GRID COMPUTING ENABLED GEOSPATIAL CATALOGUE WEB SERVICE

Aijun Chen, Assistant Research Professor  
Liping Di, Associate Professor and Director  
Yaxing Wei, Ph.D. Student  
Yang Liu, Ph.D. Student  
Yuqi Bai, Post-doc Research Associate  
Laboratory for Advanced Information Technology and Standards (LATIS)  
George Mason University (GMU)  
9801 Greenbelt Road, Suite 316-317, Lanham, MD 20706  
achen6@gmu.edu  
tlidi@gmu.edu  
ywei@gmu.edu  
yliu5@gmu.edu  
ybai@gmu.edu

Chaumin Hu, Senior Software Engineer  
Piyush Mehrotra, Senior Research Scientist  
NASA Ames Research Center  
M/S 258-6, Moffett Field, CA 94035  
chaumin6@nas.nasa.gov  
piyush.mehrotra@nasa.gov

ABSTRACT

Geospatial Catalogue Web Service is a vital service for sharing and interoperating volumes of distributed heterogeneous geospatial resources, such as data, services, applications, and their replicas over the web. Based on the Grid technology and the Open Geospatial Consortium (OGC)'s Catalogue Service – Web Information Model, this paper proposes a new information model for Geospatial Catalogue Web Service, named as GCWS which can securely provides Grid-based publishing, managing and querying geospatial data and services, and the transparent access to the replica data and related services under the Grid environment. This information model integrates the information model of the Grid Replica Location Service (RLS)/Monitoring & Discovery Service (MDS) with the information model of OGC Catalogue Service (CSW), and refers to the geospatial data metadata standards from ISO 19115, FGDC and NASA EOS Core System and service metadata standards from ISO 19119 to extend itself for expressing geospatial resources. Using GCWS, any valid geospatial user, who belongs to an authorized Virtual Organization (VO), can securely publish and manage geospatial resources, especially query on-demand data in the virtual community and get back it through the data-related services which provide functions such as subsetting, reformatting, reprojection etc. This work facilitates the geospatial resources sharing and interoperating under the Grid environment, and implements geospatial resources Grid enabled and Grid technologies geospatial enabled. It also makes researcher to focus on science, not on issues with computing ability, data location, processing and management. GCWS also is a key component for workflow-based virtual geospatial data producing.

INTRODUCTION

Grid computing has appeared as a new e-science information technology for addressing the formidable challenges associated with the complete integration of heterogeneous computing systems and data resources with the final aim of providing a global computing space with global 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 securely provide advanced distributed high-performance computing to users among one or more Virtual Organizations (VOs) which use the Authority and Authentication security policy (Foster, 2001, 2002). Currently, the most popular and widely used Grid software for Grid researcher and scientist is named as Globus which is provided through The Globus Project. The Globus Project, cooperated with Global Grid Forum (GGF), proposes the Open Grid Service Architecture (OGSA), Open Grid Service Infrastructure (OGSI) for Globus 3.0 and Web Service Resource Framework (WSRF) for Globus 4.0 as the guidelines and specifications of system design and implementation to develop fundamental technologies needed to
build computational Grid (GGF, 2004; Globus, 2004a). Now, Globus provides many functional modules both as Grid Service and non-Grid Service, such as Globus Resource Allocation Manager (GRAM & WS-GRAM) for providing a common user interface to submit a job to dispersed multiple machines, the Monitoring and Discovery Service (MDS & WS-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, GridFTP for providing a standard, reliable, high-speed, efficient and secure data access and transfer service, Metadata Catalog Service (MCS) for providing a mechanism for storing and accessing metadata of 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 or data related service, Reliable File Transfer (RFL) for performing third-party transfers across GridFTP servers, and other modules such as simple Certificate Authorization (CA) etc. (Globus, 2004a).

Open Geospatial Consortium (OGC), as a non-profit international membership based organization, devotes to the interoperability of various geospatial information systems that consist of many geospatial web services and process geospatial data. OGC Web Services (OWS) is one of initiatives proposed by OGC for addressing the above issue. OGC has successfully executed a series of web-based geospatial interoperability initiatives, including Web Mapping Testbed I (WMT-I), WMT II, and OGC Web Service Initiative 1.1 (OWS-1.1), and OWS 1.2 (Di, 2002). Those initiatives have produced a set of web-based data interoperability specifications, such as the OGC Web Mapping Service (WMS) specification which allows interactively assembling maps from multiple servers, and the OGC Web Coverage Service (WCS) specification which provides an interoperable way of accessing geospatial data from multiple coverage servers, especially those data from remote sensing, and OGC Catalogue Service - Web (CSW) specification which is based on ebRIM and aims to provide an object-oriented registry system for registering, managing and retrieval of geospatial resources, e.g. service, data and any other objects. CSW is the ganglia of the whole geospatial resources service center.

LAITS at GMU, as a member of OGC and a participant of those OGC interoperability initiatives, has implemented several OGC-specification compliant Web Services, such as Web Map Server (WMS), Web Coverage Server (WCS), Catalogue Service - Web (CSW) and so on.

To Combine the OGC OWS technologies and Grid technologies providing the registry, management and retrieval of volumes Grid-based geographically distributed resources can greatly make better and more efficient access to distributed computing and data resources, allow many data-intensive geospatial applications to improve significantly geospatial data access, management and analysis. So the Grid based Geospatial Catalogue Web Service is proposed here.

In the rest of this paper, we firstly give a summary of the information model of OGC Catalogue Service for Web application profile (CSW) and ISO standards for geospatial resources e.g. data, services and so on. Secondly we extend the OGC CSW information model to ingest ISO standards information model for successfully describing geospatial resources. Thirdly we integrated the extended information model into Grid environment to make CSW grid enabled. Next we discussed the security of CSW on the Grid. Then an integrated running environment and an implementation based on it are proposed to show how CSW function there. Finally, we conclude with a discussion of related work and future research direction.

### INFORMATION MODEL OF CSW FOR GEOSPATIAL RESOURCES

Information model is not only the abstract of real object and its relationships, but also the foundation of system implementation. Here we brief the ebRIM model and its contribution to CSW information model. ISO 19115/19119 (OGC, 2001), NASA Earth Observation System (EOS) Core System (ECS) (NASA, 1994) and FGDC information model (FGDC, 2002) used for describing geospatial data and services is outlined and as a extension to ebRIM-derived CSW to make it easily and normatively provide the services of publishing, managing and retrieving geospatial resources, especially for NASA HDF EOS data. The extended CSW information model also works as the core information model of Grid enabled Catalogue Web Service.

#### EbRIM-derived CSW information model

The ebRIM has been extended with a few extension elements in order to meet common requirements in the geospatial domain. So OGC CSW model extended the ebXML model and based on the extension model defining a Web-based common mechanism to classify, describe, register, retrieve and access the extended geospatial object. The CSW model specifies formally how domain objects are organized, constrained and interpreted based on...
conceptual structure. A high-level view of the ebRIM model and extension model for CSW appears in Figure 1 (OGC, 2004a; 2004b).

The RegistryObject class is an abstract base class used by most classes in the model. It provides minimal metadata for registry objects, such as name, object type, identifier and so on. The Association class is inherited from RegistryObject that is used to define many-to-many associations between objects in the information model. It uses an “associationType” attribute to identify the relationship between a source “RegistryObject” and a target “RegistryObject”. The “ClassificationScheme” class defines a tree structure made up of “ClassificationNode” to describe a structured way for classifying or categorizing “RegistryObject”. An “ExtrinsicObject” provides required metadata about the content being submitted to the registry, thus allowing any type of object to be cataloged. And the “CSWExtrinsicObject” class adds the “repositoryItem” attribute in order to refer to the content stored in remote repositories outside of the registry. A dataset service can be tightly-coupled with a dataset by specifying the value “operatesOn” to the “associationType” attribute. The “Slot” instances provide a dynamic way to add arbitrary attributes to a registry object.

The “CSWExtrinsicObject” class adds the optional repositoryItem attribute in order to specify the network location of a resource located in a repository that may not be intrinsic to the catalogue service. The getRepositoryItem() operation returns the content as the entity body within an HTTP response message. The Geometry class may be used to indicate the geometric characteristics of registry objects. It extends CSWExtrinsicObject and adds a few attributes based on the simple geometry model. We use the repositoryitem attribute of CSWExtrinsicObject to extend the geospatial metadata information model.

The OGC CSW defines several Web-based interfaces, the main interfaces are “CSW Discovery” and “CSW Transaction” to constraint on ‘find’, ‘bind’ and ‘publish’ registry objects at the geospatial conceptual level. Not only do the CSW interfaces provide the basic set of operations, such as add, delete, modify and query service and data offers and type descriptions, but also provide a number of specific capabilities, such as modify classification scheme, change registry object classification and so on. We provide a Grid enabled CSW compliant Web interfaces in this paper. The CSW adopts OGC filter syntax for expressing spatial query constraints in XML. This XML encoded filter is a system neutral representation of a query predicate that can be easily validated, parsed and then transformed into whatever target language is. For example, it could be transformed into a WHERE clause for a SQL SELECT statement to fetch data stored in a relational database, or an XPath or XPointer expression for fetching data from XML documents.

Extension of CSW for geospatial resources

Georeferenced resources usually include geospatial dataset, service and application. These resources can be described, published and retrieved through their corresponding metadata. Currently, there are three main metadata standards that we have to take into consideration for our research goal. They are NASA ECS metadata mainly for HDF EOS data, FGDC dataset metadata standards and its extension for remote sensing, and ISO 19115 dataset standards.
metadata and 19119 service metadata.

OGC has proposed "OpenGIS Catalogue Service Specifications 2.0 – ISO19115/ISO19119 Application Profile for CSW2.0" (OGC, 2004b). This CSW-ISO profile information model is based on the international standard for metadata description ISO 19115:2003. In addition, the catalogue uses a metadata description for service metadata based on the draft international ISO 19119:2003 standard to facilitate the management of service metadata. The main purpose of the information model is to provide a formal structure for the description of information resources that can be managed by a catalogue service that complies with the application profile.

For our current research project, we have to make our metadata information models compatible to NASA ECS, FGDC and ISO19115/19119 metadata standards and base on this information model to implement the OGC ebRIM-derived CSW for serving the geospatial resources by the OGC standard interfaces. So we did not use the CSW-ISO information models for describing the geospatial resources because it is only comply with ISO19115/19119 standards. Another reason is the complexity of the CSW-ISO information models because of the big amount of ISO19115/19119 metadata entries. There are more than 300 entries used to describe resources, but only a little of them work as core queryable entries. Lower complexity means higher efficiency and more convenient maintenance and usage for the catalogue service. Therefore, we simplify and synthesize the above mentioned three metadata information models on the efforts of doing our best to comply with the ISO standards at the same of satisfying our project requirements.

In order to comply with ISO19115/19119, the core metadata elements required for describing a geospatial dataset should be chosen. The core metadata elements include both mandatory and recommended optional elements. Using the recommended optional elements in addition to the mandatory elements will increase interoperability, allowing users to understand without ambiguity the geographic data and the related metadata provided by either the producer or the distributor. So we select all mandatory and part of the recommended core entries from ISO19115 and most of the entries of ISO19119. For successfully describing NASA HDF EOS dataset, mainly focusing on MODIS and ASTER data, we add some new entries derived from NASA ECS metadata. Some additional entries for describing the remote sensing data are added from FGDC standard extension for remote sensing which is becoming the ISO19115 Part 2. So this information model not only conforms to ISO19115/19119 standards, but also provides the support for publishing, managing and querying of NASA HDF EOS data. Figure 2 shows the new dataset metadata information model which is integrated into ebRIM-derived CSW models. Only main entries are showed here, others omitted. Service metadata information model is omitted here too.

![Figure 2. Dataset Metadata IM from ISO19115, NASA ECS and FGDC Extension for Remote Sensing](image)

For applying this metadata extension informational model to the above ebRIM-derived CSW model, two new special objects are proposed here. One is named as DatasetMetadata and another named as ServiceMetadata. We use...
the repositoryItem attribute of the extended Object "CSWExtrinsicObject" to point to the extended DatasetMetadata

object. And we extended the Service object with ServiceMetadata object by using the exactly same UUID on these
two objects for any new Service object registry.

A new metadata information model and its extension to ebRIM-derived CSW illustrated in Figure 3 are
proposed based on the above analysis.

GRID ENABLED CATALOGUE WEB SERVICE (GCWS)

In this section, we detail how to integrate the extended CSW model with Grid technologies to provide a Grid
enabled efficiently fundamental mechanism for publishing, managing and retrieving of geospatial resources. We
also examine the security of GCWS.

Integration of CSW with Grid technologies

One of the most popular Grid infrastructure software is Globus. Here, we integrate the CSW information model
with Globus information model of Replica Location Service (RLS) and Monitoring & Discovery Service (MDS) to
make CSW Grid enabled.

Globus Replica Location Service (RLS) maintains and provides distributed access to mapping information from
logical names for data items to target names. It consists of Local Replica Catalogs (LRC), which maintains mapping
between arbitrary Logical File Names (LFN) and the Physical File Names (PFN) associated with those LFNs on its
local storages system, and Replica Location Index (RLI), which contains a set of mappings from LFNs to RLCS
(Globus, 2004b). Globus Monitoring & Discovery Service (MDS) 2.2 is a Light Directory Access Protocol (LDAP)
based information infrastructure for computational Grid. It mainly consists of a configurable information provider
component called a Grid Resource Information Service (GRIS) and a configurable aggregate directory component
called a Grid Index Information Service (GIIS) (Globus, 2003; Czajkowski, 2001). MDS provides the ability of
discovering the properties of machines, operating system, file system, computing and network that you want to use
among VOs. Using this ability, an optimal selection of resources service can be obtained for any user/client request
which has been fully utilized on our work.

Every registered object in CSW information model has a repositoryItem attribute pointing to the extended
geospatial metadata information model. A mapping between the UUID of geospatial metadata object and the logical
file name (LFN) of the RLS entries is established. This LFN maps to a physical file name in RLC or a URL address
of RLC in RLI. If the LFN maps to the URL address of RLC, two times requests to RLS are needed for getting the
PFN which is the actual accessible address of user required data or services that is capable of providing the user
required data. We combine the PFN of the RLS entries with the Network interface information of MDS. Based on
this combination, any request with a PFN can be executed through GRIS or and GIIS services to know about what,
which and where computational resources among VOs is the most optimal one and can be most efficiently and
securely used. Figure 4 illustrates the integration of information model of extended CSW with the RLS and MDS of
Globus Toolkit.

**Legend**
- **IM** - Information Model
- **CSW** - Catalogue Service - Web
- **RLS** - Replica Location Service
- **MDS** - Monitoring & Discovery Service
- **VO** - Virtual Organization

**Figure 4. Integration of Information Model of CSW and Grid RLS/MDS**

**CSW Grid security enabled**

The essence of the Grid Security Infrastructure (GSI) is twofold: are you who you say you are (Authentication) and are you allowed to access the resources you are requesting for the tasks you want to perform (Authorization). Authorization grants access based on authenticated identity. That is what we used here. A big VO covering three Certificate Authorities (CAS) is established as our Grid security infrastructure. These three CAS are Committee on Earth Observation System (CEOS), Laboratory of Advanced Information and Technology Standard at George Mason University (LAITS/GMU) and NASA Information Power Grid (IPG). GCWS is implemented as a Grid Service among the VO and deployed on every machines of VO to make it accessible to every authorized user among the VO.

Every machine among the VO is issued a Host Certificate and a LDAP Database Certificate. Every Grid user on every machine of the VO is issued a User Certificate. The authorized user can transparently access any resources among the VO. It implies that user have to initialize the User Proxy by using User Certificate before it can access to the Grid enabled CSW – GCWS. Finally, a Grid enabled OGC Catalogue Service – Web is securely accessible.

**IMPLEMENTATION AND ITS RELATED RUNNING ENVIRONMENT**

Based on the above discussion, A Grid software infrastructure based OGC and ISO standard compatible prototype system has successfully been implemented with cooperating with NASA Ames Research Center (AMS) at LAITS/GMU. A very fruitful website [http://grid.laits.gmu.edu/](http://grid.laits.gmu.edu/) shows this prototype. We implement all OGC CSW interfaces including OGC_Service getCapabilities, CSW-Discovery and CSW-Publication and make them work as Grid Service under the Grid environment to achieve OGC CSW Grid enabled and Grid geospatial services enabled of supporting the geospatial community.

Figure 5 is the framework of Grid enabled OGC Web Services where the architecture of GCWS is embedded and its related services and running environments is showed. GCWS plays a very important role on this framework for the publishing, managing and retrieving of geospatial resources. The numbers show a simple scenario of a Grid client request a geospatial data by providing data requirements. The scenario is detailed as following:

1. Grid client establishes the Grid security authentication and requests geospatial data towards local Grid Web Coverage Service (GWCS) or OGC client requests geospatial data towards pure Web Coverage Service (WCS). GWCS or WCS will judge whether or not the requested data is local, if yes, GWCS or WCS process data locally and return data or the URL of data.

ASPRS 2005 Annual Conference
Baltimore, Maryland • March 7-11, 2005
2. If data is remote, GWCS transfers the request to intelligent Grid Service Mediator (iGSM). These information exchanging is happened between two Grid Services among the same VO.

3. iGSM queries GCWS and return a Logical File Name (LFN) securely.

4. Using the LFN, iGSM securely queries RLS to get back many physical addresses (PFNs) of the user required data or corresponding services which can produce user required data.

5. Using PFNs, iGSM securely retrieve MDS to get the current computational resources corresponding to every PFN among the VO. Then submit the job to the most optimal computational resources.

6. iGSM securely monitors the status of data requesting at remote GWCS and get back the error information or the URL of result data.

7. iGSM returns the URL to local GWCS.

8. Local GWCS returns the URL to Grid client. By this URL, user can get the required geospatial data.

---

**Figure 5.** A framework of Grid enabled OGC Web Services embedded GCWS

A demonstration of the prototype system is omitted here. It can be detailed and lively found at the website http://grid.laits.gmu.edu.

---

**RELATED WORK**

We have referred to a range of work in this paper. Here we give additional comments concerning the use of Grid technologies for OGC Web Services and their integration and Catalogue Service related work.

The ebXML information model contributed so much to the OGC Catalogue Service (CSW). Currently, prominent models used within web services realm include the ebXML and the UDDI model (OASIS, 2003). The development API associated with both models provides multiple query patterns: browse and drill-down, or filtered queries against specified registry objects (OGC, 2002). The UDDI model more focuses on business entities and associated service descriptions. An extend UDDI registry, which allows to record user-defined attributes about service, is described in (Shaikhali, 2003). The ebRIM, which draws on the ISO 11179 set of standards to provide comprehensive facilities for managing metadata, is more general and extensible. The OGC CSW extends the capabilities of ebXML model to address the catalogue service among geospatial community (OGC, 2004a; 2004b).

Some researches related to the integration of Grid and OGC technologies for geospatial discipline are as following: (Zhao, 2004) discusses the Grid enabled OGC WRS which is an implementation of the former version of OGC Catalogue Registry Service based on Grid technologies (Di, 2002). describes how to integrate Grid technology
with OGC Web Services for NASA HDF-EOS data. It combines the advantages of Grid technologies with OGC CSW to build Grid enabled Geospatial Catalogue Service to work as a Grid Service under the Grid secure environment for publishing, managing and discovering any geospatial resources, such as data, service and any related object through standard both Grid interface and Web interface. Now, here we proposed and implemented a new, improved, more efficient OGC Catalogue Service which works as both Grid Service and Web Service providing Grid interface for access from Grid users and Web Service for OGC user.

In the (Di, 2004), the concept of Geospatial Grid is proposed which includes geospatial Data Grid and Computational Grid and the combination of both. It discusses the characteristics of geospatial Grid and presents an approach to geospatial modeling with geospatial Grid. European Space Agency (ESA) uses the Web Service technologies to implement the Earth Observation (EO) Grid interfaces for access to Grid computing resources and large amount of satellite EO data (ESA, 2004; Fusco, 2004). The value of using OGC and Grid technologies in the deployment of geo-information services as part of the European Commission (EC) and European Space Agency (ESA) Global Monitoring for Environment and Security (GMES) Initiative is assessed by the SciSys on behalf of the British National Space Center (BNSC) and gotten a very positive result for GMES initiative (Fowell, 2004). Dutch Space and ESA’s European Space Research Institute (ESRIN) propose a Grid-based workflow management system – Grid Assist which hides the Grid technologies but provides a user friendly environment for executing distributed EO instrument simulations using Computational Grid (Dutch Space, 2004; Grim, 2004). Europe Union (EU) CrossGrid is a large Earth Science System project which is developing new Grid services and tools for interactive compute- and data-intensive applications like the flood crisis team decision support system and air pollution combined with weather forecasting (Bubak, 2004). The European DataGrid (EDG) develops and deploys a large scale Grid Testbed to examine the correspondence between EO application needs and actual and potential functionalities offered by Grids today (Petitdidier, 2004).

CONCLUSION AND FUTURE WORK

With the successfully implementation of Grid enabled Geospatial Catalogue Web Service and its related components and the running environments, We not only extend the application of Grid technologies to the Earth Science community, also extend the OGC CSW to ingest the NASA ECS, ISO and FGDC metadata standards to facilities the publishing, managing and querying of the NASA HDF EOS data via the Grid environment and Web environment. The OGC CSW is a vital service for sharing and interoprating volumes of distributed heterogeneous geospatial resources over the web. We propose a new information model for Geospatial Catalogue Web Service, named as GCWS by combining OGC CSW information model with the information model of the geospatial data metadata standards from ISO 19115, FGDC and NASA ECS and service metadata standards from ISO 19119. For making GCWS Grid enabled, this information model integrates the information model of the Grid Replica Location Service (RLS)/Monitoring & Discovery Service (MDS). A Virtual Organization (VO), which consists of three Certificate Authorities (CEOS, LAITS/GMU and IPG), is provided as the Grid based advanced Earth Science related computational environment among which any authenticated geospatial user can securely publish and manage geospatial resources, especially query on-demand data in the diverse community and get back it through the data-related services. This work greatly benefits the geospatial resources sharing and interoperating under the Grid environment, and implements geospatial resources Grid enabled and Grid technologies geospatial enabled.

The next goal of our research work is the access to virtual geospatial products based on the geospatial ontologies and workflow technologies. Geospatial Catalogue Web Service also is a key component for workflow-based virtual geospatial data producing. But currently, it only supports the access to the real geospatial data. So we need to enrich the information model of the GCWS and other information and directory services to make it provide enough information to construct the logical workflow and concrete workflow and related parameter models needed when a workflow is executed. Also the specific geospatial models and its relationships to the virtual geospatial products have to be investigated to support the geospatial workflow under the Grid environments.

ACKNOWLEDGEMENT

This project was supported 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 Geospatial Consortium (OGC) for the development of Web Coverage Server, Web Map Server and Catalogue Service - Web as a part of OGC WMT II and OWS-II.

ASPRS 2005 Annual Conference
Baltimore, Maryland o March 7-11, 2005
REFERENCES


