OAI-PMH BASED INTEROPERATION FOR SPATIAL METADATA

Haixia Mao [a,b]

[a] School of Remote Sensing and Information, Wuhan University, Wuhan 430072, China
[b] Advanced Research Centre for Spatial Information Technology, Department of Land Surveying and Geo-Informatics
The Hong Kong Polytechnic University, Hong Kong, China-
hxmao78@sohu.com

KEY WORDS: Spatial Information, Metadata, Interoperation, OAI-PMH

ABSTRACT:

The interoperation of spatial information is an effective way for sharing spatial information, in order to complete any exchanges and integrations of spatial information within and between application systems and organizations. This paper is to apply the metadata interoperability-related research outputs, namely Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), in the field of Digital library to the GIS research field. Then, the technical framework of this metadata interoperation under this protocol has been utilized to the spatial metadata interoperation, which provides a new means to transmitting, sharing, distributing and interoperating spatial information, and as well gives a new usable solution to the promotion and application of this protocol in the field of spatial information science. Furthermore, the relationship of all structured models, functioning descriptions and alternant modes of the prototype of the metadata interoperation for spatial information are also given in this paper. At last, but not least, the current research article has additionally proposed the theoretically applicable technical framework of the novel metadata interoperation for spatial information based on the extended protocol of the metadata interoperation.

1. INTRODUCTION

The basic geographic spatial information has been fully used in all different divisions in the society, such as resources, transportation, environment, land planning, land use and various applicable fields, under the promotion of the technical development of spatial information science. Along with the development of GIS technology and its wide application, each division applying GIS has already accumulated vast amounts of spatial data by the ways of self-establishment of spatial database. How to improve the usage efficiency of spatial data, to eliminate so-called "information islands" and to accelerate the establishment of fundamental facilities for spatial information are difficult problems as well as challenges currently faced in the development of geographic spatial science(Yi et al., 2000). Spatial information interoperation will be an effective way to solve this problem. At the same time, the theoretical skeleton and methods of researches of spatial information interoperation are the hot topics of the current field of spatial information science (Open GIS Forum).

Digital Library has been the newly developed research field in recent years probably to provide the highly effective administration for media resources, such as text, image, audio, video and so on, and to process unified manufacturing and storage for these resources according to specific regulations. Consequently, with a high speed, it is possible to go across different databases for enquiries and searching, and to furnish with the value-added possessed services. Among all, metadata is the key of the engineering construction of Digital Library (Gao et al., 2000). The amount of digital resources collected in libraries would be increased exponentially, whereas the same developmental trend also exists in the field of spatial information science. That is to say, with the non-stop development and improvement of the technologies of spatial data acquisition for land observations, the amount and contents of collected spatial data with different kinds and from various sources are increased tremendously. Therefore, these for the spatial data processing, managing, issuing, sharing and interoperating have led advanced requests in the future.

This paper is to apply the research consequence related to the metadata interoperation, so-called Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), in the field of Digital Library to the research field of GIS. The technical framework of metadata interoperation addressed in this protocol is applied to spatial metadata interoperation, for the implementation of propagating, sharing, distributing and interoperating metadata of spatial information, and for the proposal of a newly usable overall solution to the promotion and application of this protocol in the field of spatial information science (Vckovski, 1999).

A brief introduction of OAI-PMH-related contents and extension will be presented in the consecutive Section to satisfy the requirements of metadata interoperability for spatial information; the novel systematic design and implementation of the framework for the metadata interoperation of spatial information will be discussed in Section 3; an experiment and its corresponding experimental results will be presented in section 4; and finally, Section 5 concludes and remarks the purport of the technical framework for the metadata interoperability of spatial information.

2. OAI-PMH

2.1 OAI-PMH

Open Archive Initiative, referred as OAI in the rest of this paper, is a cooperative organization with the theme in the promotion of exploring, announcing and sharing network resources (OAI organization). The first conception was brought out at the meeting convened in Santa Fe, New Mexico in the late-1999. Soon, the 1.0 version of Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) was published in the January of 2001. The target of the OAI Organization is to implement the

interoperability among different associations for information posting on Web, through collecting this mode by metadata, in order to provide an application-independent framework of the interoperability based on metadata harvesting. It is an important new infrastructure component for supporting the services of distributed network information.

There are two classes of participants in the OAI-PMH framework: Data Providers and Service Providers. The former one receives requests from the latter one to support and announce metadata to service providers according to the format of OAI-PMH, whilst Service Providers may collect metadata from multi data providers and then supply value-added services. An organization having data resources can act as either Data Provider or Service Provider. The technical framework of OAI-PMH is illustrated as in figure 1.

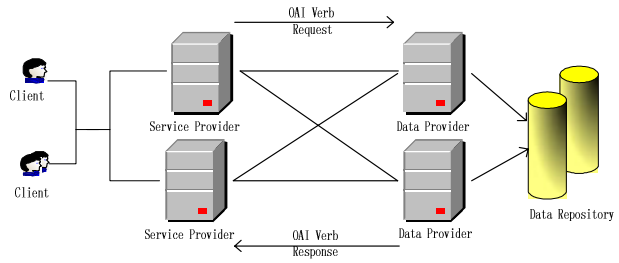


Figure 1. Structure of the OAI-PMH's technical framework

2.2 Advantages of OAI-PMH

OAI-PMH frames the XML Schema for Data Provider to distribute metadata and all Data providers support Dublin Core metadata standard at least. The OAI-PMH is the protocol of application level above HTTP protocol. It specifies the format of request and response between Service Provider and Data Provider. The concrete function of every verb is listed in table 1 below:

Verb	Description
GetRecord	to retrieve an individual metadata record from a repository
Identify	to retrieve information about a repository
List-Identifiers	an abbreviated form of ListRecords, retrieving only headers rather than records
ListMetadata-Formats	to retrieve the metadata formats available from a repository
ListRecords	to harvest records from a repository
ListSets	to retrieve the set structure of a repository, useful for selective harvesting

Table 1. Function list of verb in OAI-PMH

The advantages of OAI-PMH are as follows:
New mode of academic: The framework of OAI leads easier and broader propagation of digitized files. According to the fashion of metadata harvesting, it is possible to collect metadata information from different divisions and data venders, in order to promote information intercommunion across every corner.

Easy to implement: The design of the framework of the OAI's metadata interoperation is based on the principle of simplification. It can be very easy to implement on the networking applicable environment based on HTTP as the main visiting protocol and also to build OAI servers and harvesters in a short time.

Openness: any organizations and individuals can apply the OAI-defined framework to build the data provided server and service provided server conformed to OAI regulation.

Adoption of the openness standard for HTTP and XML: In OAI-PMH, HTTP protocol is utilized as the basic communicating protocol, with its advantage of the supports by all the currently existed Web Servers and browsers; thus, the platform-crossing and compatibility-related problems for OAI can be solved and the frame of OAI can also be simplified. Meanwhile, XML is used in OAI as the mark language of metadata. As HTTP and XML are the openness standard, adopting these two technologies does not only solve the problem of compatibility, but also ensures the openness principle of OAI (Xu, 2003).

OAI-PMH provides a feasible solution for the problem of metadata interoperation on the network. This protocol adopts the two technologies, Internet and metadata, in order to balance the conflict between the reinforcement of functions and the difficulty of implementation. Therefore, with the aid of OAI metadata acquisition, digital resources distributed and posted on the network can be shared and flown widely without the restrictions of system platforms, application procedures, academic disciplines, country boundaries and languages. At the same time, users can acquire their needed digital resources efficiently and conveniently.

2.3 Extension of OAI-PMH

The initial objective of OAI-PMH is to exercise the distributing service of the metadata of the original e-print-archive and corresponding literature. Considering the good openness and extensibility in the protocol, it is feasible to apply OAI-PMH on the theory and practice of metadata interoperation for spatial information. Moreover, considering the prospect of the construction of an actual applicable system, it will provide a novel solution for the interoperation of spatial metadata. However, OAI-PMH was designed to apply to the field of Digital Library initially (Wang et al., 2002); hence, it is essential to extend the contents of OAI-PMH in order to meet the request for the interoperation of spatial metadata. The said extension of the contents is from the three aspects listed as the follows.

(1) **Metadata standard supported by OAI-PMH**: OAI-PMH acquiescently supports Dublin Core Metadata standard in default and uses XML to process marking. For spatial metadata, it's necessary for OAI-PMH to support many other extensively-used standards of spatial metadata, such as FGDC, ISO/TC211 and so on, and as well as to establish the corresponding XML Schema. It is not complex to extend the metadata standards supported by OAI-PMH. That is to say, when Data Provider produces spatial metadata, metadata servers must support these standards and obey the rule of the XML schema. In addition, these standards and schema are the foundation to catalogue and manage spatial metadata. There are many metadata items in the metadata standards of FGDC and ISO/TC211. During designing meta-database and indexing spatial metadata, it can adjust the number of the metadata items based on concrete applicable demands and establish accurate regulations for XML schema.

(2) **Elements in OAI-PMH**: The objective of OAI-PMH is to share and distribute digital information in the field of Digital

Library, so the elements of the protocol are matched the catalogue request of the library. Focusing on the characteristics of the interoperation of spatial metadata, the current article is to modify and adjust some elements in OAI-PMH. The corresponding extension and amended contents are as following:

Concept of "Item" in OAI-PMH (ITEM): ITEM is the basic component of the repository for Data Providers' metadata. It accords with the principle of data indexing and can dynamically produce metadata information of single resource. The content of ITEM will be enriched and extended during the organizing process in spatial metadata in this paper. The partition of every ITEM is adopted based on the structuring principles of characteristics of spatial information. For example, ITEM can be divided according to districts, themes, temporal-spatial range, features of data sources such as vector and gird format and so on.

Concept of "Record" in OAI-PMH (RECORD): The RECORD is dynamically produced by the ITEM and expresses metadata with single format. It can be returned as the byte stream of XML code when answering the request of OAI-PMH. The metadata RECORD discussed in this paper should obey the XML schema of custom spatial metadata.

Concept of "Set" in OAI-PMH (SET): The SET is an optional element in OAI-PMH, which is treated as the top-container in structuring metadata for the interoperation of spatial metadata in the present research. It contains the ITEM and is organized in terms of a leveling structure. For example, spatial information can be organized according to the scale and other information.

(3) **OrderVerb of OAI-PMH**: There are six basic order verbs in OAI-PMH mentioned in Table 1. In this article, their functions are extended in the focal of the characteristics of the interoperation of spatial metadata. As particular relations exist between these order verbs, the parameters of an order verb, *ListIdentifiers*, are discussed as an example of the extension. The original parameters of *ListIdentifiers* include *from*, *until*, *metadataPrefix* and *set*. The extended parameters consisting of the parameter for data quality (data*Q*uality) and the parameter for spatial range control, such as longitude and latitude, are added in order to possess conditional restrictions of quality-referencing information and spatial reference by the searched metadata records. This paper only supports the searching of spatial information via longitude and latitude. The quality of spatial data refers to the regulation of international or specialty standard. It will be classed into different grades during the phase of cataloguing spatial metadata, and these grades will be the conditional values of the parameters of *dataQuality*.

3. IMPLEMENTATION FOR THE INTEROPERATION OF SPATIAL METADATA

3.1 Architectural Design of Meta-database

Data Provider regulated by OAI-PMH place the metadata information into metadata repositories, i.e. meta-databases. Generally speaking, a good architecture of a meta-database assists the distribution and sharing of metadata. MYSQL database is the platform of meta-databases for the interoperation of spatial metadata discussed in this paper. The structure of this typed database includes the table of SET, which records the SET information of the meta-database, the table of ITEM, which records ITEM information in a certain SET, the table of metadata, which stores the metadata records and is made up of exclusive identification of ITEM, the value of SET, the value of metadata, the established time of records and so on. The field of metadata value is adopted with a text type and stores the information of character string of the spatial metadata-recorded XML Schema. This recording obeys and fits the established regulation of the XML Schema.

3.2 The Components of the System

The prototype system is divided into five models according to the system functions, namely spatial data management model, extended OAI-PMH based applicable model, spatial metadata searching model, B-register server model and handle interpreting system. The follows simply introduce the functions of these models (Wang et al., 2003):

① **Spatial Data Management Model**: This model is a multi-divisions-cooperated system for spatial information management, collected with the inputting, editing, processing, searching, displaying and outputting functions for spatial data and spatial metadata. Then, MYSQL and ORACLE databases are used as the servers of both spatial data and spatial metadata. This model plays the role of Data Provider defined in OAI and supports the extended OAI-PMH.

② **Extended OAI-PMH-based Applicable Model**: This model practices the extended OAI-PMH. It intercommunicates with the interface of the spatial data management model described in the above and receives the requests of extended OAI-PMH sent by spatial metadata searching model described in the next paragraph.

③ **Spatial Metadata Searching Model**: This model supports the extended OAI-PMH, acting as Service Provider and Data Harvester. It periodically collects specific spatial metadata information from the spatial management model of any other registered Data Providers. As well, it can provide the value-added service of spatial metadata for the customization of application systems for users.

④ **Extended B-registering Service Model**: This model is adopted the self-registering mode and is administrated the B-address list of Data Providers. From it, spatial metadata searching model can obtain the list of Data Providers, in which it can harvest and collect the metadata ITEM from fitted Data Providers.

⑤ **Handle Interpreting System**: The named system is designed to identify spatial data and provide the interpreting mechanism of identifiers. Spatial data can be located accurately based on spatial metadata harvested by this handle interpreting system.

The intercommunication of the five models of prototype system for interoperation of spatial metadata is showed in figure 2.

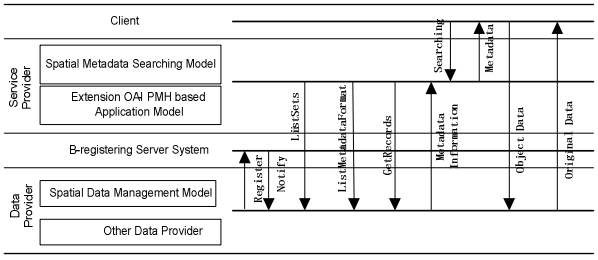


Figure 2. Intercommunication among the models of interoperation of spatial metadata

3.3 Systematic Realization

The practices and technologies of the former two models will be introduced in this section. They are the cores of the architecture of the interoperation.

① **Spatial Data Management Model**: This model is the base of the prototype system discussed in this paper supporting the extensively developed OAI-PMH. It is developed by the means of component and provides an interface related to the interoperation of spatial data and metadata. It is like an engine of spatial database to link with object-relation database.

② **Extension OAI-PMH based Application Model**: This model belonged to the process of the part for Web Server locates between the HTTP server and spatial data management model. It has been developed with the programming language of Perl 5. The six order verbs have been extended and developed based on the version1.1 of HTTP. Furthermore, this model is utilized the HTTP Server of Apache as the Web-announcing Server.

In this research, the exercises of other functional models, such as the implementation of spatial metadata searching model, have been extended for further functional development based on open coding programming. The Java programming language is adopted to extensively develop the OAIArc model of Arc Software. The B-register service model of extended OAI-PMH is practiced through JSP technology. Handle interpreting system of spatial data customizes the interpreting strategy mentioned in this paper by using Handle System for the reference. For a complete functioning, this model is still needed to be further improved and investigated.

4. CASE STUDY

Currently, the novel system of the technical framework for the metadata interoperation discussed in this chapter can be used with the IE Explorer as the procedure of client side to acquire matched and fitted metadata list and finally obtain the matched metadata records of spatial information displayed in XML mode. OAI solely offers the enquiries according to the time poke and set ascription. The present system has been extended in the aspect of the selectively collecting conditions. Similarly being shown in the following figure, searching specifications are classified into five main classes, respectively the publishing date of data collection, metadata-belonged storage, selection of OAI verbs, contents-based enquiry conditions and filtration of results.

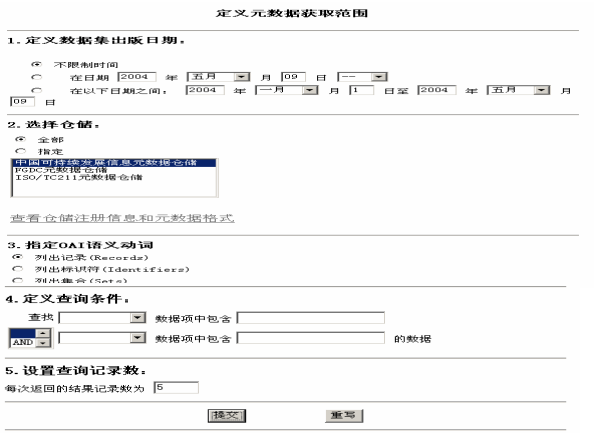
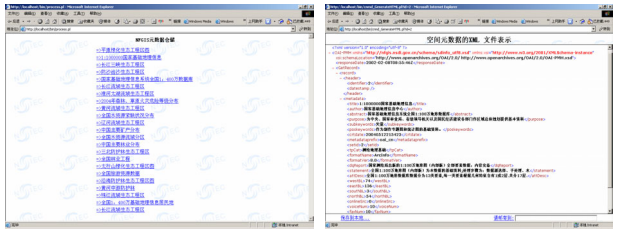


Figure 3. Interface of interoperation of spatial metadata (Interface of metadata harvesting: to define the publishing date of data sets, to select the metadata repository, to appoint an order verb, to define keywords and to set the records counts)

After users input the enquiry conditions of metadata in the enquiry interface, they submit their requests/ enquiries to Service Provider, who then send the OAI requests to Data Provider, according to the enquiry conditions and OAI verbs provided and selected by users. After searching in the data repositories, Data Provider then returns to the metadata interface, as shown in figure 4. According to the chosen metadata in the metadata theme, users then select corresponding hyperlinks. After Data Provider re-send the request to the repositories for collecting metadata (GetRecord), it would then return and display the detailed information of the metadata in IE explore in the form of XML. The displayed interface is categorized into three parts, the first of which is metadata theme, the second of which is the metadata mass presented with the XML form, and the last one is the method for metadata acquisition.



(a) (b)

Figure 4. Spatial metadata searched by OAI-PMH
(Image (a) is metadata list;
Image (b) is the XML expression of metadata)

5. CONCLUSIONS

Depending on its superiority in the field of metadata interoperation, OAI-PMH is the rapid and effective way to share and interoperate metadata with wide authorizations in various research fields. Through the extensively investigation and improvements for many aspects such as the parameters of order verb and metadata format in this research, it has been successfully introduced in the field of spatial information science. This novel concept has assisted to propose and design the novel solution to the interoperation of spatial metadata.

The results of the experiment have proved that this newly developed system is feasible from the aspects of both the speed and the effect of the interoperation of spatial metadata. As a result, it's promising to further improve and apply the OAI-PMH to the field of spatial information science.

Digital Library is an important component of Digital City. Although the characteristic of library metadata is different from that of spatial metadata, if spatial metadata can be produced according to the currently newly proposed OAI-PMH, it will be much easy to interoperate spatial metadata using OAI-PMH and needn't perform extra large amounts of works when integrating the Digital Library and Spatial Information Technology. Therefore, the introduction and examination of OAI-PMH into the field of spatial information do not only gain and import the advantages of OAI-PMH itself to the metadata interoperability for spatial data, but also benefit the integration of Digital Library and spatial information technology. Consequently, it provides not only significant meaning in the research field of spatial

information technologies but also a well further development's future.

REFERENCES

Gao, W., F., Liu, T.J. Huang, 2000, Principle and technology of digital library. Beijing: Tsing Hua University.

OAI organization, http://www.openarchives.org/ (Access on 10th September 2005).

Open GIS Forum, http://www.opengis.org/ (Access on 10th September 2005).

Vckovski, A., 1999, Interoperability and spatial information theory Goodchild, M., M. Egenhofer, R. Fegeas and C. Kottman (eds) Interoperating geographic information systems. Kluwer Academic Publishers, US. Pp. 31-38.

Wang, A.H., M. Zhang, D.Q. Yang, S.W. Tang, 2002, The framework for metadata interoperability based on OAI. Computer Engineering and Application, Vol (1), Pp5-7.

Wang, S.A., M. Wang, M. Zhang, 2003, The design and implementation of a metadata interoperable architecture supporting OAI-PMH. Computer Engineering and Application, Vol (20), Pp168-172.

Xu, R.X., 2003, Initial Research on cross-platform retrieval systems in digital library. Journal of Shanghai Jiao Tong University, Vol (37), Pp 191-194.

Yi, S.Z., Q. Li, J.C. Cheng, 2000, Sharing and interoperability of spatial information. Bulletin of Surveying and Mapping, Vol (8), Pp 17-19.

ACKNOWLEDGEMENTS

This research is supported by the project of Web Assisted Teaching for Geomatics, funded by the Hong Kong Polytechnic University.