PFG 2013 / 1, 0043–0052
Stuttgart, February 2013

Report

Federated Catalogue for Discovering Earth Observation Data

Yuanzheng Shao, Liping Di, Fairfax, VA, USA, Yuqi Bai, Beijing, China, Lingjun Kang, Huilin Wang, Fairfax, VA, USA & Chao Yang, Wuhan, China

Keywords: earth observation, federated catalogue, mediator-wrapper architecture, geospatial data discovery, CEOS WGISS Integrated Catalogue

Abstract: As Earth observation (EO) technologies develop continuously, the volumes of geospatial data archived in data centers also grow. Those geospatial data can be used in many scientific fields such as agriculture, land use and climate change. The complex features and massive amounts of EO data bring challenges on how to make better use of these data for the community and for public service. Most of legacy data center distribute the data through individual catalogue, and use heterogeneous query interface and metadata model. To find the data of interest from multiple data source for multi-disciplinary research, the scientists and data users need to handle different query languages and metadata models. To facilitate Earth observation data discovery for researchers, this paper proposes a federated catalogue to integrate multiple legacy data centers. By analysing existing data discovery mechanism, mediator-wrapper framework was adopted to implement catalogue federation. By solving the query interface translation and metadata model conversion between federated catalogues and individual data centers, a system – The CEOS WGISS Integrated Catalogue (CWIC) was implemented. CWIC was proved to be an effective tool to discovery geospatial data from multi-source data centers by interacting with standardized query interface and metadata model. The architecture and approaches proposed in this paper can be used to establish a federated catalogue system for different communities.

Zusammenfassung: *Konzept für einen Zentralkatalog für Fernerkundungsdaten.* Mit dem fortschreitenden Ausbau der Erdbeobachtung steigen auch die zu archivierenden Datenmengen. Die Daten finden in den Geowissenschaften vielfältig Verwendung, z. B. in der Landwirtschaft, beim Monitoring von Landnutzungsänderungen und beim Klimawandel. Die komplexe Datenstruktur und die riesigen Datenmengen sind eine Herausforderung an die geeignete Bereitstellung für private und öffentliche Zwecke. Die meisten staatlichen Datenzentren bieten ihre Daten über eigene Kataloge und sehr unterschiedliche Benutzerschnittstellen und Metadatenmodelle an. Für fächerübergreifende Untersuchungen müssen die Benutzer unterschiedliche Abfragesprachen einsetzen. Zur Vereinfachung wird daher in diesem Artikel ein Konzept für einen Zentralkatalog für die übergreifende Suche in vielen Datenzentren vorgestellt. Nach der Analyse verschiedener vorhandener Lösungen wurde das „Mediator-Wrapper“-Modell für den Zentralkatalog zu Grunde gelegt. Der Zentralkatalog wurde als „CEOS WGISS Integrated Catalogue (CWIC)“ implementiert, nachdem die Transformation der Abfragen (queries) und der Metadaten in eine standardisierte Form gelöst war. Inzwischen hat der CWIC seine Leistungsfähigkeit für katalogübergreifende Abfragen unter Beweis stellen können. Die in diesem Artikel vorgestellte Architektur kann zur Einrichtung von Zentralkatalogen auch in anderen Communities empfohlen werden.

1 Introduction

As the development of Earth observation technology progresses, more and more geospatial data used for scientific research are available from various geospatial agencies and government departments. Those geospatial data could be used in many scientific fields such as agriculture, land use and climate change. To find the data of interest from multiple data sources for multi-disciplinary researches, the scientists and data users need to get familiar

www.schweizerbart.de
DOI: 10.1127/1432-8364/2013/0157
1432-8364/13/0157 $ 2.50

with the individual data center and the corresponding web portal. Most of data centers require data users to register an account before searching and acquiring data, which require data users to maintain many user account information for every data center. For some archived data without online access URLs, the users need to place an order against specific data center and get the notification of data access URLs from registered email. Such work is very time-consuming and tedious, especially when the catalogues use different metadata models and catalogue interface protocols (Bai et al. 2007). Moreover, this human-computer interaction of data discovery brings more challenges for data users to build an automatic workflow for some scientific research.

To solve such challenge, it is highly required to build an integrated catalogue, also known as catalogue federation, to conduct the data discovery in a consistent way. This federated catalogue, by solving the inconsistency among multiple data centers, presents a standardized data query interface and metadata model. With the catalogue federation, data user only needs to interact with the federated catalogue to find data of interest other than working with different catalogues. This paper proposes a solution for a geospatial catalogue federation, which integrates some major legacy catalogues, to facilitate geospatial data discovery.

In order to implement a federated catalogue for geospatial data discovery, the mediator-wrapper (Wiederhold 1992) architecture was selected. This architecture has been widely used for the integrated and universal access to multiple, autonomous information sources. Distributed data sources (legacy data center) archive data and distribute them through the Internet. Such architecture has been successfully used in many research areas, such as integration of GIS data source (Stoimenov et al. 2000), digital libraries (Melnik et al. 2000), intelligent web application (Sahuguet & Azavant 2001), and XML data (Baru et al. 1999, Lin et al. 2000).

Song et al. (2006) designed a grid-enabled information integration system which can access heterogeneous and distributed data sources through the standard interface of grid computing. Stoimenov et al. (2000) provided intelligent integration of information using mediation technology, which is used for the integration of spatial data from geographic information system (GIS) databases with alphanumerical data from relational database management systems and other data sources. Hidalgo et al. (2006) described a cost model that stores values from a complete set of web source-focused parameters obtained by the web wrappers in a mediator-wrapper environment. Based on a mediator-wrapper approach, Corcho et al. (2003) designed a framework which supports an integrated view over multiple heterogeneous sources for e-commerce application. Miller & Nusser (2003) presented 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.

Most of the previous work about data discovery focused on Earth observation retrieval. Geospatial metadata, as an important and effective way to help data users to understand the retrieved data, usually attracted little attention. This paper uses the generic metadata model – ISO 19115:2003 (Geographic information – Metadata) – to describe the data granules returned from a federated catalogue. The architecture, global query interface and metadata model proposed in this paper are used by the Working Group on Information Systems and Services (WGISS) of the Committee on Earth Observation Satellite (CEOS) (WGISS 2012), and are proved to be useful for catalogue federation in Earth observation community.

The rest of this paper is structured as follows. Section 2 introduces and compares some major existing data centers and their data discovery mechanism, including their data distribution policies, query interface and metadata model. Section 3 analyses the challenges and proposes solutions to coordinate different data discovery mechanisms, including query interface and metadata model selection. Section 4 presents the federated catalogue architecture and the implementation of the CWIC system. The conclusion and discussion are drawn in section 5.

2 Existing Data Discovery Mechanism

There are some major data centers worldwide which archive massive and heterogeneous geospatial data. This paper focuses on the following seven legacy geospatial catalogues: the NASA EOS ClearingHOuse (ECHO), the NOAA's Comprehensive Large Array-data Stewardship System (CLASS), U.S. Geology Survey (USGS) – Landsat Catalogue System, the Brazil National Institute for Space Research (INPE) catalogue, the China Academy of Opto-Electronics (AOE) catalogue, the Japan Aerospace Exploration Agency (JAXA) and the Group for High-Resolution Sea Surface Temperature (GHRSST) catalogue. All of those data centers have already provided a web portal or a service to enable data user to discover data of interest. The following part of this chapter will analyze the data discovery mechanism of some of them, and pay more attention to the comparison of query interface and metadata model.

2.1 *NASA ECHO Data Catalogue*

The NASA-developed Earth Observing System (EOS) Clearinghouse (ECHO) is a spatial and temporal metadata registry and order broker built by NASA's Earth Science Data and Information System (ESDIS) that enables the science community to more easily use and exchange NASA's data and services (ECHO 2012).

1) Data query interface: ECHO developed a query language, named IIMSAQL, to enable the query on collection-level and granule-level in the ECHO system. This is an XML-based language with a detailed DTD definition. IIMSAQL defined the element for specifying collection identifier, spatial extent, temporal range, data center identifier, and many other features, such as Percentage of Cloud Cover, of the data center and collection.
2) Metadata Conceptual Model: the ECHO's metadata conceptual model is named as ECHO Earth Science Metadata Conceptual Model (EESMCM). It is developed from the EOSDIS Core System (ECS) Science Data Model. This metadata model is used to describe collection and granule. In the ECS, a granule is the smallest unit of data that is independently described and taken stock, while a collection represents a logical grouping of granules. Each Collection is identified with an Earth Science Data Type (Wei et al. 2007).

2.2 *NOAA CLASS Data Catalogue*

The CLASS system is an electronic library of NOAA environmental data. This web site provides capabilities for finding and obtaining those data. CLASS provides more than 70 data collections. The data archived in CLASS system do not have an online access URL, so the client has to register an account with email address, and the query response containing access URLs will be sent to the specified email address. Beside the web site, NOAA Enterprise Archive Access Tool (NEAAT) is also developed by NOAA to enable data query through an Application Programmer Interface. NEAAT is an access API designed to integrate various heterogeneous NOAA data systems for the purposes of data discovery and retrieval.

1) Data query interface: NEAAT supports a number of query types for both catalogue- and inventory-level. Both simple queries and compound quires are supported. There are four comparisons implemented in NEAAT API: string, algebraic, temporal and spatial. NEAAT provides Java Archive (JAR) files which expose many easy-to-use functions. Data clients need to call the corresponding function to specify query parameters and execute a data inventory query.
2) Metadata Conceptual Model: the CLASS system defines a very simple XML model to encode the returned data granule. NEAAT contains limited metadata about returned granule, but the basic spatial extent, temporal range, platform and data size information are included.

2.3 *USGS Landsat Data Catalogue*

LandSAT data has wide use in geospatial-related scientific research. The USGS offers all users the entire Landsat 1–5 and 7 archive

data at no charge using a standard data product recipe. The user can discover Landsat data through several web portals, such as EarthExplorer, GloVis and GeoBrain. Based on these portals, data users could discover and access all types of Landsat data.

1) Data query interface: the USGS releases a Web API which could be used to discover Landsat data. The access URL for this Web API is as follows: *http://edcsns17.cr.usgs.gov/EE/InventoryStream/latlong?* Some simple parameters could be specified in the data query request of the Web API, including collection identifier, spatial bounding box, temporal range and the number of returned records.
2) Metadata Conceptual Model: A simple XML model is defined to describe the response from the USGS Landsat catalogue. The response from the USGS Landsat catalogue includes information like browser image URL, data ordering URL, platform, cloud cover rate and spatial-temporal information.

2.4 *China AOE Data Catalogue*

China has launched several EO satellites, such as the CBRES series, the FengYun series and so on. The data products generated from those satellites have been widely applied to the field of agriculture and other fields. The EO data have good catalogue and archive systems in several satellite data centers (Feng et al. 2011).

To facilitate data discovery for those EO data, the AOE of the Chinese Academy of Science designed and developed an OGC Catalogue Service for the Web (CSW). AOE CSW adopts the OGC CSW standard as the query interface, and the ISO 19115 standard to describe metadata.

2.5 *JAXA Data Catalogue*

JAXA is an independent administrative institution to be able to perform all activities in the aerospace field as one organization, from basic research and development to utilization. JAXA has archived a huge amount of Earth observation data products, such as the Advanced Land Observation Satellite (ALOS), the Greenhouse gases Observing SATellite (GOSAT) and the Advanced Earth Observing Satellite (ADEOS). JAXA provides the Earth Observation Satellite Data Search/Order System (JAXA 2012) which facilitates data users to retrieve the archived data through a web portal-based interface.

Besides the web portal, JAXA also developed its CSW service to distribute its archived data based on GeoNetwork and deployed the CSW on Amazon Cloud Computing platform. Both the ebRIM profile and the ISO profile are implemented.

3 Challenges to Integrate Different Data Catalogues

Based on the analysis and the comparison in the second section, it is noticeable that heterogeneity exists among those data catalogues. To design and implement a federated catalogue, which aims to integrate different catalogues and provides a universal data discovery mechanism, a global data query language and metadata model should be selected. Moreover, the mapping between the federated catalogue and the individual catalogues also should be established. This chapter will address the challenges in data discovery mechanisms coordination and its solutions.

3.1 *Global Query Interface Selection*

In order to implement an integrated catalogue, a global query interface should be adopted. The OpenGIS CSW specification defines three protocol bindings for the catalogue service: Z39.50, CORBA/IIOP and HTTP. In this paper, the OGC CSW version 2.0.2 is implemented. The OpenGIS CSW specification proposes an OGC common catalogue query language, which has to be supported by all compliant OpenGIS catalogue services (Voges & Senkler 2007). This query language supports nested Boolean queries, text comparison operations and temporal-spatial operators. The federated catalogue in this paper implements the OGC Filter specification (Vretanos 2010).

The heterogeneity feature of query interfaces of different data catalogues brings challenges to determine queryables of a global query interface. To translate the query language from a global to a native query interface, the federated catalogue has to solve the following problems:

1) Asymmetry of metadata queryables: the queryables defined in the global query language may not have the corresponding one in a legacy data catalogue. The queryables supported by different data centers also are various. For example, NASA ECHO query API supports the granule query against cloud cover rate, but other data centers do not support such query. To solve the asymmetry of metadata queryables, the federated should select the minimal subset of all queryables implemented among different data centers. By analysing those legacy catalogues, the dataset identifier, spatial bounding box and temporal range query parameters are selected as the mandatory queryables of a federated catalogue, which are also the most important attributes of geospatial data.
2) Heterogeneity of query formats: as described in section 2, those legacy data catalogues have different query interfaces and formats. For example, NOAA CLASS uses native NEAAT API package, USGS and INPE accepts Web API query format, while ECHO only accepts IIMSAQL format. Even for the catalogues implementing a CSW service, such as AOE, JAXA and GHRSST, the queryables name for the same metadata object may also be different. To solve the heterogeneity of query formats, a federated catalogue has to translate the queryables and query format from a global query language to the corresponding query language supported by an individual data catalogue.

The HTTP protocol is selected as the protocol binding, and both HTTP GET and HTTP POST methods are implemented. The following operation shall be implemented based on the CSW specification.

GetCapabilities: this mandatory operation allows CSW clients to retrieve service metadata from a server. The response of GetCapabilities request shall be an XML document containing service metadata about the server.

DescribeRecord: this mandatory operation allows a client to discover elements of the information model supported by the target catalogue service. The operation allows some or the entire information model to be described.

GetDomain: this optional operation is used to obtain runtime information about the range of values of a metadata record element or a request parameter. The runtime range of values for a property or a request parameter is typically much smaller than the value space for that property or parameter based on its static type definition.

GetRecords: this mandatory operation is used to search the catalogue content. It supports not only requesting metadata records at different levels of detail (full, summary, and brief) and in different output schemas, but also allows for paging and sorting, as well as stating the filter criteria in different query languages.

GetRecordById: this mandatory operation retrieves the default representation of a catalogue records using their identifier. The records identifier could be retrieved from the GetRecords response.

3.2 *Global Metadata Model Selection*

The existing data centers adopted different metadata models to describe their archived data. In order to provide a universal metadata schema to describe data returned from all catalogue, it is necessary to define a global metadata model for a federated catalogue, and build the mapping from individual metadata model to the global one.

Geographical information – Metadata (ISO 19115) and its extension for imagery and gridded data (ISO 19115-2), which define comprehensive and detailed elements to describe geospatial data, are widely used in the Earth observation field (ISO 2003). Some software packages, such as GeoNetwork and ESRI GeoPortal, have overall support of the ISO 19115 metadata model. In this paper, ISO 19115 is selected as the global metadata model for the federated catalogue.

Some legacy catalogues, such as GHRSST, AOE and JAXA have already implemented the CSW service and fully support the global metadata model. For such data center, federated catalogue does not need to make any conversion for the returned metadata. But for the data center with a customized metadata model, such as the ECS model defined in ECHO, the federated catalogue has to convert the metadata from their native schema to a global one.

Since the richness of the metadata returned from different catalogues is also various, the asymmetry of the metadata model is becoming another challenge to implement a catalogue federation. The asymmetry features are expressed in two aspects: some metadata items in a legacy data center cannot be mapped to corresponding elements in a global metadata model; some mandatory elements in the global metadata model cannot find the corresponding elements in the response of a legacy catalogue. For the first problem, the current solution is to ignore the elements that cannot be mapped to global metadata. For the latter problem, the a priori knowledge is used to fill the mandatory elements in the global metadata. Taking NASA ECHO as example, the point of contact information are required in ISO 19115 metadata, but the XML response returned from ECHO does not include any contact information. For such situation, we can find the point of contact information from the NASA Distributed Active Archive Centers (DAAC) or their web portal, and fill that information to mandatory element of global metadata.

4 Federated Catalogue Implementation

4.1 *System Architecture Design*

The mediator-wrapper architecture is used to design the federated catalogue. As shown in Fig. 1, data is archived in heterogeneous models and disseminated through different interfaces to the data sources. The wrapper layer tackles the heterogeneity through providing a universal querying interface (i.e. OGC CSW) and an intermediate information model, i.e. Dublin Core, ISO 19115. Mediator, on the other hand, acts as a proxy accepting the query from data users, dispatching the request to corresponding wrapper, assembling the query results from individual data center and returning it to the data users.

Specifically, the interactions among the user, mediator, and wrapper are summarized as follows:

1) The data client queries against mediator through the OGC CSW interface.
2) The mediator parses the query and dispatches it to a related wrapper.

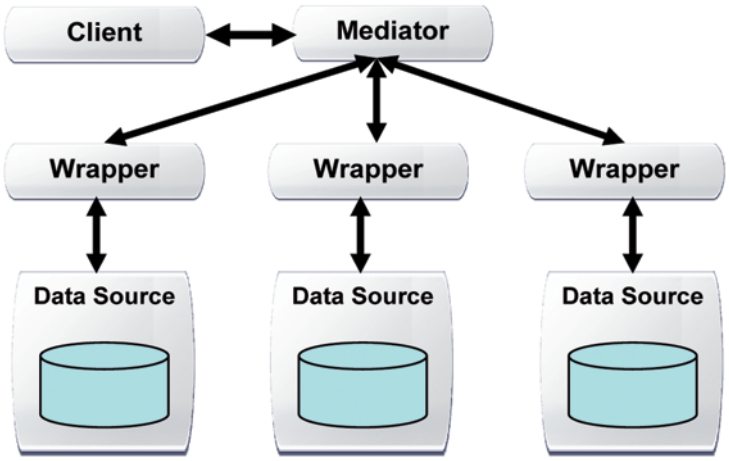

Fig. 1: The mediator-wrapper architecture.

3) The wrapper converts the request to a local query language and queries against a data catalogue. After getting a response, the wrapper converts the response to an intermediate information model and returns it to the mediator.
4) The mediator sends the response to the data users.

4.2 *CEOS WGISS Integrated Catalogue*

The Committee on Earth Observation Satellite (CEOS) is the international inter-agency organization addressing the coordination of the satellite EO programs of the world's governmental space agencies. The Working Group on Information Systems and Services (WGISS) is a subgroup of CEOS, which aims at promoting collaboration in the development of systems and services that manage and supply EO data to users world-wide. With the initiation by NOAA and NASA, WGISS started the CWIC project in 2010, aiming at creating a federated catalogue system for providing inventory-level search to catalogue systems of all major CEOS members around the world through a standard-based unified interface.

The architecture, query interface and metadata model proposed in the previous section are used in the project of CWIC. CWIC is designed to provide a single access point for the major CEOS agency catalogue systems (Di et al. 2012).

As of August 2012, seven data catalogues described in section 2 have been integrated in the CWIC service. The framework of CWIC system is shown in Fig. 2, and the interaction between major modules within this framework is presented in Fig. 3.

The harvest operation is not implemented in the CWIC service, but the following operations, defined in OGC CSW specification, have been implemented and introduced in section 3.1. They are the operations GetCapabilities, DescribeRecord, GetDomain, GetRecords and GetRecordById.

To support a query interface translation, the mapping between OGC CSW query interface and the local query interfaces is established. At the wrapper level, all queries dispatched from the mediator will be translated to a native query interface. For example, the query about an ECHO dataset will be translated to the IIMSAQL language. Regarding the metadata model conversion like query interface the mapping between the ISO 19115 and the metadata model used in legacy catalogues is also created. The issue that needs to be solved during the conversion is the asymmetry feature of different metadata.

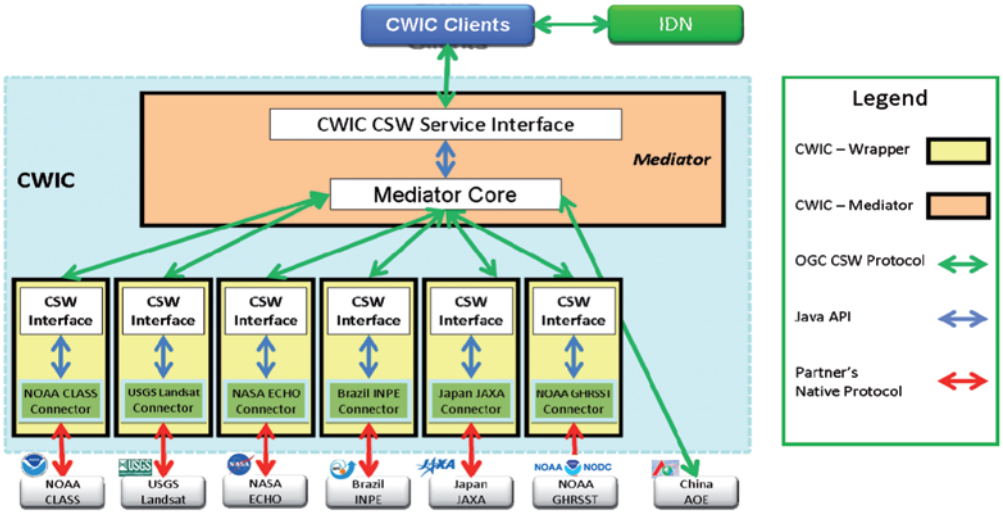

Fig. 2: The framework of CEOS WGISS Integrated Catalogue.

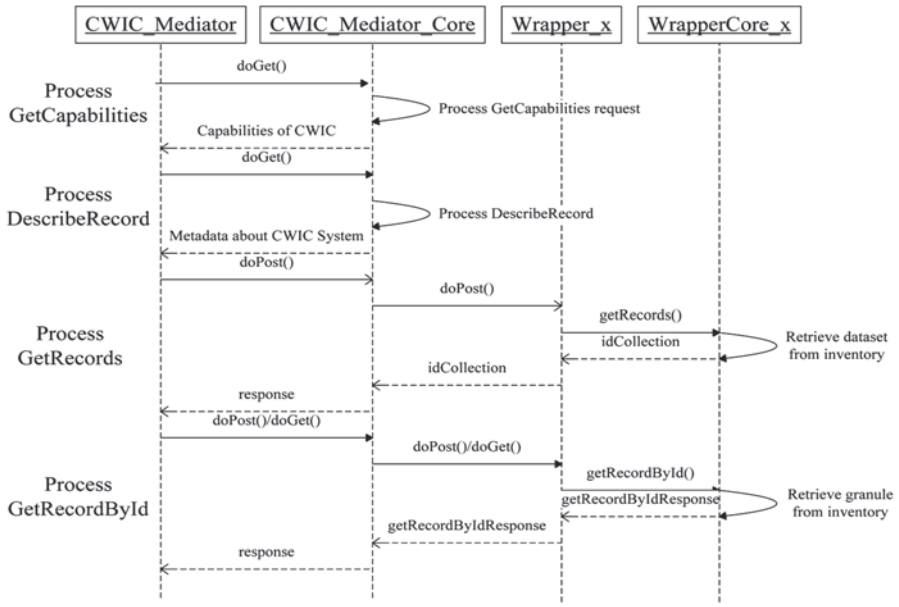

Fig. 3: Interaction between major modules of CWIC.

CWIC is supposed to provide more enriched information in its query response. For some data catalogues with overloaded information, the original response contains so much information but some of them cannot find matched elements in global metadata schema. But for some data catalogues, the query response just includes very limited information. To fix the issues of overloaded information, CWIC creates a mapping table to put more information from native metadata to global metadata. For a native response with limited information, CWIC imports pre-defined attributes or a priori knowledge for some mandatory elements such as dataset identifier and point of contact in the global metadata model.

4.3 *CWIC servable Datasets*

The entire federated catalogue of CWIC has registered their dataset information into the Global Climate Master Directory (GCMD) (GCMD 2012). Each dataset will be assigned to an individual Directory Interchange Format (DIF) Entry ID and label the dataset with CWIC attributes. The DIF Entry ID is used as the dataset identifier in the GetRecords and GetRecordById query of CWIC.

By interacting with the GCMD CSW interface, data users could get the collection identifier based on free text search and other search options provided by GCMD. Besides interacting with GCMD, data user can also retrieve all servable collection identifiers from CWIC capabilities document by sending GetCapabilities request. Once getting the collection identifier, data users and clients could build the request payload by specifying other temporal-spatial query parameters and post the GetRecords request to CWIC service. In the back end of CWIC, there is a configuration file which maintains the dataset identifier and its corresponding wrapper URL. The CWIC mediator dispatches the request to a proper wrapper based on this configuration file.

5 Conclusion and Discussion

This paper discusses the design and the implemention of a federated geospatial catalogue system that provides a single standard interface for inventory-level searches of data

archived in major CEOS member agencies. The CWIC system has successfully integrated seven data catalogues: NOAA CLASS, USGS Landsat, NASA ECHO, Brazil INPE, JAXA, NODC/GHRSST and China AOE. Within the CWIC, data user queries and retrieves data from heterogeneous data catalogues through universal interface and data model. Based on CWIC, more and more clients emerge to discover Earth observation data in the standard way from multiple heterogeneous data centers. For example, some representative clients including GMU GeoBrain, USGS LSI portal, NASA CWIC-Start and Canada CCRS. It is expected that more CWIC clients, especially domain-specific ones, will be developed by the EO communities.

Three conclusions are drawn from the design and implementation of CWIC: (1) the mediator-wrapper architecture is suitable for creating a federated catalogue system, (2) OGC CSW can be used as a universal query interface in catalogue federation, and (3) ISO 19115 and ISO 19115-2 effectively act as the standard metadata information model for such a federated geospatial catalogue system. Both the architecture and the approaches proposed in this paper are also applicable to tackle a heterogeneous problem in other communities for resource discovery.

CWIC does not adopt the caching or harvesting of the metadata from remote catalogues to improve the query performance. So currently the query performance highly depends on remote catalogue. As to the search capabilities, in order to present a universal query interface, CWIC only kept the lowest common queryables, which leads to the problem that some useful query parameters in native catalogue are unsupported in CWIC. In the current version of CWIC, the common queryables (spatial extent, temporal range, and dataset identifier) are sufficient to fulfil the demand from a client.

In the next version 2.0 of the CWIC system, more catalogue systems of CEOS member will be integrated into CWIC system. Moreover, CEOC WGISS query language will be implemented to provide more flexible query capabilities. Besides the common queryables, more native queryables will be supported in future version.

Acknowledgements

This work is supported mainly by a grant from NOAA. The authors would like to thank the CEOS WGISS CWIC task team as well as scientists of individual CEOS member agencies for their review and comments on the CEOS architectures, information models, and implementation details, and for their testing of the CWIC version 1. The authors would especially acknowledge the contributions from teams of the current seven CWIC member catalogue systems and the NASA GCMD/IDN team during the implementation of CWIC version 1.

References

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.

Baru, C., Gupta, A., Ludäscher, B., Marciano, R., Papakonstantinou, Y., Velikhov, P. & Chu, V., 1999: XML-Based information mediation with MIX. – ACM SIGMOD International Conference on Management of Data: 597–599, Philadelphia, Pennsylvania, USA.

Corcho, O., Gomez-perez, A., Leger, A., Rey, C. & Toumani, F., 2003: An Ontology-Based Mediation Architecture for E-Commerce Applications. – Proceedings of Intelligent Information Systems: 477–486, Zakopane, Poland.

Di, L., Bai, Y., Shao, Y., Wang, H., Kang, L., Warnock, A., Yapur, M. & Enloe, Y., 2012: A Global Earth Observation Data Discovery System – the CEOS WGISS Integrated Catalog. – IEEE Geoscience and Remote Sensing Society, in press.

ECHO, 2012: EOS Metadata Clearinghouse (ECHO), http://earthdata.nasa.gov/about-eosdis/system-description/eos-metadata-clearing-house-echo (10.6.2012).

Feng, L., Wang, C., Li, C. & Deng, Z., 2011: Catalogue system integration of China EO data for CWIC system. – **19th** International Conference on Geoinformatics: 1–4, Shanghai, China.

GCMD, 2012: Earth Science data and services directory: Global Change Master Directory. – http://gcmd.nasa.gov/, NASA (30.8.2012).

Hidalgo, J., Pan, A., Álvarez, M. & Guerrero, J., 2006: Efficiently Updating Cost Repository Values for Query Optimization on Web Data Sources in a Mediator/Wrapper Environment. – Lecture Notes in Computer Science: 1–12.

ISO, 2003: ISO 19115: Geographic information – Metadata. – http://www.iso.org/iso/iso_cata-

logue/catalogue_tc/catalogue_detail.htm?csnumber=26020 (30.8.2012).

JAXA, 2012: Observation/Research Result Databases. – http://www.jaxa.jp/archives/db/index_e.html, Japan Aerospace Exploration Agency (30.8.2012).

Lin, H., Risch, T. & Katchaounov, T., 2000: Object-oriented mediator queries to XML data. – **First** International Conference on Web Information Systems Engineering: 39–46, Hong Kong, China.

Melnik, S., Garcia-Molina, H. & Paepcke, A., 2000: A mediation infrastructure for digital library services. – **5th** ACM conference on Digital libraries: 123–132, San Antonio, Texas, USA.

Miller, L. & Nusser, S., 2003: An Infrastructure for Supporting Spatial Data Integration. – Federal Committee on Statistical Methodology Conference: Arlington, Virginia, USA.

Sahuguet, A. & Azavant, F., 2001: Building intelligent Web applications using light weight wrappers. – Data & Knowledge Engineering **36** (3): 283–316.

Song, J., Yoo, S., Park, C., Choi, D. & Lee., Y., 2006: The Design of a Grid-enabled Information Integration System Based on Mediator/Wrapper Architectures. – 2006 International Conference on Grid Computing & Applications: 114–120, Las Vegas, Nevada, USA.

Stoimenov, L., Djordjevic-Kajan, S. & Stojanovic, D., 2000: Integration of GIS data sources over the Internet using mediator and wrapper technology. – **10th** Mediterranean Electrotechnical Conference: 334–336, Lemesos, Cyprus.

Voges, U. & Senkler, K., 2007: OpenGIS Catalogue Services Specification 2.0.2 – ISO Metadata Application Profile (07-045). – http://www.opengeospatial.org/standards/cat, OpenGIS (10.6.2012).

Vretanos, P., 2010: OpenGIS Filter Encoding 2.0 Encoding Standards (OGC 09-026r1 and ISO/DIS 19143). – http://www.opengeospatial.org/standards/filter, OpenGIS (30.8.2012).

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

WGISS, 2012: WGISS Information Systems. – http://wgiss.ceos.org, CEOS WGISS (10.6.2012).

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

Addresses of the authors:

Dr. Yuanzheng Shao, Dr. Liping Di, Lingjun Kang, Huilin Wang, Center for Spatial Information Science and Systems, George Mason University, 10519 Braddock Road Suite 2900, Fairfax, VA 22032, USA, e-mail: {yshao3}{ldi}{lkang3}{whuilin}@gmu.edu

Dr. Yuqi Bai, Tsinghua University, 1 Qinghuayuan Road, Haidian District, Beijing 100084, China, e-mail: yuqibai@gmail.com

Chao Yang, Wuhan University, 129 Luoyu Road, Wuhan, Hubei 430079, China, e-mail: cyang9@gmu.edu

Manuskript eingereicht: September 2012
Angenommen: Oktober 2012