INFORMATION INTEGRATION BASED ON OPEN GEOSPATIAL DATABASE CONNECTIVITY SPECIFICATION

Yuejin Deng [a, *], Zhongri Tang[a], Yunhua Chen[b], Min Cai [a]

[a] Wuhan University, LIESMARS, 430079 Wuhan, Hubei, China – whudyj2010@gmail.com, standby211@qq.com
[b] SuperMap Software Co., Ltd., HaiDian District, 100096, Beijing, China - chenjunhua@supermap.com

KEY WORDS: OGDC, Data Integration, Call-Level Interface, GIS, Data provider, OGC, Simple Feature specification

ABSTRACT:

Nowadays enormous geospatial data are acquired every day with various sensors. However the geographical information resources are naturally heterogeneous. For example, there are many different types of geographical data, such as databases, data files, maps, further more these data are maintained in different software systems by different enterprises, governments and individuals according to their own application requirements. Integration and interoperation of multi-source, heterogeneous spatial database are important for complex GIS applications, such as emergent response of natural disaster. In this paper we give a kind of call-level specification of Open Geospatial Database Connectivity (OGDC) used in China, which is designed for better data sharing among different GIS software. This specification defines a framework of data providers, spatial vector objects based on OGC simple feature specification and spatial raster objects, ODBC-styled database access objects, and corresponding interfaces. The framework of OGDC consists of database level, data provider level, interface level and application level. Several data providers are developed by some Chinese GIS companies, for example, SuperMap, GeoStar, GeoBean, etc. Oracle spatial data provider is also provided by Wuhan University. So we can do some comprehensive spatial analysis in our applications by these providers using different data sources. For data users, it is unnecessary to know which data source they use. How these providers are integrated to perform data overlay is demonstrated.

1. INTRODUCTION

Geospatial information plays a very important role in making any decisions on major economic, environmental and social issues (Abbas Rajabifard, 2010). Rapid and real-time access to needed geographic data and services is useful to help solve the location based problems, such as global change and sustainable development, terrorist threats, positioning and navigation, etc. People often need to know where they are and how far it is from here to the destination in their daily life. With the advance of technologies and sensor devices, large amount of geographic information can be captured quickly to satisfy the requirements of various applications, even though acquiring geographic data is an expensive process. Geographic information is increasingly used around the world (Agustina B., Alejandra C., 2007). The era of geographic information explosion is coming. However, the capability to discover intelligent knowledge from the existed data to solve complicated problems, e.g. the global disaster prevention, is still weak. It is saying “So many sensors, so little data”. Research in geographic data integration has become one of the major subjects in GIS.

Geographical information resources are naturally heterogeneous due to technical or non-technical inconsistencies, for example, there are many different types of geographical data, such as databases, data files, and maps, further more these data are maintained in different software systems by different enterprises, governments and individuals according to their own application requirement. Geographic data may differ in the aspects of data model, schema, data types and interpretation. The heterogeneity had been classified in three forms (André R. C., 2009):
(1) Syntactic heterogeneity of different ways to encode information, different file formats or access protocols.
(2) Structural heterogeneity of different data models, schema, data type, relationship types and attributes.
(3) Semantic heterogeneity of different interpretations of data due to its context and the users understanding of the world.

The heterogeneous data are usually stored and managed by different geographic information systems in distributed environment for a particular purpose. For example, some spatial datasets ranging from fundamental datasets of cadastre, topography, roads and imagery through locational data of police, fire and points of interests to socio-economic and infrastructure data of demography, valuation, public transport and utilities, are necessary for emergency management purposes. Many of these datasets are managed by different administration units. The diversity of approaches to capture, manage, exchange and share these datasets utilized by these organizations leads to many technical and non-technical inconsistencies and heterogeneity among datasets (Hossein M., et al., 2008). On the other hand, GIS software packages use specific storage formats and data models in the past decades (André R. C., 2009).

Many applications need the capability of geographic data integration, so various data integration approaches have been researched for obtaining solutions of the complicated problems. The developers of GIS systems have made different choices about the best way of integration at different levels. Generally speaking, there are two branches of data integration, without ontology and with ontology.

In computer science, the term “ontology” is introduced as an “explicit specification of a conceptualization”, which means that the concepts and relations of the abstract model of real world have been given explicit names and definitions (Agustina B., Alejandra C., 2003). Because the ontology can be used to describe the semantic of information sources, it has long been

* LIESMARS, Wuhan University, 430079 Wuhan, Hubei, China. Whudyj2010@gmail.com, MSN: whudyj2010@hotmail.com.

used in many ways and domains for semantic heterogeneity problems (Frederico F., et al., 2003; Agustina B., et al., 2007; Gholam R. F., et al., 2008; Agustina B., 2009; Abbas R., 2010). Some advantages by applying ontologies for data integration are addressed as fellow (Agustina B., Alejandra C., 2003): the ontology provides a rich, predefined vocabulary that serves as a stable conceptual interface to the databases and is independent of the database schemas; the knowledge implied by the ontology is sufficiently comprehensive to support translation of all the relevant information sources; the ontology supports consistent management and inconsistencies detection; etc. However, the definition of rich, accurate vocabulary in a specific application domain is more complicated and laborious. The knowledge implied by the ontology or semantic of information needs to be gathered by many experts and to be consistent according to standards, so the management of ontologies is not as easy as geographic information itself.

One of the simplest approaches to exchange geographic information without ontology may be the direct translation of data between two systems. Common specifications or standards about geographic conceptual schema and model are very helpful for data transformation. The Open Geospatial Consortium (OGC) tries to lead the global development, promotion and harmonization of open standards and architectures that enable the integration of spatial and location-based data and services into user applications (Martin Klopfer, 2005). OGC had provided some standards for geographic information conceptual schema, model, and interface such as OGC Reference Model (ORM) (George Percivall, 2008), Simple feature access (John R. H., 2010), as well as Geographic Information Language (GML). Products and services which are compliant to OGC's open interface specifications enable users to freely exchange and make use of spatial information, applications and services across networks, different platforms and products.

Federated database, which provides an integrated data access, is another way to integrate multiple autonomous databases. McLeod and Heimbigner (1985) firstly defined a Federated Database Architecture (FDA). Sheth and Larson (1990) defined a federated database as a collection of cooperating component systems. Some surveyors investigated the integration of heterogeneous geographic datasets which are linked to a federated database (Matthias B., et al., 2007).

Recently based on service-oriented architecture, loosely coupled web service is a new approach to information integration in a distributed computing environment. Web services are self-contained, self-describing, modular applications that can be published, located, and invoked across the Web. The critical issue in the interoperation of web services is the adoption of a set of standard protocols that consists of WSDL (Web Services Description Language), SOAP (Simple Object Access Protocol), and UDDI (Universal Description, Discovery, and Integration) (Gholam R. F., et al. , 2008). OGC also produced some specifications for geo-processing services including Web Map Service (WMS), Web Coverage Service (WCS), Web Feature Service (WFS), and Web Processing Service (WPS), etc. More advanced researches and technologies of geo-services have been concerned, such as automatic service discovery, semantic integration in WPS, services aggregation and combination, etc. (E. Klien, et al., 2006; Simone A. L, et al., 2006; Meenakshi N., et al., 2006; André R. C., 2009; M. Lutz, et al., 2009). Jeff Waters, et al. (2009) brought out a concept of global interoperability across net-centric software systems that span a wide spectrum of political, geographical, and organizational units. Nowadays Spatial Data Infrastructures (SDI) have been widely accepted to share geospatial data among organizations (Kiehle, C., et al., 2006; Hossein M., et al., 2008; Abbas Rajabifard, 2010). Geoportals have emerged as mechanisms to support searches for geographic information relevant to specific needs, for instance, Geospatial One-Stop (GOS) which is sponsored by the U.S. Federal Government provide a single portal to geographic information (Michael F. Goodchild, et al., 2007).

In this paper, we designed an Open Geospatial Database Connectivity (OGDC) specification used to access and share heterogeneous geographic data which are manipulated by diverse GIS softwares, such as ArcGIS, GeoStar, SuperMap, and MapGIS. Our work makes it possible to directly access geographic database through universal call-level interface in developing applications, so the applications can integrate geographic data from different sources using a kind of data provider middleware developed by GIS vendors according to OGDC specification. Though similar efforts had done by OGC with Simple Features Implementation Specification for SQL, OLE/COM, and CORBA, OGDC is very like Open Database Connectivity (ODBC) applied to GIS (Robert S., et al., 1995). The OGC geographic models of simple features and spatial reference system are adopted to realize the OGDC drivers. Our approach is without ontology, however, the applications based on OGDC data providers would deal with the semantic heterogeneity issue at top levels. These fundamental OGDC data providers can also be used for web applications in distributed computing environment. This paper is organized as fellow: the next section describes the motivation and requirements of OGDC project, the framework of OGDC specification, detailed models and geographic objects used in OGDC, and the management of OGDC data providers. In section 3, it explains how OGDC data providers are integrated into GIS software and applications to increase the capability of data access. The last part gives brief summary and expected further work.

2. OPEN GEOSPATIAL DATABASE CONNECTIVITY (OGDC) SPECIFICATION

2.1 Requirements of OGDC

In China, several GIS companies developed some geographic information management system according to their own conceptual schema, model, data organization, and management approaches to help local and national government establish geospatial databases. Some frequently used software systems are GeoStar, SuperMap, MapGIS, GeoBeans, and so on. However, the information exchange and sharing among these geospatial databases had become the difficult problems for intensive use to solve real complicated problems, for example, environment protection or hazard reduction. So the national government encouraged GIS enterprises to cooperatively provide solutions of the above problems. Our project is sponsored by the national 863 programme of the Ministry of Science and Technology (MOST), and OGDC is one part of the project of geographic data integration.

From the data user's point of view, the GIS power largely relies on its capability to access and manipulate heterogeneous data. The first step in developing most applications is to search and get desired, accurate information. Data providers should have the ability to provide required data according to understandable, common geographic model. So OGDC is required to provide a framework of direct access to geospatial database without ontology at the bottom level by developing corresponding middleware of data provider, these data providers are independent of operation systems. The conceptual model and interface of OGDC should be standard and open for various applications.

2.2 Architecture of OGDC

The simple model of data integration includes two roles: data provider and data user. Each data provider can access to different types of geospatial data, such as plain files, relational databases, object-oriented databases, web services, etc. and then provides them in a universal interface so that the data user can integrate various data through common interface without knowing where the data are, and how they are formatted in storage. According to this easy-use model, the architecture of OGDC is given as figure 1.

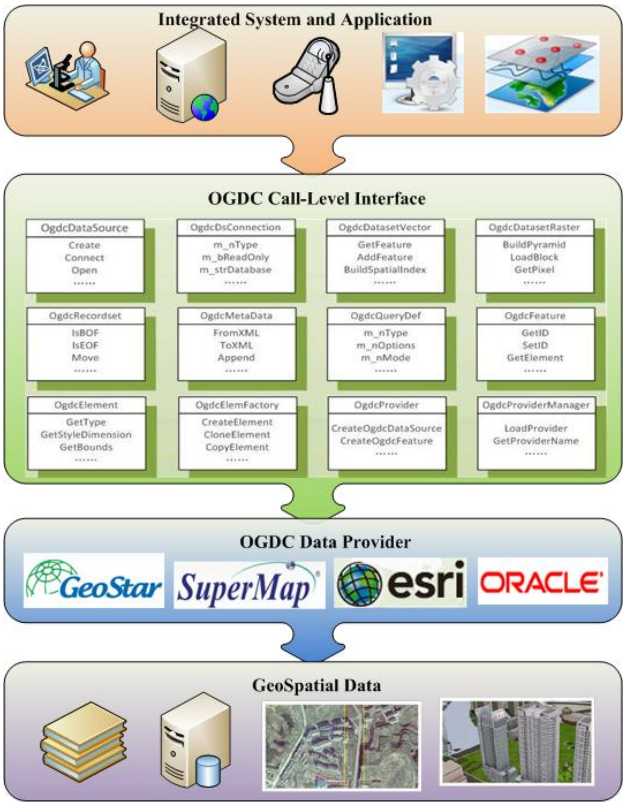

Figure 1. Architecture of OGDC

In this framework of OGDC, four tires named geospatial tire, data provider tire, call-level interface tire, and application tire are proposed. However, the OGDC specification mainly describes the database access model, geographic information objects, methods and parameters of each interface, and management of data providers, which will introduce in the next section. Data providers are required to implement the interface specified by OGDC.

2.3 Geographic Information Model and Objects of OGDC

Data modeling originally depends on how the real world is interpreted. There exist a variety of geographic data models, each model might be more appropriate to a specific problem. In this paper we prefer a common data model proposed by OGC specifications and International Organization for Standardization (ISO) 19100 series standards in order to integrate distributed sources. Our approach is object oriented, so a set of geographic information objects is defined to represent the real world geographic phenomena, as depicted in Fig. 2.

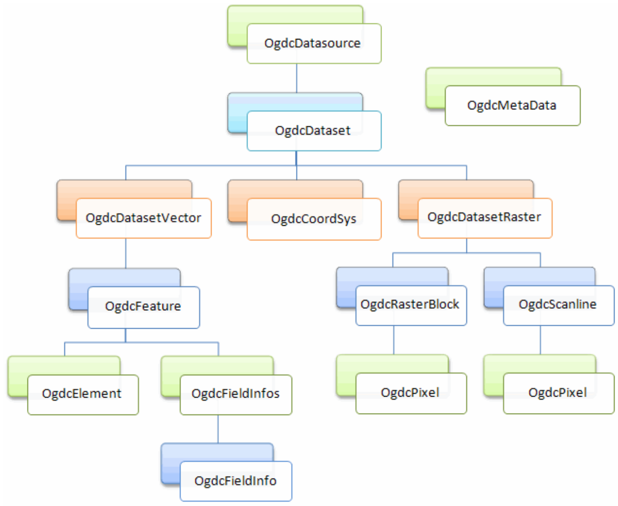

Figure 2. OGDC Data Objects

Generally speaking, the logic flow of access to geospatial database is given as follow: opening the accessible database which can be considered as a kind of data source; querying, reading and returning needed features from the database. OgdcDatasource object is designed as the entrance of access to different data sources, each data source may have several datasets and its corresponding meta data. OgdcDataset object is used to represent the collection of datasets, which may contain two different forms of geographic information:
- Vector data of discrete objects (e.g. seas or streets), represented by OgdcDatasetVector class.
- Raster data of continuous phenomena in geographical space (e.g. temperatures at certain heights), represented by OgdcDatasetRaster class.

Each dataset may be associated with spatial reference system, represented by OgdcCoordSys class. For vector dataset, OgdcFeature that consists of geometry elements and optional attributes is retrieved, and for raster datasets, OgdcRasterBlock or OgdcRasterScanline objects are retrieved. Any query results of datasets is described by OgdcRecordset class.

It is noteworthy that, some concepts of OGDC come from ODBC API and the OGC simple features specification. For example, OgdcDatasource object contains the interface to operate database including operations to open, connect, create, close database; to change database connection information; to modify the state of database; to begin, commit, or rollback transactions; to get or set database information (name, description, version, type, and so on) or metadata; to manipulate datasets; to execute SQL sentence; and to drop data source, as showed in Fig. 3. This object is designed to have methods of data source operation as many as possible, for some particular data providers, it's not necessary to implement all these interfaces, but some operations are mandatory. The meaning of each operation listed in Fig.3 is apparent by its literal name, and not interpret here.

Figure 3. Operations of OgdcDataSource

According to the OGC conceptual model, feature is one of the most important concepts, which is defined as the abstraction of real world phenomena. A feature, entity or instance, may have geometry element which describes its location property on the earth, optional time property, or some thematic properties. Geometry property which is one of the main parts of a feature may include point, curve, surface, and geometryCollection elements. OGDC contains all OGC simple feature geometry (John R. H., 2010), and uses some extended, concrete geometry objects such as arc, circle, ellipse, and so on, all these aren't included by OGC simple feature geometry model. The geometry elements supported by OGDC are listed as Fig.4.

OGDC defined a complete object and interface of access to heterogeneous geographic information, but it's impossible to describe them all in the strength of this paper. More information refers to other work of the authors.

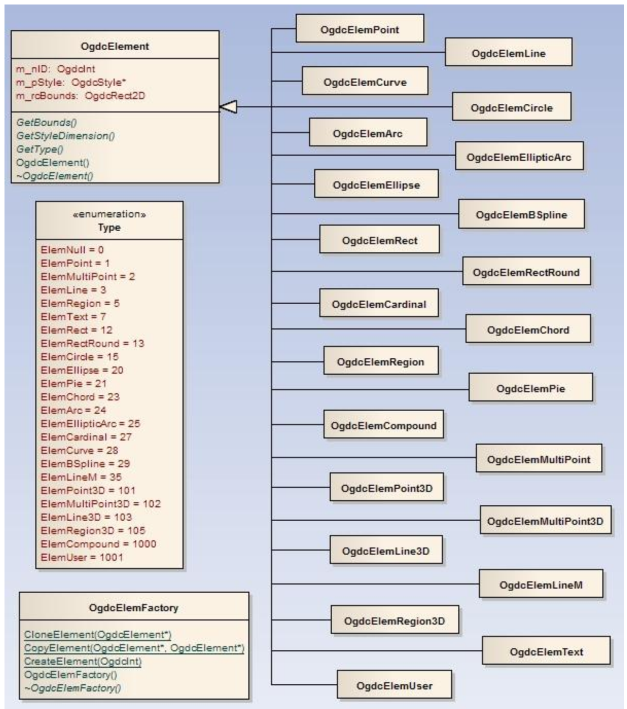

Figure 4. Geometry Elements of Features in OGDC

2.4 Management of Data Providers

Each data provider should be developed and registered according to the OGDC specification so that applications can automatically discover it. Once it is done, the data providers are managed by the operation system, and the data users can access to needed geographic data through OGDC interface.

Some data providers are developed and listed in table 1.

No.	Data provider name	Developer	Status
1	SDB Plus Engine	SuperMap Software Co., Ltd.	In use
2	SQL Server Plus		In use
3	Universal Database		In use
4	Micosoft Access Database		In use
5	Image Plugin		In use
6	GeoAccess Engine	Wuda Geoinformatics Co., Ltd.	In use
7	GeoOracleBlob		In use
8	GeoOracleSpatial		In use
9	GeoSQLServer		In use
10	GeoGlobe Engine		In use
11	GeoOracleRaster	Wuhan University	developing
12	SDX+ for Oracle	Institute Of RS, CAS.	In use
13	SDX+ for sdbplus		In use
14	SDE Data Engine	Beijing easyMap	In use
15	FME Driver	Beijing antu	developing
16	BeyondDB Driver	Institute Of GSNRR, CAS.	In use

Table 1. Some OGDC Data Providers

3. DATA INTEGRATION BASED ON OGDC

3.1 Visualization of Integrated Data

To test the validation of different OGDC data providers, we developed a tool named "OGDC viewer" to visualize the geographic data from diverse sources, as shown in Fig.5.

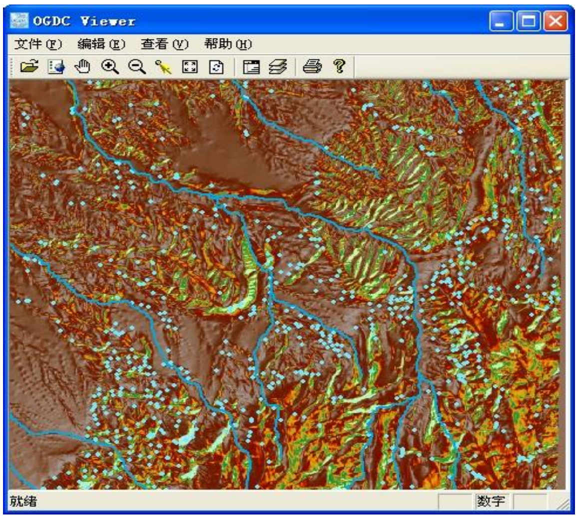

Figure 5. A tool to view geographic data – OGDC viewer

3.2 GIS Operations with Data from Different Sources

Since OGDC data providers give an approach of access to diverse data, they can be embedded into GIS softwares or other applications to support further information integration with ontology, to increase the ability of geo-processing analysis or network analysis with data of different sources. They also can be used as the fundamental components for developing geo-services in a distributed computing environment.

As an example, we plug OGDC data providers in GeoStar software which is developed by Wuda Geoinformatics Co., Ltd., China and is widely used in projects of government or enterprises. These providers can be considered as different sources of geographic data in GeoStar software, then GeoStar software integrates these data for GIS operations. For example, network analysis is performed with data from SuperMap and GeoStar providers, as shown in Fig. 6.

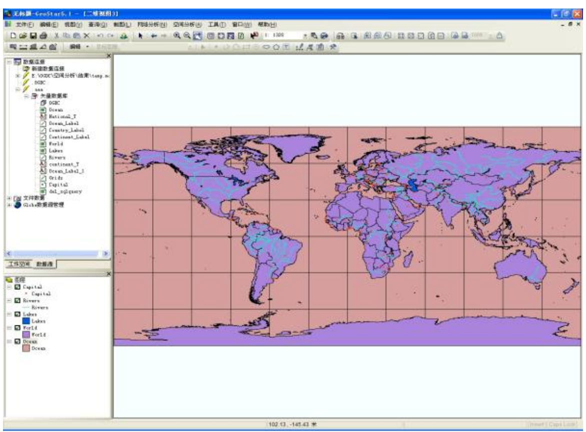

Figure 6. Embedded OGDC Data Providers in GeoStar Software

4. CONCLUSIONS

Nowadays large amount of geographic information is gathered every day over the world. Information integration is an inevitable, critical issue for many complicated applications because of different understandings of real geographic phenomenon. Although many researches focus on the methods of data integration, optimal solutions to this laborious problem are still on the way. In this paper, we showed some aspects of the OGDC specification, such as OGDC framework, geographic data objects, and the management of data providers, and then give some examples of visualization of integrated data, GIS operations with data from diverse sources. OGDC is mainly used in data access with standard interface from various sources, however, the ability of semantic integration of OGDC is weak, further research about ontology and geo-services using OGDC data providers could be achieved.

References

Abbas Rajabifard, 2010. Critical Issues in Global Geographic Information Management with a Detailed Focused on Data Integration and Interoperability of Systems and Data. Scoping Paper for the 2nd Preparatory Meeting of the Proposed UN Committee on Global Geographic Information Management, New York, USA, 10-11, May 2010.

Agustina Buccella, Alejandra Cechich, 2003. An Ontology Approach to Data Integration. Journal of Computer Science & Technology (JCS&T), 3(2), pp.62-68.

Agustina Buccella, Alejandra Cechich, 2007. Towards Integration of Geographic Information Systems. Electronic Notes in Theoretical Computer Science, 168(SPEC.ISS.), pp. 45–59.

Agustina Buccella, Alejandra Cechich, Pablo Fillottrani, 2009. Ontology-driven Geographic Information Integration: A Survey of Current Approaches. Computers & Geosciences, 35 (4) pp.710–723.

André Rocha Coimbra, 2009. Geographic Data Integration to Support Web GIS Development. Proceedings of the International Conference on Management of Emergent Digital EcoSystems, MEDES 2009, Lyon, France, pp.493-497.

E. Klien, M. Lutz and W. Kuhn, 2006. Ontology-based Discovery of Geographic Information Services — An application in disaster management. Computers, Environment and Urban Systems, 30(1), pp. 102-123.

Frederico Fonseca, Clodoveu Davis, Gilberto Câmara,2003. Bridging Ontologies and Conceptual Schemas in Geographic Information Integration. Geoinformatica,7(4),pp.355-378.

George Percivall, 2008. OGC Reference Model (ORM) . http://www.opengeospatial.org/standards/orm. (accessed 6 Sep. 2010).

Gholam Reza Fallahi, Andrew U. Frank, Mohammad Saadi Mesgari, Abbas Rajabifard, 2008. An Ontological Structure for Semantic Interoperability of GIS and Environmental Modelling. International Journal of Applied Earth Observation and Geoinformation, 10 (3), pp. 342–357.

Hossein Mohammadi,Abbas Rajabifard, Ian Williamson,2008. Spatial Data Integratability and Interoperability in the Context *of SDI. Lecture Notes in Geoinformation and Cartography, The European Information Society, pp.401-413.*

Jeff Waters, Brenda J. Powers, Marion G. Ceruti, 2009. Global Interoperability Using Semantics, Standards, Science and Technology (GIS^{3}T). Computer Standards & Interfaces 31 (6), pp. 1158 –1166.

John R. H., 2010. Implementation Standard for Geographic information – Simple Feature Access - Part 1: Common architecture. http://www.opengeospatial.org/standards/sfa (accessed 6 Sep. 2010).

Kiehle, C., Greve, K., Heier, C., 2006. Standardized Geoprocessing - Taking Spatial Data Infrastructures one step further. Proceedings of the 9th AGILE International Conference on Geographic Information Science, Visegrád, Hungary, pp. 273-282.

Martin Klopfer, 2005. Interoperability & Open Architectures: An Analysis of Existing Standardisation Processes & Procedures, Open Geospatial Consortium document: 05-049r1. http://www.opengeospatial.org/pressroom/papers.

Matthias Butenuth, Guido v. Gösseln, Michael Tiedge, Christian Heipke, Udo Lipeck, Monika Sester, 2007. Integration of heterogeneous geospatial data in a federated database. ISPRS Journal of Photogrammetry & Remote Sensing, 62 (2007), pp. 328 –346.

McLeod, Heimbigner, 1985. A Federated Architecture for Information Management. ACM Transactions on Information Systems, 3(3), pp. 253–278.

Meenakshi Nagarajan, Kunal Verma, Amit P. Sheth, John Miller, Jon Lathem,2006. Semantic Interoperability of Web Services - Challenges and Experiences. IEEE International Conference on Web Services (ICWS'06), pp. 373-382.

Michael F. Goodchild, Pinde Fu, and Paul Rich, 2007. Sharing Geographic Information: An Assessment of the Geospatial One-Stop. Annals of the Association of American Geographers, 97(2), pp. 250–266.

M. Lutz, J. Sprado, E. Klien, C. Schubert and I. Christ, 2009. Overcoming Semantic Heterogeneity in Spatial Data Infrastructures. Computers & Geosciences, 35(4), pp. 739-752.

Robert Signore, Michael O. Stegman, John Creamer, 1995. The ODBC Solution: Open Database Connectivity in Distributed Environments, McGraw-Hill, Inc., New York, NY.

Sheth, Larson, 1990. Federated Database Systems for Managing Distributed, Heterogenous, and Autonomous Databases. ACM Computing Surveys, 22(3), pp. 183–236.

Simone A. Ludwig and S.M.S. Reyhani, 2006. Semantic Approach to Service Discovery in a Grid Environment. Web Semantics: Science, Services and Agents on the World Wide Web, 4(1), pp.1-13.

Acknowledgements

This paper is supported by the project named Heterogeneous GIS Data Interoperation Technology under Grid Computing Environment, which was sponsored by the national 863 programme of the Ministry of Science and Technology of the People's Republic of China (Project No. 2007AA120501). The authors also appreciate the softwares from SuperMap Software Co., Ltd., and Wuda Geoinformatics Co., Ltd. in our experiment.