Research of Service-Oriented Data Sharing Standard for Spatial Information System

Yi Zeng and Liyun Zhang
College of Information Science and Technology, Beijing Forestry University, Beijing, China

Abstract: The research of spatial data specification and description mechanism is very important for spatial information sharing and collaboration, especially for spatial information system. Most of the studies focus on the composition of the content and implementation of specific standards, but not aims the procedure in data integration. This study is based on the data service architecture in spatial information system and analysis the relevant data specification and standards. We study the resource description mechanism and metadata strategy and summarize the relations between spatial resources description and corresponding standards. Compared with other methods, this method can expand the spatial data standard combined with operational features of the data in the grid. The study gives the application in spatial information grid platform for case study, discusses the results in detail and establishes the basis for data sharing and service coordination.

Key words: Spatial data standards, spatial data service, service-oriented data sharing

INTRODUCTION

Complex spatial data description is often accompanied all the procedures in heterogeneous spatial information collaboration and sharing services. The data submit, request and the intermediate state parameters in data processing all need to consider how to give the data integral and accurate description. Although the data standards encapsulate the differences of spatial information heterogeneity, but the attributes of the data structure itself are completely preserved, such as the satellite parameters, imaging parameters, reference systems, data types (Zhang and Gong, 2000).

Therefore, the study of spatial data specification and description mechanism is the prerequisites for spatial information sharing and services collaboration, especially in service-oriented environment. The system can have good extensibility only by applying service-oriented data resource description standards.

Currently, most of the studies about spatial data description specifications focus on the content composition and implementation methods. Considering the specific needs in spatial data service, this study takes the specific processes and operational characteristics of grid data service into research. In the following sections, the study analyzes a variety of mainstream spatial metadata of international standards first.

During the process of analysis, we take two aspects into consideration, the spatial information metadata specification and the semantic relationships, from which we explore to develop a mechanism of multi-level spatial data standard and description system. Importantly, we establish a standardized and universal descriptive model of the generic data resources and take a case study in spatial information grid test bed.

ANALYSES AND COMPOSITION OF SPATIAL METADATA STAND

Description mechanism for spatial data resources is closely related to the characteristics of semantic relationships for spatial data and the spatial metadata specification. Metadata strategy is the basis of spatial information query and interaction and the semantic analysis is the key to intelligent retrieval and exact match. Thus, both of them are closely related to the building and the adaptability of the spatial data service.

Core metadata standard of spatial data: The geospatial spatial metadata is the description of data process, operating data process and the structure and meaning of application (Wang et al., 2005). With the development of information technology, the metadata has been extended to be an indispensable powerful tool and method in the whole information processing flow. Metadata is an important factor to improve the utilization of geospatial information. At the same time, it is closely related to the quality and adaptability of the space data service (Feng and Xiao, 2001). There are many international institutions participating in the spatial metadata standards development and the related researches develop in endless. Table 1 lists some common spatial metadata standards.

Corresponding Author: Yi Zeng, Faculty of Information Science and Technology, Beijing Forestry University, Beijing, 100083, China
Currently, the more commonly used and influential spatial metadata standards are CSDGM (Content Standard for Digital Geospatial Metadata, CSDGM) developed by the U.S. Federal Geographic Data Committee and the ISO/TC211 developed by International Organization for Standardization on the basis of CSDGM. They have provided and described various aspects of the spatial data in detail. If all spatial data sources were defined in accordance with the above criteria, then the retrieval and exchange of spatial data would be a very simple job.

Unfortunately, the contents and form of organization of the metadata tends to be difficult to achieve consistent as the metadata is possessed by different organizations most of the time. At the same time, spatial data in different areas are difficult to formulate a unified metadata standard because of the complexity of itself. Therefore, in order to achieve multi-source distributed spatial data sharing in the grid environment, we must establish a reasonable spatial information metadata mechanism and an exchange strategy.

Our research is mainly on the basis of CSDGM and ISO/TC211. Besides, the ISO19115 standard is closely related to data query, access and exchange in SIG data service node. Thus, it becomes one of the most important reference standards of SIG spatial data grid description mechanism.

**Analysis of spatial data query standards:** Query standard is the key to data grid resource description specification. As a query standard on spatial data, the OGC Filter specification developed by the Z39.50 protocol OGC organizations is more sanctioned than other standards. The Z39.50 protocol is the standard protocol for network information retrieval and the original specification was developed by ANSI (He, 2002). First of all, the Z39.50 protocol has a wealth of information resources sharing interface specification, interoperability and rigorous architecture. Second, it can return query results in standardized query syntax and standardized format. Through the connection between client and server-side, it can send format inquiries to return the search results (Miao et al., 2008). Comparatively, OGC Filter is more suitable for defining the description of spatial data query conditions.

It is developed as query predicate in the system in 2005, describing the XML coding of the OGC CQL (Common Catalogue Query Language). The OGC Filter defines the structure description of certain behavior and operation in OGC, in which it can identify the OGC entity object or a subset of an example according to the geographical spatial data attribute values. Currently, many OGC specifications based on web will use Filter in XML as the standard for geographic data representation, especially WFS, Gazetteer and WRS. It is easy to verify, analyze the language of data source based on OGC Filter specification or convert it to SQL by using XML tools. For example, the language of data source can be converted to where clause of the SQL SELECT sentences to obtain data in SQL-relational database. Similarly, it also can convert to X Path or X Pointer to obtain data in the XML documents.

There are two key points about design of query standard. On one hand, the final standard must have loosely coupled definition of elements. As a result, the description of conditional information it supports can adapt all kinds of heterogeneous data sources and match the query conditions thoroughly. On the other hand, the definition of relational operators must meet the comparisons between data and pattern matching as much as possible. Comparing the two key points over, the OGC Filter specification is more suitable for describing request information of spatial data than others. Therefore, taking the demands for data service in grid into account, this study formulates SIG spatial data query specification, having the adaptability of heterogeneous data based on the OGC Filter. The specific elements are showed in Fig. 1.

![Image](image_url)
Fig. 1: XML definition of SIG data query standard

Currently, the query metadata standard established in SIG data service can meet the majority of cases for the operation of raster data and the process of vector data file-level.

**Semantic technology for improving data service:**

Semantics and ontology technology are introduced in the field of spatial information service. The main purpose is to achieve intelligence and accuracy in the process of multi-source spatial data service. Furthermore, the method can solve series problems caused by semantic relationships. Such as, spatial data service cannot be accurately selected; it relies too much on human interaction during the spatial data service; the matching mistakes in keywords query, etc. The main functions of ontology in spatial data service can be summarized as the following aspects:

- It is beneficial to heterogeneous data integration. Ontology is the common conceptual knowledge about a particular field. It can solve the heterogeneity problems in structure, syntax, systems, semantic for spatial data by building semantic relationships between ontology and itself and that between ontology and data (Jiang *et al.*, 2003). As a result, it can shield heterogeneity of data, providing users with integrate, semantically consistent space data service. For example, we can establish the standard definition of the space entity, the geometric properties and spatial attribute data by building the conceptual model for geospatial information service field and basing on ontology technology. In addition, we can describe the domain concepts and relationships between themselves, so that the computer can understand these descriptions and achieve the spatial information sharing and semantic interoperability.
- It can provide semantics-based data retrieval technology. The spatial data service can achieve the semantic retrieval of data by describing the semantic metadata of data (Shum *et al.*, 2000). Currently, semantic-based data retrieval is mainly reflected in three aspects. First of all, the data retrieval based on basic concepts, in detail, we can retrieve this conceptual semantic data according to the basic concept of ontology. Second, complex limit query based on semantic, in detail, retrieve data with the conditions by the concept of ontology and complex limit query of the construction of relation. At the third place, achieving new semantic relationships through reasoning, in a detail way, we can obtain the semantic relationships between the new data by semantic relationships between existing data and inference rules of the semantic relationships.
- It can provide data extraction technology. According to the areas of knowledge about ontology, the technology can organize the disorderly data, providing users a data collection of specific semantic knowledge structure. It is a method that organizes and extracts data according to domain knowledge by using ontology to guide. Then, the extracted data in accordance with domain knowledge are organized to become a meaningful text messages by the definition of the specific semantic templates.

To compare the main standard work for spatial data ontology in the international is the key to establish spatial data ontology. By comparing and analyzing, we can support major international standards work on the concept of ontology level. At the same time, this work ensures that the ontology of spatial data formed from this study has same foundation with other spatial data ontology research projects and achieves data interoperability. We focus on the GCMD ontology, the SWEET ontology and ISO/TC211 standards-based geographic information ontology and the related comparative analysis is shown in Table 2.
Currently, SWEET ontology is a relatively comprehensive and stable geographic information ontology standard. The data representation based on spatial information in SIG is carried out on the basis of the SWWET.

**DATA RESOURCE DESCRIPTION SYSTEM AND METADATA STRAGETY**

Here, discuss which levels of the spatial data service in the grid environment will involve the data resource description mechanism first. Then, we will propose how to select the appropriate standard and reasonable strategy.

**Service links with the relevant standard mapping relationship analysis:** As is mentioned previously, many standardization organizations are committed to the study of the spatial data description standard, then constantly introduce and improve them (Li, 2007). But as formulation and development of the standards is a very difficult job, there is not a metadata standard which can be followed by all spatial data business system. Even if we have one standard like this, we cannot ensure that the eventual content would keep up with the expression as people tend to have different application background when they organize data. Therefore, the core of our research should extract, supplement and improve the existing standards according to the analysis of the spatial data resource description and the specific needs in the grid data model implementation. Besides, we should study the mapping relations of every link with the corresponding data standard on the basis of the specific features in spatial data service. Then, establish spatial information description standards and metadata strategy of support for distributed heterogeneous data sharing.

From the perspective of spatial data service architecture, we analyze that which levels and areas of service will relate to the description of data resources and standards, as is shown in Fig. 2.

We make the data description mechanism focus on three aspects: Data online processing specifications, data exchange standards and protocols and data services published specification. It includes registrations of data resources, data query and retrieval operation standards, return results expressed specifications and so on. The core of the research about spatial data grid service is to support the spatial information of multi-source, heterogeneous, distribution. Only when each level selects resource description method standardized and adapting to the grid environment, can we ensure the data service model support the underlying basis for international standards and also ensure that information interoperability is established with heterogeneous data sources.

Based on the experience of the SIG platform in spatial information grid, we sum up the mapping relations between spatial resource description links and the corresponding standards shown in Table 3.

According to the characteristics of spatial data service processes, we have analyzed the spatial data description specifications and the mapping relations of related standards in specific links from three aspects. The related standards include the main international standards in the field of spatial information, data standards suitable for building grid services and the main reference standards of SIG data service node resources description.

**Description model of spatial data grid resource:** Grid descriptive model of the spatial data resources is a unified spatial data description specification structure. The model is used to restrict and define that data service based on grid expresses and manages spatial information metadata in which way. At the same time, based on the model, we can develop registration, search, access and other operations about spatial data in grid. Besides, the model plays a vital role in the application of data service. In addition, only when the model has perfect function, could it express variety of space metadata information, increase the standardization and versatility of grid data service and last meet the integration and sharing demand about distributed multi-source heterogeneous spatial data sources.

Table 3 illustrates the SIG data service reference standard based on the space data standard extension method. Furthermore, when it comes to the series of links about operation and exchange of data in grid, it is important to consider some special data attributes. Specific analyses are as follows:
Grid data service processes

![Diagram of Grid data service processes]

Fig. 2: Data service processes and mapping relationships of standards

Table 3: Resource description standards of grid data service node

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Major international standards</th>
<th>Applicable to grid data service standards</th>
<th>SIG data service standard node reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online data processing</td>
<td>Spatial metadata standard</td>
<td>CSDGM, GCM, DCMS, SDBCM, GDPL, ISO/TC211</td>
<td>CSDGM, ISO/TC211</td>
<td>ISO/TC211</td>
</tr>
<tr>
<td>specifications</td>
<td>Ontology construction</td>
<td>GCM, SWEET, ISO/TC211</td>
<td>SWEET, ISO/TC211</td>
<td>SWEET</td>
</tr>
<tr>
<td></td>
<td>Data query criteria</td>
<td>Z39.50, OGC Filter</td>
<td>Z39.50, OGC Filter</td>
<td>OGC Filter</td>
</tr>
<tr>
<td></td>
<td>Data acquisition standard</td>
<td>OGC WCS</td>
<td>OGC WCS</td>
<td>OGC WCS</td>
</tr>
<tr>
<td>Data exchange standards</td>
<td>Query results to return</td>
<td>OGC WCS</td>
<td>WCS coverage description</td>
<td>SIG QUERY</td>
</tr>
<tr>
<td>and protocols</td>
<td>and to the standard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Get results and back to</td>
<td>OGC WCS</td>
<td>WCS coverage description</td>
<td>WCS coverage description</td>
</tr>
<tr>
<td></td>
<td>the standard</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data service published</td>
<td>Data service loads</td>
<td>WSDL</td>
<td>WSDL</td>
<td>WSDL</td>
</tr>
<tr>
<td>specification</td>
<td>Data service published</td>
<td>UDDI</td>
<td>UDDI</td>
<td>UDDI</td>
</tr>
</tbody>
</table>

- When it comes to the registration of service, the operating structure of the data source, the organizations and released ways are necessary options to submit and describe.
- Speaking of the data query, the sensor, latitude and longitude and time range information are the described parameters of query conditions paid close attention. The parameter specification should have heterogeneous adaptability.
- For data acquisition, specific demand processing and description of access conditions are the focal points.
Every data acquisition needs to include the identification of data item, referred identification for access, image quality information, etc.

To build ontology for improving the quality of data service, the concept definition, attribute definition and relationship definition are urgent to be regulated after analyzing and comparing the appropriate anthologies.

Based on the analysis mentioned above, the study studies and proposes the grid descriptive model of SIG spatial data and forms the series specification of data resource description for grid service processes. During the process of studying, we not only have considered the needs and characteristics in every links of the data service, but also have resupplied and improved relevant standards. The specific compositions and the relationships are shown in Fig. 3.

Thus, each specific data service is actually a method of conversion between this mode and specific spatial data set. As for further establishment of geo-oriented computing and data service, this model is the basis of all application grid services.

The metadata representation on spatial data description standard has effects on searching for data, conditions matching and conditions of access. So, providing appropriate data representation is an important
ACKNOWLEDGMENTS

Most of research work was conducted in Data Technology Division, Center for Earth Observation and Digital Earth, Chinese Academy of Sciences. The authors take this opportunity to acknowledge the contributions and supports for all the colleagues. This work was supported in part by a grant from the Fundamental Research Funds for the Central Universities (YX2013-16) and China Specialized Research Fund for the Doctoral Program of Higher Education (20110014120010).

REFERENCES