Computers & Geosciences 36 (2010) 411–421

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Grid computing enhances standards-compatible geospatial catalogue service

Aijun Chen [a,b,*], Liping Di [a,1], Yuqi Bai [a,2], Yaxing Wei [a,3,4], Yang Liu [a,5,6]

[a] George Mason University, 6301 Ivy Lane, Ste. 620, Greenbelt, MD 20770, USA
[b] NASA Goddard Space Flight Center, Code 610.2, Greenbelt, MD 20771, USA

ARTICLE INFO

Article history:
Received 19 June 2008
Received in revised form
11 September 2009
Accepted 21 September 2009

Keywords:
Catalogue service
Geospatial metadata standards
Grid computing
OGC Web services

ABSTRACT

A catalogue service facilitates sharing, discovery, retrieval, management of, and access to large volumes of distributed geospatial resources, for example data, services, applications, and their replicas on the Internet. Grid computing provides an infrastructure for effective use of computing, storage, and other resources available online. The Open Geospatial Consortium has proposed a catalogue service specification and a series of profiles for promoting the interoperability of geospatial resources. By referring to the profile of the catalogue service for Web, an innovative information model of a catalogue service is proposed to offer Grid-enabled registry, management, retrieval of and access to geospatial resources and their replicas. This information model extends the e-business registry information model by adopting several geospatial data and service metadata standards—the International Organization for Standardization (ISO)'s 19115/19119 standards and the US Federal Geographic Data Committee (FGDC) and US National Aeronautics and Space Administration (NASA) metadata standards for describing and indexing geospatial resources. In order to select the optimal geospatial resources and their replicas managed by the Grid, the Grid data management service and information service from the Globus Toolkits are closely integrated with the extended catalogue information model. Based on this new model, a catalogue service is implemented first as a Web service. Then, the catalogue service is further developed as a Grid service conforming to Grid service specifications. The catalogue service can be deployed in both the Web and Grid environments and accessed by standard Web services or authorized Grid services, respectively. The catalogue service has been implemented at the George Mason University/Center for Spatial Information Science and Systems (GMU/CSISS), managing more than 17 TB of geospatial data and geospatial Grid services. This service makes it easy to share and interoperate geospatial resources by using Grid technology and extends Grid technology into the geoscience communities.

Published by Elsevier Ltd.

1. Introduction

Cataloguing has been of increasing importance in sharing available information for almost all fields, from the traditional library to the current large amounts of digital data/information, especially satellite images. Catalogues are more critical than ever to collecting, archiving, managing, sharing, and serving digital data/information. This paper focuses on geospatial data, which is more than 80% of all data in the world (Bossler et al., 2002).

The Open Geospatial Consortium (OGC) has been advancing the sharing and interoperability of geospatial resources via developing standard geospatial specifications. OGC Web Services (OWS) is one of initiatives for addressing this issue. OWS has produced the following Web-based data, services, and systems interoperability specifications:

- Web Map Service (WMS) (de La Beaujardiere, 2001),
- Web Coverage Service (WCS) (Evans, 2003),
- Web Feature Service (WFS) (Vretanos, 2002),
- a catalogue service that is based on the e-business Registry Information Model (ebRIM) and aims to provide an object-oriented registry system for registering, managing and retrieving geospatial resources. The Catalogue Service—Web, Profile (CSW) is one of the profiles of the catalogue services implementation specifications,
- Web Processing Service (WPS) (Schut and Whiteside, 2005),

* Corresponding author at: George Mason University, 6301 Ivy Lane, Ste. 620, Greenbelt, MD 20770, USA. Tel.: +1 301 614 5152; fax: +1 301 614 5268.
E-mail addresses: aijunchen@gmail.com (A. Chen), ldi@gmu.com (L. Di), ybai1@gmu.com (Y. Bai), weiy@ornl.gov (Y. Wei), Yangliu5@gmail.com (Y. Liu).
[1] Tel.: +1 301 982 0795; fax: +1 301 345 5492.
[2] Tel.: +1 301 345 3271; fax: +1 301 345 5492.
[3] Tel.: +1 865 241 3403; fax: +1 865 574 4665.
[4] Current affiliation: Environmental Data Science & Systems, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.
[5] Current affiliation: Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA.
[6] Tel.: +1 301 345 3271; fax: +1 301 345 5492.

0098-3004/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.cageo.2009.09.006

- Sensor Planning Service (SPS) (Simonis, 2007),
- Geo Processing Workflow (GPW), which connects geo-processes through service chaining and orchestration to fulfill the requirements of complicated geospatial applications (OGC, 2007).

WCS, WMS, WFS, and SPS specify standard interfaces for access to geospatial data. CSW defines the standard interfaces for registry and retrieval of geospatial resources and their related metadata information (Nebert et al., 2007). CSW plays the role of the ganglia of the geospatial resources center. GPW chains together individual services for more complicated and value-added applications based on information retrieved from CSW and data served by other OWS services. OGC specifications, some of which are becoming International Organization for Standardization (ISO) standards, are being widely used by the geoscience community.

Grid computing has been developed to address the formidable challenges associated with the data- and computing-intensive research and applications involving many distributed heterogeneous computing systems and huge volumes of data (Foster et al., 2001). The Globus Toolkit, which is compliant with the Open Grid Service Architecture (OGSA) and Web Service Resource Framework (WSRF) specifications, is the de facto Grid software. Globus Toolkit 4.0 provides many standards-compliant Grid services for data and execution management, information services, and security in a Virtual Organization (VO) (Globus, 2008).

The United States (US) National Aeronautics and Space Administration (NASA)'s Earth Observing System (EOS) has been collecting vast volumes of earth-related data from the Earth's surface, biosphere to atmosphere at a unprecedented speed of 3.5 TB each day beside data collection from other US government agencies, such as the National Oceanic and Atmospheric Administration (NOAA) and the US Geological Survey (USGS). A major challenge for geospatial research and applications is how to promote the use, usefulness, and usability of those data. Catalogue service is a key approach to this challenge, providing ways of registering, managing, sharing, and retrieving highly multidisciplinary, heterogeneous, and distributed geospatial data and services. Following the OGC Web services specifications and the ISO and Federal Geographic Data Committee (FGDC) metadata specifications allows the sharing and interoperation of data and services. Combining catalogue services with Grid technology leads to optimal use of online computing, storage, bandwidth, and other resources for successful data-, computing- and process-intensive geospatial applications. That is the pattern—an interoperable geospatial catalogue service that is Grid-supported and compliant with widely used metadata standards. This paper discusses the service and its implementation.

The system has two merits:

(a) It extends the CSW information model that was designed based on the e-business XML registry information model so that it accommodates ISO, FGDC and NASA metadata standards for geospatial resources.
(b) It tightly integrates the catalogue service extended with Grid data management and information service components to utilize the advantages of both extensions for the geoscience community.

The paper is organized as follows: related research is first discussed. Second, the information model (IM) of OGC CSW is described in detail. Then, for describing geospatial resources, the geospatial metadata standards information models from ISO, FGDC and the NASA EOS Core System are discussed from the point of view of integration into the CSW IM. This extended IM is used to design and implement a geospatial catalogue service (GCS) providing standard CSW interfaces. Geospatial metadata generation and collection are included automatically. Thirdly, the new IM is integrated with Grid modules to take advantage of Grid technology. Doing so allows GCS to work as Grid services to both take advantage of Grid-supported computing resources and manage vast volumes of Grid-enabled geospatial data and their replicas and services. Data management, information services, and Grid security are discussed. Furthermore, an implementation and its run environment integrated with other geospatial Grid services together illustrate how the geospatial Grid catalogue service works. The performances of the catalogue services in the Grid and Web environment are evaluated. Finally, conclusions are presented and future research directions are discussed.

2. Related research

OGC has done lots of work to propose and define standards to promote the implementation and interoperability of geospatial catalogue services (Martell, 2005; Neal et al., 2006; Nebert and Whiteside, 2007; Voges and Senkler, 2007). Two leading general models usually used for cataloguing digital resources are the e-business XML Registration Information Model (ebRIM) and the Universal Description, Discovery, and Integration (UDDI) models. The ebRIM model adds extensibility and flexibility to the OGC Catalogue Service—Web Profile (CSW). Its capabilities are extended for the OGC CSW to provide a standards-compliant, object-oriented catalogue service (Lesage, 2007). Neal et al. (2006) proposed "OpenGIS® Catalogue Service Implementation Specification 2.0.1—FGDC Content Standard for Digital Geospatial Metadata (CSDGM) Application Profile for CSW 2.0". This specification defines how to use the FGDC CSDGM to implement CSW for discovering and managing geospatial metadata and their related geospatial data. OGC (Voges and Senkler, 2007) has also proposed "OGC Catalogue Services Specification 2.0.1 (with Corrigendum)—ISO 19115/ISO 19119 Application Profile for CSW 2.0". This specification explains how a catalogue service based on ISO 19115/19119 is to be organized and implemented for registering, querying, and managing metadata for geospatial data. These specifications contain some elements in common with the catalogue for NASA EOS data and data collections, but they are not the most significant contributors to it. The NASA EOS Core System (ECS) metadata (NASA, 1994) are integrated into the catalogue information model to effectively access and manage NASA Earth science data.

The overall goal of the UDDI initiative is to provide a methodology for enterprises to register OGC web services in a public registry for public access. The Web Service Description Language (WSDL) data model entities can be mapped onto UDDI data model entities. However, the OGC catalogue service focuses on registering both data and services (Sonnet, 2005). Large volumes of geospatial data must be processed. Also, the OGC Catalogue Service can use different meta-information schemas through its application profiles, while UDDI defines its own fixed structure (Hilbring and Usländer, 2006). For these reasons, the OGC catalogue service was selected as the base to extend for geospatial data and services. OGC did some experiments to discover OGC services through UDDI interfaces, as well as means of mapping between metadata models used by UDDI and OGC services. The experimental result is not satisfied by professionals.

Nogueras-Iso et al. (2005) implemented an OGC catalogue service in Java based on the metadata selected from FGDC CSDGM and ISO 19115. ShaikhAli et al. (2003) extended the UDDI information model to support discovery and retrieval of geospatial services by supporting more extensive metadata, including quasi-static, stateless metadata. This extended model is used for the Solid Earth Research Virtual Observatory (SERVO Grid) (Aktas et al., 2004), but its use in the Grid environment is not discussed. Wei et al. (2007) mapped the NASA ECS metadata items to ISO 19115 items and expressed the mapped items in an XML document. The system described in this paper upgrades that work.

Grid and OGC technologies have been integrated for use in geoscience disciplines. Zhao et al. (2004) integrate OGC Web Registry Service with Grid technology. This Web registry service is an earlier version of the OGC catalogue service. Di et al. (2003) discuss the integration of OGC Web Services with Grid technology. The catalogue system discussed in the present paper upgrades both works. This paper proposes a new, improved, more efficient Grid computing-supported geospatial catalogue service and discusses the implementation. Deelman et al. (2004) have proposed a data model for capturing the complexity of the data publication and discovery process. A set of interfaces and operations based on this model has been provided to support metadata management and a Grid metadata service was implemented. The proposed metadata management service is general although the paper mentions the Dublin core metadata and FGDC metadata standards. There is no discussion on how to use the Grid metadata service with geospatial metadata standards to manage large volumes of geospatial data. The Joint Information Systems Committee (JISC), an organization of United Kingdom higher education funding bodies, has envisioned the potential advantages of Grid and OGC collaboration. JISC (2006) states, "there are significant areas where the Grid community is more advanced, and using Grid concepts could significantly improve the capability of the geospatial community to exploit existing datasets". O'Neill et al. (2004) proposed a metadata and data model, DataGrid, for the Natural Environment Research Council (NERC) for the discovery and use of data held in such British data centers as the British Atmospheric Data Centre and the British Oceanographic Data Centre. Five distinct metadata objects are defined in the metadata model to cover discovery and the high-level properties of data. The models support the NASA Global Change Master Directory (GCMD) Directory Interchange Format (DIF), the Dublin Core, the GEO profile of Z39.50 and the Catalogue Interoperability Protocols (CIP). Although their implementation is compliant with the ISO and OGC standards, they have not followed the ISO 191xx series of standards and the OGC Web services specifications. The SEE/SAW GEO projects (See/Saw Geo, 2008) supported by the JISC focus on using Grid and OGC Web service technologies to make geospatial data securely available, semantically orchestrating geospatial Web services. Bai et al. (2007) proposed a service federation of three catalogues of geospatial data to provide a uniform access interface.

In the work described here, the information model of the ebRIM is extended to accommodate ISO, FGDC and NASA metadata standards for registering and retrieving geospatial resources, especially for the large volumes of NASA Hierarchical Data Format Earth Observation System (HDF-EOS) data. Then, the Grid software, the Globus Toolkit, is investigated and a method of integrating the extended catalogue service with the Globus Toolkit at the information model level is proposed. The information models for Grid data management and the information service components are internally coupled with the information model of the catalogue service. This structure makes Grid technology transparently available to the catalogue service. When accepting a request from a user, the catalogue service can find that service that will serve the requested data at the fastest speed. For example, when several copies of a dataset are available on different machines in a Grid environment, Grid software can help find which machine has the most idle resources (e.g. CPU, memory, and storage) to process the data and return the result to the user. This system provides the geoscience community the advantages of both a catalogue service and Grid technology.

3. Geospatial catalogue service (GCS)

As the geospatial catalogue service discussed here is an instance of a Web OGC catalogue service, it must adopt service-oriented architecture (SOA) for some fundamental interactions (Nebert and Whiteside, 2004; Nebert et al., 2007):

- publishing resource descriptions so they are accessible to prospective users (publish);
- discovering resources of interest according to some set of search criteria (discover);
- interacting with the resource provider to access the desired resources (bind).

The catalogue service plays the key role of matchmaker in such architecture. By providing publication and discovery functions, it enables a requester to dynamically discover and communicate with a suitable resource provider without requiring the requester to have advance knowledge about the provider (Martell, 2005). Therefore, the essential purpose of the catalogue service is to enable a user to locate, access, and utilize geospatial resources in the Grid environment by providing facilities for retrieving, storing, and managing many kinds of geospatial and computing resources. The geospatial metadata repository managed by the catalogue can store most of those geospatial resources whose metadata are standards-compliant. Furthermore, arbitrary relationships between catalogued items can be expressed by creating associations between any two resource descriptions; for example, a service instance may be associated with descriptions of the datasets that can be acquired using the service.

The information model of the catalogue service abstractly describes geospatial resources, metadata, and their relationships. The e-business registry information model (ebRIM) (OASIS, 2005) contributes to the catalogue service's information model. The ebRIM information model is extended to construct a geospatial metadata repository that registers, shares, manages, and provides access to geospatial data services, especially NASA HDF-EOS data. The information model of the catalogue service we proposed here accommodates ISO 19115, parts I and II and 19119 (ISO, 2003, 2002, 2005), NASA ECS (NASA, 1994) and the FGDC metadata information model (FGDC, 2002). This extended information model is used to implement the catalogue service. The OWS-compliant interfaces retrieve geospatial resources and generate and absorb their metadata in an XML document automatically. The extended CSW IM is also the core of the Grid-compliant geospatial catalogue service. Its implementation is wrapped to support Grid computing.

3.1. *ebRIM-based OGC catalogue service and its information model*

The ebRIM specifies how the catalogue content is to be structured and interrelated. It constitutes a public schema for discovery and publication purposes. It provides extensible objects for meeting some specific needs of some science communities (OASIS, 2005). The ebRIM profile of the OGC Catalogue Service for the Web was adopted to directly facilitate Web-based registry and retrieval of and access to geospatial resources. Fig. 1 illustrates the extended ebRIM information model for CSW. The optional repositoryItem attribute and getRepositoryItem() operation of the "CSWExtrinsicObject" class specifies the possible location of a resource and the way to retrieve its content. Details have been described by OGC (Martell, 2005; Lesage, 2007).

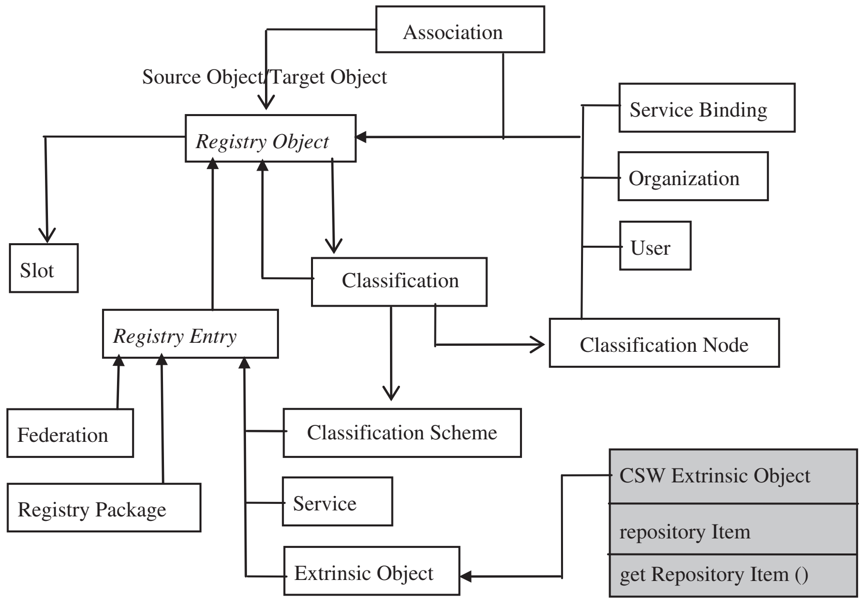

Fig. 1. High Level View of ebRIM model (OASIS, 2005) and its CSW extension elements.

3.2. Geospatial metadata and its extension to CSW

Geospatial resources can be geospatial data, services, and applications. The geospatial dataset metadata and service metadata describe those geospatial resources. In order for the catalogue service to both fully support the HDF-EOS data and follow international standards, to increase its interoperability with other geospatial catalogue systems, the CSW information model is extended to be compatible with three kinds of dataset metadata specifications and one geospatial service metadata specification.

First, the ISO dataset and service metadata specifications are ingested and seamlessly integrated with the CSW IM for describing geospatial data and service resources. All core elements are included. Then, the FGDC metadata Extensions for Remote Sensing specification is considered and some metadata elements are selected to improve the capability of CSW to describe satellite data. Finally, the heading of the HDF-EOS data files and the NASA ECS metadata specifications introduce some new elements to CSW IM, such as InstrumentName, PlatformName, ProcessingLevelID, ECSTopicKeyword, ECSDiscipline-Keyword, and LocalGranuleID. Such extension guarantees that the catalogue system described here both facilitates the publishing of and access to large volumes of HDF-EOS NASA data, and is compatible with international and US national standards. The extended IM is used as a basis for implementing the OGC ebRIM-based CSW to utilize the OGC-standard interfaces for serving geospatial resources. We did not use the information model of OGC CSW for ISO profile because of its exclusivity except for complying with ISO specifications. Therefore, the above-mentioned three metadata standards are simplified and synthesized with the goal of simultaneously being best compatible with ISO specifications and fulfilling practical demands of the system. Fig. 2 displays the new information model for geospatial dataset metadata. Only the main elements are shown here; others have been omitted. The information model for service metadata is shown in Fig. 3.

New classes are derived from the existing ebRIM-based classes, e.g. the MD_DatasetMetadata class is inherited from the CSWExtrinsicObject class that is used to describe dataset metadata. All metadata elements from the extended IM dataset metadata are added to the MD_DatasetMetadata class as attributes. The derived repositoryItem attribute can identify the URI (unified resource identifier) of geospatial data. The SV_OperationMetadata and SV_Parameter classes both come from the ISO 19119 service standard and describe geospatial service interfaces in detail. Also, the Slots class can be used to add new attributes to existing classes. The Service class from the ebRIM model can represent service metadata; however, it does not have sufficient attributes to describe Web services. The Slots class is used to add ISO 19119 elements to the Service class as attributes (Wei et al., 2005). An ebRIM-based, extended, and metadata standard-compliant information model of catalogue service is proposed based on the above analysis. It is illustrated in Fig. 4.

3.3. Geospatial catalogue service interfaces

OGC catalogue service specification 2.0.0 (Nebert and Whiteside, 2004) defines public interfaces, describing how the catalogue service provides unified operations that allow data providers, data managers, and data consumers to use the catalogue to register, manage, and discover data, services, and other resources. The design and implementation of the catalogue service is based on the Catalogue Service—Web Profile, version 2.0.0. Table 1 lists the three public interfaces. Each interface defines different numbers of operations, which can have request, response, and exception messages. The OGC_Service interface ‘getCapabilities’ queries catalogue service metadata, including ServiceIdentification, ServiceProvider, OperationsMetadata, and Contents. These metadata provide the capabilities and related information of the catalogue service. The Discovery interface provides four operations that allow clients to retrieve metadata information and geospatial resources. The Manager interface enables compatible clients to use the transaction or havestRecords operation to register metadata content (Nebert and Whiteside, 2004).

An operation request in the HTTP protocol binding usually has one of two kinds of encodings. The first is the Keyword-Value Pair (KVP) encoding. It is suitable for HTTP GET binding. The second is XML encoding, suitable for HTTP POST binding. Table 2 gives examples of both kinds of encodings of the ‘getCapabilities’ operation.

The ‘getRecords’ operation is particularly important. The primary means of geospatial resource discovery in General Catalogue Service

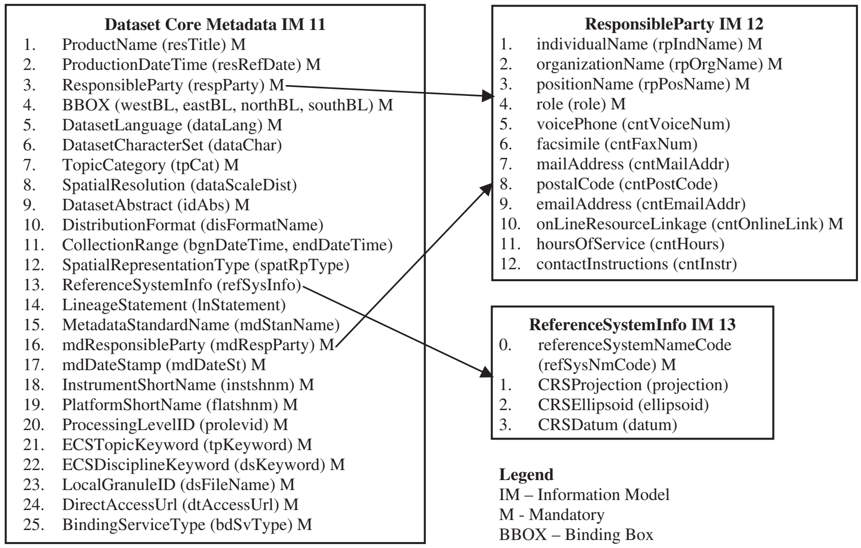

Fig. 2. Dataset metadata IM from ISO19115, NASA ECS and FGDC for remote sensing.

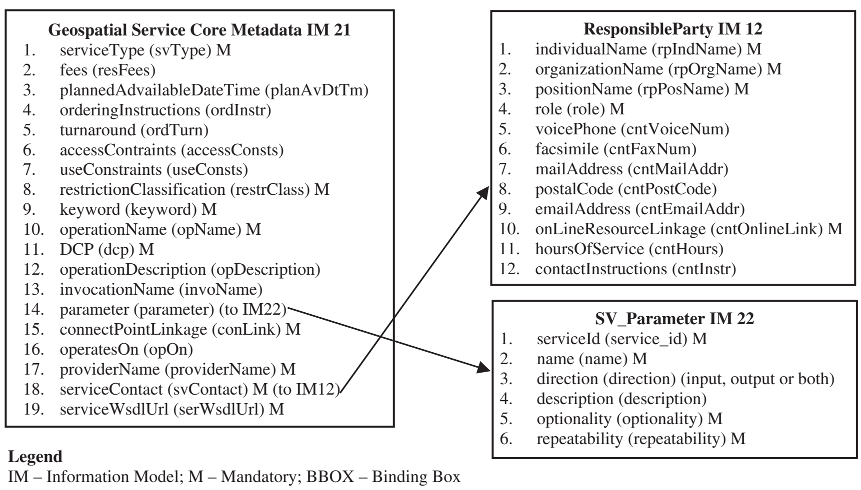

Fig. 3. Service metadata IM from ISO 19119 for geospatial services.

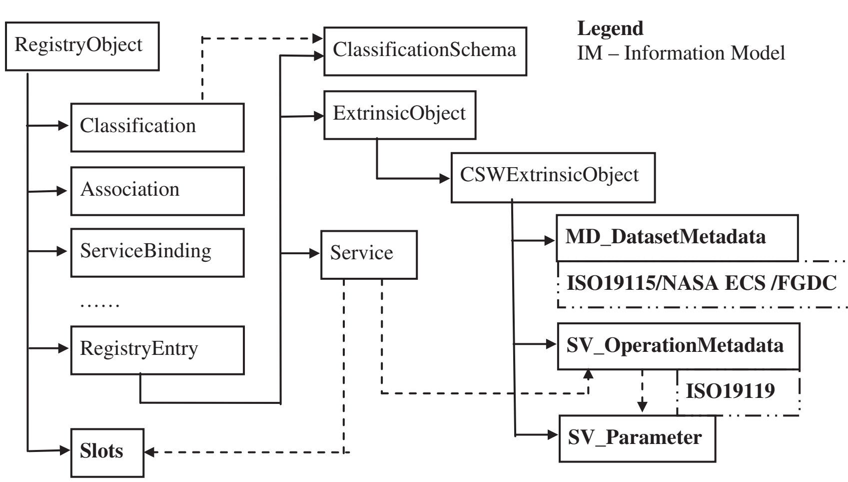

Fig. 4. Extension of ebRIM-derived CSW IM for serving geospatial resources.

2.0 are the 'search' and 'present' operations. These two operations are combined into the 'getRecords' operation, which searches the catalogue and presents the search results (Nebert and Whiteside, 2004).

4. Grid-enabled geospatial catalogue service (GeGCS)

Grid technology is applied to two levels of the geospatial catalogue service. At the information model level, the new

Table 1
Geospatial catalogue service and its interfaces, operations, and messages.

Service name	Interfaces	Operations	Messages
Geospatial Catalogue Service for Web (HTTP binding: GET/POST)	OGC_Service Discovery	getCapabilities() getRecords() describeRecord() getRegistryObjectById() getRespositoryItem()	Request, Response, Exception
	Manager	havestRecords() transaction()	

Table 2
Examples of two kinds of encoding of operation request.

Operation request	KVP encoding	XML encoding
getCapabilities	http://laits.gmu.edu/csw?request=GetCapabilities&service=CSW&version=2.0.0&outputFormat=text/xml	<GetCapabilities service=“CSW”> <version>2.0.0</version> <outputFormat>text/xml</outputFormat> </GetCapabilities>

geospatial catalogue service IM is combined with the IM of the Data Management services of Grid software. At the service level, the catalogue service is designed and developed by following Grid specifications in the Grid environment. Grid Information Services and Security components are used to optimize selection of and secure access to geospatial resources available in the Grid environment

4.1. Integration of IM of GCS with Grid data management modules

The Globus Toolkit is the de facto Grid software infrastructure. It has been used in many Grid research and applications. It consists mainly of four components: Security, Data Management, Information Services, and Execution Management. As GCS focuses on data and service management, it is integrated with the data management and information services components.

GCS and the Grid data management component are integrated at the information model level. One data management service is the replica location service (RLS), which is used to resolve the distributed multiple data replicas in the GCS. Each data replica has one and only one physical location, which is usually an URL, but all replicas appear at different sites. A replica location service consists of a replica location index (RLI) and a local replica catalog (LRC) that forms a pyramidal distributed management structure for managing and retrieving data replicas distributed at different sites in the Grid environment. The metadata for all replicas are exactly the same. So, one logical name represents all replicas of a dataset and its metadata in the GCS. The same logical name is used in the replica location service, where the logical name for all replicas of the dataset is mapped to multiple physical locations. That mapping information, which is at the bottom of the pyramid, is saved at LRC. The top of the pyramid is that root of the RLI that contains mapping information between the logical name and the entries of LRC. The monitoring and discovery information service (MDS4) includes an index service and a trigger service. The index service gathers the properties of Grid resources and offers an access interface to those data. Examples of properties are host name, memory size, operating system name and version, file system data, processor, and load data. Those properties are used to select the optimal data and services when multiple physical replicas of a dataset or multiple service instances are found in the GCS (Chen et al., 2005, 2006). Reliable file transfer and data replication services from the data management component are invoked when data and available services are not at the same nodes so that data transfer is necessary. The entire procedure for querying GCS and optimizing query results is automated. It fully integrates the geospatial catalogue service with Grid technology. Fig. 5 shows the integration.

4.2. Grid-enabling GCS

To make GCS workable in the Grid environment, GCS must be integrated with Grid at the service level, implementing a WSRF-based GCS without changing the access interfaces defined by the OGC specifications. Because the Grid service is a special Web service and the GCS is a Web service, implementing Grid-enabled GCS is specializing GCS to make it compatible with the Grid specifications. The procedure also involves migrating GCS from the Web environment to the Grid environment. The Grid service WSDL has some extensions to the WSDL standard for a Web service. We have designed and implemented a generalized command line tool that accepts Grid service WSDL documents and produces the necessary server-side Java class framework and stubs. The tool runs in Linux and Sun Solaris environments. The stubs program client codes to communicate with server-side Grid services. The server-side Java classes add necessary processing functions. The implemented Grid services are compatible with Grid service standards and can be deployed into the Globus Container and run in the Grid environment. All metadata records related to Web services in the catalogue are updated with Grid service related metadata. Table 3 illustrates part of the WSDL file of a Grid service.

The WSDL document describes the GeGCS interfaces and messages. The definitions of types, omitted here, and messages are same as for OGC CSW. However, the portTypes implemented in the CSW server are extended from wsrf:GridService portType. The binding address points to the access URL of the GeGCS.

4.3. GCS Grid security

The Grid Security Infrastructure (GSI) is one of the most important components in Grid software. GSI allows users to access those resources for which are authorized by authenticating them. GSI secures the access to GeGCS. Grid service can be deployed into the Grid service container in secure and non-secure mode. New parameters are introduced in this Grid service deployment descriptor file (∗.wsdd) to enable GeGCS

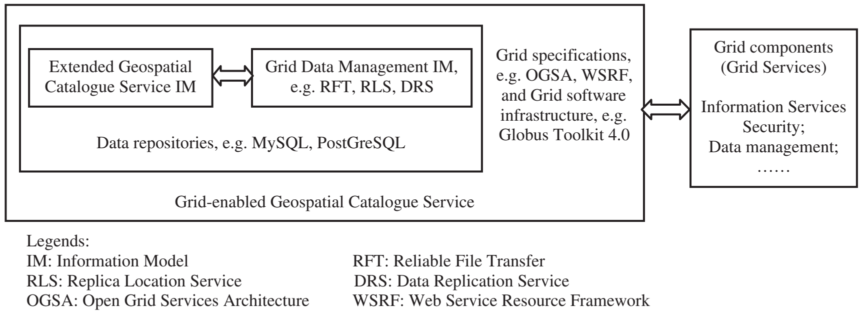

Fig. 5. Integration of Grid-enabled Geospatial Catalogue Service (GCS).

Table 3
Part of the WSDL file of a Grid service.

```
<wsdl:definitions name="GridGCS" targetNamespace="http://laits.gmu.edu/GridGCS/service" >
  <wsdl:types> … … </wsdl:types>
  <wsdl:message name="getRecordsOutputMessage">
    <wsdl:part name="parameters" element="tns:getRecordsResponse"/>
  </wsdl:message>
  <wsdl:message name="getRecordsInputMessage">
    <wsdl:part name="parameters" element="tns:getRecords"/>
  </wsdl:message>
  <wsdl:portType name="GridGCSPortType">
    <wsdl:operation name="getRecords">
      <wsdl:input message="tns:getRecordsInputMessage"/>
      <wsdl:output message="tns:getRecordsOutputMessage"/>
    </wsdl:operation>
  </wsdl:portType>
  <wsdl:binding name="GridGCSPortTypeSOAPBinding"
  type="tns:GridGCSPortType">
    <soap:binding style="document" transport="http://
  schemas.xmlsoap.org/soap/http"/>
    <wsdl:operation name="getRecords">
      <soap:operation soapAction="http://laits.gmu.edu/GridGCS/
  GridGCSPortType/getRecordsRequest"/>
      <wsdl:input>
        <soap:body use="literal"/>
      </wsdl:input>
      <wsdl:output>
        <soap:body use="literal"/>
      </wsdl:output>
    </wsdl:operation>
  </wsdl:binding>
</wsdl:definitions>
  <wsdl:service name="GridGCSService">
    <wsdl:port binding="tns:GridGCSPortTypeSOAPBinding"
  name="GridGCSPortTypePort">
      <soap:address location="https://65.123.203.147:8443/wsrf/services/
  GridGCS"/>
    </wsdl:port>
  </wsdl:service>
</wsdl:definitions>
```

to support GSI (Di et al., 2008). The added parameters appear as follows:

```
<parameter name="securityConfig"
value="org/globus/ogsa/impl/security/descriptor/gsi-security-config.xml"/>
<parameter name="authorization" value="Gridmap"/>
```

The Gridmap file saves user authorization information as follows:

"/O=Grid/OU=GMU/OU=LAITS/OU=gmu.edu/CN=Aijun Chen" achen

Any one node in the Grid environment should have at least one host certificate, one database certificate and several user certificates. All certificates should be issued by a Certificate Authority (CA). The user with a certificate can access any Grid resource at those nodes whose host and services certificates come from the same CA that issues the user certificate. Also, any user can be issued a certificate to access authorized resources in a Virtual Organization (VO). Our VO uses three CAs: from George Mason University (GMU), from the National Aeronautics and Space Administration (NASA), and the Department of Energy (DOE).

5. Implementation, running environment, performance assessment and prototype system

A prototype GeGCS has been implemented. This service, together with other Grid-enabled OGC Web services, shares and provides access to geospatial resources for the geoscience community. Performance comparison of GeGCS with GCS reveals that both kinds of services perform almost equally well.

5.1. Implementation of GeGCS and its running environment

First, the OGC CSW is implemented, then it is Grid enabled by creating a GeGCS. The interfaces listed in Table 1 are implemented using the following components of the OGC specifications for CSW. The function getCapabilities provides detailed information about the GeGCS. The harvestRecords() and transaction() interfaces are used to absorb massive amounts of metadata information into the GeGCS. The getRecords() interface allows the user to query the GeGCS metadata repository. Fig. 6 illustrates how metadata from HDF-EOS data are harvested into and are retrieved from the GeGCS by users.

A template XML file conforming to ISO 19139 is provided first as a sample XML file in which the values of all attributes and elements are empty. The template is modified to be compliant with the FGDC and NASA ECS metadata information models. ISO 19139 is a Geographic Information – Metadata – XML Schema Specification (ISO, 2007) that provides an XML schema describing how ISO 19115 metadata can be stored in XML format. A transformation program converts the metadata harvested from HDF-EOS data files to GeGCS form. The program first automatically reads ECS metadata from HDF-EOS files, then it maps ECS metadata elements into XML elements, and finally, it modifies the XML template file. The generated XML file that represents an HDF-EOS data file can be registered in two ways. The first is directly to register the XML file into the catalogue as a Dataset object through the manager interface. The second is to parse the XML file to retrieve the values of every attribute and element and then register the values into the MD_DatasetMetadata table of the catalogue. The table MD_DatasetMetadata represents a Dataset object. A client can search through Dataset objects by providing CSW-compatible request messages through the discovery interfaces. The request messages provide the user's requirements, for example, spatial bounding box, temporal

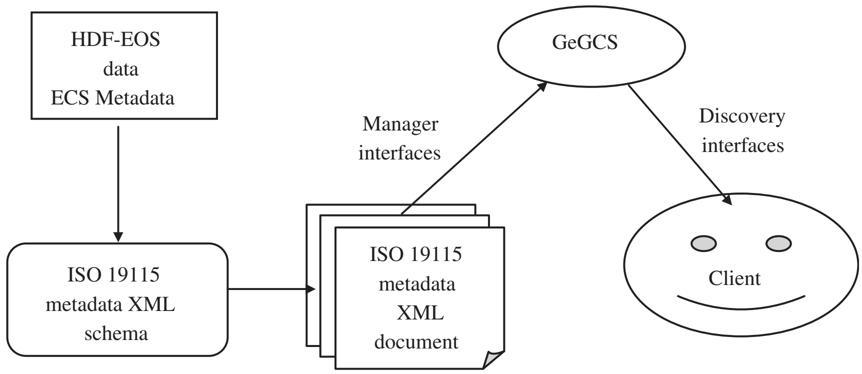

Fig. 6. Metadata harvest from HDF-EOS files to GeGCS.

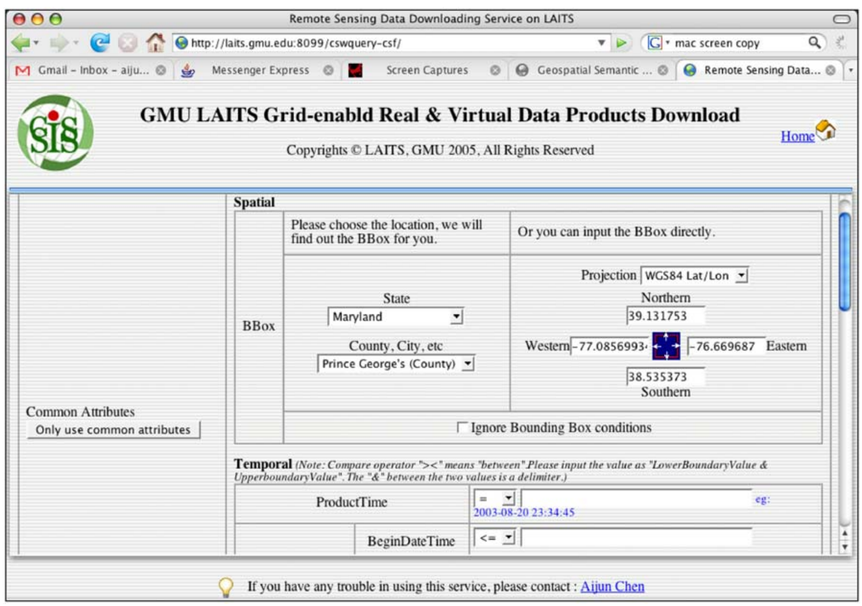

Fig. 7. Discovery Interface of Grid-enabled Geospatial Catalogue Service.

range, instrument name, and platform name. The response messages contain a list of qualifying Dataset objects in XML format (Wei et al., 2007). More than 17TB of NASA HDF-EOS data have been absorbed into the catalogue for user query and free download. Fig. 7 shows the discovery Web user interface of the GeGCS.

A CSW portal provides an OGC-standard interface for OGC-compatible client access to GeGCS. The client submits a standard request in XML format and gets back a standard response in XML format. There are also portals for other OGC Web services, for example WCS and WMS. Therefore, three interfaces are provided for every OGC Web service:

1. OGC standard interfaces without Grid support;
2. Grid WSRF-compatible interfaces in the Grid environment;
3. OGC standard interfaces with Grid support (the portals).

Fig. 8 shows the portals and interfaces in a prototype running environment. The environment consists of three layers. The outer layer is the CSW portal layer. The portals provide a single point for Web entrance to all data, services, and computing capabilities managed by the Grid. At the front end, the portals provide OGC Web services standard interfaces for any OGC clients accessing the geospatial resources managed by a geospatial Grid without knowing of the existence of the Grid. At the back end, the portals sign in as an authorized Grid user to use geospatial Grid resources to fulfill user requests. The middle layer is the GeGCS, which provides access interfaces for Grid users. The inner layer is the GCS, which provides OGC standard interfaces without the support of Grid technology.

5.2. *Performance assessment and comparison between Web Service and Grid service*

The performances of GeGCS (a Grid service) and GCS (a Web Service) were evaluated and compared. The request method, size, and the response payload were considered in assessing the overall time required for successfully completing a user request. Three types of call were tested; secure GeGCS, non-secure GeGCS, and GCS. GeGCS runs under a Globus container in the Grid environment and GCS runs under an Apache Axis container deployed in Apache Tomcat. Both GeGCS and GCS were running on the same computer, using a Linux system with the same Java packages, same database (MySQL), and the same catalogue size querying the

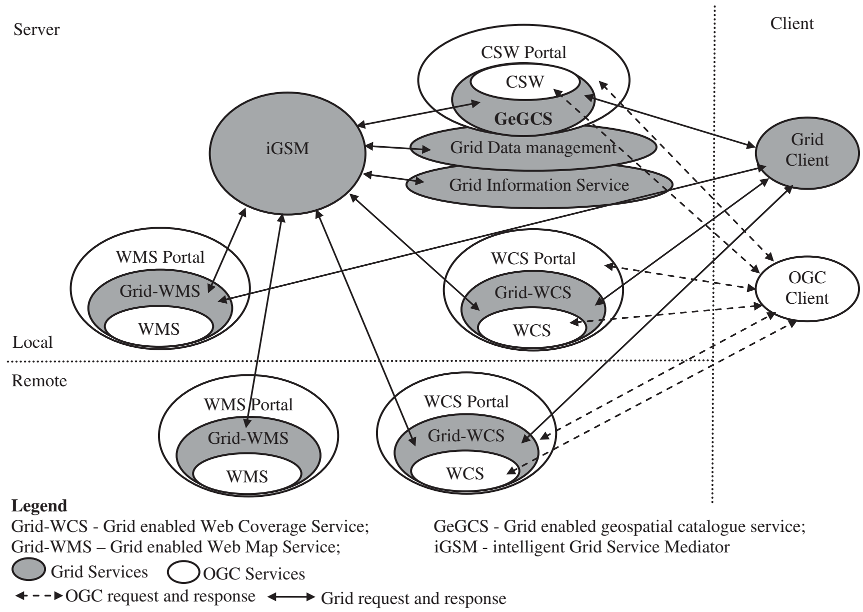

Fig. 8. Prototype environment for Grid-enabled geospatial catalogue service.

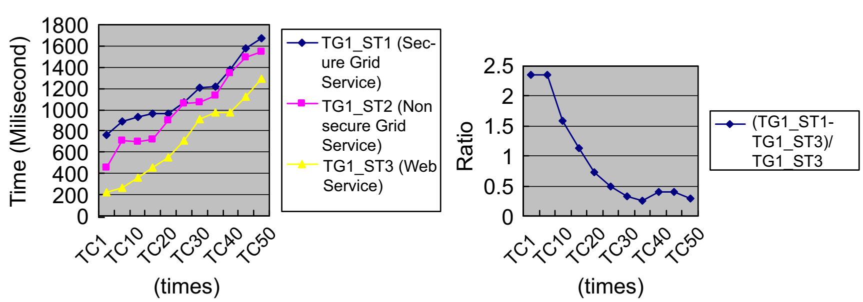

Fig. 9. Performance evaluation of secure and non-secure GeGCS (Grid Service) and GCS (Web Service).

same datasets. The evaluations were completely separate. All testing was Web-based. Two different Web-based interfaces were provided. The testing client (running Windows) and testing server (running Linux) were in the same local network with 100 M network bandwidth. Because the focus is on the performance difference between geospatial catalogue Web services and catalogue Grid service, the network latencies for both services are the same and will not affect the difference of performance. So, no matter from where both services are tested the difference of the performance should be same.

The following three types of services are assessed:

Service Type 1 (ST1: GeGCS): Grid service in secure mode;
Service Type 2 (ST2: GeGCS): Grid service in non-secure mode;
Service Type 3 (ST3: GCS): Web service without security.

Each test case (TC) deals with a specific request size and response payload, represented as TC10, TC20 etc. The request payload is the same although the numbers of data resources of interest and response sizes are different. The left panel of Fig. 9 illustrates overhead times for different kinds of services. The right panel shows the comparison of secure Grid service with Web service.

Grid security does affect overall performance. Secure Grid service requires 300 ms more response time than non-secure Grid service and 600 ms more response time than Web service. The difference in response time between secure Grid service and the other two kinds of services decreases with increasing request payload. With the increase of the request payload, the ratio of the overhead to overall response time becomes so small that the difference can be ignored. These assessment results demonstrate the advantages of the Grid services.

5.3. Prototype system

Both geospatial services and geospatial data were registered using the extended catalogue service information model. The OGC OWS web services, like WMS, WCS, WFS, and CSW, were registered in the catalogue and categorized into different catalogue taxonomies such as OGC service and ISO 19119 service taxonomies. More than 17 terabytes of LandSat satellite data from NASA have been downloaded in GeoTIFF format, stored and registered in the catalogue. The geospatial catalogue Web service is now in operation, serving 17 terabytes of data for public use. The Grid-enabled geospatial catalogue Grid service is a prototype system. Fig. 7 shows the Web user query interface of the geospatial catalogue Grid services.

6. Conclusions

We successfully take advantage of Grid technology to support and enhance extended geospatial catalogue service (GeGCS). The extended geospatial catalogue service can accommodate geospatial metadata from NASA ECS, ISO, and FGDC for facilitating the registry, management, publication, and retrieval of HDF-EOS data in Web and Grid environments. The catalogue service is integrated with Grid technology in two levels—information model level and service level. The Globus data management component, information services components and secure component involved the integration. A virtual organization, consisting of six nodes with CAs from GMU, NASA, and DOE, is established to form a Grid computational environment to verify the GeGCS. In the environment, authorized users securely access to Grid-managed geospatial resources. In particular, they can securely access to data and services from other communities or securely share their data and services with others. The research promotes the sharing and interoperation of geospatial data, services, and applications using Grid technology.

We are using GeGCS as a system catalogue to provide data, information, and services that are needed for automatically building up virtual geospatial products based on geospatial ontologies and Grid workflow technologies. GeGCS is playing a very important role in implementing and producing workflow-based virtual geospatial data based on Grid technology. The information model of GeGCS has been extended to store abstract DataTypes and ServiceTypes. The enriched GeGCS is capable of storing adequate information for constructing a logical geospatial processing workflow and instantiate it to a concrete workflow in the Grid environment. GeGCS also provides related parameters to support workflow execution using the rich geospatial resources provided by Grid.

Acknowledgements

This project was supported by grants from the NASA Earth Science Data and Information System Project (NCC5-645, PI: Dr. Liping Di), NASA Advanced Information System Technology program (NAG-13409, PI: Dr. Liping Di), NASA REASoN program (NNG04GE61A, PI: Dr. Liping Di), and US National Geospatial-Intelligence Agency NURI program (HM1582-04-1-2021, PI: Dr. Liping Di). Additional funding was received from the OGC for developing the OGC Web Services. Many thanks are given to our colleague, Dr. Barry Schlesinger, for his proofreading of the manuscript.

References

Aktas, M., Aydin, G., Donnellan, A., Fox, G., Granat, R., Grant, L., Lyzenga, G., Mcleod, D., Pallickara, S., Parker, J., Pierce, M., Rundle, J., Sayar, A., Tullis, T., 2004. iSERVO: implementing the international solid earth research virtual observatory by integrating computational grid and geographical information web services. Pure and Applied Geophysics 163 (11–12), 2281–2296, doi:10.1007/s00024-006-0137-8.

Bai, Y., Di, L., Chen, A., Liu, Y., Wei, Y., 2007. Towards a geospatial catalogue federation service. Photogrammetric Engineering & Remote Sensing 73 (6), 699–708.

Bossler, J.D., Jensen, J.R., McMaster, R.B., Rizos, C., 2002. Manual of Geospatial Science and Technology. Taylor & Francis, London, 668 pp.

Chen, A., Di, L., Wei, Y., Bai, Y., 2006. An optimized Grid-based, OGC standards-compliant collaborative software system for serving NASA geospatial data. In: Proceedings Second IEEE Systems and Software Week (SASW 2006), Columbia, MD, pp. 159–166 10.1109/SEW.2006.11.

Chen, A., Di, L., Wei, Y., Bai, Y., Liu, Y., Hu, C., Mehrotra, P., 2005. Grid computing enabled geospatial catalogue web service. In: Proceedings American Society for Photogrammetry and Remote Sensing (ASPRS), Baltimore, MD, USA, CD-ROM, 1-57083-076-2.

Deelman, E., Singh, G., Atkinson, M.P., Chervenak, A., Hong, N.P.C., Kesselman, C., Patil, S., Pearlman, L., Su, M., 2004. Grid-Based metadata services. In: Proceedings 16th International Conference on Scientific and Statistical Database Management (SSDBM 04), Santorini Island, Greece, 393 pp 〈http://doi.ieeecomputersociety.org/〉 10.1109/SSDM.20041311235.

de La Beaujardiere, J., 2001. Web map service implementation specification (OGC 01-047r2), version 1.1.0. Open Geospatial Consortium Inc., 73 pp 〈http://www.opengeospatial.org/standards/wms〉.

Di, L., Chen, A., Yang, W., Liu, Y., Wei, Y., Mehrotra, P., Hu, C., Williams, D., 2008. The development of a geospatial data Grid by integrating OGC Web services with Globus-based Grid technology. Concurrency Computation: Practice and Experience 20 (14), 1617–1635, doi:10.1002/cpe.1292.

Di, L., Chen, A., Yang, W., Zhao, P., 2003. The integration of Grid technology with OGC Web Services (OWS) in NWGISS for NASA EOS Data. In: Proceedings Workshop on Grid Applications and Programming Tools held in conjunction with Eighth Global Grid Forum, Seattle, USA, pp. 63–72.

Evans, J. (Ed.), 2003. Web coverage service (WCS), (OGC 03-065r6), version 1.0.0. Open Geospatial Consortium Inc., 67 pp 〈http://www.opengeospatial.org/standards/wcs〉.

FGDC, 2002. Content standard for digital geospatial metadata: extensions for remote sensing metadata. Federal Geographic Data Committee, Standard Working Group. URL: 〈http://www.fgdc.gov/standards/status/csdgm_rs_ex.html〉 (Accessed 18 March, 2008).

Foster, I., Kesselman, C., Tuecke, S., 2001. The anatomy of the Grid-enabling scalable virtual organizations. International Journal of High Performance Computing Applications 15 (3), 200–222.

Globus, 2008. The Globus toolkit 4.0 release manuals. URL: 〈http://www.globus.org/toolkit/docs/4.0/〉 (Accessed 2 April, 2008).

Hilbring, D., Usländer T., 2006. Catalogue services enabling syntactical and semantic interoperability in environmental risk management architectures. Managing environmental knowledge, In: Proceedings 20th International Conference on Informatics for Environmental Protection, Aachen: Shaker Verlag, Graz, Austria 〈http://www.eu-orchestra.org/docs/20060906-Orchestra Paper-EnviroInfo2006-CatalogueServiceApproach-updated.pdf〉.

ISO/TC211, 2007. Geographic information – metadata – XML schema implementation, 19139:2007. ISO Technical Committee 211. URL: 〈http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=54904〉 (Accessed 11 April, 2008).

ISO/TC211, 2005. Geographic information—services, 19119:2005. ISO Technical Committee 211. URL: 〈http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_ tc_browse.htm?commid=54904〉 (Accessed 20 March, 2008).

ISO/TC211, 2003. Geographic information—metadata, 19115:2003. ISO Technical Committee 211. URL: 〈http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_tc_browse.htm?commid=54904〉 (Accessed 13 March, 2008).

ISO/TC211, 2002. ISO draft: Geographic information – metadata – Part 2: Extensions for imagery and gridded data, ISO/WD 19115-2.2. ISO Technical Committee 211. URL: 〈http://www.isotc211.org/protdoc/211n1337/211n1337.pdf〉 (Accessed 14 March, 2008).

JISC, 2006. Grid and Open Geospatial Consortium (OGC) collision. Joint Information Systems Committee. URL: 〈http://www.jisc.ac.uk/media/documents/funding/2006/04/grid_and_ogc_collision.pdf〉 (Accessed 22 April, 2008).

Lesage, N. (Ed.), 2007. OGC cataloguing of ISO metadata (CIM) using the ebRIM profile of CS-W (OGC 07-038) version 0.1.7. Open Geospatial Consortium, Inc., USA, 95 pp 〈http://www.opengeospatial.org/standards/cat〉.

Martell, R. (Ed.), 2005. OpenGIS® catalogue services—ebRIM (ISO/TS 15000-3) profile of CSW (OGC 05-025r3), version 1.0.0. Open Geospatial Consortium, Inc., USA, 62 pp. 〈http://www.opengeospatial.org/standards/requests/29〉.

NASA, 1994. Proposed ECS (EOS Core System) core metadata standard release 2.0. December 1994. URL: 〈http://edhs1.gsfc.nasa.gov/waisdata/docsw/pdf/tp4200105.pdf〉 (Accessed 12 April, 2008).

Neal, P., Davidson, J., Westcott, B. (Ed.), 2006. OpenGIS® catalogue service implementation specification 2.0.1—FGDC CSDGM application profile for CSW 2.0. (OGC® 06-129r1), version 0.0.12 Open Geospatial Consortium, Inc., USA, 70 pp 〈http://www.opengeospatial.org/standards/cat〉.

Nebert, D., Whiteside, A., Vretanos, P. (Eds.), 2007. OpenGIS® catalogue services specification (OGC® 07-006r1), version 2.0.2 Open Geospatial Consortium, Inc., USA, 218 pp 〈http://www.opengeospatial.org/standards/cat〉.

Nebert, D., Whiteside, A. (Eds.), 2004. OGC® catalogue services specification, (OGC® 04-021r3), version 2.0.0 with corrigendum Open Geospatial Consortium, Inc., USA, 187 pp 〈http://www.opengeospatial.org/standards/cat〉.

Nogueras-Iso, J., Zarazaga-Soria, F.J., Bejar, R., Alvarez, P.J., Muro-Medrano, P.R., 2005. OGC Catalog Services: a key element for the development of spatial data infrastructures. Computers & Geosciences 31, 199–209.

OASIS, 2005. EbXML registry information model version 3.0, Organization for the Advanced Structured Information Standards (OASIS), 78 pp. URL: 〈http://www.oasis-open.org/committees/download.php/13591/docs.oasis-open.orgregrepv3. 0specsregrep-rim-3.0-os.pdf〉 (Accessed 20 March 2008).

OGC, 2007. OGC Web services, Phase 5. Open Geospatial Consortium, Inc. URL: 〈http://www.opengeospatial.org/projects/initiatives/ows-5/#gpw〉 (Accessed 9 April 2008).

O'Neill, K., Cramer, R., Gutierrez, M., Kleese van Dam, K., Kondapalli, S., Latham, S., Lawrence, B., Lowry, R., Woolf, A., 2004. A specialized metadata approach to discovery and use of data in the NERC DataGrid. In: Proceedings the UK E-Science All Hands Meeting, Nottingham, UK, 4 pp. 〈http://epubs.cclrc.ac.uk/bitstream/495/172.pdf〉 (Accessed 10 April 2008).

Schut, P., Whiteside, A., 2005. OpenGIS Web processing service (OGC® 05-007r4), version 0.4.0. 〈http://portal.opengeospatial.org/files/?artifact_id=13149&version=1&format= pdf〉 (Accessed 16 March 2008).

See/Saw Geo, 2008. The SEE GEO and SAW GEO projects. URL: 〈http://edina.ac.uk/projects/seesaw/seegeo/〉 (Accessed 10 March 2008).

ShaikhAli, A., Rana, O.F., Al-Ali, R., Walker, D.W., 2003. UDDIe: An extended registry for Web services. In: Proceedings of the Symposium on Applications and the Internet Workshops, 27–31 January 2003–Orlando, FL, USA, pp. 85–89.

Simonis, I., (Ed.), 2007. OpenGIS® sensor planning service implementation specification 1.0. (OGC® 07-014r3), version 1.0 Open Geospatial Consortium, Inc., USA, 186 pp 〈http://www.opengeospatial.org/standards/sps〉.

Sonnet, J. (Ed.), 2005. OWS2 common architecture: WSDL SOAP UDDI, (OGC 04-060r1), version 1.0.0. Open Geospatial Consortium, Inc., USA, 76 pp 〈http://www.opengeospatial.org/standards/dpd〉.

Voges, U., Senkler, K. (Eds.), 2007. OGC catalogue services specification 2.0.2—ISO metadata application profile, (OGC® 07-045), version 1.0 Open Geospatial Consortium, Inc., USA, 127 pp 〈http://www.opengeospatial.org/standards/cat〉.

Vretanos, P., 2002, Web feature service implementation specification (OGC® 02-058), version 1.0.0 Open Geospatial Consortium, Inc., 105 pp 〈http://www.opengeospatial.org/standards/wfs〉.

Wei, Y., Di, L., Zhao, B., Liao, G., Chen, A., 2007. Transformation of HDF-EOS metadata from ECS model to ISO 19115-based XML. Computers & Geosciences 33 (2), 238–247.

Wei, Y., Di, L., Zhao, B., Liao G., Chen, A., Bai, Y., Liu, Y., 2005. The design and implementation of a Grid-enabled catalogue service. In: Proceedings 2005 IEEE International Geoscience & Remote Sensing Symposium, vol. 6, Seoul, South Korea, pp. 4224–4227.

Zhao, P., Chen, A., Liu, Y., Di, L., Yang W., Li, P., 2004. Grid metadata catalog service-based OGC Web registry service. In: Pfoser, D., Cruz, I.F., Ronthaler, M. (Eds.), Proceedings 12th ACM International Workshop on Geographic Information Systems. Washington DC, USA. URL: 〈http://portal.acm.org/citation.cfm?id=1032227〉 (Accessed 12 March 2008).