Towards a Geospatial Catalogue Federation Service

Yuqi Bai, Liping Di, Aijun Chen, Yang Liu, and Yaxing Wei

Abstract

As geospatial catalogues are becoming accessible online through public query interfaces, a federation to fulfill distributed and integrated metadata discovery needs to be built. This study investigates the feasibility of federating three distinct geospatial catalogue services: the NASA Earth Observing System (EOS) ClearingHOuse (ECHO), the George Mason University (GMU) OpenGIS Catalogue Service for Web (CSW), and the U.S. Department of Energy (DOE) Earth System Grid (ESG) Simulation Data Catalogue. Challenges and problems in dealing with the metadata conceptual models, query languages, and communication protocols are analyzed. Proposed federation strategies and the operational federation system are introduced. Our results show that protocol adaptation, query dispatching, query criteria translation, and query results integration are the four main challenges in building a catalogue federation. A mediator-wrapper based approach can be adopted to build a federation service. The OpenGIS Catalogue Service specification can be used to define the internal communication protocols between the federation service and the affiliated catalogue services, and between the federation service and its clients.

Introduction

In recent decades, different agencies have developed their own geospatial catalogues to facilitate discovery, access, and sharing of large volumes of geospatial data, either observed satellite images or simulation data (Crompvoets *et al.*, 2004). These geospatial catalogues are becoming accessible online through their query interfaces (Nebert, 2000; Unidata, 2004; Tait, 2005). For scientists who conduct multi-disciplinary research, they may need to search multiple catalogues in order to find the data they need. Such work is very time-consuming and tedious, especially when the catalogues may use different metadata models and catalog interface protocols. It is very desirable if those catalogues can be integrated into a catalogue federation, which will present a well-known metadata model and interface protocol to users and hide the complexity and diversity of the affiliated catalogues behind the interface. With the federation, users only need to work with the federated catalogue to find the data they need instead of working with individual catalogues individually. This article addresses a mechanism to build such a catalogue federation service, which will integrate multiple legacy catalogues to facilitate distributed and integrated data discovery.

The mediator-wrapper (Wiederhold, 1992) architecture has been widely used for integrated and united access to multiple, autonomous information sources. As depicted in Figure 1, data sources store their data in different ways, for instance, in a database or as flat files. The data is exported over the Internet through a query interface. Each data source is wrapped by one or more source-specific wrapper components, which offer a query interface hiding the particular data model, access path, and interface technology of the source. Wrappers are used by a mediator, which offers users an integrated access through its global schema. The user poses queries against the global schema of the mediator; the mediator then distributes the query to appropriate wrappers. The wrappers transform the queries so they are understandable and executable by the data source they wrap, collect the results, and return them to the mediator. Finally, the mediator integrates the results as a user response (Naumann, 2002).

This mediator-wrapper architecture has been successfully applied in many research areas, such as database systems (Sheth and Larson, 1990; Dwyer and Larson, 1987; Litwn, 1985), taxonomy-based information sources (Tzitzikas, 2005), digital libraries (Melnik, 2000), Neuroimaging (Barillot *et al.*, 2006), XML data (Baru, 1999; Lin *et al.*, 2000), Internet search engines (Katchaounoy *et al.*, 2002), and Grid services (Alpdemir *et al.*, 2003).

Ramroop and Pascoe (2001) presented Selector Broker, a Common Request Broker Architecture (CORBA) component, focusing on geospatial information resources (datasets and/or services). It employed direct translation of metadata semantics and syntax, without going through an intermediate system, to provide distributed access to geographical data sets using the Lightweight Directory Access Protocol (LDAP). Gupta *et al.* (2002) introduced a specification for describing the spatial capabilities of geospatial sources and services in a wrapper-mediator system and described multi-step query planning based on the capability descriptions. Seamless navigation based on a mediator-wrapper was initialized for automatic and dynamic selection of a location-based sensor and a map suited to a user's context (Hosokawa and Taga, 2004). An infrastructure model, which is based on a spatial mediator that takes metadata on the information needs of the user, data sources, and tools available, as well as device characteristics (in field settings) into consideration when processing the user's request, has been presented for integrating spatial data (Miller and Nusser, 2003).

Most of the previous geospatial federation efforts have focused on geospatial data retrieval. Geospatial metadata are usually textual descriptive information about data. Most of their entries have nothing to do with spatial relationships.

Center for Spatial Information Science and Systems, George Mason University, 6301 Ivy Lane, Suite 620, Greenbelt, MD 20770 (ldi@gmu.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 73, No. 6, June 2007, pp. 699–708.

0099-1112/07/7306–0699/$3.00/0

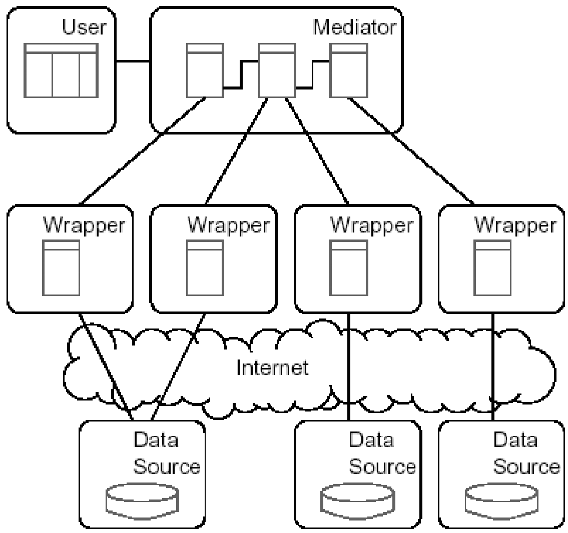

Figure 1. The mediator-wrapper architecture.

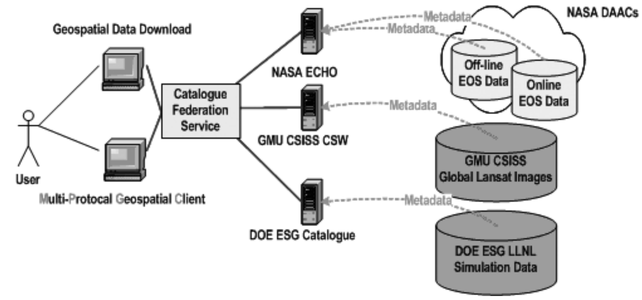

Figure 2. Federation System Context Diagram. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

Previous proposed federation mechanisms, especially query plan evaluation and query results assembling, are not very suitable for data discovery with geospatial metadata models. Little information is available about federation of diverse catalogue services in the literature. The purpose of this paper is to present through a case study how a federation of multiple, diverse catalogue services for geospatial metadata could be carried out through public discovery interfaces.

The rest of this paper is organized as follows. The case study scenario is introduced followed by a summary of the challenges in building a federation to provide distributed and integrated metadata discovery. The next section focuses on the design strategy followed by the federation implementation in detail. The advantages and limitations of the proposed solution are discussed leading to the conclusions.

Scenario

In the NASA Earth Observing System (EOS) Higher-Education Alliance (NEHEA) GeoBrain project, the principal problem in facilitating discovery of spatial data and information for teaching and research at institutions of higher education was that of how to provide a single access point for three legacy geospatial catalogue services: the NASA EOS ClearingHOuse (ECHO) (Pfister *et al.*, 2001; ECHO, 2005), the Center for Spatial Information Science and Systems (CSISS) at George Mason University (GMU) OpenGIS Catalogue Service for Web (CSW) (Bai and Di, 2005), and the U.S. Department of Energy (DOE) Earth System Grid (ESG) Simulation Data Catalogue (David *et al.*, 2005).

As depicted in Figure 2, these three catalogue services help users to discover and access, respectively, NASA EOS data, GMU CSISS global Landsat images, and DOE ESG simulation data. A Catalogue Federation Service (CFS) is needed to provide a single access point for these three underlying catalogue services and to support end users in discovery of distributed metadata through either a Web application (e.g., the Geospatial Data Download portal in GeoBrain), or stand-alone client program.

NASA ECHO

NASA has externalized the Earth Observing System Data and Information System (EOSDIS) metadata into a clearinghouse named ECHO and provided a foundation for the interoperability of externally developed user interfaces and data services. ECHO acts as middleware between data partners and client partners. Data partners, such as the NASA Distributed Active Archive Centers (DAACs), provide metadata information about their data holdings, and client partners develop software to access this information. End users use one of the ECHO clients to access this metadata information. The current operational version is ECHO-8.

Metadata Conceptual Model

ECHO's metadata conceptual model is named the ECHO Earth Science Metadata Conceptual Model (EESMCM). It is developed from the EOSDIS Core System (ECS) Science Data Model (MacHarrie and McBride, 2002). This EESMCM model deals with geospatial data at two levels: Granule and Collection. In the ECS, a Granule is the smallest unit of data that is independently described and inventoried, while a Collection represents a logical grouping of granules. Each Collection is identified with an Earth Science Data Type. Granule-level metadata is usually different for each granule, while collection-level metadata applies to all granules in a collection. Both Granule and Collection metadata are distinctly identified and described in ECHO.

Query Language

ECHO defines a query language, named IIMSAQL, for searches on metadata of collections and granules in the ECHO system. This is an XML-based language with a detailed DTD definition (GST, 2006).

As shown in Figures 3 and 4, this XML language takes "*query*" as a root element composed of:

- One "*for*" element to identify the target geospatial object to query against, either collection or granule;
- One "*dataCenterId*" element to describe the DAAC from which to search;
- One "*where*" element, which is composed of child elements that define the exact query criteria.

ECHO explicitly defines 18 query criteria for the collection-level and 25 for the granule-level. Some criterion examples are *Campaign Short Name, Processing Level, Percentage of Cloud Cover, Sensor Name, Spatial*, and *Temporal*. For each criterion, ECHO has proposed a corresponding syntax definition in IIMSAQL.

Figure 3 shows a complete ECHO query example for a collection-level search. In the "dataCenterId" part, it defines LPDAAC and GSFC, which means that this collection search

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE query
SYSTEM "http://api.echo.eos.nasa.gov/echo/dtd/IMSAQLQueryLanguage.dtd">
<query>
  <for value="collections"/>
  <dataCenterId>
     <list>
       <value>GSFCECS</value>
       <value>LPDAAC</value>
     </list>
  </dataCenterId>
  <where>
     <collectionCondition>
       <dataSetId>
        <list>
         <value>'ASTER DEM Product V002'</value>
        </list>
       </dataSetId>
     </collectionCondition>
    <collectionCondition negated="Y">
      <processingLevel>
        <list>
         <value>'1'</value>
         <value>'1A'</value>
         <value>'1B'</value>
         <value>'2'</value>
        </list>
       </processingLevel>
     </collectionCondition>
  </where>
</query>
```

Figure 3. ECHO Collection Query Example.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE query SYSTEM
"http://api.echo.eos.nasa.gov/echo/dtd/IMSAQLQueryLanguage.dtd">
<query>
  <for value="granules"/>
  <dataCenterId>
    <all/>
  </dataCenterId>
  <where>
    <granuleCondition><onlineOnly/></granuleCondition>
    <granuleCondition>
      <cloudCover>
        <range lower="10" upper="20"/ >
      </cloudCover>
     </granuleCondition>
     <granuleCondition>
       <spatial operator="RELATE">
         <Polygon>
           <LRing>
             <CList>-77.37, 38.75, -76.66, 38.75, -76.66, 39.03, -77.37, 39.03, -77.37, 38.75</CList>
           </LRing>
         </Polygon>
       </spatial>
     </granuleCondition>
     <granuleCondition>
       <temporal>
         <startDate><Date YYYY="2006" MM="01" DD="01"/></startDate>
         <stopDate><Date YYYY="2006" MM="12" DD="31"/></stopDate>
       </temporal>
     </granuleCondition>
  </where>
</query>
```

Figure 4. ECHO Granule Query Example.

will only work on those metadata holdings provided by the NASA LP DAAC or the NASA GSFC DAAC. In the "where" element, two criteria are defined: Processing Level and Dataset ID. Each query criterion is encoded as a "collection-Condition" element. The defined Processing Level criterion means that all collections are of interest except for the ones with processing level 1 or 2. The Dataset ID criterion prescribes that only those "ASTER DEM Product V002" data are to be searched.

The ECHO query example for a granule-level search is shown in Figure 4. Four criteria are defined, *Online Access Only, Percentage of Cloud Cover, Spatial*, and *Temporal*. The granule query criterion is encoded as a "granuleCondition" element. For this query, ECHO returns only those granules that have an online access address, cloud cover between 10 percent and 20 percent, cover the spatial region which is defined by the following geodetic system polygon (lower-left corner latitude: 38.75° longitude: −77.37°; upper-right corner latitude: 39.03°, longitude: −76.66°) and were collected in 2006.

ECHO supports text-matching operations, temporal data types and geospatial operators. However, it does not support Boolean queries and thus cannot make nested Boolean queries. So, complex logical queries are not supported in ECHO. The only possible logical relationship among query criteria is the logical **AND** operation.

Communication Protocol

ECHO uses the Simple Object Access Protocol (SOAP) (Mitra, 2003) as the transport protocol. ECHO-7 exposes the Session Manager and a limited set of the ECHO services as Web Services, defined using the Web Services Description Language (WSDL) (Christensen, 2001). In addition, ECHO provides several software libraries, such as "Façade" and "ECHOTalk," for client providers to interact with ECHO. These packages expose the ECHO services as a set of auto generated Java classes and functions. They free developers from having to create the ECHO messages and handling the details of the underlying SOAP communication protocol. ECHO-8 provides a new alternative API interface that is based on an industry standard Web Services interface and is compliant with the WS-I Basic Profile 1.0 (Ballinger *et al.*, 2004) for client communication.

GMU CSISS CSW System

The Center for Spatial Information Science and Systems (CSISS) at GMU has launched a pilot project to build a proof-of-concept OpenGIS standards-compliant catalogue service to promote the online sharing of its 8.5 TB global Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper (ETM+) images.

OpenGIS has published one abstract specification (Kottman, 1999), one implementation specification (Nebert and Whiteside, 2004), and two public recommendation papers (Martell, 2005; Voges and Senkler, 2005) to identify how the Geospatial Catalogue Service is envisioned and can be designed and implemented. The GMU CSW is defined and implemented by following the ebRIM application profile of the implementation specification, in particular the HTTP protocol binding specification of the (Catalogue Service for Web).

Metadata Conceptual Model

GMU defines its metadata conceptual model by combining the requirements of the ISO 19115 (ISO, 2003) and ISO 19119 (ISO, 2005) standards with the metadata implementation specifications of ISO 19139 (ISO, 2006), and the ECS metadata specification.

To facilitate on-the-fly geospatial service chaining, two types of geospatial information are maintained in the GMU CSW: geospatial data and geospatial services. Currently, only OpenGIS geospatial services, such as the Web Coverage Service (WCS) and the Web Mapping Service (WMS), can be dealt with by the GMU CSW. One characteristic of the GMU CSW is that those geospatial datasets that can be served through a particular service are always associated with that service. Given this association information, the users can customize data through standard OpenGIS service interfaces.

Besides those pre-defined ebRIM information model (OASIS, 2005) objects, the following metadata objects are defined in the GMU CSW:

1. Data Granule: Data Granule has the same meaning in the GMU CSW as in ECHO. The metadata information for each Landsat image is manipulated as a Data Granule object.
2. WCSCoverages: The coverage information for each data granule is maintained in one or more corresponding WCSCoverage elements.
3. WMS Layers: GMU CSW maintains WMS Layer information for each Data Granule to meet the user requirement that the data granule must be retrieved through the OpenGIS WMS interface.

4. Services: The metadata for each OpenGIS service can be stored in the GMU CSW. To facilitate service-related queries, the authors of this paper have proposed a detailed service classification schema, which will not be discussed in this paper, to identify the available service types.

Query Languages

The OpenGIS CSW specification proposes an OGC-Common Catalogue Query Language, which must be supported by all compliant OpenGIS Catalogue Services. This query language supports nested Boolean queries, text matching operations, temporal data types, and geospatial operators. Implementations of query languages that are transformable to this language are the OGC Filter specification (Vretanos, 2005) and the CIP (CEOS, 2005) and GEO (Nebert, 2000) profiles of Z39.50 (ISO, 1998) Type-1 queries. GMU CSW supports the OGC Filter specification. Users can define complex query criteria against any object that is defined in the aforementioned metadata conceptual model.

Communication Protocols

The OpenGIS CSW specification proposes three protocol bindings for its catalogue service: Z39.50, CORBA/IIOP, or HTTP. GMU CSW chooses to support HTTP protocol binding to support web-based applications. In this binding, the interaction between a client and a Catalogue Service Server uses a standard request-response model of the HTTP protocol. Request and response messages are encoded as keyword-value pairs (HTTP GET) within a request URI or use an XML entity-body (HTTP POST).

This protocol binding defines the semantic and syntactic implementation for the following proposed discovery related operations:

1. GetCapabilities: This mandatory operation allows CSW Clients to retrieve service metadata which describes what the service can do, what operations it supports, and how the clients can contact it, that is, its capabilities. The response is an XML document.
2. DescribeRecord: This mandatory operation enables a client to discover elements of the information model supported by the target catalogue service. The operation allows some of or the entire information model to be described.
3. GetDomain: This optional operation is used to obtain runtime information about the range of values of a metadata record element or request parameter. The runtime range of values for a property or request parameter is typically much smaller than the value space for that property or parameter based on its static type definition.
4. GetRecords: This is the principal operation used to search catalogue content. Some or all of the registry objects in the result set that satisfy the search criteria may be returned in the response message. The query and response message payload must have a content type of application/XML.
5. GetRecordById: The mandatory GetRecordById request retrieves the default representation of catalogue records using their identifiers. In addition, this operation is a subset of the GetRecords operation, and is a convenient short form for retrieving and linking to records in a catalogue.

Figure 5 shows an XML example for a request payload. In this example, the object name of interest, DataGranule, is defined in the "csw:Query" element and also in the "csw:Element" element. An "ogc: Filter" element is used to define a complex search condition composed of three predicates combined using the logical operator **AND**. The first predicate defines a targeted spatial region using the "ogc: BBOX" element. This element identifies a bounding box in EPSG:4326 Geographic Longitude-Latitude Projection with lower-left corner latitude of 38.75° and longitude of −77.37°, and upper-right corner latitude of 39.03° and longitude of −76.66°. The second predicate defines a temporal coverage composed of two predicates combined by using the logical operator **OR**. Each predicate defines a time period of interest. In particular, the "ogc: PropertyIsGreater ThenOrEqualTo" element is used to define the beginning of the time period, and the "ogc: PropertyIsLessThanOrEqual To" the end of it. The last predicate specifies that the "dataSetId" attribute should equal "MODIS/Terra Aerosol 5-Min L2 Swath 10 km V004."

```
<?xml version="1.0" encoding="UTF-8"?>
<csw:GetRecords   xmlns="http://www.opengis.net/cat/csw"
            xmlns:csw="http://www.opengis.net/cat/csw"   xmlns:ogc="http://www.opengis.net/ogc"
            xmlns:gml="http://www.opengis.net/gml"   version="2.0"
            outputFormat="text/xml; charset=UTF-8"   outputSchema="http://www.opengis.net/cat/csw"
            maxRecords="1"   startPosition="1">
 <csw:Query typeNames="DataGranule">
  <csw:ElementSetName>full</csw:ElementSetName>
  <csw:ElementName>/DataGranule/</csw:ElementName>
  <csw:Constraint version="1.0.0">
   <ogc:Filter>
    <ogc:And>
     <ogc:BBOX>
      <ogc:PropertyName>/DataGranule/BBOX</ogc:PropertyName>
      <gml:Box srsName="EPSG:4326">
       <gml:coordinates>>-77.37,38.75 -76.66,39.03</gml:coordinates>
      </gml:Box>
     </ogc:BBOX>
     <ogc:Or>
      <ogc:And>
       <ogc:PropertyIsGreaterThanOrEqualTo>
        <ogc:PropertyName>/DataGranule/beginDateTime</ogc:PropertyName>
        <ogc:Literal>2006-02-01 00:00:00</ogc:Literal>
       </ogc:PropertyIsGreaterThanOrEqualTo>
       <ogc:PropertyIsLessThanOrEqualTo>
        <ogc:PropertyName>/DataGranule/endDateTime</ogc:PropertyName>
        <ogc:Literal>2006-02-28 23:59:59</ogc:Literal>
       </ogc:PropertyIsLessThanOrEqualTo>
      </ogc:And>
      <ogc:And>
       <ogc:PropertyIsGreaterThanOrEqualTo>
        <ogc:PropertyName>/DataGranule/beginDateTime</ogc:PropertyName>
        <ogc:Literal>2006-04-01 00:00:00</ogc:Literal>
       </ogc:PropertyIsGreaterThanOrEqualTo>
       <ogc:PropertyIsLessThanOrEqualTo>
        <ogc:PropertyName>/DataGranule/endDateTime</ogc:PropertyName>
        <ogc:Literal>2006-04-30 23:59:59</ogc:Literal>
       </ogc:PropertyIsLessThanOrEqualTo>
      </ogc:And>
     </ogc:Or>
     <ogc:PropertyIsEqualTo>
      <ogc:PropertyName>/DataGranule/dataSetId</ogc:PropertyName>
      <ogc:Literal>MODIS/Terra Aerosol 5-Min L2 Swath 10km V004</ogc:Literal>
     </ ogc:PropertyIsEqualTo>
    </ogc:And>
   </ogc:Filter>
  </csw:Constraint>
 </csw:Query>
</csw:GetRecords>
```

Figure 5. GMU CSW Granule Query Example.

Figure 6 shows the GMU CSW GetRecords response payload. It always consists of one "csw: GetRecordsResponse" element. This element includes one "csw: searchStatus" element and one "csw: SearchResults" element. The metadata for the matched results is organized as subelements of "csw:SearchResults."

ESG Catalogue System

The ESG integrates supercomputers with numerous different large-scale data and analysis servers located at national laboratories and research centers. Doing so creates a powerful environment for next generation climate research. To facilitate the global sharing of large volumes of simulation datasets, Lawrence Livermore National Laboratory (LLNL) has built a proof-of-concept Catalogue System, which refers the OpenGIS CSW specification.

Metadata Conceptual Model

The ESG metadata schema is tailored specifically to climate data. It consists of an object model centered on a "Dataset," which is best described as a collection of logical records representing some computational results, such as the result of a climate model run. Datasets are characterized by a number of attributes, including geospatial and temporal bounds. Each logical record contains data for one or more climatological variables. Logical records are mapped to one or more physical files by the Globus Toolkit Replica Location Service (RLS) database.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<csw:GetRecordsResponse
        xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:csw="http://www.opengis.net/cat/csw"
        xmlns:rim="urn:oasis:names:tc:ebxml-regrep:rim:xsd:2.5" xmlns:ogc="http://www.opengis.net/ogc"
        xmlns:laitscsw="http://laits.gmu.edu/csw" version="2.0.0">
 <SearchStatus status="complete" timestamp="2007-02-19T11:34:33-05:00"/>
 <csw:SearchResults  numberOfRecordsMatched="129573" numberOfRecordsReturned="1" nextRecord="2">
  <laitscsw:DataGranule  id="urn:uuid:1ee71520-468d-1028-915e-2145c06bc3e2"
        home="http://laits.gmu.edu"  objectType="urn:uuid:5F610352-133F-4D22-A900-62D704DDA79F"
        status="Approved"  expiration="2010-01-01 00:00:00.0"  majorVersion="1"  minorVersion="0"
        stability="Dynamic"  userVersion="1.0"  mimeType="image/tiff"  isOpaque="false"  repositoryItem="" >
   <name>p014r035_7t19990923_z18_nn10.tif</name>
   <description>Landsat ETM Data, Path: 014, Row: 035, Band: 1, Date: 1999-09-23</description>
   <archiveCenter>GMU_LAITS</archiveCenter>
   <beginDateTime>1999-09-23 00:00:00.0</beginDateTime>
   <endDateTime>1999-09-23 23:59:59.0</endDateTime>
   <dataFormat>GeoTIFF</dataFormat>
   <sizeMBDataGranule>68.84</sizeMBDataGranule>
   <processingLevelID>Orthorectified</processingLevelID>
   <platformShortName>LANDSAT-7</platformShortName>
   <instrumentShortName>LANDSAT ETM</instrumentShortName>
   <sensorShortName>LANDSAT ETM</sensorShortName>
   <previewURL>http://geobrain.laits.gmu.edu/LANDSAT/Places/Maryland/p014r035_7x19990923.
            ETM-EarthSat-Orthorectified/p014r035_7t19990923.preview.jpg</previewURL>
   <browseURL>http://geobrain.laits.gmu.edu/LANDSAT/Places/Maryland/p014r035_7x19990923.
            ETM-EarthSat-Orthorectified/p014r035_7t19990923.browse.jpg</browseURL>
   <path>14</path><row>35</row>
   <OnlineAccessURLs><OnlineAccessURL>
    <URL>http://geobrain.laits.gmu.edu/LANDSAT/Places/Maryland/p014r035_7x19990923.
            ETM-EarthSat-Orthorectified/p014r035_7t19990923_z18_nn10.tif</URL>
        <URLDescription></URLDescription>
        <MimeType>image/tiff</MimeType>
   </OnlineAccessURL></OnlineAccessURLs>
   <ogc:BBOX><referenceSystemNameCode>EPSG:4326</referenceSystemNameCode>
    <westBoundingCoordinate>-77.621933</westBoundingCoordinate>
    <eastBoundingCoordinate>-74.809631</eastBoundingCoordinate>
    <northBoundingCoordinate>37.061718</northBoundingCoordinate>
    <southBoundingCoordinate>35.021454</southBoundingCoordinate>
   </ogc:BBOX>
   <ogc:BBOX><referenceSystemNameCode>EPSG:32618</referenceSystemNameCode>
    <westBoundingCoordinate>266788.500000</westBoundingCoordinate>
    <eastBoundingCoordinate>516933.000000</eastBoundingCoordinate>
    <northBoundingCoordinate>4101720.000000</northBoundingCoordinate>
    <southBoundingCoordinate>3878394.000000</southBoundingCoordinate>
   </ogc:BBOX>
  </laitscsw:DataGranule>
 </csw:SearchResults>
</csw:GetRecordsResponse>
```


Figure 6. GMU CSW Granule Query Response Example.

Query Languages
Due to the limited time available for its implementation, the ESG catalogue can support only simple query criteria that are encoded according to the OpenGIS Filter Encoding specification.

Communication Protocols
Like GMU CSW, the ESG Catalogue implements the HTTP protocol bindings as specified in the OpenGIS specification.

Challenges in Federating NASA ECHO, GMU CSW, and ESG Catalogues

The analysis above shows that there is much heterogeneity in metadata information models, query languages, and communication protocols. Building a federation service to integrate these three online Geospatial Catalogue Services faces the following challenges:

1. Protocol Adaptation: GMU CSW and the ESG catalogue support HTTP protocol (GET/POST) binding, while NASA ECHO uses SOAP to maintain the connection with the clients. The federation server should use the correct protocol when communicating with each Catalogue service. The protocol the clients may use to talk to the federation server itself is another concern. After all protocols have been defined and identified, the federation server should support protocol adaptation internally.
2. Query Dispatching: The federation server is responsible for dispatching a query to the affiliated catalogue services. A dispatching model should be defined to deal with the following issues:
   - Transparency: Whether the federation user is aware of these affiliated catalogue services and whether users can define which catalogue services are of interest in their queries.
   - Sequence: Whether the federation server dispatches the users' queries to these affiliated catalogue services in a predefined sequence, whether this sequence can be changed in runtime, and whether the federation users can define this sequence in their queries.
3. Query Translation: The translation of queries is another major issue. The federation has to deal with the following problems:
   - Metadata Query Objects: The metadata objects queried against using one set of query criteria may not have counterparts in another schema. For example, the federation service cannot fulfill queries for objects defined in GMU CSW and NASA ECHO for those simulation-specific metadata objects referenced only in the ESG catalogue schema. Another issue is that the same registry object has different names, in different schemes, e.g., *Granule* in NASA ECHO versus DataGranule in GMU CSW.
   - Query Format: Both GMU CSW and the ESG Catalogue accept queries in OGC Filter format, while ECHO only accepts IIMSAQL format. The federation server needs to transform an individual query into the different proprietary formats. The spatial query criterion and temporal query criterion are expressed differently in the NASA ECHO granule query example and the GMU CSW granule query example, as shown in Figures 4 and 5.
   - Query Language Functionality: Some complex query predicates in one query language cannot be identically expressed in another one. For example, the OGC Filter specification supports nested Boolean queries. Such queries can be supported at best with difficulty on ECHO IIMSAQL, and some cannot be supported at all.

Query Results Integration

Catalogue query results from multiple Catalogue Services may need to be integrated before being sent back to users. As these metadata results may not use the same schema, the rules the federation server uses to re-organize metadata information while keeping the original content should be well designed. Furthermore, whether the clients can define the format of the query result of interest and, if so, how, also needs to be addressed.

Proposed Federation Strategy

Protocol Adaptation

As this federation is supposed to provide a single access point to multiple, autonomous information sources, it may follow the mediator-wrapper architecture, where the federation works as a mediator, and wrappers may be deployed for communicating with specific catalogue services if protocol adaptation is needed.

From the client's point of view, the federation should operate as a Catalogue Service, except that the federation service will always delegate metadata queries to the affiliated Catalogue Services, which are stand-alone and self-maintained. The OpenGIS Catalogue Service for Web Specification (CSW)'s ebRIM profile is an OGC recommended standard for catalog interoperability and many catalogue clients have been developed with this specification. In order to promote interoperability between this federation service and other third party's OpenGIS services and clients, we use the ebRIM profile as the protocol between the federation and the catalogue clients.

For the NASA ECHO, GMU CSW, and the ESG Catalogues, as two of these three catalogue services support the OpenGIS Catalogue Service specification, the internal communication protocol between this federation service and these affiliated catalogue services may also follow this specification. In this case, a wrapper for NASA ECHO is needed to adapt the query criteria to be compliant with IIMSAQL language, and transform the query results to be compatible with the OpenGIS Catalogue Service. The advantage of this strategy is that we do not need to develop two more wrappers. Since the federation also takes user queries through this interface on the front end, this strategy will decrease the difficulty of internal protocol adaptation. And, more important, as this OpenGIS Catalogue service specification is public rather than proprietary, following this public standard will definitely ease future efforts to include more catalogue services, at least on protocol adaptation.

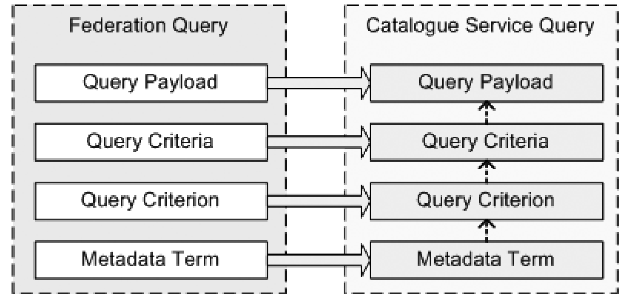

Figure 7. Federation Query Transformation. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

```
<ogc:BBOX>
  <ogc:PropertyName>/DataGranule/BBOX</ogc:PropertyName>
    <gml:Box srsName="EPSG:4326">
    <gml:coordinates>>-77.37,38.75 -76.66,39.03</gml:coordinates>
  </gml:Box>
</ogc:BBOX>
```

(a)

```
<granuleCondition>
  <spatial operator="RELATE">
    <Polygon>
      <LRing>
        <CList>-77.37, 38.75, -76.66, 38.75, -76.66, 39.03, -77.37, 39.03, -77.37, 38.75</CList>
      </LRing>
    </Polygon>
  </spatial>
</granuleCondition>
```

(b)

Figure 8. Query Criterion Transformation Example.

Query Dispatching

For dispatching queries, the federation service can choose from three patterns:

1. Opaque: In this scenario, the federation service fully controls the distributed query process, with the clients having no awareness of the affiliated Catalogue Services.
2. Translucent: The federation service may expose the affiliated Catalogue Services to the users, but the users can define neither which Catalogue Services their query can be forwarded to nor the sequence of queries.
3. Transparent: The federation service may expose the affiliated Catalogue Service to the users, and the users can define those Catalogue Services of interest and the sequence in which their queries can be dispatched.

The way the federation service exposes the affiliated Catalogue Service is another issue. Since we propose federation service work as an OpenGIS Catalogue Service, this could be achieved through the service capabilities XML file. By embedding some new elements there, federation service users may know more detailed information about those underlying Catalogue Services, such as physical address, binding information, introduction Web page URL.

Whether a particular Catalogue Service will be queried against is also dependent on whether it supports the user's query criteria. For example, since *model, experiment, run, frequency*, and *variable* are only used in the DOE ESG catalogue, if these criteria are cited in the user's query, the federation does not need to dispatch this query to NASA ECHO or GMU CSW, even if the user selects those catalogues in the transparent pattern. Generally, when a user's query includes any proprietary query condition, it will dispatch the query only to those Catalogue services capable of handling those conditions.

Query Translation

Query Translation in CFS has two aspects: semantic and syntactic. A federation usually maintains a global schema that is exposed to end-users. Metadata attribute terms in user queries always follow this global schema. Before being dispatched to an underlying affiliated catalogue service, they should be transformed appropriately. This transformation logically involves four layers: metadata term, query criterion, query criteria, and query payload, as shown in Figure 7.

Metadata term transformation aims to select an appropriate metadata object/attribute term from a local schema that is semantically equal to the current term referenced in federation query. This schema mapping information could be organized as a direct mapping (global schema – local schema) or an indirect mapping (e.g., ontology-based), and should be referenced at run time.

Query Criterion transformation deals with a single query criterion. Simple metadata term replacement may also work on this level, but there are some unresolved issues, particularly around spatial query criterion. Figure 8 depicts one example of a spatial query criterion transformation.

In the OpenGIS Filter specification, the *<ogc:BBOX>* element is defined as a convenient way of encoding the very common bounding box constraint based on the *<gml:Envelope>* geometry. It is equivalent to the spatial operation *<Not><Disjoint>* . . . *</Disjoint></Not>* meaning that the *<ogc:BBOX>* operator should identify all geometries that spatially interact with the box. In this example, *<gml:Box>* is used to identify a rectangular bounding box of interest where the lat/long coordinate of the left-bottom point is (−77.37, 38.75), and that of the upper-right point is (−76.66, 39.03). The whole *<ogc:BBOX>* defines a spatial query criterion against the *BBOX* attribute of the *DataGranule* registry object. To transforming this query criterion to be acceptable by NASA ECHO, the federation requires construction of a *granuleCondition* element that further consists of a *spatial* element. Note that, in alignment with the underlying Oracle spatial database, ECHO accepts a different type of representation for the same latitude/longitude bounding box.

At the query criteria level, a federation needs to ensure the proper relationship among criteria for a single query. Figure 9 shows a query criterion example, where Cloud Cover is examined. Figure 10 shows that this query criterion is nested within an <ogc:Not> predicate. In this case, a *negated* attribute is used to with a value of "Y" in the transformation result. Figure 11 illustrates another situation where multiple single query criteria are considered together to reach the right result.

At this level, a federation also needs to pay attention to the query language functionality of the target Catalogue Service. One example would be NASA ECHO. Since the only

```
<ogc:PropertyIsGreaterThanOrEqualTo>
  <ogc:PropertyName>/DataGranule/cloudCover</ogc:PropertyName>
  <ogc:Literal>10</ogc:Literal>
</ogc:PropertyIsGreaterThanOrEqualTo>
                    (a1)

<granuleCondition>
 <cloudCover>
  <range lower="10"/ >
 </cloudCover>
</granuleCondition>
                    (b1)
```

Figure 9. Query Criteria Transformation Example (1).

```
<ogc:Not>
<ogc:PropertyIsGreaterThanOrEqualTo>
  <ogc:PropertyName>/DataGranule/cloudCover</ogc:PropertyName>
  <ogc:Literal>10</ogc:Literal>
</ogc:PropertyIsGreaterThanOrEqualTo>
</ogc:Not>
                    (a2)

<granuleCondition negated="Y">
 <cloudCover>
  <range lower="10"/ >
 </cloudCover>
</granuleCondition>
                    (b2)
```

Figure 10. Query Criteria Transformation Example (2).

```
<ogc:And>
 <ogc:PropertyIsGreaterThanOrEqualTo>
  <ogc:PropertyName>/DataGranule/cloudCover</ogc:PropertyName>
  <ogc:Literal>10</ogc:Literal>
 </ogc:PropertyIsGreaterThanOrEqualTo>
 <ogc: PropertyIsLessThanOrEqualTo >
  <ogc:PropertyName>/DataGranule/cloudCover</ogc:PropertyName>
  <ogc:Literal>20</ogc:Literal>
 </ogc: PropertyIsLessThanOrEqualTo >
</ogc:And>
                    (a3)

<granuleCondition>
 <cloudCover>
  <range lower="10" upper="20"/ >
 </cloudCover>
</granuleCondition>
                    (b3)
```

Figure 11. Query Criteria Transformation Example (3).

possible logical relationship among query criteria is the logical **AND** operation, any user queries that contain <ogc:Or> predicates may not be fulfilled on NASA ECHO. Based on the aforementioned three steps, a federation service is capable of creating the final transformation at the query payload level by considering those non-query related elements and attributes.

Query Result Integration

A federation service needs to integrate query results from multiple underlying Catalogue Services before sending them back to the clients. It may choose to implement one of three kinds of integration mechanisms.

1. Opaque: In this case, the federation service defines, maintains and advertises a unique information model. Each query result from affiliated Catalogue Services should, if necessary, be transformed to this information model. The original metadata information can be kept in the final transformed query results.
2. Translucent: The federation service does not maintain a complete, unique information model but defines a common subset of metadata objects that are supported by all the affiliated Catalogue Services, such as name, and spatial and temporal range. The federation service transforms only this part of the metadata information, while the remaining embedded original metadata information remains unchanged in the final response.
3. Transparent: The federation service has no role in metadata integration. All the query results from affiliated Catalogue Services are simply grouped together, keeping the original metadata formats. In this scenario, the users are supposed to analyze each result fetched from federation service, since the results may not all conform to the same schema even though grouped together in one response.

Federation System

The GMU Catalogue Federation Service (CFS) product, shown in Figure 12, which is based on the design strategies discussed in the previous section have been developed at GMU.

The GMU CFS consists of two types of components: Mediator and Wrapper. The Mediator is a key component of the GMU CFS. It accepts user's queries through an OpenGIS CSW query interface. A *GMU OGC Filter Decoder* module was developed and deployed to fulfill query interpretation. Mediator introduces a *Request ID management* component, which is responsible for assigning a unique ID for each user query. This ID is kept in all the messages between the GMU CFS and all three affiliated catalogue services for each user query, facilitating tracking of the query process procedures in the multi-thread environment. To facilitate the query transformation, a *Query Evaluation* module was devised. This module is equipped with knowledge about how transformations on *Metadata Term, Query Criterion, Query Criteria*, and *Query Payload* could be carried out. Information Model Mapping Rules are hard-coded in this module. This module evaluates which Catalogue Service(s) will be dispatched. XSL Transformations are used for the necessary query transformations in this step.

For simplicity, the GMU CFS follows the opaque query-dispatching pattern. By default, it always queries first against the GMU CSW, then against the OGC CSW for ECHO and finally the ESG Catalogue. A multi-thread request and response queue is implemented to facilitate the query process. The GMU CFS follows the opaque query results integration scheme, while keeping all the original metadata information in the result response for user reference.

Figure 13 depicts the GMU CFS user query processing flow. When getting a user query in the OGC Filter format, the system assigns a unique request identifier to it for tracking use. Then the CFS system does a syntactic analysis to retrieve *maxRecords* and *startPosition*; the CFS system also performs a semantic analysis to decide which catalogue services to query against. Before calling each catalogue service, the CFS system transforms the query criteria accordingly; and

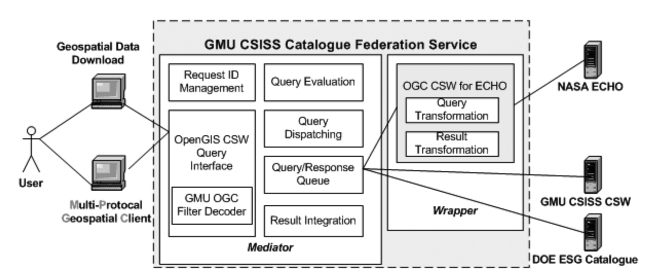

Figure 12. GMU Catalogue Federation Service. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

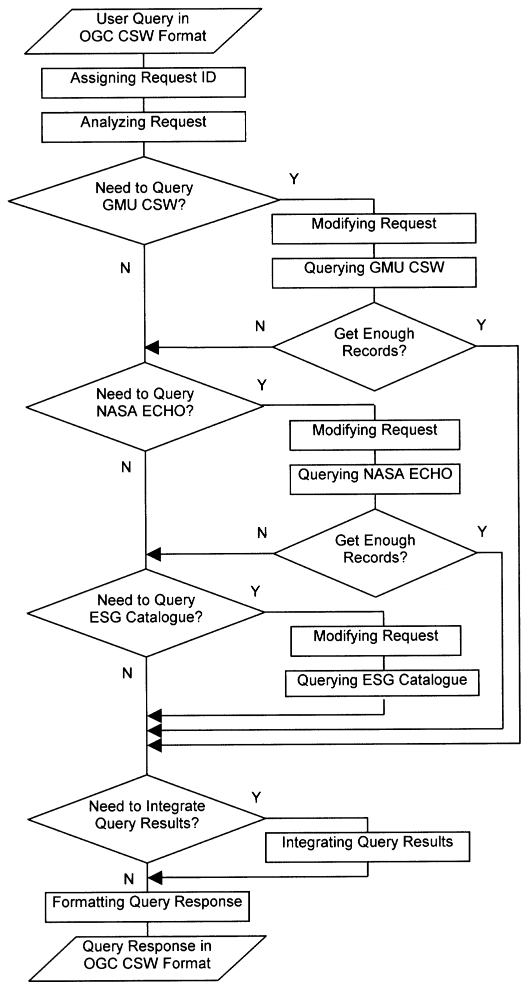

Figure 13. GMU CFS system flow diagram.

adjusts the *maxRecords* and *startPosition* dynamically according to the number of records that have already been fetched. Lastly, the results from the multiple Catalogue Services are integrated before being sent back to the users.

Wrapper: OGC CSW for ECHO

Since the GMU CFS only supports the OpenGIS CSW query interface, every encoded query needs to be transformed into ECHO IIMSAQL form to communicate with ECHO, and ECHO query results need to be transformed before being sent back to the Mediator. For better modularity and sustainability, the GMU CFS has externalized the query translation and the result transformation module for ECHO to make a wrapper, i.e., OGC CSW for ECHO. This wrapper accepts client queries, in OGC Filter format, that are forwarded by the Mediator. They are transformed to ECHO IIMSAQL format by the *Query Transformation* module. This transformation consists of two steps: a collection level search and a granule-level search. With the support of the GMU OGC Filter Decoder and the CSISS OGC CSW Core software libraries, the wrapper provides an OGC CSW interface for ECHO.

This GMU CFS system is operational in the GeoBrain projects. As a cornerstone, it helps GeoBrain partners easily find NASA EOS data and information for teaching and research purpose. It enables education community users to discover the online datasets in NASA data pools, the GMU Spatial Database, and the DOE ESG LLNL Simulation Database in a convenient and standard way.

Discussion

This GMU CFS system can integrate NASA ECHO, GMU CSW and the DOE ESG Simulation Data Catalogue. One advantage of its design is that CFS follows the OpenGIS Catalogue Service standard as the communication protocol with the underlying affiliated catalogue services. As long as new catalogue services follow this standard, they can easily be integrated into the federation system. However, integrating new legacy catalogue services cannot be plug and play. Abstracting the specific information models, the catalogue registration mechanism, and query orchestration would be new issues to consider when scaling this federation beyond these three catalogue services.

In the strategy described here, the specific information models need to be carefully evaluated before incorporation into the global schema and subsequent exposure to client users, when including new catalogue services. Efforts such as the ISO Technical Committee (TC) 211 Geospatial Metadata Standard 19115, Content Standard for Digital Geospatial Metadata (FGDC, 1998) and Dublin Core (DCMI, 2006) are attempting to standardize the geospatial metadata information model, but, in many cases, their use is voluntary (Flewelling and Egenhofer, 1999). A metadata crosswalk could be of help when mapping two distinct models, but not very suitable for one-to-many mapping. An ontology-based approach (Hakimpour and Geppert, 2001; Giger and Najar, 2003; Lutz and Klien, 2006) would provide a new way to create a global schema.

New catalogue services must be registered manually. In the design presented here, the federation service discovers the underlying catalogue services at design time, rather than at run time. This strategy greatly simplifies the mechanism for federation, and lowers the complexity of implementation. In fact, without an automatic way to fulfill information model integration, it does not make much sense to perform the automatic catalogue service registration.

Because of the complex heterogeneities of information models and query languages, an XSLT-based proprietary tool has been developed to hard-code that knowledge to transform queries and integrate results. We did not introduce formal descriptions for query orchestration. Besides these scalability issues, future work will consider development of a general cost model for parsing, evaluating, and distributing queries, and for assembling the results.

Conclusion

This paper presents research on a federation service for geospatial catalogues through a case study of building integration over three legacy catalogue services: NASA ECHO, GMU CSW, and the DOE ESG simulation catalogues. We summarized the research issues, proposed general design strategies, and introduced the GMU CFS. The following conclusions were reached:

- Protocol adaptation, query dispatching, query translation, and query results integration are the four main challenges that must be met to reach a catalogue federation
- A mediator-wrapper based approach can be adopted to build the resulting federation service
- The OpenGIS Catalogue Service specification can be used to define the internal communication protocols between the federation service and the affiliated catalogue services, and between the federation service and its clients.

Acknowledgments

This study was supported by grants from NASA through the REASoN program (NNG04GE61A; Professor Liping Di, Principal Investigator). The authors would like to thank the anonymous reviewers for their time and efforts, valuable comments, and insightful ideas.

References

Alpdemir, M.N., A. Mukherjee, A. Gounaris, A.A.A. Fernandes, N.W. Paton, and P. Watson, 2003. An experience report on designing and building OGSA-DQP: A service based distributed query processor for the grid, *Proceedings of Global Grid Forum Workshop on Designing and Building Grid Services*, 08 October, Chicago, Illinois

Bai, Y., and L. Di, 2005. GMU CSISS OGC Catalog Service for Web – Discovery Interface, URL: *http://geobrain.laits.gmu.edu/csw/discovery/*, CSISS, GMU, Greenbelt, Maryland (last date accessed: 22 February 2007).

Ballinger, K., D. Ehnebuske, M. Gudgin, M. Nottingham, and P. Yendluri, 2004. WS-I Basic Profile 1.0, URL: *http://www.ws-i.org/Profiles/BasicProfile-1.0-2004-04-16.html*, Web Services Interoperability Organization (last date accessed: 22 February 2007).

Barillot, C., H. Benali, O. Dameron, M. Dojat, A. Gaignard, B. Gibaud, S.Kinkingnéhun, J.-P. Matsumoto, M. Pélégrini-Issac, E. Simon, L.Temal, and R.Valabregue, 2006. Federating distributed and heterogeneous information sources in neuroimaging: The NeuroBase Project, *Studies in Health Technology and Informatics*, 120:3–13.

Baru, C., A. Gupta, B. Ludäscher, R. Marciano, Y. Papakonstantinou, P. Velikhov, and V. Chu, 1999. XML-Based information mediation with MIX, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 01–03 June, Philadelphia, Pennsylvania, pp. 597–599.

CEOS, 2005. Catalogue Interoperability Protocol – (CIP) Specification – Release B, URL: *http://wgiss.ceos.org/ics/documents/cip2.4/cipspec-2_4_75_6.pdf#search=%22Catalogue%20Interoperability%20Protocol%20-%20(CIP)%20Specification%20-%20Release%20B%22*, CEOS (last date accessed: 22 February 2007).

Christensen, E., F. Curberam, G. Meredith, and S. Weerawarana, 2001. Web Services Description Language (WSDL) 1.1, URL: *http://www.w3.org/TR/2001/NOTE-wsdl-20010315, W3C* (last date accessed: 22 February 2007).

Crompvoets, J., A. Bregt, A. Rajabifard, and I. Williamson, 2004. Assessing the worldwide developments of national spatial data clearinghouses, *International Journal of Geographical Information Science*, 18(7):665–689

David, B., S. Bharathi, D. Brown, K. Chanchio, M. Chen, A. Chervenak, L. Cinquini, B. Drach, I. Foster, P. Fox, J. Garcia, C. Kesselman, R. Markel, D. Middleton, V. Nefedova, L. Pouchard, A. Showshani, A. Sim, G. Strand, and D. Williams, 2005. The Earth system grid: Supporting the next generation of climate modeling research, *Proceedings of the IEEE*, 93(3): 485–495.

DCMI, 2006. DCMI Metadata Terms, URL: *http://dublincore.org/documents/dcmi-terms/*, Dublin Core Metadata Initiative (last date accessed: 22 February 2007).

Dwyer, P.A., and J.A. Larson, 1987. Some experiences with a distributed database testbed system, *Proceedings of the IEEE*, 75(5):633–648.

ECHO, 2005. Earth Observing System Clearinghouse, URL: *http://www.echo.eos.nasa.gov/*, GST, Greenbelt, Maryland (last date accessed: 22 February 2007).

FGDC, 1998. Content Standard for Digital Geospatial Metadata, URL: *http://www.fgdc.gov/metadata/csdgm/*, Federal Geographic Data Committee (last date accessed: 22 February 2007).

Flewelling, D.M., and M.J. Egenhofer, 1999. Using digital spatial archives effectively, *International Journal of Geographical Information Science*, 13(1):1–8.

Giger, C., and C. Najar, 2003. Ontology-based integration of data and metadata, *Proceedings of the 6th AGILE Conference on Geographic Information Science*, Lyon, France.

GST, 2006. IIMSAQL Query Language DTD, URL: *http://www.echo.nasa.gov/client_partners/client_tools7.shtml*, GST, Greenbelt, Maryland (last date accessed: 22 February 2007).

Gupta, A., A. Memon, J. Tran, R.P. Bharadwaja, and I. Zaslavsky, 2002, Information mediation across heterogeneous government spatial data sources, *Proceedings of the 2002 Annual National Conference on Digital Government Research*, 19–22 May, Los Angeles, California, pp. 1–6.

Hakimpour, F., and A. Geppert, 2001. Resolving semantic heterogeneity in schema integration: An ontology based approach, *International Conference on Formal Ontology in Information Systems*, Ogunquit, Maine, ACM Press.

Hosokawa, Y., N. Takahashi, and H. Taga, 2004. A system architecture for seamless navigation, *Proceedings of the 24th International Conference on Distributed Computing Systems Workshops*, 23–26 March, Tokyo, Japan, pp. 498–504.

ISO, 1998. ISO 23950: Information and documentation - Information retrieval (Z39.50) – Application service definition and protocol specification, URL: *http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=27446&ICS1=35&ICS2=240&ICS3=30*, International Standard Organization (last date accessed: 22 February 2007).

ISO, 2003. ISO 19115: Geographic Information – Metadata, URL: *http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26020&ICS1=35&ICS2=240&ICS3=70*, International Standard Organization (last date accessed: 22 February 2007).

ISO, 2005. ISO 19119: Geographic information - Services, URL: *http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39890&ICS1=35&ICS2=240&ICS3=70*, International Standard Organization (last date accessed: 22 February 2007).

ISO, 2006. ISO/CD TS 19139: Geographic information - Metadata - XML schema implementation, URL: *http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=32557&scopelist=PROGRAMME*, ISO (last date accessed: 22 February 2007).

Katchaounov, T., T. Risch, and S. Zürcher, 2002. Object-oriented mediator queries to Internet search engines, *Lecture Notes in Computer Science* (J.-M. Bruel and Z. Bellahsene, editors), Springer Verlag, Berlin, Germany, pp.176–186.

Kottman, C., 1999. OpenGIS Abstract Specification Topic 13: Catalog Services (OGC 99-113), URL: *http://www.opengeospatial.org/specs/?page=abstract*, OpenGIS (last date accessed: 22 February 2007).

Litwn, W., 1985. An overview of the multidatabase system MRDSM, *Proceedings of the ACM Annual Conference on the Range of Computing: Mid-80's Perspective*, 14–16 October, Denver, Colorado, pp. 524–533.

Lin, H., T. Risch, and T. Katchaounov, 2000. Object-oriented mediator queries to XML data, *Proceedings of the First International Conference on Web Information Systems Engineering*, 19–20 June, Hong Kong, China, pp. 2039.

Lutz, M., and E. Klien., 2006. Ontology - based retrieval of geographic information, *International Journal of Geographical Information Science*, 20(3):233–260.

Martell, R., 2005. OGC Catalogue Services – ebRIM (ISO/TS 15000-3) profile of CSW (OGC 04–017r1), URL: *https://portal.opengeospatial.org/files/?artifact_id=7048*, OpenGIS (last date accessed: 22 February 2007).

MacHarrie, P., and M. McBride, 2002. Release 6B Implementation Earth Science Data Model, URL: *http://edhs1.gsfc.nasa.gov/waisdata/rel6/html/tp4202301.html*, NASA GSFC, Greenbelt, Maryland (last date accessed: 22 February 2007).

Miller, L., and S. Nusser, 2003. An infrastructure for supporting spatial data integration, *Federal Committee on Statistical Methodology Conference*, 17–19 November, Arlington, Virginia.

Mitra, N., 2003. Simple Object Access Protocol (SOAP) 1.1, URL: *http://www.w3.org/TR/soap/*, W3C (last date accessed: 22 February 2007).

Melnik, S., H. Garcia-Molina, and A. Paepcke, 2000. A mediation infrastructure for digital library services, *Proceedings of the 5th ACM conference on Digital libraries*, 02–07 June, San Antonio, Texas, pp. 123–132.

Naumann, F., 2002. *Quality-Driven Query Answering for Integrated Information Systems*, Springer Verlag, Berlin, Germany, 166 p.

Nebert, D.D., 2000. Z39.50 Application Profile for Geospatial Metadata, URL: *http://www.blueangeltech.com/standards/GeoProfile/geo22.htm*, U.S. Geological Survey, Reston, Virginia (last date accessed: 22 February 2007).

Nebert, D., and A. Whiteside, 2004. OpenGIS Catalogue Service Implementation Specification (OGC 04–021r3), URL: *http://www.opengeospatial.org/standards/cat*, OpenGIS (last date accessed: 22 February 2007).

OASIS, 2005. OASIS/ebXML Registry Information Model, URL: *http://www.oasisopen.org/committees/regrep/documents/2.5/specs/ebrim-2.5.pdf*, OASIS (last date accessed: 22 February 2007).

Pfister, R., R. Ullman, and K. Wichmann, 2001. ECHO rediscovers and responds to NASA's Earth science user community, *Proceedings of the First International Conference on Universal Access in Human-Computer Interaction*, 05–10 August, New Orleans, Louisiana, pp. 981–985.

Ramroop, S., and R. Pascoe, 2001. Use of LDAP to partially implement the OGIS discovery service, *International Journal of Geographical Information Science*, 15(5):391–413.

Sheth, A.P., and J.A. Larson, 1990. Federated database systems for managing distributed, heterogeneous, and autonomous databases, *ACM Computing Surveys*, 22(3):183–236.

Tait, M.G., 2005. Implementing geoportals: Applications of distributed GIS, *Computers, Environment and Urban Systems*, 29(1):33–47.

Tzitzikas, Y., N. Spyratos, and P. Constantopoulos, 2005. Mediators over taxonomy-based information sources, *International Journal on Very Large Data Bases*, 14(1):112–136.

Unidata, 2004. THREDDS Dataset Inventory Catalog Specification Version 1.0, URL: *http://www.unidata.ucar.edu/projects/THREDDS/tech/catalog/InvCatalogSpec.html*, Unidata, Boulder, Colorado (last date accessed: 22 February 2007).

Voges, U., and K. Senkler, 2005. OpenGIS Catalogue Services Specification 2.0 - ISO19115/ISO19119 Application Profile for CSW 2.0 (OGC 04-038r2), URL: *https://portal.opengeospatial.org/files/?artifact_id=8305*, OpenGIS (last date accessed: 22 February 2007).

Vretanos, P.A., 2005. OpenGIS Filter Encoding Implementation Specification (OGC 04–095), URL: *http://www.opengeospatial.org/standards/filter*, OGC (last date accessed: 22 February 2007).

Wiederhold, G., 1992. Mediators in the architecture of future information systems, *IEEE Computer*, 25(3):38–49.