogeneous geospatial data to conduct geo-analysis. Traditional methods of geospatial data discovery, visualization, and delivery lack the capabilities of resource sharing and automation across systems or organizational boundaries. They require users to download the data "as-is" in their original file format, projection, and extent. Also, discovering data served by traditional methods requires prior knowledge of data location, and processing requires specialized expertise. These drawbacks of traditional methods create additional burden to users, introduce too much overhead to research, and also reduce the potential usage of the data. At the Oak Ridge National Laboratory (ORNL), researchers working on NASA-sponsored projects: Distributed Active Archive Center (DAAC) and Modeling and Synthesis Thematic Data Center (MAST-DC) have tapped into the benefits of Open Geospatial Consortium (OGC) standards to overcome the drawbacks of traditional methods of geospatial data discovery, visualization, and delivery. The OGC standards-based approach facilitates data sharing and interoperability across network, organizational, and geopolitical boundaries. Tools and services based on OGC standards deliver the data in many user defined formats and allow users to visualize the data prior to download. This paper introduces an approach taken to visualize and deliver ORNL DAAC, MAST-DC, and other relevant geospatial data through OGC standards-based Web Services, including Web Map Service (WMS), Web Coverage Service (WCS), and Web Feature Service (WFS). It also introduces a WebGIS system built on top of OGC services that helps users discover, visualize, and access geospatial data.**

Keywords—OGC, WMS, WCS, WFS, Geospatial Data, WebGIS, ORNL DAAC, MAST-DC

I. INTRODUCTION

Geospatial data are an important source for people to learn and understand the Earth. Researchers analyze geospatial data to extract information and derive knowledge about the Earth system. With the rapid advancement of computing and network technologies, geospatial data have become more and more important in many areas. The usefulness of geospatial data has been demonstrated across many diverse disciplines, such as meteorology, global climate change, agriculture, forestry, flood monitoring, fire detection and monitoring, and geology [1].

Processing in geospatial knowledge discovery includes three consecutive steps: data query, data and information assembly, and computation [2]. Data query is to locate and obtain required data from data repositories. Data and information assembly organizes the data and information from distributed data centers based on the needs of computation. Computation is to analyze and simulate the complex Earth system using data and information assembled from various data repositories. Ideally, researchers and users should only focus on computation. But in practice, nowadays, users spend much time and effort to query, access, and process geospatial data that can then be used to perform the required computation. It is estimated that more than 50% of users' time is spent on the query and assembly steps of the geospatial knowledge discovery [3]. Actually, the query and assembly tasks are duplicated efforts for users and dramatically reduce the efficiency of knowledge discovery. These problems arise from the following challenges:

- Geospatial data are voluminous. Diverse groups of users are realizing the importance of geospatial data, and many organizations and countries have started to collect geospatial data using a wide range of methods, including remote sensing, ground-based sensors, aircraft, and simulation models. Geospatial data can cover wide spatial extent and multiple time steps or extremely fine resolution images. As sensors, computer processing speed, and storage capacity all increase dramatically, the volume of geospatial data collected keeps growing rapidly. For example, NASA EOSDIS [4] has already archived more than 2 petabytes of data, with more than 3 terabytes of new data arriving each day. The huge volume of geospatial data introduces much overhead in data query, transfer, and pre-processing.
- Geospatial data are geographically distributed. Geospatial data's multi-disciplinary characteristic, its ownership issue, and its provenance make it impossible for a single organization or a data center to gather and manage all the geospatial data together. For example, NASA's Earth Science Data Systems is highly distributed and contains 11 thematic data centers and 14 Science Investigator-led Processing Systems (SIPS) as the core part and about 100 nodes in the community part [5].

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 24,2010 at 10:03:55 EDT from IEEE Xplore. Restrictions apply.

- Geospatial data are highly heterogeneous. The data file format, metadata schema, coordinate system, and many other aspects of geospatial data vary across different organizations and research areas.
- Diversity in geospatial data discovery and access methods. Because geospatial data are held by many organizations for a huge array of purposes, data centers use many different ways to store, publish, and deliver data.
- Complex geospatial problems need large quantities of geospatial data from multiple sources and locations. For example, a statistical wildfire forecast model at 8-km spatial resolution for the conterminous USA used over 1 terabytes of data obtained from measurements of the NASA Earth Observing System satellites and daily weather data provided by National Oceanic and Atmospheric Administration (NOAA) National Climate Data Center (NCDC) [6].

Besides the above challenges faced by geospatial data users, there are also challenges faced by data centers that archive, manage, and distribute geospatial data:

- Data centers need to fight with data loss and keep data longevity. Among the various reasons that cause data loss and reduce data longevity, data format obsolescence is a major issue.
- Data centers are facing a bigger and bigger user community. Users' requirements on geospatial data are diverse.
- Heterogeneity of the ways in which data centers publish and deliver geospatial data creates a big obstacle for the collaboration of data centers. Isolation of a data center is an obstacle for data sharing and can greatly reduce the usage of geospatial data.

With the growing of geospatial data user community, data centers need to find new solutions to meet users' diverse demands, reduce data loss, and make full use of the geospatial data they manage. Data centers also need to collectively improve the interoperability of their geospatial data and provide users a universal interface for geospatial data discovery and access.

At the Oak Ridge National Laboratory (ORNL), researchers working on NASA-sponsored projects, Distributed Active Archive Center (DAAC) and Modeling and Synthesis Thematic Data center (MAST-DC) have tapped into the benefits offered by Open Geospatial Consortium (OGC) standards to address the user and data center problems described above.

This paper first introduces ORNL DAAC and MAST-DC, analyzes the characteristics of the data archived in these two data centers. It then describes OGC standards, including Web Coverage Service (WCS), Web Map Service (WMS), and Web Feature Service (WFS), and how those standards are used to address user and data center problems. The paper then introduces a Web-based geospatial data visualization and download system built on top of OGC services.

II. ORNL DAAC AND MAST-DC

A. *ORNL DAAC*

The Oak Ridge National Laboratory Distributed Active Archive Center[1] (ORNL DAAC) is a NASA-sponsored source for biogeochemical and ecological data and models useful in environmental research. It archives geospatial data collected on the ground, by aircraft, by satellite, and from computer models. The extent of archived data and model products ranges from site specific to global, and durations range from days to years.

The ORNL DAAC has a unique role within NASA EOSDIS. Rather than serving as a repository for large volumes of data from a small number of remote sensing missions, it serves as a repository for small, disparate sets of data from principal investigators of field projects sponsored by NASA and other federal agencies and international organizations. Examples of the data sets archived in ORNL DAAC include Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA), SAFARI2000, EOS Land Validation, Net Primary Production (NPP). These ground, aircraft, and modeled data are critical for understanding biogeochemical dynamics. The ORNL DAAC is also a crucial link in the validation of biogeochemical inferences from remotely sensed data.

Currently, there are more than 800 data sets and about 220,000 data granules in many different file formats archived at the ORNL DAAC. The total archived data volume is about 1.1TB. A major issue is that many of the data are archived in variety of file formats, such as ASCII and binary formats. These heterogeneous file formats introduce additional processing burden for users and is an obstacle for geospatial data automation and interoperation.

Currently, the ORNL DAAC users are provided a metadata catalog system to find data, an online data order system to place data orders, and an FTP site to retrieve data and orders. It also provides several data visualization and delivery tools, such as an online ArcIMS-based GIS system and Thematic Real-time Environmental Distributed Data Services (THREDDS) Data Server. Recently we have added some OGC-based Web Services to customize the data before download, such as choosing different file format, projection, and spatial/temporal extent. In addition we have also added a standards-based user interface so that ORNL DAAC's data holdings can be easily integrated within the whole geospatial community. This paper describes the OGC-based Web Services and the user interface.

B. *MAST-DC*

The Modeling and Synthesis Thematic Data Center [2] (MAST-DC) is a component of the data system of the North American Carbon Program (NACP) data system designed to support data products and data management services needed for modeling and synthesis activities. The overall objective of MAST-DC is to provide data management support for NACP project and provide data to users and in the formats, resolutions, and projections based on their requirements.

[1] ORNL DAAC, http://daac.ornl.gov
[2] MAST-DC, http://nacp.ornl.gov

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 24,2010 at 10:03:55 EDT from IEEE Xplore. Restrictions apply.

As part of regional synthesis activities, we have collected 25 model simulation results at MAST-DC. Those model simulation results are raster spatial data layers. They include multiple variables for carbon cycle research, like Gross Primary Production (GPP), Net Ecosystem Exchange (NEE), and Respiration. Besides model simulation data, there are also observation and measurement data to be used for model-observation inter-comparison and model evaluation, including crop NPP, forest biomass, and MODIS-derived GPP, NPP, LAI, FPAR, NDVI, and EVI. Most of those data span represent a time period from year 2000 to 2005 in monthly or finer temporal resolution.

The original data of the model output and observation varies in file formats, spatial/temporal resolution and extent, projections, and data units. For example, there are HDF-EOS, netCDF, Binary, and ASCII file formats. Spatial resolution varies from finer 250 meter resolution to coarser 2.8 degree resolution. Each variable may contain multiple time steps.

MAST-DC is converting these heterogeneous data products into consistent and standard file formats, projections and resolutions to provide a standard data access interface to modelers, regional/site-level comparison researchers, and general users.

III. OGC SERVICES STANDARDS

OGC is an international consortium of companies, government agencies, research organizations, and universities participating in a consensus process to develop publicly available interface specifications. The objective of OGC standards is to promote geospatial information interoperability. OGC services standards provide a framework for developers to create applications that enable users to access and process geographic data from a variety of sources across a generic computing interface within an open information technology environment [7].

OGC services standards are loosely arranged into 4 tiers: Clients, Application Services, Processing Services, and Information Management Services tier. Among the 4 tiers, the Information Management Services tier contains services designed to store and provide access to geospatial data. It is the most important tier to the geospatial data visualization and delivery, and includes OGC standards such as WCS, WFS, and WMS.

A. *Web Coverage Service (WCS)*

The Web Coverage Service (WCS) supports electronic interchange of geospatial data as "coverages" – that is, digital geospatial information representing space-varying phenomena that relate a spatio-temporal domain to a range of properties. A WCS provides access to detailed and rich sets of geospatial information, in forms that are useful for client-side rendering, multi-valued coverages, and input into scientific models and other clients [8]. At the ORNL DAAC and MAST-DC, WCS standard is used to distribute gridded data files like LBA Land Use/Land Coverage Change data and ecosystem model simulation data.

B. *Web Feature Service (WFS)*

The Web Feature Service (WFS) provides an interface allowing access to discrete geospatial feature data encoded in Geography Markup Language (GML) [9]. Unlike coverages, geospatial features are described by a set of properties where each property can be thought of as a {name, type, value} tuple. The vector data files archived in ORNL DAAC such as FLUXNET site location information, Global River Discharge (RivDIS) field data fall into this category and can leverage WFS for data distribution.

C. *Web Map Service (WMS)*

Unlike WCS or WFS, which provide access to geospatial data itself, a Web Map Service (WMS) produces styled maps of georeferenced data, including both coverage and feature data. A "map" is defined as a visual representation of geospatial data [10]. WMS can be used to visualize the data before downloading.

As indicated by their names, OGC WCS/WFS/WMS are Web-based. Communication between OGC WCS/WFS/WMS services and their clients uses only non-proprietary Internet standards, that is, the OGC Web Services (OWS) architecture uses standard World Wide Web (WWW) protocols, namely HTTP GET and HTTP POST. Specific server operations are addressed using Uniform Resource Locators (URLs). Multipurpose Internet Mail Extensions (MIME) types are used to identify data transfer formats. The data transferred is often encoded using the Extensible Markup Language (XML), with the contents and format carefully specified using XML Schemas. Each of the OGC services contain a common mandatory operation called "GetCapabilities", which returns a standard-based XML file describing the capabilities and contents provided by a server. The self-descriptive nature of the OGC services can be used to build standard interfaces for geospatial data access in data centers.

IV. OGC STANDARDS-BASED GEOSPATIAL DATA VISULIZATION AND DOWNLOAD SYSTEM

A. *The ORNL WCS/WFS/WMS Servers*

At ORNL, several servers conforming to OGC WCS/WFS/WMS standards have been developed and deployed for discovery, visualization, and delivery of geospatial data managed in DAAC and MAST-DC. The ORNL DAAC and MAST-DC OGC WCS/WFS/WMS servers are developed using MapServer[3] – an open source and customizable mapping server.

The ORNL WCS server[4] conforms to OGC WCS standard version 1.0.0. It supports three mandatory operations: GetCapabilities, DescribeCoverage, and GetCoverage [8]. Through GetCapabilities operation, a client can get service metadata and a list of available coverages. Then through DescribeCoverage, a client can get detailed metadata

[3] MapServer, http://mapserver.org/

[4] The ORNL WCS server, http://webmap.ornl.gov/ogcbroker/wcs?service=WCS&version=1.0.0&request=GetCapabilities

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 24,2010 at 10:03:55 EDT from IEEE Xplore. Restrictions apply.

information about a specific coverage, like spatial/temporal extent and resolution, supported projections, supported file formats, and supported resampling methods. The ORNL WCS server supports several ways to customize geospatial coverage data, including spatial, temporal, and bands subsetting, reprojection, spatial resampling, and file reformatting. Three interpolation options (nearest neighbor, bilinear, and average) are available for customizing the behavior of reprojection and resampling. Five categories of file formats (GeoTIFF, Arc/Info ASCII Grid, netCDF, NITF, and ERDAS Imagine) are available for data customization. In each of those 5 categories, up to three formats (BYTE, INT16, and FLOAT32) are available based on original data's internal data type.

Right now, the ORNL WFS server[5] is a read-only server, which means it can only support mandatory discovery and retrieval operations (GetCapabilities, DescribeFeatureType, and GetFeature) while not transactional operations (LockFeature, Transaction, and GetFeatureWithLock) [9]. The ORNL WFS server conforms to OGC WFS standard version 1.0.0. Through the GetFeature operation, clients can retrieve part of the features with certain properties from a feature layer using OGC Filtering Encoding Standard (FES). The output format can be either GML2 or GML3.

The ORNL WMS server[6] conforms to OGC WMS standard version 1.1.1. It works together with the ORNL WCS and WFS servers to provide visualization capability for both gridded and vector geospatial data. The ORNL WMS server supports two mandatory operations (GetCapabilities and GetMap) and one optional operation (GetLegendGraphic) [10]. For gridded data, the ORNL WMS server supports visualization of spatial, temporal, and band subsets, and map reprojection. It also supports popular image formats, including JPEG, GIF, and PNG.

The ORNL WCS and WFS servers' data customization capability brings lot of convenience to ORNL DAAC and MAST-DC users. For example, researchers working on Moderate Resolution Imaging Spectrometer (MODIS) validation tasks may use WCS to retrieve both MODIS Land Cover and BigFoot LAI data in same spatial resolution, data format, spatial extent, projection, and at certain time positions. Researchers may also use WFS server to retrieve certain FLUXNET features site information. Even though the MODIS, BigFoot, and FLUXNET data differ in many aspects, OGC WCS and WFS can help researchers to assembly the data for direct use in their analysis. Also, researchers can visualize the geospatial data using WMS before they actually get the data.

The ORNL WCS/WFS/WMS servers and their corresponding data have been registered with several geospatial community registries, like GEOSS, NBII, ECHO, GCMD, and GMU GeoBrain[7]. Through the portals of those registries, users can search and access geospatial data provided by the ORNL WCS/WFS/WMS. Some system can even automate the retrieval of geospatial data without user interaction, like in GeoBrain, the ORNL WCS and WFS can be used as input data sources for geospatial workflows to enable the virtual data products materialization on the fly [11].

The ORNL WCS/WFS/WMS servers support many OGC standards-compatible clients, like Google Earth, NASA World Wind, OpenLayers, uDig, Quantum GIS, and ArcGIS. This helps users to access and use the geospatial data more easily and helps data centers to extend the usage of the geospatial data.

B. OGC Standards-based Geospatial Data Visualization and Download Web Application[8]

Usually OGC servers are not used by end users directly. They are designed to support client-to-server, that is, machine-to-machine communication. To help users realize the benefits brought by the ORNL OGC servers, an OGC standards-based geospatial data visualization and download Web application was developed on top of the ORNL OGC servers. The high-level architecture of this application is shown in figure 1. Boxes in round-corner rectangle shape are newly developed components and boxes in ellipsis shape are existing components in ORNL data centers.

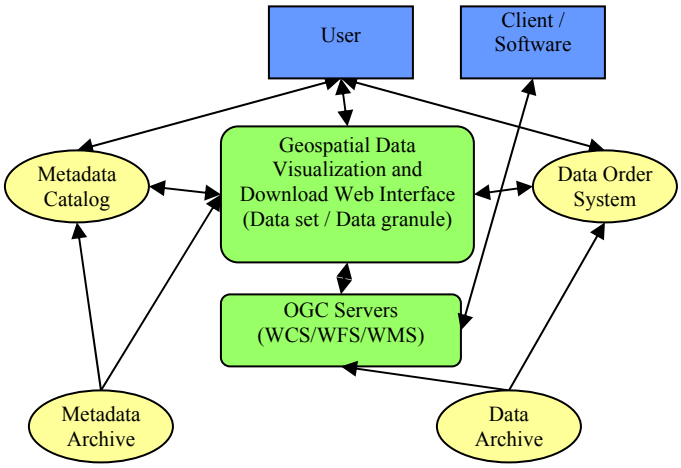

Figure 1. Architecture of OGC Standards-based Geospatial Data Visualization and Download Web Application

Geospatial data are organized into two levels in this Web application: data set and data granule. A data set contains one or more data granules that share some common properties. A dynamic web page, as shown in figure 2, was created for users to view data set level metadata information and browse all the data granules contained in a data set. Another dynamic web page, as shown in figure 3, was created for data granules. This web page contains 3 sections. The first section displays granule level metadata, like spatial extent and available time steps. The second section contains a map widget for users to preview the geospatial data. The preview images are created by the ORNL WMS server. The third section gives users the control to

[5] The ORNL WFS server, http://webmap.ornl.gov/ogcbroker/wfs?service=WFS&version=1.0.0&request=GetCapabilities

[6] The ORNL WMS server, http://webmap.ornl.gov/ogcbroker/wms?service=WMS&version=1.1.1&request=GetCapabilities

[7] GeoBrain, http://geobrain.laits.gmu.edu/

[8] OGC Standards-based Geospatial Data Visualization and Download Web Application, http://webmap.ornl.gov/wcsdown

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 24,2010 at 10:03:55 EDT from IEEE Xplore. Restrictions apply.

customize the geospatial data by setting WCS or WFS parameters.

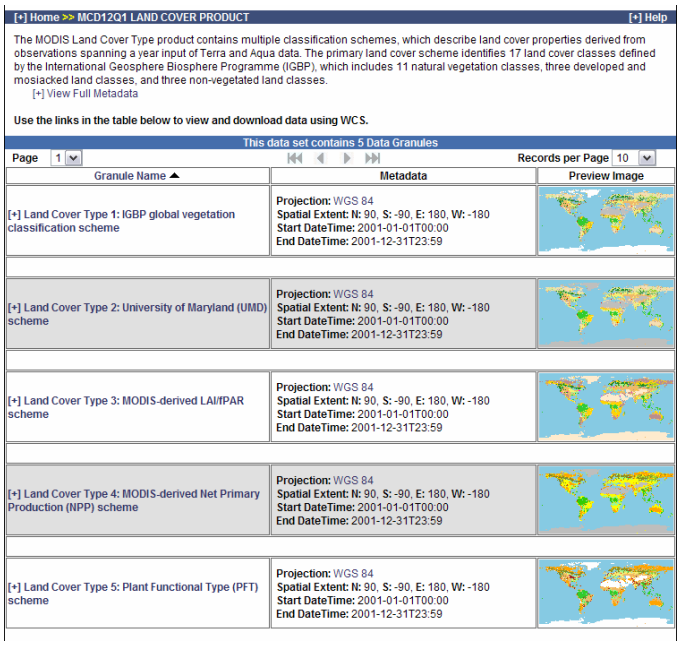

Figure 2. Dynamic Data Set Web Page

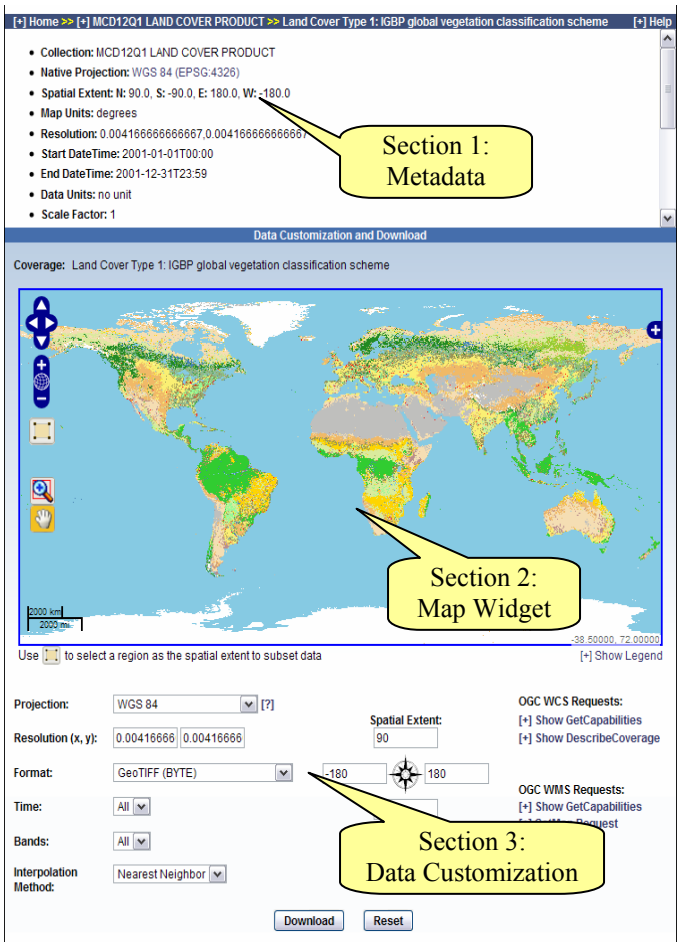

Figure 3. Dynamic Data Granule Web Page

This OGC standards-based geospatial data visualization and download Web user interface can be integrated seamlessly with other systems currently deployed in ORNL DAAC. For example, metadata system can maintain URL links to the data set web pages and data granule web pages in its data set metadata and data granule metadata respectively, so that users can choose customized data download by following those URL links displayed in the searching results. The OGC standards-based geospatial data visualization and download Web user interface can also be integrated with the ORNL DAAC data ordering system. After a user finishes customizing some geospatial data, the user can choose to add them to the ORNL DAAC shopping cart and then retrieve them together with data added directly from the ORNL DAAC archive.

V. CONCLUSION

Geospatial data are valuable to researches in many areas. But the success of today's multi-discipline research activities relies greatly on the instant availability and easy access to geospatial data. OGC's information management services standards (WCS, WFS, and WMS) provide customizable visualization and access to both geospatial coverage and feature data in a standard and easy way. The WCS/WFS/WMS servers and OGC standards-based geospatial data visualization and download Web application developed and deployed in the ORNL DAAC and MAST-DC broaden the way in which the data centers distribute geospatial data to users. By integrated into existing information management system, they help the data centers to meet users' diverse requirements on geospatial data. They also bring the interoperability for the ORNL DAAC and MAST-DC to collaborate with other organizations to form a bigger distributed data warehouse.

ACKNOWLEDGEMENT

This work is part of the ORNL DAAC and MAST-DC projects, which are supported by NASA ESDIS (NNGO4HU031) and NASA-TE (05-NACPO5-12) respectively.

REFERENCES

[1] F. Lamberti and S. Beco, "SpaceGRID - an international programme to ease access and dissemination of Earth Observation data/products: How new technologies can support Earth Observation Users Community", 22nd EARSeL Symposium & General Assembly, Prague, Czech Republic, June 4-6, 2002.

[2] L. Di, "Distributed Geospatial Information Services-Architectures, Standards, and Research Issues", 2004 International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences (ISPRS 2004), Vol. XXXV, Part 2, Commission II, 7p. (Invited), July 12-23, 2004, Istanbul, Turkey.

[3] L. Di and K. McDonald, "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, 1999, pp. 92-101.

[4] K. McDonald, L. Di, and S. W. Wharton, "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.

[5] L. Francis and M. Martha, "NASA's Strategy for Enabling the Discovery, Access, and Use of Earth Science Data for Land Science

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 24,2010 at 10:03:55 EDT from IEEE Xplore. Restrictions apply.

Research: Revisited", ORNL DAAC UWG Meeting, June 2, 2008, Crystal City, VA.

[6] Y. Wei, L. Di, B. Zhao, G. Liao, A. Chen, and Y. Bai. "Sharing of Distributed Geospatial Data through Grid Technology", Handbook of Research on Geoinformatics, Hassan A. Karimi, eds., Information Science Reference, Hershey, PA. USA, 2008, pp. 222-228.

[7] G. Percivall, "OGC's Open Standards for Geospatial Interoperability", Encyclopedia of GIS, S. Shekhar and H. Xiong, eds., Springer, 2008, pp 800-805.

[8] OGC, OGC 05-076, Web Coverage Service, Version 1.0.0, 2005.

[9] OGC, OGC 02-058, Web Feature Service Implementation Speicification, Version 1.0.0, 2002.

[10] OGC, OGC 01-068r3, Web Map Service Implementation Speicification, Version 1.1.1, 2002.

[11] Y. Wei, P. Yue, U. Dadi, M. Min, C. Hu, and L. Di, "Effective Acquisition of Geospatial Data Products in a Collaborative Grid Environment", 2006 IEEE International Conference on Services Computing (SCC 2006), IEEE Computer Society Press, Sept. 2006, pp. 455-462, doi: 10.1109/SCC.2006.46.

Authorized licensed use limited to: Oak Ridge National Laboratory. Downloaded on March 24,2010 at 10:03:55 EDT from IEEE Xplore. Restrictions apply.