Study on Data Management and Sharing Service Based Metadata and Dataset Concept
A Case Study in Environment Sciences and Ecology Area

MA Liguang, CAO Yanrong, HE Jianbang, China P.R.

Key words: metadata, dataset, environmental sciences and ecology, integration, data sharing

SUMMARY

Nowadays, data play an important role in each side of life with the information technology and advanced data acquisition methods developments. It's necessary to manage the accumulated massive datasets and developed spatial database or subject-specific database. Moreover, the integration of the dataset management with web service to realize the dataset sharing online is a big portion of new technologies innovation events.

In this paper, a data management system called EEDMIS (Environmental and Ecology Sciences Data Management Information System) based on metadata and dataset concept with three-layers infrastructure which can used to storage, manage the multi-source, multi-type and spatio-temporal Environmental Sciences and Ecology data was designed and built. Especially, the new ideas which manage the dataset based on metadata elements was proposed and the integration of dataset with standard metadata XML file was published on world wild web portal which realized the dataset sharing service. Lastly, the corresponding merits and practices of the EEDMIS were discussed as well as the prospects of the future work were described.

Study on Data Management and Sharing Service Based Metadata and Dataset Concept
A Case Study in Environment Sciences and Ecology Area

MA Liguang, CAO Yanrong, HE Jianbang, China P.R.

1. INTRODUCTION

The development of information technologies is followed by extensive requirements of the growth of vast databases and the improved data acquisition method in all industries. The emergence of magnanimity data enabled the progress of the data management tool as time goes on, which effectively realized the excessive data storage, management, display, transformation and transmission service. In the future, the dataset and knowledge sharing services based on the developing web technique will become more easily and maturely to implement data exchange.

In this paper, we proposed a new mode which integrated the dataset and metadata concept together to develop a client /server data management system called EEDMIS (Environmental and Ecology Sciences Data Management Information System), it can implement data management, display, upload /download and query functions. Moreover, a browser/server dataset sharing service system based on world wild web was published successively. The system support the relation between dataset and metadata in order to manage the data, and it is used in real environmental sciences and ecology fields. Finally, the dataset sharing and management mechanism was carried into execution successfully.

The article starts with an introduction in Chapter 1, and examines the requirements of the dataset manager system and the sharing service requirements as well as other background information in Chapter 2. The system design and relational technology which we prefer and prospect are explained in detail in Chapter 3.In Chapter 4, each module and function of the system we implemented is described respectively. Lastly, in Chapter 5 some advices and experience are put forward, and the future work is mentioned.

2. RELATION WORKS

This chapter examines the requirements of the EEDMIS and some background information about present work.

2.1 Background

With the time elapsed, we have to face a mount of data in our research work, these data include basic data which used in our study and the research results, it's urgent to storage and manage those data by computer and database from the view of data protection and reuse. Publish the dataset sharing service on web of our fields. All these requirements are in dire

needed of the integration DBMS (Database Management Information System) with the web-based information sharing system.

2.2 Requirements

In order to illustrate the requirements about of the system and show our work the future work, some functions should be considered as follow list:

－Classify the current data with the basic classification standard sequentially
－Storage and management of the dataset function with computer and database technique
－Support the spatial data upload and download real time
－The basic query function of metadata
－Dataset and information sharing service on the web
－Manage the research work and research projects and the related dataset

Additional functions include a spatial data display technique to enhance our research work results and data attributes; a spatial index technique to quickly search the dataset . In particular, the system should support multi-type data and operators between spatial data and non-spatial data with metadata.

2.3 Data

The case study data are from multi-resources and multi-type data which include spatial data and non-spatial data such as vector, grid, multi-media and documents. These data with spatial and temporal attribute summarized in table 1.

Table 1. Multi-Type study data table

DATA TYPE	SPATIAL DATA (Y/N)	TIME SEQUENCE(Y/N)
Shape	Y	N
Coverage	Y	N
CAD File	Y	N
DEM	Y	N
Remote Sensing Image	Y	Y
Documents	N	Y
Picture	N	Y
Media Data	N	Y

2.3.1 Spatial scale

The major spatial data which we deal with include remote sensing image and shapefile format, those data coverage whole China region in spatial scale. The spatial distribution attributes such as municipal administration and drainage area like Yangtze drainage area and drainage area of the Haihe River etc are presented.

2.3.2 Temporal Scale

The study data in temporal scale distributed in a long time periods. From 1980s to now, the data temporal attribute present sequentially. Especially the remote sensing image data, it is multi-spatial and temporal distributed.

2.4 Concept and Method

Some concept and method which adopted in our work and the system implement, the detailed description are below.

2.4.1 Metadata

Metadata is the data about the data. It is the resume of geo-spatial (or non-spatial) data, describing such information as availability, accessibility and transferability of the spatial (or non-spatial) data. A metadata standard can promote the proper use and effective retrieval of the data, facilitate the organization and management of data.

In this paper, we adopt two metadata standards for both the spatial dataset and non-spatial dataset respectively. The first one is science data sharing metadata standard (draft) which used for spatial dataset, and the other one is Dublin Core metadata which is used for the document files and so on. This method allows responsible person to fully describe the dataset, so the end-user can understand the assumptions, limitations and evaluate the dataset's applicability for their intended use.

2.4.2 Dataset

Dataset are defined as the minimum collection of data, it's an uncertainty concept in different research fields. In our study, the dataset represent an integrated data file package. It's may be only one file or several, but the components files belong to one and only one data type. This method has given great convenience to the management of dataset with different temporal-spatial pattern. Figure 1 illustrated the dataset elements.

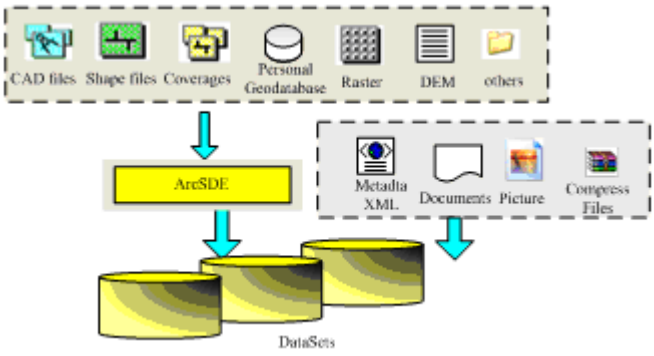

Figure 1. The basic dataset elements sketch

3. DESIGN OF THE EEDMIS

In the preview chapter we have interpreted the constructed management system in our study process. Following, we will describe the structure and the main function of the EEDMIS detailed.

3.1 Structure of the Client/Server system

Based on the requirements described in 2.2 and the initial motives, we design a system that integrates management information system (MIS), office automation system (OA), which called EEDMIS. The function of the EEDMIS developed in this paper is roughly divided into an interface manager, a data display manager, a metadata manager, an upload/download data manager, a query processing manager, a project and working flow manager and a user manager and so on.

The designed three-layers framework based on Client/Server Mode include client, middle control and database server. The whole system framework that have put forward was shown in Figure 2.

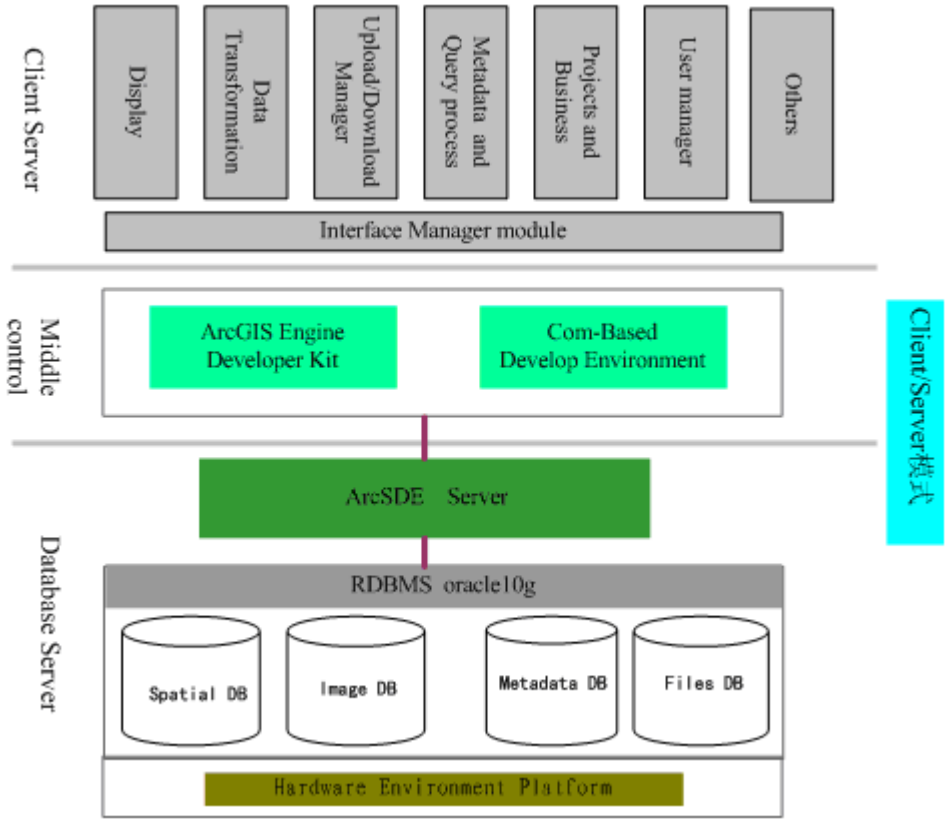

Figure 2.The Framework of the EEDMIS

The topside layer is the system display layer, all user's operation and function interface related to software are included in this part. Friends user interface, commands input window, user's requirement output etc. are implemented in this part. The middle layer is the technique development tool and the intermedium between user interface and database. It's a big bridge connecting user and database closely. The undermost layer is database and data engine for both spatial data and non-spatial dataset.

3.1.1 System Developing Environment

The system developed by Visual Basic Language, integrated the COM-Based technology and ArcGIS Engine Developer Kit. Visual Basic Language has the property of smart, great control management and easy operation property. ArcGIS Engine Developer Kit is an embedded GIS component library for building custom applications which can use Component Object Model (COM) developed by the famous company ESRI.

The application combination of Visual Basic (VB) and ArcGIS Engine (AE) will have a small memory footprint. The multi-type spatial data supported by AE draws much quicker than other windows mapping software. All these achievement is due to COM-Based technology directly on the Microsoft Foundation Class (MFC).

3.1.2 Geodatabase Storage Method

Geodatabase is a data model to manage vector, raster and gridded data and metadata. It provides a general and object-oriented spatial database schema that process common data storage and management framework for different types of data such as:

- attribute tables
- geographic features (Shape file, Coverage, Grid)
- satellite and aerial imagery
- surface modeling data
- survey measurements (CAD file)

The geodatabase is not only defined how data is stored, accessed, and managed, but also it implemented the complex business logic such as: modeling of spatial relationships between data, data validation and long transactions etc. Therefore, based on the benefits of the geodatabase data model, we have to leverage the spatial data to its full potential and maintain a consistent, accurate database with ArcSDE (spatial data engine) support. Furthermore, we will implement the collection of files (documents, reports, papers) in a file system within oracle blob field.

3.2 Structure of the Browser/Server Sharing Service

As the same initial purpose of EEDMIS design, the ideas that manager dataset by metadata which can provide the dataset sharing service on web server is point out. The project implement as one dataset accompany with one metadata (XML) file. The metadata standard has been introduced in 2.4.1 comprehensively. Here, we designed a file exchange folder server and a dataset sharing server based on http and ftp service. User can upload their dataset and metadata file (XML) to exchange files folder by network. After the quality control stage, the dataset was published on the ftp server and the corresponding metadata (XML) file was published on the website to network user. Everyone logging on our website can browse the metadata contents and the abstract of the corresponding dataset. Then, they decided to download or undowload the dataset from ftp server. The metadata page provided the download URL for user who has the sharing authorization. Figure 3 demonstrate the work flow of the dataset web sharing service.

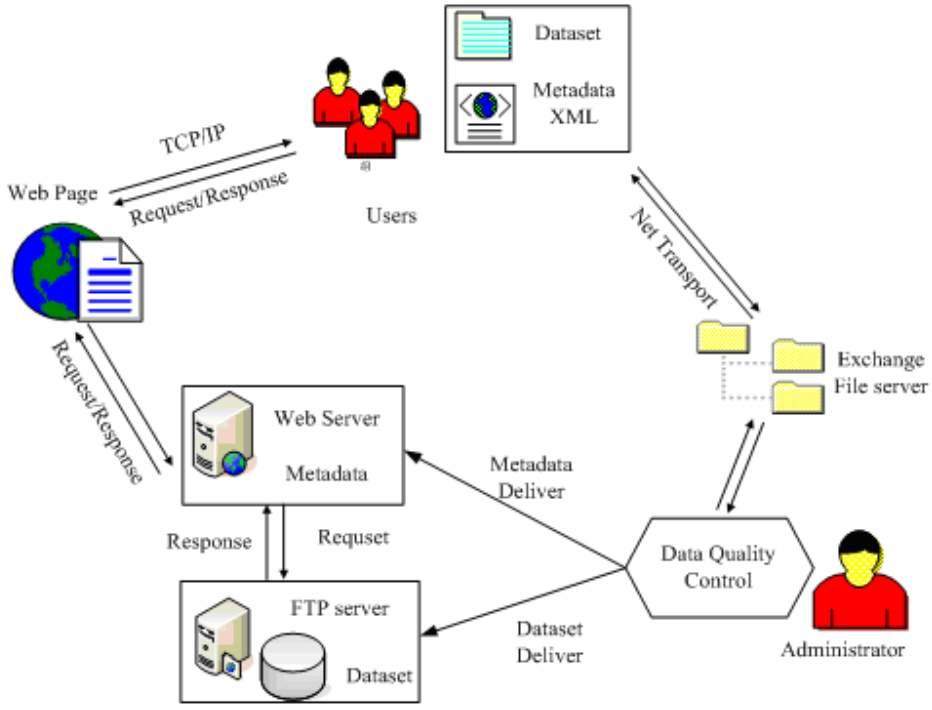

Figure3. Web-based dataset sharing service

3.3 Dataset Catalogue

All dataset will not be detailed present with list follow mode, so we established a catalogue services with the capability to classify these dataset according to a proposed ontology to facilitate finding out the appropriate dataset source and metadata (XML) file with index function.

The catalogue was a multi-hierarchy framework including spatial dimension, temporal dimension and attribute dimension. The classification catalogue is intersected among each other but one classify item is independence. The results of the classification is one dataset belong different classified items. It is necessary for user to find out the dataset quickly from different initial motive with the contribution of dataset catalogue.

4. IMPLEMENTATION OF THE EEDMIS

This Chapter states detailed descriptions of each module of the EEDMIS and the Web-based system.

4.1 Upload/DownLoad Manager of Dataset

The upload/download manger provides the data transmission function between the system and users based on local area network. The data upload and download module performs two functions. Firstly, the user can upload their prepared dataset in their memorizer by the data upload function. With spatial dataset, the ArcSDE for oracle can assist the implement of this work. With non-spatial dataset, the binary blob field provided by oracle database management system was used to store the corresponding files. Secondly, users submits the application form at the operation interface, after the authorization by system administrator, they can download the dataset from the computer server to satisfy their purpose.

4.2 Dataset Transformation Manager

The data we deal with is tempo-spatial attribute, The SDBMS (spatial database management system) based on the geodatabase storage model, multi-type, multi-resource dataset should be converted from current spatial-type, such as from coverage, shape, cad file and DEM etc. to geodatabase table when we upload the data. On the contrary, users can save the spatial dataset as they wanted such as shape file. The dataset transformation manager can convert the non-spatial dataset as compressed binary to storage in the oracle database fields. Besides, this module is necessary for user to revise and convert the metadata (XML) file by user interface dialog, then summits to the database with the dataset.

4.3 Metadata Manage and Query process

Metadata manager module include metadata edit (read and write of the xml file) tool and metadata query function. The user register the metadata standard by the edit tool, then can edit new metadata file using the registered standard (standard *.xsd file) and export the metadata content as XML file for proper dataset. The query processing accepts the SQL Query entered by user from the user interface to examine, analyze and control the SQL query statement, and uses the metadata provided by EEDMIS to process the searching, browse of the spatial dataset and the non-spatial dataset. Figure 4 illustrate the interface of Metadata Edited Tool.

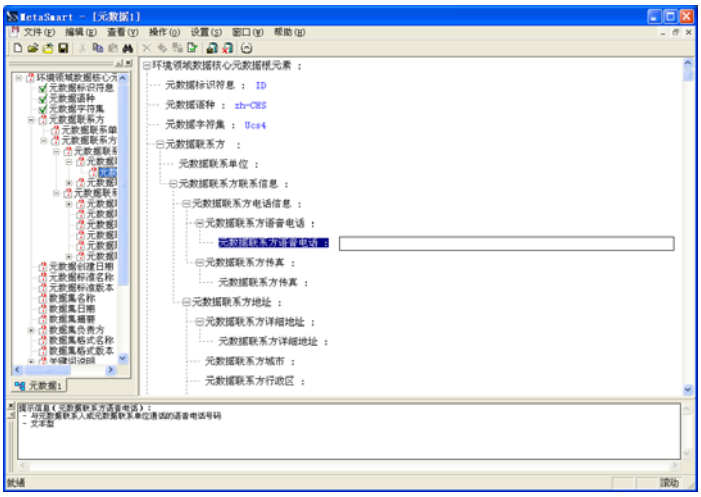

Figure 4 Metadata Edit Tool

4.4 Dataset Display Interface

The dataset display function included both the spatial data display and the attribute data browse. With SDE (Spatial Data Engine), the spatial data like vector data and imagery data are written into cache after read out from SDBMS(Spatial Database Management System), finally shown in map display window. The attribute data was shown in the table format. User can zoom in/zoom out, full map and pan the map which display in map window and identify the objects attributes. The non-spatial dataset was shown by the OLE technique such as web browser and MS documents. Figure 5 show the main interface and the spatial data display window of the EEDMIS.

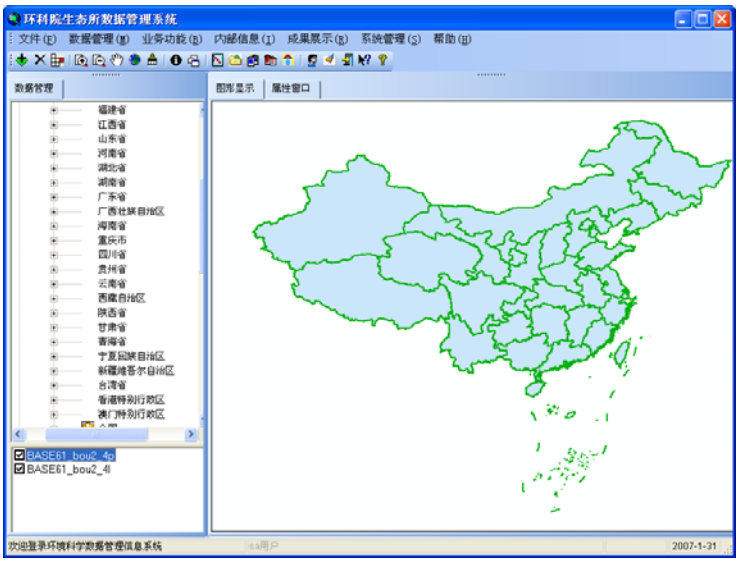

Figure 5 The main interface of EEDMIS

4.5 Metadata and dataset Sharing Service

In our web service system, all sharing services which can provide at current condition have been released on the web, metadata documents describe the dataset content published at web server. A web catalogue portal site has harvest catalogs from a constellation of current web sites and relates to reference data holdings contained at other websites. By the metadata sharing service on web, user knows what kinds of dataset have been provided and whether it meets his (or her) requirement or not. Then they can request for the dataset by summiting brevity application form or by log on identification to download the sharing dataset.

However, up to now, this method just realized the dataset sharing service simply and limitedly, so we have big progress to pursue in our future work to stress our web service robustness and expand our service region etc. .

5. CONCLUSION AND FUTURE WORK

As the world wild web technique and the computer software, hardware are developed, the system we developed is necessary and effective to manage the spatial dataset and non-spatial dataset as well as the metadata web service, the system performance is stably and very helpful to the administrator of the dataset with time saving and simplified the flow of office work that make the data reuse level hige and quickly.

Furthermore, some other pursuits should be considered in our future works. Firstly, the spatial data GIS server will be provided in our web service such as map service. Secondly, it's necessary to integrate the EEDMIS with the web service as a new platform for both spatial data service and non-spatial data service. Lastly, the spatial data model and the spatial analysia function should be developed in the system to strengthen the speciality.

REFERENCES

1. Tschangho John Kim, Metadata for geo-spatial data sharing: A comparative analysis(J) . The Annals of Regional Science, 1999, 33:171－181.
2. Jae-Kwan Yun, Joung-Joon Kim, Dong-Suk Hong, and Ki-Joon Han. Development of an Embedded Spatial MMDBMS for Spatial Mobile Devices. Web and Wireless Geographical Information Systems. 5th International Workshope, W2GIS 2005 Lausanne, Switzerland, December 2005 Proceeding
3. Ming-Hsiang Tsou. Integrating Web-based GIS and image processing tools for environmental monitoring and natural resource management(J). Journal of Geographical Systems, Volume 6, Issue 2, pp. 155－174 , 2004.
4. Hale SS, Miglarese AH, Bradley MP, etc. . Managing troubled data: coastal data partnerships smooth data integration. Environmental Monitoring and Assessment 81:133–148, 2003.

5. Hao Wu, Liguang Ma, Xianghong Hua, XinZhou Wang. GIS-based digital mining management information system: a case in Laozhaiwan gold mine. Geoscience and Remote Sensing Symposium, 2005. IGARSS '05. Proceedings. 2005 IEEE International Publication Volume: 1, page(s): 3 pp. July 2005.

BIOGRAPHICAL NOTES

MA Liguang is currently studying as a Ph.D. in Institute of Remote Sensing Application Chinese Academy of Sciences (IRSA, CAS), he get the Master degree in 2005. His interested research direction is Public Health and Geographical Information System, Spatial Data Mining, Spatial Data Sharing Policy etc.

CONTACTS

Ma Liguang
Institute of Remote Sensing Application, Chinese Academy of Sciences
No.11 Datun Road, Chaoyang District, 100101
Beijing
China
Tel. +86 1064839670
Fax +86 1064839670
Email: maliguang@irsa.ac.cn

CAO Yanrong, HE Jianbang
Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences
No.11 Datun Road, Chaoyang District, 100101, Beijing, China
Email: caoyr@lreis.ac.cn
jbhe@lreis.ac.cn